Bioinformatics, 33(16), 2017, 2573–2574
doi: 10.1093/bioinformatics/btx202
Advance Access Publication Date: 5 April 2017
Applications Note

OXFORD

Genome analysis

EzMap: a simple pipeline for reproducible analysis of the human virome

Patrick Czeczko[1], Steven C. Greenway[2] and A. P. Jason de Koning[3,*]

[1]Bioinformatics Bachelor of Health Sciences Program, [2]Department of Pediatrics, Cardiac Sciences, Biochemistry & Molecular Biology, Alberta Children's Hospital Research Institute, Libin Cardiovascular Institute of Alberta and [3]Department of Biochemistry and Molecular Biology, Department of Medical Genetics, Alberta Children's Hospital Research Institute, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 1N4, Canada

*To whom correspondence should be addressed.
Associate Editor: Inanc Birol

Received on September 27, 2016; revised on March 13, 2017; editorial decision on April 3, 2017; accepted on April 4, 2017

Abstract

Summary: In solid-organ transplant recipients, a delicate balance between immunosuppression and immunocompetence must be achieved, which can be difficult to monitor in real-time. Shotgun sequencing of cell-free DNA (cfDNA) has been recently proposed as a new way to indirectly assess immune function in transplant recipients through analysis of the status of the human virome. To facilitate exploration of the utility of the human virome as an indicator of immune status, and to enable rapid, straightforward analyses by clinicians, we developed a fully automated computational pipeline, EzMap, for performing metagenomic analysis of the human virome. EzMap combines a number of tools to clean, filter, and subtract WGS reads by mapping to a reference human assembly. The relative abundance of each virus present is estimated using a maximum likelihood approach that accounts for genome size, and results are presented with interactive visualizations and taxonomy-based summaries that enable rapid insights. The pipeline is automated to run on both workstations and computing clusters for all steps. EzMap automates an otherwise tedious and time-consuming protocol and aims to facilitate rapid and reproducible insights from cfDNA.
Availability and Implementation: EzMap is freely available at https://github.com/dekoning-lab/ezmap.
Contact: jason.dekoning@ucalgary.ca
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Metagenomic studies have largely focused on characterizing the determinants of bacterial and fungal community structures within microbiomes. Although it is not generally apparent that viruses within a human host have sufficient functional interactions to comprise an ecological community, it has become apparent that high-level snapshots of viral abundances can nevertheless be obtained by metagenomic profiling methods and that they can be indicative of host immune status. In particular, cell-free DNA (cfDNA) viromics techniques have been proposed as one way to perform such surveillance of immune function within individual patients (Burnham *et al.*, 2016; De Vlaminck *et al.*, 2013, 2015; Dinakaran *et al.*, 2014; Ngoi *et al.*, 2016; Rascovan *et al.*, 2016; Young *et al.*, 2015).

To automate exploration of viral activity profiles from cfDNA shotgun sequencing reads, we developed EzMap. EzMap allows the relative abundance of all known viruses in a sample to be determined based solely on DNA sequence. It leverages the power of various publicly available tools and the speed of parallel computation to subtract out human sequences, and then to clean, align, and analyze non-human sequence data. EzMap generates an interactive report in the form of an HTML document containing dynamically generated, interactive figures that enable uncomplicated analysis.

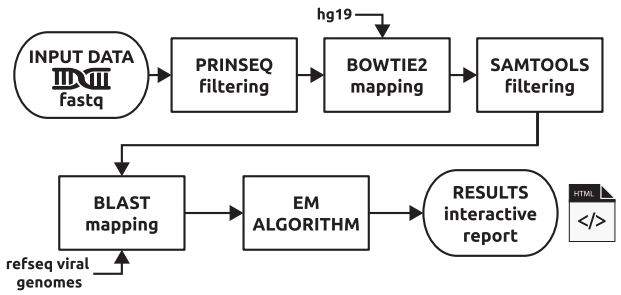

Fig. 1. Flowchart demonstrating the individual steps of the EzMap pipeline. The EM Algorithm step uses the approach of Xia *et al.* (2011), which we implemented independently

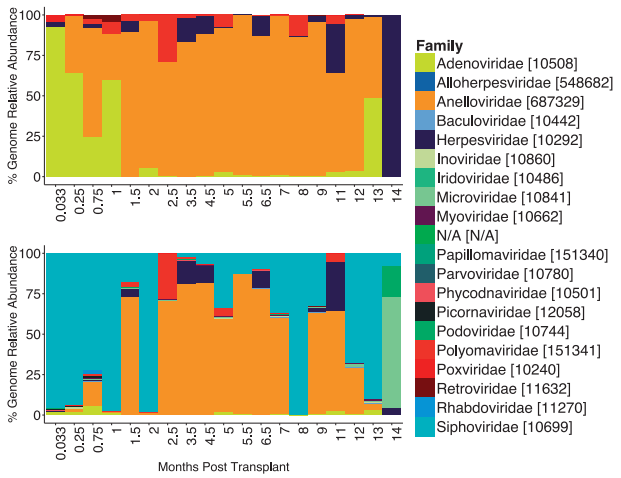

Fig. 2. Genome relative abundance across multiple time-points within the dataset from De Vlaminck *et al.* (2013). Top: six families previously highlighted (De Vlaminck *et al.*, 2013). Bottom: all viral families found (Color version of this figure is available at Bioinformatics online.)

2 Materials and methods

EzMap is partially based on a previously described pipeline (De Vlaminck *et al.*, 2013) (Fig. 1) and was written in Python for SLURM-based computing clusters and multi-core workstation computers. Installation is automated, and requires only the availability of Python 3 and the BioPython library Cock *et al.* (2009). A performance evaluation of the workstation version of EzMap can be found online in the Supplementary Results.

The pipeline workflow is shown in Figure 1 and is fully explained in the Supplementary Methods. Full details of the EM algorithm used (Xia *et al.*, 2011) can also be found there. This important step estimates genome relative abundances, which reflect the relative abundance of each viral genome in the sample. In addition, raw mapping results are output as text files to allow for custom analyses of absolute abundances. The final step of the pipeline generates an interactive report, which can be opened in any modern browser and easily shared. An interactive demo of an EzMap report is available at: http://dekoning-lab.github.io/ezmap/.

3 Discussion

We validated EzMap using the data set from De Vlaminck *et al.* (2013), in which cfDNA samples from 65 cardiac transplants were sequenced over a prospective time course after transplantation. Our re-analysis was broadly consistent with published findings; some expected variation was observed due to more viral genome sequences being currently available (Fig. 2). One noteworthy discrepancy is the appearance of herpesvirus sequences at time-points from throughout the study (Fig. 2). Interestingly, all of these were identified as HHV-6 viruses, which are known to sometimes be reactivated in transplant recipients (Burnham *et al.*, 2016; Lautenschlager and Razonable, 2012] and can be clinically important. Another discrepancy (Fig. 2, bottom) is the high prevalence of Siphoviridae viruses at nearly all time-points. This appears to be due to the recent sequencing of a number of previously uncharacterized genomes. Interestingly, these viruses have bacteria as their natural hosts, and their dynamics may therefore provide a link between the virome and bacterial microbiome communities.

Author contributions

P.C., S.G. and J.d.K. developed the method, P.C. implemented the method, P.C. and J.d.K. analyzed the data, and P.C., S.G. and J.d.K. wrote the paper.

Funding

This work was supported by startup and infrastructure support from the Alberta Children's Hospital Research Institute (to A.P.J.d.K. and S.G.) and by the Canada Foundation for Innovation (CFI LOF #31908 to A.P.J.d.K.).

Conflict of Interest: none declared.

References

Burnham,P. *et al.* (2016) Single-stranded DNA library preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma. *Sci. Rep.*, **6**, 27859.

Cock,P.J.A. *et al.* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422–1423.

De Vlaminck,I. *et al.* (2013) Temporal response of the human virome to immunosuppression and antiviral therapy. *Cell*, **155**, 1178–1187.

De Vlaminck,I. *et al.* (2015) Noninvasive monitoring of infection and rejection after lung transplantation. *Proc. Natl. Acad. Sci. USA*, **112**, 13336–13341.

Dinakaran,V. *et al.* (2014) Elevated levels of circulating DNA in cardiovascular disease patients: metagenomic profiling of microbiome in the circulation. *PLoS One*, **9**, e105221.

Lautenschlager,I. and Razonable,R.R. (2012) Human herpesvirus-6 infections in kidney, liver, lung, and heart transplantation: review. *Transpl. Int.*, **25**, 493–502.

Ngoi,C.N. *et al.* (2016) The plasma virome of febrile adult Kenyans shows frequent parvovirus B19 infections and a novel arbovirus (Kadipiro virus). *J. Gen. Virol.*, **97**, 3359–3367.

Rascovan,N. *et al.* (2016) Metagenomics and the human virome in asymptomatic individuals. *Annu. Rev. Microbiol.*, **70**, 125–141.

Xia,L.C. *et al.* (2011) Accurate genome relative abundance estimation based on shotgun metagenomic reads. *PLoS One*, **6**, e27992.

Young,J.C. *et al.* (2015) Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. *Am. J. Transplant.*, **15**, 200–209.