Metagenomic datasets of cerebrospinal fluid from a small cohort of MS/non-MS patients do not show DNA from the fungal genus *Trichosporon*

Ralf Stephan, unaffiliated

Contact: gtrwst9@gmail.com

In [2] Alonso et al used nested PCR assays together with next-generation sequencing to find *Trichosporon* species in the nervous tissue of 10 patients with MS. We deemed it possible that the fungus would be present in cerebrospinal fluid (CSF) samples. Whole metagenomic shotgun (WMGS) sequencing allows detection of any organism in a sample. With *Trichosporon* any detection would be a true positive because these fungi are not known to be on the skin, or as typical lab contamination. We screened public WMGS datasets of CSF from a cohort of 43 Canadian patients (28 MS, 13 non-MS)[1], using Kraken2[7], the ultrafast kmer-based classifier, using a fungal database augmented with all cleaned available *Trichosporon* genome assemblies from the NCBI. Blasting the marked reads against an equally augmented blastn database revealed no alignments with an evalue $\leq 1e-50$. In general Kraken2 marked not more than 5 consecutive kmers in any read, which is a clear negative.

INTRODUCTION

Only in recent years has the microbiome of different body surfaces and compartments gained importance. The gut and other parts of the skin were the natural first investigation objects. The subarachnoid space containing the CSF is usually considered sterile except during meningitis or arachnoiditis. With recent doubts about the sterility of the CNS[3], and the report of *Trichosporon* in MS patients' nerve tissue[2] we tried to add information to the discussion by

looking at public WMGS sequencing datasets of human CSF samples. The only public datasets from MS patients are well described in [1], where bacteria and virii were examined. The data is of good depth to allow reclassification for fungi. It consists of WMGS datasets of CSF from a cohort of 43 Canadian patients (28 MS, 13 non-MS).

METHODS

The read pairs of each dataset were merged using BBmerge[6] which also did the basic quality control in its default configuration (trim quality threshold 10, minimum insert size 35 bases, minimum number of overlapping bases: 12). All except 4 datasets had a merge rate > 30%. Merged and unmerged reads were joined and all reads with less than 100 bases were discarded using fastx-clipper, part of the FASTX toolkit (https://hannonlab.cshl.edu/fastx_toolkit/).

Kraken2[7] is a fast k-mer based taxonomic classifier that can be installed with a full organism database, which is very demanding on the machine, or specialized databases for all domains. We installed and used the human and an augmented fungi database. We used Kraken2 to mark reads containing human kmers and discarded them (confidence value 0.0). We collected data consisting of all RefSeq genomes of fungi, plus all assemblies of *Malassezia* (a skin fungus and possible contamination) and *Trichosporon* species. We also added some assemblies from outgroups near *Trichosporon*: species from the genera *Apiotrichum*, *Cutaneotrichosporon*, and *Vanrija*. Then we applied the process described in [4] to clean the data from possible contamination with human, bacterial, viral, plasmid, and well known lab contamination snippets (UniVec). After it, we found 433 accessions completely masked (mainly *Malassezia furfur* WGS

scaffolds) and removed them. With the remaining data a Kraken2 database was built, as well as a blastn[8] database.

Kraken2 together with our *Trichosporon* database classified the non-human sequences taxonomically, using a confidence value of 0.02 (meaning 2% of kmers in the read should be positive). There were 893 positive reads in all datasets, however, none of them had more than 5 kmers in a block, and none of them gave an evalue better than 1e-50 when blasted against our database.

We used Biopython[9] for helper scripts. The work was done on a Linux machine with i7-6700 CPU, 32 GB RAM, and a 300 GB SSD.

RESULT AND DISCUSSION

No *Trichosporon* DNA seems to be present in the datasets from [1], neither in the MS nor the non-MS samples. This may not be surprising as the central nervous system is not directly in contact with the CSF. It may be possible that destructive processes can create such a contact and hypothetically allow *Trichosporon* to transfer from the CNS to the CSF but the examined small set of data does not support this possibility.

THANKS

We like to thank Dr. Jovel for publication of the data and fruitful discussion.

REFERENCES

1. Jovel J, O'keefe S, Patterson J, Bording-Jorgensen M, Wang W, Mason AL, Warren KG, Wong GK. [Cerebrospinal Fluid in a Small Cohort of Patients with Multiple Sclerosis Was Generally Free of Microbial DNA](#). Front Cell Infect Microbiol. 2017 Jan 6;6:198. doi: 10.3389/fcimb.2016.00198. eCollection 2016. PubMed PMID: 28111617; PubMed Central PMCID: PMC5216046.
2. Alonso R, Fernández-Fernández AM, Pisa D, Carrasco L. [Multiple sclerosis and mixed microbial infections. Direct identification of fungi and bacteria in nervous tissue](#). Neurobiol Dis. 2018 Sep;117:42-61. doi: 10.1016/j.nbd.2018.05.022. Epub 2018 Jun 1. PubMed PMID: 29859870.
3. Alonso R, Pisa D, Fernández-Fernández AM, Carrasco L. [Infection of Fungi and Bacteria in Brain Tissue From Elderly Persons and Patients With Alzheimer's Disease](#). Front Aging Neurosci. 2018 May 24;10:159. doi: 10.3389/fnagi.2018.00159. eCollection 2018. PubMed PMID: 29881346; PubMed Central PMCID: PMC5976758.
4. Lu J, Salzberg SL. [Removing contaminants from databases of draft genomes](#). PLoS Comput Biol. 2018 Jun 25;14(6):e1006277. doi: 10.1371/journal.pcbi.1006277. eCollection 2018 Jun. PubMed PMID: 29939994; PubMed Central PMCID: PMC6034898.
5. Edgar RC. [Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences](#). PeerJ. 2018 Apr 18;6:e4652. doi: 10.7717/peerj.4652. eCollection 2018. PubMed PMID: 29682424; PubMed Central PMCID: PMC5910792.

6. Bushnell B, Rood J, Singer E. [BBMerge - Accurate paired shotgun read merging via overlap](#). PLoS One. 2017 Oct 26;12(10):e0185056. doi: 10.1371/journal.pone.0185056. eCollection 2017. PubMed PMID: 29073143; PubMed Central PMCID: PMC5657622.
7. Wood DE, Salzberg SL. [Kraken: ultrafast metagenomic sequence classification using exact alignments](#). Genome Biol. 2014 Mar 3;15(3):R46. doi: 10.1186/gb-2014-15-3-r46. PubMed PMID: 24580807; PubMed Central PMCID: PMC4053813.
8. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. [BLAST+: architecture and applications](#). BMC Bioinformatics. 2009 Dec 15;10:421. doi: 10.1186/1471-2105-10-421. PubMed PMID: 20003500; PubMed Central PMCID: PMC2803857.
9. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ. [Biopython: freely available Python tools for computational molecular biology and bioinformatics](#). Bioinformatics. 2009 Jun 1;25(11):1422-3. doi: 10.1093/bioinformatics/btp163. Epub 2009 Mar 20. PubMed PMID: 19304878; PubMed Central PMCID: PMC2682512.