Proposal to identify bacterial species in the serum microbiota of type II diabetics

Introduction

A commonly held view is that circulating blood is a sterile environment. However in reality, individuals go through transient bacteremia due to brushing or flossing of teeth (Maharaj et al., 2012). Such bacteria has the ability to colonize multiple infection sites, most significantly the colonization of endocardial sites by oral bacteria (Baehr, 1912). Furthermore, commonly found oral bacteria such as *Streptococcus sanguinis* and *Porphyromonas gingivalis* have been shown to cause platelet aggregation, endocarditis and have been implicated in heart disease following translocation into the bloodstream (Paik et al. 2005).

A subset of the populace that is at increased risk of bacteremia are type II diabetics. Recently, an increased abundance of bacteria or their cellular components have been observed in the bloodstream of type II diabetic patients (Amar et al., 2011). Type II diabetics exhibit numerous symptoms and have multiple potential sources of bacteremia that would need to be examined. Type II diabetics are three times more likely to develop periodontitis, an inflammatory disease accompanied by a polymicrobial infection (Emrich et al., 1991). Poor oral health and periodontitis also leads to bacterial translocation into the bloodstream and has been shown to significantly correlate with heart disease and endocardial infection (Koren et al., 2001). Further studies have also indicated that bacterial translocation from the lower intestine could also contribute to an increase in blood-borne microbes (Potgieter et al., 2015).

Traditionally the detection of blood-borne bacteria relied on culturing techniques, however this underestimates the total bacterial abundance as it does not take non-culturable bacteria into consideration. Sequence-based approaches have better detection (Amar et al., 2011). A caveat of this approach has been the inability to obtain species-level resolution. This clearly demonstrates a clear need for further study. The objective of this proposal is to quantify the bacterial burden in serum from type 2 diabetics with or without periodontitis compared to non-

diabetic subjects. Our aim is to describe the serum microbiome at a species level resolution to verify which compartment (oral and/or gut cavity) these bacterial originated from.

Aim: To carry out a species-level micrbiome analysis of serum from type II diabetics with or without periodontitis.

We hypothesize that specific species will be significantly elevated in the serum of all type II diabetics and that distinct oral species will be further elevated in diabetics with periodontitis. Using high-throughput sequencing of the bacterial 16S V1-V3 region and species-level minimal entropy decomposition (MED) analysis, we will identify bacterial species present in samples from our patient cohort. Linear discriminate analysis effect size (LEfSe) will be used to identify species that are significantly enriched between each condition (Segeta et al. 2011). This data will provide us with the first species-level analysis of the serum microbiota of T2D patients and will determine if oral species are specifically overrepresented in diabetics with periodontitis.

Experimental Design

Human serum samples have been shown to be successful in estimating blood-borne bacterial abundance (Amar et al., 2011). Using an exisiting collection of serum samples from 25 type II diabetics with periodontitis, 25 type II diabetics without periodontitis and 25 non-diabetic subjects, we will isolate total DNA from each sample with a "Complete DNA Purification" kit (Epicentre). DNA will be amplified for 16s rDNA using eubacterial (518F and 27R) primers (Muyzer et al., 1993), amplifying the V1-V3 region and sequence libraries will be assembled. Mock community controls and sterile reagent-only controls will also be utilized to verify the quality of our input samples. 16s libraries will be sequenced on an Illumina Miseq facility. Quality filtering and chimeric sequence removal will be carried out using Trimmomatic and flexbar. MED analysis will be carried out as described (Eren et al., 2015). Species assignment will be determined by oligotypes that align with >98.5% sequence similarity to rDNA genes in the Human Oral Microbiome Database (HOMD) and/or the Ribosome database project (RDP). After

determining species-level relative abundances, we will compare data from each category via LEfSe analysis, to identify bacteria that are significantly enriched in each group. This data will provide the first species-level analysis of serum microbiota of type II diabetics and will determine if oral species are significantly over-represented in type II diabetics with periodontitis.

Preliminary data- Proof of concept study

To demonstrate the ability to carry out a species level analysis of microbiota and determine relative representation in different groups, I used some publically available data (Paisse et al., 2016). This study looked at the blood of 30 healthy volunteers and looked at the microbiota in Plasma, buffy coat and Red blood cell layer. This use of raw illumina sequence data from this study would help demonstrate proof of concept for my study as it enables the comparison of different categories (i.e. plasma vs buffy coat).

Paisse et al. (2016) sequenced the V3-V4 regions of 16s rRNA using illumina Miseq. After obtaining their illumina sequence data, we carried out quality control using trimmomatic and flexbar. Their sequence data contained individual files composed of each subject's microbial DNA separated by blood component (i.e. plasma and buffy coat). Our analysis used minimal entropy decomposition (MED), which is an unsupervised version of oligotyping (Eren et al., 2015). Traditionally, 16s rDNA sequences from large datasets have been classified into operational taxonomic units (OTUs) by clustering based on a percent similarity cut-off (usually 97%) and then matching the consensus sequence the existing 16s rDNA reference databases. Which was the approach utilized by Paisse et al. (2016). This has provided reliable genus level identification, however oral bacteria such as species belonging to the genus *Streptococcus*, share >98% similarity. MED provides a greater resolution that OTU calling because it sorts sequences by minimizing sequence variation using Shannon entropy, which minimizes the impact of sequencing error and allows for species-level assignment at >98.5% similarity. This

approach of identifying oligotypes has been used to distinguish oral microbiota as shown by a published re-analysis of Human Microbiome project data (Eren et al., 2013).

The ouput of MED data provides a list of oligotypes (denoted by a sample ID) followed by their read abundance. MED also provides the consensus sequence that belongs to each oligotype, in a separate file. These sequences need to be aligned against a database in order to identify which species the sequence belongs to. For the purpose of this assignment I used the top hit from the HOMD database to make my species level calls.

Using jupyter notebook, I was able to essentially replace the sample ID with the top hit from the HOMD database (see supplemental information). I then graphed the read abundance of each species in both the Plasma and the buffy coat(Figures 1-2).
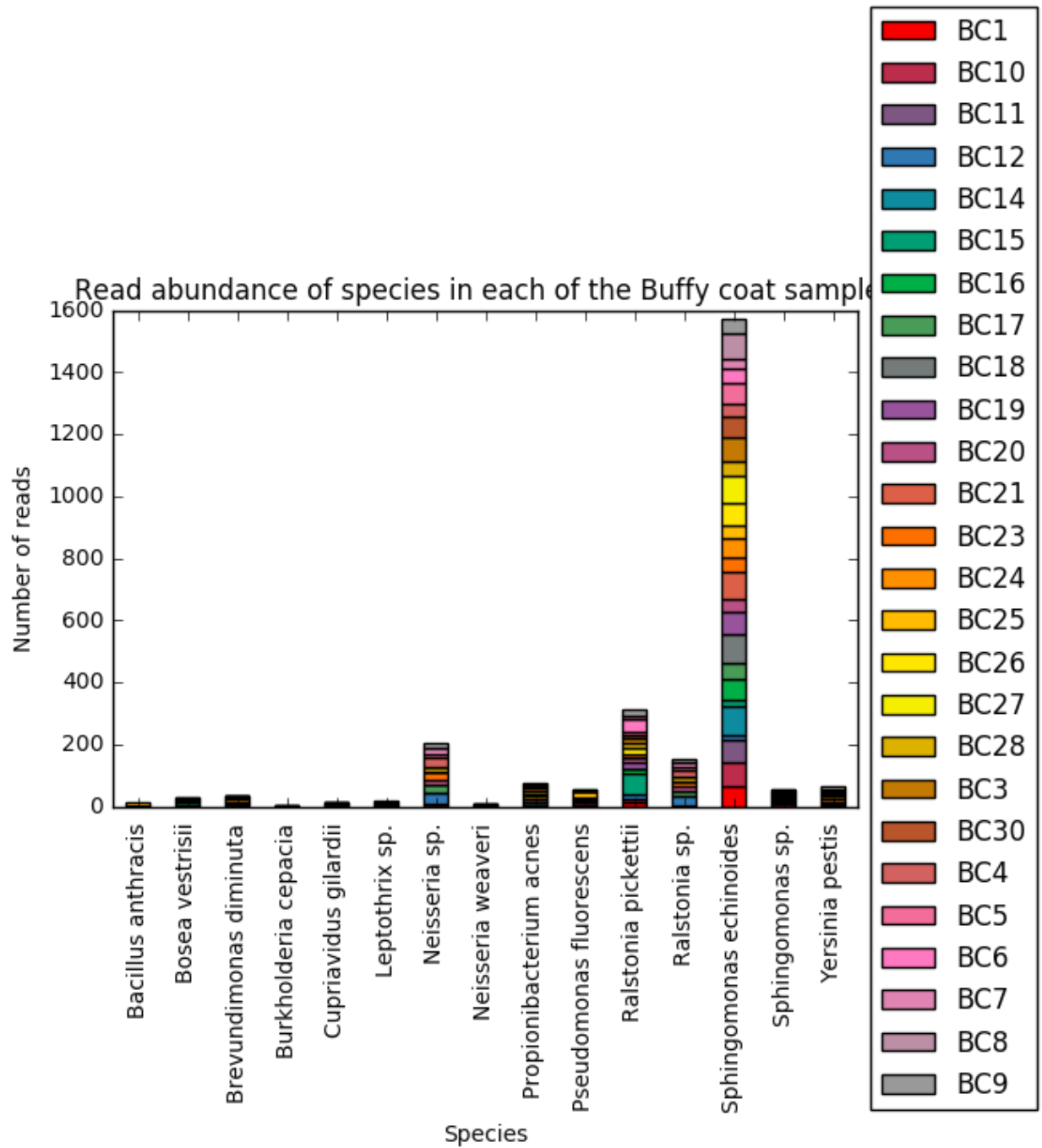
Figure 1: Graph showing the read abundance of bacterial species present in each of the plasma samples.

The results from figure 1 indicate that *Sphingomonas echinoides* has the highest relative abundance in the plasma. This species appears to be distributed between all 19 subjects. The next most abundant species appears to be *Ralstonia pickettii*.

Figure 2: Graph showing the read abundance of bacterial species present in the buffy coat samples

The results from figures 1-2 indicate that *S. echinoides* appears to have the highest abundance in both the plasma and buffy coat samples. This is consistent with the results obtained by Paisse et al.,(2006), where they found the genus sphingomonas represented in about 70% of the reads. The next most abundant species appears to be Ralstonia picketti. This appears to be the case in both the plasma and buffy coat. It also appears that there is more species diversity in the plasma samples.

The next step is to determine which species is significantly over or under-represented in either the plasma or buffy coat. In order to do this I carried out a LefSe analysis (Segeta et al. 2011).
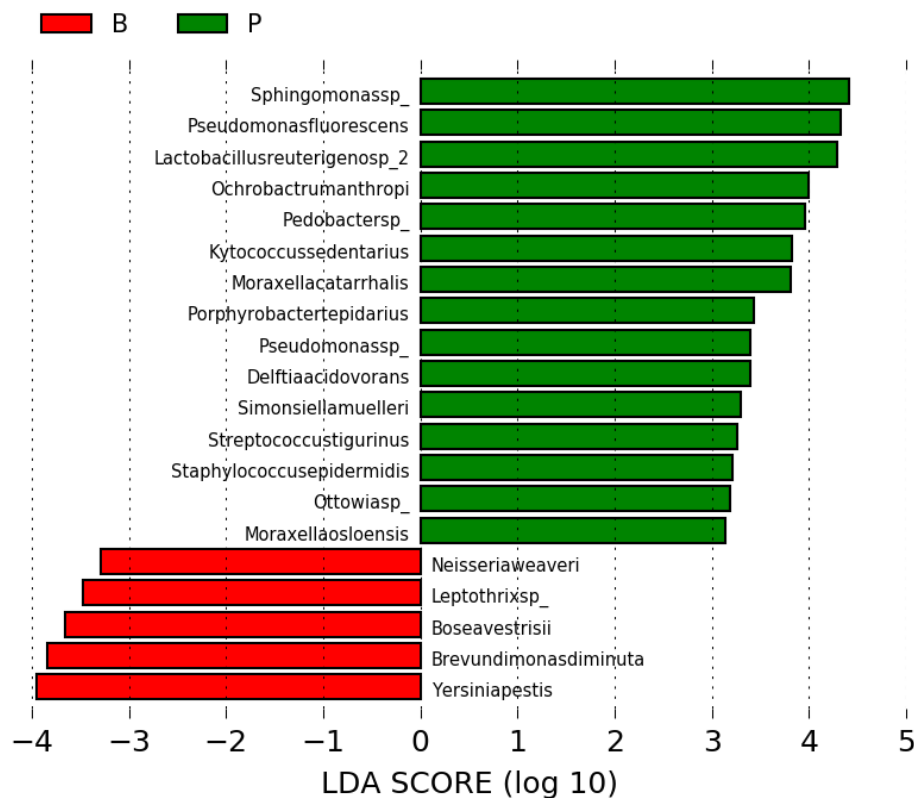

Figure 3: Results of LefSe analysis showing the species over represented in the plasma (blue) and buffy coat (red).

These results which correlate with the observations by Paisse et al., (2016), demonstrate that the approach that we propose is a valid approach that we can use for our study. We will carry out a 16s RNA analysis using MED on serum samples from diabetics with periodontitis, diabetics without periodontitis and healthy subjects. LefSe analysis will be carried out to determine if any species are over or under-represented in these groups when carrying out a pairwise comparison.

References

Amar, J., et al., *Involvement of tissue bacteria in the onset of diabetes in humans: evidence for a concept.* Diabetologia, 2011. **54**(12): p. 3055-61.

Baehr, G., *Glomerular Lesions of Subacute Bacterial Endocarditis.* J Exp Med, 1912. **15**(4): p. 330-47.

Emrich, L.J., M. Shlossman, and R.J. Genco, *Periodontal disease in non-insulin-dependent diabetes mellitus.* J Periodontol, 1991. **62**(2): p. 123-31.

Eren, A.M., et al., *Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences.* ISME J, 2015. **9**(4): p. 968-79.

Koren, O., et al., *Human oral, gut, and plaque microbiota in patients with atherosclerosis.* Proc Natl Acad Sci U S A, 2011. **108 Suppl 1**: p. 4592-8.

Maharaj, B., Y. Coovadia, and A.C. Vayej, *An investigation of the frequency of bacteraemia following dental extraction, tooth brushing and chewing.* Cardiovasc J Afr, 2012. **23**(6): p. 340-4.

Muyzer, G., E.C. de Waal, and A.G. Uitterlinden, *Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA.* Appl Environ Microbiol, 1993. **59**(3): p. 695-700.

Paik, S., Senty. L., Das, S., Noe, JC., Munro, CL. and Kitten, T. (2005). *Identification of virulence determinants for endocarditis in Streptococcus sanguinis by signature-tagged mutagenesis.* Infect Immun, 2005. 73(9):6064-74

Potgieter, M., et al., *The dormant blood microbiome in chronic, inflammatory diseases.* FEMS Microbiol Rev, 2015. **39**(4): p. 567-91.

Segata, N., et al., *Metagenomic biomarker discovery and explanation.* Genome Biol, 2011. **12**(6): p. R60.