Metaproteomics of saliva identifies human protein markers specific for individuals with periodontitis and dental caries compared to orally healthy controls (#11675)

First submission

Please read the **Important notes** below, and the **Review guidance** on the next page.
When ready **submit online**. The manuscript starts on page 3.

Important notes

Editor and deadline

Melissa Grant / 8 Aug 2016

Files	6 Figure file(s) 2 Table file(s) Please visit the overview page to **download and review** the files not included in this review pdf.
Declarations	**Involves the study of human participants/human tissue.**

Review guidelines

Please in full read before you begin

How to review

When ready **submit your review online**. The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this **pdf** and upload it as part of your review

To finish, enter your editorial recommendation (accept, revise or reject) and submit.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to **PeerJ standard**, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (See **PeerJ policy**).

EXPERIMENTAL DESIGN

- Original primary research within **Scope of the journal**.
- Research question well defined, relevant & meaningful. It is stated how research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- Data is robust, statistically sound, & controlled.
- Conclusion well stated, linked to original research question & limited to supporting results.
- Speculation is welcome, but should be identified as such.

The above is the editorial criteria summary. To view in full visit **https://peerj.com/about/editorial-criteria/**

Metaproteomics of saliva identifies human protein markers specific for individuals with periodontitis and dental caries compared to orally healthy controls

Daniel Belstrøm Corresp., [1] , **Rosa R. Jersie-Christensen** [2] , **David Lyon** [3] , **Christian Damgaard** [1] , **Lars J. Jensen** [3] , **Palle Holmstrup** [1] , **Jesper V. Olsen** [2]

[1] Section of Periodontology and Microbiology, Department of Odontology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[2] Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[3] Disease Systems Biology Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

Corresponding Author: Daniel Belstrøm
Email address: dbel@sund.ku.dk

Abstract Background: The composition of the salivary microbiota has been reported to differentiate between patients with periodontitis, dental caries and orally healthy individuals. To identify characteristics of diseased and healthy saliva we thus wanted to compare saliva metaproteomes from patients with periodontitis and dental caries to healthy individuals. **Methods:** Stimulated saliva samples were collected from 10 patients with periodontitis, 10 patients with dental caries and 10 orally healthy individuals. The proteins in the saliva samples were subjected to denaturing buffer and digested enzymatically with LysC and trypsin. The resulting peptide mixtures were cleaned up by solid-phase extraction and separated online with 2h gradients by nano-scale C_{18} reversed-phase chromatography connected to a mass spectrometer through an electrospray source. The eluting peptides were analyzed on a tandem mass spectrometer operated in data-dependent acquisition mode. **Results:** We identified a total of 35664 unique peptides from 4161 different proteins, of which 1946 and 2090 were of bacterial and human origin, respectively. The human protein profiles displayed significant overexpression of the complement system and inflammatory markers in periodontitis and dental caries compared to healthy controls. Bacterial proteome profiles and functional annotation were very similar in health and disease. **Conclusions:** Overexpression of proteins related to the complement system and inflammation seems to correlate with oral disease status. Similar bacterial proteomes in healthy and diseased individuals suggests that the salivary microbiota predominantly thrives in a planktonic state expressing no disease-associated characteristics of metabolic activity.

Metaproteomics of saliva identifies human protein markers specific for individuals with periodontitis and dental caries compared to orally healthy controls

Daniel Belstrøm[1*], Rosa R. Jersie-Christensen[2*], David Lyon[3], Christian Damgaard[1], Lars J. Jensen[3], Palle Holmstrup[1], Jesper V. Olsen[2]

*: these authors contributed equally to this manuscript.

1. Section of Periodontology and Microbiology, Department of Odontology, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark.

2. Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen, Denmark.

3. Disease Systems Biology Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen, Denmark.

Corresponding author

Daniel Belstrøm[1]

E-mail: dbel@sund.ku.dk

Abstract

Background: The composition of the salivary microbiota has been reported to differentiate between patients with periodontitis, dental caries and orally healthy individuals. To identify characteristics of diseased and healthy saliva we thus wanted to compare saliva metaproteomes from patients with periodontitis and dental caries to healthy individuals.

Methods: Stimulated saliva samples were collected from 10 patients with periodontitis, 10 patients with dental caries and 10 orally healthy individuals. The proteins in the saliva samples were subjected to denaturing buffer and digested enzymatically with LysC and trypsin. The resulting peptide mixtures were cleaned up by solid-phase extraction and separated online with 2h gradients by nano-scale C_{18} reversed-phase chromatography connected to a mass spectrometer through an electrospray source. The eluting peptides were analyzed on a tandem mass spectrometer operated in data-dependent acquisition mode.

Results: We identified a total of 35664 unique peptides from 4161 different proteins, of which 1946 and 2090 were of bacterial and human origin, respectively. The human protein profiles displayed significant overexpression of the complement system and inflammatory markers in periodontitis and dental caries compared to healthy controls. Bacterial proteome profiles and functional annotation were very similar in health and disease.

Conclusions: Overexpression of proteins related to the complement system and inflammation seems to correlate with oral disease status. Similar bacterial proteomes in healthy and diseased individuals suggests that the salivary microbiota predominantly thrives in a planktonic state expressing no disease-associated characteristics of metabolic activity.

Introduction

Saliva is a biological fluid critically involved in maintenance of oral homeostasis (Marsh et al., 2016), as qualitative and quantitative changes of saliva associates with increased frequency and severity of diseases in the oral cavity (Almstahl and Wikstrom, 1999;Dawes et al., 2015). Furthermore, saliva is easily and non-invasively collected (Giannobile et al., 2011), making it interesting to screen for biomarkers associated with oral and general health and disease status (Baum et al., 2011;Zhang et al., 2016).

In the last decade salivary biomarkers of periodontitis and dental caries have been intensively investigated (Miller et al., 2010;Yoshizawa et al., 2013). These include salivary bacterial profiles that differentiate in patients with periodontitis (Belstrom et al., 2014b;Belstrom et al., 2016;Paju et al., 2009), dental caries (Belstrom et al., 2014a;Belstrom et al., 2015;Yang et al., 2012) and orally healthy individuals. Furthermore, increased salivary levels of inflammatory protein biomarkers such as interleukin-1β (IL-1β), IL-6 and matrix metalloproteinase-8 (MMP-8) have been described to be associated with periodontal disease status (Ebersole et al., 2015;Ebersole et al., 2013;Kinney et al., 2011;Rathnayake et al., 2013). Recently, the salivary transcriptome has been assessed (Spielman et al., 2012), and some transcriptomic characteristics of saliva have been reported in patients with dental caries (Do et al., 2015). Collectively, these reports conclude that biomarkers of different biological origin may be adequately assessed in saliva samples and support the concept that the biological composition of saliva reflects individual oral health status.

Mass spectrometry-based proteomics enables characterization of the protein content in any sample, including proteins of human and bacterial origin. It thus provides the possibility for simultaneous characterization of bacterial and host specific differences of saliva associated with oral health and disease. Only three studies have so far attempted to perform metaproteomic analysis of saliva in oral health (Grassl et al., 2016;Jagtap et al., 2012;Rudney et al., 2010). To the best of our knowledge, no study has so far compared metaproteomic profiles of saliva from patients with periodontitis and dental caries to orally healthy individuals.

The aim of the present study was to characterize the salivary metaproteome in 30 saliva samples, and compare human and bacterial proteome profiles between patients with periodontitis, dental caries and orally healthy individuals. The hypothesis was that both bacterial and human subsets

of salivary metaproteome would differentiate between individuals with different oral health status.

Materials and Methods

Study population and sample collection

The study population, clinical examination and collection of saliva samples have been presented in detail (Belstrom et al., 2016). In brief, saliva production was induced by chewing on a tasteless paraffin gum, and chewing-stimulated saliva samples were collected from 10 patients with periodontitis, 10 patients with dental caries and 10 orally healthy individuals following a standardized protocol (Kongstad et al., 2013). Immediately after collection saliva samples were divided into four aliquots and stored at -80°C for further analysis. One aliquot has previously been analyzed by next-generation sequencing (the Human Oral Microbe Identification using Next Generation Sequencing, HOMI*NGS*) (Belstrom et al., 2016). All participants signed an informed consent prior to participation, and the study was approved by the regional ethical committee (H-15000856-53175) and reported to the Danish Data Authorization (2015-54-0970).

Sample preparation

The saliva proteome samples were prepared as described in (Jersie-Christensen et al., 2016) with a few modifications. Briefly, 1 ml of saliva was mixed with 1.5 ml lysis buffer (9M Guanidine hydrochloride, 10mM Chloroacetamide, 5mM *tris*(2-carboxyethyl)phosphine in 100mM Tris pH 8.5) and heated for 10 min (99°C) followed by 4 min of sonication.

Protein concentration was measured with Bradford protein assay and ranged from 1 mg/ml to 2.5 mg/ml. All samples were digested with the same amount of Lysyl Endoproteinase (Wako, Osaka, Japan) in a ratio of 1:100 w/w calculated from the highest concentration for 2h. Samples were diluted to a final volume of 10ml with 25mM Tris pH8 and digested overnight with Trypsin (modified sequencing grade, Sigma) in a 1:100 w/w ratio.

Digestion was quenched by adding 1 ml of 10% trifluoroacetic acid and centrifuged at 2000g for 5 min. The resulting soluble peptides in the supernatant were desalted and concentrated on Waters Sep-Pak reversed-phase C_{18} cartridges (one per sample) and the tryptic peptide mixtures were eluted with 40% acetonitrile (ACN) followed by 60% ACN. Peptide concentrations were determined by NanoDrop (Thermo, Wilmington, DE) measurement.

106 *Mass spectrometry analysis*

107 1.5µg peptide mixture from each sample was analyzed by online nano-scale liquid
108 chromatography tandem mass spectrometry (LC-MS/MS) in turn. Peptides were separated on an
109 in-house packed 50 cm capillary column with 1.9 µm Reprosil-Pur C_{18} beads using an EASY-
110 nLC 1000 system (Thermo Scientific). The column temperature was maintained at 50°C using an
111 integrated column oven (PRSO-V1, Sonation GmbH, Biberach, Germany). Buffer A consisted of
112 0.1% Formic Acid, and Buffer B of 80% ACN, 0.1% Formic Acid. The flow rate of the gradient
113 was 200 nl/min and started at 5% buffer B, going to 25% buffer B in 110 min, followed by a 25
114 min step going to 40% buffer B and continuing to 80% buffer B in 5 min for a 5 min wash and
115 returning to 5% in 5 min and continuing for re-equilibration for 5 min.

116 The Q Exactive HF instrument (Thermo Scientific, Bremen, Germany) was run in a data
117 dependent acquisition mode using a top 12 Higher-Collisional Dissociation (HCD)-MS/MS
118 method with the following settings. Spray voltage was set to 2 kV, S-lens RF level at 50, and
119 heated capillary at 275 °C. Full scan resolutions were set to and 60,000 at m/z 200 and scan
120 target was 3×10^6 with a maximum fill time of 20 ms. Full-scan MS mass range was set to 300–
121 1750 and dynamic exclusion to 20 s. Target value for HCD-MS/MS scans was set at 1×10^5 with
122 a resolution of 30,000 and a maximum fill time of 60 ms. Normalized collision energy was set at
123 28.

124 *Data analysis*

125 All 30 raw LC-MS/MS data files were processed together using MaxQuant version 1.5.0.38 (Cox
126 and Mann, 2008) with default settings and match between runs. The integrated Andromeda
127 peptide search engine and a reversed database approach applying a 1% FDR at both peptide and
128 protein level was used. The data was searched in two iterations analogous to a previously
129 described metaproteomics database search strategy (Jagtap et al., 2013) First, the search space
130 consisted of the full SwissProt protein database (2015) and the Human Oral Microbiome
131 database (Chen et al., 2010) (both downloaded August 2014). The resulting search output was
132 then used for reduction of the search space after filtering on different parameters. As a quality
133 control measure proteins with less than 2 unique peptides were removed. Furthermore, we
134 required proteins to be detected in at least 5 out of 30 samples. Accession numbers from the

Majority protein IDs column in the proteinGroups.txt were used to retrieve information about Lowest Common Ancestor (LCA) for each protein group entry. To find the Lowest Common Ancestor (LCA) of a protein group, accession numbers with the most peptide-associations were selected, mapped to species and their full taxonomic lineage. The lowest taxonomic rank of the intersection of the latter yielded the LCA. All LCA searches resulting in the parvorder Catarrhini (primates) were set to be human. LCAs at taxonomic rank of species and genera, as well as all of their descendants were used to create a new, reduced search space. The latter was used for the second iteration of MS data identification and quantification and all accession numbers within a protein group were used to perform LCA searches. The above functionality was achieved using Python. Species names from SwissProt and HOMD were mapped to NCBI taxonomic identifiers using UniProt (http://www.uniprot.org/docs/speclist) and NCBI resources (http://www.ncbi.nlm.nih.gov/Taxonomy/TaxIdentifier/tax_identifier.cgi), respectively. Full taxonomic lineages were retrieved from NCBI Taxonomy database dump files (ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/).

Protein intensities based on summed peptide MS signal intensities were quantile normalized using the limma package version 3.24.15 under R version 3.2.2. Only proteins identified with more than one peptide ("razor+unique") and present in more than five out of the 30 samples were considered for further analysis. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaino et al., 2016) partner repository with the dataset identifier PXD004319.

For comparative analysis of the human protein profiles, the normalized intensity values were log2 transformed. For principal component analysis (PCA), missing values were replaced with the constant value 19, representing the lowest protein intensity value measured. Analysis of significance (ANOVA) between groups was performed with the software package Perseus (http://coxdocs.org/doku.php?id=perseus:start). The resulting differentially expressed proteins were clustered using Euclidian distance after scaling the data by subtracting the mean intensity value.

Functional annotation of bacterial proteins

Bacterial proteins from HOMD were searched against Hidden Markov Models (HMMs) of bacterial Nested Orthologous Groups (NOGs) from eggNOG (Huerta-Cepas et al., 2016) using HMMscan version 3.1 (http://hmmer.org)(Eddy, 2009). For each protein query the resulting hits were restricted by two criteria. E-values had to be equal or lower than 1e-4 and a maximum overlap of 8 amino acids of HMMs was allowed (selecting hits with the lowest e-value). All corresponding NOG-names were used to retrieve Gene Ontology (GO)-terms as well as KEGG pathways from eggNOG.

KEGG pathway enrichment and characterization

To gain insights into differences between the three sample groups KEGG pathway enrichment was performed using a modified version of AGOtool (Scholz et al., 2015) Individual samples were grouped to sample categories and the three paired combinations used for the enrichment analysis. All bacterial protein groups with an LCA at rank genus or below were selected. Benjamini–Hochberg correction (FDR) of p-values was applied to correct for multiple testing. The FDR was set to 1%. The following additional filter criterion was applied. The fold change had to be equal or higher than 2 or equal or lower than 0.5.

To get a functional overview of the bacterial proteins, we characterized each individual sample group by counting the number of protein groups associated with each KEGG pathway. For visualization purposes (Fig. S2), we selected the most highly associated terms. Within each group the number of associations was converted to percentages, and the most highly associated terms retained, until a cumulative sum of 90% was reached. This reduced the number of KEGG terms from 135 to 50.

Results

General findings

Biomass analyses based on summed protein intensity measures demonstrated that approximately 95% and 3% of the total protein biomass was of human and bacterial origin, respectively (Fig. 1). Food-related proteins and proteins that could not be assigned to kingdom level comprised the rest of the biomass. We identified a total of 35664 unique peptides from 4161 proteins, of which 97% of the identified proteins could be assigned as bacterial or human proteins, with almost equal numbers of the two (Fig. 1 and Table 1). No differences in protein biomass or numbers between groups were observed (Fig. S1).

Human protein profiling

Principal component analysis (Fig. 2A) of the human proteins in saliva showed decent separation of samples from patients with periodontitis and dental caries from orally healthy individuals, based on the most decisive component of the dataset, accounting for 17.9% of the variation. The most enriched KEGG pathway in component 1 and 2 was ‘Complement and coagulation cascades’ (Fig. 2B). Component 2 also separated samples from patients with dental caries and periodontitis patients from orally healthy individuals with the component explaining 12.7% of the variation. Two of the most enriched terms in component 2 in the direction of the orally healthy individuals were KEGG pathway ‘Salivary secretion’ and GOBP ‘protein glycosylation’.

From a total of 2090 identified proteins of human origin, 60 proteins were significantly differentially expressed when performing multiple sample test (ANOVA, $p<0.05$). Hierarchical cluster analysis of the proteins nicely separated the three sample groups, although 3 periodontitis individuals cluster together with the healthy group (Fig. 3). We identified three main protein clusters. Cluster I contains human proteins that are higher expressed in both disease groups compared to controls, and 10 of 20 proteins in this cluster are associated with the GO term ‘innate immune response’ (protein name in purple). Cluster II consist of 9 proteins that distinguish the individuals with caries from the other groups. In cluster III the protein intensities in the caries group are higher than the mean, for the orally healthy group it is around the mean and for the individuals with periodontitis lower.

213 *Bacterial protein profiling*

214 Of the 1946 proteins of bacterial origin identified, approximately 92% and 34% could be
215 assigned to genus and species level, respectively (Table 1). A total of 29 different bacterial
216 genera and 81 species were identified. The five most predominant bacterial genera were
217 *Streptococcus*, *Prevotella, Veillonella*, *Rothia* and *Neisseria* collectively representing approx.
218 70% of the total bacterial mass. The five most predominant bacterial species identified were
219 *Rothia mucilanginosa*, *Veillonella atypica*, *Prevotella histicola*, *Prevotella melaninogenica* and
220 *Streptococcus salivarius*. Abundances of the 20 most predominant bacterial genera and species
221 are displayed in Fig. 4A-B. No statistically significant differences were observed between groups
222 at genus or species level. However, at genus level there is a trend of higher proportion of
223 *Veillonella* and lower proportion of *Haemophilus* were associated with dental caries and higher
224 proportions of *Fusobacterium*, *Leptotrichia* and *Selenomonas* and lower proportions of
225 *Streptococcus*, *Rothia* and *Haemophilus* were associated with periodontitis, when compared to
226 orally healthy individuals. The same trend is seen at species level where higher proportion of
227 *Veillonella atypica* and lower proportion of *Haemophilus parainfluenzae* were associated with
228 dental caries, and higher proportions of *Fusobacterium periodonticum* and *Leptotrichia wadei*
229 and lower proportions of *Haemophilus parainfluenzae* were associated with periodontitis, when
230 compared to orally healthy individuals. A full list of all bacterial genera and species identified
231 are presented in Table S1.

232 *KEGG pathway enrichment for bacterial proteins*

233 KEGG pathway enrichment analysis of bacterial proteins resulted in no significant differences
234 with the application of the previously mentioned fold-change and FDR filter criteria. The
235 characterization of functional associations of bacterial proteins is shown in Fig. S2.

236

Discussion

The purpose of the present study was to compare metaproteome profiles of saliva from patients with periodontitis or dental caries to that of orally healthy individuals, as we hypothesized that the composition of the salivary metaproteome would associate with oral health status. To the best of our knowledge, this is the first study to characterize both human and bacterial parts of the salivary metaproteome in patients with periodontitis and dental caries.

In this study, proteins of bacterial origin constituted 46% of the proteome diversity, despite only 3% of the total biomass being bacterial. This agrees with the previously reported approx. 1% of DNA in saliva being of bacterial origin (Lazarevic et al., 2012).

We identify 1946 different bacterial proteins, which is in the same range as the pioneering studies of the salivary metaproteome (Jagtap et al., 2012;Rudney et al., 2010) but substantially higher than in dental plaque (983 proteins) (Belda-Ferre et al., 2015). From the total 1946 bacterial proteins, 92% and 34% could be assigned to genus and species level, respectively (Table 1). The percentage of bacterial proteins identified at genus and species level is considerably higher than what has previously been accomplished in metaproteomic analysis of saliva (Jagtap et al., 2012;Rudney et al., 2010). *Streptococcus*, *Prevotella*, *Veillonella*, *Rothia* and *Neisseria* were the most predominant genera identified, constituting approx. 70% of the biomass across all samples (Fig 4A). This phylogenetic distribution is in line with analysis of the same samples using next-generation sequencing (Belstrom et al., 2016), and with previous metaproteomic analysis of saliva in oral health (Grassl et al., 2016;Jagtap et al., 2012;Rudney et al., 2010). By contrast, an analysis of 17 plaque samples from patients with dental caries and healthy controls by metagenomics, metatranscriptomics and metaproteomics found different bacterial compositions in dental plaque at DNA, mRNA and protein level (Belda-Ferre et al., 2015). This may reflect differences between studying the metabolically active dental plaque biofilm and the planktonic, metabolically inactive state of the salivary microbiota, and it is in concordance with the functional annotation analysis performed (Fig. S2).

Furthermore, 2090 different proteins of human origin were identified, which is more than in metaproteome profiling of dental plaque (Belda-Ferre et al., 2015) and less than a recent study that identified more than 3700 different human proteins in a mouth swab analysis (Grassl et al.,

2016). The higher number of human protein identifications in mouth swabs is probably due to swabbing the inside of the complete oral cavity including the inside of the cheek. In this study we used stimulated saliva samples, which may have diluted the concentration of proteins within the samples compared to that of unstimulated saliva (Schafer et al., 2014;Yakob et al., 2014). This will of course also affect number of identifications. Based on this finding, unstimulated saliva samples may be preferred for in-depth analysis of the salivary proteome. However, as collection of unstimulated saliva samples is considerably more intricate and time-consuming than collection of stimulated saliva samples, the feasibility of using unstimulated saliva samples for population-based biomarker screening approaches may be limited (Belstrom et al., 2016). Consequently, stimulated saliva samples were used in this study.

Data on the human profile of the salivary metaproteome showed differences between oral health and disease, as proteins involved in innate immunity and inflammatory proteins were more abundantly expressed in saliva samples from patients with periodontitis and dental caries than orally healthy individuals (Fig 3). Thus, by use of a contemporary metaproteomics approach we were able to explore that salivary expression of proteins from the innate immune system associates with periodontitis and dental caries. Interestingly, these data are in line with previous reports on periodontitis patients (Aurer et al., 1999;Cole et al., 1981). Likewise, it has been reported that active components of the complement system in the gingival crevicular fluid associates with both periodontitis (Courts et al., 1977;Schenkein and Genco, 1977) and gingivitis (Attstrom et al., 1975;Patters et al., 1989). Increased local activation of the complement system in the periodontal tissues increases vascular permeability, vasodilatation and recruitment of inflammatory cells, resulting in excessive release of reactive oxygen species, proteolytic enzymes and interleukins (Okada and Silverman, 1979;Okuda and Takazoe, 1980;Watanabe et al., 1997). Furthermore, serum levels of complement proteins, has been suggested to express a linear relationship with the degree of periodontal inflammation (Henry et al., 1987). Gingivitis is a mild form of gum disease that results in irritation, redness and swelling caused by inflammation of the gums. Thus, the abundant expression of complement proteins and inflammatory mediators in saliva might reflect either a spillover from the gingival crevicular fluid, or alternatively, mirror increased serum levels of these proteins. Notably, while the complement system has been acknowledged to have a profound role in the pathogenesis of periodontitis (Damgaard et al., 2015), the complement system seems to have limited impact on

development of dental caries. The expression of complement proteins and other inflammatory proteins in saliva from patients with dental caries is most likely associated with gingivitis in the periodontal tissues adjacent to approximal and gingival caries lesions, and presumably not directly associated with presence of dental caries as such.

Conclusion

Quantitative proteomics data from the present investigation suggest that the salivary microbiota predominantly thrives in a planktonic state with limited metabolic activity, as comparable microbial compositions of the salivary microbiota were obtained based on different omics analysis. Thus, the bacterial part of the metaproteome seems to be inadequate for biomarker analysis of periodontitis and caries. Conversely, a set of human proteins hold the potential to be used as future biomarkers of oral disease status. However, the cross-sectional study design obviously hampers the possibility to address causality of this observation. Thus, future large-scale longitudinal studies of human saliva proteome changes are warranted to reveal the full potential of quantitative proteomics of saliva as a technique to discover biomarkers of oral health and disease.

Funding

This work was in part funded by the University of Copenhagen (KU2016 programme). Work at Novo Nordisk Foundation Center for Protein Research (CPR) is funded in part by a generous donation from the Novo Nordisk Foundation (Grant number NNF14CC0001). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

320 Reference List

321

322 UniProt: a hub for protein information (2015). *Nucleic Acids Res* 43(Database issue):D204-D212.

323 Almstahl A, Wikstrom M (1999). Oral microflora in subjects with reduced salivary secretion. *J Dent Res*
324 78(8):1410-1416.

325 Attstrom R, Laurel AB, Lahsson U, Sjoholm A (1975). Complement factors in gingival crevice material
326 from healthy and inflamed gingiva in humans. *J Periodontal Res* 10(1):19-27.

327 Aurer A, Aurer-Kozelj J, Stavljenic-Rukavina A, Kalenic S, Ivic-Kardum M, Haban V (1999). Inflammatory
328 mediators in saliva of patients with rapidly progressive periodontitis during war stress induced incidence
329 increase. *Coll Antropol* 23(1):117-124.

330 Baum BJ, Yates JR, III, Srivastava S, Wong DT, Melvin JE (2011). Scientific frontiers: emerging
331 technologies for salivary diagnostics. *Adv Dent Res* 23(4):360-368.

332 Belda-Ferre P, Williamson J, Simon-Soro A, Artacho A, Jensen ON, Mira A (2015). The human oral
333 metaproteome reveals potential biomarkers for caries disease. *Proteomics* 15(20):3497-3507.

334 Belstrom D, Fiehn NE, Nielsen CH, Holmstrup P, Kirkby N, Klepac-Ceraj V *et al.* (2014a). Altered Bacterial
335 Profiles in Saliva from Adults with Caries Lesions: A Case-Cohort Study. *Caries Res* 48(5):368-375.

336 Belstrom D, Fiehn NE, Nielsen CH, Kirkby N, Twetman S, Klepac-Ceraj V *et al.* (2014b). Differences in
337 bacterial saliva profile between periodontitis patients and a control cohort. *J Clin Periodontol* 41(2):104-
338 112.

339 Belstrom D, Fiehn NE, Nielsen CH, Klepac-Ceraj V, Paster BJ, Twetman S *et al.* (2015). Differentiation of
340 salivary bacterial profiles of subjects with periodontitis and dental caries. *J Oral Microbiol* 7:27429.

341 Belstrom D, Paster BJ, Fiehn NE, Bardow A, Holmstrup P (2016). Salivary bacterial fingerprints of
342 established oral disease revealed by the Human Oral Microbe Identification using Next Generation
343 Sequencing (HOMINGS) technique. *J Oral Microbiol* 8:30170.

344 Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE (2010). The Human Oral Microbiome
345 Database: a web accessible resource for investigating oral microbe taxonomic and genomic information.
346 *Database (Oxford)* 2010:baq013.

347 Cole MF, Hsu SD, Baum BJ, Bowen WH, Sierra LI, Aquirre M *et al.* (1981). Specific and nonspecific
348 immune factors in dental plaque fluid and saliva from young and old populations. *Infect Immun*
349 31(3):998-1002.

350 Courts FJ, Boackle RJ, Fudenberg HH, Silverman MS (1977). Detection of functional complement
351 components in gingival crevicular fluid from humans with periodontal diseases. *J Dent Res* 56(3):327-
352 331.

353 Cox J, Mann M (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range
354 mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26(12):1367-1372.

355 Damgaard C, Holmstrup P, Van Dyke TE, Nielsen CH (2015). The complement system and its role in the
356 pathogenesis of periodontitis: current concepts. *J Periodontal Res* 50(3):283-293.

357 Dawes C, Pedersen AM, Villa A, Ekstrom J, Proctor GB, Vissink A *et al.* (2015). The functions of human
358 saliva: A review sponsored by the World Workshop on Oral Medicine VI. *Arch Oral Biol* 60(6):863-874.

359 Do T, Sheehy EC, Mulli T, Hughes F, Beighton D (2015). Transcriptomic analysis of three Veillonella spp.
360 present in carious dentine and in the saliva of caries-free individuals. *Front Cell Infect Microbiol* 5:25.

361 Ebersole JL, Nagarajan R, Akers D, Miller CS (2015). Targeted salivary biomarkers for discrimination of
362 periodontal health and disease(s). *Front Cell Infect Microbiol* 5:62.

363 Ebersole JL, Schuster JL, Stevens J, Dawson D, III, Kryscio RJ, Lin Y *et al.* (2013). Patterns of salivary
364 analytes provide diagnostic capacity for distinguishing chronic adult periodontitis from health. *J Clin*
365 *Immunol* 33(1):271-279.

366 Eddy SR (2009). A new generation of homology search tools based on probabilistic inference. *Genome*
367 *Inform* 23(1):205-211.

368 Giannobile WV, McDevitt JT, Niedbala RS, Malamud D (2011). Translational and clinical applications of
369 salivary diagnostics. *Adv Dent Res* 23(4):375-380.

370 Grassl N, Kulak NA, Pichler G, Geyer PE, Jung J, Schubert S *et al.* (2016). Ultra-deep and quantitative
371 saliva proteome reveals dynamics of the oral microbiome. *Genome Med* 8(1):44.

372 Henry CA, Ungchusri T, Charbeneau TD, Winford TE (1987). Relationships of serum opsonins and
373 complement in human experimental gingivitis. *J Periodontol* 58(3):177-186.

374 Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC *et al.* (2016). eggNOG 4.5: a
375 hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and
376 viral sequences. *Nucleic Acids Res* 44(D1):D286-D293.

377 Jagtap P, Goslinga J, Kooren JA, McGowan T, Wroblewski MS, Seymour SL *et al.* (2013). A two-step
378 database search method improves sensitivity in peptide sequence matches for metaproteomics and
379 proteogenomics studies. *Proteomics* 13(8):1352-1357.

380 Jagtap P, McGowan T, Bandhakavi S, Tu ZJ, Seymour S, Griffin TJ *et al.* (2012). Deep metaproteomic
381 analysis of human salivary supernatant. *Proteomics* 12(7):992-1001.

382 Jersie-Christensen RR, Sultan A, Olsen JV (2016). Simple and Reproducible Sample Preparation for Single-
383 Shot Phosphoproteomics with High Sensitivity. *Methods Mol Biol* 1355:251-260.

384 Kinney JS, Morelli T, Braun T, Ramseier CA, Herr AE, Sugai JV *et al.* (2011). Saliva/pathogen biomarker
385 signatures and periodontal disease progression. *J Dent Res* 90(6):752-758.

386 Kongstad J, Ekstrand K, Qvist V, Christensen LB, Cortsen B, Gronbaek M *et al.* (2013). Findings from the
387 oral health study of the Danish Health Examination Survey 2007-2008. *Acta Odontol Scand*.

Lazarevic V, Whiteson K, Gaia N, Gizard Y, Hernandez D, Farinelli L *et al.* (2012). Analysis of the salivary microbiome using culture-independent techniques. *J Clin Bioinforma* 2:4.

Marsh PD, Do T, Beighton D, Devine DA (2016). Influence of saliva on the oral microbiota. *Periodontol 2000* 70(1):80-92.

Miller CS, Foley JD, Bailey AL, Campell CL, Humphries RL, Christodoulides N *et al.* (2010). Current developments in salivary diagnostics. *Biomark Med* 4(1):171-189.

Okada H, Silverman MS (1979). Chemotactic activity in periodontal disease. I. The role of complement in monocyte chemotaxis. *J Periodontal Res* 14(1):20-25.

Okuda K, Takazoe I (1980). Activation of complement by dental plaque. *J Periodontal Res* 15(3):232-239.

Paju S, Pussinen PJ, Suominen-Taipale L, Hyvonen M, Knuuttila M, Kononen E (2009). Detection of multiple pathogenic species in saliva is associated with periodontal infection in adults. *J Clin Microbiol* 47(1):235-238.

Patters MR, Niekrash CE, Lang NP (1989). Assessment of complement cleavage in gingival fluid during experimental gingivitis in man. *J Clin Periodontol* 16(1):33-37.

Rathnayake N, Akerman S, Klinge B, Lundegren N, Jansson H, Tryselius Y *et al.* (2013). Salivary biomarkers of oral health: a cross-sectional study. *J Clin Periodontol* 40(2):140-147.

Rudney JD, Xie H, Rhodus NL, Ondrey FG, Griffin TJ (2010). A metaproteomic analysis of the human salivary microbiota by three-dimensional peptide fractionation and tandem mass spectrometry. *Mol Oral Microbiol* 25(1):38-49.

Schafer CA, Schafer JJ, Yakob M, Lima P, Camargo P, Wong DT (2014). Saliva diagnostics: utilizing oral fluids to determine health status. *Monogr Oral Sci* 24:88-98.

Schenkein HA, Genco RJ (1977). Gingival fluid and serum in periodontal diseases. I. Quantitative study of immunoglobulins, complement components, and other plasma proteins. *J Periodontol* 48(12):772-777.

Scholz C, Lyon D, Refsgaard JC, Jensen LJ, Choudhary C, Weinert BT (2015). Avoiding abundance bias in the functional annotation of post-translationally modified proteins. *Nat Methods* 12(11):1003-1004.

Spielman N, Ilsley D, Gu J, Lea K, Brockman J, Heater S *et al.* (2012). The Human Salivary RNA Transcriptome Revealed by Massively Parallel Sequencing. *Clin Chem*.

Vizcaino JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I *et al.* (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* 44(D1):D447-D456.

Watanabe K, Blew B, Scherer M, Burke J, Koh G, Block C *et al.* (1997). CD11b mRNA expression in neutrophils isolated from peripheral blood and gingival crevicular fluid. *J Clin Periodontol* 24(11):814-822.

Yakob M, Fuentes L, Wang MB, Abemayor E, Wong DT (2014). Salivary biomarkers for detection of oral squamous cell carcinoma - current state and recent advances. *Curr Oral Health Rep* 1(2):133-141.

Yang F, Zeng X, Ning K, Liu KL, Lo CC, Wang W *et al.* (2012). Saliva microbiomes distinguish caries-active from healthy human populations. *ISME J* 6(1):1-10.

Yoshizawa JM, Schafer CA, Schafer JJ, Farrell JJ, Paster BJ, Wong DT (2013). Salivary biomarkers: toward future clinical and diagnostic utilities. *Clin Microbiol Rev* 26(4):781-791.

Zhang Y, Sun J, Lin CC, Abemayor E, Wang MB, Wong DT (2016). The emerging landscape of salivary diagnostics. *Periodontol 2000* 70(1):38-52.

Figure 1(on next page)

Figure 1

Protein biomass and abundance across sample groups: Relative distribution as a measure of summed intensity and protein count.

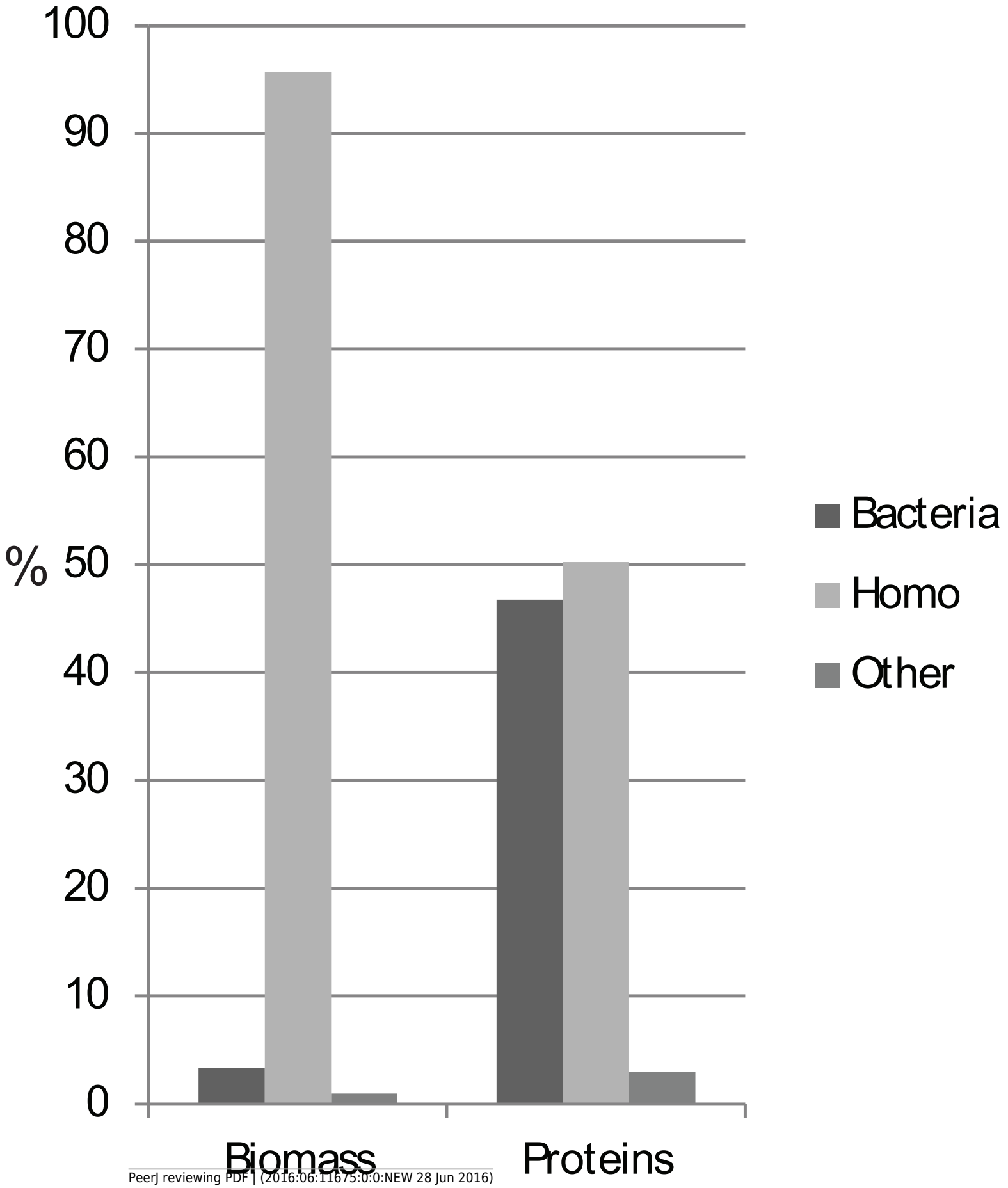
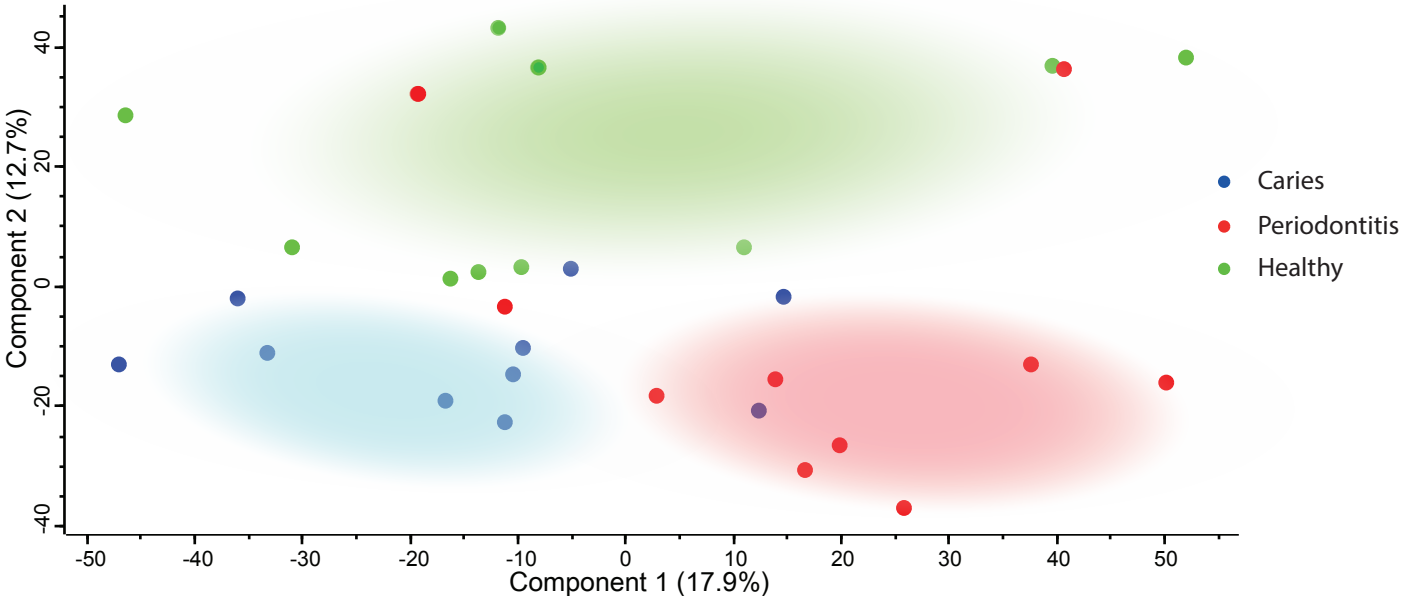

100
90
80
70
60
% 50
40
30
20
10
0
Bacteria
Homo
Other
Biomass
Proteins

Figure 2(on next page)

Figure 2

Principal Component Analysis: A) PCA plot of individuals with caries (blue), periodontitis (red) and orally healthy individuals (green). B) Loadings driving the separation of the PCA plot are mainly proteins belonging to complement and coagulation cascades (purple) for discriminating diseased from healthy individuals. Proteins belonging to salivary secretion and protein glycosylation (green) are mainly defining the healthy individuals.

A


Component 2 (12.7%)
Component 1 (17.9%)
Caries
Periodontitis
Healthy

B

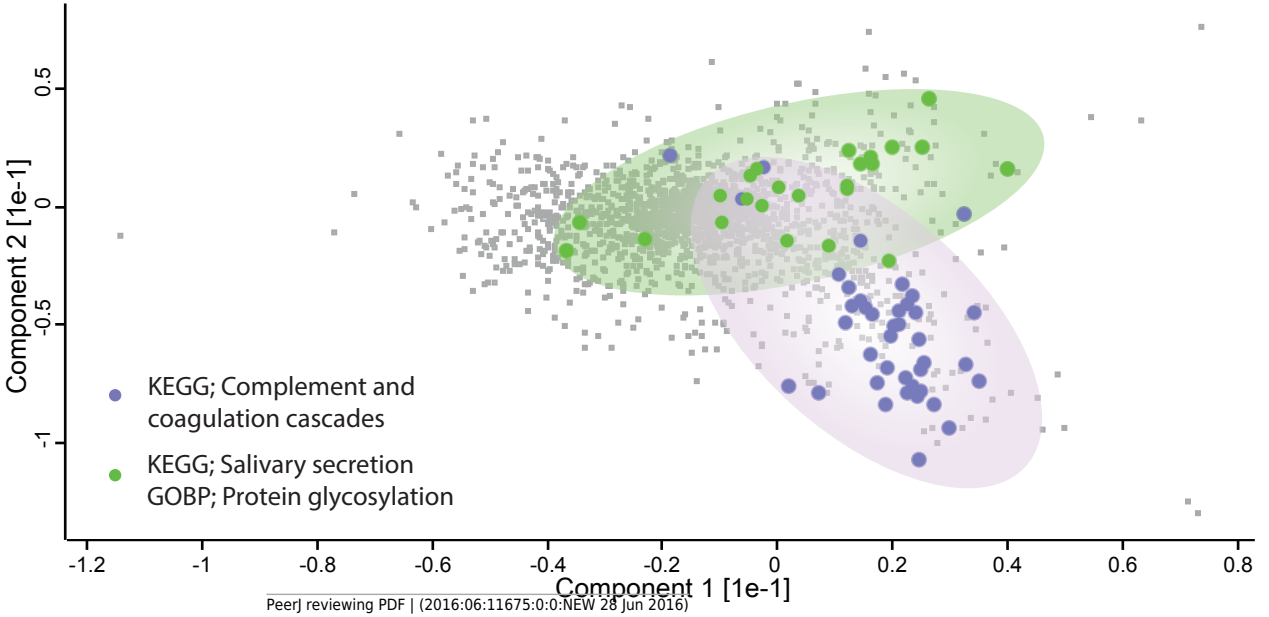

Component 2 [1e-1]
Component 1 [1e-1]
KEGG; Complement and coagulation cascades
KEGG; Salivary secretion
GOBP; Protein glycosylation

Figure 3(on next page)

Figure 3

Potential biomarkers of oral health and disease: Intensity-based heat-map of proteins significantly differentially expressed between the three groups. Protein names in purple are associated with innate immune response, protein names in orange are associated with lipid transfer.

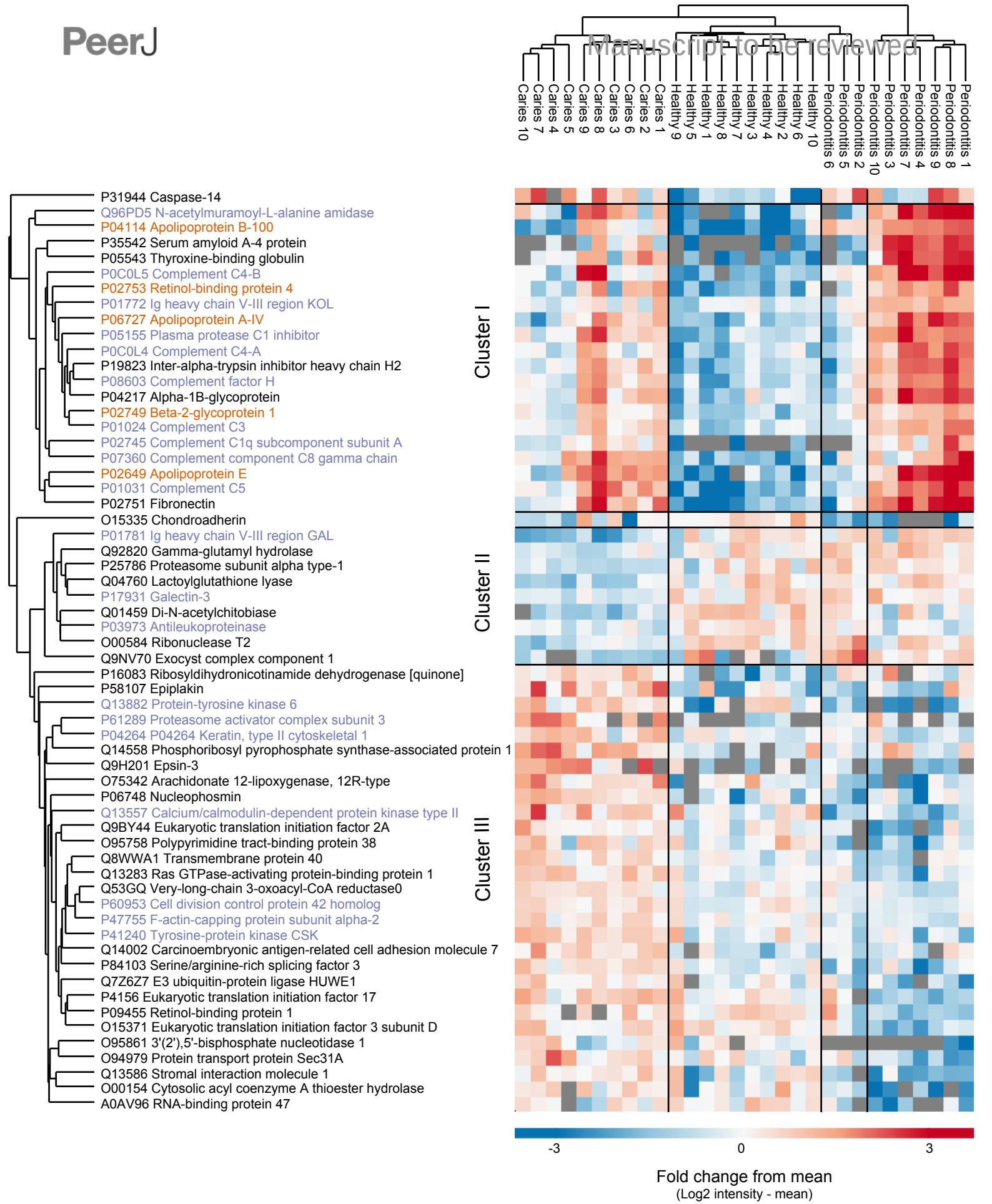
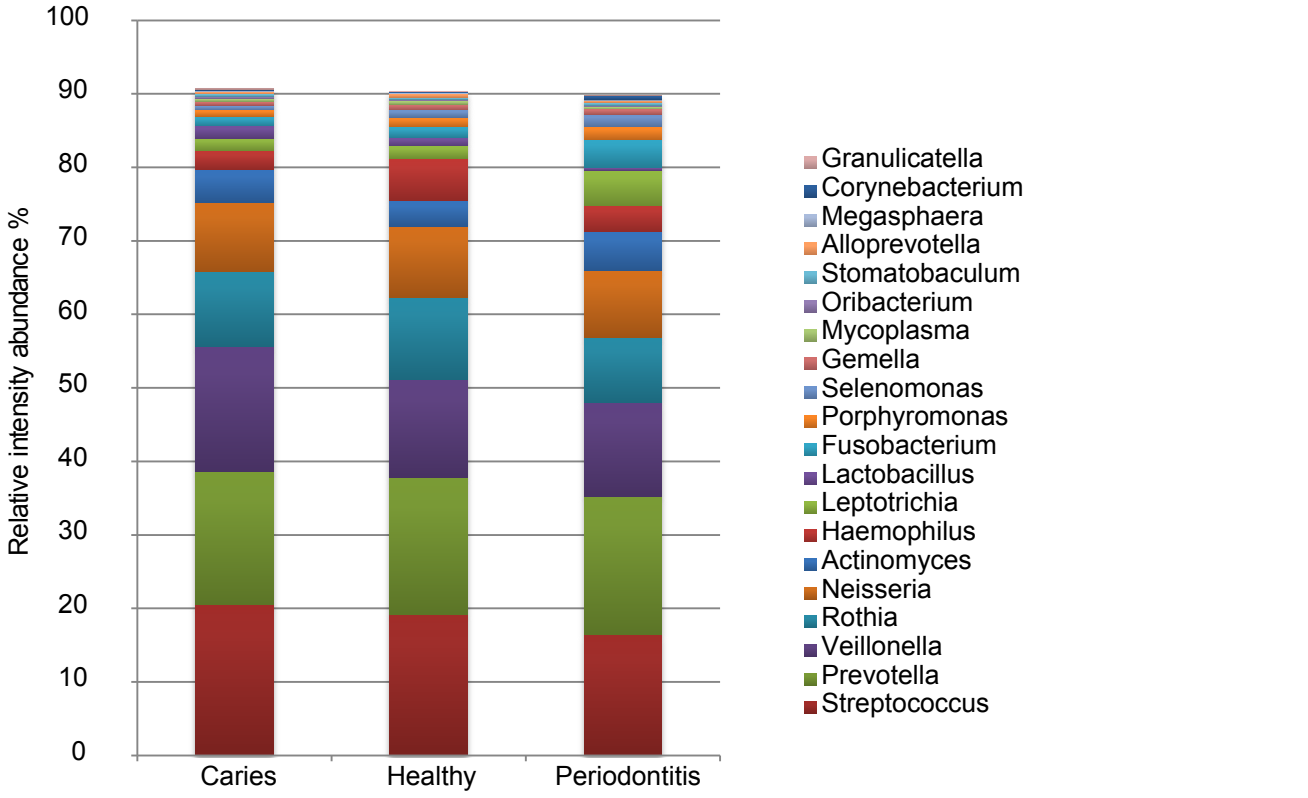


Figure 4(on next page)

Figure 4

Predominant bacterial genera and species: Relative protein intensity abundance of top 20 genera A) and species B).

A


100
90
80
70
60
50
40
30
20
10
0
Relative intensity abundance %
Caries
Healthy
Periodontitis
Granulicatella
Corynebacterium
Megasphaera
Alloprevotella
Stomatobaculum
Oribacterium
Mycoplasma
Gemella
Selenomonas
Porphyromonas
Fusobacterium
Lactobacillus
Leptotrichia
Haemophilus
Actinomyces
Neisseria
Rothia
Veillonella
Prevotella
Streptococcus

B

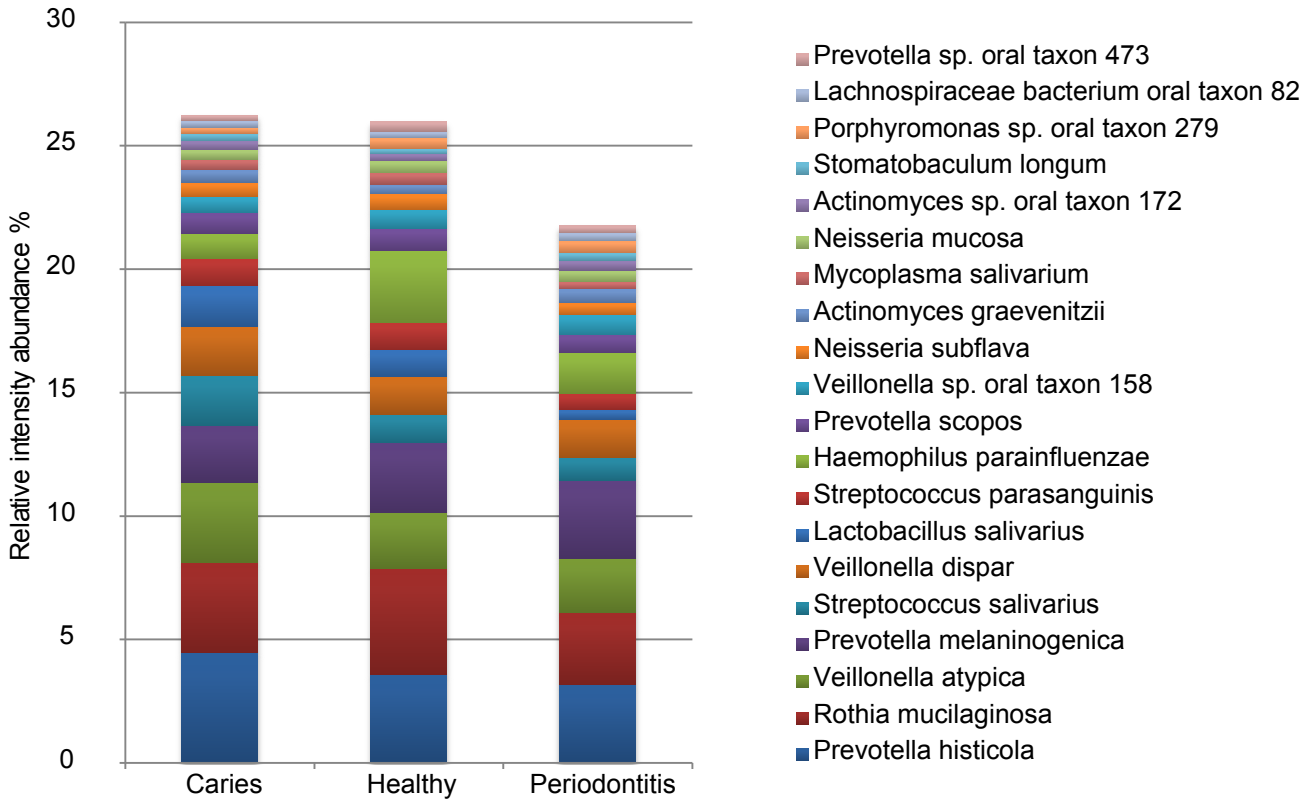

30
25
20
15
10
5
0
Relative intensity abundance %
Caries
Healthy
Periodontitis
Prevotella sp. oral taxon 473
Lachnospiraceae bacterium oral taxon 82
Porphyromonas sp. oral taxon 279
Stomatobaculum longum
Actinomyces sp. oral taxon 172
Neisseria mucosa
Mycoplasma salivarium
Actinomyces graevenitzii
Neisseria subflava
Veillonella sp. oral taxon 158
Prevotella scopos
Haemophilus parainfluenzae
Streptococcus parasanguinis
Lactobacillus salivarius
Veillonella dispar
Streptococcus salivarius
Prevotella melaninogenica
Veillonella atypica
Rothia mucilaginosa
Prevotella histicola

Table 1(on next page)

Table 1

Overview of proteins identified

Number of Proteins	Caries	Healthy	Periodontitis	Total*
Other**	125	115	120	125
Human	2084	2079	2084	2090
Bacteria	1861	1926	1924	1946
-mapped to genus level	1710 (91.9%)	1765 (91.6%)	1762 (91.6%)	1784 (91.7%)
-mapped to species level	594 (31.9%)	602 (34.1%)	609 (34.6%)	616 (34.5%)
Total	4070	4120	4128	4161

*Unique proteins.

** Food related proteins and proteins that could not be assigned to kingdom level.