Two dimensions of chemical variation of the human microbiome across body sites and in COVID-19 patients

Jeffrey M. Dick

Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring of Ministry of Education, School of Geosciences and Info-Physics, Central South University, Changsha, 410083, China.

E-mail: jeff@chnosz.net

Abstract

A better understanding of dysbiosis is a major goal of human microbiome studies, but more knowledge about chemical effects on microbial communities is needed. Oxidation-reduction and hydration-dehydration reactions are chemical processes that are important for physiological functions and, it is hypothesized here, may also influence the elemental composition of microbial proteins. Chemical metrics of biomolecules relevant to these processes are carbon oxidation state (Z_C) and stoichiometric hydration state (n_{H_2O}). I calculated these metrics for protein sequences derived from microbial genomes (multiplied by 16S rRNA-based taxonomic abundances to obtain community reference proteomes), shotgun metagenomes, and metaproteomes. Metaproteomes of gut communities are reduced (i.e., have lower Z_C) compared to oral communities. In contrast, community reference proteomes have lower n_{H_2O} in gut compared to nasal, skin, and oral communities, and metagenomes for gut and oral communities exhibit the same trend. The chemical differences for metaproteomes may be explained by physiological adjustment of protein expression levels to anaerobic, reducing conditions in the gut, whereas metagenomes and reference proteomes may reflect evolutionary adaptation to dehydrating conditions brought on by intestinal absorption of water. Community reference

proteomes, metagenome-assembled genomes (MAGs), and metaproteomes compiled from various studies yield a common trend of more reduced proteins in gut communities of COVID-19 patients compared to controls. These chemical differences imply more reducing conditions in the guts of COVID-19 patients, a finding that contrasts with oxidative conditions that have been previously associated with dysbiosis in inflammatory bowel disease and HIV infection. These results reveal how the human microbiome is shaped by multiple chemical factors over a range of timescales and suggest a new strategy for using multi-omics data to infer changes in gut redox conditions in COVID-19 patients.

Keywords: Multi-omics, chemical analysis, protein evolution, internal environment, dehydration reactions, gut redox conditions, COVID-19

Introduction

The microbiota that inhabit human body sites are major contributors to health and disease [1]. The gut has the largest population of bacterial cells, which aid digestion by providing metabolic functions that are absent from the human genome [2]; moreover, a healthy gut microbiome benefits the host by producing anti-inflammatory molecules and has a barrier function against pathogen infection [3, 4]. The taxonomic composition of the healthy microbiome varies across body sites [5] and is disrupted in some diseases, a condition known as dysbiosis. Coronavirus disease 2019 (COVID-19) is primarily a respiratory disease, but the causative agent (SARS-CoV-2) also infects the gastrointestinal tract [6], and population-based studies have revealed dysbiosis of the gut microbiota in COVID-19 patients [7]. Dysbiosis associated with COVID-19 has been proposed to impair oxygen sensing and contribute to hypoxemia (i.e., blood oxygen concentrations that are lower than the normal physiological range) [8]. However, establishing causal linkages between dysbiosis and disease is challenging [1], and new insight is needed into the interactions between microbial communities and host-associated environments.

One promising but underexplored approach is based on patterns of elemental abundance. For instance, environmental oxygen levels are potential drivers of elemental variation of protein sequences [9, 10]. This is relevant from a clinical

standpoint because oxygen concentration differs among anatomical sites in normal conditions and is altered in several diseases [11, 12]. The skin and nares are sites with more aerobic niches compared to interior sites such as the oral cavity and gut [10], and O_2 concentrations in physiologically normal tissues are lower than those in standard culture conditions [13]. Geochemical variables in the external environment influence the evolution and assembly of free-living communities, but how analogous variables in the internal environment of the body may affect host-associated communities has received far less attention [14]. Therefore, several remarks on the conceptual background for a chemical analysis of multi-omics data are timely.

First, the eco-evolutionary context of different meta-omics techniques is crucial. Communities assemble and mature at timescales of human lives, but microbial lineages have co-evolved with the ancestors of modern humans for millions of years [15, 16]. It follows that the chemical differences of proteins encoded by genomes and metagenomes provide insight into adaptation on evolutionary timescales, while metaproteomes reflect changes that occur over a range of timescales. That is, protein expression levels depend on host and microbial physiology, but the possible protein sequences are still constrained by evolution.

Second, the structure of ecosystems can be represented in terms of not only taxonomic but also chemical attributes [17]. Many previous studies have performed taxonomic and functional characterization of metagenomic DNA sequences but have left out a chemical analysis. Compared to DNA, proteins have a richer repertoire of chemical variability (20 amino acids compared to 4 nucleotides); in addition, because of their abundance in cells, proteins rather than DNA make a larger contribution to cellular energy demand due to environmental variation. Therefore, all multi-omics data in this study are presented in terms of chemical metrics of proteins.

Third is the hypothesis that protein sequences have evolved toward incomplete metastable equilibrium in an environmental context during microbe-host coevolution

over millions of years. I refer to this as the chemical adaptation hypothesis. This hypothesis acknowledges two key features of biomolecular systems: 1) metastability: in most environments (ultramafic-hosted submarine hot springs are one exception [18]), proteins are unstable with respect to amino acids and small inorganic molecules, so the comparisons here are limited to metastable protein molecules compared with each other, and 2) incomplete equilibrium: the postulated thermodynamic effects are just one of many factors that structure microbial ecosystems; other factors include enzyme function, kinetics, metabolic networks, and community assembly. Accordingly, the chemical adaptation hypothesis taken on its own only states how environmental gradients, in the absence of other effects, may explain chemical differences among protein sequences.

Fourth and finally, the choice of chemical variables to analyze should be guided by biological considerations. Redox state, which includes not only oxygen tension but other measures of redox disequilibria, is an important variable for host and microbial physiology; as emphasized by Pfister et al., "the redox environment underpins all microbial metabolisms and should be a key component of research across systems" [14]. A previous analysis of the human microbiome was made in terms of O content and C:O ratios of proteins [10]. However, the oxidation state of carbon in organic molecules is affected not only by C–O but also by C–H and C–S bonds [19], and an informative metric based on this reasoning is the average oxidation state of carbon (Z_C) [20]. Next, I identified hydration state as another important variable. Water is a substrate for numerous metabolic reactions (more than even oxygen) [21], tissue water content in mammals decreases during development from embryo to adult [22, 23] and differs between tumors and normal tissue [24], eukaryotic cells gain or lose water during transitions to cellular proliferation and dormancy [25, 26], intestinal absorption of water from chyme is a major function of the mammalian gut [27], and some microbes can tolerate extremes of hydrostatic pressure and osmotic

strength, both of which affect biomolecular hydration state monitored by biophysical techniques [28]. In contrast to detecting physical interactions of water molecules, stoichiometric hydration state (n_{H_2O}) is a chemical metric that is calculated by balancing chemical reactions using the elemental composition of proteins (see below and [29]). Associations between proteomic n_{H_2O} and the environment or physiology are evident in various systems, including 1000 km-scale salinity gradients, cancer compared to normal tissue, and growth stages of *Bacillus subtilis* biofilms [29–31].

Research in geochemical biology investigates the chemical and thermodynamic links between geochemical environments and genome evolution [32]. As an extension of geochemical biology, this study asks whether chemical conditions in human body sites and disease may influence the eco-evolutionary and physiological processes of microbial communities. Thermodynamics predicts that reducing conditions favor the evolution of reduced molecules and that dehydrating conditions favor the evolution of dehydrated molecules. The first aim of this study is to test these predictions by comparing chemical metrics for proteins calculated from multi-omics data with the chemical conditions of body sites. These outcomes of these tests establish the relevance of the chemical adaptation hypothesis to the human microbiome. This justifies the second aim of the study, which is to calculate chemical metrics for proteins in COVID-19 patients in order to infer changes in the internal chemical environment of the body.

Results

Two chemical metrics

Average oxidation state of carbon (Z_C) in organic molecules in general can be obtained by counting electrons in Lewis structures [33]. For molecules in which the heteroatoms O and S are only bonded to C or H and not to one another (this includes the standard amino acids and primary sequences of proteins but not some

post-translational modifications such as disulfide bonds, none of which are considered here), Z_C can be calculated from the elemental formula [20]. Counting water molecules, instead of electrons, is needed to calculate hydration state, so its formulation requires writing chemical reactions. Stoichiometric hydration state (n_{H_2O}) is the number of water molecules in a theoretical reaction to form a protein from thermodynamic components, normalized by the number of amino acid residues. This definition leverages the notion of thermodynamic components in order to project a vector of elemental composition (for the five elements C, H, N, O, and S) into a vector of chemical composition with the same number of components. The specific components – also known as basis species – used in this study are glutamine, glutamic acid, cysteine, H_2O, and O_2, which have been referred to as the QEC basis species [29]. The inclusion of amino acids rather than simple inorganic species permits a better separation of trends of oxidation and hydration state among proteins in representative genomes [31]; these particular amino acids were chosen because they are relatively abundant and highly connected metabolites and yield a relatively low covariation between n_{H_2O} and Z_C [29].

I calculated chemical metrics of reference proteomes for genera in the "List of Prokaryotes according to their Aerotolerant or Obligate Anaerobic Metabolism" taken from Table S1 of [34]. There is significantly higher Z_C for aerotolerant organisms compared to obligate anaerobes, but no significant difference for n_{H_2O} (Fig. 1A). This shows that evolutionary adaptation to oxygen is reflected more strongly in changes of oxidation state than hydration state of proteins.

Differences between metaproteomes and community reference proteomes

The Genome Taxonomy Database (GTDB) takes advantage of the recent growth of metagenome-assembled genomes (MAGs) as well as culture-based studies to develop

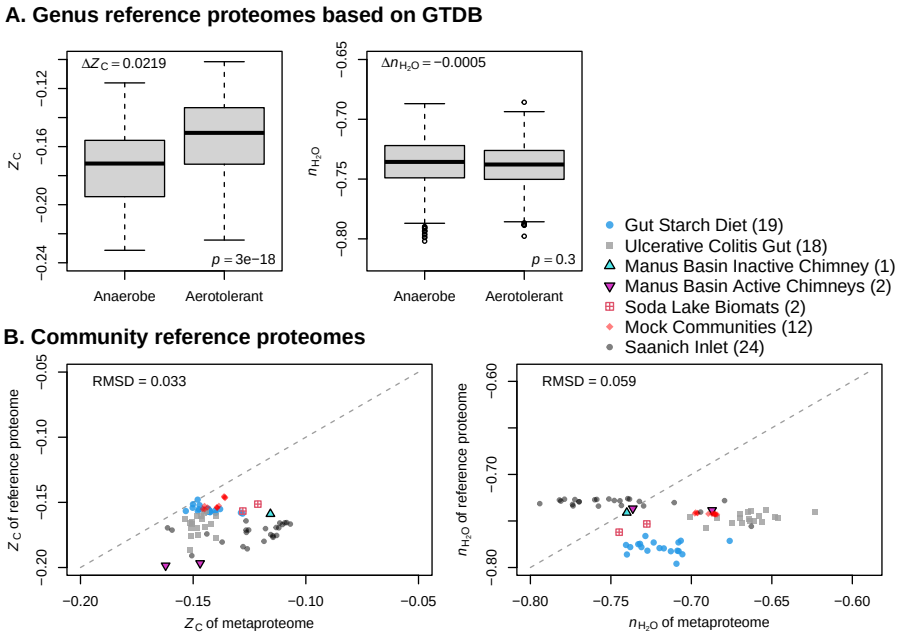

Fig. 1 Chemical metrics of reference proteomes as a function of oxygen tolerance and comparison with metaproteomes. (**A**) Carbon oxidation state (Z_C) and stoichiometric hydration state (n_{H_2O}) for reference proteomes of prokaryotic genera classified as obligately anaerobic or aerotolerant (classifications from [34], their Table S1). Delta values denote mean differences of Z_C or n_{H_2O} between anaerobic and aerotolerant genera. (**B**) Comparison of chemical metrics for metaproteomes and community reference proteomes. Theoretical 1:1 correspondence is indicated by dashed gray lines, and the number of samples in each dataset is listed in the legend.

a phylogenetically consistent taxonomy [35]. In this study, the GTDB was used to retrain the RDP Classifier and to generate reference proteomes for taxa (see Methods). Taxonomic classifications from high-throughput 16S rRNA data were then combined with the reference proteomes for all identified taxa in order to obtain the amino acid compositions of community reference proteomes.

I analyzed several published datasets that provide both 16S rRNA gene sequences and metaproteomes for the same samples. The higher Z_C for metaproteomes than for community reference proteomes (Fig. 1B) may be due to higher natural abundance of cytoplasmic than membrane proteins as well as bias against detection of membrane proteins in metaproteomic protocols (see Discussion). Furthermore, there is a

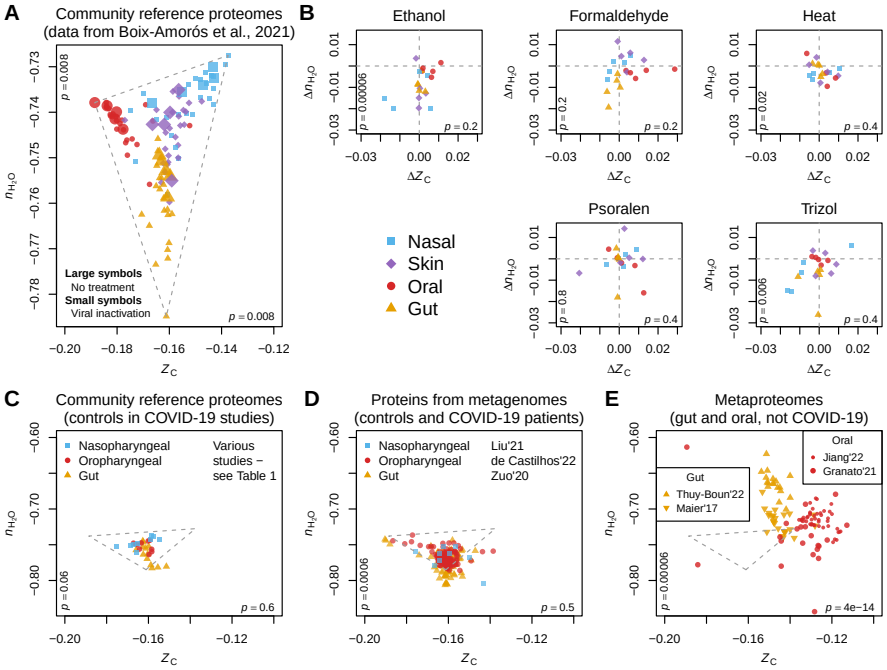

Fig. 2 Chemical variation of microbial proteins across body sites, after viral inactivation treatment, and multi-omics comparison. (**A**–**B**) Z_C and n_{H_2O} of community reference proteomes for body sites, based on 16S rRNA gene sequences from [36]. (A) Chemical metrics for nasal, oral, skin, and gut sites. Large symbols indicate untreated samples; small symbols indicate samples after viral inactivation treatment. (B) Differences of chemical metrics (ΔZ_C and Δn_{H_2O}) between untreated and viral-inactivated samples from the same subjects. (**C**–**E**) Chemical metrics for community reference proteomes, metagenomes, and metaproteomes based on data from other studies. (C) Community reference proteomes based on 16S rRNA data for controls in COVID-19 studies (Table 1). (D) Proteins predicted from metagenomes for controls and COVID-19 patients: nasal [37], oral [38], and gut [39]. (E) Metaproteomes for controls and patients in non-COVID-19 studies: ulcerative colitis [40] and dietary resistant starch [41] for gut microbiomes and oral cancer [42] and lung cancer [43] for oral microbiomes. Each point represents a sample, except for (C), where each point represents the mean of samples in a single dataset. For visual comparison, the dashed triangle representing the convex hull around the points in (A) is replicated in (C–E). *p*-values on the horizontal and vertical axes were calculated for Z_C and n_{H_2O}, respectively. *p*-values in (B) were calculated using the Wilcoxon paired test for samples from all body sites; *p*-values in other panels were calculated using the Wilcoxon unpaired test comparing data for gut and oral or oropharyngeal samples only.

large spread in n_{H_2O} of metaproteomes that is not captured by community reference proteomes.

Chemical variation of the human microbiome across body sites

The 16S rRNA dataset of Boix-Amorós et al. [36] includes samples from four body sites (nasal, skin, oral, and fecal communities); furthermore, community DNA from

each body site was sequenced before and after treatment with viral inactivation by ethanol, formaldehyde, heat, psoralen, or trizol. Considering reference proteomes for untreated communities, oral and nasal sites exhibit the lowest and highest ranges of Z_C compared to other body sites, and the skin and gut have intermediate Z_C whereas the gut has lower n_{H_2O} than other body sites (Fig. 2A). Ethanol treatment is associated with a shift to lower n_{H_2O} compared to untreated samples for communities from most body sites except the mouth (Fig. 2B). A milder signal of dehydration in community reference proteomes is apparent for heat and trizol treatment, but not for formaldehyde or psoralen.

Next, I asked whether these trends are recapitulated in independent 16S rRNA datasets from other studies as well as metagenomes and and metaproteomes. I calculated chemical metrics for community reference proteomes derived from 16S rRNA data for control subjects in COVID-19 studies (sources of data are listed Table 1). The few datasets for gut communities that fall outside of the triangular area that circumscribes values for the Boix-Amorós et al. dataset exhibit relatively low n_{H_2O} (Fig. 2C). Furthermore, proteins coded by metagenomes for various body sites of controls and COVID-19 patients largely overlap the range of chemical metrics for community reference proteomes; the main exception is again lower n_{H_2O} of gut communities (Fig. 2D). Finally, selected metaproteomes for gut communities have significantly lower Z_C than those from oral communities (Fig. 2E; none of these studies were for COVID-19). The distinct ranges n_{H_2O} of gut metaproteomes from two studies [40, 41] may be due to technical differences, so the apparently significant difference from oral metaproteomes is excluded from further discussion.

Chemical variation of the microbiome in COVID-19 patients

I analyzed publicly available 16S rRNA gene sequences for oral or oropharyngeal, nasopharyngeal, and gut (i.e., fecal) samples in COVID-19 studies (Table 1). For

Table 1 Sources of 16S rRNA gene sequence data from COVID-19 studies. n_{Control} and $n_{\text{COVID-19}}$ are the numbers of samples in control and COVID-19 groups, and numbers in the left column correspond to point symbols in Fig. 3A.

Fig. 3A	Reference	BioProject	n_{Control}	$n_{\text{COVID-19}}$
Nasopharyngeal				
1	[44]	PRJNA777915	38	38
2	[45]	PRJNA704967	10	9
3	[46]	PRJNA707350	26	63
4	[47]	PRJNA726205	7	76
5	[48]	PRJNA683617	34	215
6	[49]	PRJNA774098	12	21
7	[50]	PRJNA714242	10	32
8	[51]	PRJNA726992 and PRJNA726994	20	75
9	[52]	PRJNA673585	18	56
Oral or oropharyngeal				
1	[53]	PRJNA780671	24	30
2	[54]	PRJNA739539	140	166
3	[55]	PRJNA692359	14	25
4	[56]	PRJNA669421	54	46
5	[48]	PRJNA683617	33	220
6	[57]	PRJNA767939	15	22
7	[58]	PRJNA660302	150	242
8	[59]	PRJNA684070	44	50
9	[60]	PRJNA639286	24	97
Gut				
1	[61]	PRJNA703303	30	26
2	[62]	PRJNA758913	38	141
3	[63]	PRJNA636824	30	30
4	[64]	PRJNA705797	14	25
5	[65]	PRJDB11949 and PRJDB12349	61	45
6	[66]	PRJNA734646	18	57
7	[67]	PRJNA678695	34	50
8	[57]	PRJNA767939	15	22
9	[58]	PRJNA660302	72	100
10	[68]	PRJNA753792	22	89
11	[69]	PRJNA756849	145	104
12	[59]	PRJNA684070	32	17
13	[70]	PRJNA736160	14	15

the naso- and oropharyngeal/oral studies, various datasets exhibit either positive or negative mean differences of chemical metrics of community reference proteomes between COVID-19 patients and controls, with no clear trend in the data (Fig. 3A). Although the majority of datasets (6 of 9) for oropharyngeal communities show lower mean Z_C in COVID-19 patients, the difference is not significant for all datasets combined. In contrast, a large majority of datasets (10 of 13) for gut communities exhibit a lower mean Z_C in COVID-19 patients and the difference is significant. It is also

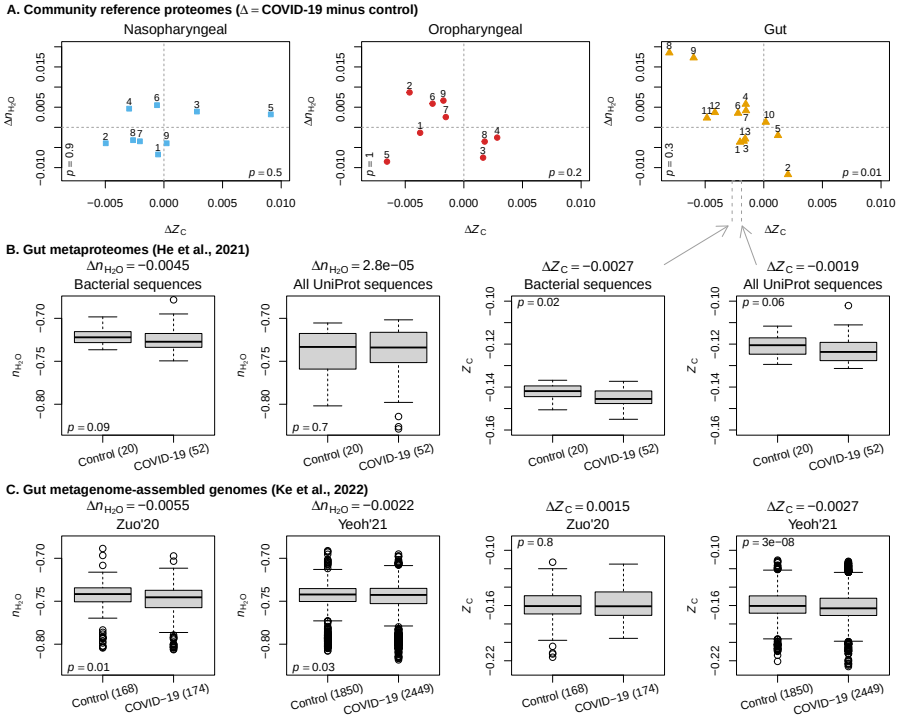

Fig. 3 Differences of chemical metrics of nasopharyngeal, oropharyngeal, and gut microbiomes between controls and COVID-19 patients. (**A**) Community reference proteomes computed from the 16S rRNA gene sequence datasets listed in Table 1. Each point represents the mean difference between controls and COVID-19 positive subjects in a single dataset. *p*-values (Wilcoxon paired test) are shown for differences of Z_C and n_{H_2O} on the horizontal and vertical axes, respectively. (**B**) Bacterial metaproteomes in controls and COVID-19 positive subjects. Metaproteomic data used to make this plot are from [71]; I filtered the list of identified proteins to use UniProt IDs for bacteria or all UniProt IDs. The arrows indicate the mean differences of metaproteomic Z_C plotted on the horizontal axis in (A). (**C**) MAGs provided by [72] based on metagenomic data originally reported by Zuo et al. [39] and Yeoh et al. [73]. In (B) and (C), *p*-values were calculated using the unpaired Wilcoxon rank-sum test; boxplots show the median (thick line), first and third quartiles (box), most extreme values within 1.5 × the interquartile range away from the box (whiskers), and outliers (points).

notable that the most extreme points correspond to large negative differences; four gut datasets have $\Delta Z_C < -0.003$ but no gut dataset has $\Delta Z_C > 0.003$.

Metaproteomic data for gut communities in controls and COVID-19 patients were reported by He et al. [71]. Carbon oxidation state for gut metaproteomes is significantly lower in COVID-19 patients than controls when considering only bacterial sequences, and marginally lower with consideration of all mapped UniProt sequences (this includes bacterial and human sequences) (Fig. 3B).

Finally, I analyzed the metagenome-assembled genomes (MAGs) provided by Ke et al. [72], which are in turn based on metagenomic data originally reported by Zuo et al. [39] and Yeoh et al. [73]. Both sets of MAGs indicate significantly lower protein n_{H_2O} for gut communities in COVID-19 compared to controls; this trend agrees with the non-significant differences of n_{H_2O} for the metaproteomes. The MAGs based on the data of Zuo et al. and Yeoh et al., respectively, show higher and lower protein Z_C for gut communities in COVID-19 patients, but only the differences computed from the Yeoh et al. dataset (which includes substantially more MAGs) are significant (Fig. 3C).

Discussion

Metagenomes and community reference proteomes (which are derived from genomes) only indicate the sequences of possibly expressed proteins, but metaproteomes identify some of the proteins that are actually present in a sample. The analysis of metaproteomes is not necessarily the best choice for all purposes in biology, because each -omics technique provides a different representation of biological processes. Specifically, the chemical metrics calculated from (meta)genomically coded proteins represent chemical differences that occur due to evolution and ecology, while metaproteomes include the effects of both eco-evolutionary and physiological processes.

On the whole, community reference proteomes and proteins predicted from metagenomes show similar ranges of Z_C and n_{H_2O} (see Fig. 2D and [74]), but metaproteomes are relatively oxidized (Fig. 1C–E and Fig. 2E). Two factors may explain this result. First, at least in *Escherichia coli*, cytoplasmic proteins are more abundant than membrane proteins [75]. Second, some metaproteomic workflows are biased against extraction and detection of membrane proteins [76]. Because membrane proteins tend to be enriched in hydrophobic amino acid residues, they have

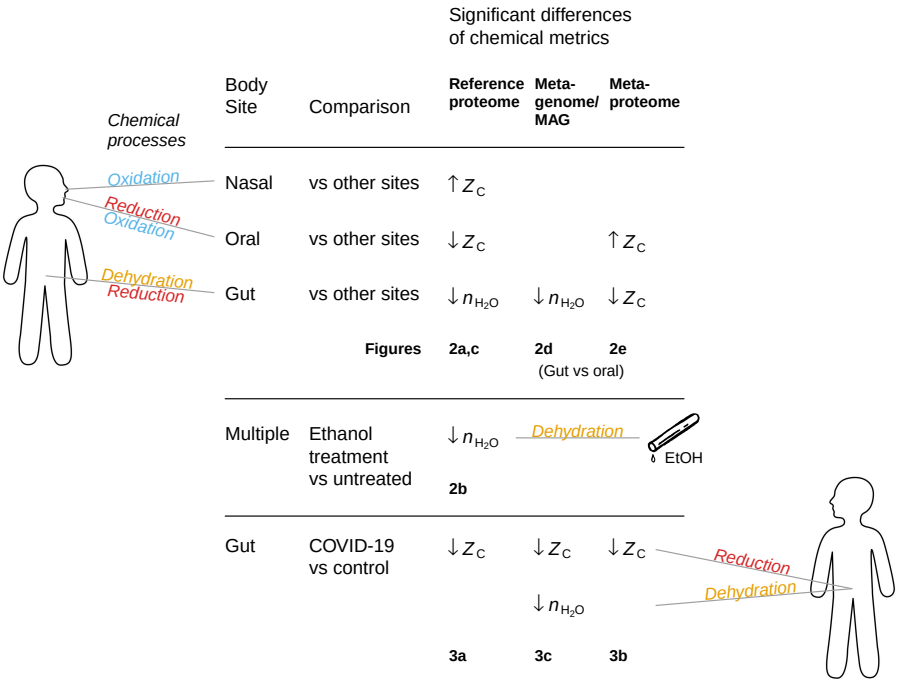

Fig. 4 Overview of chemical variation of the human microbiome inferred from multi-omics datasets. The indicated figures show significant differences of chemical metrics for proteins. These differences are manifestations of chemical processes that alter the elemental composition of proteins through evolution and ecology (for proteins identified by all -omics techniques) as well as physiology (for metaproteomes).

lower Z_C than cytoplasmic proteins [20], and a low proportion of membrane proteins in cells or technical biases against their detection would both contribute to higher metaproteomic Z_C compared to community reference proteomes and metagenomes.

Body sites

The main findings of this study are summarized in Fig. 4. In the first chemical dimension, the finding that community reference proteomes for skin and nasal communities are more oxidized than those for oral and gut communities (see also Fig. 2A) is congruent with previous analyses of metagenomic data from the Human Microbiome Project (HMP) by this laboratory [74] and another group [10]. This finding supports the notion that oxygenated habitats on the skin and in the airway are environmental filters for the evolution and growth of bacteria with genomes that code for relatively

oxidized proteins. Community reference proteomes for oral communities are more reduced than those for the gut, but metaproteomes show the opposite trend (Fig. 2A and E). Although oxygen is depleted locally in some oral habitats, the lumen of the large intestine has the lowest oxygen concentrations within the body [13, 14]. It therefore appears that low oxygen concentrations in the gut may have a larger chemical effect on protein expression (detected by metaproteomes) than on protein evolution (represented by community reference proteomes and metagenomes).

In the second chemical dimension, community reference proteomes are more dehydrated for gut communities compared to other body sites and are more dehydrated for samples treated with ethanol compared to untreated samples. The intestine has a physiological function of water absorption [27], and the dehydrating effect of ethanol is described in protocols for preparing tissue sections for histological analysis [77]. In line with metagenomic trends in salinity gradients [29], these results supply additional evidence for chemical adaptation of proteins at evolutionary timescales to dehydrating forces in microbial habitats.

COVID-19

It was previously hypothesized that dysbiosis in inflammatory bowel disease (IBD) may be driven by higher oxygen levels in the gut [12]. Likewise, gut communities in HIV patients exhibit higher abundances of aerotolerant organisms [78], and dysbiosis in several pathological conditions (but not COVID-19) has been linked with greater availability of electron acceptors [14]. In contrast to the oxidative trend that has been described for dysbiosis in IBD and HIV, the analysis of multi-omics datasets for COVID-19 in this study indicates a shift toward more reduced proteins. According to the chemical adaptation hypothesis, environmental selection for genomes with more reduced proteins requires relatively reducing, rather than oxidizing conditions. Chemical analysis of multi-omics datasets for IBD and HIV is beyond the scope of

this study, but should be performed to determine whether trends of protein Z_C for dysbiotic communities in those conditions differ from COVID-19.

These findings have important implications for understanding how gut communities respond to pathological oxygen concentrations. The intestine maintains a steep radial gradient of O_2 between the vascularized mucosa and anaerobic lumen [79]. If the radial O_2 gradient in the intestine is sensitive to blood oxygen levels, then the occurrence of hypoxemia in some COVID-19 patients may help explain the identification of relatively reduced proteins in multi-omics datasets for gut communities. However, electron acceptors other than oxygen should also be examined for a more complete understanding of the effects of the redox environment on gut dysbiosis [14].

In the second chemical dimension, proteins coded by MAGs exhibit lower n_{H_2O} for gut communities in COVID-19 patients. This provides evidence for chemical adaptation through eco-evolutionary processes to dehydrating conditions, but it is too early to tell if chemical variation of the gut microbiome can be mechanistically connected to the pathophysiology of COVID-19 that may include whole-body dehydration [80, 81]. The absence of this trend from community reference proteomes suggests possible contributions from horizontal gene transfer or species- or strain-level variability that cannot be detected by taxonomic classification of 16S rRNA sequences.

Limitations

Micro-scale gradients of oxygen concentration may be key determinants of microbiome composition [14], but the macro-scale datasets analyzed in this study do not permit resolving these features. Furthermore, the analysis in this study only compares controls and COVID-19 patients without classification according to disease severity (although some source datasets include metadata for disease severity), and no longitudinal data were analyzed here.

This analysis inherits other limitations intrinsic to the source studies, including all types of bias that affect amplicon-based, metagenomic, and metaproteomic studies [76, 82]; no corrections for bias were applied. Furthermore, cross-comparisons between different data types, such as metaproteomes and metagenomes, compounds the bias intrinsic to each experimental technique.

This study does not provide clinical guidance. Other research is needed to reveal how dietary or other interventions may alter the body's internal chemical environment and interact with microbiota (e.g. [83]).

Materials and Methods

16S rRNA data processing

Publicly available 16S rRNA gene sequencing datasets were downloaded from the NCBI Sequence Read Archive (SRA). FASTQ sequence files for paired-end sequences were merged using the "fastq_mergepairs" command of VSEARCH version 2.15.0 [84]. Quality filtering was done with maximum expected error rate of 0.005 and a sequence truncation length specific for each dataset (see Table S1 for details and sequence processing statistics). Singletons were removed and remaining sequences were subsampled at a depth of 10000. Reference-based chimera detection with the VSEARCH command "uchime_ref" was performed using the SILVA SSURef NR99 database version 138.1 [85].

To generate reference proteomes of taxa, species representatives for bacterial and archaeal genomes with at least 500 proteins were taken from the GTDB (release 207). Within each genome, the amino acid composition of all proteins was summed and divided by the number of proteins. Then, amino acid compositions for all genomes in each genus were summed and divided by the number of genomes to generate the amino acid compositions of genus reference proteomes. Similarly, amino acid compositions for all genomes in each family, order, class, and phylum were summed

and divided by the number of genomes to generate the amino acid compositions of reference proteomes for taxa at those ranks.

Taxonomic classification was performed using the RDP Classifier version 2.13 [86] retrained with the GTDB database (bac120_ssu_reps and ar53_ssu_reps in GTDB release R07-RS207). The lowest-level taxonomic classification (from genus to phylum) for each sequence was used; the numbers of taxonomic classifications were multiplied by the amino acid compositions of reference proteomes for taxa and summed to obtain the amino acid composition of the community reference proteome.

Metaproteomic data processing

For datasets with an available study-specific protein sequence database, amino acid compositions of those proteins were multiplied by protein abundances to calculate the amino acid composition of the metaproteome. For other datasets, protein sequences were obtained by using IDs for either UniProt (last accessed on 2022-08-31) or the Human Oral Microbiome Database (HOMD) [87] version 9.15 (updated date: 2022-02-07; last accessed on 2023-02-03). For processing UniProt IDs, the UniProt ID mapping tool was used with the taxonomy filter set to include only bacterial sequences (in order to exclude human proteins), and amino acid composition was computed from canonical protein sequences. For processing HOMD IDs, sequence files in the PROKKA directory were used (Genomes Annotated with PROKKA 1.14.6). Specific processing steps for each dataset are described below. Database accessions starting with PXD are ProteomeXchange identifiers [88].

Metaproteomes for comparison with community reference proteomes

Protein sequences were obtained from the UniProt database for the following metaproteomic datasets. For these datasets, acid compositions of proteins identified in each sample were summed (without abundance information) to obtain the amino acid composition of the metaproteome. *Gut Starch Diet* [41]: Metaproteomic

abundances and UniProt IDs were obtained from https://zenodo.org/record/838741. Additional mapping to the UniParc database was used to retrieve obsolete sequences. *Ulcerative Colitis* [40]: Protein IDs were extracted from PeptideEvidence fields of *.mzid.gz files on PXD022433. IDs with "." or "_" were omitted in order retain UniProt IDs.

Study-specific sequence databases were used for the following metaproteomic datasets. For these datasets, amino acid compositions of proteins identified in each sample were multiplied by protein abundances (taken from the number of peptide spectral matches) and summed to obtain the amino acid composition of the metaproteome. *Manus Basin Inactive Chimney* [89]: Protein sequence database (DEAD_Chimneys_20170720_nr_fw_rev_cont.fasta) and protein IDs (Mudpit_121204_V2_P6_CH_SM_Chimney_3b_1.pride.mztab.gz) were downloaded from PXD010074. *Manus Basin Active Chimneys* [90]: Protein sequence database (140826_StM_Chimney39surface53surface_fw_rev_cont.fasta) and protein IDs (Mudpit_121026_V2_P6_AO_SM_Chimney_1a_10.pride.mztab.gz for diffuse-flow chimney RMR-D and Mudpit_121028_V2_P6_AO_SM_Chimney_2a_1.pride.mztab.gz for focused-flow chimney RMR5) were downloaded from PXD009105. *Soda Lake Biomats* [91]: Protein sequence database (SodaLakes_AllCombined_Cluster95ID_V2.fasta) and protein IDs (GEM-(01).pep.xml and LCM-(01).pep.xml) were downloaded from PXD006343. *Mock Communities* [91]: Protein sequence database (Mock_Comm_RefDB_V3.fasta) and protein IDs (12 pep.xml files for 3 mock communities with 4 replicates each for 260-minute 1D-LC-MS/MS runs) were downloaded from PXD006118. *Saanich Inlet* [92]: Protein sequence database (SaanichInlet_LP_ORFs_2015-01-13_Filtered.fasta) and protein IDs (SBI_Metagenome2015_AllProteinsAllExperiments.txt; the ScanCount column in this table was used to estimate protein abundance) were downloaded from

PXD004433. Metaproteomic experiment names were mapped to sample names using Table S2 of [92].

Community reference proteomes were based on 16S rRNA data from publications listed above with the exception of [93] for Manus Basin Active Chimneys.

Metaproteomes for body sites and COVID-19 gut samples

For the saliva microbiome from [42], Majority Protein IDs and LFQ intensity for saliva cells were taken from Table S9 of the source publication; data for saliva supernatant were not analyzed here. For each protein, the first Majority Protein ID was used to look up the protein sequence in HOMD. Organism ID SEQF2791 (*Selenomonas* sp. HMT 136) was not found in HOMD at the time of this study, so the second Majority Protein ID was used in this case. For the oral microbiome from [43], Majority Protein IDs and LFQ intensity were taken from the proteinGroups.txt file available downloaded from PXD026727. The first Majority Protein ID was used for each protein, except for some organisms not found in HOMD at the time of this study (SEQF1058 SEQF3075 SEQF1068 SEQF1063 SEQF2480 SEQF2762 SEQF3069 SEQF2791), for which the second Majority Protein ID was used. For both of these datasets, the amino acid composition of each identified protein was multiplied by the LFQ intensity and summed to obtain the amino acid composition of the metaproteome.

For the COVID-19 gut metaproteome from [71], the file "Table S2. Global metaproteome.xlsx" was downloaded from Supplementary Files on ResearchSquare (https://doi.org/10.21203/rs.3.rs-208797/v1). UniProt IDs were extracted from the "Accession" column by keeping values starting with "trl" or "spl". Amino acid compositions of the proteins were multiplied by spectral counts and summed to obtain the amino acid composition of the metaproteome.

Metagenomic data processing

Metagenomes for body sites

For each sequencing run, adapter sequences were removed, sequences were dereplicated, human sequences were removed using bowtie2 version 2.5.0 [94] with the GRCh38 reference database (https://genome-idx.s3.amazonaws.com/bt/GRCh38_noalt_as.zip, accessed on 2022-11-17), rRNA sequences were removed using SortMeRNA version 2.1b [95], and partial protein sequences were predicted using FragGeneScan version 1.18 [96]. The amino acid compositions of all protein sequences in each run were summed and used to calculate chemical metrics. Dataset accession numbers and processing statistics are listed in Table S2.

Metagenome-assembled genomes for COVID-19 patients and controls

To analyze the MAGs generated by Ke et al. [72], which are based on metagenomic data originally reported in [39] and [73] (NCBI BioProjects PRJNA624223 and PRJNA650244), I downloaded the files MAG.zip with nucleotide sequences of MAGs (https://figshare.com/s/a426a12b463758ed6a54, accessed on 2022-10-27) and COVID19_metadata.txt with identification of MAGs from COVID-19 patients and controls (https://github.com/Owenke247/COVID-19, accessed on 2022-10-28). For each MAG, protein sequences were predicted using Prodigal [97], and the amino acid compositions of all predicted proteins were summed and used to calculate chemical metrics.

Chemical metrics

Z_C was calculated from

$$Z_C = \frac{-h + 3n + 2o + 2s}{c} \tag{1}$$

where the letters c, h, n, o, and s denote elemental abundance. n_{H_2O} was calculated from

$$\begin{bmatrix} 5 & 5 & 3 & 0 & 0 \\ 10 & 9 & 7 & 2 & 0 \\ 2 & 1 & 1 & 0 & 0 \\ 3 & 4 & 2 & 1 & 2 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} n_{Gln} \\ n_{Glu} \\ n_{Cys} \\ n_{H_2O} \\ n_{O_2} \end{bmatrix} = \begin{bmatrix} c \\ h \\ n \\ o \\ s \end{bmatrix} \tag{2}$$

where n_{Gln}, n_{Glu}, n_{Cys}, n_{H_2O}, and n_{O_2} are stoichiometric coefficients of a theoretical reaction to form a protein with formula $C_cH_hN_nO_oS_s$, and the stoichiometric matrix represents the number of elements in each of the basis species (glutamine, glutamic acid, cysteine, water, and oxygen). Amino acid compositions of community reference proteomes, metagenome-coded proteins, or metaproteomes were combined with precomputed values of Z_C and n_{H_2O} for amino acid residues (see [29]) in order to calculate the respective chemical metrics for proteins.

Statistics

Wilcoxon tests were used to calculate p-values. Tests were made for unpaired observations except for untreated and viral-inactivated samples from the same subject and body site (Fig. 2B) and means for COVID-19 and control samples from the same study (Fig. 4A), for which paired Wilcoxon tests were used. $p < 0.05$ was considered significant.

Supplementary Information

Table S1. 16S rRNA processing statistics. **Table S2.** Metagenome processing statistics.

Acknowledgements

Thanks to Jie Cao for making the drawings in Fig. 4.

Availability of data and materials

GTDB (release 207) SSU rRNA training files for RDP Classifier are archived at https://doi.org/10.5281/zenodo.7633100. Amino acid compositions of reference proteomes for taxa and functions to generate community reference proteomes are available in the chem16S R package version 0.1.3 (available at https://github.com/jedick/chem16S, archived at https://doi.org/10.5281/zenodo.7633130). RDP Classifier results, scripts for processing metagenomic and metaproteomic data (together with the computed amino acid compositions), and functions to make the plots are available in the sars16S section of the JMDplots R package version 1.2.17 (available at https://github.com/jedick/JMDplots, archived at https://doi.org/10.5281/zenodo.7634341).

Ethics approval and consent to participate

Not applicable.

Competing interests

The author declares no competing interests.

References

[1] Cho, I. & Blaser, M. J. The human microbiome: at the interface of health and disease. *Nature Reviews Genetics* **13** (4), 260–270 (2012). https://doi.org/10.1038/nrg3182 .

[2] Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464** (7285), 59–65 (2010). https://doi.org/10.1038/nature08821 .

[3] Koboziev, I., Webb, C. R., Furr, K. L. & Grisham, M. B. Role of the enteric microbiota in intestinal homeostasis and inflammation. *Free Radical Biology*

and Medicine **68**, 122–133 (2014). https://doi.org/10.1016/j.freeradbiomed.2013.11.008 .

[4] Gomaa, E. Z. Human gut microbiota/microbiome in health and diseases: a review. *Antonie van Leeuwenhoek* **113** (12), 2019–2040 (2020). https://doi.org/10.1007/s10482-020-01474-7 .

[5] Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326** (5960), 1694–1697 (2009). https://doi.org/10.1126/science.1177486 .

[6] Xiao, F. *et al.* Evidence for gastrointestinal infection of SARS-CoV-2. *Gastroenterology* **158** (6), 1831–1833 (2020). https://doi.org/10.1053/j.gastro.2020.02.055 .

[7] Yamamoto, S. *et al.* The human microbiome and COVID-19: A systematic review. *PLOS One* **16** (6), e0253293 (2021). https://doi.org/10.1371/journal.pone.0253293 .

[8] Gopal, A. B. *et al.* Silent hypoxia in COVID-19: a gut microbiota connection. *Current Opinion in Physiology* **23**, 100456 (2021). https://doi.org/10.1016/j.cophys.2021.06.010 .

[9] Acquisti, C., Kleffe, J. & Collins, S. Oxygen content of transmembrane proteins over macroevolutionary time scales. *Nature* **445** (7123), 47–52 (2007). https://doi.org/10.1038/nature05450 .

[10] Vecchio-Pagan, B., Bewick, S., Mainali, K., Karig, D. K. & Fagan, W. F. A stoichioproteomic analysis of samples from the Human Microbiome Project. *Frontiers in Microbiology* **8**, 1119 (2017). https://doi.org/10.3389/fmicb.2017.01119 .

[11] Espey, M. G. Role of oxygen gradients in shaping redox relationships between the human intestine and its microbiota. *Free Radical Biology and Medicine* **55**, 130–140 (2013). https://doi.org/10.1016/j.freeradbiomed.2012.10.554 .

[12] Rigottier-Gois, L. Dysbiosis in inflammatory bowel diseases: the oxygen hypothesis. *The ISME Journal* **7** (7), 1256–1261 (2013). https://doi.org/10.1038/ismej.2013.80 .

[13] Ast, T. & Mootha, V. K. Oxygen and mammalian cell culture: are we repeating the experiment of Dr. Ox? *Nature Metabolism* **1** (9), 858–860 (2019). https://doi.org/10.1038/s42255-019-0105-0 .

[14] Pfister, C. A. *et al.* Conceptual exchanges for understanding free-living and host-associated microbiomes. *mSystems* **7** (1), e01374–21 (2022). https://doi.org/10.1128/msystems.01374-21 .

[15] Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R. & Gordon, J. I. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature Reviews Microbiology* **6** (10), 776–788 (2008). https://doi.org/10.1038/nrmicro1978 .

[16] Moeller, A. H. *et al.* Cospeciation of gut microbiota with hominids. *Science* **353** (6297), 380–382 (2016). https://doi.org/10.1126/science.aaf3951 .

[17] Margalef, R. On certain unifying principles in ecology. *The American Naturalist* **97** (897), 357–374 (1963). https://doi.org/10.1086/282286 .

[18] Dick, J. M. & Shock, E. L. The release of energy during protein synthesis at ultramafic-hosted submarine hydrothermal ecosystems. *Journal of Geophysical Research: Biogeosciences* **126** (11), e2021JG006436 (2021). https://doi.org/10.1029/2021JG006436 .

[19] Visconti, F. & de Paz, J. M. Estimation of the carbon valence from its average formal oxidation state in the soil organic matter. *European Journal of Soil Science* **72** (5), 2225–2230 (2021). https://doi.org/10.1111/ejss.13122 .

[20] Dick, J. M. Average oxidation state of carbon in proteins. *Journal of the Royal Society, Interface* **11**, 20131095 (2014). https://doi.org/10.1098/rsif.2013.1095 .

[21] Frenkel-Pinter, M., Rajaei, V., Glass, J. B., Hud, N. V. & Williams, L. D. Water and life: The medium is the message. *Journal of Molecular Evolution* **89** (1), 2–11 (2021). https://doi.org/10.1007/s00239-020-09978-6 .

[22] Moulton, C. R. Age and chemical development in mammals. *Journal of Biological Chemistry* **57** (1), 79–97 (1923). https://doi.org/10.1016/S0021-9258(18)85527-2 .

[23] Toro-Ramos, T., Paley, C., Pi-Sunyer, F. X. & Gallagher, D. Body composition during fetal development and infancy through the age of 5 years. *European Journal of Clinical Nutrition* **69** (12), 1279–1289 (2015). https://doi.org/10.1038/ejcn.2015.117 .

[24] Penet, M.-F., Kakkad, S., Wildes, F. & Bhujwalla, Z. M. Water and collagen content are high in pancreatic cancer: Implications for quantitative metabolic imaging. *Frontiers in Oncology* **10**, 599204 (2021). https://doi.org/10.3389/fonc.2020.599204 .

[25] Munder, M. C. *et al.* A pH-driven transition of the cytoplasm from a fluid- to a solid-like state promotes entry into dormancy. *eLife* **5**, e09347 (2016). https://doi.org/10.7554/eLife.09347 .

[26] Marakhova, I. *et al.* Intracellular K^+ and water content in human blood lymphocytes during transition from quiescence to proliferation. *Scientific Reports* **9** (1), 16253 (2019). https://doi.org/10.1038/s41598-019-52571-1 .

[27] Leiper, J. B. Fate of ingested fluids: Factors affecting gastric emptying and intestinal absorption of beverages in humans. *Nutrition Reviews* **73** (S2), 57–72 (2015). https://doi.org/10.1093/nutrit/nuv032 .

[28] Knop, J.-M. *et al.* Life in multi-extreme environments: Brines, osmotic and hydrostatic pressure—a physicochemical view. *Chemical Reviews* **123** (1), 73–104 (2023). https://doi.org/10.1021/acs.chemrev.2c00491 .

[29] Dick, J. M., Yu, M. & Tan, J. Uncovering chemical signatures of salinity gradients through compositional analysis of protein sequences. *Biogeosciences* **17** (23), 6145–6162 (2020). https://doi.org/10.5194/bg-17-6145-2020 .

[30] Dick, J. M. Water as a reactant in the differential expression of proteins in cancer. *Computational and Systems Oncology* **1** (1), e1007 (2021). https://doi.org/10.1002/cso2.1007 .

[31] Dick, J. M. A thermodynamic model for water activity and redox potential in evolution and development. *Journal of Molecular Evolution* **90** (2), 182–199 (2022). https://doi.org/10.1007/s00239-022-10051-7 .

[32] Dick, J. M., Boyer, G. M., Peter A. Canovas, I. & Shock, E. L. Using thermodynamics to obtain geochemical information from genomes. *Geobiology* **21** (2), 262–273 (2023). https://doi.org/10.1111/gbi.12532 .

[33] Loock, H.-P. Expanded definition of the oxidation state. *Journal of Chemical Education* **88** (3), 282–283 (2011). https://doi.org/10.1021/ed1005213 .

[34] Million, M. & Raoult, D. Linking gut redox to human microbiome. *Human Microbiome Journal* **10**, 27–32 (2018). https://doi.org/10.1016/j.humic.2018.07.002 .

[35] Parks, D. H. *et al.* GTDB: An ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research* **50** (D1), D785–D794 (2022). https://doi.org/10.1093/nar/gkab776 .

[36] Boix-Amorós, A. *et al.* Viral inactivation impacts microbiome estimates in a tissue-specific manner. *mSystems* **6** (5), e00674–21 (2021). https://doi.org/10.1128/mSystems.00674-21 .

[37] Liu, J. *et al.* Association between the nasopharyngeal microbiome and metabolome in patients with COVID-19. *Synthetic and Systems Biotechnology* **6** (3), 135–143 (2021). https://doi.org/10.1016/j.synbio.2021.06.002 .

[38] de Castilhos, J. *et al.* Severe dysbiosis and specific *Haemophilus* and *Neisseria* signatures as hallmarks of the oropharyngeal microbiome in critically ill coronavirus disease 2019 (COVID-19) patients. *Clinical Infectious Diseases* **75** (1), e1063–e1071 (2022). https://doi.org/10.1093/cid/ciab902 .

[39] Zuo, T. *et al.* Alterations in gut microbiota of patients with COVID-19 during time of hospitalization. *Gastroenterology* **159** (3), 944–955 (2020). https://doi.org/10.1053/j.gastro.2020.05.048 .

[40] Thuy-Boun, P. S. *et al.* Quantitative metaproteomics and activity-based protein profiling of patient fecal microbiome identifies host and microbial serine-type endopeptidase activity associated with ulcerative colitis. *Molecular & Cellular Proteomics* **21** (3), 100197 (2022). https://doi.org/10.1016/j.mcpro.2022.100197 .

[41] Maier, T. V. *et al.* Impact of dietary resistant starch on the human gut microbiome, metaproteome, and metabolome. *mBio* **8** (5), e01343–17 (2017). https://doi.org/10.1128/mBio.01343-17 .

[42] Granato, D. C. *et al.* Meta-omics analysis indicates the saliva microbiome and its proteins associated with the prognosis of oral cancer patients. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1869** (8), 140659 (2021). https://doi.org/10.1016/j.bbapap.2021.140659 .

[43] Jiang, X. *et al.* In-depth metaproteomics analysis of oral microbiome for lung cancer. *Research* **2022**, 9781578 (2022). https://doi.org/10.34133/2022/9781578 .

[44] Crovetto, F. *et al.* Nasopharyngeal microbiota profiling of pregnant women with SARS-CoV-2 infection. *Scientific Reports* **12** (1), 13404 (2022). https://doi.org/10.1038/s41598-022-17542-z .

[45] Engen, P. A. *et al.* Nasopharyngeal microbiota in SARS-CoV-2 positive and negative patients. *Biological Procedures Online* **23** (1), 10 (2021). https://doi.org/10.1186/s12575-021-00148-6 .

[46] Gupta, A. *et al.* Nasopharyngeal microbiome reveals the prevalence of opportunistic pathogens in SARS-CoV-2 infected individuals and their association with host types. *Microbes and Infection* **24** (1), 104880 (2022). https://doi.org/10.1016/j.micinf.2021.104880 .

[47] Hernández-Terán, A. *et al.* Dysbiosis and structural disruption of the respiratory microbiota in COVID-19 patients with severe and fatal outcomes. *Scientific Reports* **11** (1), 21297 (2021). https://doi.org/10.1038/s41598-021-00851-0 .

[48] Merenstein, C. *et al.* Signatures of COVID-19 severity and immune response in the respiratory tract microbiome. *mBio* **12** (4), e01777–21 (2021). https://doi.org/10.1128/mBio.01777-21 .

[49] Prasad, P. *et al.* Long-read 16S-seq reveals nasopharynx microbial dysbiosis and enrichment of *Mycobacterium* and *Mycoplasma* in COVID-19 patients: a potential source of co-infection. *Molecular Omics* **18** (6), 490–505 (2022). https://doi.org/10.1039/D2MO00044J .

[50] Smith, N. *et al.* Distinct systemic and mucosal immune responses during acute SARS-CoV-2 infection. *Nature Immunology* **22** (11), 1428–1439 (2021). https://doi.org/10.1038/s41590-021-01028-7 .

[51] Shilts, M. H. *et al.* Severe COVID-19 is associated with an altered upper respiratory tract microbiome. *Frontiers in Cellular and Infection Microbiology* **11**, 781968 (2022). https://doi.org/10.3389/fcimb.2021.781968 .

[52] Ventero, M. P. *et al.* Nasopharyngeal microbial communities of patients infected with SARS-CoV-2 that developed COVID-19. *Frontiers in Microbiology* **12**, 637430 (2021). https://doi.org/10.3389/fmicb.2021.637430 .

[53] Gupta, A. *et al.* Oral dysbiosis and its linkage with SARS-CoV-2 infection. *Microbiological Research* **261**, 127055 (2022). https://doi.org/10.1016/j.micres.2022.127055 .

[54] Gao, M. *et al.* Characterization of the human oropharyngeal microbiomes in SARS-CoV-2 infection and recovery patients. *Advanced Science* **8** (20), 2102785 (2021). https://doi.org/10.1002/advs.202102785 .

[55] Iebba, V. *et al.* Profiling of oral microbiota and cytokines in COVID-19 patients. *Frontiers in Microbiology* **12**, 671813 (2021). https://doi.org/10.3389/fmicb.

2021.671813 .

[56] Miller, E. H. *et al.* Oral microbiome alterations and SARS-CoV-2 saliva viral load in patients with COVID-19. *Microbiology Spectrum* **9** (2), e00055–21 (2021). https://doi.org/10.1128/Spectrum.00055-21 .

[57] Rafiqul Islam, S. M. *et al.* Dysbiosis of oral and gut microbiomes in SARS-CoV-2 infected patients in Bangladesh: Elucidating the role of opportunistic gut microbes. *Frontiers in Medicine* **9**, 821777 (2022). https://doi.org/10.3389/fmed.2022.821777 .

[58] Ren, Z. *et al.* Alterations in the human oral and gut microbiomes and lipidomics in COVID-19. *Gut* **70** (7), 1253–1265 (2021). https://doi.org/10.1136/gutjnl-2020-323826 .

[59] Wu, Y. *et al.* Altered oral and gut microbiota and its association with SARS-CoV-2 viral load in COVID-19 patients during hospitalization. *npj Biofilms and Microbiomes* **7** (1), 61 (2021). https://doi.org/10.1038/s41522-021-00232-5 .

[60] Xu, R. *et al.* Temporal association between human upper respiratory and gut bacterial microbiomes during the course of COVID-19 in adults. *Communications Biology* **4** (1), 240 (2021). https://doi.org/10.1038/s42003-021-01796-w .

[61] Chen, Y. *et al.* Six-month follow-up of gut microbiota richness in patients with COVID-19. *Gut* **71** (1), 222–225 (2022). https://doi.org/10.1136/gutjnl-2021-324090 .

[62] Ferreira-Junior, A. S. *et al.* Detection of intestinal dysbiosis in post-COVID-19 patients one to eight months after acute disease resolution. *International Journal of Environmental Research and Public Health* **19** (16), 10189 (2022).

https://doi.org/10.3390/ijerph191610189 .

[63] Gu, S. *et al.* Alterations of the gut microbiota in patients with coronavirus disease 2019 or H1N1 influenza. *Clinical Infectious Diseases* **71** (10), 2669–2678 (2020). https://doi.org/10.1093/cid/ciaa709 .

[64] Khan, M. *et al.* Gut dysbiosis and IL-21 response in patients with severe COVID-19. *Microorganisms* **9** (6), 1292 (2021). https://doi.org/10.3390/microorganisms9061292 .

[65] Mizutani, T. *et al.* Correlation analysis between gut microbiota alterations and the cytokine response in patients with coronavirus disease during hospitalization. *Microbiology Spectrum* **10** (2), e01689–21 (2022). https://doi.org/10.1128/spectrum.01689-21 .

[66] Moreira-Rosário, A. *et al.* Gut microbiota diversity and C-reactive protein are predictors of disease severity in COVID-19 patients. *Frontiers in Microbiology* **12**, 705020 (2021). https://doi.org/10.3389/fmicb.2021.705020 .

[67] Newsome, R. C. *et al.* The gut microbiome of COVID-19 recovered patients returns to uninfected status in a minority-dominated United States cohort. *Gut Microbes* **13** (1), 1926840 (2021). https://doi.org/10.1080/19490976.2021.1926840 .

[68] Romani, L. *et al.* The relationship between pediatric gut microbiota and SARS-CoV-2 infection. *Frontiers in Cellular and Infection Microbiology* **12**, 908492 (2022). https://doi.org/10.3389/fcimb.2022.908492 .

[69] Schult, D. *et al.* Gut bacterial dysbiosis and instability is associated with the onset of complications and mortality in COVID-19. *Gut Microbes* **14** (1), 2031840 (2022). https://doi.org/10.1080/19490976.2022.2031840 .

[70] Zhou, Y., Zhang, J., Zhang, D., Ma, W.-L. & Wang, X. Linking the gut microbiota to persistent symptoms in survivors of COVID-19 after discharge. *Journal of Microbiology* **59** (10), 941–948 (2021). https://doi.org/10.1007/s12275-021-1206-5 .

[71] He, F. *et al.* Fecal multi-omics analysis reveals diverse molecular alterations of gut ecosystem in COVID-19 patients. *Analytica Chimica Acta* **1180**, 338881 (2021). https://doi.org/10.1016/j.aca.2021.338881 .

[72] Ke, S., Weiss, S. T. & Liu, Y.-Y. Dissecting the role of the human microbiome in COVID-19 via metagenome-assembled genomes. *Nature Communications* **13** (1), 5235 (2022). https://doi.org/10.1038/s41467-022-32991-w .

[73] Yeoh, Y. K. *et al.* Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut* **70** (4), 698–706 (2021). https://doi.org/10.1136/gutjnl-2020-323020 .

[74] Dick, J. M. & Tan, J. Chemical links between redox conditions and estimated community proteomes from 16S rRNA and reference protein sequences. *Microbial Ecology* (2022). https://doi.org/10.1007/s00248-022-01988-9 .

[75] Masuda, T., Saito, N., Tomita, M. & Ishihama, Y. Unbiased quantitation of *Escherichia coli* membrane proteome using phase transfer surfactants. *Molecular & Cellular Proteomics* **8** (12), 2770–2777 (2009). https://doi.org/10.1074/mcp.M900240-MCP200 .

[76] Leary, D. H., Hervey, W. J., Deschamps, J. R., Kusterbeck, A. W. & Vora, G. J. Which metaproteome? the impact of protein extraction bias on metaproteomic analyses. *Molecular and Cellular Probes* **27** (5), 193–199 (2013). https://doi.org/10.1016/j.mcp.2013.06.003 .

[77] Lai, M. & Lü, B. Tissue preparation for microscopy and histology. *Comprehensive Sampling and Sample Preparation* **3** (3.04), 53–93 (2012). https://doi.org/10.1016/B978-0-12-381373-2.00070-3 .

[78] Dubourg, G. *et al.* Gut microbiota associated with HIV infection is significantly enriched in bacteria tolerant to oxygen. *BMJ Open Gastroenterology* **3** (1), e000080 (2016). https://doi.org/10.1136/bmjgast-2016-000080 .

[79] Zheng, L., Kelly, C. J. & Colgan, S. P. Physiologic hypoxia and oxygen homeostasis in the healthy intestine. a review in the theme: Cellular responses to hypoxia. *American Journal of Physiology-Cell Physiology* **309** (6), C350–C360 (2015). https://doi.org/10.1152/ajpcell.00191.2015 .

[80] George, C. E. *et al.* COVID-19 symptoms are reduced by targeted hydration of the nose, larynx and trachea. *Scientific Reports* **12** (1), 4599 (2022). https://doi.org/10.1038/s41598-022-08609-y .

[81] Hultström, M. *et al.* Dehydration is associated with production of organic osmolytes and predicts physical long-term symptoms after COVID-19: a multicenter cohort study. *Critical Care* **26** (1), 322 (2022). https://doi.org/10.1186/s13054-022-04203-w .

[82] Nearing, J. T., Comeau, A. M. & Langille, M. G. I. Identifying biases and their potential solutions in human microbiome studies. *Microbiome* **9** (1), 113 (2021). https://doi.org/10.1186/s40168-021-01059-0 .

[83] Lee, J.-Y., Tsolis, R. M. & Bäumler, A. J. The microbiome and gut homeostasis. *Science* **377** (6601), eabp9960 (2022). https://doi.org/10.1126/science.abp9960 .

[84] Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016). https://doi.org/10.7717/peerj.2584 .

[85] Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research* **41** (D1), D590–D596 (2012). https://doi.org/10.1093/nar/gks1219 .

[86] Cole, J. R. *et al.* Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Research* **42** (D1), D633–D642 (2014). https://doi.org/10.1093/nar/gkt1244 .

[87] Chen, T. *et al.* The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database* **2010**, baq013 (2010). https://doi.org/10.1093/database/baq013 .

[88] Deutsch, E. W. *et al.* The ProteomeXchange consortium in 2020: Enabling 'big data' approaches in proteomics. *Nucleic Acids Research* **48** (D1), D1145–D1152 (2020). https://doi.org/10.1093/nar/gkz984 .

[89] Meier, D. V. *et al.* Microbial metal-sulfide oxidation in inactive hydrothermal vent chimneys suggested by metagenomic and metaproteomic analyses. *Environmental Microbiology* **21** (2), 682–701 (2019). https://doi.org/10.1111/1462-2920.14514 .

[90] Pjevac, P. *et al.* Metaproteogenomic profiling of microbial communities colonizing actively venting hydrothermal chimneys. *Frontiers in Microbiology* **9**, 680 (2018). https://doi.org/10.3389/fmicb.2018.00680 .

[91] Kleiner, M. *et al.* Assessing species biomass contributions in microbial communities via metaproteomics. *Nature Communications* **8** (1), 1558 (2017).

https://doi.org/10.1038/s41467-017-01544-x .

[92] Hawley, A. K. *et al.* A compendium of multi-omic sequence information from the Saanich Inlet water column. *Scientific Data* **4** (1), 170160 (2017). https://doi.org/10.1038/sdata.2017.160 .

[93] Reeves, E. P. *et al.* Microbial lipids reveal carbon assimilation patterns on hydrothermal sulfide chimneys. *Environmental Microbiology* **16** (11), 3515–3532 (2014). https://doi.org/10.1111/1462-2920.12525 .

[94] Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9** (4), 357–359 (2012). https://doi.org/10.1038/nmeth.1923 .

[95] Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28** (24), 3211–3217 (2012). https://doi.org/10.1093/bioinformatics/bts611 .

[96] Rho, M., Tang, H. & Ye, Y. FragGeneScan: Predicting genes in short and error-prone reads. *Nucleic Acids Research* **38** (20), e191 (2010). https://doi.org/10.1093/nar/gkq747 .

[97] Hyatt, D. *et al.* Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11** (1), 119 (2010). https://doi.org/10.1186/1471-2105-11-119 .