frontiers
in Genetics

Challenges, Strategies, and Perspectives for Reference-Independent Longitudinal Multi-Omic Microbiome Studies

Susana Martínez Arbas[1]*, Susheel Bhanu Busi[1], Pedro Queirós[1], Laura de Nies[1], Malte Herold[2], Patrick May[1], Paul Wilmes[1,3], Emilie E. L. Muller[4] and Shaman Narayanasamy[1]

[1]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg, [2]Department of Environmental Research and Innovation, Luxembourg Institute of Science and Technology, Belvaux, Luxembourg, [3]Department of Life Sciences and Medicine, Faculty of Science, Technology and Medicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg, [4]Université de Strasbourg, UMR 7156 CNRS, Génétique Moléculaire, Génomique, Microbiologie, Strasbourg, France

In recent years, multi-omic studies have enabled resolving community structure and interrogating community function of microbial communities. Simultaneous generation of metagenomic, metatranscriptomic, metaproteomic, and (meta) metabolomic data is more feasible than ever before, thus enabling in-depth assessment of community structure, function, and phenotype, thus resulting in a multitude of multi-omic microbiome datasets and the development of innovative methods to integrate and interrogate those multi-omic datasets. Specifically, the application of reference-independent approaches provides opportunities in identifying novel organisms and functions. At present, most of these large-scale multi-omic datasets stem from spatial sampling (e.g., water/soil microbiomes at several depths, microbiomes in/on different parts of the human anatomy) or case-control studies (e.g., cohorts of human microbiomes). We believe that longitudinal multi-omic microbiome datasets are the logical next step in microbiome studies due to their characteristic advantages in providing a better understanding of community dynamics, including: observation of trends, inference of causality, and ultimately, prediction of community behavior. Furthermore, the acquisition of complementary host-derived omics, environmental measurements, and suitable metadata will further enhance the aforementioned advantages of longitudinal data, which will serve as the basis to resolve drivers of community structure and function to understand the biotic and abiotic factors governing communities and specific populations. Carefully setup future experiments hold great potential to further unveil ecological mechanisms to evolution, microbe-microbe interactions, or microbe-host interactions. In this article, we discuss the challenges, emerging strategies, and best-practices applicable to longitudinal microbiome studies ranging from sampling, biomolecular extraction, systematic multi-omic measurements, reference-independent data integration, modeling, and validation.

Keywords: microbiome, metatranscriptomics, metaproteomics, time-series, metagenomics, metabolomics, *de novo* assembly

INTRODUCTION

Advances in the study of microbial communities have highlighted their important role in natural processes, including those considered as ecosystem services for humankind (Bodelier, 2011). Complex dynamics in microbiomes at the level of composition and structure, as well as function (Heintz-Buschart and Wilmes, 2018) stem from constant adaptation of a given community toward fluctuations of abiotic and biotic factors. However, the fate of these microbial consortia in the face of perturbations is often not understood nor predictable (Muller, 2019). Longitudinal approaches are necessary to understand microbial community dynamics, as they may offer valuable insights into temporal trends and consequences of environmental forcings, when used in tandem with host-derived (Heintz-Buschart et al., 2016; Lloyd-Price et al., 2019; Mars et al., 2020) or environmental (Law et al., 2016; Herold et al., 2020) data. Longitudinal studies can be conducted using diachronic or synchronic approaches (Costa Junior et al., 2013). Herein, we discuss the capacity of longitudinal diachronic approaches as a critical tool toward studying microbial communities. We will further focus on multi-omics longitudinal studies, which leverage the power of the entire high-throughput meta-omic spectrum, namely meta-genomics (MG), -transcriptomics (MT), -proteomics (MP), and -metabolomics (MM), as they are now more feasible and affordable than ever before (Narayanasamy et al., 2015).

Overall, longitudinal multi-omics will enhance our understanding of microbial community dynamics, which could potentially bring about positive outcomes in biomedicine, biotechnology, and for the environment. However, various aspects must be considered when conducting longitudinal multi-omic microbiome studies, ranging from experimental design, bioinformatic processing, modeling, and validation. In this article, we explore challenges, considerations, and potential solutions for such studies, based on recent advances and reports (Law et al., 2016; Lloyd-Price et al., 2019; Herold et al., 2020; Martínez Arbas et al., 2021), which are applicable to both microbe-centric (e.g., soil, water) or host-centric (e.g., human gut) systems. Finally, although this article focuses on specifically longitudinal multi-omic microbiome studies, the content is generally applicable to any large-scale microbiome studies.

MULTI-OMIC CONSIDERATIONS AND EXPERIMENTAL DESIGN FOR LONGITUDINAL STUDIES

Integration of multi-omic microbiome datasets has been routinely performed, with notable instances, including studies on type-1 diabetes (Heintz-Buschart et al., 2016), cancer (Kaysen et al., 2017), healthy human gut (Tanca et al., 2017), Crohn's disease (Erickson et al., 2012), and activated sludge (Muller et al., 2014; Roume et al., 2015; Yu et al., 2019). These studies clearly demonstrate the maturity of the current microbiome multi-omics toolbox. Despite this, and to the best of our knowledge, equivalent multi-omic surveys based on extensive longitudinal microbiome sampling remain rather limited. **Table 1** lists several relevant studies of longitudinal (at least six timepoints) and multi-omic (at least two omic levels, excluding 16S amplicon sequencing) microbiome datasets.

The famous adage "*absence of evidence is not evidence of absence*" (Altman and Bland, 1995) could likely be a prelude to most microbiome studies. Hence, we discuss these studies in the context of reference-independent bioinformatics

---

TABLE 1 | Longitudinal multi-omic microbiome datasets and studies.

System	Sample type	Duration*	Frequency*	Total of samples	MG	MT	MP	MM	Complementary data	Studies
Human gut microbiome	Stool samples from 132 humans; healthy or with Crohn's disease or ulcerative colitis	1 year	Bi-weekly	2,965	x	x	x	x	Host genomics, transcriptomics bisulfite sequencing, serologic profiles, diet surveys, and fecal calprotectin	Lloyd-Price et al., 2019 Ruiz-Perez et al., 2021
	Stool samples of 77 individuals	6 months	Monthly	474	x			x	Host transcriptome, metabolome, cytokines, methylome, dietary survey, and physiology	Blasche et al., 2021
Activated sludge	Floating sludge islets from a single anoxic tank	1.5 year	Weekly	53	x	x	x	x	Temperature, pH, oxygen concentration, conductivity, inflow, nitrate concentration, and extracellular metabolites	Herold et al., 2020 Martínez Arbas et al., 2021
	Full- and lab-scale activated sludge	2.5 months	Weekly	10	x	x			Temperature, pH, redox potential and dissolved oxygen	Law et al., 2016

Longitudinal multi-omic data must be of least six timepoints and at least two meta-omic readouts excluding 16S amplicon sequencing. Omics data derived from host(s) are considered separate from the microbial meta-omic spectra.
*Approximate values.

approaches, centered around *de novo* assemblies of sequencing data (MG and MT), subsequently complemented by additional omics (MP and MM, depending on their availability; **Figure 1**). Reference-independent approaches offer asymmetric advantages and opportunities in discovering novel microbial taxa and/or functionalities (Celaj et al., 2014; Narayanasamy et al., 2015; Lapidus and Korobeynikov, 2021), compared to reference-dependent methodologies (Sunagawa et al., 2013; Treangen et al., 2013). Moreover, the integration of multi-omics has been shown to yield superior output compared to single omic studies. For instance, the co-assembly of MG and MT sequencing reads was shown to improve the quality of assembled contigs (Narayanasamy et al., 2016), which in turn improves taxonomic annotation, gene calling/annotation, binning, metabolic pathway (re) construction (Muller et al., 2018; Zhou et al., 2020;

Zimmermann et al., 2021), and quantification of features, e.g., taxa/genes (Narayanasamy et al., 2016). Similarly, MP spectra searches are more effective when performed against gene databases derived from MG assemblies of the same sample/environment, compared to generic databases, thus improving the recruitment of measured peptides (Tanca et al., 2016; Heyer et al., 2017; Timmins-Schiffman et al., 2017). Moreover, such a reference-independent approach may be necessary for microbial communities that are not well characterized and lack extensive unified genome or gene catalogues, such as those available for the human gut microbiome (Li et al., 2014; Almeida et al., 2021). However, most microbial communities are heterogeneous, which further complicates downstream multi-omic data processing, integration, curation, transformation, and modeling (Jiang et al., 2019). Therefore, the adherence toward standards
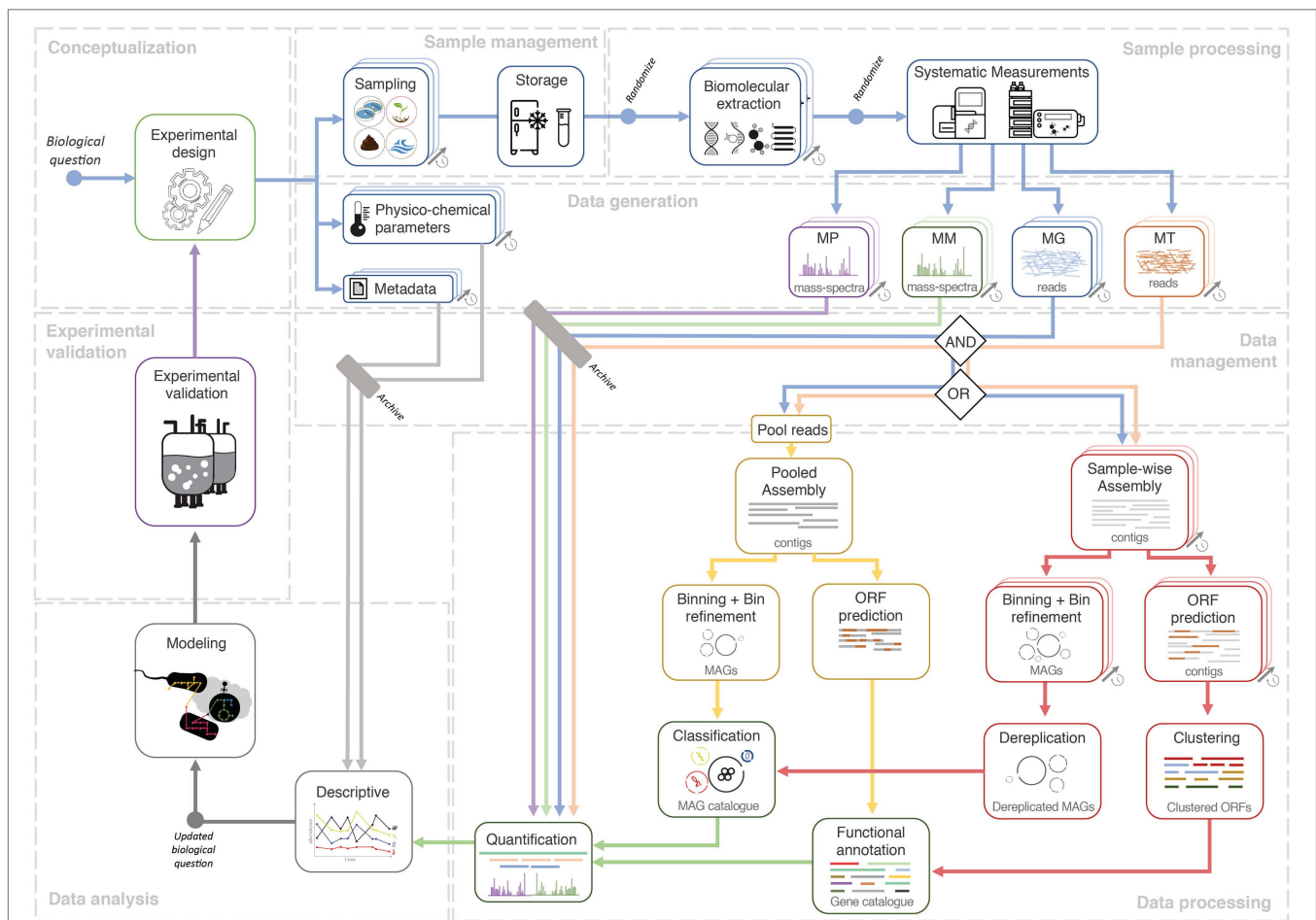

FIGURE 1 | Systems ecology workflow for longitudinal multi-omic microbiome studies. A study conceptualized *via* an experimental design phase and an initial biological question which is then followed by sample collection, sample management, and systematic high-throughput measurements. The next-generation sequencing (NGS) data could either undergo aggregated processing (yellow track) involving a pooled *de novo* assembly of NGS reads from all longitudinal samples, to eventually yield a metagenome assembled genome (MAG) and/or gene catalogue *via* binning and gene calling, respectively. In the dereplication approach (red track), data from each sample are first processed in a sample-wise manner, namely the steps of *de novo* assembly, binning, and gene calling. The resulting MAGs and predicted ORFs are then merged through a process called dereplication which generates the catalogue. The availability of a catalogue allows quantification whereby the output could be used for descriptive analyses which could potentially lead to updated or entirely novel biological questions. Quantified values, combined with descriptive analyses, could then be used within dynamic or metabolic models (gray track). Validation of models could lead to further *in situ* longitudinal experimental designs. Finally, all data (raw input, output, metadata) and code (not depicted) should be archived under a data and code management strategy. Free icons were used from https://www.flaticon.com (creators: Freepik, Gregor Cresnar, Freepik, and Smashicons).

and best-practices, spanning from sampling to data analyses is important to the outcome of a project. Accordingly, **Figure 1** illustrates the potential lifecycle of a longitudinal multi-omic microbiome study.

Longitudinal multi-omic studies require systematic and thorough study designs that consider sampling parameters (Gerber, 2014; Cao et al., 2017; Liang et al., 2020), metadata, and complementary measurements, such as physico-chemical parameters or questionnaires (Kumar et al., 2014), all of which affect downstream analyses. Sampling parameters, such as duration and frequency, are dictated by the inherent properties of a given microbial system. For instance, the sampling duration when studying gut microbiome development of neonates could span from birth until a "mature" gut microbiome composition is achieved (Stewart et al., 2018), which may vary from subject to subject. Naturally-occurring microbial systems that are exposed to the environment may exhibit annual cyclical behavior based on seasonality and, therefore, could be sampled for at least one complete season-to-season cycle (Johnston et al., 2019). Sampling frequency may be determined by the dynamics and/or generational-timescale of a given system. For instance, the human gut microbiome is known to exhibit daily fluctuations, and therefore could be sampled on a daily basis within a given temporal study (David et al., 2014), while activated sludge systems are known to exhibit (approximately) weekly doubling periods and thus could be sampled on a weekly basis (Herold et al., 2020; Martínez Arbas et al., 2021). Based on the recommendations of Sefer et al. (2016), if biological replicates are either not feasible (i.e., $n = 1$) or limited (i.e., low n) (Herold et al., 2020), one should ideally opt for higher frequency (dense) longitudinal sampling, and less dense sampling if biological replicates were available (i.e., high n), e.g., a cohort of patients (Lloyd-Price et al., 2019). Equidistant sampling is required by many downstream mathematical frameworks, such as cross-correlation or local similarity analysis (Faust et al., 2015), and thus should be strived for, as much as possible. However, the datasets listed in **Table 1**, albeit extensive and resource intensive, are not perfectly equidistant, further highlighting the practical challenges for longitudinal sampling *in situ*, including, but not limited to, accessibility, consistent biomass availability, and cost.

SAMPLE, DATA AND CODE MANAGEMENT

It is crucial to limit potential biases linked to longitudinal data, e.g., in extended time-series; samples are stored for long periods, while multiple personnel may be involved in sample collection, handling, storage, and documentation. Hence, clear guidelines and standardization must be established, as they are key factors that potentially affect downstream processes and overall outcome (Blekhman et al., 2016; Schoenenberger et al., 2016).

Biomolecular extraction from a single sample is ideal over multiple extractions from subsamples (Roume et al., 2013a). Advantageously, commercial kits for concomitant extraction of

multiple biomolecules are available, including reports proposing adapted methods for extracting various biomolecules, such as DNA, total RNA, small RNA, protein, and metabolites (Peña-Llopis and Brugarolas, 2013; Roume et al., 2013b; Thorn et al., 2019). The availability of sufficient biomass (Eisenhofer et al., 2019) lysis-, homogenization-(Machiels et al., 2000; Santiago et al., 2014; Fiedorová et al., 2019) and preservation- (Borén, 2015; Hickl et al., 2019) methods are key factors that determine effectiveness to comprehensively recover all intracellular and/or extracellular biomolecules. Next, biomolecular extraction should be automated, whenever possible. While evaluations have shown that it may not necessarily provide better quality results compared to a human operator (Phillips et al., 2012), the output is more consistent (Fidler et al., 2020). In the same vein, omic readouts should also be generated on a single platform (s) as unique batches to ensure consistent output quality.

Batch effects are often overlooked in omic studies (de Goffau et al., 2021), but can be minimized during stages of sample processing by including randomization, sample tracking, and extensive documentation (Leek et al., 2010). Sample randomization implemented within batches of biomolecular extraction and high-throughput measurements could help discriminate batch effects and temporal variation, i.e., different sets of randomly selected samples from different timepoints could be treated together at each different step (Oh et al., 2019). Additionally, batch effects could be mitigated using downstream analytical (Wang and Cao, 2019) and computational methods (Gibbons et al., 2018; McLaren et al., 2019).

A potential effective experimental measure for minimizing and elucidating batch effects is the inclusion of mock/control samples during both the extraction and high-throughput measurements (Bokulich et al., 2016; Hornung et al., 2019; ATCC Mock Microbial Communities, 2020). Samples with low biomass, e.g., from neonates, glacier-streams, or acid-mine drainage, should include extraction blanks as negative controls, which are extremely valuable to discriminate contaminants arising from kits and reagents (Salter et al., 2014; Heintz-Buschart et al., 2018; Wampach et al., 2018; Weyrich et al., 2019). Furthermore, spike-ins could be helpful for downstream quantification (Zinter et al., 2019). Importantly, replicates can be used within downstream statistical frameworks (Sokal, 1995; Anderson, 2017; Kuznetsova et al., 2017; Mallick et al., 2021) to understand both within- and between-sample heterogeneity, thereby minimizing mischaracterisation of contaminants or findings driven by batch effects (de Goffau et al., 2021).

Longitudinal and multi-omic studies yield large datasets, where data processing and analyses are typically time and resource intensive. These rich datasets may be reused to study multiple aspects of a given microbial system (**Table 1**). Therefore, equal emphasis should be placed on designing bioinformatic workflows and code/data management strategies to improve reproducibility and transparency. For example, peer-review journals have begun mandating "data availability" sections and links to code repositories in adherence to project/coding best practices and standards (Sandve et al., 2013; Bokulich et al., 2020), further improving posterior data integration and analysis in the short-term, while improving scaling-up

and knowledge transfer in the long run (Shahin et al., 2017; Wilson et al., 2017). In addition, format-free archival repositories, such as Zenodo could be used for non-standard data types,[1] for instance simulated raw data, physico-chemical measurements, intermediate data, large tables, and archived Github repositories. Despite this, reports indicate that 26% of bioinformatics tools are no longer available (Mangul et al., 2019), while gaps in available raw data (Jurburg et al., 2020) and metadata (Schriml et al., 2020) still exist.

CONSTRUCTION OF LONGITUDINAL GENE AND GENOME REFERENCE CATALOGUES

Microbiomes may be studied from a gene-centric perspective (Roume et al., 2015), which requires read or contig-level taxonomic classification (Segata et al., 2012; Wood and Salzberg, 2014), ORF prediction (Hyatt et al., 2010; Rho et al., 2010), and gene annotation (Seemann, 2014; Buchfink et al., 2015; Franzosa et al., 2018; Queirós et al., 2020). Metagenome assembled genomes (MAGs) provide genomic context and can be obtained through binning (Chen et al., 2020; Yue et al., 2020) followed by taxonomic classification (Bremges et al., 2020; Chaumeil et al., 2020) and functional annotation. In that regard, several tools exist that improve the binning process by automating the selection of highest-quality MAGs (bins) and/or performing MAG refinement (Broeksema et al., 2017; Sieber et al., 2018; Uritskiy et al., 2018). These tools enable ensemble binning approaches, balancing out the strengths and weaknesses of different binning methods (Chen et al., 2020; Yue et al., 2020).

Features (i.e., taxa or genes) appear in varying quantities, in different timepoints of longitudinal meta-omic studies. It is challenging to link and track features from one timepoint to another without any given point of reference. Therefore, the construction of what we term as "representative longitudinal catalogues" (hereafter referred to as catalogues) of MAGs/genes, provides a non-redundant representative base to link features from the different longitudinal samples (Herold et al., 2020; Martínez Arbas et al., 2021). The outcome of any downstream analysis is highly reliant on the quality of the MAGs and genes within a catalogue, which further depends on the quality of large-scale bioinformatic processing (e.g., *de novo* assembly and binning). **Figure 1** illustrates two methods of constructing such catalogues, which are through aggregated processing of data from all samples or through de-replicating the output from individually processed sample data (i.e., sample-wise processing). A third alternative to these methods could be the representation of non-redundant genes in pangenomes from MAGs annotated at the species-level (Tettelin et al., 2005; Delmont and Eren, 2018), collected across all timepoints. This allows for identifying any varying patterns especially in the context of environmental factors and phylogenetic constraints influencing gene acquisition and/or genome-streamlining (Tettelin et al., 2005). Given that

others have highlighted the catalogue building methodologies (Qin et al., 2010; Nayfach et al., 2020; Almeida et al., 2021); here, we elaborate methods discussed above in the context of both gene- and MAG-centric strategies.

The general advantage of the aggregated processing approach is simplicity, whereby a single run is required for all the large-scale bioinformatic processing steps (**Figure 1**). Moreover, pooled assemblies have been shown to be effective (Magasin and Gerloff, 2015), especially in the advent of highly efficient *de novo* assemblers (Li et al., 2016) and digital normalization (Brown et al., 2012). However, pooling reads from a large number of samples increases the complexity of the *de novo* assembly process, especially for complex communities. It also requires substantial computational resources, while potentially resulting in lower quality contigs, MAGs, and genes (Chen et al., 2020).

The dereplication method (**Figure 1**) is applied after independent sample-wise large-scale bioinformatic processing (Evans and Denef, 2020). Predicted ORFs could be de-replicated through clustering (Li and Godzik, 2006; Edgar, 2010; Mirdita et al., 2019), producing a gene catalogue (Li et al., 2014). On the contrary, the dereplication of MAGs is more complex, requiring several steps: binning from sample-wise *de novo* assemblies to generate MAGs, curation of high-quality MAGs (Parks et al., 2015), and dereplication of MAGs (Olm et al., 2017; Wampach et al., 2018) to select the most representative MAGs of the longitudinal data (Uritskiy et al., 2018; Chen et al., 2020). In general, dereplication methods are particularly advantageous for longitudinal microbiome studies with many deeply sequenced samples (Herold et al., 2020; Martínez Arbas et al., 2021).

Although not systematically evaluated, one caveat worth considering when constructing a catalogue based on *de novo* assemblies, binning, and dereplication is the potential loss of resolution in population-level diversity (Kashtan et al., 2014; Evans and Denef, 2020; Quince et al., 2020), which may include single nucleotide variants, copy number variants, strains, and auxiliary gene content (Evans and Denef, 2020) potentially impacting important downstream steps, such as integration of metaproteomic data (Tanca et al., 2016) or time-resolved strain tracking (Brito and Alm, 2016; Zlitni et al., 2020). To the best of our knowledge, the extent of the impact has yet to be systematically investigated. In our opinion, several strategies can be applied to overcome this issue, including the usage of a comparative genomics methodology, i.e., pangenomes (Delmont and Eren, 2018), even opt for (re) assemblies of read subsets associated to particular taxa or MAGs of interest (Albertsen et al., 2013), or the application of strain-level analysis tools (Anyansi et al., 2020).

Overall, choosing the specific methods for constructing a longitudinal catalogue depends on various factors, including the biological question, complexity of the community (van der Walt et al., 2017), number of samples, and sequencing depth. To the best of our knowledge, a comparison between an aggregated processing approach and a dereplication approach has yet to be conducted. Such a comparison would further help to inform researchers on selecting the best strategy for longitudinal analyses.

---

[1]https://zenodo.org

QUANTIFICATION AND NORMALIZATION

Longitudinal catalogues provide compositional information of community taxa and potential functions. However, the relative quantification of community members and functionalities is key in harnessing the power of longitudinal microbiome data, as it allows the observation of community taxa/functional dynamics and could be used in downstream modeling. In that regard, quantifying MG and MT sequencing data is a standard process of aligning reads (Li and Durbin, 2009) to relevant catalogues, and then quantifying features of interest (e.g., population/gene relative genomic abundance, gene expression) based on those alignments, providing information on community structure, functional potential, and gene expression. Complementarily, MP data provide functional insights, whereby several methods are available for the quantification of such data (Delogu et al., 2020; Pible et al., 2020), while identification and quantification of metabolites through MM data (Kapoore and Vaidyanathan, 2016; Mallick et al., 2019; Røst et al., 2020) provide insights on the community phenotype (s). However, *in situ* measurements of substrate uptake through labeling-based approaches (Starr et al., 2018) are challenging. Therefore, specific metabolites of interest could be indirectly linked to members of a microbial community by proportionally assigning the relative contribution of a MAG to a given (re) constructed metabolic pathway based on genomic abundance or gene/protein expression (Noecker et al., 2016; Blasche et al., 2021).

Normalization of quantified values is required to enable community structure and function comparisons between timepoint samples. The selection of normalization methods is important as it affects downstream analytical steps. There are several methods to normalize longitudinal MG and MT data, from the generation of compositional data to log-ratios and differential rankings (Chen et al., 2018; Pereira et al., 2018; Morton et al., 2019). Additionally, one should also inspect the data for potential confounding batch effects and take it into consideration when performing normalization (Gibbons et al., 2018; McLaren et al., 2019; Coenen et al., 2020). In summary, effective relative quantification and normalization will serve as a strong basis for downstream modeling approaches, and the development of robust methods for absolute quantification will be decisive in the future.

ANALYSIS OF COMMUNITY CHARACTERISTICS AND DYNAMICS

Generally, microbiome omic data are complex, as it is (i) compositional, e.g., provided as relative abundances, which require specific considerations when selecting statistical analyses (Gloor et al., 2017), (ii) highly sparse, such that the interpretation of zero-values generated from sampling, biological, or technical processes heavily affects data-derived conclusions (Silverman et al., 2020), and (iii) high dimensional, which increases modeling difficulty due to the influence of feature selection that heavily affect potential predictions (Bolón-Canedo et al., 2016). Furthermore, multi-omic studies may contain gaps within the omic spectrum, such that certain samples may not be represented within a certain omic layer (Lloyd-Price et al., 2019). Despite introducing complexity, the complementary use of different omics could improve analysis outcomes and add predictive power to models (Muller et al., 2013; Fondi and Liò, 2015). Longitudinal data introduce another layer of complexity, i.e., time dependencies, such that one timepoint is dependent on the previous timepoints, rendering conventional statistical analyses unsuitable as they assume samples to be independent (Coenen et al., 2020). This is further compounded by the fact that samples from longitudinal *in situ* studies are often low in number and non-equidistant (Park et al., 2020). Imputation may be used to supplement missing values (i.e., omic measurements or timepoints; Jiang et al., 2020).

Initial exploration of the microbiome dynamics can be assessed through ordination analyses, where high dimensional population structure data are visualized in a two-dimensional space to observe the trajectory of the samples and the behavior of the system, i.e., metastability, cycles, and alternative states (Gonze et al., 2018). Then, community member relationships may be inferred using, e.g., correlation methods (Faust et al., 2012; Friedman and Alm, 2012; Weiss et al., 2016). Unfortunately, correlations may be insufficient to assess complex community interactions, whereby the application of modeling approaches would be necessary to resolve those relationships (Fisher and Mehta, 2014; Trosvik et al., 2015; Ridenhour et al., 2017). Modeling could serve as a means of integrating several layers of omic data (Lloyd-Price et al., 2019; Ruiz-Perez et al., 2021) further elucidating microbial interplay beyond species abundances and functional potential.

Extensive literature of statistical and mathematical frameworks for multi-omic and/or longitudinal microbiome data is currently available. For instance, Noor et al. (2019) review the integration of multi-omics data from data-driven and knowledge-based perspectives. Coenen et al. (2020) discuss approaches to characterize temporal dynamics and to identify periodicity of populations and putative interactions between them, while Faust et al. (2018) propose a classification scheme for better model selection. Bodein et al. (2019) provide a multivariate framework to integrate longitudinal and multi-omics data, while Park et al. (2020) discuss the development of models and software tools for time-series metagenome and metabolome data. Overall, the application of these methodologies should be tailored toward specific hypotheses and studies, for which data exploration is essential to select modeling approaches that fit the type, quality, and quantity of the data.

More recently, the emergence of studies which track microbiome dynamics of cohorts over time, i.e., multiple individuals/sites (Carmody et al., 2019; Lloyd-Price et al., 2019; Mars et al., 2020), necessitates the ability to discriminate variation stemming from the same individual/environment compared to those from different individuals/environments. In such cases, multi-level statistical modeling (also known as mixed-effects/ hierarchical models) is able to account for repeated sampling or nested variation across a sample population (Sokal, 1995; Anderson, 2017; Kuznetsova et al., 2017; Mallick et al., 2021). Most notably Lloyd-Price et al. (2019) extensively applied such

methods to associate multi-omic microbiome signatures with host-derived molecular profiles in a cohort of 132 individuals. Other instances include multi-omic longitudinal studies that combine murine and human datasets to unveil the adaptation of gut microbiomes to raw and cooked food (Carmody et al., 2019) and the identification of therapeutic targets for irritable bowel syndrome (Mars et al., 2020). Finally, there are newer methodologies that apply similar/related statistical frameworks to modeling multi-omic data (Mallick et al., 2021).

The validation of the models remains one of the most challenging issues. Mathematical models combined with culture of synthetic microbial communities are commonly utilized to study mechanisms behind host-microbiome interactions (Moejes et al., 2017). It is also possible to validate interactions between microbes by, e.g., applying environmental perturbations in controlled conditions (Law et al., 2016; Herold et al., 2020). These explorations may result in a further understanding of the role of biotic and abiotic factors in shaping microbiomes, in relation to community phenotypes found in nature, biotechnological processes (Law et al., 2016; Herold et al., 2020), or host-associated microbiomes (Moejes et al., 2017; Garza et al., 2018).

CONCLUSION

Longitudinal microbiome studies combined with integrated multi-omic measurements provide unprecedented opportunities to study microbial community dynamics, both structurally and functionally. In tandem with evolving high-throughput technologies, e.g., long-read sequencing (Moss et al., 2020; Wickramarachchi et al., 2020), these studies will become important tools in the exploration and potential exploitation of microbial consortia. We described strategies to mitigate the various challenges associated with such studies, encompassing study design, best practices, practical considerations, and bioinformatics processing and modeling. While longitudinal multi-omics datasets are currently scarce (**Table 1**), we are confident that it will increasingly become more common, similar to how we are increasingly transitioning from single omics to multi-omic (Noor et al., 2019). Longitudinal microbiome multi-omics will serve as an important tool for further improving analytical methods, which will in turn lead to relevant biomedical, biotechnological, and environmental outcomes.

AUTHOR CONTRIBUTIONS

SMA and SN outlined the manuscript and coordinated the writing process. LdN, SN, and SMA prepared the figure. All authors contributed to the writing, reviewing, and editing of the manuscript. All authors approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W., and Nielsen, P. H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* 31, 533–538. doi: 10.1038/nbt.2579

Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z. J., et al. (2021). A unified catalog of 204, 938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* 39, 105–114. doi: 10.1038/s41587-020-0603-3

Altman, D. G., and Bland, J. M. (1995). Statistics notes: absence of evidence is not evidence of absence. *BMJ* 311:485. doi: 10.1136/bmj.311.7003.485

Anderson, M. J. (2017). "Permutational multivariate analysis of variance (PERMANOVA)," in *Wiley Stats Ref: Statistics Reference Online*. eds. N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri and J. L. Teugels (Chichester, UK: John Wiley & Sons, Ltd.), 1–15.

Anyansi, C., Straub, T. J., Manson, A. L., Earl, A. M., and Abeel, T. (2020). Computational methods for strain-level microbial detection in colony and metagenome sequencing data. *Front. Microbiol.* 11:1925. doi: 10.3389/fmicb.2020.01925

ATCC Mock Microbial Communities (2020). Available at: https://www.atcc.org/en/Products/Microbiome_Standards.aspx (Accessed November 30, 2020).

Blasche, S., Kim, Y., Mars, R. A. T., Machado, D., Maansson, M., Kafkia, E., et al. (2021). Metabolic cooperation and spatiotemporal niche partitioning in a kefir microbial community. *Nat. Microbiol.* 6, 196–208. doi: 10.1038/s41564-020-00816-5

Blekhman, R., Tang, K., Archie, E. A., Barreiro, L. B., Johnson, Z. P., Wilson, M. E., et al. (2016). Common methods for fecal sample storage in field studies yield consistent signatures of individual identity in microbiome sequencing data. *Sci. Rep.* 6:31519. doi: 10.1038/srep31519

Bodein, A., Chapleur, O., Droit, A., and Lê Cao, K.-A. (2019). A generic multivariate framework for the integration of microbiome longitudinal studies with other data types. *Front. Genet.* 10:963. doi: 10.3389/fgene.2019.00963

Bodelier, P. L. E. (2011). Toward understanding, managing, and protecting microbial ecosystems. *Front. Microbiol.* 2:80. doi: 10.3389/fmicb.2011.00080

Bokulich, N. A., Rideout, J. R., Mercurio, W. G., Shiffer, A., Wolfe, B., Maurice, C. F., et al. (2016). Mockrobiota: a public resource for microbiome bioinformatics benchmarking. *mSystems* 1:e00062-16. doi: 10.1128/mSystems.00062-16

Bokulich, N. A., Ziemski, M., Robeson, M. S., and Kaehler, B. D. (2020). Measuring the microbiome: best practices for developing and benchmarking microbiomics methods. *Comput. Struct. Biotechnol. J.* 18, 4048–4062. doi: 10.1016/j.csbj.2020.11.049

Bolón-Canedo, V., Sánchez-Maroño, N., and Alonso-Betanzos, A. (2016). Feature selection for high-dimensional data. *Prog. Artif. Intell.* 5, 65–75. doi: 10.1007/s13748-015-0080-y

Borén, M. (2015). "Sample preservation Through heat stabilization of proteins: principles and examples," in *Proteomic Profiling Methods in Molecular Biology*. ed. A. Posch (New York, NY: Springer), 21–32.

Bremges, A., Fritz, A., and McHardy, A. C. (2020). CAMITAX: taxon labels for microbial genomes. *Giga Science* 9:giz154. doi: 10.1093/gigascience/giz154

Brito, I. L., and Alm, E. J. (2016). Tracking strains in the microbiome: insights from metagenomics and models. *Front. Microbiol.* 7:712. doi: 10.3389/fmicb.2016.00712

Broeksema, B., Calusinska, M., McGee, F., Winter, K., Bongiovanni, F., Goux, X., et al. (2017). ICoVeR – an interactive visualization tool for verification and refinement of metagenomic bins. *BMC Bioinformatics* 18:233. doi: 10.1186/s12859-017-1653-5

Brown, C. T., Howe, A., Zhang, Q., Pyrkosz, A. B., and Brom, T. H. (2012). A Reference-Free Algorithm for Computational Normalization of Shotgun Sequencing Data. arXiv [Preprint].

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Cao, H.-T., Gibson, T. E., Bashan, A., and Liu, Y.-Y. (2017). Inferring human microbial dynamics from temporal metagenomics data: pitfalls and lessons. *BioEssays* 39:1600188. doi: 10.1002/bies.201600188

Carmody, R. N., Bisanz, J. E., Bowen, B. P., Maurice, C. F., Lyalina, S., Louie, K. B., et al. (2019). Cooking shapes the structure and function of the gut microbiome. *Nat. Microbiol.* 4, 2052–2063. doi: 10.1038/s41564-019-0569-4

Celaj, A., Markle, J., Danska, J., and Parkinson, J. (2014). Comparison of assembly algorithms for improving rate of metatranscriptomic functional annotation. *Microbiome* 2:39. doi: 10.1186/2049-2618-2-39

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848

Chen, L.-X., Anantharaman, K., Shaiber, A., Eren, A. M., and Banfield, J. F. (2020). Accurate and complete genomes from metagenomes. *Genome Res.* 30, 315–333. doi: 10.1101/gr.258640.119

Chen, L., Reeve, J., Zhang, L., Huang, S., Wang, X., and Chen, J. (2018). GMPR: A robust normalization method for zero-inflated count data with application to microbiome sequencing data. *PeerJ* 6:e4600. doi: 10.7717/peerj.4600

Coenen, A. R., Hu, S. K., Luo, E., Muratore, D., and Weitz, J. S. (2020). A primer for microbiome time-series analysis. *Front. Genet.* 11:310. doi: 10.3389/fgene.2020.00310

Costa Junior, C., Corbeels, M., Bernoux, M., Píccolo, M. C., Siqueira Neto, M., Feigl, B. J., et al. (2013). Assessing soil carbon storage rates under no-tillage: comparing the synchronic and diachronic approaches. *Soil Tillage Res.* 134, 207–212. doi: 10.1016/j.still.2013.08.010

David, L. A., Materna, A. C., Friedman, J., Campos-Baptista, M. I., Blackburn, M. C., Perrotta, A., et al. (2014). Host lifestyle affects human microbiota on daily timescales. *Genome Biol.* 15:R89. doi: 10.1186/gb-2014-15-7-r89

de Goffau, M. C., Charnock-Jones, D. S., Smith, G. C. S., and Parkhill, J. (2021). Batch effects account for the main findings of an in utero human intestinal bacterial colonization study. *Microbiome* 9:6. doi: 10.1186/s40168-020-00949-z

Delmont, T. O., and Eren, A. M. (2018). Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ* 6:e4320. doi: 10.7717/peerj.4320

Delogu, F., Kunath, B. J., Evans, P. N., Arntzen, M. Ø., Hvidsten, T. R., and Pope, P. B. (2020). Integration of absolute multi-omics reveals dynamic protein-to-RNA ratios and metabolic interplay within mixed-domain microbiomes. *Nat. Commun.* 11:4708. doi: 10.1038/s41467-020-18543-0

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R., and Weyrich, L. S. (2019). Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends Microbiol.* 2, 105–117. doi: 10.1016/j.tim.2018.11.003

Erickson, A. R., Cantarel, B. L., Lamendella, R., Darzi, Y., Mongodin, E. F., Pan, C., et al. (2012). Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. *PLoS One* 7:e49138. doi: 10.1371/journal.pone.0049138

Evans, J. T., and Denef, V. J. (2020). To dereplicate or not to dereplicate? *mSphere* 5:e00971-19. doi: 10.1128/mSphere.00971-19

Faust, K., Bauchinger, F., Laroche, B., de Buyl, S., Lahti, L., Washburne, A. D., et al. (2018). Signatures of ecological processes in microbial community time series. *Microbiome* 6:120. doi: 10.1186/s40168-018-0496-2

Faust, K., Lahti, L., Gonze, D., de Vos, W. M., and Raes, J. (2015). Metagenomics meets time series analysis: unraveling microbial community dynamics. *Curr. Opin. Microbiol.* 25, 56–66. doi: 10.1016/j.mib.2015.04.004

Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., et al. (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* 8:e1002606. doi: 10.1371/journal.pcbi.1002606

Fidler, G., Tolnai, E., Stagel, A., Remenyik, J., Stundl, L., Gal, F., et al. (2020). Tendentious effects of automated and manual metagenomic DNA purification protocols on broiler gut microbiome taxonomic profiling. *Sci. Rep.* 10:3419. doi: 10.1038/s41598-020-60304-y

Fiedorová, K., Radvanský, M., Němcová, E., Grombiříková, H., Bosák, J., Černochová, M., et al. (2019). The impact of DNA extraction methods on stool bacterial and fungal microbiota community recovery. *Front. Microbiol.* 10:821. doi: 10.3389/fmicb.2019.00821

Fisher, C. K., and Mehta, P. (2014). Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PLoS One* 9:e102451. doi: 10.1371/journal.pone.0102451

Fondi, M., and Liò, P. (2015). Multi-omics and metabolic modelling pipelines: challenges and tools for systems microbiology. *Microbiol. Res.* 171, 52–64. doi: 10.1016/j.micres.2015.01.003

Franzosa, E. A., McIver, L. J., Rahnavard, G., Thompson, L. R., Schirmer, M., Weingart, G., et al. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* 15, 962–968. doi: 10.1038/s41592-018-0176-y

Friedman, J., and Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* 8:e1002687. doi: 10.1371/journal.pcbi.1002687

Garza, D. R., van Verk, M. C., Huynen, M. A., and Dutilh, B. E. (2018). Towards predicting the environmental metabolome from metagenomics with a mechanistic model. *Nat. Microbiol.* 3, 456–460. doi: 10.1038/s41564-018-0124-8

Gerber, G. K. (2014). The dynamic microbiome. *FEBS Lett.* 588, 4131–4139. doi: 10.1016/j.febslet.2014.02.037

Gibbons, S. M., Duvallet, C., and Alm, E. J. (2018). Correcting for batch effects in case-control microbiome studies. *PLoS Comput. Biol.* 14:e1006102. doi: 10.1371/journal.pcbi.1006102

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J. (2017). Microbiome datasets are compositional: and this is not optional. *Front. Microbiol.* 8:2224. doi: 10.3389/fmicb.2017.02224

Gonze, D., Coyte, K. Z., Lahti, L., and Faust, K. (2018). Microbial communities as dynamical systems. *Curr. Opin. Microbiol.* 44, 41–49. doi: 10.1016/j.mib.2018.07.004

Heintz-Buschart, A., May, P., Laczny, C. C., Lebrun, L. A., Bellora, C., Krishna, A., et al. (2016). Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nat. Microbiol.* 2:16180. doi: 10.1038/nmicrobiol.2016.227

Heintz-Buschart, A., and Wilmes, P. (2018). Human gut microbiome: function matters. *Trends Microbiol.* 26, 563–574. doi: 10.1016/j.tim.2017.11.002

Heintz-Buschart, A., Yusuf, D., Kaysen, A., Etheridge, A., Fritz, J. V., May, P., et al. (2018). Small RNA profiling of low biomass samples: identification and removal of contaminants. *BMC Biol.* 16:52. doi: 10.1186/s12915-018-0522-7

Herold, M., Arbas, S. M., Narayanasamy, S., Sheik, A. R., Kleine-Borgmann, L. A. K., Lebrun, L. A., et al. (2020). Integration of time-series meta-omics data reveals how microbial ecosystems respond to disturbance. *Nat. Commun.* 11:5281. doi: 10.1038/s41467-020-19006-2

Heyer, R., Schallert, K., Zoun, R., Becher, B., Saake, G., and Benndorf, D. (2017). Challenges and perspectives of metaproteomic data analysis. *J. Biotechnol.* 261, 24–36. doi: 10.1016/j.jbiotec.2017.06.1201

Hickl, O., Heintz-Buschart, A., Trautwein-Schult, A., Hercog, R., Bork, P., Wilmes, P., et al. (2019). Sample preservation and storage significantly impact taxonomic and functional profiles in metaproteomics studies of the human gut microbiome. *Microorganisms* 7:367. doi: 10.3390/microorganisms7090367

Hornung, B. V. H., Zwittink, R. D., and Kuijper, E. J. (2019). Issues and current standards of controls in microbiome research. *FEMS Microbiol. Ecol.* 95:fiz045. doi: 10.1093/femsec/fiz045

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Jiang, D., Armour, C. R., Hu, C., Mei, M., Tian, C., Sharpton, T. J., et al. (2019). Microbiome multi-omics network analysis: statistical considerations,

limitations, and opportunities. *Front. Genet.* 10:995. doi: 10.3389/fgene.2019.00995

Jiang, R., Li, W. V., and Li, J. J. (2020). mbImpute: an accurate and robust imputation method for microbiome data. *Genomics* [Preprint]. doi: 10.1101/2020.03.07.982314

Johnston, J., LaPara, T., and Behrens, S. (2019). Composition and dynamics of the activated sludge microbiome during seasonal nitrification failure. *Sci. Rep.* 9:4565. doi: 10.1038/s41598-019-40872-4

Jurburg, S. D., Konzack, M., Eisenhauer, N., and Heintz-Buschart, A. (2020). The archives are half-empty: an assessment of the availability of microbial community sequencing data. *Commun. Biol.* 3:474. doi: 10.1038/s42003-020-01204-9

Kapoore, R. V., and Vaidyanathan, S. (2016). Towards quantitative mass spectrometry-based metabolomics in microbial and mammalian systems. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* 374:20150363. doi: 10.1098/rsta.2015.0363

Kashtan, N., Roggensack, S. E., Rodrigue, S., Thompson, J. W., Biller, S. J., Coe, A., et al. (2014). Single-cell genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. *Science* 344, 416–420. doi: 10.1126/science.1248575

Kaysen, A., Heintz-Buschart, A., Muller, E. E. L., Narayanasamy, S., Wampach, L., Laczny, C. C., et al. (2017). Integrated meta-omic analyses of the gastrointestinal tract microbiome in patients undergoing allogeneic hematopoietic stem cell transplantation. *Transl. Res.* 186, 79–94. doi: 10.1016/j.trsl.2017.06.008

Kumar, R., Eipers, P., Little, R. B., Crowley, M., Crossman, D. K., Lefkowitz, E. J., et al. (2014). Getting started with microbiome analysis: sample acquisition to bioinformatics. *Curr. Protoc. Hum. Genet.* 82, 18.8.1–18.8.29. doi: 10.1002/0471142905.hg1808s82

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Lapidus, A. L., and Korobeynikov, A. I. (2021). Metagenomic data assembly – the way of decoding unknown microorganisms. *Front. Microbiol.* 12:613791. doi: 10.3389/fmicb.2021.613791

Law, Y., Kirkegaard, R. H., Cokro, A. A., Liu, X., Arumugam, K., Xie, C., et al. (2016). Integrative microbial community analysis reveals full-scale enhanced biological phosphorus removal under tropical conditions. *Sci. Rep.* 6:25719. doi: 10.1038/srep25719

Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., et al. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* 11, 733–739. doi: 10.1038/nrg2825

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., et al. (2014). An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* 32, 834–841. doi: 10.1038/nbt.2942

Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., et al. (2016). MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* 102, 3–11. doi: 10.1016/j.ymeth.2016.02.020

Liang, Y., Dong, T., Chen, M., He, L., Wang, T., Liu, X., et al. (2020). Systematic analysis of impact of sampling regions and storage methods on Fecal gut microbiome and metabolome profiles. *mSphere* 5:e00763-19. doi: 10.1128/mSphere.00763-19

Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., et al. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 569, 655–662. doi: 10.1038/s41586-019-1237-9

Machiels, B. M., Ruers, T., Lindhout, M., Hardy, K., Hlavaty, T., Bang, D. D., et al. (2000). New protocol for DNA extraction of stool. *Bio Techniques* 28, 286–290. doi: 10.2144/00282st05

Magasin, J. D., and Gerloff, D. L. (2015). Pooled assembly of marine metagenomic datasets: enriching annotation through chimerism. *Bioinformatics* 31, 311–317. doi: 10.1093/bioinformatics/btu546

Mallick, H., Franzosa, E. A., McIver, L. J., Banerjee, S., Sirota-Madi, A., Kostic, A. D., et al. (2019). Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. *Nat. Commun.* 10:3136. doi: 10.1038/s41467-019-10927-1

Mallick, H., Rahnavard, A., McIver, L. J., Ma, S., Zhang, Y., Nguyen, L. H., et al. (2021). Multivariable association discovery in population-scale meta-omics studies. *Microbiology* [Preprint]. doi: 10.1099/mic.0.001031

Mangul, S., Martin, L. S., Eskin, E., and Blekhman, R. (2019). Improving the usability and archival stability of bioinformatics software. *Genome Biol.* 20:47. doi: 10.1186/s13059-019-1649-8

Mars, R. A. T., Yang, Y., Ward, T., Houtti, M., Priya, S., Lekatz, H. R., et al. (2020). Longitudinal multi-omics reveals subset-specific mechanisms underlying irritable bowel syndrome. *Cell* 182, 1460–1473. doi: 10.1016/j.cell.2020.08.007

Martínez Arbas, S. M., Narayanasamy, S., Herold, M., Lebrun, L. A., Hoopmann, M. R., Li, S., et al. (2021). Roles of bacteriophages, plasmids and CRISPR immunity in microbial community dynamics revealed using time-series integrated meta-omics. *Nat. Microbiol.* 6, 123–135. doi: 10.1038/s41564-020-00794-8

McLaren, M. R., Willis, A. D., and Callahan, B. J. (2019). Consistent and correctable bias in metagenomic sequencing experiments. *elife* 8:e46923. doi: 10.7554/eLife.46923

Mirdita, M., Steinegger, M., and Söding, J. (2019). MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* 35, 2856–2858. doi: 10.1093/bioinformatics/bty1057

Moejes, F., Succurro, A., Popa, O., Maguire, J., and Ebenhöh, O. (2017). Dynamics of the bacterial community associated with *Phaeodactylum tricornutum* cultures. *Processes* 5:77. doi: 10.3390/pr5040077

Morton, J. T., Marotz, C., Washburne, A., Silverman, J., Zaramela, L. S., Edlund, A., et al. (2019). Establishing microbial composition measurement standards with reference frames. *Nat. Commun.* 10:2719. doi: 10.1038/s41467-019-10656-5

Moss, E. L., Maghini, D. G., and Bhatt, A. S. (2020). Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat. Biotechnol.* 38, 701–707. doi: 10.1038/s41587-020-0422-6

Muller, E. E. L. (2019). Determining microbial niche breadth in the environment for better ecosystem fate predictions. *mSystems* 4:e00080-19. doi: 10.1128/mSystems.00080-19

Muller, E. E. L., Faust, K., Widder, S., Herold, M., Arbas, S. M., and Wilmes, P. (2018). Using metabolic networks to resolve ecological properties of microbiomes. *Curr. Opin. Syst. Biol.* 8, 73–80. doi: 10.1016/j.coisb.2017.12.004

Muller, E. E. L., Glaab, E., May, P., Vlassis, N., and Wilmes, P. (2013). Condensing the omics fog of microbial communities. *Trends Microbiol.* 21, 325–333. doi: 10.1016/j.tim.2013.04.009

Muller, E. E. L., Pinel, N., Laczny, C. C., Hoopmann, M. R., Narayanasamy, S., Lebrun, L. A., et al. (2014). Community-integrated omics links dominance of a microbial generalist to fine-tuned resource usage. *Nat. Commun.* 5:5603. doi: 10.1038/ncomms6603

Narayanasamy, S., Jarosz, Y., Muller, E. E. L., Heintz-Buschart, A., Herold, M., Kaysen, A., et al. (2016). IMP: a pipeline for reproducible reference-independent integrated metagenomic and metatranscriptomic analyses. *Genome Biol.* 17:260. doi: 10.1186/s13059-016-1116-8

Narayanasamy, S., Muller, E. E. L., Sheik, A. R., and Wilmes, P. (2015). Integrated omics for the identification of key functionalities in biological wastewater treatment microbial communities. *Microb. Biotechnol.* 8, 363–368. doi: 10.1111/1751-7915.12255

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., et al. (2020). A genomic catalog of earth's microbiomes. *Nat. Biotechnol.* 39, 499–509. doi: 10.1038/s41587-020-0718-6

Noecker, C., Eng, A., Srinivasan, S., Theriot, C. M., Young, V. B., Jansson, J. K., et al. (2016). Metabolic model-based integration of microbiome taxonomic and metabolomic profiles elucidates mechanistic links between ecological and metabolic variation. *mSystems* 1:e00013-15. doi: 10.1128/mSystems.00013-15

Noor, E., Cherkaoui, S., and Sauer, U. (2019). Biological insights through omics data integration. *Gene Regul.* 15, 39–47. doi: 10.1016/j.coisb.2019.03.007

Oh, S., Li, C., Baldwin, R. L., Song, S., Liu, F., and Li, R. W. (2019). Temporal dynamics in meta longitudinal RNA-Seq data. *Sci. Rep.* 9:763. doi: 10.1038/s41598-018-37397-7

Olm, M. R., Brown, C. T., Brooks, B., and Banfield, J. F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* 11, 2864–2868. doi: 10.1038/ismej.2017.126

Park, S.-Y., Ufondu, A., Lee, K., and Jayaraman, A. (2020). Emerging computational tools and models for studying gut microbiota composition and function. *Tissue Cell Pathw. Eng.* 66, 301–311. doi: 10.1016/j.copbio.2020.10.005

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). Check M: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114

Peña-Llopis, S., and Brugarolas, J. (2013). Simultaneous isolation of high-quality DNA, RNA, miRNA and proteins from tissues for genomic applications. *Nat. Protoc.* 8, 2240–2255. doi: 10.1038/nprot.2013.141

Pereira, M. B., Wallroth, M., Jonsson, V., and Kristiansson, E. (2018). Comparison of normalization methods for the analysis of metagenomic gene abundance data. *BMC Genomics* 19:274. doi: 10.1186/s12864-018-4637-6

Phillips, K., McCallum, N., and Welch, L. (2012). A comparison of methods for forensic DNA extraction: Chelex-100® and the QIAGEN DNA Investigator Kit (manual and automated). *Forensic Sci. Int. Genet.* 6, 282–285. doi: 10.1016/j.fsigen.2011.04.018

Pible, O., Allain, F., Jouffret, V., Culotta, K., Miotello, G., and Armengaud, J. (2020). Estimating relative biomasses of organisms in microbiota using "phylopeptidomics". *Microbiome* 8:30. doi: 10.1186/s40168-020-00797-x

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65. doi: 10.1038/nature08821

Queirós, P., Delogu, F., Hickl, O., May, P., and Wilmes, P. (2020). Mantis: flexible and consensus-driven genome annotation. *Bioinformatics* [Preprint]. doi: 10.1101/2020.11.02.360933

Quince, C., Nurk, S., Raguideau, S., James, R., Soyer, O. S., Summers, J. K., et al. (2020). Metagenomics strain resolution on assembly graphs. *Bioinformatics* [Preprint]. doi: 10.1101/2020.09.06.284828

Rho, M., Tang, H., and Ye, Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res.* 38:e191. doi: 10.1093/nar/gkq747

Ridenhour, B. J., Brooker, S. L., Williams, J. E., Van Leuven, J. T., Miller, A. W., Dearing, M. D., et al. (2017). Modeling time-series data from microbial communities. *ISME J.* 11, 2526–2537. doi: 10.1038/ismej.2017.107

Røst, L. M., Brekke Thorfinnsdottir, L., Kumar, K., Fuchino, K., Eide Langørgen, I., Bartosova, Z., et al. (2020). Absolute quantification of the central carbon metabolome in eight commonly applied prokaryotic and eukaryotic model systems. *Metabolites* 10:74. doi: 10.3390/metabo10020074

Roume, H., Heintz-Buschart, A., Muller, E. E. L., May, P., Satagopam, V. P., Laczny, C. C., et al. (2015). Comparative integrated omics: identification of key functionalities in microbial community-wide metabolic networks. *Npj Biofilms Microbiomes* 1:15007. doi: 10.1038/npjbiofilms.2015.7

Roume, H., Heintz-Buschart, A., Muller, E. E. L., and Wilmes, P. (2013b). "Sequential isolation of metabolites, RNA, DNA, and proteins from the same unique sample," in *Methods in Enzymology*. ed. E. F. DeLong (Cambridge, Massachusetts, United States: Elsevier), 219–236.

Roume, H., Muller, E. E., Cordes, T., Renaut, J., Hiller, K., and Wilmes, P. (2013a). A biomolecular isolation framework for eco-systems biology. *ISME J.* 7, 110–121. doi: 10.1038/ismej.2012.72

Ruiz-Perez, D., Lugo-Martinez, J., Bourguignon, N., Mathee, K., Lerner, B., Bar-Joseph, Z., et al. (2021). Dynamic Bayesian networks for integrating multi-omics time series microbiome data. *mSystems* 6:e01105-20. doi: 10.1128/mSystems.01105-20

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., et al. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 12:87. doi: 10.1186/s12915-014-0087-z

Sandve, G. K., Nekrutenko, A., Taylor, J., and Hovig, E. (2013). Ten simple rules for reproducible computational research. *PLoS Comput. Biol.* 9:e1003285. doi: 10.1371/journal.pcbi.1003285

Santiago, A., Panda, S., Mengels, G., Martinez, X., Azpiroz, F., Dore, J., et al. (2014). Processing faecal samples: a step forward for standards in microbial community analysis. *BMC Microbiol.* 14:112. doi: 10.1186/1471-2180-14-112

Schoenenberger, A. W., Muggli, F., Parati, G., Gallino, A., Ehret, G., Suter, P. M., et al. (2016). Protocol of the Swiss Longitudinal Cohort Study (SWICOS) in rural Switzerland. *BMJ Open* 6:e013280. doi: 10.1136/bmjopen-2016-013280

Schriml, L. M., Chuvochina, M., Davies, N., Eloe-Fadrosh, E. A., Finn, R. D., Hugenholtz, P., et al. (2020). COVID-19 pandemic reveals the peril of ignoring metadata standards. *Sci. Data* 7:188. doi: 10.1038/s41597-020-0524-5

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sefer, E., Kleyman, M., and Bar-Joseph, Z. (2016). Tradeoffs between dense and replicate sampling strategies for high-throughput time series experiments. *Cell Syst.* 3, 35–42. doi: 10.1016/j.cels.2016.06.007

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* 9, 811–814. doi: 10.1038/nmeth.2066

Shahin, M., Ali Babar, M., and Zhu, L. (2017). Continuous integration, delivery and deployment: a systematic review on approaches, tools, challenges and practices. *IEEE Access* 5, 3909–3943. doi: 10.1109/ACCESS.2017.2685629

Sieber, C. M. K., Probst, A. J., Sharrar, A., Thomas, B. C., Hess, M., Tringe, S. G., et al. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* 3, 836–843. doi: 10.1038/s41564-018-0171-1

Silverman, J. D., Roche, K., Mukherjee, S., and David, L. A. (2020). Naught all zeros in sequence count data are the same. *Comput. Struct. Biotechnol. J.* 18, 2789–2798. doi: 10.1016/j.csbj.2020.09.014

Sokal, R. R. (1995). *Biometry: The Principles and Practice of Statistics in Biological Research.* 3rd Edn. New York: W.H. Freeman.

Starr, E. P., Shi, S., Blazewicz, S. J., Probst, A. J., Herman, D. J., Firestone, M. K., et al. (2018). Stable isotope informed genome-resolved metagenomics reveals that Saccharibacteria utilize microbially-processed plant-derived carbon. *Microbiome* 6:122. doi: 10.1186/s40168-018-0499-z

Stewart, C. J., Ajami, N. J., O'Brien, J. L., Hutchinson, D. S., Smith, D. P., Wong, M. C., et al. (2018). Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* 562, 583–588. doi: 10.1038/s41586-018-0617-x

Sunagawa, S., Mende, D. R., Zeller, G., Izquierdo-Carrasco, F., Berger, S. A., Kultima, J. R., et al. (2013). Metagenomic species profiling using universal phylogenetic marker genes. *Nat. Methods* 10, 1196–1199. doi: 10.1038/nmeth.2693

Tanca, A., Abbondio, M., Palomba, A., Fraumene, C., Manghina, V., Cucca, F., et al. (2017). Potential and active functions in the gut microbiota of a healthy human cohort. *Microbiome* 5:79. doi: 10.1186/s40168-017-0293-3

Tanca, A., Palomba, A., Fraumene, C., Pagnozzi, D., Manghina, V., Deligios, M., et al. (2016). The impact of sequence database choice on metaproteomic results in gut microbiota studies. *Microbiome* 4:51. doi: 10.1186/s40168-016-0196-8

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc. Natl. Acad. Sci. U. S. A.* 102, 13950–13955. doi: 10.1073/pnas.0506758102

Thorn, C. E., Bergesch, C., Joyce, A., Sambrano, G., McDonnell, K., Brennan, F., et al. (2019). A robust, cost-effective method for DNA, RNA and protein co-extraction from soil, other complex microbiomes and pure cultures. *Mol. Ecol. Resour.* 19, 439–455. doi: 10.1111/1755-0998.12979

Timmins-Schiffman, E., May, D. H., Mikan, M., Riffle, M., Frazar, C., Harvey, H. R., et al. (2017). Critical decisions in metaproteomics: achieving high confidence protein annotations in a sea of unknowns. *ISME J.* 11, 309–314. doi: 10.1038/ismej.2016.132

Treangen, T. J., Koren, S., Sommer, D. D., Liu, B., Astrovskaya, I., Ondov, B., et al. (2013). MetAMOS: a modular and open source metagenomic assembly and analysis pipeline. *Genome Biol.* 14:R2. doi: 10.1186/gb-2013-14-1-r2

Trosvik, P., de Muinck, E. J., and Stenseth, N. C. (2015). Biotic interactions and temporal dynamics of the human gastrointestinal microbiota. *ISME J.* 9, 533–541. doi: 10.1038/ismej.2014.147

Uritskiy, G. V., DiRuggiero, J., and Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6:158. doi: 10.1186/s40168-018-0541-1

van der Walt, A. J., van Goethem, M. W., Ramond, J.-B., Makhalanyane, T. P., Reva, O., and Cowan, D. A. (2017). Assembling metagenomes, one community at a time. *BMC Genomics* 18:521. doi: 10.1186/s12864-017-3918-9

Wampach, L., Heintz-Buschart, A., Fritz, J. V., Ramiro-Garcia, J., Habier, J., Herold, M., et al. (2018). Birth mode is associated with earliest strain-conferred gut microbiome functions and immunostimulatory potential. *Nat. Commun.* 9:5091. doi: 10.1038/s41467-018-07631-x

Wang, Y., and Cao, K.-A. L. (2019). Managing batch effects in microbiome data. *Brief. Bioinform.* 21, 1954–1970. doi: 10.1093/bib/bbz105

Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., et al. (2016). Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J.* 10, 1669–1681. doi: 10.1038/ismej.2015.235

Weyrich, L. S., Farrer, A. G., Eisenhofer, R., Arriola, L. A., Young, J., Selway, C. A., et al. (2019). Laboratory contamination over time during low-biomass sample analysis. *Mol. Ecol. Resour.* 19, 982–996. doi: 10.1111/1755-0998.13011

Wickramarachchi, A., Mallawaarachchi, V., Rajan, V., and Lin, Y. (2020). MetaBCC-LR: metagenomics binning by coverage and composition for long reads. *Bioinformatics* 36, i3–i11. doi: 10.1093/bioinformatics/btaa441

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., and Teal, T. K. (2017). Good enough practices in scientific computing. *PLoS Comput. Biol.* 13:e1005510. doi: 10.1371/journal.pcbi.1005510

Wood, D. E., and Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15:R46. doi: 10.1186/gb-2014-15-3-r46

Yu, K., Yi, S., Li, B., Guo, F., Peng, X., Wang, Z., et al. (2019). An integrated meta-omics approach reveals substrates involved in synergistic interactions in a bisphenol A (BPA)-degrading microbial community. *Microbiome* 7:16. doi: 10.1186/s40168-019-0634-5

Yue, Y., Huang, H., Qi, Z., Dou, H.-M., Liu, X.-Y., Han, T.-F., et al. (2020). Evaluating metagenomics tools for genome binning with real metagenomic datasets and CAMI datasets. *BMC Bioinformatics* 21:334. doi: 10.1186/s12859-020-03667-3

Zhou, Z., Tran, P. Q., Breiser, A. M., Liu, Y., Kieft, K., Cowley, E. S., et al. (2020). METABOLIC: high-throughput profiling of microbial genomes for functional traits, biogeochemistry, and community-scale metabolic networks. *bioRxiv* [Preprint]. doi: 10.1101/2020.10.27.357558

Zimmermann, J., Kaleta, C., and Waschina, S. (2021). gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biol.* 22:81. doi: 10.1186/s13059-021-02295-1

Zinter, M. S., Mayday, M. Y., Ryckman, K. K., Jelliffe-Pawlowski, L. L., and DeRisi, J. L. (2019). Towards precision quantification of contamination in metagenomic sequencing experiments. *Microbiome* 7, 62. doi: 10.1186/s40168-019-0678-6

Zlitni, S., Bishara, A., Moss, E. L., Tkachenko, E., Kang, J. B., Culver, R. N., et al. (2020). Strain-resolved microbiome sequencing reveals mobile elements that drive bacterial competition on a clinical timescale. *Genome Med.* 12:50. doi: 10.1186/s13073-020-00747-0