A peer-reviewed version of this preprint was published in PeerJ on 26 March 2019.

García-Jiménez B, Wilkinson MD. 2019. Robust and automatic definition of microbiome states. PeerJ 7:e6657 https://doi.org/10.7717/peerj.6657

Automatic definition of robust microbiome sub-states in longitudinal data

Beatriz García-Jiménez [Corresp., 1, 2] , **Mark D Wilkinson** [1]

[1] Biological Informatics Group, Centro de Biotecnologıa y Genomica de Plantas UPM-INIA, Universidad Politécnica de Madrid, Madrid, Spain

[2] Systems Biotechnology Group, Centro Nacional de Biotecnología, Spanish National Research Council, Madrid, Spain

Corresponding Author: Beatriz García-Jiménez
Email address: beatriz.garcia@csic.es

The analysis of microbiome dynamics would allow us to elucidate patterns within microbial community evolution; however, microbiome state-transition dynamics have been scarcely studied. This is in part because a necessary first-step in such analyses has not been well-defined: how to deterministically describe a microbiome's "state". Clustering in states have been widely studied, although no standard has been concluded yet. We propose a generic, domain-independent and automatic procedure to determine a reliable set of microbiome sub-states within a specific dataset, and with respect to the conditions of the study. The robustness of sub-state identification is established by the combination of diverse techniques for stable cluster verification. We reuse four distinct longitudinal microbiome datasets to demonstrate the broad applicability of our method, analysing results with different taxa subset allowing to adjust it depending on the application goal, and showing that the methodology provides a set of robust sub-states to examine in downstream studies about dynamics in microbiome.

Automatic Definition of Robust Microbiome Sub-states in Longitudinal Data

Beatriz García-Jiménez[1,2] and Mark D. Wilkinson[1]

[1]**Biological Informatics Group, Centro de Biotecnología y Genómica de Plantas UPM-INIA, Universidad Politécnica de Madrid, Campus Montegancedo, 28223-Pozuelo de Alarcón (Madrid), Spain.**
[2]**Current affiliation: Systems Biotechnology Group, Centro Nacional de Biotecnología, Spanish National Research Council, C/ Darwin nº 3, Campus de Cantoblanco, 28049-Madrid, Spain.**

Corresponding author:
Beatriz García-Jiménez[1,2]

Email address: beatriz.garcia@csic.es

ABSTRACT

The analysis of microbiome dynamics would allow us to elucidate patterns within microbial community evolution; however, microbiome state-transition dynamics have been scarcely studied. This is in part because a necessary first-step in such analyses has not been well-defined: how to deterministically describe a microbiome's "state". Clustering in states have been widely studied, although no standard has been concluded yet. We propose a generic, domain-independent and automatic procedure to determine a reliable set of microbiome sub-states within a specific dataset, and with respect to the conditions of the study. The robustness of sub-state identification is established by the combination of diverse techniques for stable cluster verification. We reuse four distinct longitudinal microbiome datasets to demonstrate the broad applicability of our method, analysing results with different taxa subset allowing to adjust it depending on the application goal, and showing that the methodology provides a set of robust sub-states to examine in downstream studies about dynamics in microbiome.

INTRODUCTION

This manuscript addresses an important challenge in microbiome analysis: identification and description of longitudinal microbiome variability and dynamics (Gilbert et al., 2016; Bashan et al., 2016; Bradley and Pollard, 2017). Microbiomes are distinct between different cavities (or more generally, distinct environments) (Bashan et al., 2016). One widely-accepted view of the microbiome within a given cavity is to consider it from a global perspective - that among all individuals there is one shared state, broadly-defined, but internally-variable and dynamic (Gibbons et al., 2017). With respect to the gut microbes, there were initial attempts to resolve distinct microbial population structures that were associated with "habits" such as lifestyle, culture, or eating - called "enterotypes". Albeit, there is a big discussion about enterotypes (whether, how many and which ones) (Costea et al., 2018). Some studies are agree with discrete states (Zhou et al., 2014; Turroni et al., 2017), while other ones with gradients (MacDonald et al., 2012; Gibbons et al., 2017). It is well-recognized that distinct and predictable microbial compositions are associated with important traits such as health (Gilbert et al., 2016; Shankar, 2017). Nevertheless, even if there were no universally-distinct enterotypes, this would not imply that microbiomes cannot exist in long- or short-term stable sub-states. Moreover, within a given individual, studies suggest the existence of such stable steady-states, both in experimental data (Turroni et al., 2017) and in modeling approaches (Stein et al., 2013; Bashan et al., 2016), not considering a state as a constant community composition but an average along a period of time (Chan et al., 2017). As such, the proposition of a single, but continuously varying, microbiome state within a given environment fails to recognize subtle differences that may have significant biological consequences. Thus, the definition of several sub-states could help to capture slight microbiome shifts, happened even in a general stability situation, and undetectable in other approximations, where only one state or a few with strong divergence were considered.

Faust et al. (2015) reported that microbial diversity was quite stable over time, in a stable environment, but that stability could be disrupted by a) external perturbations, b) direct modifications or c) transient perturbations, all of which cause the microbial community to change. Longitudinal studies focused on microbiome changes over time, in a particular habitat, with known interventions, would elucidate the mechanisms behind microbial state transitions, and from this information, universal or context-specific interventions could be determined (Bashan et al., 2016), and used to manipulate the microbiome in desirable ways. Thus a domain-independent (e.g. human/animal body cavities, soils, industrial microbial communities in bioreactors, etc.) and flexible approach to analysis of longitudinal microbiome variability and dynamics is needed.

In this manuscript, we define a microbiome "sub-state" as a collection of constraints satisfied by the microbiota of a given sample, that are not satisfied by other samples, allowing them to be reliably distinguished from one another. Thus, over time, and depending on perturbations, a dynamic microbial community could move through different microbiome sub-states, even within a single "enterotype" as defined by Arumugam et al. (2011b) (or its conceptual equivalent in non-gut environments). Thus, we consider microbiome sub-states to be akin to biomarkers, and importantly, these biomarkers differ between individuals or even within the same one, as "a measurable indicator of a biological state or environmental exposure" (Gorvitovskaia et al., 2016). Finally, our definition of microbiome sub-state describes the microbial community composition over a specific fragment of time.

There are relatively few studies about temporal dynamics within any microbiome site (Gajer et al., 2012; Ding and Schloss, 2014), and for those studies, the approach to microbiome sample clustering is assorted and somewhat *ad-hoc*. This masks the fact that the definition of states in a temporal sequence of microbiomes is non-trivial, and has only been attempted within specific datasets (Koren et al., 2013). Therefore, we believe that a generally-applicable algorithm that detects robust and stable microbiome sub-states within any input dataset, would be a beneficial contribution to the community, and would aid in the cross-replication and comparison of studies in the future. We report such a methodology here.

The inputs to our pipeline are Operational Taxonomic Unit (OTU) vectors, where each OTU represents a group of species considered indistinguishable by the OTU grouping process, and whose abundance in a set of samples has then been computed. Given that these abundance distributions are effectively continuous, the list of possible OTU combinations as possible sub-states clearly must be simplified to ensure that the space of the OTU vector is so large as to be computationally intractable. On the other hand, increasing simplification results in an vanishingly small numbers of sub-states, which are insufficiently granular as to occur in any "biologically meaningful"association. This is, therefore, the key consideration when defining the approximation for grouping similar OTU vectors, for example, by machine learning clustering approaches. Effectively, our choice of clustering parameters should be guided by the desire to identify several well-populated microbiome sub-states both within and between individuals, which can then be used as the basis of a model associating these sub-states with various biologically interesting phenomenon. It should be noted that this is a distinct goal from that of most microbiome studies, which consider the stable-state microbiome as a single entity (e.g. the enterotype studies). As such, the 'default' clustering parameters that appear in most published approaches do not match our problem requirements, and must be re-considered from scratch.

Our overall procedure for defining sub-states, described in detail below, consists of applying a clustering algorithm to the OTU data, taking a metagenomics beta diversity metric as the distance measure between samples. We then attempt to robustly define the optimal number of clusters based on a comparison between several distance measures, distinct algorithms and different clustering scores.

Metagenomics sample clustering has been achieved for different studies using a variety of approaches, in terms of distance measure, algorithm and number of clusters. For example, as a distance measure, the Jensen-Shanon Distance (JSD) (or its root squared, rJSD, as in (Arumugam et al., 2011b)) is the most frequently used (Gajer et al., 2012); although the cophenetic or the Euclidean distance are also sometimes applied. Several clustering algorithms have been used to group metagenomics samples, such as PAM, Agnes, Hclust, or Dirichlet Multinomial Mixture, with different linkage options (Ding and Schloss, 2014). For determination of the number of clusters, diverse assessment criteria have been used in the literature: the average Silhouette width (SI), Calinski-Harabasz (CH) index, Laplace approximation, etc. In the specific case of enterotypes, clustering of samples was applied (Arumugam et al., 2011b). According their tutorial (Arumugam et al., 2011a), they computed the distance as the root square of the JSD, with the PAM algorithm and selecting the number of clusters with the CH index combined with a SI assessment.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26657v1 | CC BY 4.0 Open Access | rec: 10 Mar 2018, publ: 10 Mar 2018

Conversely, Gajer et al. (2012) applied hierarchical clustering with JSD and SI assessment.

This manuscript describes an algorithm consisting of several consecutive steps, where the most robust set of sub-states (or clusters) is generated for a given longitudinal microbiome dataset. Briefly, the variable factors that are combined to generate the robustness of our algorithm include: five different distance measures (JSD, rJSD, Bray-Curtis, Morisita-Horn and Kulczynski), two clustering scored (SI and Prediction Strength (PS) scores) followed by an additional bootstrapping process (evaluated with the Jaccard similarity score), and two distinct clustering approaches (PAM and Hclust).

The usability of the proposed new algorithm is verified through the re-analysis of four previously-published longitudinal microbiome datasets, where we contribute new insights into the dataset structure, and show that sub-state distribution could be further reused by other kinds of downstream associative analyses of the same datasets.

The contribution of this work, therefore, is to provide an objective and robust mechanism for identifying, and tracking, distinct sub-states and sub-state changes within an individual microbiome time series. This will enable downstream analyses about what, how and why transitions between sub-states happen, and this could in-turn address the challenge of applying microbiota dynamics to important objectives such as personalized medicine (Gilbert et al., 2016), sustainable agriculture or industrial production (Valseth et al., 2017).

MATERIALS AND METHODS

Algorithm: Automation of the robust sub-states definition

This section describes a brief overview of our automatic procedure to define microbiome sub-states. Our robust clustering methodology takes a normalized OTU matrix as input: a) a phyloseq object (McMurdie and Holmes, 2013) or b) a BIOM format file (McDonald et al., 2012), together with their corresponding metadata that identifies each sample and taxa.

Our clustering procedure is based on: 1) Koren et al. (2013)'s study 2) bootstrapping in clustering, and 3) similarity measures from (Barwell et al., 2015). Koren et al. (2013) recommends, for the definition of enterotypes, testing multiple approaches and comparing results. Following these suggestions, our algorithm first finds clusters using two different methods (PAM and Hclust), with five different distance metrics (JSD, root-JSD, Bray-Curtis, Morisita-Horn and Kulczynski) each, and with nine different seed cluster numbers, k (in the range from 2 to 10).

From the output of this first step, we begin to identify the most robust results through a novel assessment approach that utilizes the following criteria:

1. Choosing the k number of clusters (from 2 to 10) with the highest average Silhouette width (SI) among all combinations of pairs of beta diversity measures, with the score being above the SI threshold (0.25), and

2. Checking if that k value also passes the Prediction Strength (PS) threshold (0.80) for robustness, or

3. Confirming if those k clusters are stable according to the Jaccard threshold (0.75) from a bootstrapping process

Those criteria are applied to the output of the 90 combinations (2 algorithms x 5 distances x 9 k values) per dataset, in order to discard those that are not robust and/or not reproducible. Next sections explains the steps in detail, and the relevant factors within our automatic procedure. In particular, we adapt the Koren et al. (2013) approach to the distinct problem of defining microbiome states that exhibit short-term transitions within the same individual, rather than the long-term stereotypes common to enterotype studies, with sparse transitions and where each individual has just one associated state.

The final output of our algorithm is a phyloseq object with a new variable defining the cluster identifier into which each sample has been grouped, a file with <sampleID, clusterID> pairs, and the robust clustering assessment graphs (described below).

Clustering approaches

We selected these two algorithms as representatives of the two most common clustering approaches: a) PAM (Kaufman and Rousseeuw, 1990), as a partitioning approach, and b) Hclust (Kaufman and Rousseeuw, 1990), as an agglomerative hierarchical approach.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26657v1 | CC BY 4.0 Open Access | rec: 10 Mar 2018, publ: 10 Mar 2018

We selected PAM (Partitioning Around Medoids) (Kaufman and Rousseeuw, 1990) because it is an improved approach to the well-known non-deterministic *k*-medoids (Kaufman and Rousseeuw, 1987). This improvement is achieved in two ways. First, PAM selects *k* medoids among the input instances in a greedy phase, rather than the random selection of the *k*-medoids approach (Reynolds et al., 2006). The greedy phase takes each new medoid to minimize the objective function, that is, the sum of distances between each instance and its medoid. Therefore, PAM is a deterministic algorithm and does not need to be run multiple times, as is the case for *k*-medoids. Second, it spreads out the rest of the instances among the defined medoids according to the minimum distance-to-medoid criterion, using the values defined in the distance matrix. PAM then selects each possible pair of instances <medoid, not-medoid>, and evaluates whether the swapping between different clusters results in a smaller value for the objective function. This final step is repeated until the set of medoids do not change.

Hclust, a hierarchical clustering algorithm, adopts a *bottom-up* approach (Kaufman and Rousseeuw, 1990). Hclust begins with an independent cluster per instance. The two nearest clusters are then grouped, in an iterative way, until it arrives at a single cluster, which is therefore the root of an inverted tree structure. We chose *average* (i.e. UPGMA) distance between cluster members as the linkage criteria used to compute the distance between two clusters, rather than single or complete linkage which takes only one cluster element into account when computing the distance (Kaufman and Rousseeuw, 1990).

Beta diversity metrics or distance measures

We took Koren et al. (2013)'s study as reference, because they evaluated beta diversity metrics' influence on the clustering results of microbiome samples, which is similar to our goal here. They recommended comparing at least 2 or 3 different distance measures in the clustering process; we chose to compare 5 distinct measures. Those were chosen from a large list available in the R vegan package (version 2.3-1) (Oksanen et al., 2015), based on suggestions from the work of Koren et al. (2013) and a comprehensive study ranking all available beta diversity measures with abundance data (Barwell et al., 2015), which compared multiple quantitative and qualitative properties.

First, we selected well-extended Jensen-Shanon Distance (**JSD**), **rootJSD** and **Bray-Curtis**, due to they are used in our reference study (Koren et al., 2013). Then, in addition, because our method is independent of the availability of a phylogenetic tree associated to the OTU count matrix, we choose 2 additional metrics from Barwell et al. (2015) to replace the phylogenetic tree-dependent Unifrac metric (weighted and unweighted) used by (Koren et al., 2013). Although there was a precedent study with 23 presence-absence beta diversity metrics (Koleff et al., 2003), we decided to focus on the richer metrics with continuous species abundance Barwell et al. (2015). In general, abundance metrics are less biased than presence-absence ones when under-sampling. (Barwell et al., 2015) compares 29 beta diversity measures with 23 assorted properties. We chose the **Morisita-Horn** and **Kulczynski** metrics from amongst the almost thirty analyzed metrics, taking into account the overall ranking and some specific individual properties that are more important for our use of beta diversity metrics as distance measures in clustering. Morisita is the highest scored beta diversity measure according to the comprehensive set of properties analyzed in Barwell et al. (2015); although we must select the Morisita-Horn implementation (the third best scored) since we work with normalized relative abundances. In addition, both Morisita and Horn-Morisita have been described as being "able to handle different sample sizes"(Wolda, 1981), which is an important characteristic in our studies, where the re-used longitudinal microbiome datasets vary dramatically in size. Kulczynski, meanwhile, is the next-best ranked metric among those available in the R vegan package, ranking sixth out of 29 metrics. Kulczynski is characterized as Pareto-dominant, and "found to have a robust linear (proportional) relationship until ecological distances became large"(Faith et al., 1987).

Clustering assessment scores

Koren et al. (2013) recommend using at least two assessment clustering scores. The three different, complementary clustering scores we selected, and how they are included in our automatic procedure, are as follows:

1. **SI: average Silhouette width**: First, we search for the *k* number of clusters with the best SI, with *k* limited to the range of 2 to 10. This first step was also taken in a previous study, computing microbiome states in a particular dataset (Gajer et al., 2012). Here, we compute the average of all possible combinations of SI values for two different distance measures and each *k*, selecting that

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26657v1 | CC BY 4.0 Open Access | rec: 10 Mar 2018, publ: 10 Mar 2018

one with the highest average. The average SI must be greater than 0.25 in all selected measures, because it is the minimum threshold for 'sensible' clusters, according to (Rousseeuw, 1987). This score takes into account the similarity between the samples in the same and in the nearest clusters. If the selected pair of distance measures does not outperform the following robustness constraint, the next best combination is checked.

2. **PS: Prediction Strength**: Although Koren et al. (2013) indicate that clustering selection could be restricted just to SI score in small datasets, we include this alternative, PS, which is also used in Koren et al. (2013). Our method runs 100 repetitions where the dataset is split into 2 halves and clustering is applied on both; then we search for a correspondence between both group of clusters, classifying points in half A to cluster in half B and vice-versa. Each pair is considered as well classified if both points are classified to the same cluster in the other half. The score is the frequency of correct classification pairs.

3. **Jaccard similarity**: Though the computation of PS implies some kind of bootstrapping, our methodology allows an alternative, explicit bootstrapping step, to verify the stability of the clusters selected with the previous scores. This bootstrapping consists of a resampling with replacement, where clustering is computed over the whole dataset and, in addition, again 100 resamples. Since the Jaccard score compares groups of elements, we computed the similarity of the original cluster with each resampling cluster. Thus, the resulting similarity score is the mean of the size of the intersection divided by the size of the union of samples.

In summary, a total of 18,090 different clustering processes are executed to decide the best microbiome sub-states in a specific dataset. 18.090 comes from 9 ($k = 2 : 10$) potential number of clusters x 5 distance measures (JSD, rJSD, Bray-Curtis, Morisita-Horn and Kulczynski) x 201 assessment scores (1 SI + 100 PS + 100 Jaccard) x 2 clustering algorithms (PAM and Hclust).

Clustering and the distances between OTU vectors were computed with the implementation of different R packages, including the *distance* function in the *phyloseq* R package (v1.19.1) (McMurdie and Holmes, 2013), the *pam* function in the *cluster* R package (v2.0.6), the *hclust* function in the *stats* R package (v3.4.3). Those algorithms take, as input, a distance matrix, where we use the metagenomics beta diversity measures comparing samples, rather than a $\{samples\ x\ features\}$ matrix as required by other algorithms. The first clustering assessment was computed with the *silhouette* function in the *cluster* R package; the robustness evaluation is computed with the *prediction.strength* function (Tibshirani and Walther, 2005) and the corresponding bootstrapping scored by Jaccard similarity with the *clusterboot* function (Hennig, 2007, 2008), both in *fpc* R package (v2.1-11).

RESULTS

This section describes the application in identifying robust clusters in previously published longitudinal microbiome datasets. For each reused dataset, we first show its microbiome sub-state definition and evaluation, followed by a brief interpretation of the clusters (if feasible).

Human gut microbiome sub-states

The dataset from David et al. (2014) is, to our knowledge, the longest and most frequently sampled longitudinal study of the human gut microbiome in healthy subjects. Briefly, it consists of near-daily stool sampling of two distinct subjects, throughout an entire year, including 493 gut samples with 4746 taxa. The input OTU table with absolute abundances was kindly shared by the authors in a personal communication. Dataset details and availability are provided in the Data Citation section.

This and the subsequent sub-sections show the three complementary clustering assessment steps defined in section *Algorithm: Automation of the robust sub-state definition* (corresponding to the three columns in Fig 1), with two different clustering approaches, and five distinct diversity metrics (corresponding to the colored lines).

Figure 1 contains the following analyses:

1. The first column shows the results of the algorithms attempt to choose the most suitable number of clusters, k, according to distinct beta diversity measures (i.e. distance among samples) scored according to SI. The selected k value (from 2 to 10) must report the highest average SI in the best

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26657v1 | CC BY 4.0 Open Access | rec: 10 Mar 2018, publ: 10 Mar 2018

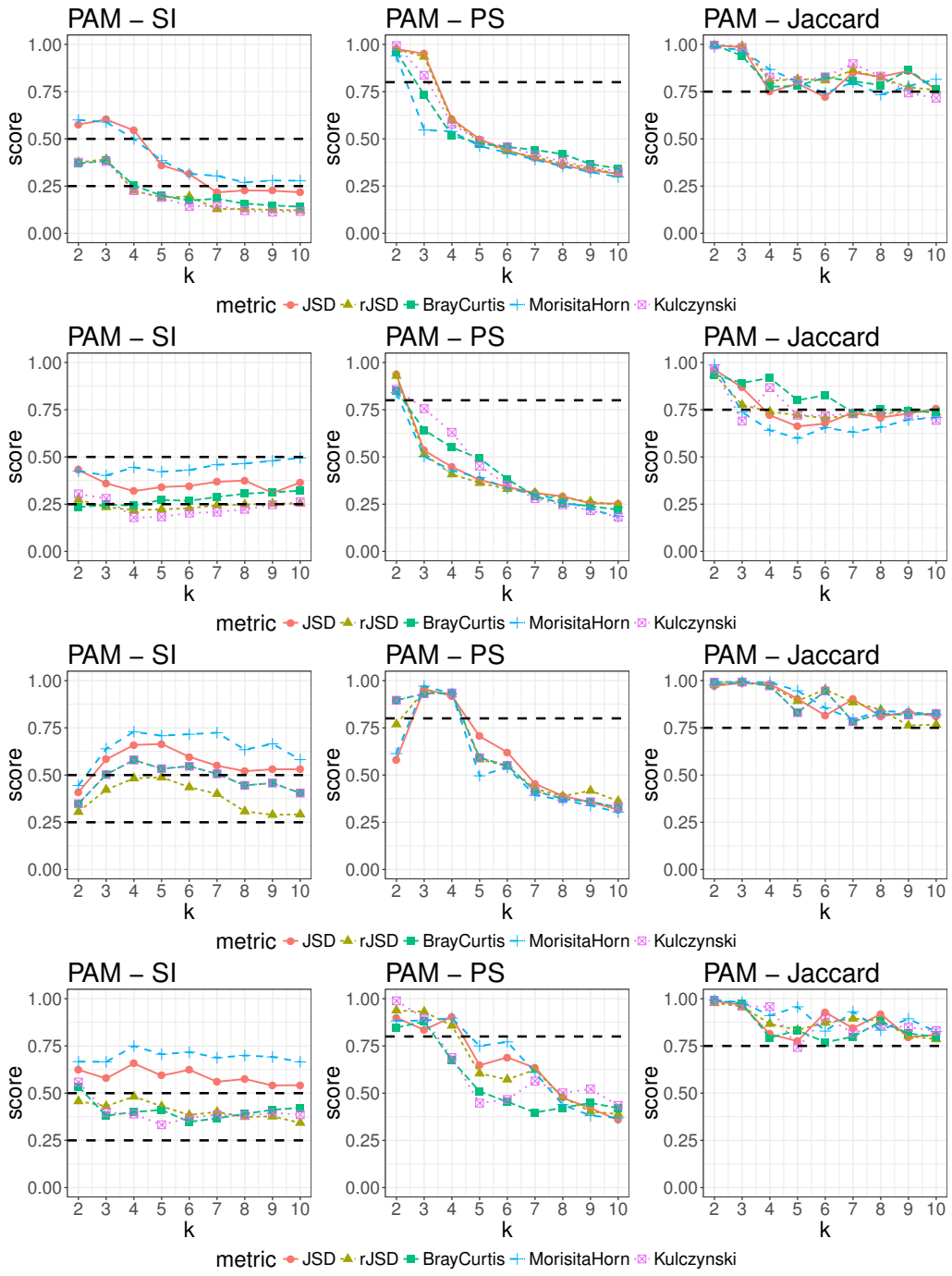

Figure 1. Robust clustering evaluation, with PAM algorithm, in different datasets. From top to bottom: (1) Human gut microbiome (David et al., 2014), (2) Chick gut (Ballou et al., 2016), (3) Vagina (Gajer et al., 2012), (4) Preterm infant gut (La Rosa et al., 2014).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26657v1 | CC BY 4.0 Open Access | rec: 10 Mar 2018, publ: 10 Mar 2018

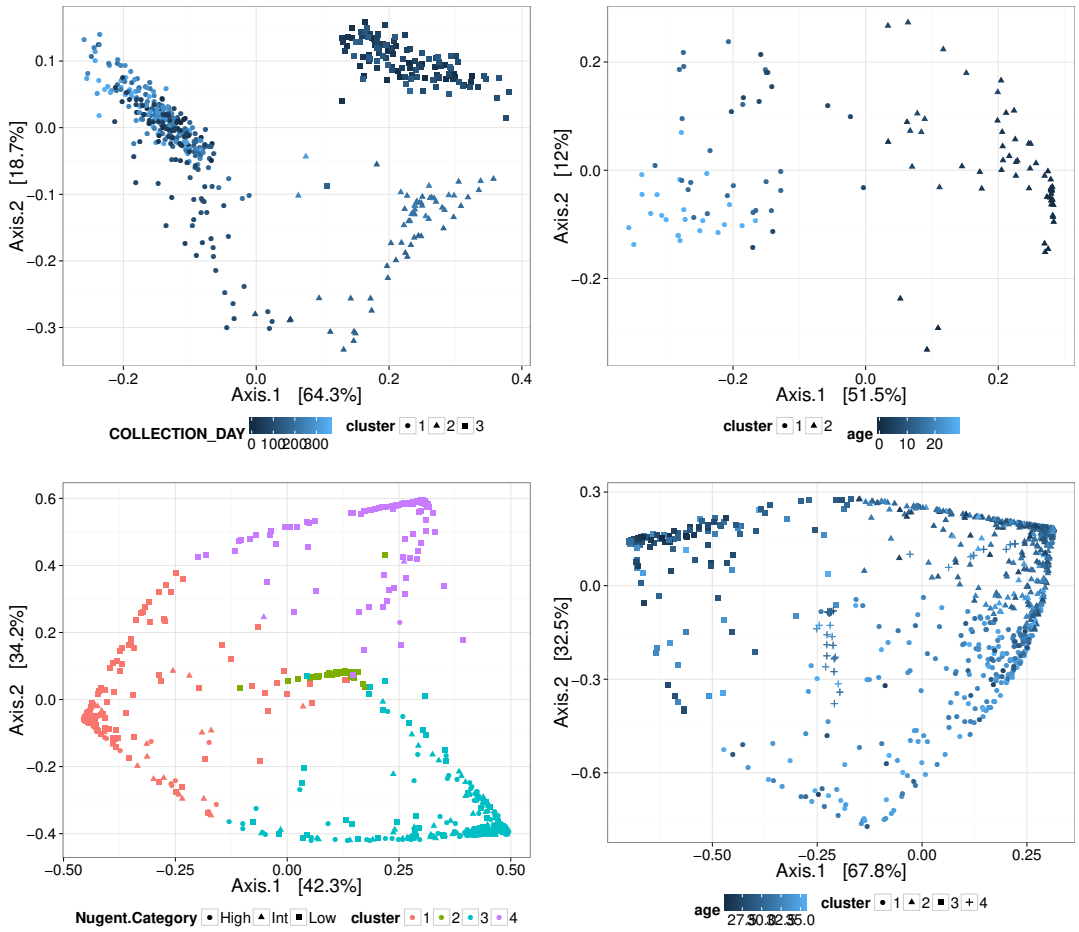

Figure 2. Clusters represented as Principal Coordinates graphs. A (up left): Human Gut. B (up right): Chick Gut. C (bottom left): Vagina. D (bottom right): Preterm Infant Gut.

pair of two beta diversity measures. In Fig 1.1, first column, the selected number of clusters is *k=3*. With 3 clusters, SI takes its highest value (0.602) when utilizing the PAM algorithm (top row), and the JSD metric; the remainder metrics are also higher than the minimum threshold, with Morisita-Horn being the second best metric, scoring above the threshold for strong clusters.

2. The second column allows us to check if the *k* value chosen by SI in the former step is sufficiently robust, by testing it against the PS criterion of being greater than 0.8. The second column of Fig 1.1 shows that JSD and rJSD in PAM with *k=3* satisfies the robustness test with PS=0.950 and PS=0.935. At *k=4*, however, the PS value decreases to the point where the threshold is not passed using any other metric, despite being acceptable based on the SI.

3. Finally, the third column reinforces the stability of the selected *k* clusters, by testing if the Jaccard similarity of the chosen diversity measures exceed 0.75. The last column of Fig 1.1 verifies the stability criterion, with Jaccard=0.986 for the selected *k*=3 clusters for JSD and, in this case, also for all remaining beta diversity metrics.

Cluster interpretation: Fig 2 shows, as Principal Coordinates (PCoA) graphs, the final set of clusters selected by our algorithm for each dataset. Fig 2.A shows the 3 sub-states (clusters) that could be associated to an annotated biological phenomenon. In this case, the associations are with: {*subject A*, *subject B* before dysbiosis, *subject B* after dysbiosis}. No further interpretation was possible using this complicated dataset because insufficient additional metadata was available to test for association.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26657v1 | CC BY 4.0 Open Access | rec: 10 Mar 2018, publ: 10 Mar 2018

Chick gut microbiome sub-states

The second dataset was generated by Ballou et al. (2016). It analyses the response of the developing chick gut microbiome to different treatments (salmonella vaccine and/or probiotics) during their first month of life. The dataset consists of 119 samples with 1583 taxa. The samples include six time points, with 4 or 5 subjects per each of the four treatment combinations. Our goal, therefore, was to define the natural groupings of chick gut microflora. Dataset details and availability are provided in the Data Citation section.

Clustering assessment: Fig 1.2 shows that two clusters are identified by our method; *k=2* presents the highest SI value (0.431) in PAM with the JSD metric (with additional metrics above the 0.25 limit) (first column), and also passes the robustness threshold of PS with 0.935 (second column) and the stability threshold of Jaccard similarity with 0.964 (third column).

Cluster interpretation: Fig 2.B shows that the 2 clusters largely correspond to *immature/young* chicks and *mature/adult* chicks, reinforcing the conclusion of the original manuscript, which showed chick age to be the primary differential factor among samples.

Vaginal microbiome sub-states

The Gajer et al. (2012) dataset consists of 937 samples and 330 OTUs, corresponding to 32 women, with samples collected twice per week for 16 weeks. Dataset details and availability are provided in the Data Citation section. In this case, the original data counts are already pre-processed, and normalized to a sum of 100 per sample, as relative abundances. This contrasts with the previous datasets, where a normalization procedure was applied by us before entering the data into our algorithm.

Clustering assessment: In the Gajer et al. (2012) dataset, our methodology determines that the strongest sample groups appear with 4 clusters, with the highest SI value (0.730) resulting from the clustering with PAM and Morisita-Horn metric. They similarly fulfil the robustness and stability criterion, with PS=0.931 and Jaccard=0.990, respectively, as Fig 1.3 shows.

Cluster interpretation: Our method defines 4 clusters (see Fig 2.c) rather than the 5 clusters identified in the original manuscript (Gajer et al., 2012). Their cluster labeled as cluster IV-A disappears, being distributed among the other bigger clusters; approximately half of samples go to cluster I (our cluster no.4), some go to IV-B (our cluster no.3), and a few to cluster III (our cluster no.1). We note that, using our algorithm, the Prediction Strength decreases dramatically when passing from *k=4* to *k=5* clusters. Our cluster no.3 is that associated with bacterial vaginosis, and overlaps significantly with the disease cluster IV-B in Gajer et al. (2012)). The remaining clusters correspond to a (nominally) healthy vaginal microbiome.

Preterm infant gut microbiome sub-states

The dataset from La Rosa et al. (2014) has 922 samples and 29 OTUs at the class level. That collection includes data from 58 preterm babies, along different time points, for their first month and a half of life. Dataset details and availability are provided in the Data Citation section.

Clustering assessment: Without any previous attempt to distribute samples into groups for this dataset (La Rosa et al., 2014), our robust clustering methodology determined that there were optimally 4 clusters that could be robustly partitioned (see Fig 1.4). The highest SI value (0.749) is achieved with PAM algorithm and Morisita-Horn metric for *k=4*, with all remaining metrics higher than the SI threshold. The robustness and stability of these 4 clusters are also verified by outperforming the Prediction Strength and Jaccard similarity limits (0.897$>$0.8 and 0.911$>$0.75, respectively).

Cluster interpretation: When we analyse the microbial composition of the 4 clusters shown in Fig 2.D, we found that cluster no.2 contains $\sim$50% of the samples, cluster no.1 contains $\sim$25%, cluster no.3 $\sim$20% and cluster no.4 has $<$5% of the samples. Cluster no.3 appears to include the set of youngest babies (see the darkest squares in Fig 2.D); cluster no.1 the oldest babies (see the clearest circles); and cluster no.2 those being those of intermediate age. Analyzing the microbial composition of samples in each cluster, we found cluster no.1 and 3 are mainly enriched in Firmicutes, cluster 2 in Proteobacteria and cluster 4 in Bacteroidetes. This group distribution is in agreement with the La Rosa et al. (2014) results, where they suggest a beginning-of-life with primarily Bacilli (phylum: Firmicutes, as in our cluster no.3 with the youngest babies); subsequently, there would have a Gammaproteobacteria prevalence (phylum: Proteobacteria, as in our cluster no.2); and finally, the infants would have Clostridia as the dominant species (phylum: Firmicutes, corresponding to our cluster no.1). We could not determine a biological association for the small cluster 4, based on the metadata captured by (La Rosa et al., 2014).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26657v1 | CC BY 4.0 Open Access | rec: 10 Mar 2018, publ: 10 Mar 2018

Taxa subset

In this section, we analyze what happens when we take distinct subsets of taxa and apply our approch to find robust clusters in the different datasets. First, we select only the dominant taxa, that is a percentage (exactly we took the 1%) of the most frequent taxa; and also the complementary subset with the non-dominant taxa (number of taxa available in table 1). Second, we aggregate taxa at a higher resolution, i.e. at genus taxonomic level, when possible. Then, we again build clusters with all, dominant and non-dominant taxa at this genus level to compare with the previous results at species level, or the most detailed available one (see table 2).

Table 1 shows a summary with the results comparing clustering with different taxa subsets. The first row corresponds to the default results reported by our method (all taxa) and already explained in previous sections. The other two rows in table 1 point out that the final number of sub-states trends towards only plus/minus one variation when the clustering takes a subset of taxa. Mostly, clusters with dominant taxa seems better according to clustering assessment measures. Moreover, it is relevant to note these best evaluated clusters are found with a maximum of 13 taxa. Some studies have found that enterotypes distinctions were primarily the result of different proportions of a small number of dominant species (Gorvitovskaia et al., 2016), probably similar to our results with dominant taxa. Sometimes, the additional appeared clusters could be outliers more than a new sub-state. For example, in Ballou et al. (2016) with dominant taxa, in the third new cluster only there are 6 samples, from unrelated young chicks. Clusters without those dominant taxa are clearly worse in assessment measures; even not finding robust clusters in one case (Gajer et al. (2012)). However, some studies have reported relevant interactions for microbiome dynamics among rare taxa (Claussen et al., 2017). So, these clusters could be useful in some particular cases.

Table 1. Robust clustering results with all, dominant and non-dominant taxa. Clustering quality scores (SI>0.25 and (PS>0.8 or Jaccard>0.75)). Not clusters found in Gajer2012 with non-dominant taxa, due to not 2 distance measures with SI>threshold.

no.clusters SI\PS\Jaccard no.taxa	David2014	Ballou2016	Gajer2012	LaRosa201
All	3 0.60/0.95/0.99 4746	2 0.43/0.94/0.96 1583	4 0.73/0.93/0.99 298	4 0.75/0.90/0.91 29
Dominant	2 0.70/0.99/0.99 13	3 0.83/0.71/0.81 6	5 0.79/0.48/0.88 13	3 0.91/0.98/0.99 3
Non-dominant	3 0.60/0.97/0.99 4733	6 0.33/0.33/0.76 1577	- - 285	4 0.87/0.64/0.82 26

Table 2 summarizes the clustering results after aggregating taxa at genus level. Our procedure only found clusters in two datasets, because of Gajer et al. (2012) dataset was not provided with taxonomic information in the original publication, and La Rosa et al. (2014) dataset does not allow the taxa simplification at genus level because the samples have taxonomy associated at a higher level (i.e. class level). SI values are lower than before aggregation at genus level reporting weaker clusters, while PS and Jaccard scores are preserved at high values, meaning high robustness in the new found sub-states. In many studies, the OTU tables are aggregated at genus level, therefore this alternative way to find sub-states could be useful in some specific scenarios.

Figure 3 shows the distribution of samples in clusters with the six considered alternative subsets of taxa in the gut microbiome dataset (see supplementary figure S2 for corresponding results in the chick gut dataset). In the bottom row (genus aggregation), in the first and second columns (all and dominant taxa) the clusters are difficult to distinguish. The differences in the general shape of the scatterplots among different cases is also due to the distance measure, being JSD in all the cases where there are 3 clusters. Apparently, it is difficult to say whether one of these three clustering configurations in 3 sub-states is clearly better than others. Hence, more knowledge or meta-data about the specific domain

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26657v1 | CC BY 4.0 Open Access | rec: 10 Mar 2018, publ: 10 Mar 2018

Table 2. Robust clustering results after genus aggregation with all, dominant and non-dominant taxa. Clustering quality scores (SI>0.25 and (PS>0.8 or Jaccard>0.75))

no.clusters SI\PS\Jaccard no.taxa	David2014	Ballou2016
All	2 0.58/0.93/0.96 386	2 0.63/0.93/0.98 180
Dominant	2 0.58/0.85/0.96 10	2 0.67/0.96/0.99 9
Non-dominant	3 0.62/0.89/0.99 376	4 0.44/0.56/0.81 171

where sub-states would be applied might be necessary to conclude which clustering configuration is more suitable than others.

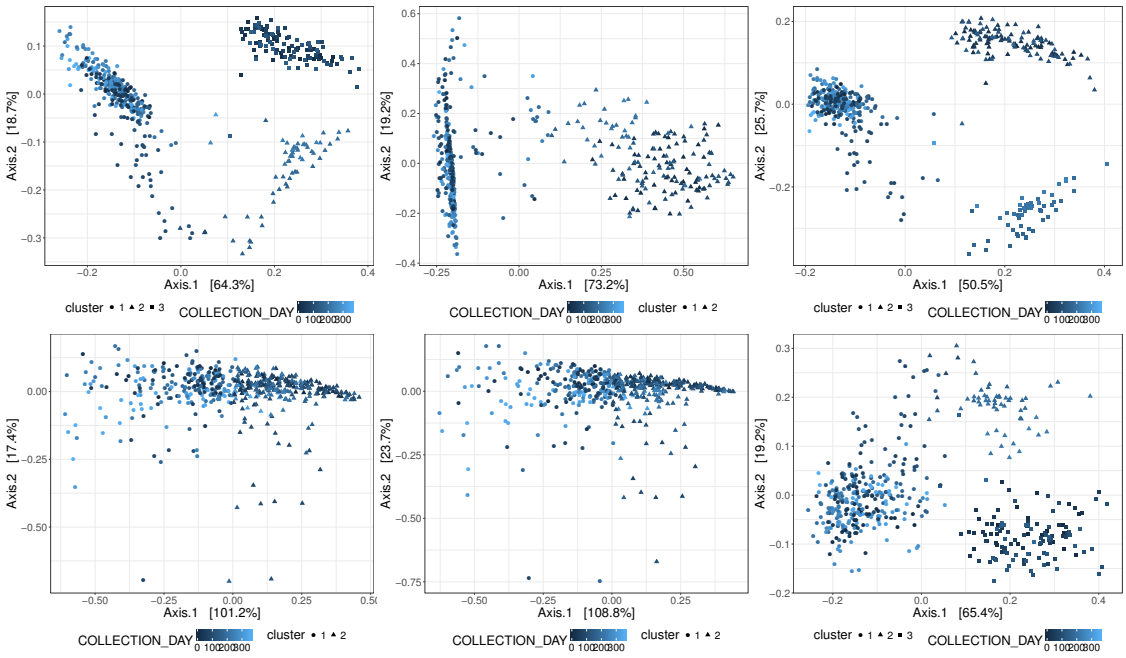

Figure 3. Clusters in Human Gut with different number of taxa, represented as Principal Coordinates graphs. Top row: default taxonomic level (i.e. species), bottom row: genus aggregation. Columns from left to right: all, dominant and non-dominant.

DISCUSSION

Our analysis suggests that, in general, an agglomerative approach (such as PAM) is more suitable than a hierarchical partitioning (such as Hclust) in our microbiome clustering scenario, where we are trying to optimize the number of distinct sub-states to make more precise biological associations, for example, for the purpose of defining biomarkers. As such, we attempt to avoid singletons or very small clusters (with size < 5). In our analyses, Hclust tends to often generate just 2 clusters, according to the limit established by the robustness PS score (see central columns of supplementary figure S1), leading us to suggest that it

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26657v1 | CC BY 4.0 Open Access | rec: 10 Mar 2018, publ: 10 Mar 2018

is not suitable for this task. The final bootstrapping step, measured in Jaccard similarity terms, determines that PAM partitions represent valid and stable clusters, with almost all k values. This contrasts with Hclust, where many k values with different distance measures do not reach the minimum of 0.75 of mean Jaccard similarity. Additionally, PS seems to be the most restrictive score we can apply to discriminate between a viable partition of microbiome samples, and other non-robust possibilities where the scores fall below its threshold of 0.8 (see central columns of Figure 1). Regarding beta diversity measures, JSD and Morisita-Horn are the metrics that usually resulted in the highest SI values, in whatever number of taxa.

We want to contribute to standardization in microbiome, promoted by the International Human Microbiome Standards (IHMS) (Morton et al., 2017) and the Microbiome Quality Control Project (MBQC) (Costea et al., 2017) with a reproducible pipeline to find sub-states in longitudinal microbiome, and ensure comparability. We do not combine data from different studies, to avoid mix different ways of extracting DNA and pre-processing it, but we does combine samples from different subjects, increasing the diversity of the subjects and, consequently, their wider applicability.

Within the four datasets used in this manuscript, our novel method found varying numbers of microbiome sub-states depending on the particular dataset composition: one with 2 clusters (chick gut); another with 3 clusters (human gut); and two datasets with 4 clusters (vaginal microbiome and preterm infant gut). These results provide evidence that our algorithm is applicable to domains with a wide range of diversity and heterogeneity. There are many papers talking about stability and dynamics in microbiome, and many of them are focused in human gut microbiome, while our approach is wider, applicable to whatever microbiome. There is not an objective way to check the correctness of our cluster definition, due to a lack of gold standard datasets with groups definition. However, we find biologically-relevant clusters although our datasets were not designed to be analysed in this way, and are perhaps too small for this kind of analysis.

There are multiple choices when clustering metagenomics data, and someones do not consider them statistically significant enough grouped. However, the same weak statistics arised when clustering different bodysites, and the difference in microbiomes of distinct body sites is scientifically widely accepted (Costea et al., 2018). Moreover, even though the longitudinal microbiome data would evolve more continuously than in a discrete way, a discretization approach could be plausible as a simplification to make viable computational modeling analysis about the influence of external perturbations on microbiome dynamics. In fact, many real processes (including many biological ones) are not discrete in time, although they are simplified as discrete to allow their modeling to be studied with computational and mathematical models (Faust and Raes, 2012; Faust et al., 2015), as we are claiming in our proposal of clustering longitudinal microbiome data in sub-states.

In the result section, we compare the clustering decisions with different sets of taxa applied to real longitudinal metagenomics data. Focusing on less abundant taxa to characterize microbiome clusters is not a common practice, even though some studies highlight interesting events about that: Claussen et al. (2017) found there are interactions between low abundances (i.e. rare) taxa, and Martí et al. (2017) concludes that the most abundant taxa are less volatile than the less abundant ones. Thus, it could imply that giving importance to not-dominant species when sub-states definition, as our approach allows, could evince shifts in microbiome not visible using only-dominant-based sub-states. The default decision in our pipeline is with all the available taxa in the data, although different sets of taxa can be considered ({dominant, without dominant taxa} x {genus or species level}), allowing the users to select the clusters at the level most suitable to their goal.

The concept of a "microbiome biomarker" (Gorvitovskaia et al., 2016) aligns well which our observations of microbiome sub-states being dependent on perturbations. In fact, our clustering pipeline is likely best-suited to identifying sub-states of dysbiosis or otherwise "perturbed" situations, and less suitable for studying non-perturbed steady-state microbiomes expressing more continuous dynamic changes (Gibbons et al., 2017) versus the discrete state-changes resulting from disease, dysbiosis or other perturbations.

CONCLUSIONS

This manuscript describes an automatic algorithm that determines a set of distinct microbiome sub-states in longitudinal data, given an available set of a particular cavity microbiome time series. Our novel methodology is characterized by a robust, objective, transparent and reproducible assessment of the quality of the identified sub-states. The algorithm is flexible with regards to the data source and taxa, and may be applied to a wide range of investigations over diverse species and intervention scenarios.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26657v1 | CC BY 4.0 Open Access | rec: 10 Mar 2018, publ: 10 Mar 2018

In summary, there is not an standard method to define clusters or sub-states in microbiome, even less in longitudinal microbiome datasets. After reviewing and analyzing many studies, with different procedures and distinct results about clustering gut microbiome datasets, Costea et al. (2018) concludes any procedure is not preferred versus others, but it depends on the conditions in each particular case, however all of them have allowed to carry out a particular microbiome analysis. Likely, the best procedure will depend on the posterior application of those clusters. Therefore, given there is not a universal standard, in our defined scenario about clustering longitudinal microbiome datasets of several subjects together, we decided to run a wide variety of those clustering procedures and to select the best one in each case, based on clustering assessment measures, to help users to take the decision, facilitating the use of the resulting sub-states.

As an additional outcome from our analyses, we make available the data files containing the sub-states we defined, to enhance the data provided by the datasets authors, and allow additional posterior analyses, for example, to elucidate the causes of transitions between different microbiome sub-states.

Natural variations and stress factors modify microbiota composition (Weiss and Hennet, 2017), however it is an open problem to discover which element(s) lead a specific shift and how and why that transitions happen. Thus, our new pipeline could help to develop novel methods to predict how to move from one to another of these microbiome sub-states depending on external perturbations.

Code and data availability

Our algorithm, implemented in R, is freely available at GitHub: https://github.com/wilkinsonlab/robust-clustering-metagenomics. Our output data files are available at Zenodo: http://doi.org/10.5281/zenodo.167376

Data Citation

This section describes the sources of the different input datasets.

Human gut microbiome (David et al., 2014):

- Metadata: David, L. et al. *Genome Biology*. Additional file 18: `https://static-content.springer.com/esm/art:10.1186/13059-016-0988-y/MediaObjects/13059_2016_988_MOESM18_ESM.csv` or `http://web.mit.edu/ldavid/www/MF_David_FWD_LAD_2014_06_10.xlsx` (2014)

- OTU table: it was kindly shared by the authors in a personal communication

- Raw data: *EBI/ENA* ERP006059 (2014)

Chick gut microbiome (Ballou et al., 2016):

Qiita 10291 (2016)

Vaginal microbiome (Gajer et al., 2012):

- OTU table and metadata: Gajer P. et al. *Science* Supplementary table S2 (2012)

- Raw data: *SRA* SRA026073 (2012).

Preterm infant gut microbiome (La Rosa et al., 2014):

OTU table and metadata: La Rosa, P.S. et al. *PNAS* Supporting Information, Dataset_S01: `http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1409497111/-/DCSupplemental/pnas.1409497111.sd01.xlsx`

ACKNOWLEDGEMENTS

Thanks to the authors of David et al. (2014) for kindly providing the OTU table privately and to the authors of Ballou et al. (2016) for pleasantly answering our questions about their datasets and metadata. All authors had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

REFERENCES

Arumugam, M., Raes, J., and al., E. (2011a). Enterotyping tutorial. http://enterotype.embl.de.

Arumugam, M., Raes, J., Pelletier, E., et al. (2011b). Enterotypes of the human gut microbiome. Nature, 473(7346):174–80.

Ballou, A. L., Ali, R. A., Mendoza, M. A., Ellis, J. C., Hassan, H. M., Croom, W. J., and Koci, M. D. (2016). Development of the chick microbiome: How early exposure influences future microbial diversity. Frontiers in Veterinary Science, 3(2).

Barwell, L. J., Isaac, N. J. B., and Kunin, W. E. (2015). Measuring β-diversity with species abundance data. Journal of Animal Ecology, 84(4):1112–1122.

Bashan, A., Gibson, T. E., Friedman, J., Carey, V. J., Weiss, S. T., Hohmann, E. L., and Liu, Y.-Y. (2016). Universality of human microbial dynamics. Nature, 534(7606):259–262.

Bradley, P. H. and Pollard, K. S. (2017). Proteobacteria explain significant functional variability in the human gut microbiome. Microbiome, 5(1):36.

Chan, S. H. J., Simons, M. N., and Maranas, C. D. (2017). SteadyCom: Predicting microbial abundances while ensuring community stability. PLOS Computational Biology, 13(5):e1005539.

Claussen, J. C., Skiecevičienė, J., Wang, J., Rausch, P., Karlsen, T. H., Lieb, W., Baines, J. F., Franke, A., and Hütt, M.-T. (2017). Boolean analysis reveals systematic interactions among low-abundance species in the human gut microbiome. PLOS Computational Biology, 13(6):e1005361.

Costea, P. I., Hildebrand, F., Arumugam, M., Bäckhed, F., Blaser, M. J., Bushman, F. D., de Vos, W. M., Ehrlich, S., Fraser, C. M., Hattori, M., Huttenhower, C., Jeffery, I. B., Knights, D., Lewis, J. D., Ley, R. E., Ochman, H., O'Toole, P. W., Quince, C., Relman, D. A., Shanahan, F., Sunagawa, S., Wang, J., Weinstock, G. M., Wu, G. D., Zeller, G., Zhao, L., Raes, J., Knight, R., and Bork, P. (2018). Enterotypes in the landscape of gut microbial community composition. Nature Microbiology, 3(1):8–16.

Costea, P. I., Zeller, G., Sunagawa, S., Pelletier, E., Alberti, A., Levenez, F., Tramontano, M., Driessen, M., Hercog, R., Jung, F.-E., Kultima, J. R., Hayward, M. R., Coelho, L. P., Allen-Vercoe, E., Bertrand, L., Blaut, M., Brown, J. R. M., Carton, T., Cools-Portier, S., Daigneault, M., Derrien, M., Druesne, A., de Vos, W. M., Finlay, B. B., Flint, H. J., Guarner, F., Hattori, M., Heilig, H., Luna, R. A., van Hylckama Vlieg, J., Junick, J., Klymiuk, I., Langella, P., Le Chatelier, E., Mai, V., Manichanh, C., Martin, J. C., Mery, C., Morita, H., O'Toole, P. W., Orvain, C., Patil, K. R., Penders, J., Persson, S., Pons, N., Popova, M., Salonen, A., Saulnier, D., Scott, K. P., Singh, B., Slezak, K., Veiga, P., Versalovic, J., Zhao, L., Zoetendal, E. G., Ehrlich, S. D., Dore, J., and Bork, P. (2017). Towards standards for human fecal sample processing in metagenomic studies. Nature Biotechnology, 35(11):1069–1076.

David, L. A., Materna, A. C., Friedman, J., Campos-Baptista, M. I., Blackburn, M. C., Perrotta, A., Erdman, S. E., and Alm, E. J. (2014). Host lifestyle affects human microbiota on daily timescales. Genome biology, 15(7):R89.

Ding, T. and Schloss, P. D. (2014). Dynamics and associations of microbial community types across the human body. Nature, 509(7500):357–60.

Faith, D. P., Minchin, P. R., and Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance. Vegetatio, 69(1-3):57–68.

Faust, K., Lahti, L., Gonze, D., de Vos, W. M., and Raes, J. (2015). Metagenomics meets time series analysis: unraveling microbial community dynamics. Current Opinion in Microbiology, 25:56–66.

Faust, K. and Raes, J. (2012). Microbial interactions: from networks to models. Nature Reviews Microbiology, 10(8):538–550.

Gajer, P., Brotman, R. M., Bai, G., Sakamoto, J., Schütte, U. M. E., Zhong, X., Koenig, S. S. K., Fu, L., Ma, Z. S., Zhou, X., Abdo, Z., Forney, L. J., and Ravel, J. (2012). Temporal dynamics of the human vaginal microbiota. Science translational medicine, 4(132):132ra52.

Gibbons, S. M., Kearney, S. M., Smillie, C. S., and Alm, E. J. (2017). Two dynamic regimes in the human gut microbiome. PLOS Computational Biology, 13(2):e1005364.

Gilbert, J. A., Quinn, R. A., Debelius, J., Xu, Z. Z., Morton, J., Garg, N., Jansson, J. K., Dorrestein, P. C., and Knight, R. (2016). Microbiome-wide association studies link dynamic microbial consortia to disease. Nature, 535(7610):94–103.

Gorvitovskaia, A., Holmes, S. P., and Huse, S. M. (2016). Interpreting Prevotella and Bacteroides as biomarkers of diet and lifestyle. Microbiome, 4:15.

Hennig, C. (2007). Cluster-wise assessment of cluster stability. Computational Statistics & Data Analysis, 52(1):258–271.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26657v1 | CC BY 4.0 Open Access | rec: 10 Mar 2018, publ: 10 Mar 2018

Hennig, C. (2008). Dissolution point and isolation robustness: Robustness criteria for general cluster analysis methods. Journal of Multivariate Analysis, 99(6):1154–1176.

Kaufman, L. and Rousseeuw, P. (1987). Clustering by means of medoids. In Statistical Data Analysis Based on the L1 Norm and Related Methods, pages 405–416.

Kaufman, L. and Rousseeuw, P. (1990). Finding Groups in Data: An Introduction to Cluster Analysis. Wiley-Interscience.

Koleff, P., Gaston, K. J., and Lennon, J. J. (2003). Measuring beta diversity for presence-absence data. Journal of Animal Ecology, 72(3):367–382.

Koren, O., Knights, D., Gonzalez, A., Waldron, L., Segata, N., Knight, R., Huttenhower, C., Ley, R. E., Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K., Arumugam, M., Raes, J., Pelletier, E., Paslier, D. L., Yamada, T., Caporaso, J., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F., Filippo, C. D., Cavalieri, D., Paola, M. D., Ramazzotti, M., Poullet, J., Dethlefsen, L., Relman, D., McDonald, D., Price, M., Goodrich, J., Nawrocki, E., Desantis, T., Lozupone, C., Knight, R., Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., Tibshirani, R., Walther, G., Rousseeuw, P., Wu, G., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y., Costello, E., Lauber, C., Hamady, M., Fierer, N., Gordon, J., Dominguez-Bello, M., Costello, E., Contreras, M., Magris, M., Hidalgo, G., Koenig, J., Spor, A., Scalfone, N., Fricker, A., Stombaugh, J., Milligan, G., Cooper, M., Claesson, M., Jeffery, I., Conde, S., Power, S., O'Connor, E., Ravel, J., Gajer, P., Abdo, Z., Schneider, G., Koenig, S., Kuczynski, J., Lauber, C., Walters, W., Parfrey, L., Clemente, J., Kumar, P., Brooker, M., Dowd, S., Camerlengo, T., Caporaso, J., Lauber, C., Costello, E., Berg-Lyons, D., Gonzalez, A., Ley, R., Turnbaugh, P., Klein, S., Gordon, J., Turnbaugh, P., Hamady, M., Yatsunenko, T., Cantarel, B., Duncan, A., Costello, E., Gordon, J., Secor, S., Knight, R., Crawford, P., Crowley, J., Sambandam, N., Muegge, B., Costello, E., Jumpertz, R., Le, D., Turnbaugh, P., Trinidad, C., and Bogardus, C. (2013). A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial Community Structures in Human Microbiome Datasets. PLOS Comput Biol, 9(1):59–65.

La Rosa, P. S., Warner, B. B., Zhou, Y., Weinstock, G. M., Sodergren, E., Hall-Moore, C. M., Stevens, H. J., Bennett, W. E., Shaikh, N., Linneman, L. A., Hoffmann, J. A., Hamvas, A., Deych, E., Shands, B. A., Shannon, W. D., and Tarr, P. I. (2014). Patterned progression of bacterial populations in the premature infant gut. Proceedings of the National Academy of Sciences, 111(34):12522–12527.

MacDonald, N. J., Parks, D. H., and Beiko, R. G. (2012). Rapid identification of high-confidence taxonomic assignments for metagenomic data. Nucleic Acids Research, 40(14):e111–e111.

Martí, J. M., Martínez-Martínez, D., Rubio, T., Gracia, C., Peña, M., Latorre, A., Moya, A., and P. Garay, C. (2017). Health and Disease Imprinted in the Time Variability of the Human Microbiome. mSystems, 2(2):e00144–16.

McDonald, D., Clemente, J. C., Kuczynski, J., Rideout, J. R., Stombaugh, J., Wendel, D., Wilke, A., Huse, S., Hufnagle, J., Meyer, F., Knight, R., and Caporaso, J. G. (2012). The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. GigaScience, 1(1):7.

McMurdie, P. J. and Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE, 8(4):e61217.

Morton, J. T., Toran, L., Edlund, A., Metcalf, J. L., Lauber, C., and Knight, R. (2017). Uncovering the Horseshoe Effect in Microbial Analyses. mSystems, 2(1):e00166–16.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., and Wagner, H. (2015). vegan: Community Ecology Package. R package version 2.3-1.

Reynolds, A. P., Richards, G., de la Iglesia, B., and Rayward-Smith, V. J. (2006). Clustering Rules: A Comparison of Partitioning and Hierarchical Clustering Algorithms. Journal of Mathematical Modelling and Algorithms, 5(4):475–504.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20:53–65.

Shankar, J. (2017). Insights into study design and statistical analyses in translational microbiome studies. Annals of Translational Medicine, 5(12):249–249.

Stein, R. R., Bucci, V., Toussaint, N. C., Buffie, C. G., Rätsch, G., Pamer, E. G., Sander, C., and Xavier, J. B. (2013). Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota. PLoS Computational Biology, 9(12):e1003388.

Tibshirani, R. and Walther, G. (2005). Cluster Validation by Prediction Strength. Journal of Computational

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26657v1 | CC BY 4.0 Open Access | rec: 10 Mar 2018, publ: 10 Mar 2018

and Graphical Statistics, 14(3):511–528.

Turroni, S., Rampelli, S., Biagi, E., Consolandi, C., Severgnini, M., Peano, C., Quercia, S., Soverini, M., Carbonero, F. G., Bianconi, G., Rettberg, P., Canganella, F., Brigidi, P., and Candela, M. (2017). Temporal dynamics of the gut microbiota in people sharing a confined environment, a 520-day ground-based space simulation, MARS500. Microbiome, 5(1):39.

Valseth, K., Nesbø, C. L., Easterday, W. R., Turner, W. C., Olsen, J. S., Stenseth, N. C., and Haverkamp, T. H. A. (2017). Temporal dynamics in microbial soil communities at anthrax carcass sites. BMC Microbiology, 17(1):206.

Weiss, G. A. and Hennet, T. (2017). Mechanisms and consequences of intestinal dysbiosis. Cellular and Molecular Life Sciences, 74(16):2959–2977.

Wolda, H. (1981). Similarity indices, sample size and diversity. Oecologia, 50(3):296–302.

Zhou, Y., Mihindukulasuriya, K. A., Gao, H., La Rosa, P. S., Wylie, K. M., Martin, J. C., Kota, K., Shannon, W. D., Mitreva, M., Sodergren, E., and Weinstock, G. M. (2014). Exploration of bacterial community classes in major human habitats. Genome Biology, 15(5):R66.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26657v1 | CC BY 4.0 Open Access | rec: 10 Mar 2018, publ: 10 Mar 2018