Identification of multidimensional Boolean patterns in microbial communities

George Golovko 1,2*, Khanipov Kamil 1,2†, Levent Albayrak 1,2, Anna M. Nia 3, Renato Salomon Arroyo Duarte 4, Sergei Chumakov 4 and Yuriy Fofanov 1,2

Abstract

Background: Identification of complex multidimensional interaction patterns within microbial communities is the key to understand, modulate, and design beneficial microbiomes. Every community has members that fulfill an essential function affecting multiple other community members through secondary metabolism. Since microbial community members are often simultaneously involved in multiple relations, not all interaction patterns for such microorganisms are expected to exhibit a visually uninterrupted pattern. As a result, such relations cannot be detected using traditional correlation, mutual information, principal coordinate analysis, or covariation-based network inference approaches.

Results: We present a novel pattern-specific method to quantify the strength and estimate the statistical significance of two-dimensional co-presence, co-exclusion, and one-way relation patterns between abundance profiles of two organisms as well as extend this approach to allow search and visualize three-, four-, and higher dimensional patterns. The proposed approach has been tested using 2380 microbiome samples from the Human Microbiome Project resulting in body site-specific networks of statistically significant 2D patterns as well as revealed the presence of 3D patterns in the Human Microbiome Project data.

Conclusions: The presented study suggested that search for Boolean patterns in the microbial abundance data needs to be pattern specific. The reported presence of multidimensional patterns (which cannot be reduced to a combination of two-dimensional patterns) suggests that multidimensional (multi-organism) relations may play important roles in the organization of microbial communities, and their detection (and appropriate visualization) may lead to a deeper understanding of the organization and dynamics of microbial communities.

Keywords: Microbiome, Multidimensional Boolean patterns, Microbial communities, Co-exclusion, Co-presence, Pattern-specific score

* Correspondence: gegolovk@utmb.edu
† George Golovko and Khanipov Kamil contributed equally to this work.
1Department of Pharmacology and Toxicology, University of Texas Medical Branch–Galveston, Galveston, TX 77555-0144, USA
2Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch–Galveston, Galveston, TX 77555-0144, USA
Full list of author information is available at the end of the article
Background
Identification of complex multidimensional patterns of abundances/appearances among members of microbial communities (MC) is the key to understand, control, and (in the future) design beneficial microbial communities as well as guide microbial transplantation and personalize antimicrobial and probiotic treatments. Since members of microbial communities can be simultaneously involved in multiple relations that altogether will determine their abundance, not all significant relations between organisms are expected to be manifested as visually uninterrupted patterns and be detected using traditional correlation, mutual information, principal coordinate analysis, or covariation-based approaches. They, however, might be identified and described using Boolean two-, three-, and higher dimensional patterns.

Non-continuous multidimensional patterns
To a certain extent, complex relations between microorganisms within microbial communities (MC) can be recovered by observing their abundances as well as monitoring how they change in response to internal and external perturbations/variables [1]. While initial microbiome characterization studies have been focused on detection of particular organisms under different conditions (healthy vs diseased state), recent studies employ pairwise microbial interaction network analysis to provide a deeper understanding of interactions in MC [2–5]. Traditional methods are generally used to recover pairwise relations between microorganisms in MC which include mutual information-based approaches such as MIC [6], Pearson’s or Spearman’s correlation [7, 8], and covariation. Several computational tools utilizing mutual information, correlation, and covariation techniques have become an essential part of advanced analysis to identify interaction patterns in microbial communities [9–13]. Tools like SparCC [14], developed to infer correlation networks from compositional data, and CoNet [15], which uses an ensemble method to combine information from several different standard comparison metrics, have become widely accepted by the scientific community. While methods based on mutual information (MI), such as MIC [16], are capable of identifying nonlinear and non-continuous pairwise relations [17], they lack the ability to discriminate against intuitively difficult to interpret patterns and can miss some important relationships such as mutual exclusion among microorganisms [18].

Some members of MC can be simultaneously involved in multiple relations which together determine their conditions (healthy vs diseased state), recent studies employ pairwise microbial interaction network analysis to provide a deeper understanding of interactions in MC [2–5]. Traditional methods are generally used to recover pairwise relations between microorganisms in MC which include mutual information-based approaches such as MIC [6], Pearson’s or Spearman’s correlation [7, 8], and covariation. Several computational tools utilizing mutual information, correlation, and covariation techniques have become an essential part of advanced analysis to identify interaction patterns in microbial communities [9–13]. Tools like SparCC [14], developed to infer correlation networks from compositional data, and CoNet [15], which uses an ensemble method to combine information from several different standard comparison metrics, have become widely accepted by the scientific community. While methods based on mutual information (MI), such as MIC [16], are capable of identifying nonlinear and non-continuous pairwise relations [17], they lack the ability to discriminate against intuitively difficult to interpret patterns and can miss some important relationships such as mutual exclusion among microorganisms [18].

Some members of MC can be simultaneously involved in multiple relations which together determine their conditions (healthy vs diseased state), recent studies employ pairwise microbial interaction network analysis to provide a deeper understanding of interactions in MC [2–5]. Traditional methods are generally used to recover pairwise relations between microorganisms in MC which include mutual information-based approaches such as MIC [6], Pearson’s or Spearman’s correlation [7, 8], and covariation. Several computational tools utilizing mutual information, correlation, and covariation techniques have become an essential part of advanced analysis to identify interaction patterns in microbial communities [9–13]. Tools like SparCC [14], developed to infer correlation networks from compositional data, and CoNet [15], which uses an ensemble method to combine information from several different standard comparison metrics, have become widely accepted by the scientific community. While methods based on mutual information (MI), such as MIC [16], are capable of identifying nonlinear and non-continuous pairwise relations [17], they lack the ability to discriminate against intuitively difficult to interpret patterns and can miss some important relationships such as mutual exclusion among microorganisms [18].

Some members of MC can be simultaneously involved in multiple relations which together determine their conditions (healthy vs diseased state), recent studies employ pairwise microbial interaction network analysis to provide a deeper understanding of interactions in MC [2–5]. Traditional methods are generally used to recover pairwise relations between microorganisms in MC which include mutual information-based approaches such as MIC [6], Pearson’s or Spearman’s correlation [7, 8], and covariation. Several computational tools utilizing mutual information, correlation, and covariation techniques have become an essential part of advanced analysis to identify interaction patterns in microbial communities [9–13]. Tools like SparCC [14], developed to infer correlation networks from compositional data, and CoNet [15], which uses an ensemble method to combine information from several different standard comparison metrics, have become widely accepted by the scientific community. While methods based on mutual information (MI), such as MIC [16], are capable of identifying nonlinear and non-continuous pairwise relations [17], they lack the ability to discriminate against intuitively difficult to interpret patterns and can miss some important relationships such as mutual exclusion among microorganisms [18].
abundance. For example, in many environmental microbial communities, functions vital to the whole community (e.g., nitrogen fixation) are often performed by a single species [19, 20], so the abundance pattern between these organisms and other members of MC will not be represented by correlation, but are rather expected to exhibit a Boolean pattern. A Boolean one-way relation pattern is exhibited when the presence of “dependent” microorganism(s) requires the presence of a “provider,” but not vice versa (Fig. 1b). Similarly, other pairwise relations such as co-presence and co-exclusion may be represented as non-continuous Boolean patterns (Fig. 1a, c). For pairwise relations, there are a total of \(2^n\) possible combinations (\(2^{2^n}\), where \(n = \) number of variables/organisms), given that there are four Boolean functions (constant function true, negation function, identity function, and constant function false) for every Boolean variable [21]. However, out of \(2^4\) possible combinations of the presence/absence profiles between two organisms, only four may be interpreted as possible relations: co-presence, co-exclusion, and two one-way relations (organism 1 needs organism 2 to survive and vice versa). It is also important to keep in mind that if the cooperation of several organisms is required to maintain a single metabolic pathway, their abundances will fit into multi-dimensional Boolean patterns, such as multidimensional co-presence (Fig. 1g).

**Complications of using mutual information as a score for non-continuous patterns**

In Boolean patterns, two microorganisms’ abundances can vary without affecting the pattern (e.g., variation in abundance within the same quadrant), and the pattern strength can be defined based on the fraction of observations located in four quadrants of two-dimensional space: \(p_{00}, p_{01}, p_{10}, p_{11}\) (Fig. 1a–c). However, it is important to mention that since different roles played by microorganisms in MC may require different minimal abundances. The appropriate calculation of the \(p_{ij}\) will require identification of the microorganism-specific thresholds, so \(p_{ij}\) becomes a function of four variables: \(p_{ij}(x_1, x_2, \epsilon_1, \epsilon_2)\), where \(x_1\) and \(x_2\) are the abundance profiles of two microorganisms under consideration and \(\epsilon_1\) and \(\epsilon_2\) are the corresponding presence/absence thresholds.

The most obvious choice to define the strength of a Boolean pattern would be by using a mutual information score (MIS):

![Fig. 2 Mutual information score values for different pattern types. MIS values for “ideal” a co-presence, b co-exclusion, and c one-way patterns. Effect of disbalance on d co-presence, e co-exclusion, and f one-way relations on MIS values.](image-url)
MIS\( (X_1, X_2) = \max (MI(X_1, X_2, \varepsilon_1, \varepsilon_2)) \)
\[ = \max_{\varepsilon_1, \varepsilon_2} (H_1(X_1, \varepsilon_1) + H_2(X_2, \varepsilon_2) - H_{12}(X_1, X_2, \varepsilon_1, \varepsilon_2)) , \]
where \( H_1(X_i, \varepsilon_i) \) and \( H_2(X_j, \varepsilon_j) \) correspond to one- and two-dimensional entropies.

The use of mutual information to identify such patterns, however, has several significant disadvantages. The best possible (maximal) MIS value is not the same for different pattern types. For example, while MIS value for “ideal” co-presence and co-exclusion patterns is 0.693 (Fig. 2a, b), the score for “ideal” one-way pattern is only 0.174 (Fig. 2c). Moreover, a small disbalance between fractions of points located in four partitions \( p_{00}, \ldots, p_{11} \) can significantly affect the MIS value. For example, while two co-exclusion patterns may be intuitively obvious (Fig. 2a, d), a disbalance between \( p_{00} \) and \( p_{11} \) can cause a significant drop in the MIS value. A similar observation can be made for co-exclusion (Fig. 2b, e). The use of MIS value can be especially misleading in the case of one-way relation patterns. An MIS value may be extremely low (0.055) for patterns which can be clearly interpreted as one-way relation (Fig. 2f). These observations suggest that the use of MIS to identify non-continuous Boolean patterns may result in missing certain intuitively obvious patterns. The presented work is an attempt to introduce an alternative, pattern-specific approach, to estimate the strength and statistical significance of two- and higher dimensional patterns between members of microbial communities.

Methods
Pattern-specific strength score

The basic idea of the proposed approach is to estimate the pattern score by counting the fraction of observations belonging to the pattern under investigation. Assuming that \( p_{00} + p_{01} + p_{10} + p_{11} = 1 \) and the presence/absence threshold can be microorganism specific, the strength of each pattern can be defined as the following:

\[
S_{\text{co-presence}} = \max_{\varepsilon_1, \varepsilon_2} (p_{00} + p_{11});
\]
\[
S_{\text{co-exclusion}} = \max_{\varepsilon_1, \varepsilon_2} (p_{00} + p_{10} + p_{01});
\]
\[
S_{\text{one-way}} = \max_{\varepsilon_1, \varepsilon_2} (p_{00} + p_{01} + p_{11}); (\text{where } X_2 \text{ depends on } X_1)
\]
\[
S_{\text{one-way}} = \max_{\varepsilon_1, \varepsilon_2} (p_{00} + p_{10} + p_{11}); (\text{where } X_1 \text{ depends on } X_2).
\]

It is important to mention that the presence of \( p_{00} \) is required to distinguish co-presence patterns from cases when both organisms are simply present in all samples. Co-presence patterns require the existence of co-absence between the microorganisms in the sample set. Additionally, co-exclusion and one-way relation pattern scores include \( p_{00} \) because mutual absence does not contradict the pattern.

While presence/absence threshold optimization allows considering that different microorganisms may have various minimal abundance thresholds to interact with the MC, this approach also can produce misleading results. For example, a perfect co-presence score may be achieved by increasing the presence/absence threshold to the point where all the observations will be counted as absent: \( \varepsilon_1 \geq (X_1) \) and \( \varepsilon_2 \geq (X_2) \).

This effect can be minimized by requiring a proportion of experimental observations in quadrants contributing to the pattern under consideration to be above a predefined minimal threshold (\( m \)):

\[
S_{\text{co-presence}} = \max_{\varepsilon_1, \varepsilon_2} (p_{00} + p_{11}), \text{ where } p_{00} > m; p_{11} > m;
\]
\[
S_{\text{co-exclusion}} = \max_{\varepsilon_1, \varepsilon_2} (p_{00} + p_{10} + p_{01}), \text{ where } p_{01} > m;
\]
\[
p_{10} > m;
\]
\[
S_{\text{one-way}} = \max_{\varepsilon_1, \varepsilon_2} (p_{00} + p_{01} + p_{11}), \text{ where } p_{01} > m, p_{11} > m;
\]
\[
(\text{where } X_2 \text{ depends on } X_1);
\]
\[
S_{\text{one-way}} = \max_{\varepsilon_1, \varepsilon_2} (p_{00} + p_{10} + p_{11}), \text{ where } p_{10} > m, p_{11} > m;
\]
\[
(\text{where } X_1 \text{ depends on } X_2).
\]

Non-trivial multidimensional patterns

The proposed approach can be further extended to identify more complex multidimensional patterns. For example, in some 3D patterns, the presence or absence of one organism may define the kind of 2D patterns exhibited between two other organisms. Figure 1f shows a case where organisms 2 and 3 will be co-present if organism 1 is present and co-exclude if this organism is absent:

\[
S_{\text{pattern}} \lambda = \max_{\varepsilon_1, \varepsilon_2, \varepsilon_3} (p_{00} + p_{11} + p_{01} + p_{001}), \text{ where } p_{11} > m; p_{01} > m; p_{001} > m.
\]

Similar to 2D patterns, not all combinations of \( p_{ijk} \) values can be interpreted as possible relations between microorganisms. Some 3D patterns can be the direct result of three pairwise 2D patterns: for example, the pairwise co-exclusion pattern between three organisms will unambiguously lead to a 3D co-exclusion pattern (Fig. 1d):

\[
S_{\text{3D co-exclusion type 1}} = \max_{\varepsilon_1, \varepsilon_2, \varepsilon_3} (p_{00} + p_{10} + p_{01} + p_{001}), \text{ where } p_{10} > m; p_{01} > m; p_{001} > m.
\]

Is a direct consequence of its 2D patterns:

\[
S_{\text{co-exclusion} 1,2} = \max_{\varepsilon_1, \varepsilon_2} (p_{00} + p_{10} + p_{01}), \text{ where } p_{01} > m; p_{10} > m;
\]
\[
S_{\text{co-exclusion} 1,3} = \max_{\varepsilon_1, \varepsilon_2} (p_{00} + p_{10} + p_{01}), \text{ where } p_{10} > m; p_{00} > m;
\]
where \( p_{001} > m; p_{010} > m \).

Three-dimensional \textit{co-exclusion} patterns may additionally be observed in a very different way where each pair of organisms is \textit{co-present} only if the third one is absent (Fig. 1e):

\[
S_{\text{3D co-exclusion type } 2} = \max_{e_1, e_2, e_3} (p_{000} + p_{011} + p_{101} + p_{111}),
\]

where \( p_{110} > m; p_{101} > m; p_{011} > m \).

In fact, every 2D pattern has at least one non-trivial 3D analog which can be interpreted as the relation between organisms and not derived directly from any 2D combination:

\[
S_{\text{3D co-presence}} = \max_{e_1, e_2, e_3} (p_{000} + p_{111}), \ \text{where} \ p_{111} > m;
\]

\[
p_{000} > m;
\]

\[
S_{\text{3D co-presence}} = \max_{e_1, e_2, e_3} (p_{0000} + p_{1111}), \ \text{where} \ p_{1111} > m;
\]

\[
p_{0000} > m;
\]

or for one-way relations:

\[
S_{\text{3D one-way relation}} = \max_{e_1, e_2, e_3} (p_{000} + p_{111} + p_{001} + p_{010}),
\]

where \( p_{111} > m; p_{001} > m; p_{010} > m; \) (organism 1 requires two others to be present).

\[
S_{\text{3D one-way relation}} = \max_{e_1, e_2, e_3} (p_{0000} + p_{1111} + p_{0001} + p_{0100}),
\]

where \( p_{1111} > m; p_{0001} > m; p_{0010} > m; p_{0100} > m; \) (organism 1 requires three others to be present).

Additionally, some high dimensional patterns can reflect interesting new relations which exist only in higher dimensions. Figure 1f presents a case where two microorganisms (\( X_1 \) and \( X_2 \)) follow \textit{co-exclusion} patterns in the presence of the third (\( X_3 \)), as well as \textit{co-presence} in its absence; Fig. 1g shows a case where three organisms can be present only all together or individually:

\[
S_{\text{3D all together or alone}} = \max_{e_1, e_2, e_3} (p_{000} + p_{111} + p_{001} + p_{010} + p_{100}),
\]

where \( p_{111} > m; p_{100} > m; p_{001} > m; p_{010} > m; \) (three organisms present only all together or individually).

### Statistical significance and type 1 error

It is important to keep in mind that an arbitrary choice of the presence threshold (\( m \)) and minimal score (\( S_{\text{min}} \)) above which patterns are considered to be present can significantly affect the results of the analysis in both: a number of detected patterns and their statistical significance (e.g., type 1 error). Lowering these thresholds increases chances for patterns to appear randomly, and this can be detected by comparing the results produced by real data against a randomized (shuffled) dataset. Table 1 provides an example of the number of two-dimensional \textit{one-way} relation patterns identified in original and shuffled (and renormalized) datasets from the Human Microbiome Project (genus level, mid-vagina samples) [22, 23].

The choice of the shuffling method reflects the underlying assumption about what would be considered as the random alternative to the observed dataset (zero model) [14, 24]. For example, the shuffling of the abundance values across the whole dataset reflects the assumption of the total randomness of the appearances of all the values across samples and organisms. While this model preserves the overall distribution of the abundance

| Population threshold/score | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 | 0.10 |
|---------------------------|------|------|------|------|------|------|------|------|------|------|------|
|                           | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  |
|                           | 1/0  | 1/0  | 1/0  | 1/0  | 1/0  | 1/0  | 1/0  | 1/0  | 1/0  | 1/0  | 1/0  |
|                           | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  |
|                           | 2/0  | 2/0  | 2/0  | 2/0  | 2/0  | 2/0  | 2/0  | 2/0  | 2/0  | 2/0  | 2/0  |
|                           | 4/0  | 4/0  | 4/0  | 4/0  | 4/0  | 4/0  | 4/0  | 4/0  | 4/0  | 4/0  | 4/0  |
|                           | 7/0  | 7/0  | 7/0  | 7/0  | 6/0  | 5/0  | 4/0  | 4/0  | 4/0  | 4/0  | 4/0  |
|                           | 1/0  | 1/0  | 1/0  | 1/0  | 1/0  | 1/0  | 1/0  | 1/0  | 1/0  | 1/0  | 1/0  |
|                           | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  |
|                           | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  | 0/0  |

Table 1: The number of two-dimensional \textit{one-way} relation patterns identified in shuffled and real (original) mid-vagina samples. Bold font reflects the score/threshold combinations where no patterns have been observed in simulated (shuffled) data.
values, it does not take into consideration that some organisms may always be present in low abundance and others can become highly dominated species in the community. In order to reflect this property on microbial abundance data, shuffling across individual OTU profiles has been implemented in the presented method and used in all the examples shown in this manuscript.

The shuffling approach has been implemented as part of all pattern-specific computational pipelines (see the “Methods” section) to make sure that the search for the patterns in real data is performed only for the presence ($m$) and minimal score ($S_{\text{min}}$) thresholds for which the number of specific patterns in shuffled data is equal to zero. The presented method, however, allows a variety of modifications including less strict type 1 error requirements. The next versions of the software will include the ability to perform multiple shuffling types as well as the ability to perform shuffling multiple times.

Implementation
The presented method is able to identify three types of 2D patterns (co-presence, co-exclusion, and one-way relations) as well as three types of 3D patterns shown on Fig. 1e–g. The codebase was developed in C++, and the executable files and source code are available on GitHub (https://github.com/kkhanipov/MultidimensionalBooleanPatterns).

In order to improve performance in the proposed implementation, the patterns in shuffled data for all the combinations of presence threshold ($m$) and minimal score ($S_{\text{min}}$) are calculated during the first step of the analysis, so search for patterns in real data can be performed in a limited search space where zero patterns are detected in randomized (shuffled) data.

To evaluate performance, the presented source code was compiled using a GCC compiler version 6.3.1 under Linux CentOS 6.7. Sixteen HMP OTU files were used for the identification of 2D and 3D patterns on 4× AMD Opteron 8 core processors, 512 GB RAM, and 30 TB of storage system. Search for the two-dimensional co-presence, co-exclusion, and one-way relation patterns for all tested samples took between 1 and 3 min, and the memory footprint did not exceed 50 Mb of RAM. However, search for the 3D patterns may take hundreds of hours and requires a higher level of parallelization and a high-performance computing environment.

Data acquisition
The microbial community compositions used for this analysis originated from the NIH Human Microbiome Project [22, 23] and contained 18 datasets associated with 16 body sites. Microbial profiles for 2910 samples have been downloaded from the project website as of December 2016 in text format (HMQCP–QIIME community Profiling v13 OTU table). Samples representing significantly low (less than 2000) and significantly high (over 50,000) number of sequencing reads were excluded from the analysis. The microbial profiles of the remaining 2380 samples, varying from 67 for posterior fornix to 200 for antecubital fossa, have been normalized against the total number of reads in each sample and transformed into relative abundance profiles merged to genus taxonomy level for each body site resulting in 619 profiles. Analysis has been performed for each body site individually. For each body site under consideration, genera present in less than 5% of samples have been excluded from the analysis.

Boolean patterns detected in Human Microbiome Project data
All three types of 2D patterns have been identified in virtually every type of sample of the Human Microbiome Project data (Additional File 1). The largest number of patterns (all types included) has been detected in supragingival plaque, tongue dorsum, stool, and subgingival plaque datasets (Fig. 3a–d and Additional File 2). No apparent correlation has been observed between the number of patterns and the total number of samples nor the number of OTUs in the datasets. It is important to mention that while all of the observed 2D patterns pass statistical significance criteria, the overall size and complexity of the resulting networks depend on the pattern score threshold (Fig. 3e). In the interaction patterns of HMP supragingival (Fig. 3a), the Fusobacteria genus (green) node has 11 one-way relationships with other taxa. Additionally, there are another 12 such significant patterns with lower scores as shown in the supplementary file “Additional File 1.” Fusobacteria nucleatum is a well-known pathogen that was not only found in the subgingival and supragingival plaques [25], but also previously characterized in vitro [26] and in vivo [27] dental plaque biofilms. Interestingly, one group found out that F. nucleatum is unable to grow as a single species and builds a mutualistic relationship with other members of local microbiota such as Aggregatibacter. The presence of this pattern is in agreement with previous findings that orofacial odontogenic infections are usually polymicrobial [28]. While the Fusobacterium-Aggregatibacter pattern has not been identified in our cohort of samples, our findings of one-way relation patterns between Fusobacterium and other members of microbial communities yield rather interesting results. Thirteen such patterns showed interaction between Fusobacteria spp. and other known pathogens. For example, Catonella and Clostridiales spp. have been previously associated as uniquely present in the patients with caries [29]. Another pattern includes a one-way relation with Tannerella spp., which is also known as periodontal pathogen [30], as well as very well-studied Dialister spp., and their role as...
periodontopathic bacteria [31]. Finally, while species of identified Johnsonella genus are not known directly related to dental diseases, they have been linked to chronic obstructive pulmonary disease (COPD) which strengthen proposed method as a tool for hypothesis discovery in microbial communities [32].
While the available HMP data does not possess enough precision power to pinpoint the exact pathogenic strains that are contained within the samples and which happen to be within the pathogenic genera, family, or even order, we believe these observations of interaction between Fusobacteria and other members of microbial communities are not random and may benefit from further analysis and validation. Improvement of high-throughput sequence technologies, decrease of cost, and availability of the high-quality public data will close this data precision gap.

Some 3D patterns have also been observed in buccal mucosa, supragingival plaque, and merged retroauricular crease datasets (see example in Table 2).

### Discussion and conclusion

Identification of interaction patterns in microbial communities is essential to further our understanding of...
relationships in microbial communities. Knowledge of the interactions between specific organisms can help transition microbiomes between enterotypes and better predict microbial responses due to perturbations (e.g., targeted antimicrobials, probiotics, prebiotics). Ability to manipulate microbial communities in terms of community members and their functions will open new opportunities for precision medicine and personalized treatments. Thus, the development of systematic and statistically sound methods for interaction pattern identification is a necessary step to understand the structure of microbiomes and the processes by which they evolve.

Since correlation and MI-based approaches can miss important multidimensional patterns and produce misleadingly low scores for certain intuitively obvious patterns, the proposed method could serve as a useful addition to the set of tools available for the microbiome analysis. It is important to keep in mind, however, that the presence of statistically significant patterns between the abundance of two and more organisms must be interpreted very carefully and treated more like an indication of a potential interaction and requires independent experimental validation. Additionally, datasets from different environments or conditions should not be analyzed simultaneously for patterns, since this may result in false interaction patterns such as co-exclusion due to the datasets being of different nature with different compositions. The comparison of interactions between different conditions should be done between the calculated interaction pattern sets (networks).

The visualization of multidimensional patterns involving multiple organisms, however, remains a significant challenge. Traditionally, the graph (network) representation of the patterns between organisms in microbial communities represents each OTU as the node and pairwise relationships as edges [14, 16, 24, 33]. We believe that one of the possible ways to visualize 2D, 3D, and higher dimensional patterns could be by using a multi-layer network (multi-layer graph) which in contrast with traditional graphs (networks) can simultaneously include nodes of different types [34], such as OTUs and multidimensional patterns. Figure 4 shows an example of such a representation for two- and three-dimensional patterns in attached keratinized gingiva samples from the Human Microbiome Project.

The presented approach can also be extended by including a variety of physical (pH, temperature, oxygen concentration) and biochemical (antimicrobial susceptibility, nutrient, and metabolite concentration) variables into the search for multidimensional patterns. We also believe that it can be extended to the simultaneous analysis of multi-omics data, such as protein and mRNA expression in both microbial communities and the mammalian host.

Supplementary information
Supplementary information accompanies this paper at https://doi.org/10.1186/s40168-020-00853-6.

Acknowledgements
Not applicable

Authors’ contributions
YF and LA designed the model and the computational framework. YF, GG, and KK carried out the implementation. GG, KK, AN, and DR performed the calculations. YF, GG, KK, and AN wrote the manuscript with input from all authors. GG, KK, SC, and YF conceived the study and were in charge of the overall direction and planning. All authors read and approved the final manuscript.

Funding
This work was partially supported by the National Institute of Health (award number R03DE028596-01) and Texas Commission on Environmental Quality (award number M1901103).

Availability of data and materials
The datasets analyzed during the current study are available in the Human Microbiome Project repository (https://portal.hmpdacc.org/) [35]. All data generated or analyzed during this study are included in this published article and its supplementary information files. Executable files and source code are available on GitHub (https://github.com/kkhanipov/MultidimensionalBooleanPatterns).

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Author details
1Department of Pharmacology and Toxicology, University of Texas Medical Branch–Galveston, Galveston, TX 77555-0144, USA. 2Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch–Galveston, Galveston, TX 77555-0144, USA. 3Department of Molecular Biophysics, University of Texas Medical Branch–Galveston, Galveston, TX 77555-0144, USA. 4Department of Physics, University of Guadalajara, Revolucion, 1500 Guadalajara, Jalisco, Mexico.

Received: 22 August 2019 Accepted: 4 May 2020
Published online: 11 September 2020

References
1. Paily O, Shankar V. Application of multivariate statistical techniques in microbial ecology. Mol Ecol. 2016;25(5):1032–57.
2. Cardinale M, Grube M, Erlacher A, Quehenberger J, Berg G. Bacterial networks and co-occurrence relationships in the lettuce root microbiota. Environ Microbiol. 2015;17(1):239–52.
3. Barberan A, Bates ST, Casamayor EO, Fierer N. Using network analysis to explore co-occurrence patterns in soil microbial communities. Isme J. 2012; 6(2):343–51.
4. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT, et al. Human genetics shape the gut microbiome. Cell. 2014;159(4):789–99.
5. Belström D, Constancias F, Liu Y, Yang L, Drautz-Moses DI, Schuster SC, Kohli GS, Jakobsen TH, Holmstrup P, Groskov M. Metagenomic and metatranscriptomic analysis of saliva reveals disease-associated microbiota...
in patients with periodontitis and dental caries. NPJ Biofilms Microbiomes. 2017;3:23.
6. Cover TM, Thomas JA. Elements of information theory. New York: Wiley; 1991.
7. Pearson K. Note on elements of information theory. J. Biometrika. 1957;44(3/4):470–81.
8. Schöler A, Jacquier S, Vestergaard G, Schulz S, Schlote M. Analysis of soil microbial communities based on amplicon sequencing of marker genes. PLoS One. 2018;9(1):38–54.
9. Baud D, Pattaroni C, Vulliemoz N, Castella V, Marsland BJ, Stojanov M. Sperm microbiota and its impact on semen parameters. Front Microbiol. 2019;10.
10. Rothman JA, Andrikopoulos C. Floral and foliar source affect the bee nest microbial community. Microb Ecol: McFrederick QS; 2018.
11. Mandakovic D, Rojas C, Maldonado J, Latorre M, Travisany D, Delage E, Bihouee A, Jean G, Diaz FP, Fernandez-Gomez B, et al. Structure and co-occurrence patterns in microbial communities under acute environmental stress reveal ecological factors fostering resilience. Sci Rep. 2018;8(1):5875.
12. Faust K, Sathirapongsasuti JB, Judd SD, Sathirapongsasuti JB, Judd SD, et al. Microbial co-occurrence relationships in the human microbiome. PLoS One. 2012;7(7):e45124.
13. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, Lozupone C, Follows MJ, Valois FW, Waterbury JB. Iron conservation by reduction of metalloenzyme inventories in the marine diazotroph Crocosphaera watsonii. Proceedings of the National Academy of Sciences. 2011;108(7):2184–9.
14. Alaybay L, Khanipov K, Golovko G, Fofanov Y. Detection of multidimensional co-exclusion patterns in microbial communities. Bioinformatics. 2018;34(21):3695–701.
15. consortium hmp. Microbiome. 2017;5(1):27.
16. consortium hmp. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):215–21.
17. consortium hmp. Microbiome. 2017;5(1):549–56.
18. consortium hmp. Microbiome. 2018:430462.
19. consortium hmp. Microbiome. 2017;5(1):23.
20. consortium hmp. Microbiome. 2012;486(7402):207–14.
21. consortium hmp. Microbiome. 2012;486(7402):196–205.
22. consortium hmp. Microbiome. 2021:205136–70.
23. consortium hmp. Microbiome. 2021:205136–70.
24. consortium hmp. Microbiome. 2021:205136–70.
25. consortium hmp. Microbiome. 2021:205136–70.
26. consortium hmp. Microbiome. 2021:205136–70.
27. consortium hmp. Microbiome. 2021:205136–70.
28. consortium hmp. Microbiome. 2021:205136–70.
29. consortium hmp. Microbiome. 2021:205136–70.
30. consortium hmp. Microbiome. 2021:205136–70.
31. consortium hmp. Microbiome. 2021:205136–70.
32. consortium hmp. Microbiome. 2021:205136–70.
33. consortium hmp. Microbiome. 2021:205136–70.
34. consortium hmp. Microbiome. 2021:205136–70.
35. consortium hmp. Microbiome. 2021:205136–70.
36. consortium hmp. Microbiome. 2021:205136–70.
37. consortium hmp. Microbiome. 2021:205136–70.
38. consortium hmp. Microbiome. 2021:205136–70.
39. consortium hmp. Microbiome. 2021:205136–70.
40. consortium hmp. Microbiome. 2021:205136–70.
41. consortium hmp. Microbiome. 2021:205136–70.
42. consortium hmp. Microbiome. 2021:205136–70.
43. consortium hmp. Microbiome. 2021:205136–70.
44. consortium hmp. Microbiome. 2021:205136–70.
45. consortium hmp. Microbiome. 2021:205136–70.
46. consortium hmp. Microbiome. 2021:205136–70.
47. consortium hmp. Microbiome. 2021:205136–70.
48. consortium hmp. Microbiome. 2021:205136–70.
49. consortium hmp. Microbiome. 2021:205136–70.
50. consortium hmp. Microbiome. 2021:205136–70.
51. consortium hmp. Microbiome. 2021:205136–70.
52. consortium hmp. Microbiome. 2021:205136–70.
53. consortium hmp. Microbiome. 2021:205136–70.
54. consortium hmp. Microbiome. 2021:205136–70.
55. consortium hmp. Microbiome. 2021:205136–70.
56. consortium hmp. Microbiome. 2021:205136–70.
57. consortium hmp. Microbiome. 2021:205136–70.
58. consortium hmp. Microbiome. 2021:205136–70.
59. consortium hmp. Microbiome. 2021:205136–70.
60. consortium hmp. Microbiome. 2021:205136–70.
61. consortium hmp. Microbiome. 2021:205136–70.
62. consortium hmp. Microbiome. 2021:205136–70.
63. consortium hmp. Microbiome. 2021:205136–70.
64. consortium hmp. Microbiome. 2021:205136–70.
65. consortium hmp. Microbiome. 2021:205136–70.
66. consortium hmp. Microbiome. 2021:205136–70.
67. consortium hmp. Microbiome. 2021:205136–70.
68. consortium hmp. Microbiome. 2021:205136–70.
69. consortium hmp. Microbiome. 2021:205136–70.
70. consortium hmp. Microbiome. 2021:205136–70.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.