Ecological Structure Comparison Between Hydrogen Producing and Anaerobic Digestion Prokaryotic Consortia

Marcelo Navarro Díaz[1], Idania Valdez-Vazquez[2], Ana E. Escalante[1*]
[1]Laboratorio Nacional de Ciencias de la Sostenibilidad (LANCIS), Instituto de Ecología, Universidad Nacional Autónoma de México
[2]Instituto de Ingeniería, Unidad Juriquilla, Universidad Nacional Autónoma de México.
* anaelena.escalante@gmail.com

ABSTRACT

To date, high hydrogen yields by dark fermentation has been challenging because tight control of methanogenic microbial assemblies used as inocula, is needed and rarely achieved. This control is commonly seek through aggressive pretreatment of such inocula (e.g. activated sludge, agricultural disposal) which has been somewhat successful despite little knowledge on the effects on the microbial composition and dynamics that we believe is key to overcome challenges like long-term stability and up-scaling of production. Specifically, we propose that ecological analyses with a systemic view of microbial communities composition coupled with modeling of their dynamics and interactions among members is a key step towards a sustainable biohydrogen production. In this work, we reviewed scientific publications corresponding to 104 hydrogen producing (dark fermentation) and 99 methanogenic (anaerobic digestion) experimental settings and constructed databases including microbial composition and environmental conditions for each experiment (e.g. maximum H_2 yield, temperature, pH). With the hydrogen-producing database we performed multivariate (PCA and CCA) and indicator species analyses to investigate correlation of performance (hydrogen production) and culture conditions with key species in these communities. Comparison between hydrogen producing and methanogenic consortia was based on the inferred ecological (co-occurrence) networks derived from the databases. Networks were compared in terms of their robustness, modularity and other topological aspects in order to unravel important interactions between populations of these communities and unknown roles of some groups. From the multivariate analyses we found that richness of the microbial consortia can explain 20% of the variance among experiments, and that richer consortia present higher H_2 yields. We also found that Ruminococcaceae and Oxalobacteraceae families were associated to high hydrogen yields and represent important nodes in the reconstructed networks. In addition, topological features of the anaerobic digestion network suggest that these communities are more robust (i.e. less centrality, higher diameter, density and connectance) than hydrogen producing networks. Overall, multivariate and network analyses showed that more diverse communities (number of different microbial groups) have systemic properties that make them more robust in terms of long term stability.

Keywords: dark fermentation; ecological networks; biohydrogen production.

1. Introduction

The urgent necessity of finding alternatives to fossil fuels has driven the development of sustainable forms of producing less polluting forms of energy. Hydrogen is a high-energy fuel that can be used for generating electricity, in transport vehicles and as a domestic or industrial fuel while its combustion results mainly in water and NO_x traces[1]. Nowadays, hydrogen is mainly produced by high-energy demand processes like water electrolysis and methane reforming [2] so recent research has focused on less costly and more sustainable hydrogen production.

Dark fermentation is a very promising process to produce hydrogen at large scale. It involves the heterofermentation of carbohydrate-rich substrates by prokaryotic consortia and is derived from a more general process of anaerobic digestion that consists of four phases (hydrolysis, acidogenesis, acetogenesis and methanogenesis). In dark fermentation, inocula are pretreated in order to eliminate methanogenic populations and, as a consequence, the process stops at the acidogenesis phase where, along

with volatile fatty acids (VFA's) and alcohols, hydrogen is produced. Following the pretreatment, overall diversity is reduced mainly to Firmicutes species (specially *Clostridium* and *Bacillus* species due to their ability to form spores and resist pretreatment conditions [3]). Yet, pretreatment of inocula has little understood ecological consequences to the process stability (i.e. process collapsing or metabolic changes) and performance.

Dark fermentation allows the coupling of fuel production with sustainable waste management. In addition to hydrogen, VFA´s can be used in hydrogen production by photo-fermentation or for producing biopolymers [4] and the solid residues can be used as compost for agricultural fields. However, dark fermentation faces important obstacles that preclude large-scale implementation, namely, stability issues and low yields that make the process unprofitable.

To overcome many of the obstacles in biohydrogen production it is important to investigate the microbial component of bioreactors from an ecological and systemic perspective. Ecological interactions are important in the sense that they drive communities dynamics and impact organisms evolution [5], potentially leading to variations in reactors performance.

Current knowledge about microbial interactions in hydrogen-producing reactors is mainly focused in granules formation (e.g. *Streptococcus* species[6]), anoxic conditions maintenance (e.g. *Klebsiella* species[6]) and substrate hydrolysis (e.g. *Pseudomonas* and *Bifidobacterium* species[7]). Beyond this processes very little is known or has been investigated. For example, there is very little information on other types of interactions among microbial consortia of the hydrogen-producing bioreactors and we neither know which members or interactions are lost with pretreatments of anaerobic digestion consortia, nor how this impacts ecological processes. Moreover, there are not reviews that explore microbial composition beyond lists of members to evaluate potential interactions or correlations among bioreactor performance, microbial composition and culture conditions to better understand community dynamics.

Multivariate and networks analyses can be the first step into modelling and understanding complex biological systems like microbial communities. On the one hand, multivariate analysis are a classic approach to relate culture conditions and species composition that is underused in microbial ecology [8]. On the other hand, topological network analyses can help to understand the systems functioning [9] and graphically they help to make inferences about important relations between biological entities. A step further can be taken and mathematical models (i.e. Boolean or differential equations) can be designed to study the dynamic properties of such systems [10].

In this work we reviewed scientific publications corresponding to 104 hydrogen producing (dark fermentation) and 99 methanogenic (anaerobic digestion) experimental settings, with which information we constructed databases with composition and environmental conditions, and analyzed it quantitatively. The goals of this review were: (i) to investigate the statistical relationship between microbial compositions in hydrogen producing experiments and culture conditions; (ii) to reconstruct the co-occurrence networks of hydrogen and anaerobic digestion consortia and (iii) to compare both networks in order to find significant changes in their structure.

2. Materials and methods

2.1 Literature search

We conducted a key word search in SCOPUS using "dark fermentation" and "microbial" for hydrogen producing experiments and "anaerobic digestion" and "microbial" for methane producing experiments. We did not restrict the search in terms of publication dates but experiments consisting of two or more digestion phases were excluded from the analyses.

2.2 Data extracted from the literature

Attributes registered for the experiments of hydrogen producing reactors were: hydrogen yield, temperature, pH and microbial composition. For the experiments of anaerobic digestion reactors we only recorded microbial composition. For all cases, we only registered information at the time of maximum production (either hydrogen or methane) based on the assumption of equivalent

ecological status thus potential coexistence of the recorded members of the communities. When possible (43 out of 104 hydrogen producing experiments), the hydrogen yield data was transformed to mol H_2/mol hexose $_{consumed}$.

2.3 Microbial composition data

Microbial composition was analyzed in terms of presence/absence data based on 16S rDNA gene sequence differences as the standard for species identity in prokaryotes [11]. For experiments where no 16S rDNA gene sequences were available, we recorded the reported species identities without any further analysis. For these cases, we assigned a Genbank's accession number only if the closest relative (at least 97% 16S rDNA sequence identity) was also reported. If the similarity between the sequences was between 95 and 97% we recorded the genera identity and if it was between 90 and 95% we recorded the family identity. Then, we downloaded the sequences using the accession numbers and grouped them according to species identities using UCLUST[12] and QIIME[13] (95% similarity cut). The multivariate analyses and network inferences were performed at the species, genera, family and phyla levels, although we only could find and report ecological patterns at the family level.

2.4 Multivariate and statistical analyses

To visualize and identify patterns of environmental variables (ecological parameters of the reactors) and evaluate their correspondence with microbial composition, we followed multivariate and ordination approaches [8].We first standardized parameters of culture conditions and hydrogen yield via z-score transformations [8]. We then performed Principal Component Analysis (PCA) to visualize environmental variance of the different experiments, and Canonical Correspondence Analysis (CCA) was performed to model OTU (family level) response to selected environmental parameters. All these analyses were performed using vegan package[14] implemented in R software [15].

An indicator species analysis was carried out using community composition and hydrogen yield. In order to classify each experiment we categorized the hydrogen yield information using 3 quantiles: low, medium and high yield. We used R[15] and indicspecies package [16]for this analysis.

2.5 Network analysis

To investigate the network properties of the studied microbial communities we reconstructed co-occurrence networks using the complex networks reconstruction algorithm implemented in Sets2Networks [17]. For these analyses we used only experiments where two or more families were reported. In addition we considered as statically significant those correlations which probability value was 0.70 or higher. The average number of neighbors, clustering coefficient, network density, centralization and diameter were measured from the resulting networks using *networks analyzer* function implemented in Cytoscape software [18]. Other topological analyses such as modularity and average path length were measured performed using Gephi software [19]. Finally, we measured connectance as an important topological parameter for ecological networks [20] using the following formula:

$$Connectance = \frac{L}{S^2} \quad (3)$$

Where L is the number of links and S the number of nodes.

3. Results and Discussion

3.1 Multivariate and statistical analyses of microbial composition

Ordination and multivariate analyses showed some important correlations between bioreactor H_2 yield, culture conditions and microbial composition of biohydrogen reactors. On the other hand, PCA results (Figure 1) showed that although there is not a clear clustering of the different sites with respect to culture conditions, features like species richness and H_2 yield explained most of the variance observed among the bioreactors analysed. We identified three components that accounted for most of the variance observed among the analyzed experiment. The first component accounted for 35.08% of the total variation, mainly involving temperature and pH; the second component of the PCA accounted for 27.40% of the total variation, mainly involving H_2 yield;

and the third component accounted for 20.09% of the total variation mainly involving richness. It is worth to mention that reactors that were maintained at high temperature (maximum temperature: 70°C) tend to have a medium to high H_2 yields (Med: 1.6-2.16 mol H_2/ mol $hexose_{consumed}$; High: 2.16-3.2 mol H_2/ mol $hexose_{consumed}$). On the other hand, the CCA (Figure 2) showed that families present at high temperatures tend to be associated with some of the highest H_2 yields. CCA also showed that some families where associated with high hydrogen yields (Ruminococcaceae and Oxalobactereceae). These families were also linked to high hydrogen yield sites in the indicator species analysis (although this result was not statistically significant; p value= 0.405 and 0.31). Streptococcaceae, Bacillaceae and Pseudomonaceae were not associated with the highest yields of hydrogen. This is surprising since these families have members associated with granules formation and anoxic conditions maintenance that promote Clostridiaceae species growth [3]. It is worth mentioning that Streptococcaceae and Pseudomonadaceae were located at the same point on the graph meaning that they tend to co-occur in most experiments with different culture conditions.

Clostridiaceae family was not associated with any of the two CCA axes implicating first, that Clostridiaceae species can grow under a wide variety of environmental conditions, and second, that not all of its members produce hydrogen. These results are in accordance with some studies where high hydrogen yield is achieved even without inoculum pretreatment which reduces the diversity to mainly Clostridiaceae species[21]. Thermoanaerobacteraceae family correlated with low hydrogen yields, although it is known that its presence (in high temperature experiments) is associated with high hydrogen production [6][22]. Enterobacterecae family was related to high hydrogen yield which was expected since some species like *Klesiella pneumoniae* are associated with granules formation which may promote hydrogen producers growth [6]. Finally some other families like Flavobacteraceae where associated with high H_2 yield although little is known about their role in hydrogen production. It has to be noted here that the number of studies that were reviewed may not be numerically representative of all the taxonomical groups present in hydrogen producing bioreactors, thus caution should be taken regarding major generalizations.

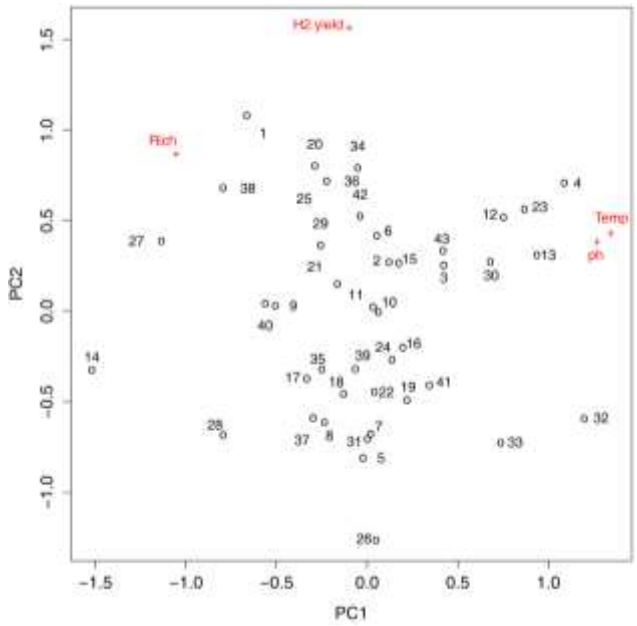

Figure 1. Biplot generated from a principal component analysis (PCA) of the standardized environmental variables (pH; Temperature, Temp; Maximum hydrogen yield, H_2 yield) and species richness (Rich). The different consortia (experiments) are shown in the graph as open circles (1-43). Vectors show scaled environmental variables. The first component of the PCA analysis accounted for 35.08% of the total variation, mainly involving temperature and pH. The second component of the PCA accounted for 27.40% of the total variation, mainly involving H_2 yield. The third component accounted for 20.09% of the total variation mainly involving richness.

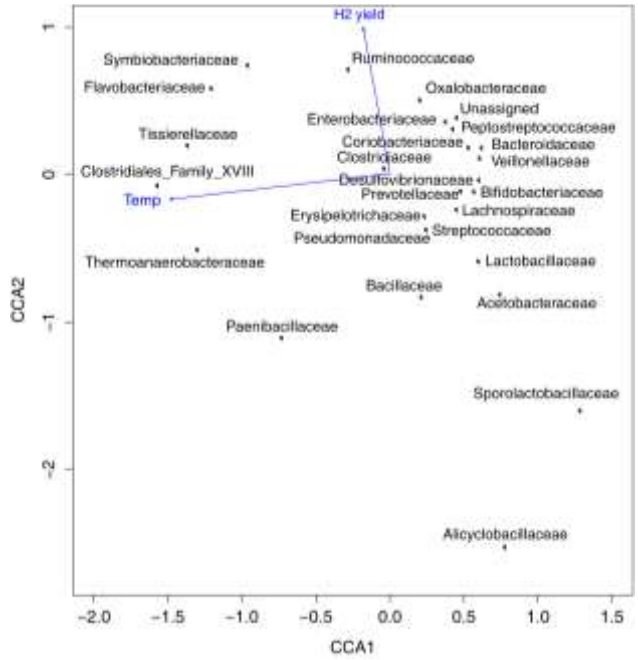

Figure 2. Canonical correspondence analysis (CCA) modeling OTUs (family level) response to environmental parameters of hydrogen producing bioreactors. Text corresponds to family names for the prokaryotes found in the experiments included in the analysis (n=43), and vectors show scaled environmental variables. Two explanatory variables are shown (Temp=temperature, H_2 yield=maximum hydrogen production). The first axis accounted for 4.4% of the total variation, mainly involving temperature. The second axis accounted for 7.34% of the total variation, mainly involving H_2 yield.

3.2 Network analysis

At the family level, the hydrogen producing consortia network (Figure 3) showed great centrality (centrality ranges from 0 to 1) being the Clostridiaceae node the most important of the network (given by its betweenness centrality values). Other important nodes in the network are Bacillacaeae and Thermonaerobacteraceae, the latter being especially important at high temperature cultures. The anaerobic digestion network (Figure 4) showed a very central node (Clostridiaceae) but in addition, other nodes are also significantly central (Methanosarcinaceae, Methanomicrobiaceae and Methanosaetaceae).

The topological attributes of both networks (Table 2) are similar, but anaerobic digestion network tends to be bigger, more connected, less modular and much less central. Both networks are almost of the same size (number of nodes and network diameter) which can be due to the lack of statically significant co-occurrence interactions that results from the sample size. Since anaerobic digestion consortia are more diverse (in terms of number of different microbial species) than the hydrogen consortia, more sets of experiments (samples) may be required to establish significance to certain co-occurrences. Connectance of the anaerobic digestion network is greater but both networks have low connectance, even when compared with macroorganisms networks [20]. This is not surprising since the data used in this review is not from high-throughput sequencing and the nodes where defined at a family level, both conditions greatly reduce the recovered diversity and can prevent the detection of co-occurrence patterns that might occur at lower taxonomic levels.

The intersection network (Figure 5) showed the 12 shared nodes and co-ocurrences between both networks. It is worth noting that in both networks the co-occurrence between Clostridiaceae, Bacillaceae and Thermoanaerobacteriaceae members was significant. This might indicate and important interaction between these three families. The presence of Lactobacillaceae members is expected due to the origins of many of the inocula and substrates (i.e. wastewater treatment plants, animal feces and food

waste). Also the presence of Ruminococcaceace members (gen: *Ruminococcus, Ethanoligenens*) in both networks in addition to its probable importance in hydrogen producing consortia requires further research. Oxalobacteraceae (gen: *Janthinobacterium*) along with Ruminococcaceae, are somewhat important nodes in the hydrogen network (given their degree=3 for both; the average degree for all the nodes in the network is 2.7 excluding Clostridiaceae). Although members of both families are commonly found in hydrogen producing experiments little is known about their ecological role in these communities.

Table 2. Topological attributes of the resulting networks at a family level.

Attribute	Hydrogen Network	Anaerobic digestion Network
Avg. number of neighbors	3.548	4.343
Avg. path length	2.043	2.112
Clustering coefficient	0.453	0.536
Connectance	0.057	0.05864
Modularity	0.312	0.210
Network density	0.118	0.128
Network centralization	0.836	0.676
Network diameter	4	4
Nodes	31	35
Unique nodes	19	23
Shared nodes	12	

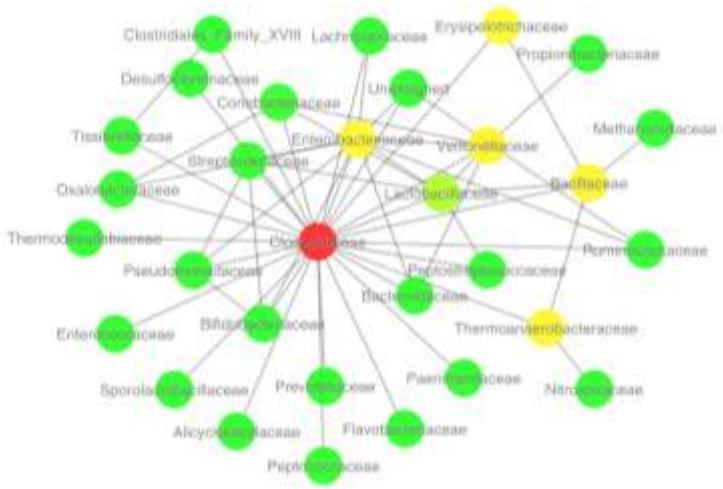

Figure 3. Co-occurence network of the hydrogen producing consortia at a family level. Color accounts to betweeness centrality of the nodes from low (green) to high values (red).

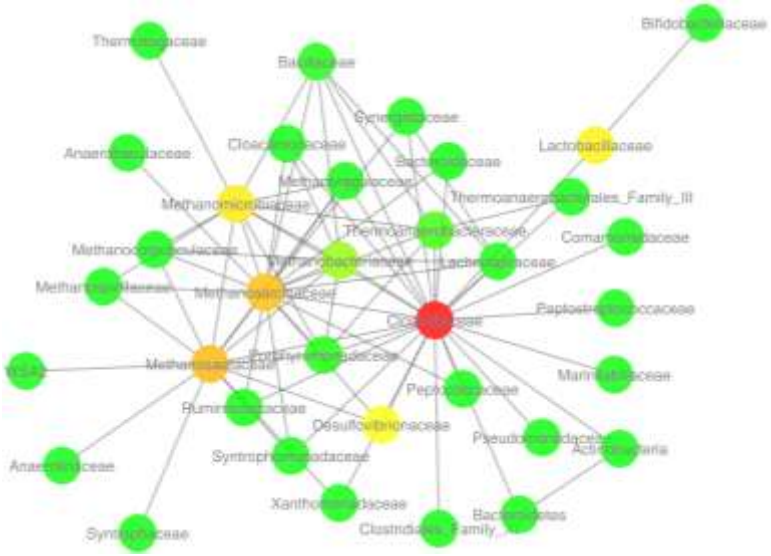

Figure 4. Co-occurrence network of the anaerobic digestion consortia at a family level. Color accounts to betweenness centrality of the nodes from low (green) to high values (red).

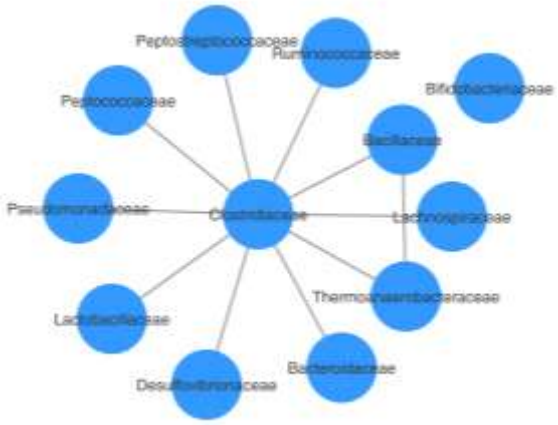

Figure 5. Intersection network between hydrogen producing and anaerobic digestion consortia.

4. Summary and perspectives

Overall, multivariate and network analyses showed that more diverse communities in hydrogen producing bioreactors have systemic properties that make them more robust in terms of long term stability.

- Multivariate and statistical analyses of hydrogen producing bioreactors showed correspondence of richness with high hydrogen yield. Although caution should be taken given relatively small sample size and methodological biases in species identification (mainly DGGE profiling), the reason behind this richness-high H_2 yield correspondence might be that, as more *Clostridium* species are present more chances that hydrogen is produced from a more wide variety of substrates [23].
- Multivariate Canonical Correspondence Analysis (CCA) identified Ruminococcaceae and Oxalobacteraceae families as associated to high hydrogen yields, and also conformed important nodes in the reconstructed networks. Further research has to be conducted to determine the role these groups have in hydrogen consortia since little is known at this respect. Also, it remained unclear the role in hydrogen production of some families that are associated with improving conditions in the bioreactors. Apparently, these families have a positive impact not only in hydrogen producing microorganisms but also on "cheaters" that take advantage from these conditions.
- Network analysis did not show major topological differences between hydrogen-producing and methanogenic consortia. However, the anaerobic digestion (methanogenic) network had topological features that suggest that these communities are more robust (i.e. less centrality, higher diameter, density and connectance) [20]. This is concordant with the hypothesis that more rich communities tend to be more robust in terms of resilience and resistance (although in microbial communities the evidence for this is mixed) [24].

Finally, it is important to remark that although big efforts have been made in the research community in order to explore hydrogen-producing microbial consortia, it is evident that to gain a deeper understanding a different approach should be followed. Specifically, high-throughput sequencing should be considered when investigating these consortia because in that way we can capture, with more detail and deeper sampling, information about its composition and the abundance of the members of these communities. With this information we can discover patterns at a finer scale and we could be able to model mathematically these communities, which is important since we can use models to predict changes in reactors performance given some information about culture conditions. Also, we can use this information in the construction of engineered synthetic microbial consortia that could significantly improve this process performance and profitableness.

5. Acknowledgements

This work was funded by Consejo Nacional de Ciencia y Tecnología (CONACyT grant 16891) as a scholarship to MN (No. 21236).

6. References

[1] Momirlan M, Veziroglu T. The properties of hydrogen as fuel tomorrow in sustainable energy system for a cleaner planet. Int J Hydrogen Energy 2005;30:795–802.

[2] Salvi BL, Subramanian K a., Panwar NL. Alternative fuels for transportation vehicles: A technical review. Renew Sustain Energy Rev 2013;25:404–19.

[3] Show KY, Lee DJ, Tay JH, Lin CY, Chang JS. Biohydrogen production: Current perspectives and the way forward. Int J Hydrogen Energy 2012;37:15616–31.

[4] Lee WS, Chua ASM, Yeoh HK, Ngoh GC. A review of the production and applications of waste-derived volatile fatty acids. Chem Eng J 2014;235:83–99.

[5] Freilich S, Kreimer A, Meilijson I, Gophna U, Sharan R, Ruppin E. The large-scale organization of the bacterial network of ecological co-occurrence interactions. Nucleic Acids Res 2010;38:3857–68.

[6] Hung C-H, Chang Y-T, Chang Y-J. Roles of microorganisms other than Clostridium and Enterobacter in anaerobic fermentative biohydrogen production systems--a review. Bioresour Technol 2011;102:8437–44.

[7] Weiland P. Biogas production: current state and perspectives. Appl Microbiol Biotechnol 2010;85:849–60.

[8] Ramette A. Multivariate analyses in microbial ecology. FEMS Microbiol Ecol 2007;62:142–60.
[9] Newman M. The structure and function of complex networks. SIAM Rev 2003;45:167–256.
[10] Faust K, Raes J. Microbial interactions: from networks to models. Nat Rev Microbiol 2012;10:538–50.
[11] Stackebrandt E, Goebel BM. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. Int J Syst Bacteriol 1994;44:846–9.
[12] Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010;26:2460–1.
[13] Caporaso J, Kuczynski J, Stombaugh J. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7:335–6.
[14] Oksanen J, Blanchet G, Kindt R, Legendre P, Minchin PR, O'Hara RB, et al. vegan: Community Ecology Package 2013.
[15] R Core Team. R: A Language and Environment for Statistical Computing 2013.
[16] De Cáceres M, Legendre P. Associations between species and groups of sites: indices and statistical inference. Ecology 2009;90:3566–74.
[17] Clark NR, Dannenfelser R, Tan CM, Komosinski ME, Ma'ayan A. Sets2Networks: network inference from repeated observations of sets. BMC Syst Biol 2012;6:89.
[18] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.
[19] Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. ICWSM 2009:361–2.
[20] Dunne J a, Williams RJ, Martinez ND. Food-web structure and network theory: The role of connectance and size. Proc Natl Acad Sci U S A 2002;99:12917–22.
[21] Kim D-H, Wu J, Jeong K-W, Kim M-S, Shin H-S. Natural inducement of hydrogen from food waste by temperature control. Int J Hydrogen Energy 2011;36:10666–73.
[22] Rittmann S, Herwig C. A comprehensive and quantitative review of dark fermentative biohydrogen production. Microb Cell Fact 2012;11:115.
[23] Tracy BP, Jones SW, Fast AG, Indurthi DC, Papoutsakis ET. Clostridia: the importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. Curr Opin Biotechnol 2012;23:364–81.
[24] Shade A, Peter H, Allison SD, Baho DL, Berga M, Bürgmann H, et al. Fundamentals of microbial community resistance and resilience. Front Microbiol 2012;3:417.