

Topological Data Analysis of Escherichia coli O157:H7 and Non-O157 Survival in Soils

ABASIOFIOK MARK IBEKWE, Jincal Ma, David E Crowley, Ching-Hong Yang, Alexis M. Johnson, Tanya C. Petrossian⁴ and Pek Y. Lum

Journal Name: Frontiers in Cellular and Infection Microbiology
ISSN: 2235-2988
Article type: Original Research Article
Received on: 27 May 2014
Accepted on: 18 Aug 2014
Provisional PDF published on: 18 Aug 2014
www.frontiersin.org: www.frontiersin.org
Citation: Ibekwe AM, Ma J, Crowley DE, Yang C, Johnson AM, Petrossian4 TC and Lum PY(2014) Topological Data Analysis of Escherichia coli O157:H7 and Non-O157 Survival in Soils. *Front. Cell. Infect. Microbiol.* 4:122. doi:10.3389/fcimb.2014.00122
Copyright statement: © 2014 Ibekwe, Ma, Crowley, Yang, Johnson, Petrossian4 and Lum. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](http://creativecommons.org/licenses/by/2.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

This Provisional PDF corresponds to the article as it appeared upon acceptance, after rigorous peer-review. Fully formatted PDF and full text (HTML) versions will be made available soon.

Topological Data Analysis of *Escherichia coli* O157:H7 and Non-O157 Survival in Soils

A. Mark Ibekwe^{1,*}, Jincai Ma^{1,2}, David E. Crowley², Ching-Hong Yang³, Alexis M. Johnson⁴,
Tanya C. Petrossian⁴, Pek Yee Lum⁴

¹ USDA-ARS U. S. Salinity Laboratory, Riverside, CA 92507

² Department of Environmental Sciences, University of California, Riverside, CA 92521

³ Department of Biological Sciences, University of Wisconsin, Milwaukee, WI 53211

⁴ Ayasdi, Inc., Menlo Park, CA 94025

*Corresponding author:

A. Mark Ibekwe

USDA-ARS-U. S. Salinity Laboratory

450 W. Big Springs Rd

Riverside, CA 92507.

Phone: 951-369-4828

Fax: 951-342-4964

E-mail: Mark.Ibekwe@ars.usda.gov

Key words: Shiga toxin, contamination, survival time, fresh produce, organic, conventional

Running Title: Topology of *E. coli* O157:H7 and non-O157 survival in Soils

Abstract

Shiga toxin-producing *E. coli* O157:H7 and non-O157 have been implicated in many foodborne illnesses caused by the consumption of contaminated fresh produce. However, data on their persistence in soils are limited due to the complexity in datasets generated from different environmental variables and bacterial taxa. There is a continuing need to distinguish the various environmental variables and different bacterial groups to understand the relationships among these factors and the pathogen survival. Using an approach called Topological Data Analysis (TDA); we reconstructed the relationship structure of *E. coli* O157 and non-O157 survival in 32 soils (16 organic and 16 conventionally managed soils) from California (CA) and Arizona (AZ) with a multi-resolution output. In our study, we took a community approach based on total soil microbiome to study community level survival and examining the network of the community as a whole and the relationship between its topology and biological processes. TDA produces a geometric representation of complex data sets. Network analysis showed that Shiga toxin negative strain *E. coli* O157:H7 4554 survived significantly longer in comparison to *E. coli* O157:H7 EDL 933, while the survival time of *E. coli* O157:NM was comparable to that of *E. coli* O157:H7 EDL 933 in all of the tested soils. Two non-O157 strains, *E. coli* O26:H11 and *E. coli* O103:H2 survived much longer than *E. coli* O91:H21 and the three strains of *E. coli* O157. We show that there are complex interactions between *E. coli* strain survival, microbial community structures, and soil parameters.

Introduction

Food-borne outbreaks associated with contaminated produce have heightened concerns about the adequacy of control measures for the safe production of fresh fruits and vegetables. In the past decade, there have been over 70 fresh produce-related outbreaks in the United States, and the risk and burden is continuous (Lynch et al., 2009; Brandl 2006). These vegetables have been implicated in approximately 20 outbreaks resulting in approximately 700 illnesses and 20 deaths between 1996 and 2006 (Allerberger and Sessitsch, 2009; Doyle and Erickson, 2008). Although there are leafy green vegetable associated outbreaks caused by *Salmonella* and *Cyclospora*, a majority of them have been due to food contamination with *Escherichia coli* O157:H7 (Sivapalasingam et al 2004). The most likely mechanisms of *E. coli* O157: H7 contaminations include contamination from soil amendments (i.e., manure, compost and compost teas), water (irrigation or flooding/runoff from adjacent land), wildlife, and airborne deposition from off-farm activities such as cattle/dairy and manure/composting operations (Franz et al., 2011, 2008, Fremaux et al., 2008). One of the worst incidents to date was a multistate *Escherichia coli* O157:H7 outbreak in August and September 2006, which was associated with consumption of fresh, bagged spinach that was traced to a field in California (CALFERT, 2007a&b; Jay et al., 2007; Cooley et al., 2007). During this outbreak, the CDC reported over 200 illnesses, 104 hospitalizations and 3 deaths.

Although *E. coli* O157:H7 is reported to be the predominant STEC serotype in the United States, more than 200 non-O157 STEC serotypes have been identified in animals or foods (Karch et al., 2005). Approximately, 60 of these serotypes have been incriminated in human diseases. Recent epidemiological studies have recognized additional non-O157 serotypes, including O26, O45, O91, O103, O104, O111, O113, O121, and O145, among STEC strains that were linked to severe human disease in the United States, Europe and parts of Latin America

(Brooks et al., 2005; Bettelheim 2007; Caprioli et al., 2005; Mathusa et al., 2010; Beutin and Martin 2012).

The mechanisms by which the pathogen is introduced into the produce are not fully understood; however, it is hypothesized that plants become contaminated when grown in fields fertilized with improperly treated manure (Beuchat, 1999) or flood irrigation with water contaminated with cattle feces or contact with contaminated surface runoff (Hillborn et al., 1999; Ibekwe et al., 2004). Depending on the soil properties and environmental factors, the survival time of *E. coli* O157:H7 in soils varies from one week to six months, and even longer in some extreme cases (Ibekwe et al., 2007, 2010; Ibekwe and Ma, 2011; Ma et al., 2011; Maule 2000; Mubiru et al., 2000; Franz et al., 2008; Jiang et al., 2002; Semenov et al., 2008).

In this study, we integrated environmental data with microbial community to assess relationships among these factors and the pathogen survival. To this end, we propose a systematic evaluation of the relative effectiveness of current and potential new intervention strategies to reduce or prevent contamination of produce by employing a new analysis method called topological data analysis (TDA) (Carlsson et al., 2009; Lum et al., 2013), to uncover environmental variables that are correlated with survival of *E. coli* O157. TDA is based on an area of mathematics called topology and its implementation allows topological techniques to be used to discover subtle signals or “shape” in complex data such as this dataset. This approach has been used in the past to discover hard-to-identify signal in other complex datasets around viral evolution, breast cancer, diabetes and effects on the metagenome due to environmental stress (Chan et al., 2013; Sarikonda et al., 2013; Probst et al., 2014; Nicolau et al., 2011). We used TDA to reconstruct the relationship structure of *E. coli* O157:H7 and non-O157 survival in 32 soils (16 organic, 16 conventional) from California (CA) and Arizona (AZ) with a multi-resolution output. We show

that differential survivability of various *E. coli* strains are dependent on microbial community structures and soil parameters.

MATERIALS AND METHODS

Datasets and bacterial strains for the dataset.

Environmental and metagenomic data were obtained from three studies of the survival pattern of *E. coli* O157:H7 and non O157 from produce growing region of California and Arizona. The first study (Ma et al., 2012) examined the effects of environmental variables on the survival of *E. coli* O157:H7 EDL 933. The second study (Ma et al., 2013) examined the effects of 454 FLX-derived sequences from the same soils on survival of *E. coli* O157:H7 EDL933. The third study (Ma et al., 2014) examined the effects of environmental variables on the survival of *E. coli* O157:H7 and non O157. All of the *E. coli* strains used in this study are described in Table 1. All soil properties are as reported by Ma et al. (2012).

Collection, characterization, inoculation of soils samples, and survival

Soil samples were collected from three major fresh produce growing areas: Salinas Valley California, Imperial Valley, southern California, and Yuma, Arizona (Ma et al., 2012). *E. coli* O157:H7 culture, a 1.0 ml aliquot was transferred into a 250 ml flask containing 100 ml LB (Luria-Bertani) broth, and incubated at 37 °C for 18 h to achieve early stationary phase. The cells were harvested by centrifugation at 3500 g (Beckman, Brea, CA), washed three times using 10 mM phosphate buffer (10 mM, pH 7.2), and finally resuspended in deionized water, and cells

were added in soils to a final density of about 0.5×10^7 CFU per gram soil dry weight (gdw^{-1}) according to a method slightly adapted from Franz et al., (2008). About 500 g of the inoculated soil was transferred to a plastic bag which was closed but which had some holes at the top to allow air exchange for survival studies. The inoculated soils were sampled (1 g) at days 0, 3, 6, 10, 14, 20, 27, 34, 40, and 48 to determine the survival of *E. coli* O157 and non O157 over time. Details of the experimental procedure had previously been described (Ma et al., 2012).

Soil DNA Extraction, Pyrosequencing and sequence data analysis

Community DNA was extracted from 32 leafy green-producing soils using a Power Soil Extraction Kit (MO BIO Laboratories, CA) with the bead-beating protocol supplied by the manufacturer. The quality and concentration of the soil DNA were assessed using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, DE). The overall size of the soil DNA was checked by running an aliquot of soil DNA on a 1.0% agarose gel. The soil DNA samples (15.0 μl) were then submitted to Research and Testing Laboratories (Lubbock, TX) for PCR optimization and pyrosequencing analysis. Bacterial tag-encoded FLX amplicon pyrosequencing were carried out as previously described (Acosta-Martinez et al., 2008; 2010). The 16S universal Eubacterial primers 530F (5'-GTG CCA GCM GCN GCG G) and 1100R (5'-GGG TTN CGN TCG TTG) were used for amplifying the ~600 bp region of 16S rRNA genes. Primer and PCR optimizations were done at the Research and Testing Laboratories (Lubbock, TX) according to the protocol described previously (Acosta-Martinez et al., 2010; Nonnenmann et al., 2010; Gontcharova et al., 2010). All FLX related procedures were performed following Genome Sequencer FLX System manufacturers instructions (Roche, NJ, USA). Bacterial pyrosequencing

population data were further analyzed by performing multiple sequence alignment techniques using the `dist.seqs` function in MOTHUR, version 1.9.1 (Schloss et al., 2009).

Data Analysis

All data were analyzed using the Ayasdi software (<http://www.ayasdi.com>). The Ayasdi software uses TDA as a framework for a large repertoire of statistical and machine learning methods. The description of the implementation of TDA as a software is described in detail in the following publication (Lum et al., 2013). Briefly, the output consists of a topological network with nodes and edges, where nodes are collections of data points and an edge connects any two nodes that have one or more common data points. In this analysis, the mathematical functions (called “lenses” in the software) used are principal metric SVD 1 and 2. Principal metric SVD lenses are used when the distance metric used is non-Euclidean. Statistical test used to look at significance between sub-networks or groups is the non-parametric Kolmogorov-Smirnov test (KS score). Variables used in the analysis are the following: chemical (Sodium (Na), iron (Fe), potassium (K), electrical conductivity or salinity (EC), copper (Cu), assimilable organic carbon (AOC), total nitrogen (TN), calcium (Ca), Nickel (N), organic carbon (OC), microbial biomass carbon (MBC), sulfate (SO₄), water holding capacity (WHC), magnesium (Mg), zinc (Zn), phosphate (PO₄), molybdenum (Mo), physical (sand, clay, silt, and bulk density) and biological (time till detection for *E. coli* O157:H7 EDL933 [ttd(d)], time till detection for *E. coli* O157:H7 strain 4555 [ttd (d) O157-4554], time till detection for *E. coli* O157:H7 non-motile strain 4555 [ttd (d) O157NM], time till detection for *E. coli* O91 [ttd (d) O91], time till detection for *E. coli* O26 [ttd (d) O26], operation taxonomic units (OTUs), *Nitrospira*, diversity index (H'),

Proteobacteria, Alphaproteobacteria, Chloroflexi, Bacteroidetes, Acidobacteria, Actinobacteria, Gemmatimonadetes, Firmicutes, Verrucomicrobia, Deltaproteobacteria, Gammaproteobacteria, Planctomycetes, Betaproteobacteria).

RESULTS

Soil sample site similarities and management network

Using the properties of physical, chemical, and biological characteristic of these soil samples as variables, we clustered the soil samples using TDA. The resulting network represents the soil samples clustering into sub-networks. Figure 1A shows 4 sub-networks A, B, C and D with B and C connecting to form a larger sub-network. There is also a singleton (1 node comprising of 2 soil samples from Salinas that stood apart from everything else). The network can also be colored by various factors and characteristics such as location and soil management type for visualization (Fig 1). In addition, we can also apply statistics to probe what factors distinguished our soils into sub-networks. We found that “location” was one of the key differences between the sub-networks (Kolmogorov-Smirnov test $PV < 0.0003$). In order to visualize the effect of “location” on the soil samples, Fig 1 is colored by “location”. We show that soil samples from the Salinas areas (A) completely formed a separate sub-network from soil samples from the Imperial and Yuma areas (B, C, and D) as indicated by the color. This indicates that physical, chemical, and biological characteristic of these soil samples collectively are quite different from location to location, especially the soil samples from Salinas, which formed a distinct sub-network (A). Soil samples from Yuma and Imperial are closer to each, forming a sub-network that looks like a dumb bell, with some samples from Imperial clustering at left side of dumb bell

(B) and the rest of the network comprised of a mixture between samples from Yuma and Imperial. Interestingly, physical, chemical, and biological properties measured of these soil samples did not differentiate between conventional and organic soil management as seen from the non-enrichment of any one type of soil management in the network (also see Table 2, where the P-value for soil management as a differentiating factor between those sub-networks was 0.4126). To further investigate sub-network D and the singleton, another network analysis was performed using the same distance metric and mathematical lenses but at a lower resolution (20 instead of 30). Sub-network D, which comprised of samples from Yuma and Imperial, became part of sub-network C (Fig 1C). The singleton however remained a singleton, indicating that these samples are fundamentally different from the rest of the samples due to unknown reasons including quality of the samples.

Statistical analysis to identify key distinguishers of these sub-networks were performed on all numerical columns on all data points (Table 2) including detection times, biodiversity measures, management, location, sand, silt, clay, soil pH, bulk density, assimilable organic carbon (AOC), organic carbon (OC), microbial biomass carbon (MBC), electrical conductivity (EC), chemical compound (Na^+ , K^+ , Ca^{2+} , Mg^{2+} etc.) and bacterial phyla. Soil sand content was significantly higher ($P= 0.0027$) for soils from the Salinas Valley area (Fig. 2A), whereas silt and clay contents were significantly higher ($P= 0.0403$ for silt and 0.0012 for clay) in soils from the Imperial and Yuma Valley areas (Fig. 2 B&C). Soil pH was between 6.7 and 8.0, with significantly higher pH ($P = 0.025$) occurring in the Yuma/Imperial Valley areas (Fig. 2D). Soil bulk density values ranged between 1.22 and 1.63 mg, with soils from the Salinas Valley having significantly higher bulk densities ($P = 0.0162$); Fig. 2E). Statistical tests indicated that total iron, PO_4 and calcium were significantly higher ($P = 0.0299$; 0.0399 ; $5.08\text{E-}4$, respectively) in Salinas

Valley samples than samples from Yuma and Imperial Valleys. On the other hand, sodium and sulfate were significantly higher ($1.05E-07$; -05 , respectively) in Yuma and Imperial Valley samples. No differences were observed among the locations in soil contents of total nitrogen (TN).

Survival behavior of *E. coli* O157:H7 in soils

Next, we investigated survival of different *E. coli* strains in these different soil sub-networks.

The network remains the same but we can now probe the network to see if any survival variables show any significant trends between these sub-networks. To do this we colored the same network by the length of survival of *E. coli* O157:H7 EDL933, *E. coli* O157:NM, and *E. coli* O157 strain 4554 (stx-) across the topological network to observe if differences exist in the soil networks. The shortest survival time (ttd) was observed for *E. coli* O157:H7 EDL933 (13.8 to 32.6 days) while the longest was observed for *E. coli* O157:NM (20.6 to 56.0 days) and *E. coli* O157:H7 strain 4554 as intermediate at 21.1 to 45.0 days (Fig. 3 a, b, c). Figure 3 is colored by survival time of the indicated strain for all the soils. We also performed statistical test on the survival times and show that the survival time of *E. coli* O157:H7 EDL933 was significantly longer in soils from the Salinas Valley area ($8.18E-05$), whereas the survival time of *E. coli* O157:NM and the stx- *E. coli* O157:H7 strain 4554 were not significantly different in soils from the Salinas Valley area (0.0995 and 0.4823, respectively) and in soils from the Yuma and Imperial Valley region (Table 2). Furthermore, the coloring pattern indicates no differences in survival (ttd) between organic and conventional soils from Imperial Valley and Salinas. Survival time was much shorter in the organic soil than the conventional soils with *E. coli* O157:NM. This can be observed by the deep blue color (Fig. 3. B).

Survival of non-O157 in soils was longer than *E. coli* O157:H7 except *E. coli* O91:H21. It was found that two non-O157 strains, *E. coli* O26:H21 and *E. coli* O103:H2 survived much longer than *E. coli* O91:H21. The three non-O157 strains survived significantly longer (*E. coli* O91:H21: 4.12E-06, O26:H21:0.079, and O103: 2.84E-06) in soils from the Salinas Valley region than in soils from the Yuma and Imperial Valleys (Table 2). There were no differences of survival between organic and conventionally managed soils with the non O157 strains. In the current study no isogenic strains (with and without stx) were used. When the six *E. coli* O157 and non-O157 strains were grouped together on the same scale it was shown that *E. coli* O103:H2 survived the longest in all the soils, followed by *E. coli* O26:H21 (Fig. 4).

Bacterial abundance and distribution as revealed by 454 pyrosequencing

We then analyzed how the abundance and distribution of different bacterial phyla based on pyrosequencing from the 32 soils collected from the three regions associated with the different regions clustered in the networks. By coloring the same networks with now the abundance of the different bacterial phyla, we show that there are marked differences between the distributions of the different bacterial phyla (Fig. 5). As shown in Figure 5A the nodes that are colored red indicate significantly higher ($P = 0.097$) percentage of *Acidobacteria* (see Table 2 for details of the phyla) in soils from the Salinas Valley area than soils from Yuma and Imperial Valleys. These soils also contained a higher percentage of *Deltaproteobacteria* ($P = 8.55E-06$), *Alphaproteobacteria* ($P = 0.0047$) (Fig. 5 d, f) as seen by the color scheme. Significant differences in *beta* ($P = 1.95E-05$) and *Gammaproteobacteria* ($P = 6.92E-05$) were also observed in soils from Yuma/Imperial Valleys and soils collected from Salinas Valley area (Fig. E&G). Significant differences were also found in *Proteobacteria* ($P = 1.63E-04$) (Fig. 5 C),

Actinobacteria ($P = 0.024$) (Fig. 5B), and no significant differences in *Firmicutes* ($P = 0.1989$) (Fig. 5H) from the three regions. Further analysis showed that *Actinobacteria*, *Proteobacteria*, *Acidobacteria*, and *Bacteroidetes* were the dominant phyla among the bacterial communities in soils, and these four phyla accounted for about 75% of the total bacterial composition based on pyrosequencing (Fig. S1). The current analysis has produced the same trends as the results obtained with our previous analysis that was based on correlation between survival time and dominant bacterial communities (Ma et al., 2013). In this earlier related study, stepwise multiple regression analysis was conducted and the results showed that EC, TN, and AOC were the most important factors impacting the survival of *E. coli* O157:H7 in all soils tested (Fig. S2), with EC showing the most negative effect ($P < 0.001$) on survival and TN and AOC showing positive effects ($P < 0.01$) (Ma et al., 2012).

Discussion

Analysis, interpretation, and visualization of complex data are major tasks confronting researchers today. Most data are presented in tabular formats after traditional statistical analysis. To better understanding the influence of selected soil properties and their impact on bacteria growth, we used TDA as implemented by the Ayasdi software. TDA can analyze disparate datasets in one setting, as well as presents topological networks as an informative visualization for understanding and interpretation. The TDA approach is sensitive to both large and small scale patterns that often fail to be detected by other analysis methods, such as principal component analysis, (PCA), multidimensional scaling, (MDS), and cluster analysis (Carlsson et al, 2009; Lum et al 2013). In addition, we note that PCA and MDS produce 2-D scatterplots that are often hard to separate more subtle signal from noise. In addition, clustering methods produce

distinct, unrelated groups that may obscure signal that is better captured using TDA, which is inherently suited to look for continuity in signal.

The three key concepts of topological analysis methods include coordinate freeness, which means that topology has the capability to measure properties of intrinsic shapes of data which is independent of the coordinate system. Coordinate free representations are vital when one is studying data collected with different technologies such as pyrosequencing or from survival data as we have used in this study or from different laboratories when the methodologies cannot be standardized (Lum et al., 2013). This is very critical to a study such as ours where the data collected are not from one uniform platform. As mentioned earlier, TDA has also been applied to various different studies to uncover complex signals (Romano et al., 2014; Chan et al., 2013; Sarikonda et al., 2013; Lum et al., 2013; Nicolau et al., 2011).

We have demonstrated that location is an important factor that we found to be associated with high survival of certain bacteria strains. Recent studies of metabolic network topologies across the bacterial tree of life revealed marked variation in network cluster and identified several genetic and environmental determinants affecting metabolic clustering (Parter et al., 2007). These authors showed that reduced metabolic cluster in single-species networks is associated with organisms inhabiting less variable environments. Our analysis, however, presents a unique characterization of microbial community-level cluster and demonstrates consistent differences that are associated with survival of *E. coli* O157:H7 from different locations. It should be noted that the correlation of certain bacterial phyla (*Actinobacteria* and *Acidobacteria*) with higher survival of *E. coli* O157:H7 does not necessarily mean causation of higher survival, and therefore, should be extrapolated very carefully. As discussed by Greenblum et al. 2012, in silico models of microbial communities are currently still scarce (Oberhardt et al., 2009) and

mostly focus on simulated communities comprising a handful of species and on pair-wise interactions among community members (Klitgord and Segre, 2009; Stoloy et al., 2007; Wintermute and Silver, 2010; Freilich et al., 2010). Experimental validation at the species or gene level of model components and parameters may be necessary for a successful and accurate understanding of individual species effects on survival. In essence, this study represents an important step in the development of a metagenomic systems biology approach. Such an approach can potentially advance metagenomic research in the same way systems biology advanced genomics, appreciating not only the parts list of a system but the complex interactions among parts and the impact of these interactions on function and dynamics.

In summary, the TDA networks identified various environmental factors that correlate with increased or decreased survival of *E. coli* O157 in the three regions. In particular, we have identified a group of environmental factors such as EC, TN, AOC, etc. that consistently may enhance or inhibit survival of this pathogen from the three regions, and these factors were in agreement with some of our earlier studies from the same locations (Ma et al., 2012; 2013). We note that the effects of different environmental factors and bacterial community were easily detected by TDA because of the inherent ability of the analysis environment that allows analysis of all these factors simultaneously. Often times classical clustering approaches by themselves will miss these subtle signals because of the need to place data points into one cluster or another. This could end up highlighting only the most obvious signals while breaking up the more subtle ones.

As we move towards better understanding of how *E. coli* O157:H7 contamination could occur in the food chain, we believe a more holistic approach such as looking at all possible available factors together is important. However, because this creates complexity, there is a

need to apply different approaches. We used here an approach to allow not only the mathematical analysis needed to uncover small signal but also the ability to visualize these complex relationships.

Acknowledgements

This research was supported by CSREES NIFA Agreement No., 2008-35201-18709 and the 214 Manure and Byproduct Utilization Project of the USDA-ARS. We thank Drs Jorge Fonseca of the University of Arizona Yuma, Mark Trent, UC-Davis, Imperial Agricultural Experiment Station, and James McCreight of USDA-ARS Salinas CA for providing soil samples for this study. We also thank Damon Baptista for technical help. Mention of trademark or propriety products in this manuscript does not constitute a guarantee or warranty of the property by the USDA and does not imply its approval to the exclusion of other products that may also be suitable.

References:

- Allerberger, F., Sessitsch, A. (2009). Incidence and microbiology of salad-borne disease. CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources. 4: 1-13
- Acosta-Martínez V, Dowd SE, Bell CW, Lascano R, Booker JD, Zobeck TM *et al.* (2010). Microbial community composition as affected by dryland cropping systems and tillage in a semiarid sandy soil. *Diversity*. 2: 910–931.
- Acosta-Martinez V, Dowd S, Sun Y, Allen V. (2008). Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. *Soil Biol Biochem* 40: 2762–2770.
- Beuchat, L. R. (1999). Survival of *Escherichia coli* O157:H7 in bovine feces applied to lettuce and ineffectiveness of chlorinated water as a disinfectant. *J. Food Prot.* 62:845-849.
- Brandl, M.T. (2006). Fitness of human enteric pathogens on plants and implications for food safety. *Annu Rev Phytopathol* 44:367-392.
- Brooks JT, Sowers EG, Wells JG, Greene KD, Griffin PM, et al. (2005). Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983-2002. *J Infect Dis* 192: 1422-1429.
- Bettelheim KA (2007). The non-O157 Shiga-toxigenic (verocytotoxigenic) *Escherichia coli*; under-rated pathogens. *Crit Rev Microbiol* 33: 67-87.
- Beutin L, Martin A (2012). Outbreak of Shiga toxin-producing *Escherichia coli* (STEC) O104:H4 infection in Germany causes a paradigm shift with regard to human pathogenicity of STEC strains. *J Food Prot.* 75: 408-418.
- Chan JM, Carlsson G, Rabadan R (2013). Topology of viral evolution. *Proc Natl Acad Sci U S*

A.110:18566-18571.

- Caprioli A, Morabito S, Brugère H, Oswald E (2005). Enterohaemorrhagic *Escherichia coli*: emerging issues on virulence and modes of transmission. *Vet Res* 36: 289-311.
- California Food Emergency Response Team. 2007a. Investigation of an *Escherichia coli* O157:H7 outbreak associated with Dole pre-packaged spinach. California Department of Public Health Food and Drug Branch, Sacramento, CA. [CALFERT]
- California Food Emergency Response Team, 2007b. Environmental Investigation of *Escherichia coli* O157:H7 outbreak associated with Taco Bell restaurants in northeastern states.
- Carlsson G. (2009) Topology and Data. *Cull Amer Math Soc* 46:255, 255-308.
- Cooley, M., Carychao, D., Crawford-Miksza, L., Jay, M. T., Myers, C., Rose, C., Keys, C., Farrar, J., (2007). Incidence and tracking of *Escherichia coli* O157:H7 in a major produce production region in California. *PLoS One* 2 (11), e1159, doi:10.1371
- Doyle, M.P. and Erickson, M.C. (2008). Summer meeting 2007 – the problems with fresh produce: an overview. *J Appl Microbiol.* 105: 317–330. doi: 10.1111/j.1365-2672.2008.03746.x.
- Feng P, Dey M, Abe A, Takeda T. (2001). Isogenic strain of *Escherichia coli* O157:H7 that has lost both Shiga toxin 1 and 2 genes. *Clin Diagn Lab Immunol.* 8: 711-717.
- Fields PI, Blom K, Hughes HJ, Hesel LO, Feng P, Swaminathan B. (1997). Molecular characterization of the gene encoding H antigen in *Escherichia coli* and development of a PCR-restriction fragment length polymorphism test for identification of *E. coli* O157:H7 and O157:NM. *J Clin Microbiol.* 35: 1066-70.

- Franz E, Semenov AV, Termorshuizen AJ, de Vos OJ, Bokhorst JG, van Bruggen AHC. (2008). Manure-amended soil characteristics affecting the survival of *E. coli* O157:H7 in 36 Dutch soils. *Environ Microbiol.* 10: 313–327.
- Franz E, van Hoek AHAM, Bouw E, Aarts HJM. (2011). Variability of *Escherichia coli* O157 Strain Survival in Manure-Amended Soil in Relation to Strain Origin, Virulence Profile, and Carbon Nutrition Profile. *Appl Environ Microbiol.* 77: 8088-8096.
- Fremaux B, Prigent-Combaret C, Delignette-Muller ML, Mallen B, Dothol M, Gleizal A, et al. (2008). Persistence of Shiga toxin-producing *Escherichia coli* O26 in various manure-amended soil types. *J Appl Microbiol.* 104: 296-304.
- Freilich S, et al. (2010). The large-scale organization of the bacterial network of ecological co-occurrence interactions. *Nucleic Acids Res* 38:3857–3868.
- Gontcharova V., Young E., Sun Y., Wolcott R D., Dowd S E. (2010). A comparison of bacterial composition in diabetic ulcers and contralateral intact skin. *The Open Microbiol J.* 4:8-19.
- Greenblum., S, Turnbaugh PJ Borenstein E. (2012). Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *PNAS.* 109:594-599.
- Probst AJ, Lum PY, John B, Dubinsky EA, Piceno YM, Tom LM, Andersen GL, He Z Todd Z. DeSantis TZ. (2014). Microarray of 16S rRNA gene probes for quantifying population differences across microbiome samples: *In* Microarrays: Current technology, innovations and applications. Chapter 5. Editor. He Z. University of Oklahoma, Norman, OK, USA.
- Hillborn ED, Mermin JH, Mshar PA, Hadler JL, Voetsch A, Wojtkunski C, Swartz M, Mshar R, Lambert-Fair MA, Farrar JA, Glynn MK and Slutsker L (1999). A multistate outbreak of *Escherichia coli* O157:H7 infections associated with consumption of mesclun lettuce.

- Arch Internal Med. 159: 1758–1764. Ibekwe AM, Grieve CM, Yang C-H. (2007). Survival of *Escherichia coli* O157:H7 in soil and on lettuce after soil fumigation. *Can J Microbiol.* 53: 623-635.
- Ibekwe AM, Watt PM, Shouse PJ, Grieve MC. (2004). Fate of *Escherichia coli* O157:H7 in irrigation water on soil and plants as validated by culture method and real-time PCR. *Can. J. Microbiol.* 50(12):1007-1014.
- Ibekwe AM, Papiernik SK, Grieve CM, Yang C-H. (2010). Quantification of persistence of *Escherichia coli* O157:H7 in contrasting soils. *Int J Microbiol* doi:10.1155/2011/421379.
- Ibekwe AM, Ma J. (2011). Effects of fumigants on microbial diversity and persistence of *E. coli* O15:H7 in contrasting soil microcosms. *Sci Total Environ.* 409: 3740-3748.
- Ito H, Terai A, Kurazono H, Takeda Y, Nishibuchi M. (1990). Cloning and nucleotide sequencing of Vero toxin 2 variant genes from *Escherichia coli* O91:H21 isolated from a patient with the hemolytic uremic syndrome. *Microb Pathog.* 8: 47-60.
- Jay MT, Cooley M, Carychao D, Wiscomb GW, Sweitzer RA, Crawford-Miksza L, Farrar J A, Lau DK, O'Connell J, Millington A, Asmundson RV, Atwill ER, Mandrell RE. (2007). *Escherichia coli* O157:H7 in feral swine near spinach fields and cattle, Central California Coast. *Emerg Infect Dis.* 13:1908–1911.
- Jiang X, Morgan J, Doyle MP. (2002). Fate of *Escherichia coli* O157:H7 in manure-amended soil. *Appl Environ Microbiol.* 68: 2605–2609.
- Karch H, Tarr PI, Bielaszewska M. 2005. Enterohaemorrhagic *Escherichia coli* in human medicine. *Int J Medical Microbiol.* 295: 405-18.
- Klitgord N, Segrè D (2010) Environments that induce synthetic microbial ecosystems. *PLOS Comput Biol* 6:e1001002.

- Louie M, Read S, Simor AE, Holland J, Louie L et al. (1998). Application of multiplex PCR for detection of non-O157 verocytotoxin-producing *Escherichia coli* in bloody stools: Identification of serogroups O26 and O111. *J Clin Microbiol.* 36: 3375-3377.
- Lum PY, Singh G, Lehman A, Ishkanov T, Vejdemo-Johansson M, Alagappan M, Carlsson J, Carlsson G. (2013). Extracting insights from the shape of complex data using topology. *Scientific Reports.*3:1236 doi:10.1038/srep01236.
- Lynch MF, Tauxe R V, Hedberg CW. (2009). The growing burden of foodborne outbreaks due to contaminated fresh produce: risks and opportunities. *Epidemiol Infect.* 137: 307-15.
- Ma J, Ibekwe AM, Yi X, Wang H, Yamazaki A et al. (2011) Persistence of *Escherichia coli* O157:H7 and Its Mutants in Soils. *PLoS ONE* 6: e23191.
- Ma J, Ibekwe AM, Crowley DE, Yang C-H. (2012). Survival of *Escherichia coli* O157:H7 in major leafy green producing soils. *Environ Sci Tech.* 46: 12154–12161.
- Ma J, Ibekwe AM, Yang C-H, Crowley DE. (2013). Influence of bacterial communities based on 454 pyrosequencing on the survival of *Escherichia coli* O157:H7 in soils. *FEMS Microbiol Ecol.* 84:542-554.
- Ma J, Ibekwe AM, Crowley DE, Yang C-H. (2014). Persistence of *Escherichia coli* O157 and non-O157 strains in agricultural soils. *Sci Total Environ.* 490: 822–829.
- Mathusa EC, Chen Y, Enache E, Hontz L. (2010). Non-O157 Shiga toxin-producing *Escherichia coli* in foods. *J Food Prot* 73: 1721-1736.
- Maule A. (2000). Survival of verocytotoxigenic *Escherichia coli* O157 in soil, water and on surfaces. *Symp Ser Soc Appl Microbiol* 29: 71S-78S.

- Mubiru DN, Coyne MS, Grove JH. (2000). Mortality of *Escherichia coli* O157:H7 in two soils with different physical and chemical properties. *J Environ Qual* 29: 1821–1825.
- Nicolau M, Levine AJ, Carlsson G. (2011). Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proc Natl Acad Sci USA* 108: 7265–7270.
- Nonnenmann M W, Bextine B, Dowd SE, Gilmore K, Levin JL. (2010). Culture-independent characterization of bacteria and fungi in a poultry bioaerosol using pyrosequencing: A new approach. *J Occup Environ Hyg* 7: 693–399.
- Oberhardt MA, Palsson BØ, Papin JA. (2009) Applications of genome-scale metabolic reconstructions. *Mol Syst Biol* 5:320.
- Parter M, Kashtan N, Alon U. (2007). Environmental variability and modularity of bacterial metabolic networks. *BMC Evol Biol* 7:169.
- Perna NT, Plunkett G, 3rd, Burland V, Mau B, Glasner JD, Rose DJ, et al. (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* . 409: 529-33.
- Romano D, Nicolau M, Quintin EM, Mazaika PK, Lightbody AA, Cody Hazlett H, Piven J, Carlsson G, Reiss AL. (2014). Topological methods reveal high and low functioning neuro-phenotypes within fragile X syndrome. *Hum Brain Mapp* doi: 10.1002/hbm.22521.
- Sarikonda G, Pettus J, Phatak S, Sachithanatham S, Miller JF, Wesley JD, Cadag E, Chae J, Ganesan L, Mallios R, Edelman S, Peters B, von Herrath M. (2013). CD8 T-cell reactivity to islet antigens is unique to type 1 while CD4 T-cell reactivity exists in both type 1 and type 2 diabetes. *J Autoimmun.* pii: S0896-8411(13)00150-9. doi: 10.1016/j.jaut.2013.12.003.
- Semenov AV, Franz E, van Overbeek L, Termorshuizen AJ, van Bruggen AHC. (2008).

- Estimating the stability of *Escherichia coli* O157:H7 survival in manure amended soils with different management histories. *Environ Microbiol* 10: 1450–1459.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.
- Sivapalasingam S, Friedman CR, Cohen L, Tauxe RV. (2004). Fresh produce: a growing cause of outbreaks of foodborne illness in the United States, 1973 through 1997. *J. Food Protect.* 67:2342–2353.
- Stolyar S, et al. (2007). Metabolic modeling of a mutualistic microbial community. *Mol Syst Biol* 3:92.
- van Elsas JD, Semenov AV, Costa R, Trevors JT. (2011). Survival of *Escherichia coli* in the environment: fundamental and public health aspects. *ISME J* **5**: 173–183.
- Wintermute EH, Silver PA. (2010). Emergent cooperation in microbial metabolism. *Mol Syst Biol.* 6:407.

Table 1: *E. coli* O157 and non-O157 strains used for the study

Strain*	Source	<i>stx</i> ₁	<i>stx</i> ₂	<i>eae</i>	<i>hlyA</i>	Reference
<i>E. coli</i> O26:H11	cow, Ontario, Canada	-	+	+	+	(Louie et al., 1998)
<i>E. coli</i> O103:H2	cow, Ontario, Canada	+	+	+	+	(Louie et al., 1998)
<i>E. coli</i> O91:H21	Human, OH, USA	+	-	-	+	(Ito et al., 1990)
<i>E. coli</i> O157 NM	-, AL, USA	+	+	+	+	(Fields et al., 1997)
<i>E. coli</i> O157:H7 4554	cow, Japan	-	-	+	+	(Feng et al., 2001)
<i>E. coli</i> O157:H7 EDL933	human, USA	+	+	+	+	(Perna et al., 2001)

*stx*₁, a gene coding for Shiga toxin1, *stx*₂, a gene coding for Shiga toxin2, *eae*, a gene coding for intimin, and *hlyA*, a gene coding for hemolysin. "+" and "-" indicate a gene was identified and not identified in a given *E. coli* strain, respectively. *E. coli* O157:H7 4554 is therefore *stx* negative.

-, indicates the source was not identified.

*Adapted from Ma et al., 2014.

Table 2: Kolmogorov-Smirnov test and t-test to identify soil and biological properties that best differentiate between Salinas Valley and Imperial/Yuma Valley locations.

Column Name*	Signed KS-score	KS-score	t-test p-value
clay	-0.8571	0.8571	0.0012
Na+	-1	1	1.05E-07
Fe	0.7321	0.7321	0.0299
ttd(d)	0.8125	0.8125	8.18E-05
K+	-0.5446	0.5446	0.0848
EC	-1	1	9.07E-06
<i>Nitrospira</i>	0.75	0.75	0.0014
Cu	-0.3660	0.3660	0.0607
tdd(d)_O1574554	-0.6666	0.6666	0.4823
Diversity index (<i>H'</i>)	0.7321	0.7321	0.0026
Molybdenum	-0.8125	0.8125	3.62E-04
<i>Proteobacteria</i>	-0.8125	0.8125	1.63E-04
WSOC	0.6696	0.6696	0.0259
T-N	0.375	0.375	0.1693
Ca++	-0.875	0.875	5.08E-04
Ni	0.4375	0.4375	0.1660
<i>Alphaproteobacteria</i>	0.6875	0.6875	0.0047
OC	-0.6517	0.6517	0.2614
MBC	-0.3839	0.3839	0.3520
SO4--	-0.9375	0.9375	4.55E-05
pH	-0.5714	0.5714	0.0259
<i>Chloroflexi</i>	0.5446	0.5446	0.03441
tdd(d)_O157NM	0.6666	0.6666	0.0995
sand	0.7321	0.7321	0.0027
Bulk density	-0.5446	0.5446	0.0162
<i>Bacteroidetes</i>	0.7321	0.7321	0.0042
WHC	-0.5267	0.5267	0.2032
tdd(d)_O91	0.875	0.875	4.12E-06
<i>Acidobacteria</i>	1	1	0.0197
<i>Actinobacteria</i>	0.6071	0.6071	0.0240
Mg++	-0.7142	0.7142	0.0012
tdd(d)_O26	1	1	0.0796
<i>Gemmatimonadetes</i>	0.75	0.75	0.0012
silt	-0.4821	0.4821	0.0403
<i>Firmicutes</i>	0.4821	0.4821	0.1989
<i>Verrucomicrobia</i>	0.4375	0.4375	0.3981
<i>Deltaproteobacteria</i>	0.9375	0.9375	8.55E-06
<i>Gammaproteobacteria</i>	-0.8125	0.8125	6.92E-05
Zn	-0.5892	0.5892	0.0056

<i>Planctomycetes</i>	0.8571	0.8571	0.0044
PO4---	1	1	0.0399
tdd(d)_O103	0.875	0.875	2.84E-06
<i>Betaproteobacteria</i>	0.75	0.75	1.95E-05
OTUs	-0.75	0.75	0.0015
location	0.875	0.875	3.04E-06
Management	-0.258	0.258	0.4128

*Meaning of abbreviations under column names: Sodium (Na), iron (Fe), time till detection for *E. coli* O157:H7 strain 933 [ttd(d)], potassium (K), electrical conductivity (EC), copper (Cu), time till detection for *E. coli* O157:H7 strain 4555 [ttd (d) O157-4554], assimilable organic carbon (AOC), total nitrogen (TN), calcium (Ca), Nickel (N), organic carbon (OC), microbial biomass carbon (MBC), sulfate (SO₄), time till detection for *E. coli* O157:H7 non-motile strain 4555 [ttd (d) O157NM], water holding capacity (WHC), time till detection for *E. coli* O91 [ttd (d) O91], magnesium (Mg), time till detection for *E. coli* O26 [ttd (d) O26], zinc (Zn), phosphate (PO₄), and operation taxonomic units (OTUs). Signed KS score: the minus sign indicates that the attribute indicated in the column name is on average smaller in value in Salinas Valley compared to Imperial/Yuma Valley locations.

Legend

Figure 1A. Sample-sample relationships in a topological network. Using physical, chemical, and biological characteristics of the samples, we obtained a network that comprised of 4 sub-networks (A, B, C and D) and a singleton (single node comprising of 2 samples). The coloring here is by location, where each location is given a color (Salinas is red, Imperial Valley is green and Yuma is blue). Nodes that have a mixture of soils have colors in between as depicted in the color bar. Sub-networks structure indicates that physical, chemical, and biological characteristics primarily segregate the samples according to location, with Salinas being most different from soils from Yuma and Imperial Valley, Figure 1B. The same network is colored by soil management type (organic (represented by 0) versus conventional (represented by 1)). The red nodes represent samples with conventional soil management while the blue nodes represented the organic soil management. The green and orange colored nodes represented mixed organic and conventional soil management with varying percent of mixture of the two types of management. Figure 1C: Another network is built using the same parameters except for resolution. The soil samples are analyzed at a lower resolution to ask if structure D and the singleton will merge with any part of the sub-networks. Sub-network D, which comprised of samples from Yuma and Imperial, became part of sub-network B (colored nodes). Samples from sub-network D are not part of the grey nodes. The singleton however remained a singleton. The size of each node reflects the number of data points contained in the node. For 1A and 1B, the distance metric and filters were Person correlation and Principal Metric SVD and secondary metric SVD. Metric: Norm Correlation; Lens: Principal Metric SVD value (Resolution 30, Gain 4.0x, Equalized) Secondary Metric SVD Value (Resolution 30, Gain 4.0x, Equalized). For 1C, all analysis parameters remained the same except for resolution (20 instead of 30).

Figure 2: Network colored by soil physical property concentrations- (A) sand, (B) clay, (C) silt, (D) pH, (E) bulk density. The legend below the figures shows ranges in concentrations of some of the soil physical properties. The distance metric and filters are as shown in the legend of Figure 1.

Figure 3. Topological network data analysis of survival of *E. coli* O157:H7 and non O157:H7 across the sub-networks identified in Fig 1A. Survival of *E. coli* O157:H7 EDL933, *E. coli* O157:NM, and *E. coli* O157 strain 4554 (stx-) and shown in Figures 3 A, B, &C. Survival of non-O157 strains, *E. coli* O26:H21, *E. coli* O103:H2, and *E. coli* O91:H21 are shown in figure 3 D, E, G. Grey nodes represent missing ttd (d) measurements for *E. coli* O157:NM, *E. coli* O157

strain 4554, *E. coli* O26:H21, *E. coli* O103:H2, and *E. coli* O91:H21. The distance metric and filters are as shown in Figure 1.

Figure 4: Survival of six *E. coli* O157 and non-O157 strains grouped together on a normalized scale showed that *E. coli* O103:H2 survived the longest in all the soils, followed by *E. coli* O26:H21. Grey nodes are as explained in Figure 3. The distance metric and filters are as shown in Figure 1.

Figure 5. The abundance and distribution of different bacterial phyla based on pyrosequencing and their correlation with survival. A: soils from Salinas Valley region, sub-networks B and C: Yuma and the Imperial region. Figure 5A-G represent the coloring of the abundance of *Actinobacteria*, *Acidobacteria*, *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, *Alpha-*, *Beta-*, *Delta-*, and *Gammaproteobacteria*, respectively. The distance metric and filters are as shown in Figure 1.

Supporting information

Topological Data Analysis of *Escherichia coli* O157:H7 and Non-O157 Survival in Soils

A. Mark Ibekwe^{1,*}, Jincai Ma^{1,2}, David E. Crowley², Ching-Hong Yang³, Alexis M. Johnson⁴, Tanya C. Petrossian⁴, Pek Yee Lum⁴

1 USDA-ARS U. S. Salinity Laboratory, Riverside, CA 92507

2 Department of Environmental Sciences, University of California, Riverside, CA 92521

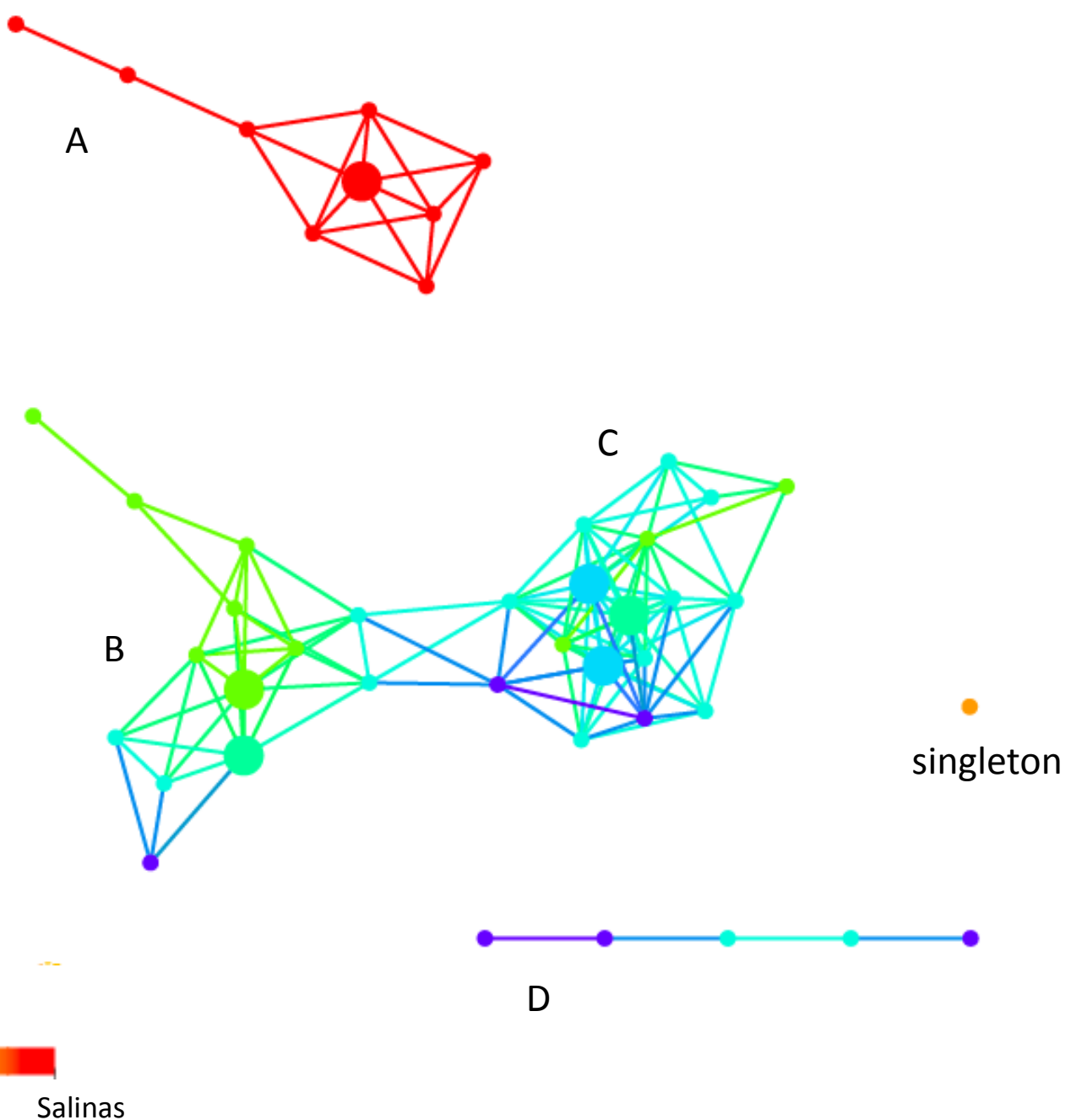
3 Department of Biological Sciences, University of Wisconsin, Milwaukee, WI 53211

4Ayasdi, Inc., Menlo Park, CA

Fig. S1. Abundance of dominated phyla revealed by 454-pyrosequencing targeting 16s rRNA genes. Different letters denotes significant differences ($P = 0.05$) among the three soils.

Fig. S2: Factorial biplot by the principal component 1 and 2 (PC1, 41.6%; PC2 26.0%) resulting from the principle component analysis (PCA) performed on soil properties and survival (*ttf*) of *E. coli* O157:EDL933 in soils. Salinity (EC, dS m⁻¹), water holding capacity (WHC, %), silt content (%), clay content (%), total nitrogen (TN, %); OC, organic carbon (%); AOC, assimilable organic carbon in soil water extract (mg kg⁻¹); MBC, microbial biomass carbon (mg kg⁻¹).

Fig 1A



Yuma Imperial Salinas

Fig 1B

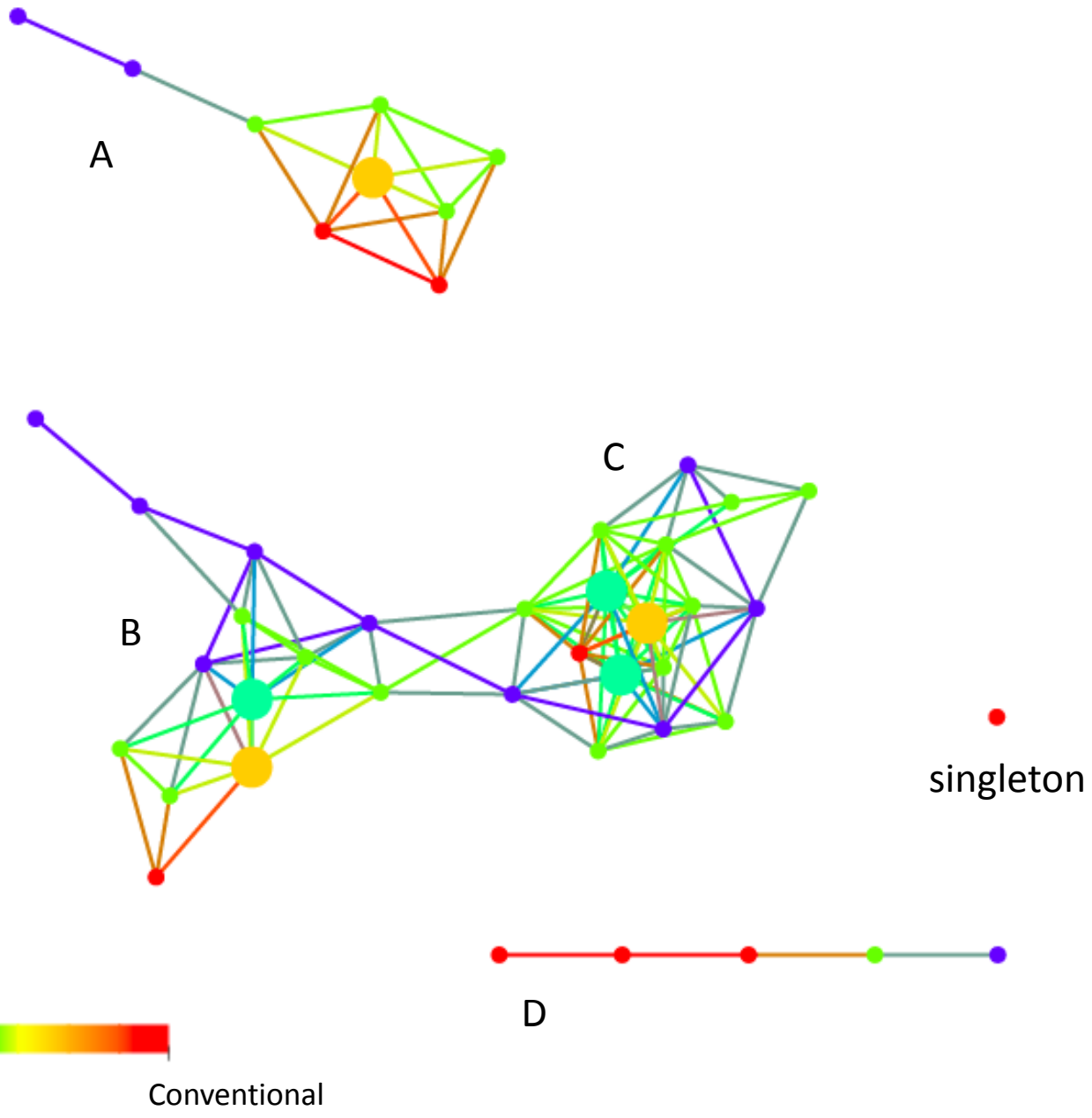


Fig 1C

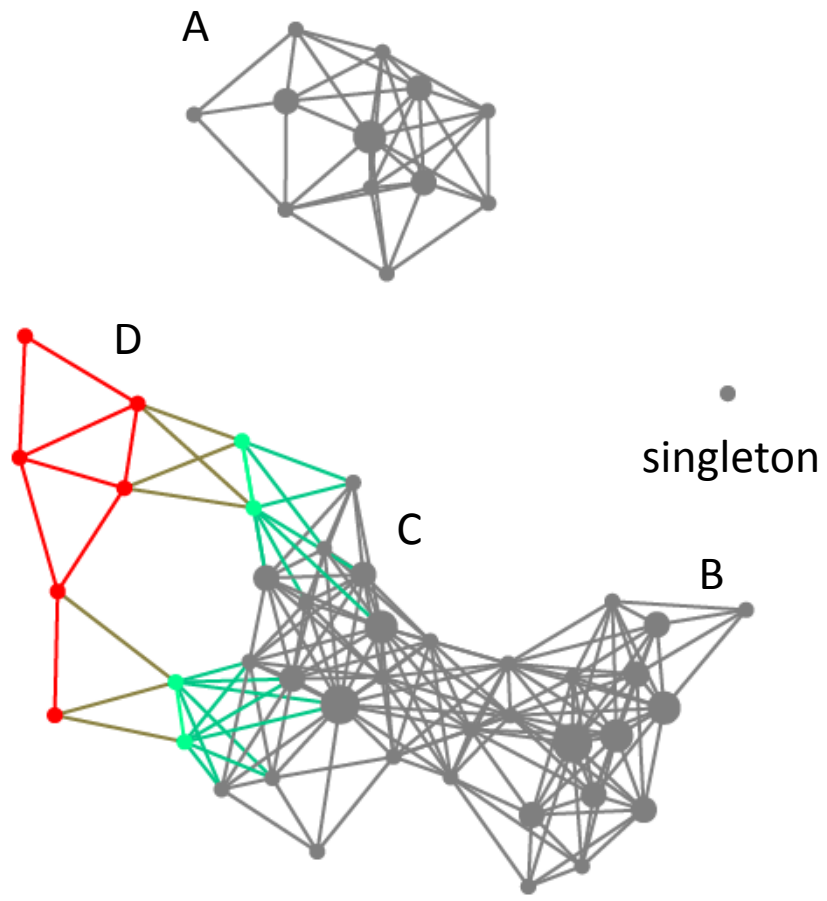
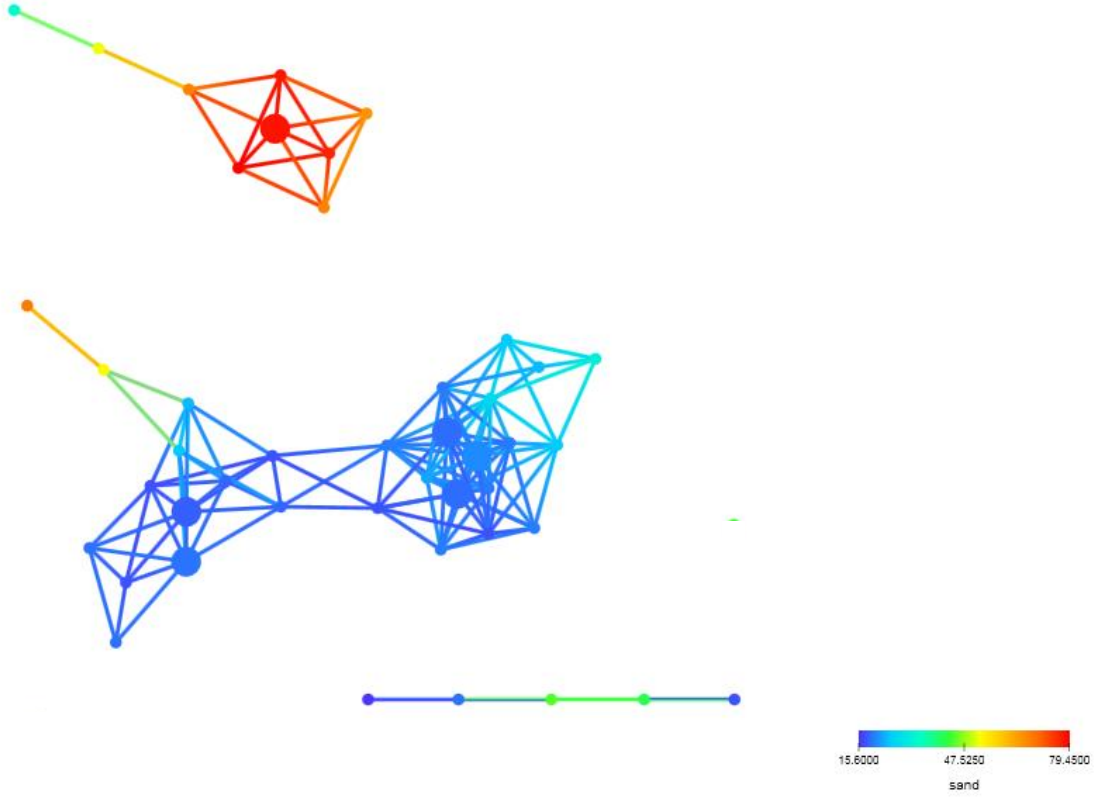
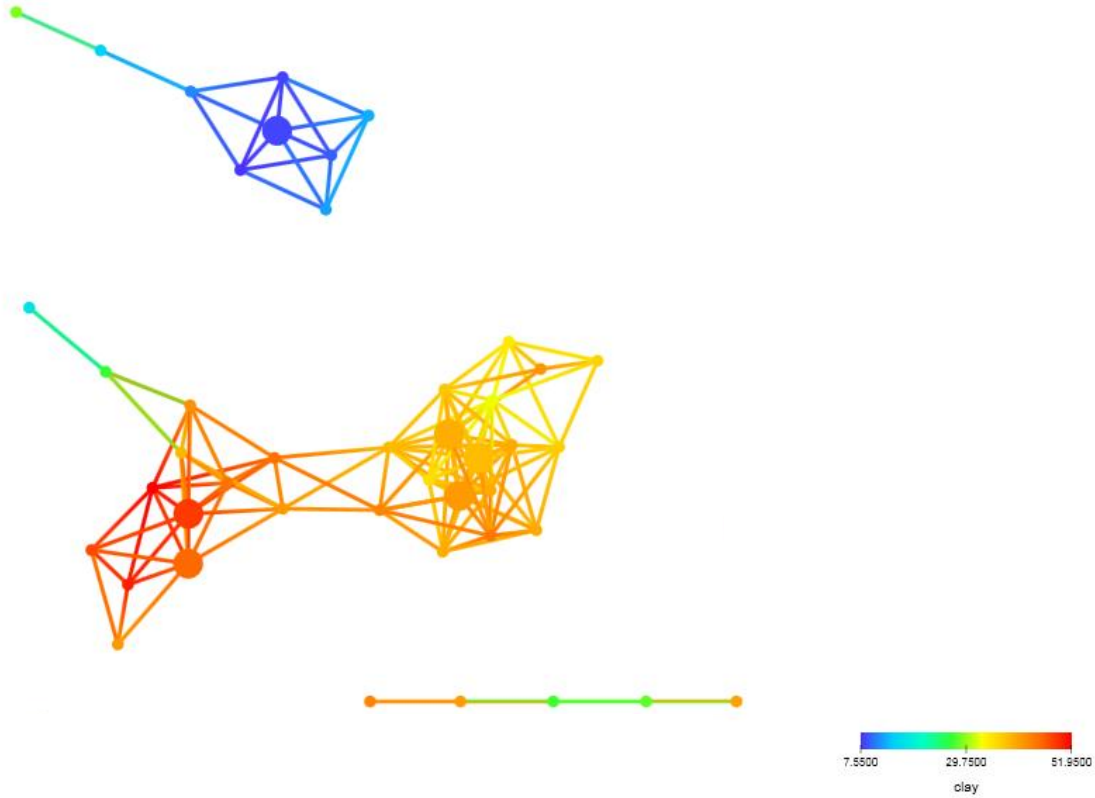


Fig. 2A



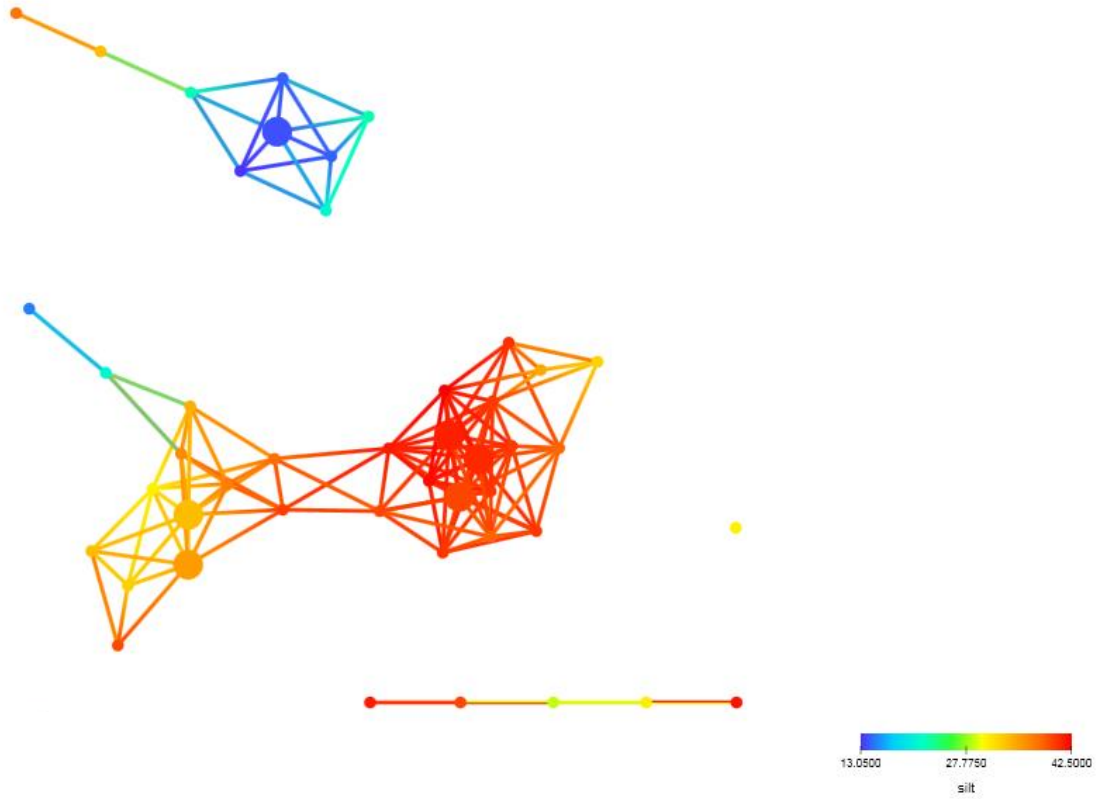
Colored by sand composition

Fig.2B clay



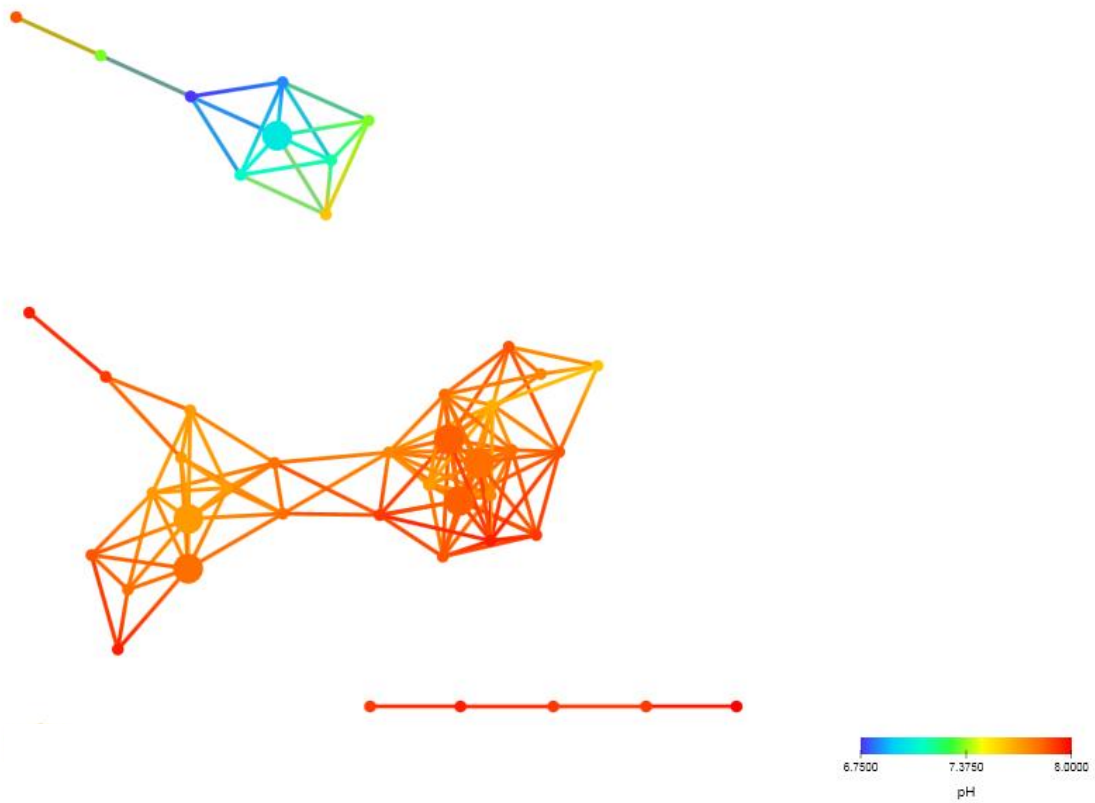
Colored by clay composition

Fig. 2c silt



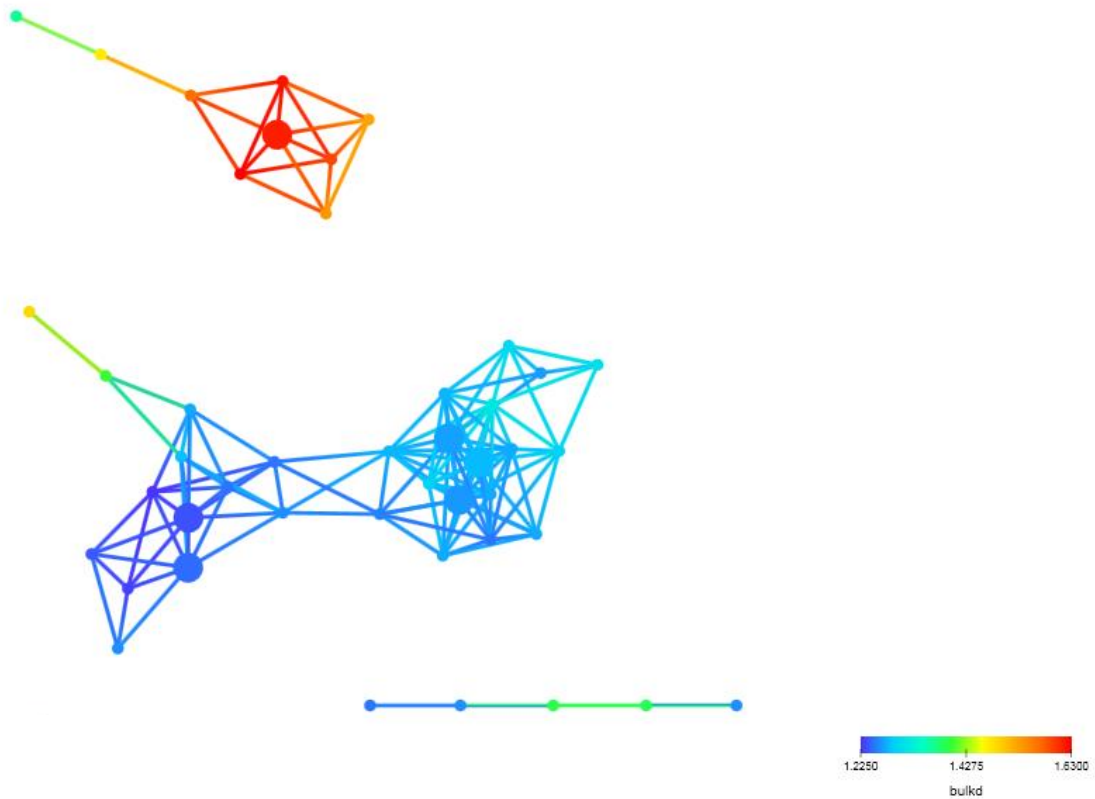
Colored by silt composition

Fig. 2 D pH



Colored by pH

Fig.2E bulk density



Colored by bulk density

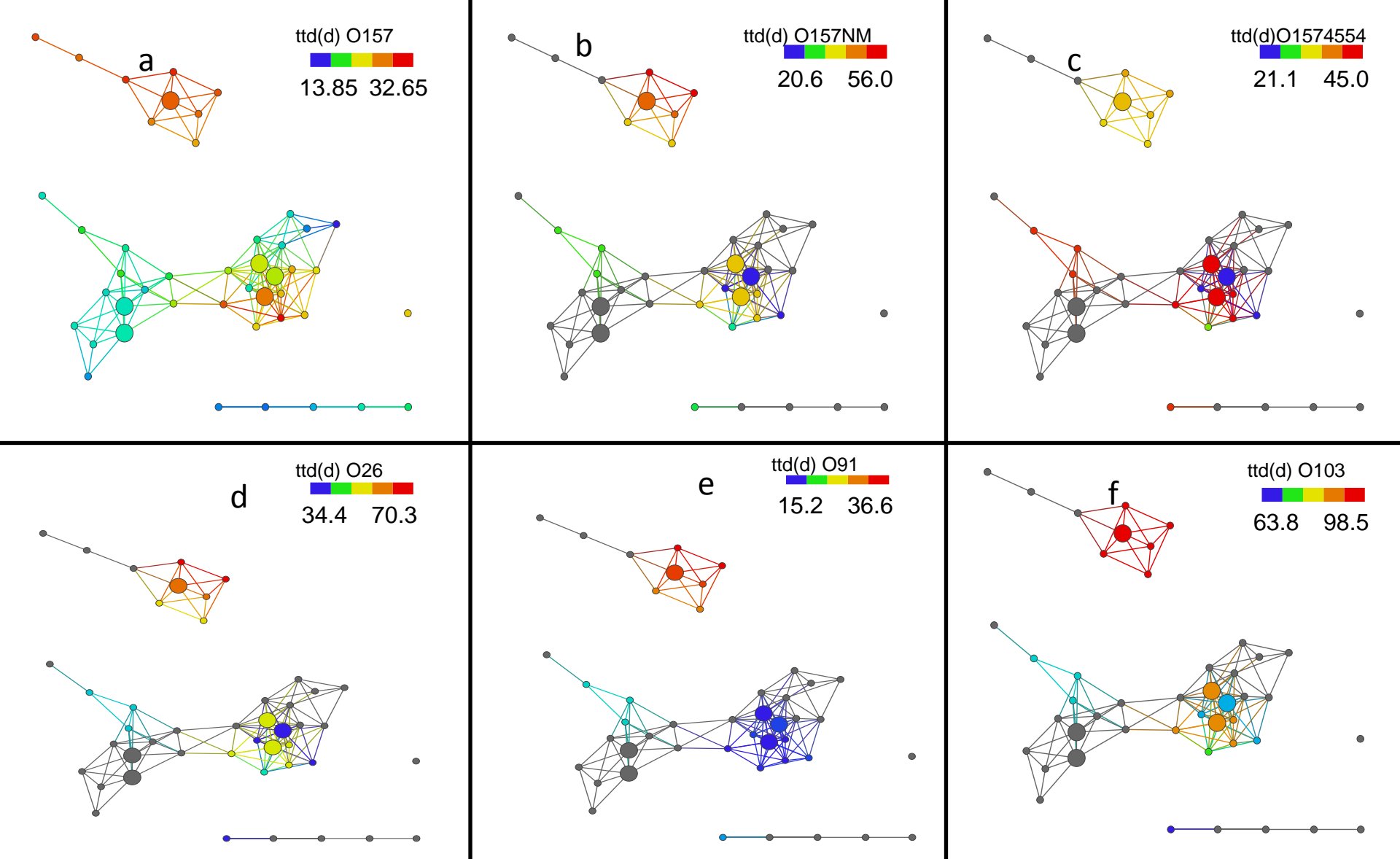


Figure 3: Survival patterns

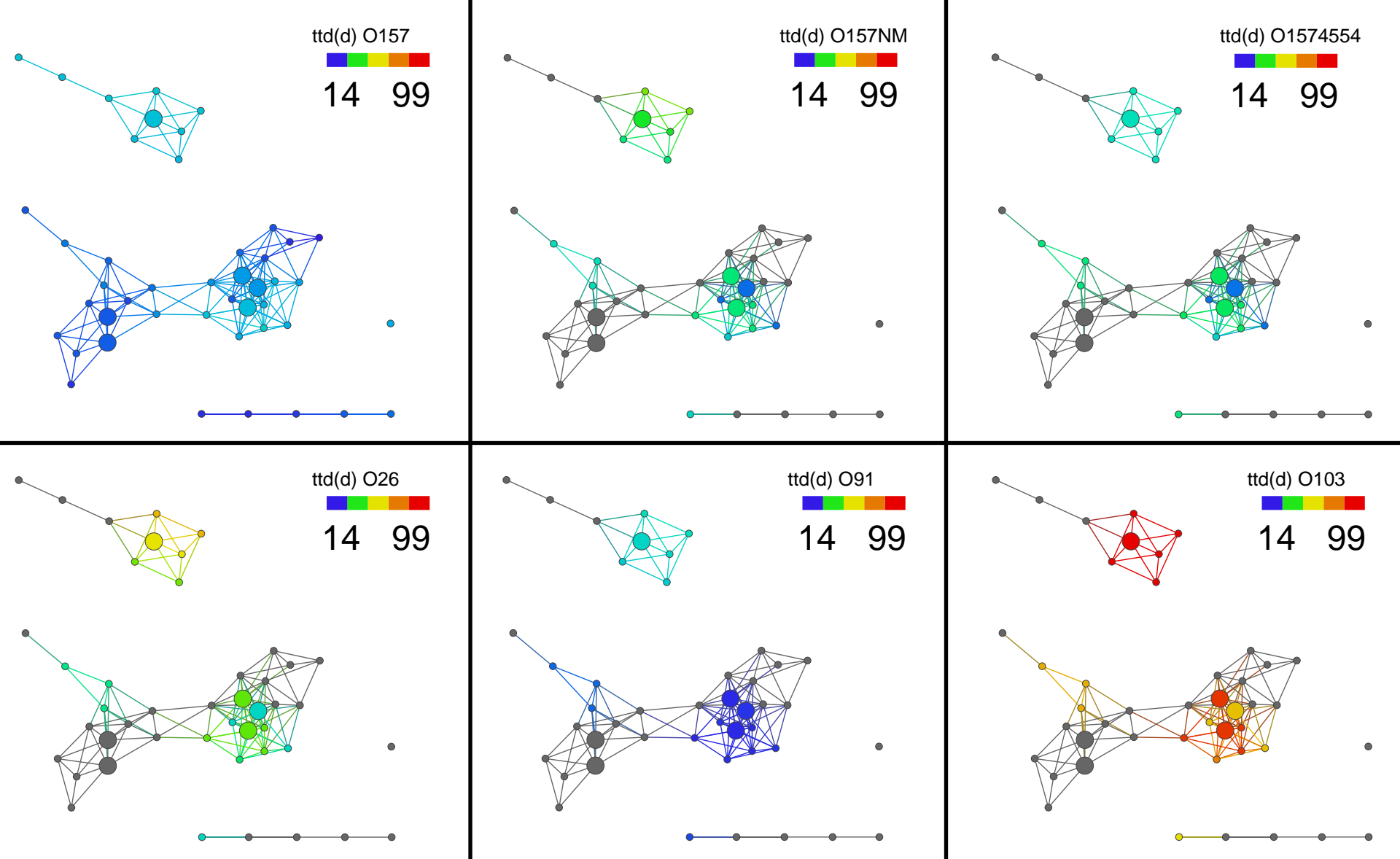
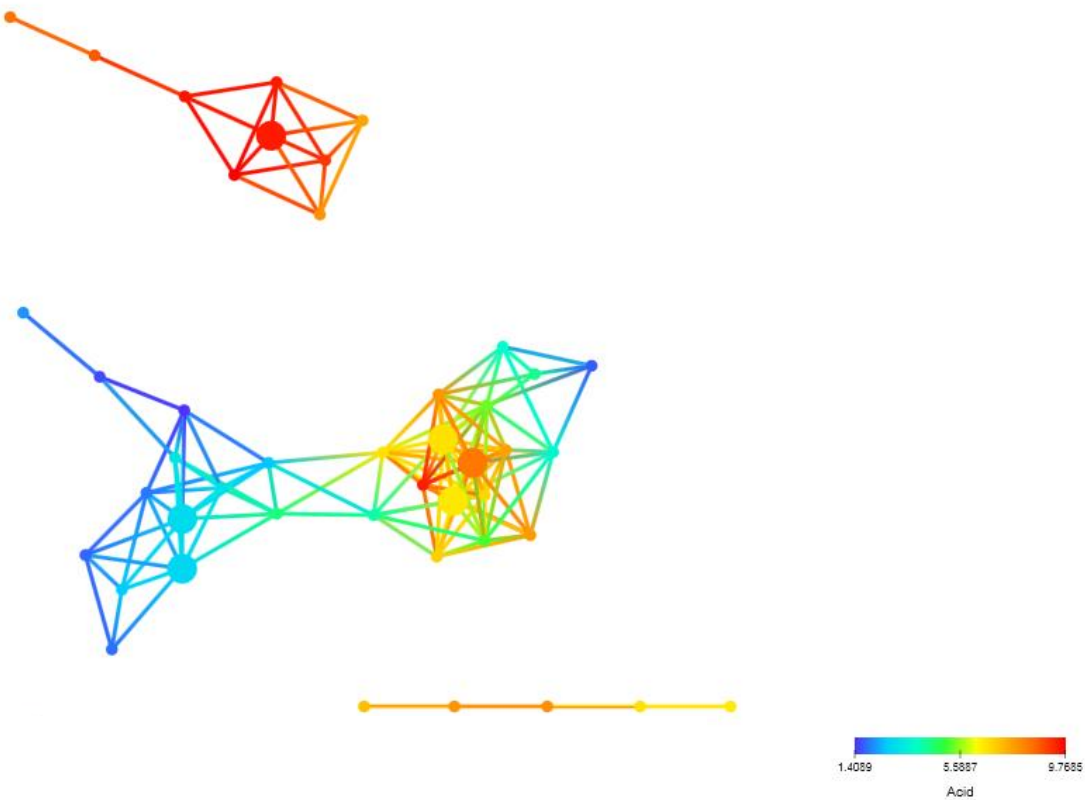


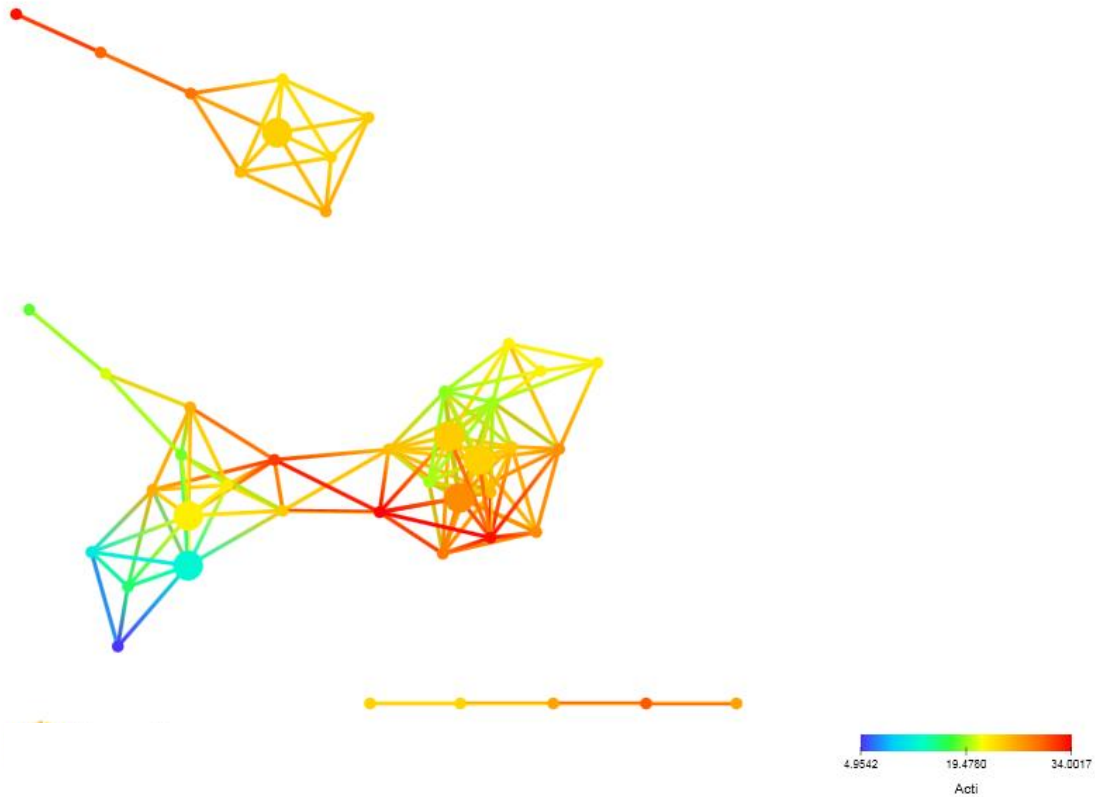
Fig. 4; survival patterns

Fig. 5A: *Acidobacteria*



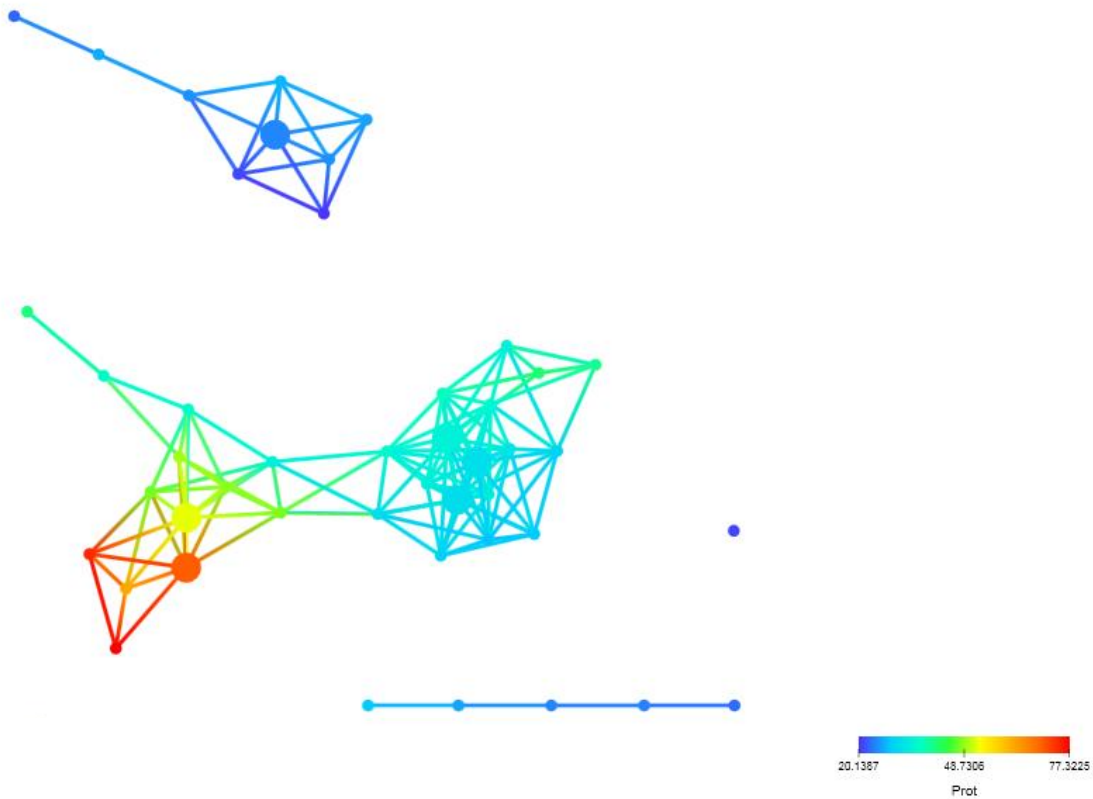
Colored by *Acidobacteria* composition

FIG. 5B: *Actinobacteria*



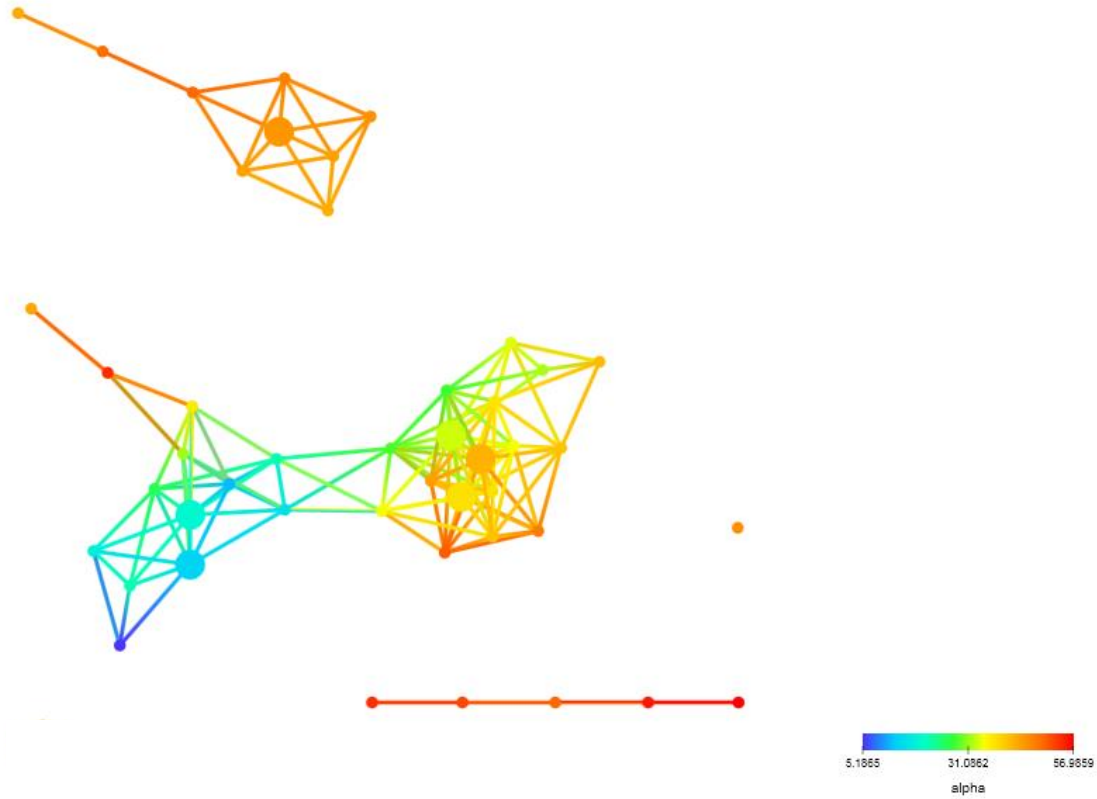
Colored by *Actinobacteria* composition

FIG. 5C: *Proteobacteria*



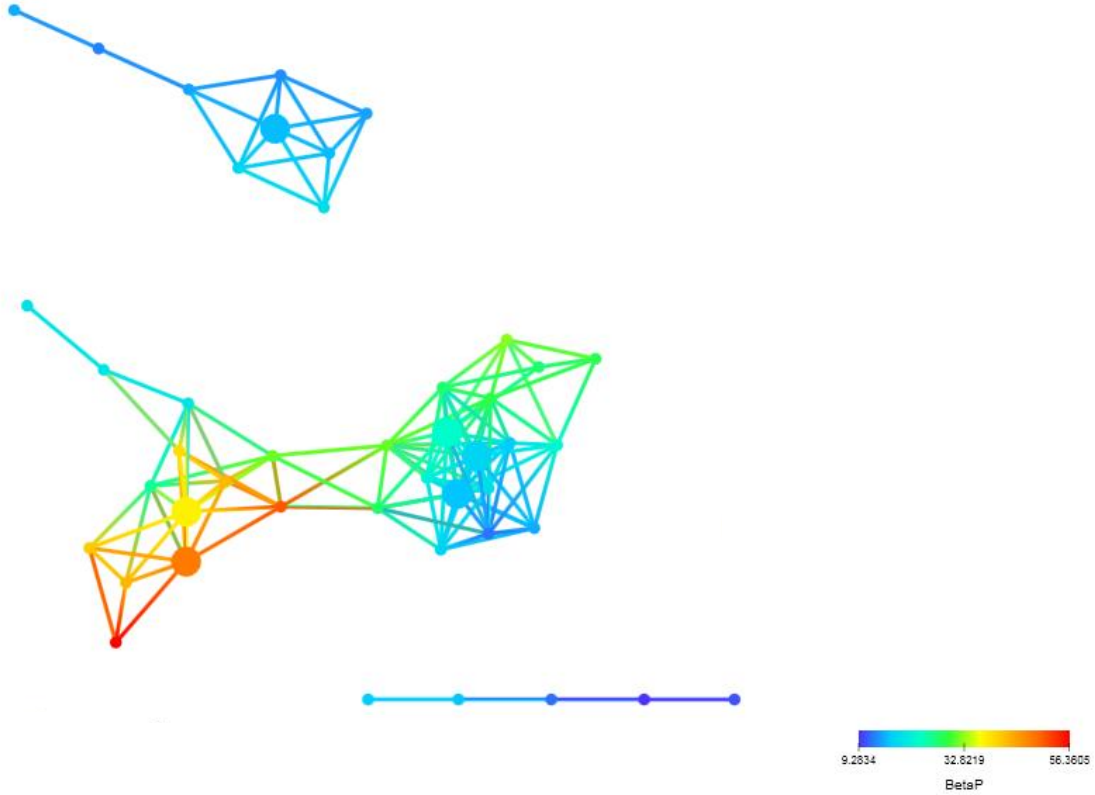
Colored by *Proteobacteria* composition

FIG. 5D: *Alphaproteobacteria*



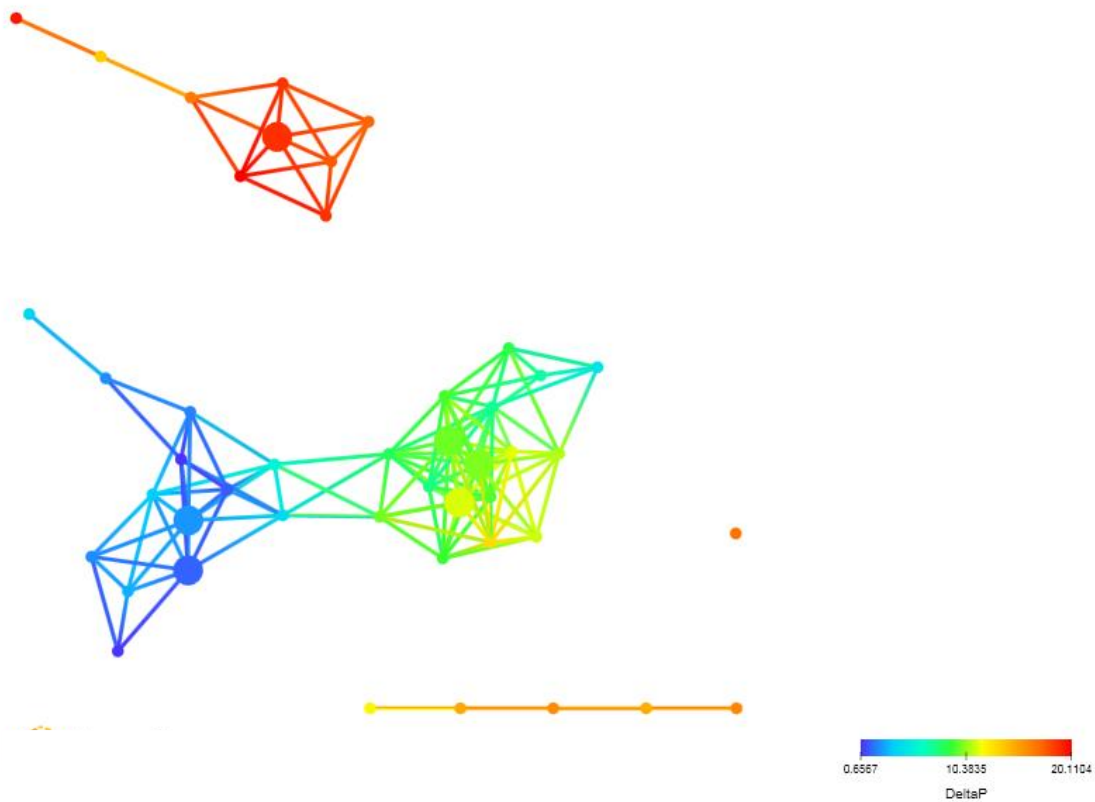
Colored by *Alphaproteobacteria*
composition

FIG.5E: *Betaproteobacteria*



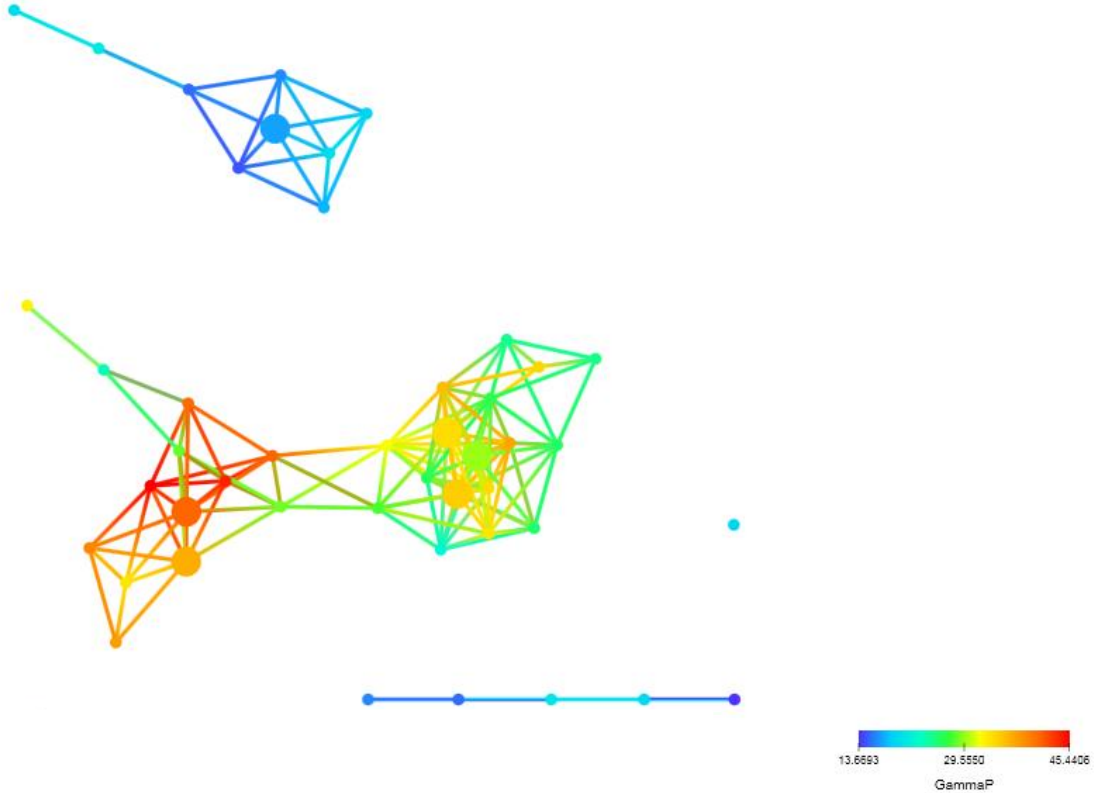
Colored by *Betaproteobacteria*
composition

FIG.5F: *Deltaproteobacteria*



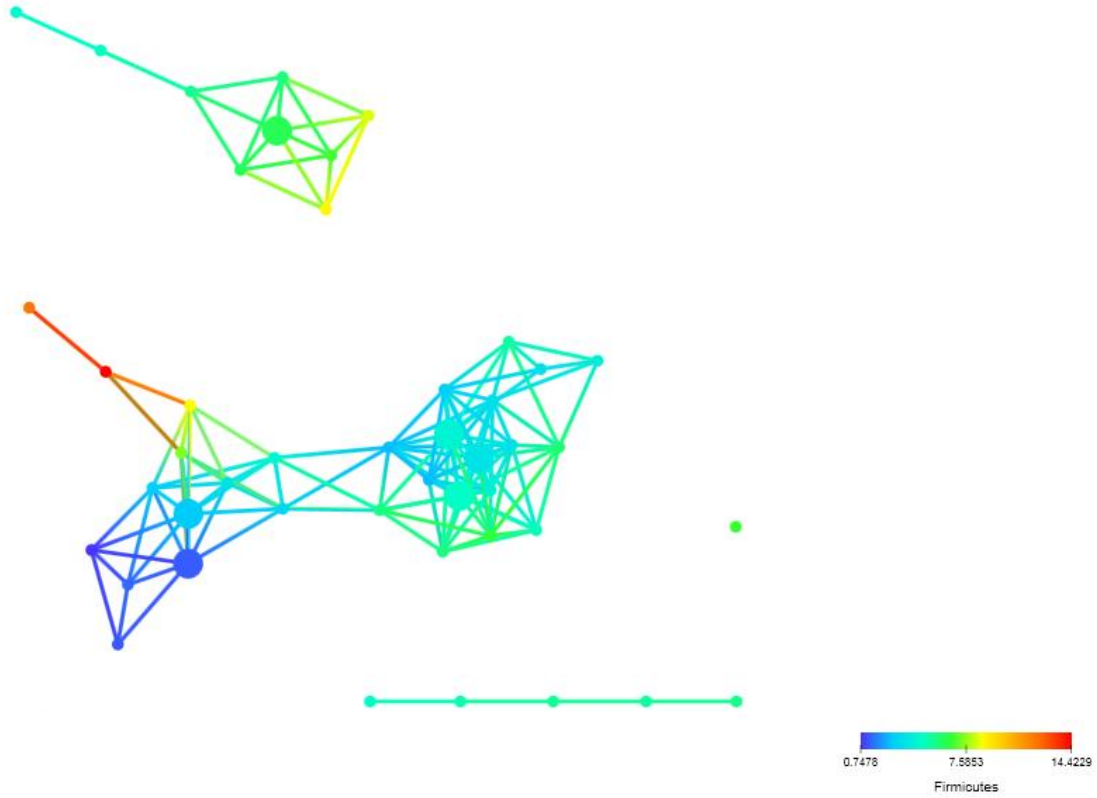
Colored by *Deltaproteobacteria*
composition

FIG.5G: *Gammaproteobacteria*



Colored by *Gammaproteobacteria* composition

FIG 5H: *Firmicutes*



Colored by *Firmicutes*
composition

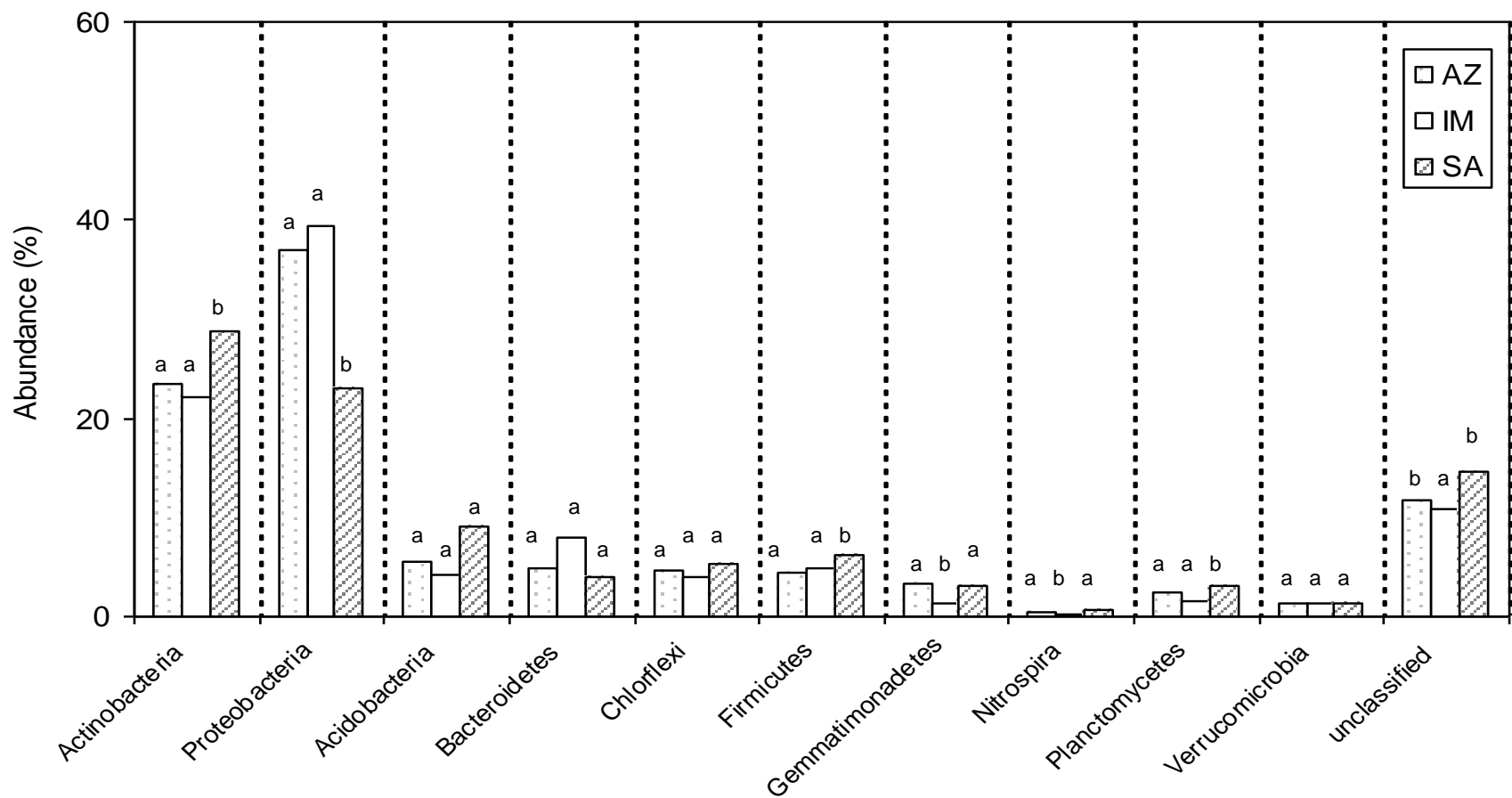


Fig. S1

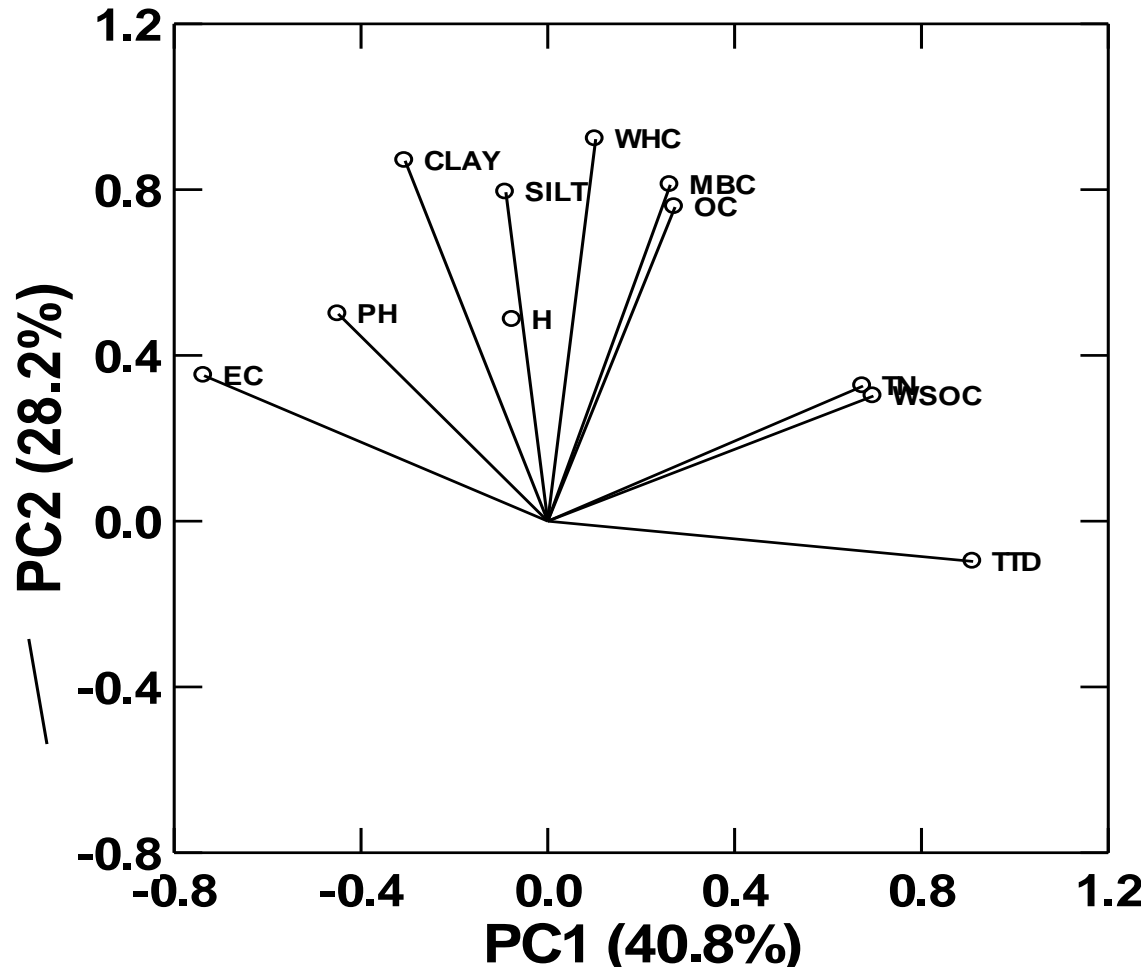


Fig. S2

