Canadian Journal of Microbiology

Physiological traits and relative abundance of species as explanatory variables of co-occurrence pattern of cultivable bacteria associated with chia seeds

Journal:	<i>Canadian Journal of Microbiology</i>
Manuscript ID	cjm-2019-0052.R1
Manuscript Type:	Article
Date Submitted by the Author:	14-May-2019
Complete List of Authors:	Jaba, Asma; INRS, Institut Armand-Frappier Dagher, Fadi; Agri-Neo Inc Hamidi Oskouei, Amir Mehdi; Agri-Neo Inc Guertin, Claude; INRS, Institut Armand-Frappier Constant, Philippe; Institut national de la recherche scientifique, Institut Armand-Frappier
Keyword:	Microbial ecology, Microbiome, Microbial communities
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)

SCHOLARONE™
Manuscripts

Article to submit to “Canadian Journal of Microbiology”

Title: Physiological traits and relative abundance of species as explanatory variables of co-occurrence pattern of cultivable bacteria associated with chia seeds

Authors: Asma Jaba^a, Fadi Dagher^b, Amir Mehdi Hamidi Oskouei^b, Claude Guertin^{a*}, Philippe Constant^{a*}

^aInstitut National de la Recherche Scientifique-Institut Armand-Frappier, 531 boulevard des Prairies, Laval (Québec), Canada, H7V 1B7.

^bAgri-Neo Inc., 435 Horner Avenue, Unit 1, Toronto, Ontario Canada, M8W 4W3.

*Corresponding authors: INRS-Institut Armand-Frappier, 531 boulevard des Prairies, Laval (Québec), Canada, H7V 1B7, Phone: (450) 687-5010, FAX: (450) 686-5566, Email addresses: Claude.Guertin@iaf.inrs.ca, Philippe.Constant@iaf.inrs.ca

Abstract

Deciphering the rules defining microbial community assemblage is envisioned as a promising strategy to improve predictions of pathogens colonization and proliferation in food. Despite the increasing number of studies reporting microbial co-occurrence patterns, only a few attempts were made to challenge them in experimental or theoretical frameworks. Here, we tested the hypothesis that observed variations in co-occurrence patterns can be explained by taxonomy, relative abundance and physiological traits of microbial species. PCR amplicon sequencing of taxonomic markers was first conducted to assess distribution and co-occurrence patterns of bacterial and fungal species found in 25 chia (*Salvia hispanica* L.) samples originating from eight different sources. The use of nutrient-rich and oligotrophic media enabled isolation of 71 strains encompassing 16 bacterial species, of which five corresponded to phylotypes represented in the molecular survey. Tolerance to different growth inhibitors and antibiotics was tested to assess physiological traits of these isolates. Divergence of physiological traits and relative abundance of each pair of species explained 69% of the co-occurrence profile displayed by cultivable bacterial phylotypes in chia. Validation of this ecological network conceptualization approach to more food products is required to integrate microbial species co-occurrence patterns in predictive microbiology.

Keywords

Microbial ecology, microbiome, microbial communities.

1. Introduction

Foodborne pathogens or contaminants may arise following accidental exposure and infection of food during transport or storage. Persistence and proliferation of these allochthonous microorganisms are then driven by abiotic and biotic features characterizing food matrix. The relevance of abiotic features of food products in determining the structure of their microbiota is demonstrated by predictive microbiology approaches implementing sophisticated models incorporating environmental factors to predict the proliferation of microorganisms (Mejlholm et al. 2010; Tenenhaus-Aziza and Ellouze 2015; Zoellner et al. 2018). On the other hand, the impact of biotic features including microbe-microbe interactions has received less attention, even though experimental evidence supports the notion that species co-occurrence patterns are non-random (Berg 2015). Pioneering work was the application of the checkerboard principle showing that species co-occurrence patterns observed in plants and animals hold in microbial communities (Horner-Devine et al. 2007). The reliance of the checkerboard principle on presence/absence datasets complicates the implementation of the approach for microbial community profiles obtained by high-throughput sequencing technologies. Indeed, these portraits are compositional matrices represented by a finite number of sequences retrieved from environmental DNA resulting in inverse relationships between the relative abundance and the variance of sequence features and the occurrence of several challenging to handle zero values (Kaul et al. 2017). These specificities regarding microbial community data assembly were considered in the implementation of several algorithms designed to compute species co-occurrence relying on their relative abundance with positive (co-presence) and negative (co-exclusion) relationships (Friedman and Alm 2012; Kurtz et al. 2015). However, such potential microbe-microbe interactions are more suggestive than definitive owing to neutral processes

selecting coexisting species from their fitness in the habitat, without relying on interactions *per se* (Bell 2005).

Despite the constraints imposed by molecular tools, implementation of microbial species correlation matrix into theoretical frameworks is expected to identify general principles or rules deciphering assembly and structure of microbial communities. An emerging strategy is to relate pairwise correlation score of microbial species to their physiological traits selected by habitat constraints. Indeed, incongruence between phylogenetic distance and physiological traits suggest the latter are the most significant parameter to infer microbial species interactions with their environment (Ho et al. 2013; Krause et al. 2014). Under this framework, the co-occurrence of microbial species can be computed from molecular profiles, and the resulting pairwise correlation coefficients can be related to physiological traits of individual species expressed either as whole genome sequence similarity (Kamneva 2017), predicted functions (Mandakovic et al. 2018; Zelezniak et al. 2015) or approaches relying on phenotype characterization of cultivable members of the communities, as proposed here.

In recent years, chia (*Salvia hispanica* L.) has acquired great interest among consumers mainly due to the nutritional benefits that are claimed for this crop (Salgado-Cruz et al. 2013). Chia is an annual herbaceous plant that belongs to the Lamiaceae family generally cultivated in tropical regions producing 1.8 mm length per 1.2 mm width seeds (Ixtaina et al. 2008; Muñoz et al. 2012). They are often consumed raw, and some recalls were experienced due to contamination with foodborne pathogens (Tamber et al. 2016). In this study, the hypothesis that phylogenetic distance, physiological traits, and relative abundance of species exert a filtering effect on

microbial assemblages was explored using a combination of molecular and cultivation methods. PCR-amplicon sequencing of taxonomic marker genes was first conducted to obtain a portrait bacterial and fungal community structure associated with chia seeds and explore co-occurrence between individual species. Efforts were then invested in building a representative collection of bacterial isolates associated with chia. Finally, bacterial isolates for which the 16S rRNA gene sequence was identical to amplicon sequence variant (ASV) detected in the molecular survey were characterized. Pairwise phylogenetic distance, distance based on physiological traits, and relative abundances were computed for all combinations of isolates to test the relevance of these variables to explain observed interspecific co-occurrences in the molecular survey.

2. Materials and methods

2.1 Samples

Twenty-five samples of non-sprouted chia from Argentina (n = 7), Paraguay (n = 4), Bolivia (n = 5), Ecuador (n = 3), Nicaragua (n = 1), Mexico (n = 1), and a Canadian distributor (n = 3).

Another sample was obtained from Nativas® Organics (n = 1), a California based seed processor and distributor where used in this study. Chia samples originating from the same site arise from different batches. These seeds were stored at room temperature in plastic bags for less than two weeks prior to genomic DNA extraction.

2.2 Microbial community structure

Total genomic DNA was extracted from chia samples using Fast DNA Spin Kit® (MP Biomedicals, Santa Ana, California, USA) according to specifications of the manufacturer, including two successive mechanical lysis steps performed with a FastPrep-24® homogenizer

(MP Biomedicals, Santa Ana, California, USA). Based on preliminary tests of the extraction procedure on different amounts of chia (100, 200 or 300 mg), samples were processed using 100 mg chia because seed mucilage impaired genomic DNA extraction from 200 and 300 mg samples. The abundance of 16S rRNA gene of bacteria quantified by qPCR with the primers Eub338 (Lane 1991) and Eub518 (Muyzer et al. 1993) was $1.2 (2.4) \times 10^6$ copies $g_{(dw)}^{-1}$ and the internal transcribed spacer (ITS) region of the nuclear ribosomal repeat unit of fungi quantified by qPCR with the primers ITS1 (Gardes and Bruns 1993) and ITS4 (White et al. 1990) was $4.3 (5.3) \times 10^5$ copies $g_{(dw)}^{-1}$ (data not shown). Extracted DNA samples were subjected to marker genes PCR amplification, library preparation and high-throughput sequencing on Illumina MiSeq PE-250 platform at the McGill University and Genome Quebec Innovation Center (Montreal, Canada). Barcoding sequences and PCR procedures are described in table S1 (cjm-2019-0052.R1suppl). Primers B969F and BA1406R were used to PCR-amplify V6-V8 region of bacterial 16S rRNA gene (Comeau et al. 2011), while the primers ITS3_KYO2F and ITS4_KYO3R were used for the second internal transcribed spacer (ITS2) flanking the 5.8S and 28S rRNA genes in fungi (Toju et al. 2012). Raw sequencing reads were proceeded using the software Usearch version 10 (Edgar 2010). Paired reads were assembled to a total length varying between 400 and 500 nucleotides for bacteria and 190 and 450 nucleotides for fungi. In all 2,691,709 bacterial and 1,583,794 fungal merged read sequences were subject to quality control. Maximum mismatch threshold in the overlapped region of assembly was set using default parameters (5 for bacteria, and 10 for fungi) and primers were removed from each sequence. Sequences having less than one erroneous base were accepted for downstream quality control steps. Reads were then dereplicated, singletons were discarded and denoised using Unoise3 (Edgar 2016b). Sequences shorter than 366 and 150 nucleotides for bacteria and fungi were

discarded, respectively. Classification of the resulting filtered sequences was undertaken using two successive clustering approaches implemented in Unoise3 (Edgar 2016b). In the first approach, single sequences were clustered into ASVs corresponding to a classification at the 100% sequence identity threshold (Callahan et al. 2017). This clustering approach was efficient to assign ASV sequence retrieved from the molecular survey to the 16S rRNA gene sequence of a bacterial isolate. Indeed, unambiguous assignment of each isolate to a single ASV was mandatory for the theoretical framework of microbial species co-occurrence. ASVs representing less than 0.005% of read counts were removed before the second clustering approach grouping ASV sequences into Operational Taxonomic Units (OTUs) using a 97% identity level cutoff. This scheme was necessary to increase the prevalence of phylotypes in chia samples in order to fulfill standard requirements of ecological network analyses. Indeed, the OTU clustering method reduced the number of null values in bacterial and fungal relative frequency datasets. The clustering procedures led to the identification of 187 and 115 ASVs grouped into 91 and 90 OTUs for bacteria and fungi (Table S1; cjm-2019-0052.R1suppla), respectively. Taxonomic affiliation of ASVs and OTUs was predicted by k-mer similarity of ASV representative sequences to the RDP (Ribosomal Database Project) version 16 training set of 16S rRNA gene for bacteria (Cole et al. 2014) and RDP Warcup training set version 2 for fungi (Deshpande et al. 2016), altogether considering taxa agreed upon by over 80% of bootstrap replications using the SINTAX algorithm (Edgar 2016a). Reads assigned to (i) chloroplasts and (ii) spurious bacterial and fungal OTUs identified as unknown bacterium and fungus were nonspecific and thus removed from the dataset. Raw sequence reads were deposited in the Sequence Read Archive of the National Center for Biotechnology Information under the Bioproject PRJNA485274.

2.3 Isolation of bacteria

Two attempts were conducted to isolate bacteria associated with chia, using five chia samples randomly selected for each isolation procedure. For the first approach, 100 mg of seeds were transferred into 900 μL saline solution (0.85% NaCl). Seeds were thoroughly mixed for one minute and 100 μL of the mixture was inoculated on oligotrophic R2A agar plate containing (in g L^{-1}) proteose peptone (3.0), sodium pyruvate (0.30), yeast extract (0.50), dibasic potassium phosphate (0.30), casein acid hydrolyzate (0.50), magnesium sulfate heptahydrate (0.05), glucose (0.50), soluble starch (0.50), and agar (15) with a final pH of 7.0. The second approach was aimed at isolating copiotrophic bacteria, using tryptone soya medium (TSB-A) containing (in g L^{-1}) Bacto[®] Tryptone (17.0), Bacto[®] Soytone (3.0), glucose (2.5), sodium chloride (5.0), dipotassium hydrogen phosphate (2.5), and agar (15). Pre-incubation of 100 mg chia seeds was done in 5 ml tryptone soya broth supplemented with the antifungal cycloheximide (100 mg/ml) for three days at 25°C under 150 rpm agitation from which 100 μL was spread on TSB agar plate. Incubation of R2A and TSB-A plates was performed at 30°C. Individual colonies were isolated and purified through three successive transfers on agar plates (R2A or TSB-A). The biomass of axenic culture was washed from agar plate with 5 ml glycerol (20% v/v). One aliquot was collected for storage at -80°C and the residual volume was subjected to centrifugation for 10 min (20,000 g, 4°C), and the resulting bacterial pellet was mixed with 100-200 mg silica beads (150-212 μm diameter), 1 ml TEN buffer (50 mM Tris-HCl, 10 mM EDTA, 150 mM NaCl, pH 8.0) and 20 μL SDS (20% w/v) for genomic DNA extraction. Mechanical cell lysis was performed in two successive times with FastPrep-24[®] homogenizer (MP Biomedicals, Santa Ana, California, USA) for 45 seconds at 6.5 m s^{-1} , with a 5-minute incubation on ice between cycles. The lysed cell mixture was centrifuged (20,000 g for 10 minutes) and the recovered

aqueous phase was treated with RNase A ($20 \mu\text{g ml}^{-1}$) for 10 minutes at room temperature before adding 500 μL phenol: chloroform: isoamyl alcohol solution (25: 24: 1, pH 7.0) to purify DNA. The aqueous phase obtained after centrifugation (20,000 g for 10 minutes) was mixed with 500 μL chloroform: isoamyl alcohol solution (24: 1) and centrifuged at 20,000 g for 2 minutes at 4°C . After collecting the aqueous phase (approximately 500 μL), a 167 μL volume of ammonium acetate (10 M) was added, incubated for 20 minutes on ice and centrifuged (20,000 g for 15 minutes). The supernatant was collected, and nucleic acids were precipitated with 1 ml ethanol (100%) at -20°C . The pellet was recovered after centrifugation (20,000 g for 15 minutes at 4°C) and DNA was solubilized in 100 μL of sterile nuclease-free water. Quality of genomic DNA was confirmed by 1% (w/v) agarose gel electrophoresis. Extracted DNA aliquots were stored at -20°C .

2.4 Molecular identification and classification of bacterial isolates

Genomic DNA extracted from axenic cultures was subjected to PCR amplification of bacterial 16S rRNA gene using the primers 27F and 1492R (Lane 1991; Turner et al. 1999). PCR products were shipped to McGill University and Genome Quebec Innovation Center (Montreal, Canada) for Sanger sequencing using the primer 1492R. Gene sequences of the isolates were thoroughly examined, and ambiguous bases were corrected on Chromas[®] version 2.6.5 (Technelysium Pty Ltd, South Brisbane, Australia). The sequences from isolates and representative sequence of each ASV detected in the whole molecular survey were aligned with Muscle algorithm (Edgar 2004). The alignment of sequences obtained counted 700 positions. The software Bioedit version 7.0.3 (Hall 1999) was used to generate a pairwise identity matrix to cluster isolates into species (97% identity cutoff between 16S rRNA gene sequence of bacterial isolates) and find which bacterial

isolate is assigned to retrieved ASV sequences (100% identity cutoff between ASV and isolate 16S rRNA gene sequences) in the molecular survey. Basic Local Alignment Search Tool was used to retrieve 16S rRNA sequences from the National Center for Biotechnology Information (NCBI) database (<http://www.ncbi.nlm.nih.gov/>) similar to those of bacterial isolates. Only genomic information from type material was considered, and three entries from different genera showing the highest identity score to query were kept for phylogenetic analyses. Phylogenetic analyses were conducted using 16S rRNA sequences retrieved from the molecular survey (OTUs), bacterial isolates, and NCBI. The resulting 147 sequences were imported in the software Mega 6.0 (Tamura et al. 2013) and aligned using Muscle algorithm (Edgar 2004), resulting in 302 consensual nucleic acid positions in the final dataset. The phylogenetic tree was computed by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura and Nei 1993).

2.5 Physiological traits of bacterial isolates

Physiological traits of bacterial isolates were evaluated using GEN III MicroPlate® (Biolog Inc, Hayward, California, USA) comprising 94 phenotypic reactions, with 71 corresponding to carbon sources and 23 corresponding to sensitivity tests to chemical inhibitors. The use of carbon substrates or resistance to chemical inhibitors by isolates results in the appearance of a purple color. This purple coloration is produced by reducing the indicator, tetrazolium, to a colored formazan compound. Qualitative reading of the plate was conducted by the assignment of 0 (no color) or 1 (purple color) binary score to each well after 24-36 hours incubation at 30°C. For each plate, wells were inoculated with 100 µL bacterial suspension consisting of 3-mm diameter colony forming unit harvested from R2A agar in 15 mL inoculation fluid. Selection of

inoculation fluid (protocol A or B) was based on the response of negative control of each strain, with protocol B utilized in the case of appearance of a purple color in the negative control well included in GEN III MicroPlate® with protocol A. Physiological traits of bacterial strains were defined by their response to 23 chemical sensitivity assays encompassing tolerance to acidic pH (pH 5 and 6), salt (NaCl 1, 4, and 8%), growth inhibitors (Niaproof 4®, guanidine hydrochloride, sodium lactate, D-serine, lithium chloride, potassium tellurite, sodium butyrate, sodium bromate, tetrazolium violet, and tetrazolium blue), and antibiotics (troleandomycin, rifampicin, minocycline, lincomycin, vancomycin, nalidixic acid, aztreonam, and fusilic acid).

2.6 Statistical analyses

Statistical analyses were performed with the software R version 3.3.0 (R Development Core Team 2008) using the “*vegan*” package (Oksanen et al. 2012) unless otherwise stated. The potential relationship between the geographic origin of chia samples and microbial community structure was first examined following two complementary approaches. Firstly, alpha diversity indices and species richness estimator of bacterial and fungal community structure were computed, and Kruskal-Wallis tests were executed with the “*stats*” package (R Development Core Team 2008). Secondly, the contribution of geographic origin of chia samples to partition Bray-Curtis dissimilarity matrix of bacterial and fungal OTU relative abundance data were computed by permutational multivariate analysis of variance (Permanova) and visualized with principal coordinate analysis (PCoA) using the “*Phyloseq*” package (McMurdie and Holmes 2013). Co-occurrence network of bacterial and fungal OTU was examined using SparCC coefficient computed in the “*SpiecEasi*” package (Kurtz et al. 2015) parameterized with a threshold of 0.1 and a maximum number of iterations of 20 (Friedman and Alm 2012). The

significance of pairwise SparCC coefficients was determined through 1000 bootstraps and this resulted to minimal coefficient value of 0.6 to obtain a pseudo P-value of 0.05. Although correlation matrix derived from a subset of bacterial and fungal species detected in a minimal number of samples (*e.g.*, 40-50%) is a current practice to avoid analyses involving rare species whose counts are uncertain, that approach leads to substantial loss of data. Using that approach, 86% of bacterial and fungal OTU were excluded and not relevant for subsequent isolation efforts. If one further considers the small proportion of cultivable microorganisms in the environment, these constraints leave little room to obtain experimental evidence supporting co-occurrence patterns. Therefore, a second approach was used to compute bacterial species co-occurrence using the whole OTU dataset associated with bacteria to achieve a trade-off between confidence regarding computed pairwise correlations and the probability to obtain relevant isolates for the proposed co-occurrence model framework. Statistical analyses related to bacterial isolates first included the elaboration of rarefaction curves to assess the isolation effort of the bacteria on R2A and TSB-A cultivation media. Comparison of isolates based on physiological traits was performed using a Jaccard distance matrix generated on the binary dataset.

Multiple linear regression analyses based on the theoretical framework of Poisot et al. (2015) were computed to evaluate the contribution of abundance, physiological traits, and taxonomy in explaining co-occurrence pattern of bacterial species observed in the 25 chia samples:

$$A_{(i,j)} = \tau_{(i,j)} + PD_{(i,j)} + N_{(i,j)} \quad \text{(equation 1)}$$

where bacterial species are isolates i and j assigned to OTU i and j in the molecular survey, A_{ij} is the pairwise SparCC coefficient between the OTU i , and j computed using the whole OTU table, $\tau_{(i,j)}$ is the Jaccard distance of physiological traits between isolates i and j , $PD_{(i,j)}$ is the phylogenetic distance computed as the difference score (score of 0 corresponds to 100% identity) between 16S rRNA partial gene sequences of isolates i and j after pairwise alignment and $N_{(i,j)}$ is a term representing the relative abundance of OTU i and j :

$$N_{(i,j)} = \left(\frac{\sum_{n=1}^{25} x_i}{\sum_{n=1}^{25} x_j} \right) \times (\sum_{n=1}^{25} x_i + \sum_{n=1}^{25} x_j) \tag{equation 2}$$

where $\sum_{n=1}^{25} x_i$ and $\sum_{n=1}^{25} x_j$ correspond to the sum of the relative abundance of x_i and x_j in the 25 chia samples, respectively. By convention, the smallest value between $\sum_{n=1}^{25} x_i$ and $\sum_{n=1}^{25} x_j$ is used as the numerator for the first term of the equation. This calculation was elaborated to adjust the ratio of the relative abundance of both species with the sum of their relative abundance with the rationale that the probability for two bacterial OTU species to be in close proximity and interact in chia increases with their absolute abundance (Cardinale et al. 2015; Malakar et al. 2003). The theoretical framework also was tested using Spearman correlations coefficient in replacement of SparCC score to challenge the model.

3. Results

3.1 Microbial community structure and co-occurrence between microbial species

A total of 25 chia samples originating from eight different sources were obtained for this study. According to species richness estimator (Chao1), the marker gene PCR-amplicon sequencing

effort was sufficient to recover the whole diversity of microbial communities (Table 1). Bacterial communities were dominated by the phyla *Proteobacteria* (91.8%), *Firmicutes* (3.5%) and *Chlamydiae* (1.7%), *Actinobacteria* (1.2%), *Bacteroidetes* (1.2%), and *Chloroflexi* (0.6%), while most of the fungi were represented by *Ascomycota* (99.9%). The bacterial OTU B6 and B17 affiliated to Burkholderiaceae and the fungal OTU 1 affiliated to Pleosporaceae comprised a core microbiome detected in all samples. Species richness and evenness of microbial communities could not be discriminated by origin (Table 1). The potential impact of chia sample origin on the beta diversity of microorganisms was further explored by computing PCoA, with the first two axes explaining, respectively, 32 and 41% of the variation in bacterial (Figure 1A) and fungal (Figure 1B) community profiles. Dispersion of microbial community profiles in the reduced space of the PCoA showed no unambiguous patterns explained by sample origin. The relationship between sample origin and microbial community profiles was further examined using Permanova analyses, considering the whole community instead of the previous reduced space, computed on chia seed samples represented by at least three independent samples. The results indicated no significant contribution of sample origin in explaining composition of bacterial communities ($R^2 = 0.23$; $P = 0.08$) whereas a significant contribution was observed for fungi ($R^2 = 0.28$; $P = 0.02$).

Co-occurrence of microbial species was first explored using a restricted list of ubiquitous OTUs detected in more than 11 (44%) chia samples (Figure 2). The relative abundance of these ubiquitous species varied between 0.05-34% and 0.01-42% for bacteria and fungi, respectively. Network structure was fragmented, with the occurrence of three modules comprising between two and five members (Figure 2). The fungal OTU F132 showed the highest connectivity, with

positive co-occurrence with fungal OTU 49 and negative co-occurrence with fungal OTU-F6 and OTU-F9. In the second approach, microbial species co-occurrence was analyzed using the whole bacterial and fungal OTU datasets to achieve a trade-off between confidence regarding computed pairwise correlations and the probability to obtain relevant isolates for the proposed co-occurrence model framework ($A_{(i,j)}$; Table S2; cjm-2019-0052.R1suppla).

3.2 Isolation of bacteria associated with chia

Individual colony forming units propagated on R2A and TSB-A cultivation media were isolated for downstream DNA extraction and 16S rRNA gene sequencing. All isolates encompassed the Firmicutes and Proteobacteria (Figure 3). In contrast to conventional approaches where differences in phenotypic traits are used to guide isolation efforts, isolation and sequencing of 16S rRNA of all colonies enabled reliable quantification (Table S3; cjm-2019-0052.R1supplc). Indeed, clustering of 16S rRNA sequence of isolates at the 97% identity cut-off was used to compare the number of individual species with theoretical estimates of the whole diversity of cultivable bacteria (Figure 4A). For TSB-A medium, 14 isolates were clustered into five different species mainly represented by *Stenotrophomonas* sp., *Enterobacter* sp. and *Bacillus* sp. (Figure 4B). Additional isolation efforts were not attempted since the number of retrieved species corresponds to 100% species richness estimators (Figure 4A). A broader diversity was observed on R2A agar, with 57 isolates clustered into 14 different species, expected to represent 85% cultivable representatives according to Chao1 species richness estimator (Figure 4A). As observed with TSB-A medium, *Bacillus* spp. and *Enterobacter* spp. were the most represented isolates in R2A medium (Figure 4B).

Since this work seeks to examine the contribution of phylogenetic distance, abundance, and physiological traits in shaping co-occurrence noticed in molecular profile, an exact concordance between the 16S rRNA gene sequence of isolates and PCR amplicons was imposed to bridge cultivation-dependent and cultivation-independent datasets. In all, the 16S rRNA gene sequence of five isolates encompassing alpha- and gamma-Proteobacteria was 100% identical to quality-controlled sequences retrieved from the molecular survey, and were namely *Sphingomonas* sp. AJ3, *Enterobacter* sp. AJ7, *Rhizobium* sp. AJ32, *Sphingomonas* sp. AJ28, and *Methylobacterium* sp. AJ8. Except for *Sphingomonas* sp. AJ28, isolates corresponded to members of non-ubiquitous species (detected in less than 44% of samples). Isolates whose 16S rRNA sequence was more than 97% identical to PCR amplicons classified into more than one OTU were prone to ambiguous assignation and were not considered for further analyses.

Physiological traits of the five selected bacterial isolates were defined by their response to 23 chemical sensitivity assays encompassing tolerance to acidic pH, salt, growth inhibitors, and antibiotics. Selection of ecological-relevant physiological traits is complicated by the inability to define metabolic features conferring fitness to microbial species sharing the same niche. The 23 chemical sensitivity assays were then selected owing to their ease to measure, their general applicability in microbial diagnostics and potential role in defining suitable habitat for microorganisms. The results of individual assays were integrated to compute a Jaccard distance matrix of these selected physiological traits between isolates ($\tau_{(i,j)}$). *Sphingomonas* sp. AJ3 was the most sensitive and *Enterobacter* sp. AJ7 was the most resistant to chemicals. Clusterisation of the phenotype profiles was not explained by the taxonomic distance of isolates since

Sphingomonas sp. AJ28 demonstrated a more similar profile with *Rhizobium* sp. AJ32 than *Sphingomonas* sp. AJ3 (Figure 5).

3.3 Model framework for co-occurrence of microbial species

The five selected isolates and their OTU counterparts in the molecular survey were included in the theoretical framework aimed at testing whether abundance, physiological traits, and taxonomy explain the co-occurrence pattern of bacteria species associated with chia (Table S4; cjm-2019-0052.R1suppld). For the analysis, pairwise SparCC coefficients were retrieved from the co-occurrence analysis computed using the whole bacterial database (Table S2; cjm-2019-0052.R1supplb). None of the three independent variables alone explained observed variations in pairwise correlations between the five isolates (Figure 6A). The application of forward stepwise multiple regression analyses showed that model including physiological traits and abundance terms offered the best performance, explaining 69% variations of observed pairwise SparCC correlation coefficients (Figure 6B). According to model parameters, co-occurrence on chia sample is expected for two species displaying dissimilar traits, but this relationship is lowered when the abundance term ($N_{(i,j)}$) of both species is elevated (Table 2).

4. Discussion

The low cost and convenience of high-throughput sequencing technologies have contributed to facilitating our ability to characterize microbial communities in the environment. It is expected that investigation into the biodiversity and interactions among the members of the microbial communities associated with food will lead to novel advanced biocontrol technologies to establish beneficial microorganisms protecting food against rotting and pathogens (Berg 2015;

De Filippis et al. 2018; Teplitski et al. 2011). This approach has been widely applied to study microbial community dynamics in food fermentation and food processing environment, offering the potential to identify biomarkers for product quality (Bokulich et al. 2016). In contrast, very few attempts have been made to characterize the microbiome of fresh food (Jackson et al. 2013; Leff and Fierer 2013; Ottesen et al. 2013). From the best of our knowledge, this is the first report on the composition of microbial communities associated with ready-to-eat dry products. The number of OTU was in the same order of magnitude than fresh fruits and vegetables (Leff and Fierer 2013; Wassermann et al. 2017), with bacterial and fungal OTU ranging between 7-40 and 10-40 per chia sample, respectively. Although only one survey of fungi is available, it showed the dominance of the fungal lineage *Ascomycota* in tomatoes (Ottesen et al. 2013). Our results are in agreement with this survey to suggest that species richness of fungal communities associated with chia is of the same magnitude than species richness of bacteria.

Even though there is an increasing number of reports on microbial species co-occurrence patterns in food (Chaillou et al. 2015), only a few attempts were made to disentangle their mechanisms. Experimental validation of species interactions suggested by co-occurrence patterns is complicated by the fact that functioning and metabolism of species examined in axenic culture can be significantly altered in the presence of other species (Ho et al. 2014). Nevertheless, both molecular survey and *in vitro* experiments demonstrated that *Paracoccus aminovorans* promotes growth of *Vibrio cholerae* in the human gut (Midani et al. 2018), supporting the interest of co-cultures to validate potential interactions inferred through co-occurrence analyses. In this study, only five bacterial isolates were unambiguously assigned to OTU retrieved from the molecular survey. Except for *Enterobacter* sp. AJ7 representative of ubiquitous OTU 24 in the molecular

survey, the isolates represented non-abundant OTU, precluding the application of standard arbitrary sample prevalence cutoffs to select a subset of OTU in co-occurrence profiles computing efforts. A SparCC correlation matrix was thus computed using all OTU represented in the molecular survey.

In contrast to approaches aimed at challenging potential microbe interaction in co-cultures, we proposed a theoretical framework to conceptualize co-occurrence patterns by taking into account characteristics of isolates. The results support the hypothesis that relative abundance, phylogenetic distance, and physiological traits explain observed co-occurrence patterns of microbial communities associated with chia seeds. The relevance of the proposed model is further supported through an application of the theoretical framework to co-occurrence pattern of cultivable bacterial OTU expressed using Spearman correlation coefficient instead of SparCC coefficient (Table S5; cjm-2019-0052.R1supple). According to the model, dissimilarity of physiological traits increases the strength of co-occurrence. As expected owing to functional redundancy in bacteria (Ho et al. 2013; Krause et al. 2014), phylogenetic distance term with species abundance or in conjunction with both species abundance and physiological traits lowered model performances. Interpretation of the negative influence of abundance ($N_{(i,j)}$) on co-occurrence in the model is shaded by the compositional nature of OTU datasets, which do not allow inferring the concept of a relationship between species abundance and their interactions.

This study is the first attempt to test the relevance of physiological traits, phylogenetic distance, and relative abundance of bacterial species to explain their co-occurrence patterns in the molecular survey. At this stage, the results are suggestive owing to a limited number and

diversity of isolates representative of microbial OTUs detected in the molecular survey. Cultivation bias that led to the isolation of Proteobacteria and Firmicutes strains combined with the absence of fungi isolate impairs a sound evaluation of the relevance of phylogenetic distance in explaining species co-occurrence. Positive correlations were frequently observed for phylogenetically-close microbial species expected to share similar ecological traits in different habitats encompassing soil, lettuce and human gut (Barberan et al. 2012; Cardinale et al. 2015; Faust et al. 2012). The structure of microbial communities reported in this study represents the legacy of microbial successions that took place along the whole production chain of chia. Based on previous investigations on dairy products and meats, these successions are expected to be in part driven by filtering effects selecting species whose metabolism is compatible with physicochemical conditions prevailing in chia seeds such as low water activity, in addition to stochastic dissemination of environmental species across the different post-harvest stages including processing and distribution (Chaillou et al. 2015; Guidone et al. 2016). The former mechanism is supported by the observation of a core microbiome represented by two bacterial species affiliated to Burkholderiaceae and one fungus species affiliated Pleosporaceae in the 25 chia samples originated from different sources. Nevertheless, stochastic dissemination of microbial species is supported by a weak relationship between species distribution and chia origin as well as low connectivity of co-occurrence patterns which is a particularity distinguishing food from soil or host environments (Parente et al. 2018; Parente et al. 2016). Finding the exact origin of detected bacterial and fungi species is beyond the scope of this study, impairing a sound evaluation of spatial and temporal variations of co-occurrence patterns in chia. As a consequence, it remains unclear whether the role of physiological traits and abundance in deciphering observed bacterial species co-occurrence was the result of a filtering effect taking

place in chia or in the environmental reservoir of microbes that contaminated chia during the whole supply chain. Despite these limitations, the approach we used should be considered in future investigations aimed at conceptualizing microbial species co-occurrence patterns.

5. Acknowledgements

This work has been supported by a Natural Sciences and Engineering Research Council of Canada Engage grant (493409-16) and a Natural Sciences and Engineering Research Council of Canada Engage Plus grant (508707-17) to PC. The authors are grateful to Sarah Piché-Choquette who introduced AJ to the bioinformatics tools utilized in this study and submitted raw sequence reads to the Sequence Read Archive repository (National Center for Biotechnology Information). The authors wish to acknowledge the contribution of the McGill University and Genome Quebec Innovation Centre (Montréal, Canada) for PCR amplicon library preparation and sequencing services.

6. References

Barberan, A., Bates, S.T., Casamayor, E.O., and Fierer, N. 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME Journal*, **6**(2): 343-351.

Bell, G. 2005. The co-distribution of species in relation to the neutral theory of community ecology. *Ecology*, **86**(7): 1757-1770.

Berg, G. 2015. Beyond borders: Investigating microbiome interactivity and diversity for advanced biocontrol technologies. *Microbial Biotechnology*, **8**(1): 5-7.

- 477 Bokulich, N.A., Lewis, Z.T., Boundy-Mills, K., and Mills, D.A. 2016. A new perspective on
478 microbial landscapes within food production. *Current Opinion in Biotechnology* **37**: 182-
479 189.
- 480 Callahan, B.J., McMurdie, P.J., and Holmes, S.P. 2017. Exact sequence variants should replace
481 operational taxonomic units in marker-gene data analysis. *The ISME Journal*, **12**: 2639-
482 2643.
- 483 Cardinale, M., Grube, M., Erlacher, A., Quehenberger, J., and Berg, G. 2015. Bacterial networks
484 and co-occurrence relationships in the lettuce root microbiota. *Environmental*
485 *Microbiology*, **17**(1): 239-252.
- 486 Chaillou, S., Chaulot-Talmon, A., Caekebeke, H., Cardinal, M., Christieans, S., Denis, C.,
487 Hlne Desmonts, M., Dousset, X., Feurer, C., Hamon, E., Joffraud, J.-J., La Carbona, S.,
488 Leroi, F., Leroy, S., Lorre, S., Mac, S., Pilet, M.-F., Prvost, H., Rivollier, M., Roux, D.,
489 Talon, R., Zagorec, M., and Champomier-Vergs, M.-C. 2015. Origin and ecological
490 selection of core and food-specific bacterial communities associated with meat and
491 seafood spoilage. *The ISME Journal*, **9**: 1105-1118.
- 492 Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro,
493 A., Kuske, C.R., and Tiedje, J.M. 2014. Ribosomal Database Project: data and tools for
494 high throughput rRNA analysis. *Nucleic Acids Research*, **42**(D1): D633-D642.
- 495 Comeau, A.M., Li, W.K.W., Tremblay, J.-., Carmack, E.C., and Lovejoy, C. 2011. Arctic
496 ocean microbial community structure before and after the 2007 record sea ice minimum
497 [online]. *PLoS ONE*, **6**(11): e27492. doi:10.1371/journal.pone.0027492.
- 498 De Filippis, F., Parente, E., and Ercolini, D. 2018. Recent past, present, and future of the food
499 microbiome. *Annual Review of Food Science and Technology*, **9**(1): 589-608.

- 500 Deshpande, V., Wang, Q., Greenfield, P., Charleston, M., Porras-Alfaro, A., Kuske, C.R., Cole,
501 J.R., Midgley, D.J., and Tran-Dinh, N. 2016. Fungal identification using a Bayesian
502 classifier and the Warcup training set of internal transcribed spacer sequences.
503 *Mycologia*, **108**(1): 1-5.
- 504 Edgar, R. 2016a. SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS
505 sequences [online]. bioRxiv, doi: <https://doi.org/10.1101/074161>.
- 506 Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
507 throughput. *Nucleic Acids Research*, **32**(5): 1792-1797.
- 508 Edgar, R.C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*,
509 **26**(19): 2460-2461.
- 510 Edgar, R.C. 2016b. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon
511 sequencing [online]. bioRxiv, <http://doi.org/10.1101/081257>.
- 512 Faust, K., Sathirapongsasuti, J.F., Izard, J., Segata, N., Gevers, D., Raes, J., and Huttenhower, C.
513 2012. Microbial co-occurrence relationships in the human microbiome [online]. *PLOS*
514 *Computational Biology*, **8**(7): e1002606. doi:10.1371/journal.pcbi.1002606.
- 515 Friedman, J., and Alm, E.J. 2012. Inferring correlation networks from genomic survey data
516 [online]. *PLOS Computational Biology*, **8**(9): e1002687.
517 doi:10.1371/journal.pcbi.1002687.
- 518 Gardes, M., and Bruns, T.D. 1993. ITS primers with enhanced specificity for
519 basidiomycetes-application to the identification of mycorrhizae and rusts. *Molecular*
520 *Ecology*, **2**(2): 113-118.

- 521 Guidone, A., Zotta, T., Matera, A., Ricciardi, A., De Filippis, F., Ercolini, D., and Parente, E.
522 2016. The microbiota of high-moisture mozzarella cheese produced with different
523 acidification methods. *International Journal of Food Microbiology* **216**: 9-17.
- 524 Hall, T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program
525 for Windows 95/98/NT. *In* *Nucleic Acids Symp. Ser.* 1999. [London]: Information
526 Retrieval Ltd., c1979-c2000. pp. 95-98.
- 527 Ho, A., Kerckhof, F.-M., Luke, C., Reim, A., Krause, S., Boon, N., and Bodelier, P.L.E. 2013.
528 Conceptualizing functional traits and ecological characteristics of methane-oxidizing
529 bacteria as life strategies. *Environmental Microbiology Reports*, **5**(3): 335-345.
- 530 Ho, A., de Roy, K., Thas, O., De Neve, J., Hoefman, S., Vandamme, P., Heylen, K., and Boon,
531 N. 2014. The more, the merrier: heterotroph richness stimulates methanotrophic activity.
532 *The ISME Journal*, **8**(9): 1945-1948.
- 533 Horner-Devine, M.C., Silver, J.M., Leibold, M.A., Bohannan, B.J., Colwell, R.K., Fuhrman,
534 J.A., Green, J.L., Kuske, C.R., Martiny, J.B., and Muyzer, G. 2007. A comparison of
535 taxon co-occurrence patterns for macro-and microorganisms. *Ecology*, **88**(6): 1345-1353.
- 536 Ixtaina, V.Y., Nolasco, S.M., and Tomás, M.C. 2008. Physical properties of chia (*Salvia*
537 *hispanica* L.) seeds. *Industrial Crops and Products*, **28**(3): 286-293.
- 538 Jackson, C.R., Randolph, K.C., Osborn, S.L., and Tyler, H.L. 2013. Culture dependent and
539 independent analysis of bacterial communities associated with commercial salad leaf
540 vegetables [online]. *BMC Microbiology*, **13**(1): 274. doi:10.1186/1471-2180-13-274.
- 541 Kamneva, O.K. 2017. Genome composition and phylogeny of microbes predict their co-
542 occurrence in the environment [online]. *PLOS Computational Biology*, **13**(2): e1005366.
543 doi:10.1371/journal.pcbi.1005366.

- 544 Kaul, A., Mandal, S., Davidov, O., and Peddada, S.D. 2017. Analysis of microbiome data in the
545 presence of excess zeros [online]. *Frontiers in Microbiology*, **8**: 2114.
546 doi:10.3389/fmicb.2017.02114.
- 547 Krause, S., Le Roux, X., Niklaus, P.A., Van Bodegom, P.M., Lennon, J.T., Bertilsson, S.,
548 Grossart, H.-P., Philippot, L., and Bodelier, P.L.E. 2014. Trait-based approaches for
549 understanding microbial biodiversity and ecosystem functioning [online]. *Frontiers in*
550 *Microbiology*, **5**: 251. doi:10.3389/fmicb.2014.00251.
- 551 Kurtz, Z.D., Müller, C.L., Miraldi, E.R., Littman, D.R., Blaser, M.J., and Bonneau, R.A. 2015.
552 Sparse and compositionally robust inference of microbial ecological networks [online].
553 *PLOS Computational Biology*, **11**(5): e1004226. doi:10.1371/journal.pcbi.1004226.
- 554 Lane, D. 1991. 16S/23S rRNA sequencing. *Nucleic acid techniques in bacterial systematics*.
555 John Wiley & Sons, New York. pp. p. 115-175.
- 556 Leff, J.W., and Fierer, N. 2013. Bacterial communities associated with the surfaces of fresh fruits
557 and vegetables [Online]. *PLOS ONE*, **8**(3): e59310. doi:10.1371/journal.pone.0059310.
- 558 Malakar, P.K., Barker, G.C., Zwietering, M.H., and van't Riet, K. 2003. Relevance of microbial
559 interactions to predictive microbiology. *International Journal of Food Microbiology*,
560 **84**(3): 263-272.
- 561 Mandakovic, D., Rojas, C., Maldonado, J., Latorre, M., Travisany, D., Delage, E., Bihouée, A.,
562 Jean, G., Díaz, F.P., Fernández-Gómez, B., Cabrera, P., Gaete, A., Latorre, C., Gutiérrez,
563 R.A., Maass, A., Cambiazo, V., Navarrete, S.A., Eveillard, D., and González, M. 2018.
564 Structure and co-occurrence patterns in microbial communities under acute
565 environmental stress reveal ecological factors fostering resilience [online]. *Scientific*
566 *Reports*, **8**(1): 5875. doi:10.1038/s41598-018-23931-0.

- 567 McMurdie, P.J., and Holmes, S. 2013. phyloseq: An R package for reproducible interactive
568 analysis and graphics of microbiome census data [online]. PLOS ONE, **8**(4): e61217.
569 doi:10.1371/journal.pone.0061217.
- 570 Mejlholm, O., Gunvig, A., Borggaard, C., Blom-Hanssen, J., Mellefont, L., Ross, T., Leroi, F.,
571 Else, T., Visser, D., and Dalgaard, P. 2010. Predicting growth rates and growth boundary
572 of *Listeria monocytogenes* — An international validation study with focus on processed
573 and ready-to-eat meat and seafood. International Journal of Food Microbiology, **141**(3):
574 137-150.
- 575 Midani, F.S., Weil, A.A., Chowdhury, F., Begum, Y.A., Khan, A.I., Debela, M.D., Durand,
576 H.K., Reese, A.T., Nimmagadda, S.N., Silverman, J.D., Ellis, C.N., Ryan, E.T.,
577 Calderwood, S.B., Harris, J.B., Qadri, F., David, L.A., and LaRocque, R.C. 2018. Human
578 gut microbiota predicts susceptibility to *Vibrio cholerae* infection. The Journal of
579 Infectious Diseases, **218**(4): 645-653.
- 580 Muñoz, L.A., Cobos, A., Diaz, O., and Aguilera, J.M. 2012. Chia seeds: Microstructure,
581 mucilage extraction and hydration. Journal of Food Engineering, **108**(1): 216-224.
- 582 Muyzer, G., de Waal, E.C., and Uitterlinden, A.G. 1993. Profiling of complex microbial
583 populations by denaturing gradient gel electrophoresis analysis of polymerase chain
584 reaction-amplified genes coding for 16S rRNA. Applied and Environmental
585 Microbiology, **59**(3): 695-700.
- 586 Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Simpson, G.,
587 Solymos, P., Henry, M., Stevens, H., and Wagner, H. 2012. vegan: community ecology
588 package. R package version 2.0-4. Available from [http://cran.r-](http://cran.r-project.org/package=vegan)
589 [project.org/package=vegan](http://cran.r-project.org/package=vegan).

- 590 Ottesen, A.R., González Peña, A., White, J.R., Pettengill, J.B., Li, C., Allard, S., Rideout, S.,
591 Allard, M., Hill, T., Evans, P., Strain, E., Musser, S., Knight, R., and Brown, E. 2013.
592 Baseline survey of the anatomical microbial ecology of an important food plant: *Solanum*
593 *lycopersicum* (tomato) [online]. BMC Microbiology, **13**(1): 114. doi:10.1186/1471-2180-
594 13-114.
- 595 Parente, E., Zotta, T., Faust, K., De Filippis, F., and Ercolini, D. 2018. Structure of association
596 networks in food bacterial communities. Food Microbiology, **73**: 49-60.
- 597 Parente, E., Cocolin, L., De Filippis, F., Zotta, T., Ferrocino, I., O'Sullivan, O., Neviani, E., De
598 Angelis, M., Cotter, P.D., and Ercolini, D. 2016. FoodMicrobionet: A database for the
599 visualisation and exploration of food bacterial communities based on network analysis.
600 International Journal of Food Microbiology, **219**: 28-37.
- 601 Poisot, T., Stouffer, D.B., and Gravel, D. 2015. Beyond species: why ecological interaction
602 networks vary through space and time. Oikos, **124**(3): 243-251.
- 603 R Development Core Team. 2008. R: A language and environment for statistical computing.
604 Edited by R.F.f.S. Computing, Vienna, Austria.
- 605 Salgado-Cruz, M.d.l.P., Calderón-Domínguez, G., Chanona-Pérez, J., Farrera-Rebollo, R.R.,
606 Méndez-Méndez, J.V., and Díaz-Ramírez, M. 2013. Chia (*Salvia hispanica* L.) seed
607 mucilage release characterisation. A microstructural and image analysis study. Industrial
608 Crops and Products, **51**: 453-462.
- 609 Tamber, S., Swist, E., and Oudit, D. 2016. Physicochemical and bacteriological characteristics of
610 organic sprouted chia and flax seed powders implicated in a foodborne salmonellosis
611 outbreak. Journal of Food Protection, **79**(5): 703-709.

- 612 Tamura, K., and Nei, M. 1993. Estimation of the number of nucleotide substitutions in the
613 control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology*
614 *and Evolution*, **10**(3): 512-526.
- 615 Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: Molecular
616 evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, **30**(12):
617 2725-2729.
- 618 Tenenhaus-Aziza, F., and Ellouze, M. 2015. Software for predictive microbiology and risk
619 assessment: a description and comparison of tools presented at the ICPMF8 Software
620 Fair. *Food Microbiology*, **45**: 290-299.
- 621 Teplitski, M., Warriner, K., Bartz, J., and Schneider, K.R. 2011. Untangling metabolic and
622 communication networks: interactions of enterics with phytobacteria and their
623 implications in produce safety. *Trends in Microbiology*, **19**(3): 121-127.
- 624 Toju, H., Tanabe, A.S., Yamamoto, S., and Sato, H. 2012. High-coverage ITS primers for the
625 DNA-based identification of ascomycetes and basidiomycetes in environmental samples
626 [online]. *PLoS ONE*, **7**(7): e40863. doi:10.1371/journal.pone.0040863.
- 627 Turner, S., Pryer, K.M., Miao, V.P., and Palmer, J.D. 1999. Investigating deep phylogenetic
628 relationships among cyanobacteria and plastids by small subunit rRNA sequence
629 analysis. *Journal of Eukaryotic Microbiology*, **46**(4): 327-338.
- 630 Wassermann, B., Rybakova, D., Müller, C., and Berg, G. 2017. Harnessing the microbiomes of
631 Brassica vegetables for health issues [online]. *Scientific Reports*, **7**(1): 17649.
632 doi:10.1038/s41598-017-17949-z.

- 633 White, T.J., Bruns, T., Lee, S., and Taylor, J. 1990. Amplification and direct sequencing of
634 fungal ribosomal RNA genes for phylogenetics. PCR protocols: a guide to methods and
635 applications **18**(1): 315-322.
- 636 Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D.R., Bork, P., and Patil, K.R. 2015.
637 Metabolic dependencies drive species co-occurrence in diverse microbial communities.
638 Proceedings of the National Academy of Sciences of the United States of America,
639 **112**(20): 6449-6454.
- 640 Zoellner, C., Al-Mamun, M.A., Grohn, Y., Jackson, P., and Worobo, R. 2018. Post-harvest
641 supply chain with microbial travelers: A novel farm-to-retail microbial simulation and
642 visualization framework. Applied and Environmental Microbiology,
643 doi:10.1128/aem.00813-18.
- 644
- 645

Figure caption

Figure 1. Principal coordinate analysis of bacterial (A) and fungal (B) communities in chia samples.

Figure 2. Ecological network of ubiquitous bacterial and fungal OTU in chia samples. The OTU are represented by the nodes and edges illustrate significant pairwise correlation (SparCC, pseudo P-value < 0.05). Nodes are colored according to the taxonomic affiliation of the OTU, the numbers appearing next to the nodes refer to OTU identifier and the size of the nodes is scaled according to the logarithm of OTU sequence reads count in the 25 chia samples. The color of the edges indicates positive (green) or negative (red) correlations. Raw read counts of bacterial and fungal OTU tables were combined prior co-occurrence analysis.

Figure 3. Maximum likelihood phylogenetic tree of 16S rRNA gene sequence of isolates and OTU retrieved from the molecular survey. An overview of the taxonomic classification of 16S rRNA gene sequence encompassing three phyla is presented in the first panel (A) with the magnification of (B) alpha- and (C) gamma-Proteobacteria clusters in the secondary panels. Bootstrap values (%) are represented in red characters for the nodes that are supported by at least 50% iterations. The five isolates whose 16S rRNA sequences were 100% identical to sequences retrieved from the molecular survey are shown in bold characters.

Figure 4. Evaluation of the isolation efforts and taxonomic distribution of the isolates. (A) Rarefaction curves were computed for quantitative analysis of isolation efforts on TSB-A and

R2A media. The number of observed species (n) and Chao1 species richness estimator are presented. (B) Proportion of bacterial genera represented by isolates encompassing Alpha-Proteobacteria, Firmicutes, and Gamma-Proteobacteria.

Figure 5. Physiological traits of the five selected bacterial isolates. The heatmap shows positive (1) and negative (0) results for chemical sensitivity assays encompassing tolerance to acidic pH, salt, growth inhibitors, and antibiotics. The UPGMA agglomerative clustering of bacterial isolates is based on a Jaccard distance matrix calculated with results of the chemical sensitivity assays.

Figure 6. Theoretical framework to explain co-occurrence patterns of the five selected bacterial isolates in chia. (A) Single and multiple linear regressions were computed using $PD_{(i,j)}$, $N_{(i,j)}$, and $\tau_{(i,j)}$ as independent variables to explain variation in SparCC pairwise coefficient ($A_{(i,j)}$) observed between each pair of OTUs represented by the eight isolates. Multiple R-square values are presented for each regression analysis. The symbols ** denote p-values lower than 0.05. (B) Linear regression between observed SparCC coefficients and predicted coefficients with $\tau_{(i,j)}$ and $N_{(i,j)}$ terms (see Table 2 for equation parameters).

Table 1. General information related to molecular profile of bacterial and fungal communities associated with chia samples.

None of the estimated parameters showed significant difference between samples of different sources (Kruskal-Wallis test; $p > 0.05$).

Source	n	Bacteria					Fungi				
		n _{reads}	n _{OTU} [†]	Shannon [†]	Simpson [†]	Chao [†]	n _{reads}	n _{OTU} [†]	Shannon [†]	Simpson [†]	Chao [†]
Paraguay	4	5541	13 ± 5	1.5 ± 0.5	0.66 ± 0.18	16 ± 7	266 467	26 ± 9	1.3 ± 0.6	0.52 ± 0.26	28 ± 10
Argentina	7	6076	17 ± 8	2.1 ± 0.5	0.81 ± 0.09	18 ± 8	503 894	20 ± 10	0.7 ± 0.6	0.33 ± 0.29	26 ± 10
Nicaragua	1	674	13	2.0	0.85	13	46 492	14	1.4	0.74	16
Bolivia	5	5061	18 ± 13	1.8 ± 0.7	0.73 ± 0.20	23 ± 13	328 465	22 ± 6	1.1 ± 0.5	0.56 ± 0.25	26 ± 11
Ecuador	3	2848	15 ± 7	2.0 ± 0.6	0.80 ± 0.14	16 ± 9	47 564	16 ± 4	1.2 ± 0.1	0.57 ± 0.11	19 ± 8
Mexico	1	356	7	1.0	0.46	8	3398	10	0.08	0.02	11
Canada	3	1919	11 ± 4	1.7 ± 0.3	0.74 ± 0.89	14 ± 9	216 744	18 ± 6	0.26 ± 0.26	0.11 ± 0.12	23 ± 6
USA	1	31 140	16	1.79	0.75	22	109 226	25	1.77	0.77	28

[†]For chia sources where n > 1, mean and standard deviation are provided.

Table 2. Multiple regressions showing the relationship of species co-occurrence ($A_{(i,j)}$) with physiological traits ($\tau_{(i,j)}$) and relative abundance ($N_{(i,j)}$). Equations were derived with observations of five selected bacterial isolates represented in the molecular survey.

Equations [†]	Multiple R ² (p-value)	Residual
$A_{(i,j)} = 0.49(\pm 0.15)\tau_{(i,j)} - 0.080(\pm 0.027)N_{(i,j)} - 0.27(\pm 0.10)$	0.69 (0.02)	0.08
$A_{(i,j)} = 0.49(\pm 0.16)\tau_{(i,j)} - 0.27(\pm 1.32)PD_{(i,j)} - 0.09(\pm 0.06)N_{(i,j)} - 0.23(\pm 0.21)$	0.69 (0.06)	0.08
$A_{(i,j)} = 0.44(\pm 0.17)\tau_{(i,j)} + 1.4(\pm 0.70)PD_{(i,j)} - 0.46(\pm 0.16)$	0.57 (0.05)	0.09

[†]Confidence interval of the coefficients is provided in parentheses.

Figure 1. Principal coordinate analysis of bacterial (A) and fungal (B) communities in chia samples.

185x99mm (300 x 300 DPI)

Figure 2. Ecological network of ubiquitous bacterial and fungal OTU in chia samples. The OTU are represented by the nodes and edges illustrate significant pairwise correlation (SparCC, pseudo P-value < 0.05). Nodes are colored according to the taxonomic affiliation of the OTU, the numbers appearing next to the nodes refer to OTU identifier and the size of the nodes is scaled according to the logarithm of OTU sequence reads count in the 25 chia samples. The color of the edges indicates positive (green) or negative (red) correlations. Raw read counts of bacterial and fungal OTU tables were combined prior co-occurrence analysis.

91x99mm (300 x 300 DPI)

Figure 3. Maximum likelihood phylogenetic tree of 16S rRNA gene sequence of isolates and OTU retrieved from the molecular survey. An overview of the taxonomic classification of 16S rRNA gene sequence encompassing three phyla is presented in the first panel (A) with the magnification of (B) alpha- and (C) gamma-Proteobacteria clusters in the secondary panels. Bootstrap values (%) are represented in red characters for the nodes that are supported by at least 50% iterations. The five isolates whose 16S rRNA sequences were 100% identical to sequences retrieved from the molecular survey are shown in bold characters.

190x278mm (300 x 300 DPI)

Figure 4. Evaluation of the isolation efforts and taxonomic distribution of the isolates. (A) Rarefaction curves were computed for quantitative analysis of isolation efforts on TSB-A and R2A media. The number of observed species (n) and Chao1 species richness estimator are presented. (B) Proportion of bacterial genera represented by isolates encompassing Alpha-Proteobacteria, Firmicutes, and Gamma-Proteobacteria.

209x89mm (300 x 300 DPI)

Figure 5. Physiological traits of the five selected bacterial isolates. The heatmap shows positive (1) and negative (0) results for chemical sensitivity assays encompassing tolerance to acidic pH, salt, growth inhibitors, and antibiotics. The UPGMA agglomerative clustering of bacterial isolates is based on a Jaccard distance matrix calculated with results of the chemical sensitivity assays.

90x79mm (300 x 300 DPI)

Figure 6. Theoretical framework to explain co-occurrence patterns of the five selected bacterial isolates in chia. (A) Single and multiple linear regressions were computed using $PD(i,j)$, $N(i,j)$, and $\tau(i,j)$ as independent variables to explain variation in SparCC pairwise coefficient ($A(i,j)$) observed between each pair of OTUs represented by the eight isolates. Multiple R-square values are presented for each regression analysis. The symbols ** denote p-values lower than 0.05. (B) Linear regression between observed SparCC coefficients and predicted coefficients with $\tau(i,j)$ and $N(i,j)$ terms (see Table 2 for equation parameters).

93x194mm (300 x 300 DPI)