CrossMark

ORIGINAL ARTICLE

Takeshi Miki ⓘ · Taichi Yokokawa · Po-Ju Ke ·
I-Fang Hsieh · Chih-hao Hsieh · Tomonori Kume ·
Kinuyo Yoneya · Kazuaki Matsui

Statistical recipe for quantifying microbial functional diversity from EcoPlate metabolic profiling

Abstract EcoPlate quantifies the ability of a microbial community to utilize 31 distinct carbon substrates, by monitoring color development of microplate wells during incubation. Well color patterns represent metabolic profiles. Previous studies typically used color patterns representing average values of three technical replicates on the final day of the incubation and did not consider substrate chemical diversity. However, color fluctuates during incubation and color varies between replicates, undermining statistical power to distinguish differences among samples in microbial functional composition and diversity. Therefore, we developed a protocol to improve statistical power with two approaches. First, we optimized data treatment for color development during incubation and technical replicates. Second, we incorporated chemical structural information for the 31 carbon substrates into the computation. Our framework implemented as the protocol in the R environment is able to compare the statistical power among different calculation methods. When we applied it to data from aquatic microcosm and forest soil systems, we observed substantial improvement in statistical power when we incorporated temporal patterns during incubation instead of using only endpoint data. Using maximum or minimum values of technical replicates also sometimes gave better results than averages. Incorporating chemical structural information based on fuzzy set theory could improve statistical power but only when relative color density information was considered; it was not seen when the pattern was first binarized into the presence or absence of metabolic activity. Finally, we discuss research directions to improve these approaches and offer some practical considerations for applying our methods to other datasets.

Takeshi Miki and Taichi Yokokawa equally contributed.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11284-017-1554-0) contains supplementary material, which is available to authorized users.

T. Miki · C. Hsieh
Institute of Oceanography, National Taiwan University, No. 1 Sec. 4 Roosevelt Rd, Taipei 10617, Taiwan

T. Miki · C. Hsieh
Research Center for Environmental Changes, Academia Sinica, 128 Academia Road, Section 2, Nankang 11529 Taipei, Taiwan

T. Yokokawa
Research and Development Center for Marine Biosciences, Japan Agency for Marine-Earth Science and Technology, 2-15 Natsushima-cho, Yokosuka City 237-0061, Japan

P.-J. Ke
Department of Biology, Stanford University, 450 Serra Mall, Stanford, CA 94305, USA

I.-F. Hsieh
Department of Biology, Boston University, 5 Cummington Mall, Boston, MA 02215, USA

C. Hsieh
Department of Life Science, Institute of Ecology and Evolutionary Biology, National Taiwan University, No. 1 Sec. 4 Roosevelt Rd, Taipei 10617, Taiwan

C. Hsieh
National Center for Theoretical Sciences, No. 1 Sec. 4 Roosevelt Rd, Taipei 10617, Taiwan

T. Kume
School of Forestry and Resource Conservation, National Taiwan University, No. 1 Sec. 4 Roosevelt Rd, Taipei 10617, Taiwan

K. Yoneya
Department of Agricultural Science, Kindai University, 3327-204, Nakamachi, Nara City 631-8505, Japan

K. Matsui (✉)
Department of Civil and Environmental Engineering, Kindai University, 3-4-1 Kowakae, Higashiosaka City 577-8502, Japan
E-mail: kmatsui@civileng.kindai.ac.jp
Tel.: +81-(0)6672102332

Introduction

Various community profiling methods have been developed to understand ecological structures and functions. DNA-based profiling techniques offer advantages for characterizing the species composition of a community (Woese et al. 1990; Giovannoni and Stingl 2005; Minamoto et al. 2012), whereas RNA-based methods allow the evaluation of active genes in the ecosystem (Baldwin et al. 2001; Moran 2015). Other profiling methods broaden understanding of the in situ ecological and ecosystem functions based on metabolic activities. These latter methods include community-level physiological profiling using the commercially available BIOLOG™ microplate (Preston-Mafham et al. 2002), quantifying multiple enzyme activities (Osono 2007; Siggins et al. 2012), and chemical profiling of plant secondary metabolites, such as defense chemicals and volatiles (Kuhlisch and Pohnert 2015). Common features of the datasets employed by these methods are that they are multivariate and used for calculating similarity (Anderson et al. 2011) and diversity indices (Petchey and Gaston 2006; Villeger et al. 2008; Laliberte and Legen-

dre 2010). Standard statistical indices and methods have been developed to extract ecological patterns (signatures) in some fields, including metacommunity (Pillar and Duarte 2010), microbial ecology (Ramette 2007; Lozupone et al. 2011), chemical ecology (Kuhlisch and Pohnert 2015), and biodiversity estimation (Legendre and Gallagher 2001; Chao et al. 2014).

The statistical analysis of microbial metabolic profiling using BIOLOG has not been standardized (but see Garland et al. 2007), despite increasing use of this method for evaluating functional composition and multifunctionality (Miki et al. 2014 and references therein). BIOLOG EcoPlate (Biolog, Hayward, CA, USA) is generally used to measure the ability of a bacterial community to utilize carbon substrates (Choi and Dobbs 1999). An EcoPlate is a 96-well microplate composed of triplicates of 31 response wells with different sole carbon sources (along with three blank wells as controls). Utilization of each carbon source is quantified by color development in each well (Fig. 1a). The pattern of color development is treated as multivariate (e.g. functional composition) but can be converted to univariate (e.g. threshold-based multifunctionality) (Byrnes et al. 2014).

EcoPlate color development patterns are highly variable due to variability in initial species composition, variability in initial community size (inoculum size), and temporal shifts in color development pattern (Fig. 1a) and species composition (Konopka et al. 1998; Preston-
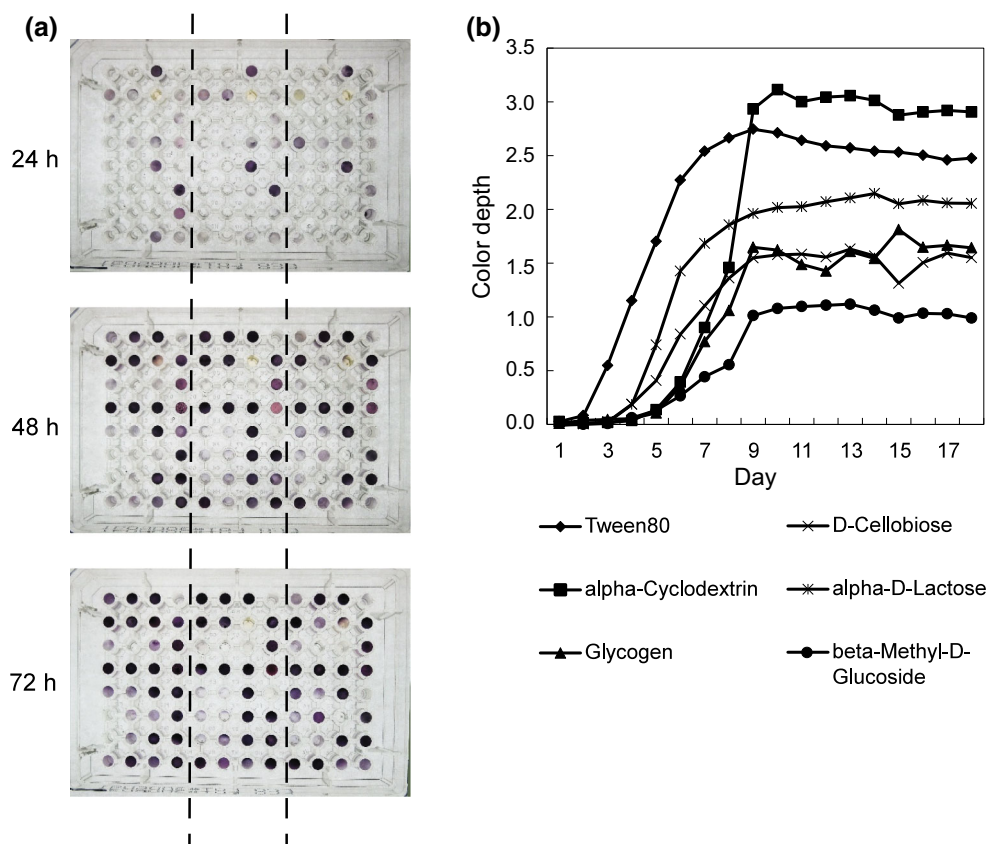
Fig. 1 Examples of EcoPlate color development and its time evolution. **a** Example of temporal shifts in color development by EcoPlate. **b** Color development normalized by blank wells. Data were from a forest soil sample obtained in July 2015

Mafham et al. 2002; Stefanowicz 2006) during incubation. Due to its small volume (100 μL/well), the community size can be less than 10,000 individuals/well, which is a source of nonnegligible stochasticity in initial species composition among the 96 wells in a plate. This situation could result in variation of color development among triplicates of an identical substrate within a plate (Zhou et al. 2013). Scientists tend to use the average value from the triplicate to reduce uncertainty; however, this averaging process may lead to loss of the ecological signature. Even with identical species composition, the difference in initial community size should affect color development (Lawley and Bell 1998; Preston-Mafham et al. 2002). The standard protocol recommends normalizing the color of each well by the average well color development (AWCD) (Garland et al. 2007). However, the AWCD itself could be an ecological indicator of microbial function, such that normalization may lead to some loss of signature.

In addition, temporal shifts of species composition as well as physicochemical conditions (e.g. oxygen concentration, pH, and nutrient concentration) during incubation add uncertainty and difficulty in interpreting the results. A practical solution would be to incubate all samples at common and relatively warm temperatures within a limited period; then, the developed color could be regarded as the potential of the in situ species composition. However, this approach could introduce new bias, because incubation under a common temperature would obscure the temperature-dependent response of the community at the in situ temperature. This problem is linked to another concern, regarding when and how we should evaluate color density (see Fig. 1b). Even if the final color density is similar, the rate to reach the maximum depends on the substrate. Some studies recommend fitting a logistic growth curve and evaluating the slope of the growth and the maximum color density (Garland et al. 2007; Muniz et al. 2014). However, this approach is computationally intensive and might lead to some loss of ecological signature owing to the need to fit a smooth curve.

Another issue regards whether it is ecologically relevant to treat color development from the 31 carbon substrates independently. The 31 substrates can be categorized into several distinct groups, such as amino acids, carbohydrates, and polymers (Hai et al. 2016). More specifically, some substrates are closely related to each other according to metabolic pathways or genes involved. Or more simply, the similarity in terms of chemical structure may act as a simple proxy for metabolic similarity. It is expected that bacteria respond more similarly to the substrates with higher metabolic similarity. Therefore, such dissimilarities among substrates should be incorporated into differences in profiling patterns, analogous to phylogenetic profiling. For example, the Unifrac distance (Lozupone et al. 2011) considers the evolutionary distance (on phylogenetic trees) between species when evaluating the dissimilarity of community composition. Such a method remains unexplored for BIOLOG profiling patterns.

In this study, we focused on two questions. (1) How to identify the best method for profiling the color development pattern from 96-well microplates? (2) How is the information of chemical structure of carbon substrates incorporated, and does such incorporation improve the profile of microbial functions? For the first question, we compared results without using time-series data (using data from the final day of incubation only) to results obtained with time-series data (integrating color development or taking the maximum value along the time evolution). In addition, we compared performances of three metrics (maximum, minimum, and average) using triplicate measurements. We hypothesized that using information along the time evolution of color would improve the quantification. For the second question, assuming that chemical structure similarity is a proxy of metabolic similarity, we weighted the EcoPlate patterns (Dixon 2003) using dissimilarity between carbon substrates, which we calculated based on chemical structure using chemoinformatic tools (Guha 2007). We hypothesized that including chemical dissimilarity would improve quantification. Our objective is to illustrate a framework allowing identify the best method for profiling the color development pattern for a given dataset from the 96-well microplate.

To illustrate our framework, we used two datasets: field soil and aquatic microcosm systems. The basic idea for the evaluation was to compare the explanatory power (R^2 values) under the same statistical model, $Y \sim X_1 + X_2 + \ldots$, where X_k is an explanatory variable (temperature or treatment) and Y is the univariate or multivariate index calculated from EcoPlate patterns (e.g. multifunctionality or functional composition). With different calculation methods, we had different Y value sets, which allowed comparison of explanatory power. Our framework should be generalizable to other datasets.

Methods

Data source

Forest soil experiments

We collected soil samples from a pure Moso bamboo stand in the National Taiwan University Experimental Forest (23.6667N, 120.7833E), located in central Taiwan. Three trenching plots (1 m × 1 m) along a 400-m² plot within the bamboo stand were established in January 2013 (Lin et al. 2017). Connections between living roots and aboveground parts of the plot were cut off, and regrowth of new root into the trenched plots was prevented.

To assess the seasonality of soil microbial function, we collected soil samples in different months as representative of different seasons: December 2014 for winter (dry period), March and May 2015 for early and late spring, respectively (aboveground growing season), and July 2015 for summer. Soil was sampled from the upper layer

(0–10 cm deep) with the soil core, litter was removed (if present), and mixed well to reduce the heterogeneity of microbial community composition. One core was collected from each trenching plot, and three cores in total were collected. Three cores also were collected from a control plot outside the trenching plots. We prepared 1:1000 dilutions from 5-g subsamples, inoculated the EcoPlate with these diluted subsamples, incubated the plates at in situ temperature, and conducted daily measurements up to 30 days. Detailed methods are available in Electronic Supplementary Material (ESM1; Additional methods A) and Hsieh et al. (2016).

Aquatic microcosm experiments

Aquatic microcosm experiments were detailed in Miki et al. (2014). We used 20 isolated bacteria strains from a eutrophic pond (33.8698N, 132.7718E, Matsuyama, Japan), which were isolated by R2A agar plates, to test the effect of initial species loss (from 20 strains to 19 strains) on bacterial multifunctionality under the controlled environment; we did not intend to reconstruct any in situ bacterial community in the pond. We prepared control microcosms with 20 isolates and microcosms with 19 isolates (representing 20 species combinations). We predicted the gene compositions for every microcosm using phylogenetic information of these 20 isolates. For EcoPlate incubation, concentration of each isolate was around 10^3 cells per 100 μL, and the total concentration of each well was around 10^4 cells per 100 μL. In this study, the time series of EcoPlate color development pattern was not available; we measured the pattern on the final (seventh) day of the incubation only.

BIOLOG: EcoPlate

The first step was to inoculate samples into the BIOLOG EcoPlate, either as aqueous samples or after suspension. Utilization of each carbon source during incubation was coupled with the conversion of triphenyl tetrazolium chloride to triphenyl formazan (TPF), such that carbon utilization could be quantified by the color development of TPF in each well. Color absorbance of each well was determined by the optical density at 595 nm by using the microplate reader (Multiskan FC, Thermo Scientific). Detailed information of the EcoPlate, including a list of the 31 carbon substrates, is available at http://www.biolog.com/pdf/milit/00A_012_EcoPlate_Sell_Sheet.pdf.

Data processing

Data from EcoPlate time-series

All of the following processes were conducted in the R environment (ESM1: Additional methods B). Each plate corresponds to one sample. One experiment consisted of measurements on multiple days from each EcoPlate sample (Fig. 2a). We had three options for methods to quantify the signals from the experiment: the temporal maximum, final endpoint, or temporal integration method. For the temporal maximum method, we took the average, maximum, or minimum value of the triplicate from each substrate for each measurement day (Fig. 2b). This option generated a matrix (Fig. 2c), with rows representing different measurement days and columns representing different substrates (i.e., control + 31 substrates = 32 values). For each substrate (and control) in this matrix, we chose only the maximum values among measurement days. This method yielded 32 values for further analyses (Fig. 2d).

For the final endpoint method, we took the measurement taken on the final day of the incubation only, ignoring all other data in the time series. This method yielded 96 values (triplicates of 31 substrates + triplicate of the control) (Fig. 2e). For the temporal integration method, we calculated the cumulative amount of color development by integrating the color density development curve. For normalization, we divided the integrated value by the integration period. This method yielded 96 values (Fig. 2e). For the final endpoint and temporal integration methods, the next step was to take the average, maximum, or minimum of the triplicate values from 96 values, which finally resulted in 32 values (Fig. 2f). In summary, we had nine vectors depending on the calculation method (Fig. 2d, f).

Basis for multifunctionality and functional dissimilarity

We first normalized the color values of substrates by the color values of the control by subtraction, converting 32 values (Fig. 2f) into 31 values for each sample. This normalization was applied to all samples (in different treatments and/or measurement campaigns), resulting in a matrix of sample number × 31 substrates, $\mathbf{E_C}$ (Fig. 3a), for each calculation method. Thus, we had nine matrices representing different ways of calculation, as explained in Fig. 2.

$\mathbf{E_C}$ was used to develop various matrices for comparing multifunctionality (functional diversity) and functional dissimilarity among samples, with the goal of understanding the effect of environmental factors on these indices. One option was to convert $\mathbf{E_C}$ to binary matrix $\mathbf{E_C^B}$ through quantile-based binarization (Byrnes et al. 2014). When color density exceeded the quantile-based threshold value T, the value was converted to 1 (presence of metabolism) or 0 otherwise (absence of metabolism). This binarization step, which was necessary to calculate quantile-based (integer) multifunctionality (Byrnes et al. 2014), yielded a vector $\mathbf{M_F}$ of $1 \times$ sample number (Fig. 3b). Summing the non-zero elements gave an integer (0–31) multifunctionality.

Alternative approaches were to calculate the functional dissimilarity matrix $\mathbf{D_F}$ from $\mathbf{E_C^B}$ via the Jaccard dissimilarity measure (Fig. 3c), or to keep the continu-
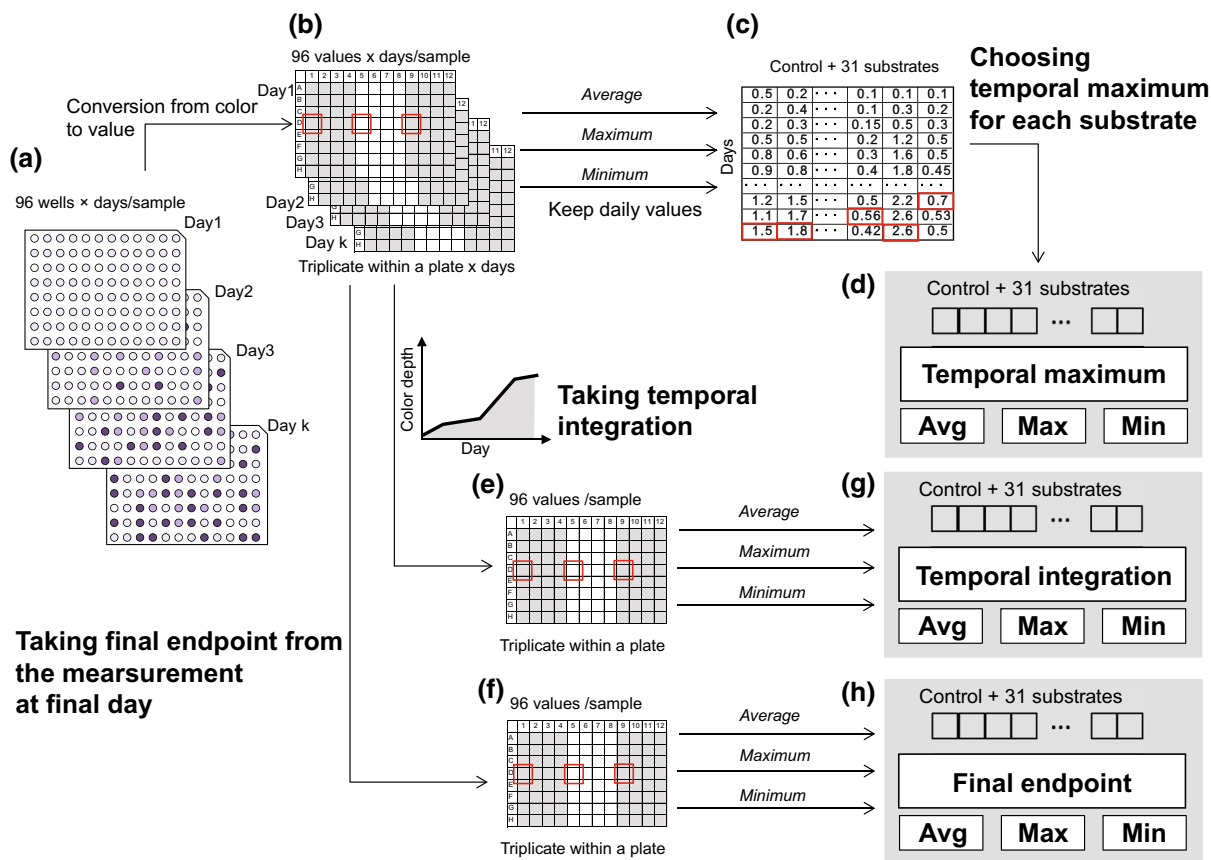
Fig. 2 Diagram illustrating different ways of generating matrices for further analysis. **a** Raw data with 96 wells × measurement days. **b** Color density data for each well of each measurement day were converted into numerical values of 96 values × measurement days. There were nine different ways of generating matrices from this dataset for further analyses. **c** For each measurement day, the average, maximum, or minimum value of the triplicate was used, yielding measurement days × numerical values of 32 (average, maximum, or minimum) values (control + 31 substrates) (**d**). From the time series of color density for each substrate (+ control), the temporal maximum value during the measurement period was selected and converted into 32 values, generating three vectors of "temporal maximum" each for average, maximum, or minimum of the triplicate. Alternatively, raw data were converted into 96 numerical values (**e**) by integrating the color development curve to calculate the cumulative color development or **f** by using data from the final measurement day only. Values in **e** and **f** were converted into 32 values by taking the average, maximum, or minimum of the triplicate, resulting in three vectors of **g** "temporal integration" or **h** "final endpoint"

ous values (E_C) and calculate the functional dissimilarity matrix (D_F) via, for instance, the Bray–Curtis dissimilarity measure. As baseline for the analysis, we used these univariate vectors and multivariate matrices, denoted as M_{FUW} (unweighted multifunctionality), D_{FBUW} (binarized unweighted functional dissimilarity), and D_{FUW} (continuous unweighted functional dissimilarity) (see ESM1: Additional methods C). These methods do not incorporate chemical structural information of the substrates and, thus, assume that the substrates are independent and equally informative.
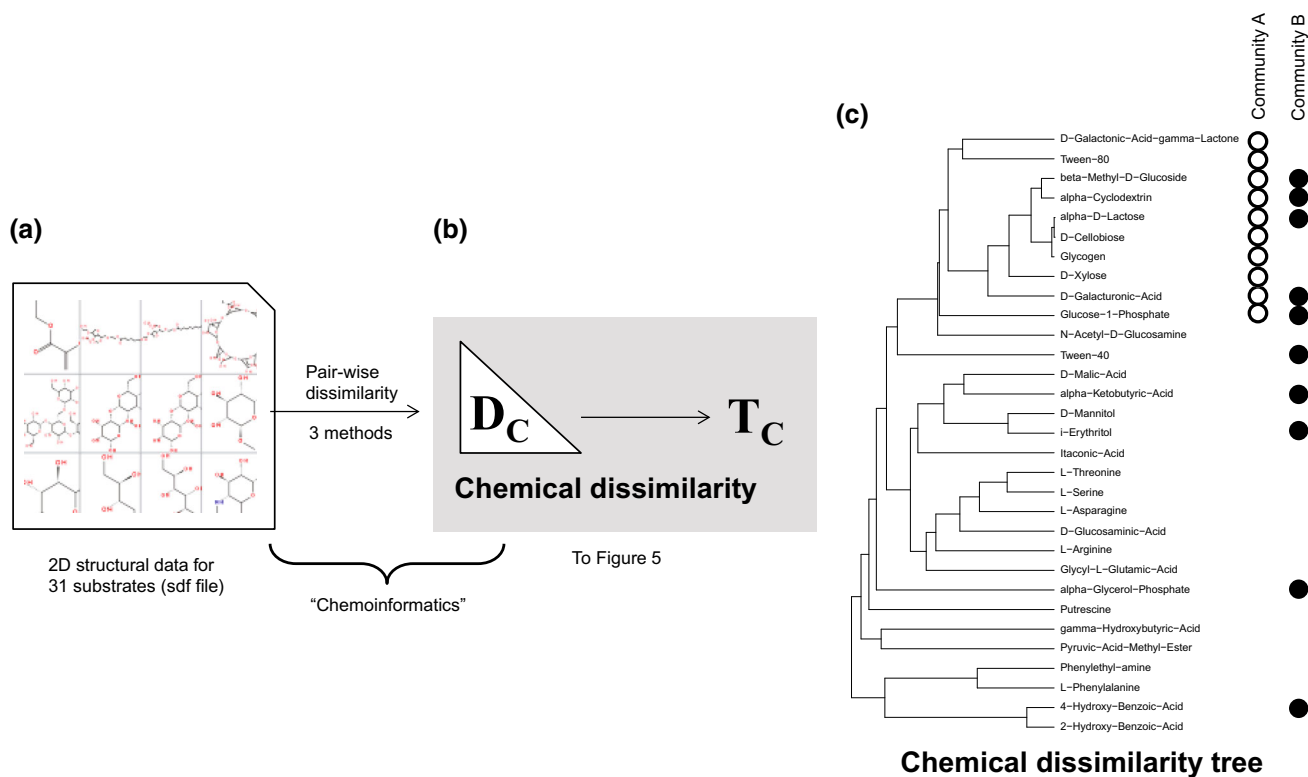
Chemical similarity and clustering trees

To incorporate chemical structural information, we needed to calculate pairwise dissimilarities among the 31 carbon substrates (Willett et al. 1998; Nikolova and Jaworska 2004; Consonni and Todeschini 2012; Todeschini et al. 2012; Floris et al. 2014). As the first step, we downloaded the two-dimensional structural data of each substrate (Fig. 4a) as an sdf file from the public database PubChem (or FooDB if data were not available in PubChem). We compiled these files into a single sdf file (Additional methods D). As the second step, we used the complied sdf file as input for two chemoinformatic tools: R package *rcdk* (which relies on the CDK Java library for chemoinformatics) (Guha and Charlop-Powers 2016) and the online ChemMine Tool (Backman et al. 2011, http://chemminetools.ucr.edu/tools/). When using *rcdk*, we applied standard and extended fingerprinting methods to calculate pairwise dissimilarity. This step resulted in two chemical dissimilarity matrices D_C (Fig. 4b), denoted chemical dissimilarities *a* and *b*, from standard fingerprinting method and extended fingerprinting methods, respectively. Using the hierarchical
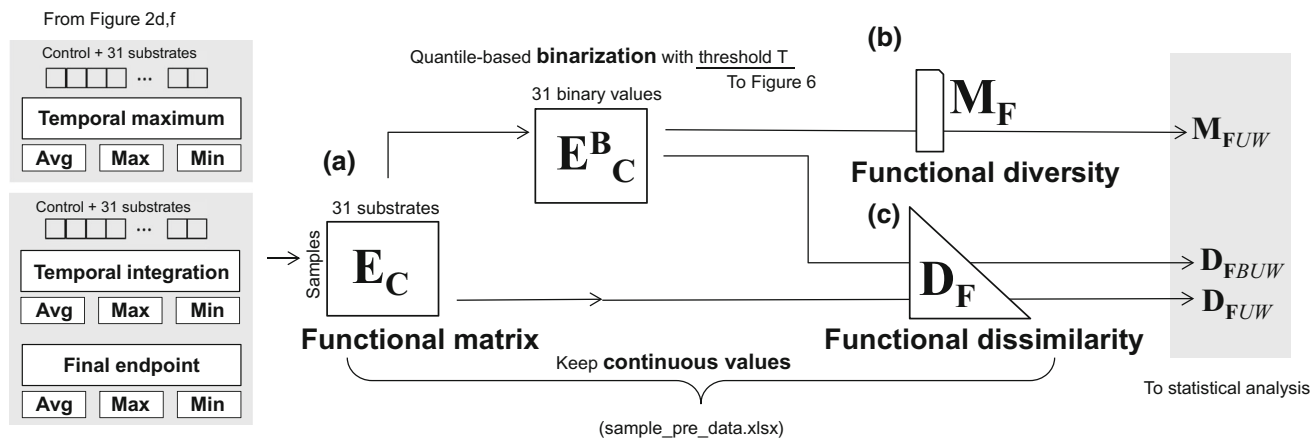
Fig. 3 Diagram illustrating different approaches for calculating multifunctionality and functional dissimilarity. **a** Functional matrix $\mathbf{E_C}$, with number of samples × 31 substrate values (normalized by control) was obtained from each of nine different methods for processing raw data (Fig. 2d, f). Binarized functional matrix $\mathbf{E_C^B}$ was obtained through quantile-based binarization and used to obtain **b** the functional diversity vector $\mathbf{M_F}$, integer (0–31) multifunctionality × number of samples (unweighted multifunctionality, $\mathbf{M}_{FUW}$), or **c** the matrix of functional dissimilarity among samples $\mathbf{D_F}$ (binarized unweighted functional dissimilarity, $\mathbf{D}_{FBUW}$). Alternatively, continuous values of $\mathbf{E_C}$ were used to obtain the unweighted functional dissimilarity $\mathbf{D}_{FUW}$
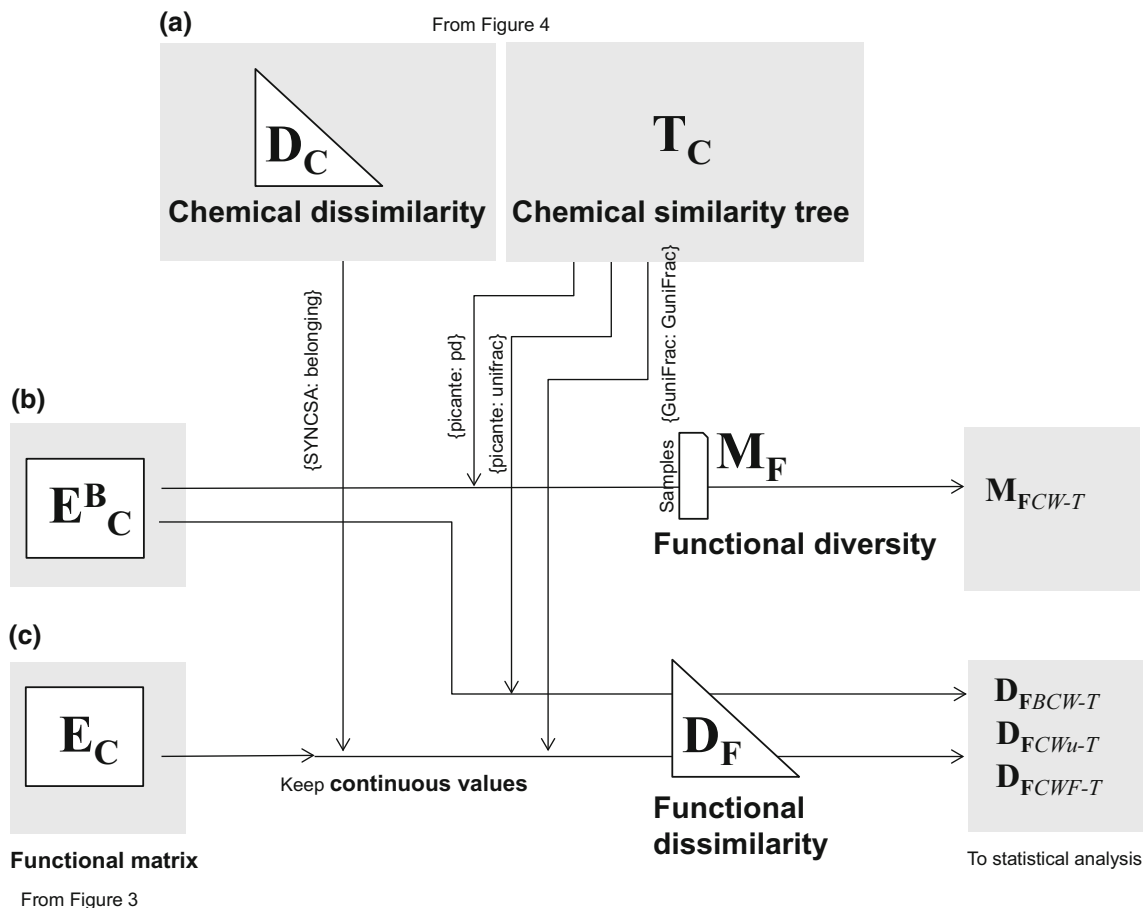

Fig. 4 Diagram illustrating the chemoinformatic approach to calculating chemical similarity and its clustering tree. **a** Two-dimensional structural data for 31 substrates were obtained from web databases. **b** From these structural data, the matrix for chemical dissimilarity among the 31 substrates $\mathbf{D_C}$ was obtained by using chemoinformatic tools (R function and online tools). **c** Chemical dissimilarity matrix was transformed into the hierarchically clustered chemical similarity tree $\mathbf{T_C}$ and converted into phylogenetic tree format. Open and filled circles are hypothetical examples of functional composition for communities A and B, respectively

clustering function in ChemMine tool, we obtained another pairwise dissimilarity matrix (available in SI), denoted as chemical dissimilarity c.

As the third step, we converted dissimilarity matrix $\mathbf{D_C}$ into a hierarchically clustered $\mathbf{T_C}$ tree with the format of phylogenetic tree (.ph) (Fig. 4c). To obtain the

Fig. 5 Diagram illustrating different approaches for incorporating chemical information into functional indices. **a** Either the chemical dissimilarity matrix or chemical similarity tree may be used to compute chemically weighted functional indices. **b** From the binarized functional matrix $\mathbf{E}_C^B$, the chemically weighted functional diversity (multifunctionality) vector ($\mathbf{M}_{FCW-T}$) or the chemically weighted functional dissimilarity matrix ($\mathbf{D}_{FBCW-T}$) can be obtained by using the function *pd* or *unifrac*, respectively, in the *picante* library. **c** From the functional matrix with continuous values $\mathbf{E}_C$, the chemically weighted functional dissimilarity matrix can be obtained in two different ways, by using the function *GuniFrac* in the *GuniFrac* library ($\mathbf{D}_{FCWu-T}$) or the function in the *SYNCSA* library ($\mathbf{D}_{FCF-T}$). The subscript T represents the chemical dissimilarity **a**, **b** or **c**

most informative tree, we chose the method that realized the highest cophenetic correlation between distances on tree $\mathbf{T}_C$ and distance matrix $\mathbf{D}_C$ (Petchey and Gaston 2006) among the eight clustering methods in function *hclust* in the *vegan* package (Dixon 2003; Oksanen et al. 2017). These steps resulted in three tree-shapes (ESM2: Fig. S1). The tree-a and tree-b were highly correlated (Mantel correlation $r = 0.972$), whereas tree-c was less correlated with the others (correlation with tree-a: $r = 0.3693$, with tree-b: $r = 0.3716$).

Concept of chemically weighted index

To describe the ecological meaning of chemical dissimilarity weighting, we considered the binary multifunctionality ($\mathbf{M}_{FUW}$) as an example. Suppose that we have two microbial communities with binarized multifunctionality values equal to 10 (i.e., both microbial com-

munities can decompose 10 carbon substrates but with different combinations). Considering the degree of chemical dissimilarity between these 10 substrates (Fig. 4c), we judge that the set of 10 substrates in community B has higher diversity than the set in community A. Therefore, the chemically weighted multifunctionality of A is smaller than that of B. The same rationale can be applied for functional dissimilarity (Roberts 1986; Lozupone et al. 2011).

Calculating chemically weighted multifunctionality and functional dissimilarity

We applied the concept described by Faith (1992) to chemical similarity trees $\mathbf{T}_C$ (Fig. 5a) and binarized functional matrix $\mathbf{E}_C^B$ (Fig. 5b) through the function *pd* in R package *picante* (Kembel et al. 2010), resulting in the chemically weighted multifunctionality vector

M_{FCW-T} ($T = a$, b, or c) (Fig. 5b). We used three methods to convert the functional matrix (E_C^B or E_C) into the chemically weighted functional dissimilarity matrix. For the first method, we applied the Unifrac distance (picante: unifrac) to the chemical similarity trees T_C (Fig. 5a) and the binarized functional matrix E_C^B (Fig. 5b), resulting in the binarized chemically weighted functional dissimilarity matrix D_{FBCW-T} ($T = a$, b, or c) (Fig. 5c). For the second method, we applied the weighted Unifrac distance (GuniFrac: GuniFrac) (Chen 2012) with two different weights ($u = 0.5$ and 1.0) to the chemical similarity trees T_C (Fig. 5a) and continuous-valued functional matrix E_C (Fig. 5c), generating the chemically weighted functional dissimilarity matrix D_{FCWu-T} ($T = a$, b, or c, $u = 0.5$, or 1). For the third method, we directly used D_C (Fig. 5a) to convert the continuous value matrix E_C via fuzzy-weighting (SYNCSA: belonging) (Roberts 1986 and equation in Additional methods D) into the weighted functional matrix. This weighted functional matrix could be converted to a functional dissimilarity matrix via, for example, the Bray–Curtis dissimilarity measure D_{FCWF-T} (chemically weighed functional dissimilarity matrix by fuzzy weight: $T = a$, b, or c) (Fig. 5c).

Statistical analysis

Statistical models

Depending on the target index (univariate vector M_F for multifunctionality or multivariate matrix D_F for functional dissimilarity) and dataset (forest soil or aquatic microcosm system), we prepared different statistical models. For the data from aquatic microcosm experiments, we only had multifunctionality vectors. We tested the hypothesis that the multifunctionality of the community would decrease linearly with decreasing functional gene diversity in the community ($M_F \sim$ reduction of functional gene diversity). For the data from forest soil experiments, we had both M_F and D_F. We hypothesized that both multifunctionality and functional dissimilarity could be explained by treatment (control or trenching), month, and their interactions. Therefore, we had the univariate linear model ($M_F \sim$ treatment + month + treatment × month) and multivariate models ($D_F \sim$ treatment + month + treatment × month), respectively. For the multivariate models, we used distance-based redundancy analysis (db-RDA) with function *capscale*. For the univariate linear model, we used R^2 values to represent the performance of each model. For db-RDA, we used the fraction of the constrained variation relative to the total variation. We do not show results from PERMANOVA because the resulting R^2 values using the function *adonis* were identical to the ratios of constrained relative to total variations for db-RDA (see Additional methods B and ecopl_comparison_EcolRes.R).

Permutation test for chemically weighted indices

A higher R^2 value from the chemically weighted index than from the chemically unweighted baseline does not automatically imply that the incorporation of chemical similarity improves statistical power. Theoretically, even a randomly generated dissimilarity matrix D_C or dissimilarity tree T_C can generate high R^2 values because the random weighting could reduce data dispersion, resulting in higher R^2 values, as occurs with logarithmic transformation. To exclude this possibility, we shuffled each element in dissimilarity matrix D_C and generated dissimilarity trees. These randomly generated matrices and trees were used for the same statistical models to obtain permutated R^2 values (R^2_{perm}) and compared with the original observed R^2 values (R^2_{obs}). With 9999 permutations, we calculated the probabilities of $R^2_{perm} \geq R^2_{obs}$ (denoted $P_{perm,U}$) and of $R^2_{perm} \leq R^2_{obs}$ (denoted $P_{perm,L}$). When $P_{perm,U}$ or $P_{perm,Ls}$ was less than 0.025, we interpreted R^2_{obs} as being significantly different from the random case (two-tailed test; see Additional methods D for way of calculation). Note, however, this permutation test cannot be used to compare results with and without chemical information.
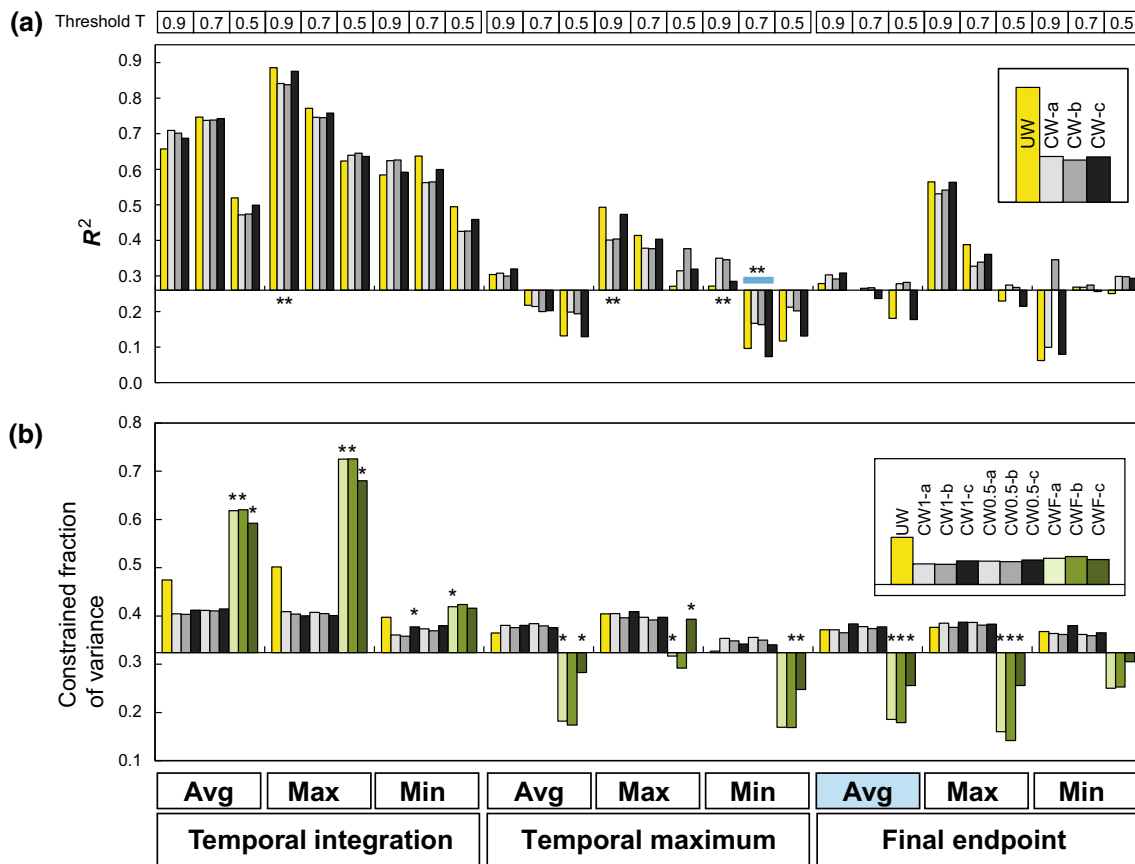
Results

When applying different calculation methods to the binarized multifunctionality from the forest soil samples, the temporal integration method performed best (highest R^2 values) among the temporal integration, temporal maximum, and final endpoint methods (Fig. 6a), for the chemically unweighted approaches (M_{FUW}). The temporal maximum method performed better than the final endpoint method. Among the average, maximum, and minimum values of the triplicates, results using the maximum values were generally the best. These results indicate that the commonly used method (i.e., final endpoint method with triplicate averaging) gave the lowest performance. However, for microcosm experiments, using the minimum value of the triplicates (Fig. S2c in ESM2) performed better than using the average or maximum value (Fig. S2a, b in ESM2).

Using chemically weighted multifunctionality did not improve the performance for forest soil samples (Fig. 6 a) or microcosm experiments (Fig. S2 in ESM2). The statistical power of the linear model with multifunctionality calculated from chemically weighted data (M_{FCW-T}) was not significantly different from those calculated with the randomly generated chemical weights ($P_{perm} > 0.05$), with very few cases in which R^2 values were significantly different from the random case.

Finally, multiple methods were applied to calculate functional dissimilarity indices for the forest soil experiments. Using integration with the maximum value of triplicates performed best in both binarized (Fig. S3 in ESM2) (D_{FBUW}) and continuous (Fig. 6b) (D_{FUW})

(a) Threshold T | 0.9 | 0.7 | 0.5 | 0.9 | 0.7 | 0.5 | 0.9 | 0.7 | 0.5 | 0.9 | 0.7 | 0.5 | 0.9 | 0.7 | 0.5 | 0.9 | 0.7 | 0.5 | 0.9 | 0.7 | 0.5 | 0.9 | 0.7 | 0.5 | 0.9 | 0.7 | 0.5

Fig. 6 Results of statistical models linking multifunctionality and functional composition with month and treatment effects in forest soils. **a** Statistical power (R^2) of the linear model (binarized multifunctionality $\sim$ treatment $\times$ month) for different calculation methods. **b** Statistical power (constrained fraction of variance) of the distance-based RDA (redundancy analysis) model (functional dissimilarity $\sim$ treatment $\times$ month) for different calculation methods. Vertical and horizontal axes cross at a position corresponding to the average statistical power from default calculation methods (i.e., "Final endpoint and taking average of triplicates") in each panel. T values ($= 0.9$, 0.7, and 0.5) represent quantile-base threshold for banalization. Bar with asterisk indicates $P_{\text{perm,U}} < 0.025$ or $P_{\text{perm,Ls}} < 0.025$. Examples of permutation distribution from a portion of the results (blue bar in panel **a**) are shown in Fig. S6. The abbreviations are the same as in Fig. 5

functional dissimilarity. Although including information of chemical structure did not yield better performance in the binarized cases (Fig. S3 in ESM2), fuzzy-weighting of continuous color density data gave results that were statistically different from those calculated with randomly generated chemical weights (Fig. 6b). Only when the data were integrated (i.e., temporal integration method) did the fuzzy-weighting method result in higher explanatory power than the chemically unweighted cases.

Discussion

Overview

We provide a framework to evaluate performances of multiple calculation methods for improving the statistical power of EcoPlate incubation experiments. The statistical power of the temporal integration method was greater than the power obtained from using data only on the final date of incubation (Figs. 6, S3 in ESM2). This result supports our first hypothesis that considering the time evolution of color development would improve the quantification of multifunctionality and functional composition. Using the maximum value for each substrate was the best choice for data processing of the triplicate data within an EcoPlate for the forest soil samples (Figs. 6, S2 in ESM2), whereas the minimum value was the best choice for aquatic microcosms (Fig. S3 in ESM2). The inconsistency of the statistical performance from different processing of the triplicate may indicate the need to identify the best solution for a given system following our statistical framework. Or, it is also possible that more conclusive recommendation for best method may be reached if larger sets of data are examined in the future.

For the second hypothesis, the ability of chemical dissimilarity information to improve statistical power depended on how the data were processed. When bina-

rized values were used for multifunctionality and functional dissimilarity, the incorporation of chemical dissimilarity information did not improve statistical performance (Figs. 6a, S2, S3 in ESM2). When continuous values were used to evaluate functional dissimilarity employing the temporal integration method, the incorporation of chemical information via fuzzy-weighting improved the results, whereas the generalized Unifrac distance did not (Fig. 6b). In contrast, with nonintegrated data (temporal maximum and final endpoint scenarios), fuzzy-weighting worsened the statistical performance.

One implication of this result is that fuzzy-weighting should not be used in the absence of daily measurements during incubation. However, the reason why the statistical performance was worse with fuzzy-weighting remains unclear. Another implication of our result is that differences in microbial functions under different conditions become less clear after incubation for several weeks, and that such a functional convergence can be more clearly detectable with the fuzzy-weighting method.

Practical remarks and cautionary notes on our method

When applying our protocol and the R script to new datasets, we recommend researchers to carefully compare the performances of all available calculation methods. Our results do not imply that the best approach found in our case study is the best for all datasets; rather, the properties of the dataset and statistical model (hypothesis) should be carefully considered. For our datasets, we assumed that multifunctionality and functional composition should be different depending on explanatory factors (treatment, month, and gene diversity), and we tried to find the method that generated the highest statistical power. We strongly recommend comparing the performances of the proposed diverse calculation methods for each dataset (c.f. Anderson et al. 2011). Furthermore, our method does not resolve the problem of temporal changes in species composition along the incubation period. In addition, we used several threshold values for calculating binary multifunctionality. We recommend trying a continuous change of the threshold value (Byrnes et al. 2014).

Resource availability

For comparing methods, we used the R environment, with the script attached in ESM1 (Additional methods B–D) and ESM3. All of the results, including analyses in the ESM1, could be reproduced. Input datasets for the script were the raw data of 96-well EcoPlate color development patterns in text format. One does not need to perform any pre-calculation using the microplate reader software. Before applying the R script, one should check the time evolution of AWCD during incubation period. Instability of the AWCD might be a sign of malfunction of incubation, due, for instance, to drought of well waters or fluctuating temperature.

Theoretical remarks and future directions

While previous studies have used the average values of triplicates within a plate, we found that the maximum values of triplicates can gave better statistical power in the case of soil experiment. It is reasonable to assume that the EcoPlate color development pattern represents the potential functionality rather than in situ-realized functional rates. Therefore, the maximum of the triplicate likely better represents the potential (maximum) metabolic rate of the community for each substrate than the average.

It is not immediately obvious why the minimum of the triplicate performed better for the aquatic microcosm samples. The integrating method gave higher statistical power by distinguishing fast and slow color development rates even when the maximum color density was identical. One question regards how long the optimal integration period should be. If the period is too long after maximum color is achieved, then the rate information will be masked. In addition, long incubation, which would be necessary for the natural samples from low temperature environments (e.g., La Ferla et al. 2017), potentially confounds color development pattern due to the production of secondary metabolites by incubated bacteria or decomposition (oxidation) of the reduced tetrazolium dye, as well as the temporal changes in species composition (see "Introduction"). Our additional analysis demonstrated that intermediate periods (5–10 days) gave the highest R^2 values from the soil samples (Fig. S4 in ESM2). However, the optimal choice may be highly dependent on datasets and incubation temperature.

Another question regards how chemical dissimilarity information improved statistical power. Chemical similarity calculated from the two-dimensional molecule structure does not necessarily imply similarity in interactions between the chemical and organisms (Todeschini et al. 2012). In fact, the shape of the similarity tree is highly dependent on the method (Fig. S1 in ESM2). In addition, we can generate similarity trees based on the similarity of microbial response to different substrates (Fig. S5 in ESM2). When we compared these two different types of trees (Figs. S1 vs. S5 in ESM2), we found no correlation between them (Mantel correlation on dissimilarity matrices, $P > 0.05$). This could be partly explained by the gap between chemical structural dissimilarity and metabolic dissimilarity. For example, in the chemical dissimilarity tree (Fig. S1a in ESM2), glycogen is clustered with other sugar molecules that require different metabolic pathways to be processed (e.g. Lactose and Cellobiose), while the close relationship between glycogen and glucose-1-phophate in the color development similarity (Fig. S5c in ESM2) linked

to the fact that glucose-1-phosphate is the direct downstream product of glycogen in glycogenolysis. Another confusing result is that indices obtained by using information from a randomly generated similarity tree could give greater statistical power (R^2) than those without chemical information (Fig. S6 in ESM2). This is why the permutation test is needed to confirm if results with chemical information are statistically different from those with random trees (Fig. S6 in ESM2).

Future research should focus on improving our method of calculating chemical dissimilarity. To this end, we propose two methods. First, the similarity of microbial response to different substrates (Fig. S5 in ESM2) could be better defined if the EcoPlate color development patterns from many isolate monocultures rather than environmental assemblages were to be used. Data compiled from past publications and/or additional experiments using isolates will be needed. Second, the similarity could be better defined if we were to focus on the metabolic pathways involved in metabolism of each substrate (e.g., KEGG; Kanehisa and Goto 2000). Greater overlap between metabolic pathways could indicate higher similarity in microbial responses to different carbon substrates. Once we obtain a highly reliable tool to evaluate similarity between the 31 substrates in EcoPlate, we could apply this tool to FF and GN plates (95 substrates) (Preston-Mafham et al. 2002), and to much more diverse chemical substrates for proposing a new combination of 31 or 95 substrates to better characterize microbial metabolism. Similarly, our approach of chemical-similarity weighting could be applied to plant metabolites to improve characterization (e.g., of plant defense chemical diversity). These methods will be developed for better quantification of the functional patterns of various types of communities.

References

Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL, Sanders NJ, Cornell HV, Comita LS, Davies KF, Harrison SP, Kraft NJB, Stegen JC, Swenson NG (2011) Navigating the multiple meaning of beta diversity: a road map for the practicing ecologist. Ecol Lett 14:19–28. https://doi.org/10.1111/j.1461-0248.2010.01552.x

Backman TW, Cao Y, Girke T (2011) ChemMine tools: an online service for analyzing and clustering small molecules. Nucleic Acids Res 39:W486–W491. https://doi.org/10.1093/nar/gkr320

Baldwin IT, Halitschke R, Kessler A, Schittko U (2001) Merging molecular and ecological approaches in plant–insect interactions. Curr Opin Plant Biol 4:351–358

Byrnes JEK, Gamfeldt L, Isbell F, Lefcheck JS, Griffin JN, Hector A, Cardinale BJ, Hooper DU, Dee LE, Duffy JE (2014) Investigating the relationship between biodiversity and ecosystem multifunctionality: challenges and solutions. Methods Ecol Evol 5:111–124. https://doi.org/10.1111/2041-210X.12143

Chao A, Chiu C-H, Jost L (2014) Unifying species diversity, phylogenetic diversity, functional diversity and related similarity/differentiation measures through Hill numbers. Annu RevEcol Evol Syst 45:297–324. https://doi.org/10.1146/annurev-ecolsys-120213-091540

Chen J (2012). GUniFrac: Generalized UniFrac distances. R package version 1.0. http://CRAN.R-project.org/package=GUniFrac

Choi KH, Dobbs FC (1999) Comparison of two kinds of Biolog microplates (GN and ECO) in their ability to distinguish among aquatic microbial communities. J Microbiol Methods 36:203–213

Consonni V, Todeschini R (2012) New similarity coefficients for binary data. MATCH Commun Math Comput Chem 68:581–592

Dixon P (2003) VEGAN, a package of R functions for community ecology. J Veg Sci 14:927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x

Faith D (1992) Conservation evaluation and phylogenetic diversity. Biol Conserv 61:1–10. https://doi.org/10.1016/0006-3207(92)91201-3

Floris M, Manganaro A, Nicolotti O, Medda R, Mangiatordi GF, Benfenati E (2014) A generalizable definition of chemical similarity for read-across. J Cheminform. https://doi.org/10.1186/s13321-014-0039-1

Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination of species data. Oecologia 129:271–280. https://doi.org/10.1007/s004420100716

Garland JL, Campbell CD, Mills AL (2007) Physiological profiling of microbial communities, vol 11. ASM manual of environmental microbiology. ASM, Washington, pp 126–138. https://doi.org/10.1128/9781555815882.ch11

Giovannoni SJ, Stingl U (2005) Molecular diversity and ecology of microbial plankton. Nature 437:343–348

Guha R (2007) Chemical informatics functionality in R. J Stat Softw. https://doi.org/10.18637/jss.v018.i05

Guha R, Charlop-Powers Z (2016) Package 'rcdk' ftp://202.38.95.110/CRAN/web/packages/rcdk/rcdk.pdf

Hai DN, Duc HT, Hanh NK, Bettarel Y, Lam NN (2016) Analysis of community level physiological profile of bacteria in NHA Trang bay in dry season applying BIOLOG ECOPLATES. Proceedings of vast-IRD symposium on marine science

Hsieh IF, Kume T, Lin MY, Cheng CH, Miki T (2016) Characteristics of soil CO_2 efflux under an invasive species, Moso bamboo, in forests of central Taiwan. Trees 30:1749–1759. https://doi.org/10.1007/s00468-016-1405-6

Kanehisa M, Goto S (2000) KEGG Koto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30

Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO (2010) Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464

Konopka A, Oliver L, Turco RF Jr (1998) The use of carbon substrate utilization patterns in environmental and ecological microbiology. Microb Ecol 35:103–115

Kuhlisch C, Pohnert G (2015) Metabolomics in chemical ecology. Nat Prod Rep 32:937–955. https://doi.org/10.1039/C5NP00003C

La Ferla R, Azzaro M, MichaudL Caruso G, LoGiudice A, Paranhos R, Cabral AS, Conte A, Cosenza A, Maimone G, Papale M, Rappazzo AC, Guglielmin M (2017) Prokaryotic abundance and activity in permafrost of the northern Victoria land and upper Victoria Valley (Antarctica). Microb Ecol 74:402–415

Laliberte E, Legendre P (2010) A distance-based framework for measuring functional diversity from multiple traits. Ecology 91:299–305

Lawley T, Bell C (1998) Kinetic analyses of Biolog community profiles to detect changes in innoculum densities and species diversity of river bacterial communities. Can J Microbiol 44:588–597

Lin MY, Hsieh IF, Lin PH et al (2017) Moso bamboo (*Phyllostachys pubescens*) forests as a significant carbon sink? A case study based on 4-year measurements in central Taiwan. Ecol Res. https://doi.org/10.1007/s11284-017-1497-5

Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R (2011) UniFrac: an effective distance metric for microbial community comparison. ISME J 5:169–172. https://doi.org/10.1038/ismej.2010.133

Miki T, Yokokawa T, Matsui K (2014) Biodiversity and multi-functionality in a microbial community: a novel theoretical approach to quantify functional redundancy. Proc Royal Soci B. https://doi.org/10.1098/rspb.2013.2498

Minamoto T, Yamanaka H, Takahara T, Honjo MN, Kawabata Z (2012) Surveillance of fish species composition using environmental DNA. Limnology 13:193–197

Moran MA (2015) The global ocean microbiome. Science 350 0, aac8455 https://doi.org/10.1126/science.aac8455

Muniz S, Lacarta J, Pata MP, Jimenez JJ, Navarro E (2014) Analysis of the diversity of substrate utilization of soil bacteria exposed to Cd and earthworm activity using generalised additive models. PLoS One. https://doi.org/10.1371/journal.pone.0085057

Nikolova N, Jaworska J (2004) Approaches to measure chemical similarity—a review. QSAR Comb Sci 22:1006–1026

Oksanen J, Guillaume Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, M. Stevens MHH, Szoecs E, Wagner H (2017). vegan: Community Ecology Package. R package version 2.4-3. http://CRAN.R-project.org/package=vegan

Osono T (2007) Ecology of ligninolytic fungi associated with leaf litter decomposition. Ecol Res 22:955–974. https://doi.org/10.1007/s11284-007-0390-z

Petchey OL, Gaston KJ (2006) Functional diversity: back to basics and looking forward. Ecol Lett 9:741–758. https://doi.org/10.1111/j.1461-0248.2006.00924.x

Pillar VD, Duarte LDS (2010) A framework for metacommunity analysis of phylogenetic structure. Ecol Lett 13:587–596. https://doi.org/10.1111/j.1461-0248.2010.01456.x

Preston-Mafham J, Boddy L, Randerson PF (2002) Analysis of microbial community functional diversity using sole-carbon-source utilization profiles—a critique. FEMS Microbiol Ecol 42:1–14. https://doi.org/10.1111/j.1574-6941.2002.tb00990.x

Ramette A (2007) Multivariate analyses in microbial ecology. FEMS Microbiol Ecol 62:142–160. https://doi.org/10.1111/j.1574-6941.2007.00375.x

Roberts DW (1986) Ordination on the basis of fuzzy set theory. Vegetatio 66:123–131. https://doi.org/10.1007/BF00039905

Siggins A, Gunnigle E, Abram F (2012) Exploring mixed microbial community functioning: recent advances in metaproteomics. FEMS Microbiol Ecol 80:265–280. https://doi.org/10.1111/j.1574-6941.2011.01284.x

Stefanowicz A (2006) The biolog plate technique as a tool in ecological studies of microbial communities. Pol J Environ Stud 15:669–676

Todeschini R, Consonni V, Xiang H, Holliday J, Buscema M, Willett P (2012) Similarity coefficients for chemoinformatics data: overview and extended comparison using simulated and real data sets. J Chem Inf Model 52:2884–2901. https://doi.org/10.1021/ci300261r

Villeger S, Mason NW, Mouillot D (2008) New multidimensional functional diversity indices for a multifaceted framework in functional ecology. Ecology 89:2290–2301

Willett P, Barnard JM, Downs GM (1998) Chemical similarity searching. J Chem Inf Comput Sci 38:983–996. https://doi.org/10.1021/ci9800211

Woese CR, Kandler O, Wheelis MC (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA 87:4576–4579

Zhou J, Liu W, Deng Y et al (2013) Stochastic assembly leads to alternative communities with distinct functions in a bioreactor microbial community. mBio. https://doi.org/10.1128/mbio.00584-12