bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Higher-order dissimilarity in biodiversity:
Identifying dissimilarities of spatial or temporal dissimilarity structures

Running title: Higher-order dissimilarity

Ryosuke Nakadai[1], Keita Fukasawa[1], Taku Kadoya[1,2], Fumiko Ishihama[1]

[*] Correspondence: Ryosuke Nakadai

[1] Biodiversity division, National Institute for Environmental Studies, Onogawa 16-2, Tsukuba, Ibaraki, 305-8506, Japan

[2] Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Ibaraki, 305-8571, Japan

ORCID

Ryosuke Nakadai 0000-0002-9512-8511

Keita Fukasawa 0000-0002-9563-457X

Taku Kadoya 0000-0002-7816-5484

Fumiko Ishihama 0000-0001-8515-5914

Acknowledgements

We thank Dr. Satoshi Aoki for his advice on the choice of coalescent simulator.

Author contributions

RN, FI, KF, and TK designed the study; RN conceived the idea of higher-order dissimilarity, analysed the datasets, and wrote the first draft of the manuscript, with significant inputs from KF, FI, and TK; KF and RN developed the methodology to quantitatively evaluate the idea

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

introduced in the present study; and FI conceived the basic concept of the methodology. All authors have contributed substantially to the final version of this manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Elucidating biodiversity patterns and their background processes is critical in biodiversity science. Dissimilarity, which is calculated based on multivariate biological quantities, is a major component of biodiversity. As the availability of spatial and temporal biodiversity information increases, the scope of dissimilarity studies has been expanded to cover various levels and types of spatio-temporal biodiversity facets (e.g. gene, community, and ecosystem function), and diverse pairwise dissimilarity indices have been developed. However, further development of the dissimilarity concept is required in comparative studies on spatio-temporal structures of biodiversity compositional patterns, such as those exploring commonalities of biogeographical boundaries among taxa, compared to the conventional ones to consider higher dimensions of dissimilarity: dissimilarity of dissimilarity structures. This study proposes a novel and general concept, high-order dissimilarity (HOD), for quantitatively evaluating the dissimilarities of spatial or temporal dissimilarity structures among different datasets, proposes specific implementation of HOD as operational indices, and illustrates potential resolution of scientific and practical questions by means of HOD. Our conceptual framework on HOD extends the existing framework of biodiversity science, and is versatile, with many potential applications in the acquisition of more valuable information from ever-increasing biodiversity data.

Keywords:

autocorrelation, beta diversity, dissimilarity matrix, genetic differentiation, multivariate analysis

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Nature's patterns are ubiquitous regardless of the level of biological organisation, and the elucidation of their determinants has been a long-standing and fundamental issue in biodiversity science. Dissimilarity is a key element of such patterns, and has been studied in terms of spatial and temporal differences and changes in various biological entities, mostly at the same level of biological organisation (Anderson et al. 2011). This encompasses differences and changes in genetic alleles, species composition, interaction networks, and ecosystem functions (Whittaker 1960, 1972; Raymond and Rousset 1995; Miki et al. 2018; Mori et al. 2015; Poisot et al. 2012). The scope of dissimilarity is continuously expanding, covering various levels and types of biodiversity facets, which has stimulated the development of conceptual and analytical frameworks for the analysis of the determinants of dissimilarity patterns using various dissimilarity indices (Koleff et al. 2003; Lozupone and Knight 2005; Baselga 2010; Rocchini et al. 2018; Mammola and Cardoso 2020). Notably, despite the varieties, all these indices have a common form in that one dissimilarity value is calculated from two vectors of multivariate biological quantities (i.e. pairwise dissimilarity indices), and the calculation of dissimilarity values for all pairs of vectors yields a dissimilarity matrix (Anderson et al. 2011).

Some of the spatio-temporal structures of biodiversity composition patterns, however, are beyond the scope of pairwise dissimilarity and identifying, quantifying, and understanding them necessitates higher-order considerations. In a spatial context, higher-order considerations are useful for the analysis of categorically different dissimilarity matrices, such as among assemblages of different taxonomic groups and among species (i.e. comparative biogeography and phylogeography). For example, biogeographical researchers have attempted to identify common biogeographical borders among taxa (Wallace 1894; Holt et al. 2013; Whittaker et al. 2013; Komaki 2021); similarities or dissimilarities in spatial composition dissimilarities among multiple taxa need to be evaluated. In the temporal context, higher-order consideration would provide a framework for quantitatively evaluating dissimilarity in compositional changes (i.e. temporal asynchrony of compositional changes), such as among sites and different trophic groups. For example, temporal changes in

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

community composition (i.e. temporal beta diversity: Legendre 2019; Nakadai 2022) have recently been expanded to multiple sites at large spatial scales (Magurran et al. 2019; Nakadai 2020; Gotelli et al. 2022), thus necessitating quantitative evaluation of inter-site dissimilarity between temporal dynamics of community composition measured using pairwise comparison. Furthermore, loss and biotic homogenisation in species composition and intraspecific haplotypes have been studied in-depth since the identification of human-induced impacts on biodiversity within the context of conservation (Olden 2006; Valtonen et al. 2017). Moreover, quantifying the temporal change of spatial dissimilarity structure from a baseline at a time point is essential to assessing the impacts. Despite the apparent importance and scientific need for higher-order consideration in dissimilarities, general concepts and formal statistical methods for quantitative comparisons across conventional pairwise dissimilarities have not been fully developed.

To address this limitation, we introduce a general framework for considering dissimilarities at a higher order (hereafter referred to as higher-order dissimilarity; HOD) than that of conventional pairwise dissimilarities. HOD is a natural extension of the pairwise dissimilarity indices and considers differences between dissimilarity matrices. In addition, it can be a novel building block in the studies of the dissimilarities of spatio-temporal structures of biodiversity compositional patterns. First, in the present study, we formally define the novel and general concept, HOD, to quantitatively evaluate dissimilarities of spatial or temporal dissimilarity structures among different datasets (Fig. 1a). Second, we designed the concept of HOD and proposed its implementation as a general statistical method. Third, we demonstrated the advantages of the HOD concept by applying it to actual patterns, such as long-term and/or large-spatial hypothetical monitoring datasets. Finally, we discuss the potential impacts of the concept of “higher-order dissimilarity” and the developed analytical framework on a wide range of research fields, including understanding complex spatio-temporal structures in biodiversity and future methodological challenges.

Higher-order dissimilarity

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In biodiversity science, one dissimilarity value is generally calculated from two vectors of multivariate biological quantities (e.g. community composition or allele frequency). Specifically, a community dissimilarity index (e.g. Bray-Curtis dissimilarity; Odum 1950) and a genetic distance index (e.g. F_{st}; Wright 1969) were calculated based on vectors of species abundance in two target communities and vectors of genetic allele frequency in two target populations, respectively. In addition, multiple vectors are summarised into a matrix called a site-species (or time-species) matrix (e.g. compositional vectors for sites or time points). The calculation of dissimilarity values for all pairs of vectors yields a dissimilarity matrix, which is a conventional pairwise dissimilarity; we call this first-order dissimilarity (Fig. 1a). Here, we extend the first-order dissimilarity to formally define HOD as dissimilarities of spatial or temporal dissimilarity structures (Fig. 1a). We also define 'order' as nested times for calculating dissimilarity, e.g. conventional pairwise dissimilarity is the first order; and dissimilarity of dissimilarity matrices the second order. By definition, the concept can be extended to a higher order where appropriate, such as to the third order (i.e. dissimilarity of second-order dissimilarity matrices).

The concept of HOD is a generalisation of the traditional analytical approach to identifying boundaries and their commonalities in the context of biogeography. In addition, HOD targets the spatial structure, which has been the primary subject of biogeography, and temporal structure. Specifically, to spatially and temporally distinguish HODs, we use the terms 'spatial HOD' and 'temporal HOD' (Fig. 1b). Given the recent rapid growth of spatio-temporal monitoring data, HODs regarding temporal variation in spatial structure (i.e. temporal change in spatial HOD, Fig. 1b) or spatial variation in temporal structure (i.e. spatial change in temporal HOD, Fig. 1b) will be applicable in future analyses. To elucidate the concept, the subsequent sections mainly focus on second-order dissimilarities, which are similar to spatial and temporal HODs.

Implementation of the higher-order dissimilarity

When sampling points comprising two biodiversity datasets are obtained at the same spatial or temporal location and correspond precisely one-to-one, their dissimilarity matrix elements

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

also correspond one-to-one. However, this is not always true. For example, when comparing two phylogeographic patterns, the sample points of each species are usually in different locations. In such a case, the elements of dissimilarity matrices do not correspond one-to-one, and temporal or spatial proximity information is required for comparison. To accommodate the biodiversity dataset with different sampling designs, we classified the calculation methods of HOD with and without spatial or temporal proximity information as types 1 and 2 of the HOD indices, respectively. Type 1 HOD indices were only adopted for exhaustive datasets; therefore, pairs of datasets were collected at identical sampling points. Unlike type 1 indices, type 2 HOD was applicable to an 'irregular dataset' (i.e. a dataset where the two compositional vectors to be compared were observed at different points spatially and temporally, and did not correspond on a one-to-one basis). The most basic calculation of the HOD index is determining the average value of the differences between the dissimilarity matrices of the same pairs of sites or time points (mean absolute differences; Fig. 2), which is a type 1 index. Technically, the HOD index can be calculated as the commonly used Mantel correlation coefficient (Mantel correlation coefficient *r*; Fig. 2) calculated from two spatial or temporal dissimilarity matrices as a type 1 index. While these two indices are technically conventional methods, they are often impractical for calculating the dissimilarity of dissimilarity structures owing to the complexity of the comprehensive collection of data from all sampling points, as we assume for conceptualisation of the present study. Particularly, in our comparison of the temporal HOD using a long-term monitoring dataset, we could not use the two conventional indices if some places of the monitoring sites changed during the period because perfectly matched datasets were assumed in the calculation of their dissimilarity. In the case of spatial HOD, genetic samples of two species for comparison are often taken at different sites, and we cannot apply the two indices in this case. This can be addressed using spatial or temporal information.

Type 2 HOD indices are newly introduced analytical concepts in the present study, and are calculated using two distance matrices of dissimilarity with spatial or temporal information. Here, we introduce two HOD indices of type 2 based on bivariate Moran's *I* and Geary'*C*

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

indices (Eckardt and Mateu 2021). Our concept of HOD is general and is thus capable of any metrics of first-order dissimilarity other than those employed here.

In the following section, we introduce a new calculation method that enables the quantification of HOD by applying the indices of spatial cross-correlation developed in previous studies to the distance space rather than the actual space. First, we define the spatial and temporal distance between two pairs of sampling points that correspond to elements of two dissimilarity matrices to expand the indices of spatial statistics to calculate the HOD. Considering a dissimilarity value from each of two dissimilarity matrices, X_{ij} and X'_{kl}, and then the relevant two pairs of sampling points are at hand which we denote ($\mathbf{s}_i$, $\mathbf{s}_j$) and ($\mathbf{s}'_k$, $\mathbf{s}'_l$). Between ($\mathbf{s}_i$, $\mathbf{s}_j$) and ($\mathbf{s}'_k$, $\mathbf{s}'_l$), there are two exhaustive sets of point-wise distances: ($d_{ik} = |\mathbf{s}_i - \mathbf{s}'_k|$, $d_{jl} = |\mathbf{s}_j - \mathbf{s}'_l|$) and ($d_{il} = |\mathbf{s}_i - \mathbf{s}'_l|$, $d_{jk} = |\mathbf{s}_j - \mathbf{s}'_k|$) (Fig. 2). From these sets of distances, the distance between the pairs of sampling points is expressed as follows:

$$\Delta_{\mathrm{ij,kl}} = \min((d_{ik}^2 + d_{jl}^2)^{1/2}, (d_{il}^2 + d_{jk}^2)^{1/2}).$$

For the special case where $\mathbf{s}_i = \mathbf{s}'_k$ and $\mathbf{s}_j = \mathbf{s}'_l$, this distance metric is equal to 0. Another special case is $\mathbf{s}_i = \mathbf{s}_j$ and $\mathbf{s}'_k = \mathbf{s}'_l$, where the distance becomes $\sqrt{2}$ × (Euclidean distance between $\mathbf{s}_i$ and $\mathbf{s}'_k$). With this generalisation of distance for pairs of sampling points, we can calculate the weight matrix for dissimilarity values and apply the conventional measures of spatial cross-correlation to type 2 HOD. To simplify the calculation of the HOD, we converted dissimilarity matrices into vectors $\mathbf{x} = \mathrm{vec}(\mathbf{X})$ and $\mathbf{x'} = \mathrm{vec}(\mathbf{X'})$, and denoted their sizes as n and n', respectively. Correspondingly, the distance between the pair of sampling points $\mathbf{\Delta}$ was reshaped to a matrix with n rows and n' columns.

Bivariate Moran's I (Wartenberg 1985; Lee 2001; Eckardt and Mateu 2021) is a popular measure of spatial cross-correlation. The specification for dissimilarity values is expressed as follows:

$$I = \frac{nn'}{\sum_i \sum_j w(\Delta_{ij})} \times \frac{\sum_i \sum_j w(\Delta_{ij})(x_i - \mu_x)(x'_j - \mu_{x'})}{\sqrt{\sum_i (x_i - \mu_x)^2 \sum_j (x'_j - \mu_{x'})^2}}, \quad (1)$$

where $w(\Delta_{ij})$ is an arbitrary spatial weight which is a function of Δ_{ij} and x_i and x'_j are the ith and j^{th} elements of $\mathbf{x}$ and $\mathbf{x'}$, respectively. μ_x and $\mu_{x'}$ are the mean values of $\mathbf{x}$ and $\mathbf{x'}$,

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

respectively. Generally, $w(\Delta_{ij})$ is a decreasing or step function of Δ_{ij}. If the dissimilarity values of the point pairs that are close in space (or time) show similar deviations from their means, the positive values are calculated, which indicates that the two sets of data compared have spatially and temporally similar dissimilarity structures, or *vice versa*.

Another option for a spatial cross-correlation measure alternative to Moran's *I* is bivariate Geary's *C* (Eckardt and Mateu 2021):

$$C = \frac{(nn'-1)}{2\sum_i \sum_j w(\Delta_{ij})} \times \frac{\sum_i \sum_j w(\Delta_{ij})(x_i - x'_j)^2}{\sqrt{\sum_i (x_i - \mu_x)^2 \sum_j (x'_j - \mu'_x)^2}} . \quad (2)$$

The definition of the parameters is the same as that in Eq. (1). The bivariate Geary's *C* tends to be large when the squared deviations between two dissimilarities are large, where w_{ij} is large, and correlates negatively with Moran's *I*.

To test the relationships among the HOD indices, we applied these four indices to both a simple sample case and 100 simulated datasets, including 37 hexagon-grided communities (Fig. S1) using the R package, 'mobsim' (May et al. 2018; see Supplementary text 1 for details on simulation). The distance between adjacent hexagons was set to 1, and we applied a step function where the weights were equal to 1 for pairs of sampling points where the points in the pair were identical; or one was identical and the other adjacent (i.e. Δ_{ij} = 0 or 1); the distance between adjacent hexagons was set to 0 for all other site pairs. For all pairs of the four HOD indices, we tested the relationships using Mantel tests, and found that all four indices were highly correlated with each other (Fig. S2); specifically, all absolute values of the Mantel correlation were above 0.9.

To check the properties of the HOD indices, we summarised the specific values of spatial HOD for pairs of simple sample cases (Table 1). The spatial weight function for type 2 HODs was identical to that in the simulation above. Within each type of HOD index, the indices are correlated. Specifically, the correlations between mean absolute differences and mantle *r*, and between bivariate Moran's *I* and bivariate Geary's *C* are positive and negative, respectively. Conversely, the relationships between type 1 and type 2 HOD indices are inconsistent with

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

respect to the robustness of the spatial autocorrelation. Specifically, for pairs that contain at least one dataset with low spatial autocorrelation, the calculated values of HOD differ considerably between types 1 and 2. HOD type 2 is affected by spatial (or temporal) autocorrelation because it is weighted in spatial or temporal distances, unlike HOD type 1, which is based solely on simple dissimilarity values.

All the samples presented in the present study used unscaled values (i.e. absolute values) of dissimilarity to allow for comparisons among similar datasets. When applying the HOD approach to two different types of datasets, it would be appropriate to standardise values within each dissimilarity matrix to compare dissimilarity structures. For example, when calculating the HOD of the genetic structure between different species, a standardisation process is necessary because the absolute value of the genetic distance between differentiated and non-differentiated species is different.

Resolvable questions through the higher-order dissimilarity approaches

Theoretically, HOD concepts are applicable to any type of dissimilarity dataset if the relative position (i.e. sites or time points) can be determined based on the absolute distance between pairs of datasets. For example, the monitoring of the genetic structure is one of the most urgent targets of biodiversity monitoring (Hoban et al. 2021; O'Brien et al. 2022) although no effective monitoring framework has been developed. In this context, the HOD provides an ideal framework for the quantitative evaluation of genetic structure temporally (Fig. 3). As frequently probable examples, we show two hypothetical scenarios of genetic structure changes: one is a stable scenario (orange line and dots) and the other is a genetically homogenised scenario (blue line and dots) in Fig. 3 (see Supplementary Text 2 for detailed information about construction for visualisation). Fig. 3(a) indicates temporal changes in the values of the HOD type 1 index (i.e. mean absolute differences) to time 0 (starting point), and the lower Fig. 2(b) represents mapped overviews of changes in genetic structure over time, from which we calculated the HOD values. The scarcity of such spatially coordinated data on genetic structures over a wide area will soon be addressed, particularly following the

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

development of environmental DNA technology (Tsuji et al. 2022), and our framework will be an essential contribution to genetic diversity monitoring.

To show the advantages of HOD more comprehensively, we summarised the potential applications of HOD in Table 2, including five types of datasets: community composition, genetic frequency, ecosystem function, interaction network, and phylogeny. Most of the cases are straightforward extensions of those shown in Fig. 1b. One special case is that HOD approaches are applicable to phylogenetic patterns, despite the difficulty in extracting single-dimensional values; for example, the output of non-metric multidimensional scaling based on multiple traits. Evidently, HOD approaches can be applied to a wide variety of datasets in community ecology, biogeography, phylogeography, and macroevolution. This fact further emphasises the importance of developing a field of study for HOD.

The main focus has been on comparisons between dissimilarity matrices with similar properties, but this is not necessarily the case. Examples include comparisons among communities of different trophic levels and comparisons between community structures and genetic structures in community members. In traditional approaches, revealing the relative importance of the driving processes of local communities, such as abiotic vs. biotic factors and/or top-down vs. bottom-up processes as drivers of community structures, has been a major concern (Smith et al. 2010; HilleRisLambers et al. 2012). The HOD approach is a different perspective that directly evaluates the similarity of patterns without explicitly considering the driving processes. Low values of HOD among data from different trophic levels of communities and different hierarchies of genetic and community data allowed us to generate novel hypotheses with more direct evidence, as the processes shaping them are considered common among the compared targets (taxonomic differences; Fig. 1b). Conversely, high HOD values suggest that the process of constructing the overall system is complex and contingent. From an application perspective, high values of spatial HOD also suggest that achieving conservation through the establishment of protected areas, setting focal species for conservation, or implementing a single conservation policy would be difficult, as the driving processes differ among taxa and levels of biological entities.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Future challenges

Despite the significance of the concept of the HOD, there are several methodological challenges. In this paper, we discussed HOD, assuming an ideal situation where the target sites (or time points) have been exhaustively examined without any missing values. However, in empirical studies, it is often unsatisfactory, and this is one of the reasons underlying the failure to devise HOD in the past. To overcome this challenge, candidate approaches are broadly divided into two categories. One is the interpolation of missing data through modelling approaches, such as generalised dissimilarity modelling and species distribution modelling. Generalised dissimilarity modelling (GDM) is an approach that specialises in dissimilarities, differences, and distances (Ferrier et al. 2007). If target dissimilarities are modelled by other variables (e.g. environment and geography) using GDM, the predicted outputs can be applied to HOD. Another approach, type 2 HOD, which is weighted by spatial or temporal distances, can theoretically consider missing values. However, further testing is required to determine how many missing values are allowed for a sufficiently accurate evaluation and what the most appropriate spatial or temporal weighting process is for each target system.

In the HOD approaches proposed here, we cannot distinguish between parts of continuous patterns and truly disjunctive patterns in dissimilarities; for example, either simple distance-decay patterns or patterns due to isolation by geographic barriers. Therefore, where the commonality of spatial disconnections is the main concern, combining HOD approaches with existing approaches would be more effective in the identification of disjunct barriers (e.g., Soltis et al. 2006). In temporal structures, approaches separating trend components have been used (Cowpertwait and Metcalife 2009), and potentially combined with such methods, it may be possible to evaluate continuous and disjunctive dissimilarities separately.

In existing studies on dissimilarity, unique conceptual and analytical frameworks have been developed for dissimilarity datasets because of the non-independence problem between dissimilarity values which share an original value (Anderson et al. 2011). To overcome the problem of non-independence among values, statistical methods that deal with dissimilarity matrices, such as the Mantel test (Mantel 1967) and certain developed methods (Lichstein

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2007, Ferrier et al. 2007, Legendre and Legendre 2012, Anderson and Walsh 2013), have been used to calculate p-values by permutation procedures. Permutation-based hypothesis testing approaches adjusted to the HOD framework need to be developed to test the statistical significance of the similarity or dissimilarity between two dissimilarity matrices, which is a critical issue for future research on HOD.

Although we mainly focused on the second-order dissimilarities in the present study, the applicability of third- and higher-order dissimilarities would increase in situations where the amount and dimension of biodiversity information continues to grow. To illustrate the benefit of HOD, we showed temporal changes of genetic structure (Fig. 3). The figure shows a pattern in a single species, and thus, second-order dissimilarity (i.e. space×time). If this type of dataset is available for multiple species, the similarity and dissimilarity of temporal changes in spatial structure in population genetics can be quantitatively evaluated among species using third-order dissimilarities (i.e. space×time×species). This simple example tells us that the breadth of areas that the HOD concept can adopt is rapidly expanding, aligning with the forthcoming explosive increase in available dissimilarity information.

Conclusions

The rapid increase in the availability of spatial and temporal biodiversity information necessitates comparative studies on spatio-temporal structures of biodiversity compositional patterns require further development of the dissimilarity concept over conventional ones to consider higher dimensions of dissimilarity: dissimilarity of dissimilarity structures. Our study introduced the concept of higher-order dissimilarity (HOD), which can account for dissimilarities of spatial or temporal dissimilarity structures. For example, this framework can be applied to various types of broad-scale biodiversity monitoring datasets and enables the evaluation of temporal changes in the spatial HOD to a baseline. Furthermore, the concept is applicable even if the compared datasets have completely different origins (e.g. different taxonomic groups and hierarchies of biological organisation, Fig. 1b), as long as they are summarised into dissimilarity information at their own level. We mainly focused on the biodiversity dataset in the present study; the concept of HOD is applicable to all kinds of

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dissimilarity matrices, not just those based on biodiversity. For example, various contemporary issues arise in the social ecological system, and this HOD will be useful in examining the relationships between different layers, such as the economy, culture, and biodiversity. Therefore, the HOD concept and related approaches pave the way for novel dimensional approaches that deal with large amounts of dissimilarity information.

Data accessibility statement

Simulated data and codes will be uploaded after acceptance of the manuscript to figshare.

References

Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L. et al. (2011) Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecology Letters*, 14, 19–28.

Anderson, M.J., & Walsh, D.C. (2013) PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing? *Ecological Monographs*, 83, 557–574.

Baselga, A. (2010) Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, 19, 134–143.

Cowpertwait, P.S., & Metcalfe, A.V. (2009) Introductory time series with R. Springer Science & Business Media.

Gotelli, N.J., Moyes, F., Antão, L.H., Blowes, S.A., Dornelas, M., McGill, B.J. et al. (2022) Long□term changes in temperate marine fish assemblages are driven by a small subset of species. *Global Change Biology*, 28, 46–53.

Eckardt, M., & Mateu, J. (2021) Partial and semi□partial statistics of spatial associations for multivariate areal data. *Geographical. Analysis*, 53, 818–835.

Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, 13, 252-264.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

HilleRisLambers, J., Adler, P.B., Harpole, W.S., Levine, J.M., & Mayfield, M.M. (2012) Rethinking community assembly through the lens of coexistence theory. *Annual Review of Ecology, Evolution and Systematics*, 43, 2012.

Hoban, S., Campbell, C.D., da Silva, J. M., Ekblom, R., Funk, W.C., Garner, B.A., et al. (2021) Genetic diversity is considered important but interpreted narrowly in country reports to the Convention on Biological Diversity: Current actions and indicators are insufficient. *Biological Conservation*, 261, 109233.

Holt, B.G., Lessard, J.P., Borregaard, M.K., Fritz, S.A., Araújo, M.B., Dimitrov, D. (2013) An update of Wallace's zoogeographic regions of the world. *Science*, 339, 74-78.

Koleff, P., Gaston, K.J., & Lennon, J.J. (2003) Measuring beta diversity for presence–absence data. *Journal of Animal Ecology*, 72, 367-382.

Komaki, S. (2021) Widespread misperception about a major East Asian biogeographic boundary exposed through bibliographic survey and biogeographic meta□analysis. *Journal of Biogeography* 48, 2375-2386.

Lee, S. I. (2001) Developing a bivariate spatial association measure: an integration of Pearson's r and Moran's I. *Journal of Geographical Systems*, 3, 369–385.

Legendre, P. (2019) A temporal beta-diversity index to identify sites that have changed in exceptional ways in space–time surveys. *Ecology and Evolution*, 9, 3500–3514

Legendre, P. & Legendre, L. (2012) Numerical ecology, 3rd edn. *Elsevier*. Amsterdam

Lichstein, J. W. (2007) Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecology* 188, 117–131.

Lozupone, C., & Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, 71, 8228–8235.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27(2_Part_1), 209-220.

May, F., Gerstner, K., McGlinn, D.J., Xiao, X., & Chase, J.M. (2018) mobsim: An R package for the simulation and measurement of biodiversity across spatial scales. *Methods in Ecology and Evolution*, 9, 1401–1408.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Magurran, A.E., Dornelas, M., Moyes, F., & Henderson, P.A. (2019) Temporal β diversity—A macroecological perspective. *Global Ecology and Biogeography*, 28, 1949-1960.

Mammola, S., & Cardoso, P. (2020) Functional diversity metrics using kernel density n□dimensional hypervolumes. *Methods in Ecology and Evolution*, 11, 986–995.

Miki, T., Yokokawa, T., Ke, P. J., Hsieh, I., Hsieh, C. H., Kume, T. et al. (2018) Statistical recipe for quantifying microbial functional diversity from EcoPlate metabolic profiling. *Ecological. Research*, 33, 249-260.

Mori, A. S., Isbell, F., Fujii, S., Makoto, K., Matsuoka, S., & Osono, T. (2016) Low multifunctional redundancy of soil fungal diversity at multiple scales. *Ecology Letters*, 19, 249-259.

Nakadai, R. (2020) Degrees of compositional shift in tree communities vary along a gradient of temperature change rates over one decade: Application of an individual□based temporal beta□diversity concept. *Ecology and Evolution* 10, 13613-13623.

Nakadai, R. (2022) Development of novel temporal beta-diversity indices for assessing community compositional shifts accounting for changes in the properties of individuals. *Ecological Indicators*, 144, 109427.

O'Brien, D., Laikre, L., Hoban, S., Bruford, M.W., Ekblom, R., Fischer, M.C. et al. (2022) Bringing together approaches to reporting on within species genetic diversity. *Journal of Applied Ecology*, 59, 2227–2233.

Odum, E.P. (1950) Bird populations of the Highlands (North Carolina) Plateau in relation to plant succession and avian invasion. *Ecology* 31, 587–605.

Olden, J.D. (2006). Biotic homogenization: a new research agenda for conservation biogeography. *Journal of Biogeography*. 33, 2027-2039.

Poisot, T., Canard, E., Mouillot, D., Mouquet, N., & Gravel, D. (2012) The dissimilarity of species interaction networks. *Ecology Letters*, 15, 1353-1361.

Raymond, M., & Rousset, F. (1995). An exact test for population differentiation. *Evolution* 49, 1280-1283.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Rocchini, D., Luque, S., Pettorelli, N., Bastin, L., Doktor, D., Faedi, N. et al. (2018) Measuring β diversity by remote sensing: A challenge for biodiversity monitoring. *Methods in Ecology and Evolution*, 9, 1787-1798.

Smith, J.E., Hunter, C.L., & Smith, C. M. (2010) The effects of top–down versus bottom–up control on benthic coral reef community structure. *Oecologia*, 163, 497-507.

Soltis, D.E., Morris, A.B., McLachlan, J.S., Manos, P.S., & Soltis, P.S. (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, 15, 4261–4293.

Tsuji, S., Shibata, N., Sawada, H., & Watanabe, K. (2022) Differences in the genetic structure between and within two landlocked Ayu groups with different migration patterns in Lake Biwa revealed by environmental DNA analysis. *Environmental DNA*, https://doi.org/10.1002/edn3.345

Valtonen, A., Hirka, A., Szőcs, L., Ayres, M. P., Roininen, H., & Csóka, G. (2017). Long term species loss and homogenization of moth communities in Central Europe. *Journal of Animal Ecology*, 86, 730–738.

Wartenberg, D. (1985) Multivariate spatial correlation: a method for exploratory geographical analysis. *Geographical Analysis*, 17, 263-283.

Wallace, A.R. (1894) What are zoological regions? *Nature*, 49, 610–613.

Whittaker R.H. (1960) Vegetation of the Siskiyou mountains, Oregon and California. *Ecological Monographs* 30, 279–338

Whittaker R.H. (1972) Evolution and measurement of species diversity. *Taxon*, 21, 213–251

Whittaker, R.J., Riddle, B.R., Hawkins, B.A., & Ladle, R.J. (2013) The geographical distribution of life and the problem of regionalization: 100 years after Alfred Russel Wallace. *Journal of Biogeography*, 40, 2209-2214.

Wright, S. (1969) Genetics and the Evolution of Populations. Vol. 2: The Theory of Gene Frequencies.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure legends

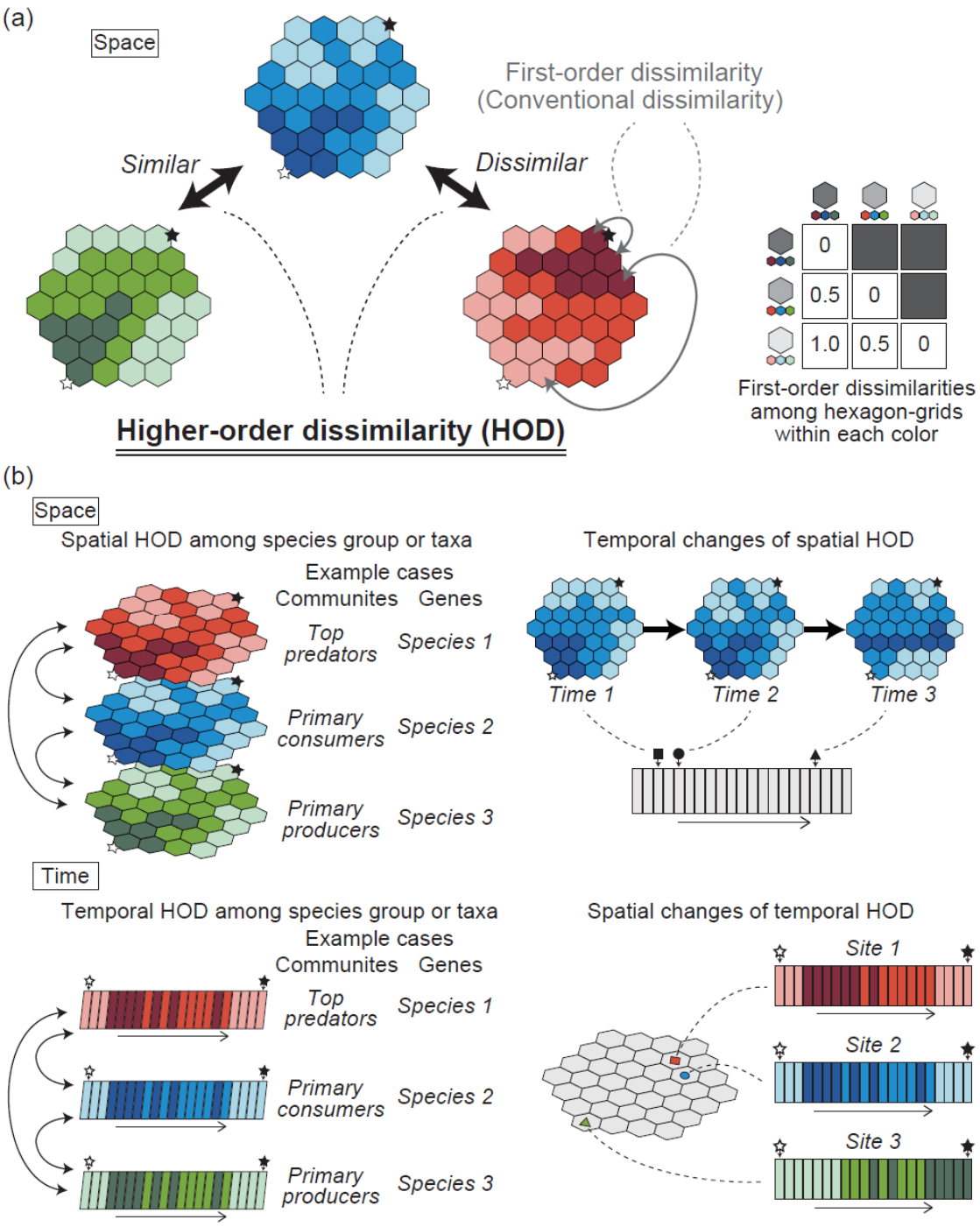
Figure 1 Schematic of higher-order dissimilarity (HOD). A spatial case of HOD is shown in (a). Potential analytical framework for HOD based on spatial and temporal comparisons (i.e., spatial and temporal HOD) are shown in (b). For visualization, only three compositional types are hypothesized, which correspond to three level of darkness in each colour (e.g., blue, green, and red). Specifically, we hypothesized that dissimilarities between sites (or time points) of same level of darkness were zero (i.e., identical composition), those between sites of one darker or lighter difference were 0.5, and those between the darkest and lightest sites were 1.0. The grey arrows indicate conventional dissimilarity (i.e., first-order dissimilarity) and the black arrows indicate HOD. Black star and white star among compared datasets indicate identical sites or time points, respectively.

Figure 2 Schematic of specific analytical procedures for two types of higher-order dissimilarity (HOD), specifically one without spatial information and the other with the information. For visualization, only three compositional types are hypothesized, which correspond to three level of darkness in each colour (e.g., blue, green, and red). Specifically, we hypothesized that dissimilarities between sites (or time points) of same level of darkness were zero (i.e., identical composition), those between sites of one darker or lighter difference were 0.5, and those between darkest and lightest sites were 1.0. The asterisk with 1(*1) showed the specific calculation of distance between pair of sites.

Figure 3 Sample application of higher-order dissimilarity (HOD) to genetic monitoring using two simulated datasets based on two types of scenarios from time 0 to time 100. In both scenarios, the stable conditions are hypothesized by time 30 (shared section in light green). After time 30, genetic structure is stable in one scenario (upper section in orange) but homogenized in the other (lower section in blue). The change of calculated type 1 of HOD index (i.e., mean absolute differences) compared to the value at time 0 is shown in (a). The genetic structure is visualized in (b), and the colours were determined using the two axes of NMDS (details see Supplementary text 2).

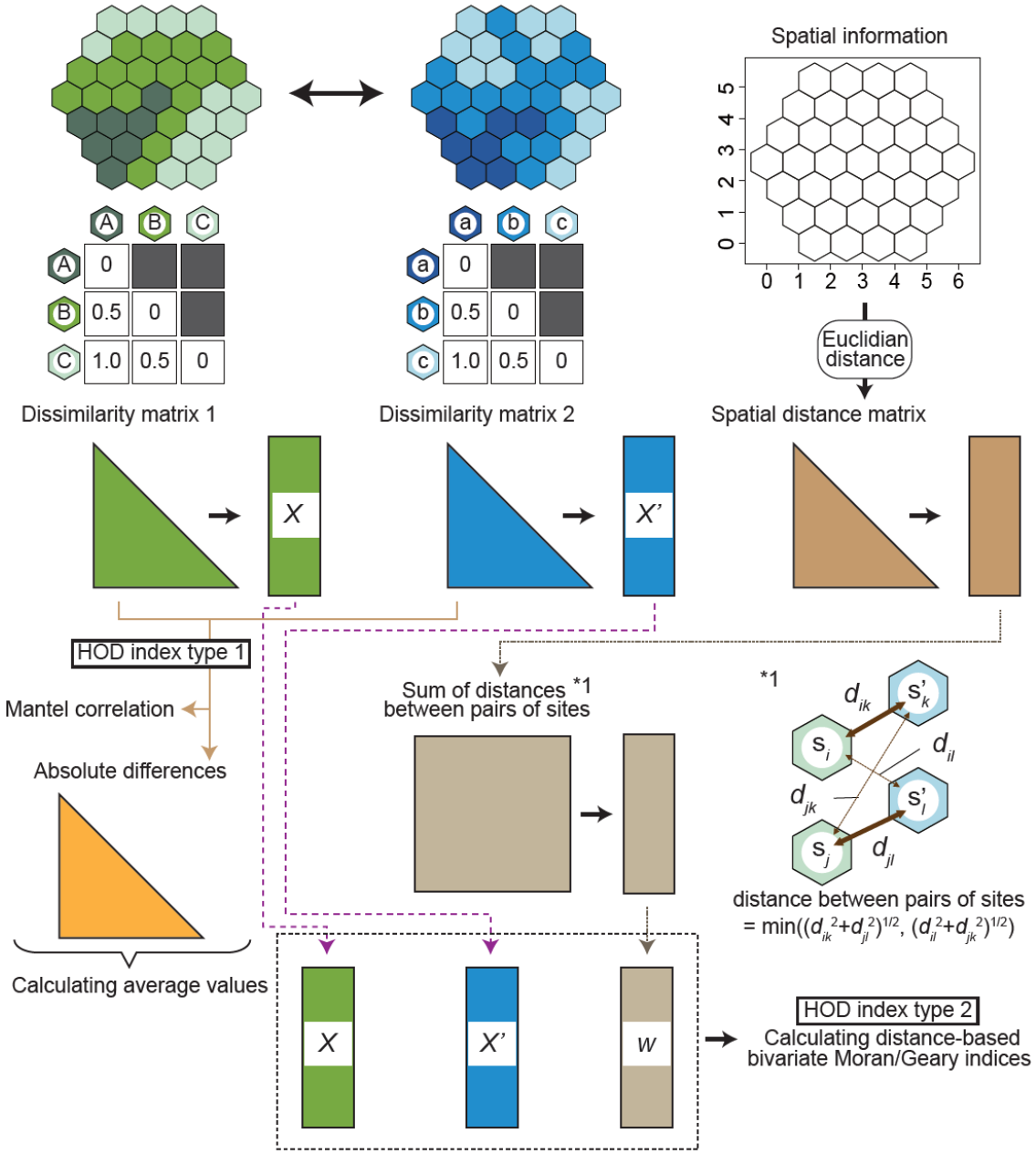
bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1



bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2



bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3

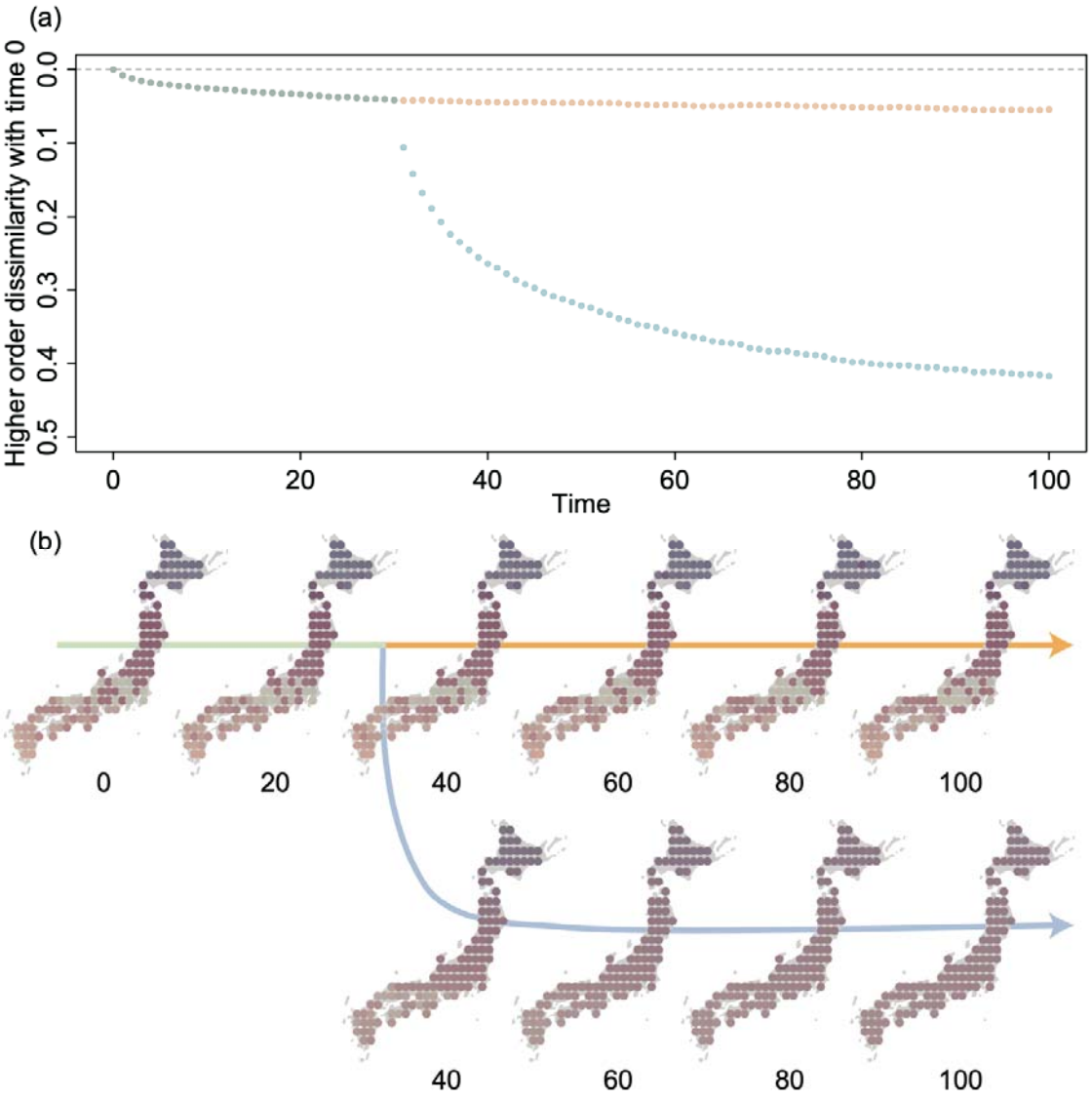
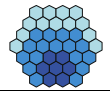


Table 1 Simple sample cases to show properties of HOD indices. Patterns X and Y represent pairs for calculating HODs. Mantel *r* (X) and (Y) indicate the spatial autocorrelation, thus correlation between spatial distances and dissimilarity values for each pattern X and Y. Four HOD indices are shown, which are mean absolute differences and mantel *r* (with *P*-value) as type 1 and bivariate Moran's *I* and Geary's *C* as type 2. For visualization purpose, only three compositional types are hypothesized, which correspond to three level of darkness in each colour (e.g., blue, green, and red). Specifically, we hypothesized that dissimilarities between sites (or time points) of the same level of darkness were zero (i.e., identical composition), those between sites of one darker or lighter difference were 0.5, and those between the darkest and lightest sites were 1.0.

Association	Pattern		Mantel (Space)				HOD type 1			HOD type 2	
	X	Y	*r* (X)	*P*-value	*r* (Y)	*P*-value	Absolute mean differences	Mantel *r*	*P*-value	Bivariate Moran *I*	Bivariate Geary *C*
A-A			0.0367	0.0013	0.0367	0.0013	0.0000	1.0000	0.0000	0.6539	0.3405
B-B			0.0263	0.0162	0.0263	0.0162	0.0000	1.0000	0.0000	0.4783	0.5107
C-C			-0.0173	0.1089	-0.0173	0.1089	0.0000	1.0000	0.0000	0.1594	0.8622
A-B			0.0367	0.0013	0.0263	0.0162	0.3108	0.1759	0.0036	0.1322	0.8594
B-C			0.0263	0.0162	-0.0173	0.1089	0.3649	-0.0886	0.0796	-0.0421	1.0473
C-A			-0.0173	0.1089	0.0367	0.0013	0.3423	0.0301	0.5580	0.0120	0.9959

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A-A'			0.0367	0.0013	0.0367	0.0013	0.3183	0.1318	0.0183	0.1345	0.8599
B-B'			0.0263	0.0162	0.0263	0.0162	0.3453	0.0097	0.8685	-0.0075	0.9964
C-C'			-0.0173	0.1089	-0.0173	0.1089	0.3348	0.0405	0.3932	0.0131	1.0097
A-C'			0.0367	0.0013	-0.0173	0.1089	0.3288	0.0910	0.0809	0.0547	0.9548

93

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2 Potential applications of higher-order dissimilarity (HOD) approach

Target axis of HOD	Targets of comparison	Question
(a) community composition		
Time	Space	How do temporal changes of community composition vary in space? Do closer sites show more similar community compositional change across time?
Time	Taxa	Which taxa share similar changes of community composition in time?
Space	Time	How does spatial composition change over time? When did large change of community compositional structure happen?
Space	Taxa	Which taxa share similar community composition structure in space?
(b) population genetics		
Time	Space	How does temporal genetic frequency change in space? Do closer sites show more similar changes in genetic frequency across time?
Time	Taxa	Which taxa share similar changes of genetic frequency over time?
Space	Time	How does the spatial genetic structure change over time? When did a large change in genetic structure occur?
Space	Taxa	Which taxa share similar changes of genetic frequency in space?
(c) ecosystem function		
Time	Space	How does temporal ecosystem function change in space? Do closer sites show more similar change of ecosystem function across time?
Space	Time	How does spatial genetic structure change over time? When did a large change in the genetic structure occur?
(d) interaction network		
Time	Space	How does the temporal network structure change in space? Do closer sites show more similar changes in network structure across time?

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Space	Time	How does spatial network structure change in time? When did large change of network structure happened?
(e) macroevolution		
Phylogeny	Trait	Which traits share similar evolutionary changes across macroevolutionary history?

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.22.517446; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Appendices

Figure S1 Mantel correlations among absolute mean differences and mantel *r* as type 1 and bivariate Moran's *I* and Geary's *C* as type 2 (9,999 times permutation)

Figure S2 Hexagon grids used in simulation

Supplementary text 1 Detailed information on the simulation of community datasets and calculation of dissimilarity matrices to test the properties of HOD indices.

Supplementary text 2 Detailed information on simulated genetic structure across four major islands in Japan