

# **Effective number of different populations: a new concept and how to use it**

Evsey Kosman<sup>\*</sup>, Frida Feijen<sup>§†</sup> and Jukka Jokela<sup>§†</sup>

<sup>\*</sup> Institute for Cereal Crops Research, School of Plant Sciences and Food Security, The George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel; <sup>§</sup> ETH Zurich, Department of Environmental Systems Science, Institute of Integrative Biology (IBZ), 8092, Zürich, Switzerland; <sup>†</sup> EAWAG, Aquatic Ecology, 8600 Dübendorf, Switzerland.

## **Correspondence**

Evsey Kosman

Email: kosman@tauex.tau.ac.il

## **Running headline**

Effective number of different populations

## **Acknowledgments**

We are grateful to Prof. William B. Sherwin (UNSW Australia) for critical reading of the manuscript and helpful suggestions. We also thank anonymous reviewers and editors for commenting on the manuscript that allowed to improve its quality.

This study is supported by Swiss National Science foundation (grant #310030\_207589 to JJ).

Collection of the data presented in the paper was supported by a grant from the Swiss National Science foundation (grant # 31003A\_166667 to JJ). We are thankful to the reviewers for their constructive criticism and very valuable and beneficial suggestions. We are also indebted to Iliya Bulayevsky for his leading in development of the user-friendly software LOCUS and FDAT.

## **Data and software availability**

22 Data used in this study are available at  
23 <https://datadryad.org/stash/dataset/doi:10.5061/dryad.2ngflvhnw>.  
24 User-friendly software LOCUS and FDAT (Functional Diversity Analysis Tools) can be  
25 downloaded at <https://en-lifesci.tau.ac.il/profile/kosman>. The software needs a programming  
26 environment of the Microsoft.NET Framework, which is an integral Windows component.

## 27 **Conflict of interest statement**

28 All authors declare that they have no conflicts of interest.

## 29 **Author Contributions**

30 Evsey Kosman conceived the idea of measuring beta variation and designed formal methodology of  
31 data analysis; Jukka Jokela controlled the proper biological interpretations of the suggested metrics  
32 and approaches; Frida Feijen and Jukka Jokela collected the data; Evsey Kosman and Jukka Jokela  
33 led the writing of the manuscript. All authors analyzed the data, contributed critically to the drafts  
34 and gave final approval for publication.

35

# Effective number of different populations: a new concept and how to use it

## Abstract

1. Widely used methods to assess population genetic structure and differentiation rely on independence of marker loci. Following the assumption, the common metrics, for example  $F_{ST}$ , evaluate genetic structure by averaging across loci. Common metrics do not use information in the associations among loci at the individual level and are often criticized for failing to measure true differentiation even when loci segregate independently.
2. We introduce a new concept to measure  $\beta$ -variation (Effective Number of Different Populations, ENDP). It requires the following steps: (a) calculation of a proper dissimilarity between genetic profiles of all individuals; (b) calculation of suitable pairwise distances between the samples based on the dissimilarities between individuals; (c) calculation of diversity (in terms of Hill numbers) and dispersion of samples based on the pairwise distances between samples; (d) ENDP is then estimated as a combination of the diversity and dispersion. ENDP estimates  $\beta$ -variation independently of within-sample  $\alpha$ -variation. This new concept differs from the existing standard where  $\beta$ -diversity is estimated based on the ‘partition of variation’ scheme ( $\beta = \gamma - \alpha$  or  $\beta = \gamma/\alpha$ ).
3. Estimates of ENDP are obtained by evaluating information in the available genetic profiles of individuals including association of loci. Therefore, ENDP can be used even in an absence of panmixia.
4. We illustrate the use of this concept by analyzing the population genetic structure of a sexual species (a trematode parasite) occupying connected populations across a broad geographic area. Analysis is complicated by two coexisting cryptic sister clades and the potentially mixed-mating system of this hermaphroditic parasite.

24 **Keywords:** assignment-based distance, beta variation, differentiation, dispersion, diversity, Hill  
25 numbers, *Atriophallophorus winterbourni*

26

## 27 Introduction

28 Discovering the genetic structure of populations is one of the key applications of population genetic  
29 markers. Not surprisingly, methods aimed at assessing the extent of difference among subdivided  
30 populations are numerous and have nearly always been central part of the standard population  
31 genetics toolkits. Historically, the first  $F_{ST}$  measure (Wright, 1951), as its many later analogues,  
32 aimed at understanding the divergence of populations in relation to evolutionary processes (Nei,  
33 1973; Excoffier et al., 1992; Slatkin, 1995). Later, one of the specific applications has been to  
34 estimate the partitioning of genetic variation within and among subdivided populations (Nei &  
35 Chesser, 1983; Hedrick, 2005; Meirmans & Hedrick, 2011). A comprehensive review of methods  
36 aimed at differentiation of molecular diversity with an emphasis on information (entropy) analysis  
37 can be found in Sherwin et al. (2017).

38         The  $F_{ST}$  measures were developed for single loci. Many of the commonly used multilocus  
39 estimates evaluate each locus independently with further averaging across loci ignoring information  
40 that is in the associations between loci (i.e., multilocus genotypes) and between alleles within a  
41 diploid (or polyploid) locus. Thus,  $F_{ST}$  and its relatives ( $G_{ST}$ ,  $G'_{ST}$ ,  $G''_{ST}$ ,  $\varphi_{ST}$ ,  $R_{ST}$ ) are not sensitive to  
42 divergence among populations that exist only due to differences in association of alleles and/or loci  
43 (allele frequencies are equal in all populations). For example, two populations  $P_1$  and  $P_2$  consisting  
44 of individuals with different binary genotypes at four loci (1010 and 0101 in  $P_1$ , and 1111, 1100,  
45 0011 and 0000 in  $P_2$ ) are indistinguishable with  $F_{ST}$  and its relatives, if frequencies of each binary  
46 allele 1 and 0 are equal in  $P_1$  and  $P_2$  (e.g.  $P_1$  and  $P_2$  consist of four individuals each with the above-  
47 mentioned genotypes:  $P_1 = \{1010, 1010, 0101, 0101\}$  and  $P_2 = \{1111, 1100, 0011, 0000\}$ ;  
48 frequencies of all binary alleles equal 0.5).

49         The classical approach works best in a fully recombining panmictic sexual population.  
50 However, the classical approach may work less well in populations with clonal reproduction, a

mixed mating system or where unknown cryptic species coexist. To measure variation and divergence of such populations, new metrics that use information based on associations of loci have emerged during the last decades (Kosman, 1996; Gregorius et al., 2003; Gillet et al., 2004; Kosman & Leonard, 2007; Kosman, 2014). These metrics also include measures aimed at evaluating the extent and significance of differences among populations (Gillet & Gregorius, 2008; Gregorius, 2010; Gillet, 2013; Kosman et al., 2014; Gultyaeva et al., 2020; Czajowski et al., 2021).

Jost (2008) criticized shortcomings of the standard metrics that are commonly called measures of “differentiation” ( $F_{ST}$ ,  $G_{ST}$ ,  $\phi_{ST}$ ,  $R_{ST}$ ) because they can provide unrealistic estimates of the differences in the structure of the populations, especially if the within-population variation is very high. Therefore, using the term “differentiation” for those measures seems inappropriate and confusing. Second, these estimates are unintuitive and can even be misleading (see Jost 2008). To be more specific, it is possible that these measures do not reach their maximum values, could be far away from maximum and approach zero (indication of no differentiation), even for populations that do not share any alleles. The latter problem was resolved to some extent by  $G'_{ST}$  and  $G''_{ST}$  metrics (Hedrick, 2005; Meirmans & Hedrick, 2011), and solved for a separate locus with introduced by Jost (2008) measure of differentiation  $D$  that reaches its maximum 1 when differentiation is complete. Nevertheless, new ideas are still needed for finding an intuitively acceptable approaches to measuring variation among populations especially in a case of multilocus genotypes.

Variation within a population (below we refer to population as ‘OU’, i.e., Operational Unit) could be thought of and described in different ways. There are two major facets of variation – diversity and dispersion (Gregorius & Gillet, 2015). **Diversity** is about individual types within a given OU, when all nonidentical types are considered equally distant, while **dispersion** is about an overall relationship between individual types based on pairwise dissimilarities between them. These attributes of variation are independent in the sense that OUs can be equally diverse for a wide range

of dispersion estimates, and values of dispersion can vary from extremely small to extremely large for highly diverse OUs. However, when diversity is low, dispersion estimates are also small, whereas high dispersion estimates predetermine large values of diversity.

**Differentiation** is a common but ambiguously used term. In a general context, differentiation is about the overall relationship among several OUs considered together as a group (e.g., a metapopulation defined as a group of populations) and refers to how a total variation of that group can be partitioned among and within those OUs. Classical measures of “differentiation” ( $F_{ST}$ ,  $G_{ST}$ ,  $\phi_{ST}$ ,  $R_{ST}$ ) are based on assessment of the extent to which variation of individuals within the group of OUs (e.g., all individuals of metapopulation) exceeds the corresponding average variation within each constituent OU. However, as we pointed out above, when diversity within each OU is very high (e.g. large number of equally frequent alleles), such “differentiation” measures are counterintuitive because they deliver very small scores even when OUs are completely different (e.g., populations share no alleles). Therefore, we would not recommend using the term “differentiation” in such a general context and suggest replacing it by “**structural variation**” among OUs. We propose to use the term “differentiation” for a much more specific context (see below) requesting that estimates of ‘true’ differentiation must increase with (i) a rise of an overall difference between OUs (dispersion of OUs), and (ii) a higher regularity of distribution of pairwise differences between OUs (diversity of OUs), provided that all other characteristics of relationships among OUs being identical.

The measures of biological variation proposed in this paper combine the diversity and dispersion perspectives with the diversity component being conceptually similar to metrics developed by Hill (1973) and Jost (2007, 2008) advocating the use of numbers equivalents for estimating diversity. Such measures can be used, for example, to conclude and compare the effective numbers of different species within a community, or effective number of different

communities within a landscape. According to Jost (2008), the properties of the corresponding diversity measures, when applied to alleles of genotypes, satisfy the expectations for answering population genetic questions in providing intuitively correct answers to a series of practical and theoretical questions. The main idea of Hill's approach is the multiplicative nature of diversity partitioning.

$$(total\ diversity) = (diversity\ within\ subunits) \times (diversity\ among\ subunits)$$

which allows independent estimates of within- and among-subunit components (Jost, 2007, 2008). In other words, the effective number of alleles, genotypes, or any chosen attribute in a set of OUs equals the product of the corresponding effective number per OU and the effective number of distinct OUs. Such diversity estimates are intuitive, easy to interpret and can be used in various applications (e.g., for management of populations and in conservation biology). The effective number of distinct populations is an absolute measure of population differentiation. Based on the proportion of total diversity that is contained in the average population in terms of effective numbers, Jost (2008) introduced a new non-negative measure of differentiation  $D$  that reaches its maximum 1 when differentiation is complete. Conceptual aspects of diversity partitioning and measuring diversity components based on the most general definition of effective numbers (Hill numbers are a partial case) were thoroughly considered by Gregorius (2016).

For multilocus genotypes, differentiation  $D$  is obtained by averaging across all loci. Then  $D$  reflects the average differentiation within separate loci in a given set of populations rather than differentiation between the populations due to differences in distribution and association of alleles among loci in multilocus genotypes. If two populations have identical allele distributions at each locus but non-identical association of those alleles into the corresponding multilocus genotypes, then no differentiation is detected ( $D = 0$ ). The same shortcoming characterizes all commonly used  $F_{ST}$  related measures ( $G_{ST}$ ,  $G'_{ST}$ ,  $G''_{ST}$ ,  $\phi_{ST}$ ,  $R_{ST}$ ) that do not actually measure differentiation. Chao et



al. (2015) further demonstrated that the heterozygosity-based “differentiation” measures, such as  $G_{ST}$  and Jost’s  $D$ , do not possess two of the essential monotonicity properties: differentiation never decrease when (i) a new unshared allele is added to a population, and (ii) when some copies of a shared allele are replaced by copies of an unshared allele. Thus, while being more intuitive, Jost’s “differentiation” metric  $D$  is not free of the shortcomings of the standard measures (violation of monotonicity property, inability to take into account association between loci) and may deliver inadequate estimates and even miss the actual difference between populations.

Nearly all papers cited above and many others (Heller & Siegismund, 2009; Ryman & Leimar, 2009) debate the pros and cons of a variety of “differentiation” measures considering numerous critical examples. A part of the problem is that there are two different perspectives to partitioning total genetic variation - **differentiation** and **apportionment** (Gregorius, 2009, 2010, 2016; Gregorius & Gillet, 2015), although separation between them is not clearly made.

**Differentiation** among populations describes a tendency of the same allele or genotype to occur in the same population reporting a maximum when all populations consist of unique alleles (genotypes) (i.e., populations do not share alleles, but each population may be polymorphic for each locus). Jost  $D$  is assumed to be an example of a differentiation measure although it has its own shortcomings.

**Apportionment**, on the other hand, describes a tendency of individuals with different alleles or genotypes to occur in different populations. Maximum apportionment is reached when each population is fixed for a different allele (or genotype), i.e., populations are monomorphic but have different genotypes. This means that maximum of differentiation among populations is necessary but not sufficient condition of maximum apportionment (if all genotypes are considered equally dissimilar). Thus, apportionment metrics measure the extent of fixation of distinct alleles or genotypes among populations (e.g. fixation index  $F_{ST}$ ).

There are a few immediate consequences of theoretical and practical importance for geneticists for considering the dual perspectives of differentiation and apportionment. First,  $F_{ST}$ -like indices (e.g.,  $G_{ST}$ ,  $G'_{ST}$ ,  $G''_{ST}$ ,  $\phi_{ST}$ ,  $R_{ST}$ ) provide a kind of apportionment (fixation) estimates based on variance partitions, even if they are commonly declared and used as measures of differentiation among populations. Second, Jost's "differentiation" metric  $D$  (Jost 2008) is actually closer to measuring differentiation among populations, not apportionment. This may explain, at least in part, inconsistency in some results obtained with  $D$  and the  $F_{ST}$  based measures. Third, valid differentiation measures can reach their maximum (absolute differentiation) independently of the degree of genetic variation within populations, i.e., even if the populations are not fixed to alternative alleles or genotypes (such situation is impossible with  $F_{ST}$  and  $G_{ST}$ ).

In this paper our purpose is to further expand the differentiation perspective for studies of population structure. The idea is to express diversity of populations in terms of the **effective number of equally distant populations**. This allows estimation of differentiation in a way that is independent of both total diversity ( $\gamma$ -diversity) of a given metapopulation and diversity within its constituents ( $\alpha$ -diversity). Determining the effective number is based on pairwise genetic distances between populations, though only the proportional contributions of those distances to the total sum of distances are utilized. Such diversity index depends only on the relative position of populations to each other in the given genetic landscape and measures regularity of relationships among populations. Therefore, an identical value of diversity index is returned for any metapopulation consisting of the same number of populations, even if all pairwise genetic distances (magnitudes of genetic differences) change proportionally (e.g., for two sets of three populations with relationships among the populations represented geometrically by two similar shaped but different size triangles). For example, if each of three populations is fixed to a single binary genotype at six loci in two metapopulations  $A = \{(100000), (001000), (000010)\}$  and  $B = \{(110000), (001100), (000011)\}$ ,

then  $A$  and  $B$  are of identical diversity among their constituent populations, although pairwise genetic differences between the three populations in  $A$  are two times smaller than those in  $B$ .

To distinguish between two different metapopulations with the same diversity (as measured in terms of effective number of equally distant populations), the diversity concept must be integrated with the dispersion concept. The dispersion component of variability is expressed in terms of genetic distances between populations. Combined metrics of diversity and dispersion components will be then called **the Effective Number of Different Populations** (ENDP). Such metrics are completely predetermined by pairwise genetic distances between populations, their magnitudes and regularity of distribution, and deliver exhaustive estimates of variation among populations within the corresponding metapopulation. Basic principles of our approach are similar to those developed by Scheiner et al. (2017) for ecological communities (Gregorius and Kosman (2018) considered a more general case of integration of the diversity and dispersion concepts).

We test the relevance of the suggested metrics with two empirical data sets. First, we use data published by Feijen et al. (2022) describing population and species structure of the New Zealand trematode parasite species in the genus *Atriophallophorus* spp. using nuclear SNP markers and mitochondrial haplotypes based on a part of the NADH5 gene. This parasite uses the snail *Potamopyrgus antipodarum* as its intermediate host and waterfowl as the definitive host. The parasite has a sexual stage in the definitive host while the reproduction in the snail host is clonal. Feijen et al. (2022) found support for cryptic species structure in the parasite populations by applying computationally demanding multispecies coalescent models on a subset of individual parasites ( $N = 52$ ) [Bayes Factor Delimitation (Leache et al., 2014)]. They further used regression analyses on pairwise genetic distances among individuals ( $N=462$ ). Both analyses supported the conclusion that the samples represent at least two distinct species that coexist in broad geographic range (see figure 2 in Feijen et al., 2022). Here we use the same subset of genotypes and the full set

of genotypes as in the two analysis by Feijen et al. (2022) to calculate both the effective number of equally distant populations and the ENDP in samples that are known to represent two coexisting cryptic species.

Second, we applied the new metric to assess population genetic structure of the common species, *Atriophallophorus winterbourni*. We asked what the effective number of equally distant and different populations is in these locations which cover the geographic regions of South Island of New Zealand. We contrast our results to a more detailed analysis of connectedness of these populations presented in Feijen et al (2022).

We use these data to raise the question whether it would be reasonable to incorporate estimates of ENDP into analyses aiming to understand diversity and structure of populations using genetic markers. An important reason for selection of those data was the fact that they were already analyzed with other state-of-the-art tools that allow a direct and effective comparison of the new delivered results with those reported previously. We also discuss the rationale and applicability of these metrics.

## Materials and methods

We develop metrics for measuring structural variation in a metapopulation based on a matrix of pairwise genetic distances between the populations. Distances between the populations are measured using the dissimilarity-based approaches (Kosman & Leonard, 2007; Kosman, 2014) although other distances can also be applied. This approach requires a proper assessment of dissimilarity between individual genotypes.

### Dissimilarity between individual genotypes

Choice of a suitable dissimilarity measure is a key factor for valid analysis of genetic variation. The selection depends on ploidy of a given organism and the type of molecular markers used for

estimating genetic variation (Kosman & Leonard, 2005; Kosman & Jokela, 2019). Here, we use nuclear SNP polymorphism of *Atriophallophorus* spp. (Feijen et al., 2022) to examine population genetic structure. Since SNPs are codominant markers and *Atriophallophorus* spp. is a diploid organism, we calculated dissimilarity between the SNP genotypes ( $\delta$ ) according to eqn. 3 in Kosman and Leonard (2005) or eqn. 6 in Kosman and Jokela (2019). Here, the dissimilarity between two genotypes at one diploid locus equals 1, 0.5 and 0, if the genotypes do not share any allele, share one allele, or have identical pair of alleles, respectively. Then the average across all loci delivers dissimilarity  $\delta$  between the two multilocus genotypes.

#### Distance between populations

The most used genetic distance measures between populations are based on allele frequencies, averaging independent estimates at each locus over all loci [e.g. Nei's genetic distances (Nei, 1972)]. Allele-frequency based measures do not consider possible associations between different loci, so that two populations with no shared genotypes can be declared identical if they share the same alleles at equal frequencies. Therefore, considering associations between loci would be important for metrics of genetic distances between populations.

The two types of distances based on dissimilarities between individuals are calculated by averaging individual dissimilarities (both between and within populations) and by assignment of individuals from two populations based on their dissimilarities without the effect of dissimilarities within populations (Kosman, 2014). The average-based approach (distance of average differences,  $DAD_\rho$ , eqn. 2 in Kosman and Leonard (2007)) may have undesirable mathematical properties for some dissimilarity measures  $\rho$  as  $DAD_\rho$  can be negative or zero for distinct populations. For example,  $DAD_m$ , which is the distance of average differences for the simple mismatch coefficient  $m$ , can be zero for distinct populations as it is identical to Nei's minimum genetic distance (Kosman & Leonard, 2007). Therefore, the distance of average differences does not properly work in the case

of association between loci. An alternative, the assignment-based genetic distance ( $KB$ ) developed by Kosman (1996) and Gregorius et al. (2003), is a generalization of the mathematical notion of distance between two sets of scattered points (Kosman, 2014). Kosman distance ( $KB$ ) can distinguish between populations where linkage of markers is variable for a same set of alleles, and it is suitable for comparison of populations with strong linkage patterns as is the case for asexual or mixed mode of reproduction, or with cryptic structure due to unidentified coexisting species.

One strength of dissimilarity-based methods is the ability to deal with missing data. Dissimilarity between a given pair of genotypes can be calculated using all the data that are available for both individuals (only loci with missing genotypes are excluded).

We applied the dissimilarity-based distances  $DAD_\delta$  and  $KB_\delta$  to measure genetic differences between the parasite populations *Atriophallophorus* spp. (SNP markers), where  $\delta$  is dissimilarity between the multilocus SNP genotypes mentioned beforehand in the previous section. Since the mode of parasite reproduction is mixed with prevailing outcrossing, we used the  $DAD_\delta$  distance as the benchmark for calculations assuming that association between loci is minimal, if any. As Feijen et al. (2022) also discovered a cryptic species structure in their *Atriophallophorus* spp. samples, we also calculated effective numbers based on  $KB_\delta$  distances. This is to show how dissimilarity-based distances,  $DAD_\delta$  and  $KB_\delta$ , can be used to study structural variation in cases where it is not known if there are groups within-populations that differ in their linkage structure.

## Metrics of variation

### **Diversity**

We first construct metrics of variability similarly to Scheiner et al. (2017). For a set of  $S$  Operational Units (OUs; single populations in our analysis), let  $d_{ij}$  be any distance between  $i$ th and  $j$ th OUs ( $0 \leq d_{ij} \leq 1$ ,  $d_{ij} = d_{ji}$ ,  $d_{ii} = 0$ ;  $i, j = 1, 2, \dots, S$ ). For any non-negative parameter  $q \neq 1$ ,

we calculate an extent of homogeneity of pairwise distances as effective number of ordered pairs of OUs according to Hill (1973):

$${}^qH = \left( \sum_{i=1}^S \sum_{j \neq i=1}^S f_{ij}^q \right)^{1/(1-q)}, \quad (1)$$

whereas for  $q = 1$

$${}^1H = \lim_{q \rightarrow 1} {}^qH = \exp\left(- \sum_{i=1}^S \sum_{j \neq i=1}^S f_{ij} \log f_{ij}\right), \quad (2)$$

where  $f_{ij} = d_{ij} / \sum_{i=1}^S \sum_{j \neq i=1}^S d_{ij}$  is the proportional contribution of the ordered pair  $(i, j)$  into the total distance between all pairs of OUs (we assume that  $f_{ij} \log f_{ij} = 0$  by definition. if  $f_{ij} = 0$ ).  ${}^qH$  equals a hypothetical number of ordered equally distant pairs of different OUs ( $d_{ij} > 0, i \neq j$ ) that generate the same Hill number as the given set of  $S^2 - S$  pairs. This measure increases when variability in distances decreases, and range of  ${}^qH$  is between 0, if all  $d_{ij} = 0$  (by definition), and its maximum  $S^2 - S$ , when all  $d_{ij} \neq 0$  are equal for  $i \neq j$  ( $S$  values  $d_{ii} = 0$ ). Then diversity within the given set of OUs is obtained as solution of quadratic equation  $({}^qD)^2 - {}^qD = {}^qH$ :

$${}^qD = \frac{1 + \sqrt{1 + 4 {}^qH}}{2}, \quad (3)$$

and expressed in terms of effective number of equally distant types of OUs (Scheiner et al., 2017). Values of  ${}^qD$  range from 1 to  $S$ , when all OUs are “identical” (all  $d_{ij} = 0$ ) and all non-identical OUs are equidistant ( $d_{ij} = \text{const} \neq 0$ ), respectively. Note,  ${}^qD$  gets smaller for larger  $q$ , and equal effect of all pairwise distances on the effective numbers is obtained just for  $q = 1$ .

A kind of evenness of the OUs distribution is determined as

$${}^qE = {}^qD / S \quad (4)$$

with a range  $[1/S, 1]$ . It is useful to transform this estimate onto the unit interval for comparison of sets with different numbers of OUs:

$${}^qE' = ({}^qD - 1) / (S - 1) \quad (4')$$

287 with a range  $[0, 1]$ . So, diversity  ${}^qD$  increases with evenness and can be decomposed to the product  
 288 of evenness and richness (number of OUs):

$$289 \quad {}^qD = {}^qE \times S \quad \text{or} \quad (5)$$

$$290 \quad {}^qD = 1 + {}^qE' \times (S - 1). \quad (5')$$

291 More accurately,  ${}^qD$  and  ${}^qE$  ( ${}^qE'$ ) should be called diversity (**effective number of equally**  
 292 **distant populations (OUs)**) and evenness of order  $q$ , respectively.

293 Diversity  ${}^qD$  reflects regularity of OUs distribution in a relevant space. It is determined by  
 294 proportions  $f_{ij}$  and does not depend on actual distances  $d_{ij}$  between OUs in a sense that if all  
 295 distances are subject to enlargement to the same extent,  ${}^qD$  remains unchanged since  ${}^qH$  does so.  
 296 Thus, the effective number of equidistant OUs serves as an invariant of configuration of the given  
 297 set in space (diversity perspective), while the degree to which OUs are similar to each other is not  
 298 considered (dispersion perspective). Therefore, the diversity reveals an important component of  
 299 biological variation, but not the complete structure of the metapopulation. Next, we will  
 300 complement the diversity with dispersion perspective for a comprehensive description of variability  
 301 within a set of OUs.

### 302 ***Integration of diversity and dispersion***

303 Theoretical aspects of dispersion and its relationship to diversity were broadly considered in  
 304 Gregorius and Kosman (2017, 2018). To develop overall metrics of variation, we incorporate two of  
 305 the most basic and tangible dispersion estimates. The first one is the Average Distance Within  
 306 ( $ADW$ ) a set of OUs

$$307 \quad ADW = \sum_i^S \sum_j^S d_{ij} / S^2 \quad (6)$$

308 with a range from 0 to  $(S - 1)/S$ , or its derivative  $ADW'$  obtained by transformation of  $ADW$  onto  
 309 the unit interval ( $0 \leq ADW' \leq 1$ )



$$ADW' = \frac{S}{S-1} \times ADW = \frac{S}{S-1} \times \sum_i^S \sum_j^S d_{ij} / S^2. \quad (6')$$

The second metric of dispersion is Kosman's assignment-based measure  $KW$  (Kosman, 1996, 2014; Kosman & Leonard, 2007) that has a range  $[0, 1]$  and can be considered as generalization of the mathematical definition of the diameter of a set of scattered points.

Finally, we combine diversity ( ${}^qD$ ) and dispersion ( $ADW$  or  $ADW'$ , and  $KW$ ) estimates into integrated metrics of overall structural variation that we call **the effective number of different populations** (ENDP), or OUs:

$${}^qD(ADW) = 1 + {}^qD \times ADW = 1 + S \times {}^qE \times ADW = 1 + (S - 1) \times {}^qE \times ADW', \quad (7)$$

$${}^qD(KW) = 1 + \frac{S-1}{S} \times {}^qD \times KW = 1 + (S - 1) \times {}^qE \times KW \quad (8)$$

with a range from 1 to  $S$ . A general form of eqns. 7-8 is

$${}^qD(M) = 1 + (S - 1) \times {}^qE \times M \quad (9)$$

for any dispersion metrics  $M$  with  $[0,1]$  range. The immediate consequence is that even if diversity is maximal ( ${}^qD = S$ ), i.e., all OUs are equally distant (evenly distributed), the effective number of different OUs  ${}^qD(M)$  decreases and approaches to 1 when OUs are closer to each other (dispersion decreases and tends to 0). According to (9), the effective numbers of different OUs  ${}^qD(M)$  can be represented as a decomposition of the three generally independent basic components: simple richness of a given set ( $S$ ), evenness ( ${}^qE$ ), and dispersion ( $M$ ). The effective number of different OUs could be conceived as the number of equidistant OUs needed to obtain the same dispersion and variability in pairwise distances as those observed in the given set of OUs (where OUs may not be equally distant).

The suggested approaches to estimating variation can be thought of as reducing the actual number of OUs (richness) in two steps. Analyzing regularity of OUs distribution, richness ( $S$ ) decreases to the effective number of distinct equidistant OUs ( ${}^qD$ ) due to deviations from a perfect

evenness. Then, considering a magnitude of similarity between OUs (dispersion) results in further richness decline from  ${}^qD$  to the effective number of different OUs ( ${}^qD(M)$  for dispersion  $M$ ). Thus, combining both the diversity and dispersion perspectives, overall variation of a set of OUs is expressed in terms of reduction of its simple estimate (richness) to perhaps the most exhaustive one – **the effective number of different units**. The effective numbers of different and equidistant units are equal only in two extreme cases: for a set consisting of one unit (trivial situation), and when all units are maximally distant.

To make a comparison of structural variation of sets with different numbers of OUs, relative estimates of the effective numbers ( $1 \leq EN \leq S$ ) are useful and reached by the linear transformation of  $EN$  onto the unit interval

$$nEN = (EN - 1)/(S - 1). \quad (10)$$

$nEN$  increases with increasing variation  $EN$  and can be considered the metric of structural differentiation of OUs. The relative effective number of equally distant OUs ( $nD$ ) is obtained for  $EN = {}^qD$  from (10), i.e.  ${}^qnD = {}^qE'$  is evenness from (4'), while the relative effective number of different OUs  ${}^qnD(M)$  is attained with  $EN$  from the absolute estimate  ${}^qD(M)$  (eqn 9). These relative estimates ( $nEN$ ) range from 0 (no differentiation) to 1 (completely structured set of OUs) when the corresponding effective number equals 1 and  $S$ , respectively. Both the metrics  $EN$  and  $nEN$  of variation among populations are totally independent of variability within the populations because the latter was not even involved in generation these metrics of differentiation. This independence is reached using conceptually different approach comparing with those of Jost (2008, eqns. 8 and 10, p. 4021), which could be referred to as approaches based on the partitioning of diversity within and among OUs. Thus, the suggested metrics of structural differentiation  $nEN$  (10) are completely different from classical measures of differentiation.

## Data and differentiation among parasite populations

We tested the new metrics with a published dataset on genetic structure of a diploid trematode parasite *Atriophallophorus* spp. (Feijen et al., 2022). *Atriophallophorus* has a snail-bird life cycle. It reproduces sexually in the bird definitive host. The adult worms are hermaphrodites but evidence supports outcrossing as main mode of reproduction (Feijen, 2020). The parasite reproduces asexually in the snail intermediate host. Feijen et al. (2022) reports a phylogeographic analysis of the most common *Atriophallophorus* species, *A. winterbourni*, but the study also revealed a previously unknown sister species coexisting with *A. winterbourni* (Feijen et al., 2022). This putative species remains undescribed. The study covered a wide geographic range (South Island of New Zealand) and applied both nuclear and mitochondrial markers in a detailed phylogeographic analysis of the studied populations. Here, we use these data to ask what the ENDP is when calculated with the new metrics we present. We first test how the new method performs when we apply it to samples representing the two main species. In our analyses we mainly refer to figure 2, figure 3, and figure S4 of the publication (Feijen et al., 2022). We use the same data that they analyzed for species delimitation among *Atriophallophorus* spp. We then limit the analysis to the most common species *A. winterbourni* and contrast effective numbers of equally distant populations ( ${}^qD$ ) to ENDP ( ${}^qD(M)$ ). Only polymorphic SNP loci were used in the analysis.

We estimated the variation among these parasite populations as follows:

1. We calculated the dissimilarity between the SNP genotypes ( $\delta$ ) according to eqn. 3 and the corresponding algorithm on p. 421 in Kosman and Leonard (2005) or eqn. 6 in Kosman and Jokela (2019). In the case of missing data, the corresponding loci were ignored for each pair, and a dissimilarity value was obtained on the reduced number of loci with available data for both individuals in the pair.

2. We computed the average-based and assignment-based distances using the  $\delta$ -dissimilarity ( $DAD_\delta$  and  $KB_\delta$ , respectively) between all pairs of populations.
3. We calculated the effective number of equally distant populations (diversity  $^1D$ ) according to eqns. 2–3 for distances  $d = DAD_\delta$  and  $d = KB_\delta$ , and  $q = 1$ . Then the diversity-based estimates of differentiation ( $nEN$ ) were obtained for  $EN = ^1D$  from eqn 10.
4. We calculated the dispersion of the parasite populations ( $ADW_{DAD_\delta}$  and  $ADW_{KB_\delta}$ ) using eqn. 6 ( $ADW$  based on distances  $d = DAD_\delta$  and  $d = KB_\delta$ ).
5. We calculated the ENDP (structural variation  $^1D(ADW)$ ) according to (7) for  $q = 1$  for the corresponding pairs of diversity  $^1D$  and dispersion  $ADW$  estimated with distances  $d = DAD_\delta$  and  $d = KB_\delta$ . Then the corresponding assessments of structural differentiation ( $nEN$ ) were obtained according to (10) with  $EN = ^1D(ADW)$ .

## Results

### *Application of effective numbers of populations to mixed populations of cryptic species*

Based on the species delimitation analysis, Feijen et al (2022) concluded that at least two species of *Atriophallophorus* parasites were found in the studied populations. We calculated that the ENDP ( $^1D(ADW_{DAD})$ ,  $^1D(ADW_{KB})$ ) in the set of samples grouped by the six major mitochondrial haplotype groups was 1.40 when based on the distance of average differences ( $DAD_\delta$ ) and 2.05 for the assignment-based genetic distance ( $KB_\delta$ ) (Table 1). While the difference between these metrics is 32%, here the assignment-based distance seems to match the expectation of at least two species particularly well and average-based distance seems to underestimate the number of inferred OUs.

As the calculation of these metrics does not demand as many computational resources as the Bayes Factor Delimitation models that Feijen et al., (2022) used, we were able to expand the analysis to a larger dataset used in the regression analysis in Feijen et al., (2022). Our results are very similar

to the results reported by Feijen et al. (Figure 1, Table 1). Interestingly, the ENDP was not affected by the sample size (Table 1). This indicates that these metrics are robust to variation in sample size assuming the samples still represent the different OUs (here, haplotype groups).

Our results illustrate that the ENDP captures the underlying genetic structure in *Atriophallophorus* clade (Figure 1). Although the species is sexual, it seems that in this case the association-based *KB* distance was more strongly in agreement with previous analyses than distance of average differences (*DAD*). This may be due to low gene flow between the species emphasizing the differences between the species that appear as strong linkage (association between loci) when haplotype groups are compared. Note also that the effective number of equally distant populations, which reflects the diversity, was close to maximum defined by the six haplotype groups (Table 1). Interestingly, when diversity was calculated based on average (*DAD*) or association-based (*KB*) distance the estimates only differed by 6% (Table 1). Analysis of number of equally distant populations does not capture the cryptic species structure in the clade, probably because it treats all haplotype groups independently of the magnitude of differences between them. In this case using the additional information from dispersion was therefore essential to describe the previously inferred structural variation among the haplotype groups.

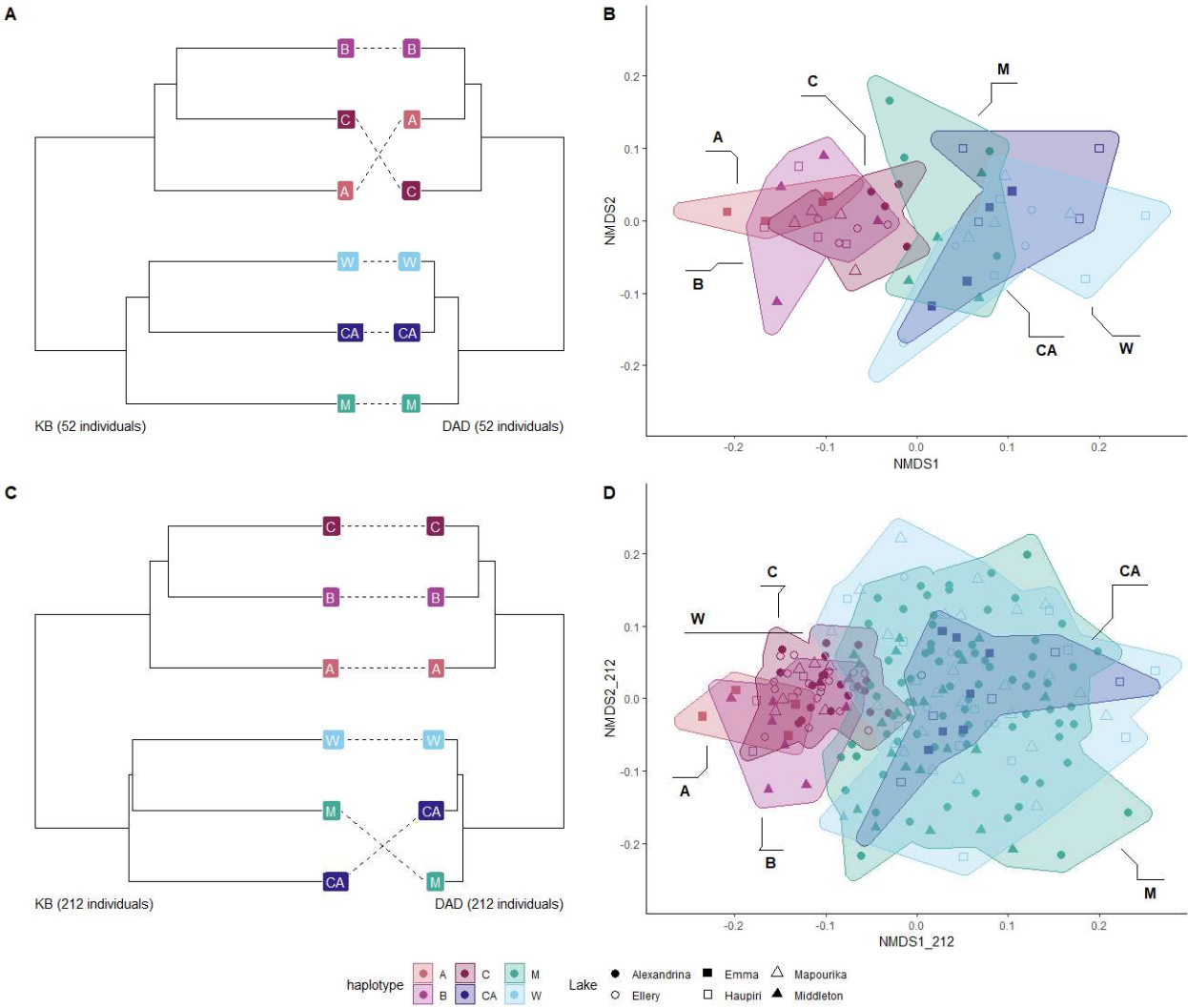


Figure 1. UPGMA dendrograms and NMDS plots of the two datasets (A, B: 52 individuals; C, D: 212 individuals). Panels A and C are based on pairwise *KB* (left) and *DAD* (right) distances between the six major mitochondrial haplotype groups reported in Feijen et al. (2022). Note that *DAD* topology in A is congruent with the tree shown in figure 2c in Feijen et al. (2022), while the top clade (haplotype groups B, C, A) show a different structure obtained with the *DAD* and *KB* distances in A and C. Panels B and D show NMDS plots calculated based on pairwise distances between individuals. The haplotype group for each sample is indicated in the label.

427 Table 1. Variability among the trematode *Atriophallophorus* spp. collections.

Type of variation	Variation parameters	“cryptic” species/populations identified based on mt-haplotype groups (Feijen et al., 2022)		<i>Atriophallophorus</i> populations (natural lakes)
		52 genotypes	212 genotypes	306 genotypes
		24 loci	24 loci	24 loci
		6 hapl. groups	6 hapl. groups	10 lakes
<b>Effective number of equally distant populations (Diversity)</b>	${}^1D_{DAD}^a$	5.544	5.445	9.783
	${}^1D_{KB}$	5.914	5.911	9.990
<b>Dispersion</b>	$ADW_{DAD}^b$	0.071	0.058	0.015
	$ADW'_{DAD}^b$	0.085	0.070	0.017
	$ADW_{KB}$	0.178	0.169	0.171
	$ADW'_{KB}$	0.217	0.203	0.190
Evenness	${}^1E_{DAD}^c$	0.924	0.908	0.978
	${}^1E_{KB}$	0.986	0.985	0.999
	${}^1E'_{DAD} = {}^1nD_{DAD}^c$	0.909	0.889	0.976
	${}^1E'_{KB} = {}^1nD_{KB}$	0.983	0.985	0.999
<b>ENDP, effective number of different populations (Structural variation)</b>	${}^1D(ADW_{DAD})^a$	1.396	1.316	1.146
	${}^1D(ADW_{KB})$	2.053	1.999	2.698
Extent of differentiation	${}^1nD(ADW_{DAD})^c$	0.079	0.063	0.016
	${}^1nD(ADW_{KB})$	0.211	0.200	0.189

428 <sup>a</sup> effective number (eqns. 3, 7 - 9);

429 <sup>b</sup> dispersion (eqns. 6, 6'; Kosman, 1996; Kosman & Leonard, 2007);

430 <sup>c</sup> evenness (eqns. 4, 4');

431 <sup>c</sup> extent of differentiation - normalized ENDP (eqn. 10).

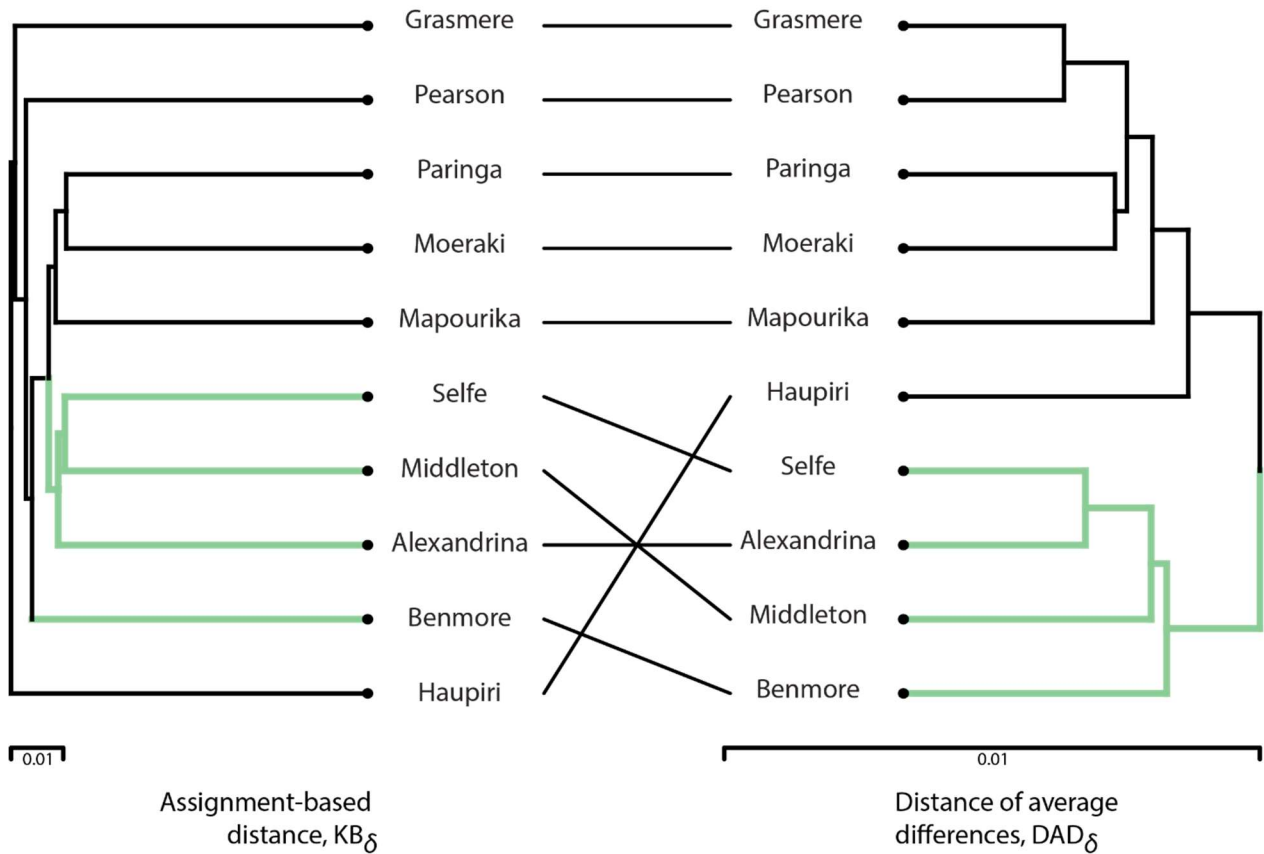


Figure 2. UPGMA trees of *Atriophallophorus winterbourni* populations from 10 lakes on the South Island of New Zealand. Data are the same as presented in the Table S4.1 of Feijen et al. (2022), with the exception that the lakes with small samples (less than 10 individuals) were excluded from the analyses. The colors of the branches correspond to two main clusters identified in the Structure analysis presented in Feijen et al. (2022; Figure 3d). Effective numbers of different populations based on the  $DAD$  and  $KB$  distances are 1.15 and 2.70, respectively.



439 *Application of effective number of populations to geographically separate populations of single*  
440 *species, Atriophallophorus winterbourni*

441 Feijen et al. (2022) presented genetic pairwise  $F_{ST}$  and structure analyses for 15 lake populations of  
442 *Atriophallophorus winterbourni*. Their first discovery was that the nuclear marker-based estimates  
443 for population structure were much less than mitochondrial marker-based estimates. Their main  
444 conclusion was that in the past the populations were likely separated in glacial refugia in the north  
445 and south of the Island and that the present population differentiation in nuclear and mitochondrial  
446 markers is maintained due to low level of cross-alpine migration. Average nuclear  $F_{ST}$  was low, and  
447 together with analysis of migration patterns using isolation by distance tests and marginal  
448 approximation of structured coalescence (phylogeographic analysis based on mitochondrial markers  
449 applying Mascot 2.1.2. in BEAST 2.6.5. [see details in Feijen et al. (2022)], the conclusion was that  
450 even if the mitochondrial  $F_{ST}$  estimates were high, there is a considerable nuclear geneflow among  
451 all populations at present.

452 Our analysis using the *DAD* distance suggested that the ENDP in these data is 1.15  
453 supporting the view that there may have been two distinct glacial refugia, but the nuclear  
454 marker-based differentiation among the population is currently weak. However, using the  
455 association-based *KB* distance the ENDP was 2.70 (Table 1). Figure 2 illustrates differences in  
456 relationships among the populations between the two estimates. In this case analysis based on the  
457 distance of average differences *DAD* reflects the expected structural variation better than the  
458 association-based *KB* distance. This may be expected as the data represent large outbred sexual  
459 populations that are in HW equilibrium showing no signal of linkage disequilibrium (Feijen et al.,  
460 2022).

## Discussion

Assessing genetic structure of populations requires that the chosen measures reflect the biological processes that affect local genetic variability and divergence among populations (Bohonak, 1999). Relevant processes shaping population genetic structure are well understood but capturing these processes to a single metric is difficult. For example, species mating system has consequences for the expected genetic variability of populations (Holsinger, 1992; Rieseberg & Burke, 2001), variation in population size affects the strength of genetic drift (Wang et al., 2016), and local adaptation may promote divergence of genes under selection (Yeaman & Whitlock, 2011). Metapopulations consist of local populations of different sizes, which may be connected by highly asymmetric gene flow (Harrison & Hastings, 1996; Morrissey & de Kerckhove, 2009). Recently evolved mating barriers may also lead to cryptic species structure that is yet unnoticed and further complicates the analysis of population genetic structure (Baker et al., 1995). Ideally, the chosen metric would be robust in the sense that there is no unrecognizable bias by specific biological processes or possible sampling errors. It would be very valuable if the metrics recorded would guide the inclusion and exclusion of alternative hypotheses to explain the observed patterns. It is unlikely that a single metric can capture all aspects of population structure, processes defining divergence of populations and methodological caveats that handicap our conclusions. Inference from several alternative metrics might allow concluding how the populations are structured, which processes are relevant and how the analyses can be refined to address specific follow-up questions.

We aimed to show how beta variation among populations can be estimated independently of alpha variation within populations, to evaluate how metrics incorporating both the diversity (based on Hill numbers) and dispersion facets of variation can be used as beta variation estimates, and how they are best constructed to evaluate population genetic data from natural populations that differ in the processes that shape the population genetic structure. We focused on evaluating both diversity

and dispersion emphasizing that both are important. The second aspect that we examined is the difference between average (*DAD*) and association-based (*KB*) distance measures (Kosman & Leonard, 2007) when deriving effective numbers estimates. We showed that estimates of the ENDP based on the *DAD* distance are well suited for situations where studied OUs have low compatibility barriers generating association due to assortative mating (or fertility) patterns. If compatibility barriers (i.e., cryptic species) exist, then the *KB* distance used in calculating the ENDP capture the structural variation better.

We argue that the analysis of population genetic structure, genetic variability of populations and assessment of the conservation value of local populations would benefit from inclusion of both the diversity and dispersion aspect of structural variation when estimating genetic relationships of populations in a metapopulation (beta variation). We use examples from population genetics, but these same approaches can be utilized in study of biological communities using functional traits (Scheiner et al., 2017; Kosman et al., 2019). We believe that in this sense the recognition of diversity and dispersion perspective to variation is integrative and common to both genetics and ecology. It would be important to examine how such integration is best achieved and if there is a link between genetic and functional diversity, or genetic and functional dispersion. Here, we recognize the debate on the link between biodiversity and ecosystem function (Grime, 1997; deLaplante & Picasso, 2011). Maybe the anomalous results from the tests of this central hypothesis are actually due to lack of consideration of diversity and dispersion aspects of the taken measures. Are the used measures of diversity also capturing the dispersion of taxa that would best map on dispersion of ecosystem function? In other words, the metrics that measure dispersion (or metrics that combine both dispersion and diversity) might be closer to the objectives for testing the biodiversity-ecosystem function hypothesis.

Our main interest was to ask how we best characterize structural variation in populations using population genetic markers. The classical approach in population genetics relies on a kind of apportionment (not differentiation!) measures (like  $F_{ST}$  and its relatives) that strictly deal with the diversity aspect of variation and are blind to dispersion. This does not seem a limitation when considering only one locus and assuming that all alleles are equally dissimilar. However, the limitations of the classical approach become real when one considers markers where the extent of similarity between different alleles at one locus may vary (e.g., microsatellites, Kosman & Jokela, 2019). At present, most genetic data consist of multilocus genotypes (e.g., any sequence of any kind). When examining such data, it is very easy to agree that not all genotypes are equally dissimilar; therefore, an analysis using information on variation in dissimilarity to support conclusions on structural variation of populations may be a useful addition. Using dissimilarity is implicit in coalescence models of evolution where evaluation of the shortest approach to ancestral type requires understanding of evolutionary distances of the derived types (Rosenberg & Nordborg, 2002). Evident power of coalescence-based models is one of the reasons why we argue that also studies on structural variation of populations (population genetic structure/diversity) would greatly benefit from incorporation of the dispersion component into measuring of overall variation.

Another known shortcoming of applying the classical (apportionment) metrics to measuring differentiation among populations is the dependence of those metrics on variation within the populations (this is why they do not assess the differentiation) (Jost, 2008; Gregorius, 2014). The great advantage of using numbers equivalents to estimate variation within (alpha) and among units (beta) is that those estimates are independent (Jost, 2007). However, even the modified metrics developed for measuring differentiation (e.g., Jost's  $D$ ) still depend on diversity within populations (e.g., counterintuitively Jost's  $D$  cannot reach its maximum value 1 even if two populations do not share any alleles, but at least one of them is not fixed). The approach we advocate here (combining

diversity and dispersion) to derive differentiation measures based on effective numbers of different OUs, provides efficient and tangible tools for analyzing relationships among populations, and allows comparisons across studies.

Our two examples illustrate how the effective numbers approach can be used in ecological genetics evaluating structural variation in natural populations. We emphasize the difference between assessments of the effective numbers of different OUs with average-based and association-based distance measures between the OUs. In some cases, where populations are large, outcrossing and not under strong selection or drift, metrics based on the distance of average differences are capturing the processes affecting structural variation among populations. This was the situation in our second example where geographically widespread species was inferred to have been divided into two major regions that had somewhat less gene flow between regions than within regions. In our first example, what was long assumed a single species in fact consisted of coexisting cryptic species that were morphologically similar but evolutionarily diverged (Feijen et al., 2022). Such cases are very demanding to discover with data that are collected to test hypotheses assuming a single species. Here, the proxy we used to construct evolutionary prior groups was the mitochondrial haplotype memberships. Finding such a prior grouping factor requires collection of additional data and processes such as incomplete lineage sorting may complicate matters further (Maddison & Knowles, 2006; Pedraza-Marrón et al., 2019). For this case we showed that association-based ENDP captured the assumed cryptic species structure and could have been used to motivate further species delimitation studies with high confidence. Of course, here we have the advantage of hindsight as such analyses were already done (Feijen et al., 2022).

The analyses we present require that it is possible to have prior assumptions of OUs. We believe that collecting data with assumed a prior structure in mind is a much more productive approach than assuming no structure. Everything in biology speaks for assuming memberships of

556 groups for observed individuals even if everything in statistics is based on constructing null models  
557 for assuming such groups/structures do not exist. For example, membership in the population can  
558 be assumed by spatial location, or by mitochondrial haplotype identity, as we show in our  
559 examples. Both spatial priors and haplotype identities can cross species boundaries, but they might  
560 still be useful starting points for structural analysis. Here, our first example relied on using priors  
561 based on haplotype groups, and the second relied on population membership. We believe that the  
562 power of using the suggested approach is that one can reduce the priors to the most likely number  
563 of different (genetically, functionally etc.) groups among the OUs in question thus providing  
564 important information about the structure in the data based on the corresponding estimate of  
565 effective number of different OUs. This is a philosophically different approach than asking the data  
566 (blindly) how many groups emerge when some clustering algorithm is applied. We think it is rare  
567 not to have a good candidate for prior grouping. Most data are collected assuming population  
568 membership. Therefore, asking about the effective number is a logical thing to do when analyzing  
569 the data. Most data are assigned to more populations than in fact are there since for most species the  
570 migration patterns and effective geneflow are not known partly due to the lack of conceptually  
571 sound methods of population delineation. This is an issue that is like the inference we receive from  
572 population size (number of individuals) and effective population size (number of individuals  
573 contributing to the next generation). We see value in assigning population memberships a priori and  
574 validating that count post hoc with effective numbers metrics and suggest this should be part of our  
575 routine beta diversity estimates when conducting studies on biodiversity, genetic diversity or  
576 functional diversity of populations.

## 577 References

- 578 Baker, A. J., Daugherty, C. H., Colbourne, R., & McLennan, J. L. (1995). Flightless brown kiwis of  
579 New Zealand possess extremely subdivided population structure and cryptic species like small  
580 mammals. *Proceedings of the National Academy of Sciences of the United States of America*,  
581 92, 8254-8258.
- 582 Bohonak, A. J. (1999). Dispersal, gene flow, and population structure. *Quarterly Review of Biology*,  
583 74, 21-45.
- 584 Chao, A., Jost, L., Hsieh, T. C., Ma, K. H., Sherwin, W. B., & Rollins, L. A. (2015). Expected  
585 Shannon entropy and Shannon differentiation between subpopulations for neutral genes under  
586 the finite island model. *PLoS ONE*, 10, e0125471.
- 587 Czajowski, G., Kosman, E., Slowacki, P., Park, R. F., & Czembor, P. (2021). Assessing new SSR  
588 markers for utility and informativeness in genetic studies of brown rust fungi on wheat,  
589 triticale, and rye. *Plant Pathology*, 70, 1110-1122.
- 590 deLaplante, K., & Picasso, V. (2011). The biodiversity-ecosystem function debate in ecology.  
591 *Philosophy of Ecology* (eds K. deLaplante, B. Brown & K.A. Peacock), pp. 169-200. North-  
592 Holland, Amsterdam.
- 593 Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from  
594 metric distances among DNA haplotypes - application to human mitochondrial-DNA  
595 restriction data. *Genetics*, 131, 479-491.
- 596 Feijen, F. (2020). Looking-glass parasites. *Evolution and ecology of the Red Queen's parasite*,  
597 *Atriophallophorus winterbourni* (Blasco-Costa et al., 2019). ETH Zurich.
- 598 Feijen, F., Zajac, N., Vorburger, C., Blasco-Costa, I., & Jokela, J. (2022). Phylogeography and cryptic  
599 species structure of a locally adapted parasite in New Zealand. *Molecular Ecology*, 31, 4112-  
600 4126.

601 Gillet, E. M. (2013). DifferInt: compositional differentiation among populations at three levels of  
602 genetic integration. *Molecular Ecology Resources*, 13, 953-964.

603 Gillet, E. M., & Gregorius, H.-R. (2008). Measuring differentiation among populations at different  
604 levels of genetic integration. *BMC Genetics*, 9, 60.

605 Gillet, E. M., Gregorius, H.-R., & Ziehe, M. (2004). May inclusion of trait differences in genetic  
606 cluster analysis alter our views? *Forest Ecology and Management*, 197, 149-158.

607 Gregorius, H.-R. (2009). Distribution of variation over populations. *Theory in Biosciences*, 128, 179-  
608 189.

609 Gregorius, H.-R. (2010). Linking diversity and differentiation. *Diversity*, 2, 370-394.

610 Gregorius, H.-R. (2014). Partitioning of diversity: the "within communities" component. *Web*  
611 *Ecology*, 14, 51-60.

612 Gregorius, H.-R. (2016). Effective numbers in the partitioning of biological diversity. *Journal of*  
613 *Theoretical Biology*, 409, 133–147.

614 Gregorius, H.-R., & Gillet, E. M. (2015). Classifying measures of biological variation. *PLoS ONE*,  
615 10, e0115312.

616 Gregorius, H.-R., Gillet, E. M., & Ziehe, M. (2003). Measuring differences of trait distributions  
617 between populations. *Biometrical Journal*, 45, 959-973.

618 Gregorius, H.-R., & Kosman, E. (2017). On the notion of dispersion: from dispersion to diversity.  
619 *Methods in Ecology and Evolution*, 8, 278-287.

620 Gregorius, H.-R., & Kosman, E. (2018). Structural type diversity: measuring structuredness of  
621 communities by type diversity. *Theoretical Ecology*, 11, 383-394.

622 Grime, J. P. (1997). Ecology - Biodiversity and ecosystem function: The debate deepens. *Science*,  
623 277, 1260-1261.



624 Gultyaeva, E. I., Shaydayuk, E. L., & Kosman, E. G. (2020). Regional and temporal differentiation  
625 of virulence phenotypes of *Puccinia triticina* Eriks. from common wheat in Russia during the  
626 period 2001-2018. Plant Pathology, 69, 860-871.

627 Harrison, S., & Hastings, A. (1996). Genetic and evolutionary consequences of metapopulation  
628 structure. Trends in Ecology & Evolution, 11, 180-183.

629 Hedrick, P. W. (2005). A standardized genetic differentiation measure. Evolution, 59, 1633-1638.

630 Heller, R., & Siegismund, H. (2009). Relationship between three measures of genetic differentiation  
631 GST, DEST and G'ST: how wrong have we been? Molecular Ecology, 18, 2080.

632 Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. Ecology, 54,  
633 427-432.

634 Holsinger, K. E. (1992). Ecological models of plant mating systems and the evolutionary stability of  
635 mixed mating systems. Ecology and evolution of plant reproduction. New approaches (ed. R.  
636 Wyatt), pp. 169-191. Chapman and Hall, New York.

637 Jost, L. (2007). Partitioning diversity into independent alpha and beta components. Ecology, 88,  
638 2427-2439.

639 Jost, L. (2008). G(ST) and its relatives do not measure differentiation. Molecular Ecology, 17, 4015-  
640 4026.

641 Kosman, E. (1996). Difference and diversity of plant pathogen populations: A new approach for  
642 measuring. Phytopathology, 86, 1152-1155.

643 Kosman, E. (2014). Measuring diversity: from individuals to populations: mini-review. European  
644 Journal of Plant Pathology, 138, 467-486.

645 Kosman, E., Ben-Yehuda, P., & Manisterski, J. (2014). Diversity of virulence phenotypes among  
646 annual populations of wheat leaf rust in Israel from 1993 to 2008. Plant Pathology, 63, 563-  
647 571.

648 Kosman, E., Burgio, K. R., Presley, S. J., Willig, M. R., & Scheiner, S. M. (2019). Conservation  
649 prioritization based on trait-based metrics illustrated with global parrot distributions.  
650 Diversity and Distributions, 25, 1156-1165.

651 Kosman, E., & Jokela, J. (2019). Dissimilarity of individual microsatellite profiles under different  
652 mutation models: Empirical approach. Ecol Evol, 9, 4038-4054.

653 Kosman, E., & Leonard, K. J. (2005). Similarity coefficients for molecular markers in studies of  
654 genetic relationships between individuals for haploid, diploid, and polyploid species.  
655 Molecular Ecology, 14, 415-424.

656 Kosman, E., & Leonard, K. J. (2007). Conceptual analysis of methods applied to assessment of  
657 diversity within and distance between populations with asexual or mixed mode of  
658 reproduction. New Phytologist, 174, 683-696.

659 Leache, A. D., Fujita, M. K., Minin, V. N., & Bouckaert, R. R. (2014). Species delimitation using  
660 genome-wide SNP data. Syst Biol, 63, 534-542.

661 Maddison, W. P., & Knowles, L. L. (2006). Inferring phylogeny despite incomplete lineage sorting.  
662 Systematic Biology, 55, 21-30.

663 Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: F(ST) and related  
664 measures. Molecular Ecology Resources, 11, 5-18.

665 Morrissey, M. B., & de Kerckhove, D. T. (2009). The maintenance of genetic variation due to  
666 asymmetric gene flow in dendritic metapopulations. American Naturalist, 174, 875-889.

667 Nei, M. (1972). Genetic distance between populations. American Naturalist, 106, 283-+.

668 Nei, M. (1973). Analysis of gene diversity in subdivided populations. Proceedings of the National  
669 Academy of Sciences of the United States of America, 70, 3321-3323.

670 Nei, M., & Chesser, R. K. (1983). Estimation of fixation indices and gene diversities. Annals of  
671 Human Genetics, 47, 253-259.

672 Pedraza-Marrón, C. d. R., Silva, R., Deeds, J., Van Belleghem, S. M., Mastretta-Yanes, A.,  
 673 Domínguez-Domínguez, O., Rivero-Vega, R. A., Lutackas, L., Murie, D., Parkyn, D.,  
 674 Bullock, L. H., Foss, K., Ortiz-Zuazaga, H., Narváez-Barandica, J., Acero, A., Gomes, G., &  
 675 Betancur-R, R. (2019). Genomics overrules mitochondrial DNA, siding with morphology on  
 676 a controversial case of species delimitation. *Proceedings of the Royal Society B: Biological*  
 677 *Sciences*, 286, 20182924.

678 Rieseberg, L. H., & Burke, J. M. (2001). The biological reality of species: gene flow, selection, and  
 679 collective evolution. *Taxon*, 50, 47-67.

680 Rosenberg, N. A., & Nordborg, M. (2002). Genealogical trees, coalescent theory and the analysis of  
 681 genetic polymorphisms. *Nature Reviews Genetics*, 3, 380-390.

682 Ryman, N., & Leimar, O. (2009). GST is still a useful measure of genetic differentiation - a comment  
 683 on Jost's. *Molecular Ecology*, 18, 2084-2087.

684 Scheiner, S. M., Kosman, E., Presley, S. J., & Willig, M. R. (2017). Decomposing functional  
 685 diversity. *Methods in Ecology and Evolution*, 8, 809-820.

686 Sherwin, W. B., Chao, A., Jost, L., & Smouse, P. E. (2017). Information theory broadens the spectrum  
 687 of molecular ecology and evolution. *Trends in Ecology and Evolution*, 32, 948 - 963.

688 Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies.  
 689 *Genetics*, 139, 457-462.

690 Wang, J., Santiago, E., & Caballero, A. (2016). Prediction and estimation of effective population size.  
 691 *Heredity (Edinb)*, 117, 193-206.

692 Wright, S. (1951) .The genetical structure of populations. *Ann Eugen*, 15, 323-354.

693 Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-  
 694 selection balance. *Evolution*, 65, 1897-1911.