

APPLICATION

dispRity: A modular R package for measuring disparity

Thomas Guillaume^{1,2}¹Department of Life Sciences, Imperial College London, Ascot, UK²School of Biological Sciences, University of Queensland, St. Lucia, Queensland, Australia**Correspondence**

Thomas Guillaume, Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot SL5 7PY, UK.

Email: guillert@tcd.ie

Funding information

European Union Seventh Framework Programme, Grant/Award Number: FP/2007-2013; Discovery Project, Grant/Award Number: DP170103227

Handling Editor: Timothée Poisot

Abstract

1. Biological data is multivariate in essence: many traits in organisms covary with each other in space and time. This causes biologists to either reduce these to a manageable number of variables or, increasingly, to use multivariate toolkits. One such toolkit is based on creating a multidimensional space where the variables are the axes. It is then possible to measure diverse aspects of the distribution of some observation (e.g. species) in this space. For example, if studying morphology, one can create a morphospace for two groups of species, measure the volume occupied by each of these groups and then test whether these two volumes are significantly different or not.
2. There are as many definitions of these multidimensional spaces, metrics and tests as there are questions that can be tackled with such methods. Many of these methods are implemented in specific software or R packages. However, the definition of the space, metric and test is often dependent on the software/package and authors points of view or specific questions. This can unfortunately hamper researchers' ability to apply different methods that best suits their specific questions.
3. Here I present the `dispRity` package, a flexible R package for performing multidimensional analysis. It allows users to define each step of the analysis (whether it is the space, the metric or the test) through a highly modular architecture where each definition can be passed as a function. It also provides a tidy interface through the `dispRity` object, allowing users to easily run reproducible multivariate analysis.
4. The `dispRity` package also comes with an extend manual regularly updated following users' questions or suggestions. Furthermore, the package contains some simulation tools (e.g to simulate complex multidimensional space or morphological data). Finally, it also contains a suite of utility functions to work with `dispRity` objects aimed at helping users to develop their own multidimensional metrics and/or tests.

KEYWORDS

disparity, disparity-through-time, ecology, multidimensionality, ordination, palaeobiology

1 | INTRODUCTION

Biological data are complex. To understand the ecology and evolution of species, we must use multiple variables that inevitably covary

with each other through time and space. One solution to this problem is to analyse these data in a multivariate framework (e.g. Díaz et al., 2016; Price, Friedman, & Wainwright, 2015). Such analyses aim to capture the complex multidimensionality of biological data,

while still providing outputs that are interpretable. These multivariate analyses can be used to investigate changes in morphological diversity through time (e.g. Close, Friedman, Lloyd, & Benson, 2015), competitive replacement scenarios (e.g. Brusatte, Benton, Ruta, & Lloyd, 2008), relationships among form and function (e.g. Díaz et al., 2016) and even to describe the entirety of possible shapes for a group of organisms (e.g. Raup, 1966). The biological variables in such analyses are equally diverse, including morphological traits (discrete traits like the presence or absence of a character, e.g. Close et al., 2015; or continuous traits such as lengths, e.g. Price et al., 2015), life history traits (e.g. Díaz et al., 2016), or even ecosystem properties (e.g. Donohue et al., 2013).

In all these analysis, each set of multivariate traits forms a multidimensional space. This space is represented as a matrix where rows are regarded as samples or observations (e.g. specimens, field sites, etc.) and columns are variables or some transformation thereof (e.g. embedding, scaling, ordination, etc.). These multidimensional spaces can be defined in many ways, for example as a pairwise distance matrix (Lloyd, 2016 and references therein; e.g. in Close et al., 2015), or as outputs from an ordination, whether it being a principal components analysis (PCA, Hotelling, 1933; e.g. in Zelditch, Swiderski, & Sheets, 2012), a metric scaling (PCO, PCoA, Torgerson, 1958; e.g. in Brusatte et al., 2008) or a non-metric scaling (MDS, NMDS, Shepard, 1962; e.g. in Donohue et al., 2013; Liow, 2004). The name we give to the multidimensional space tends to vary with the kinds of traits used to construct it. For example, when using morphological traits, the space will be a morphospace, when using ecological traits it may be referred to as an ecospace, etc.

One can then measure how the observations are distributed within this space to answer related questions (e.g. “does group A occupy more space than group B?”). This requires the definition of a proxy for space occupancy: the disparity metric (or index; Hopkins & Gerber, 2017) which can be measured in a multitude of ways. For example, one could use a metric based on the variance or the range of each axis of space (Ciampaglio, Kemp, & McShea, 2001; Wills, 2001), a distance (e.g. Euclidean) measured between observations (Foote, 1993, 1996), a more direct approximation of the hyper volume (Cornwell, Schwikl, & Ackerly, 2006; Donohue et al., 2013), or many more (e.g. Navarro, 2003).

Finally, all these different multidimensional spaces and their associated disparity metrics can be used in an equal variety of statistical tests such as nonparametric multivariate analyses of variance (NPMANOVA, Anderson, 2001; e.g. in Brusatte et al., 2008) multidimensional permutation tests (Manly, 1997; e.g. in Díaz et al., 2016) or even, less rigorously, by looking at the confidence interval overlaps between disparity measurements. In summary, there are many different ways to perform each step of a multidimensional analysis, making analyses of complexity ever more complex.

In theory, this multitude of ways to generate and define multidimensional spaces, measure disparity within and analyse these metrics is not an issue, in fact, it allows researchers to choose both the most appropriate method for their question or data, or even to test their question using multiple methods. In practice, however, this is

hampered by existing software implementations. Although many software packages exist for multidimensional analysis (e.g. Adams, Collyer, & Kaliontzopoulou, 2018; Adams & Otárola-Castillo, 2013; Bouxin, 2005; De Caceres, Oliva, Font, & Vives, 2007; Harmon, Weir, Brock, Glor, & Challenger, 2008; Lloyd, 2016; Navarro, 2003; Oksanen et al., 2007), package maintainers/software developers choose their preferred definition of multidimensional space and disparity metric to best fit their needs (i.e. data, hypothesis, etc.) making the implementations sometimes hard to adapt to different needs. For example, in the excellent and widely used *geomorph* package, morphological disparity analysis uses the `morphol.disparity` function that defines the multidimensional space as the ordination of the Procrustes transform of the morphometric data, the disparity metric as the relative sum of the diagonal of the covariance of the ordination scores (Procrustes variance), and uses permutation tests (Adams et al., 2018; Adams & Otárola-Castillo, 2013; Zelditch et al., 2012). This is ideal for testing volume based hypothesis (e.g. “does groups A and B have the same volume?”), but in other cases may not be appropriate in non-volume-based hypothesis (e.g. “do they occupy the same location?”). This can lead to inappropriate analyses by users confined by the existing software implementations.

The aim of the *disparity* package is to avoid such problems by providing a flexible framework for studying multidimensional data. This package is based on a modular architecture where each decision in multidimensional analysis (which data, metric and test) can be specified by the user. It implements many commonly used disparity metrics, as well as providing a simple interface for users to implement their own disparity functions. The package is described here for the use of discrete morphological data disparity analysis but can be generalised to any type of multidimensional data (see the glossary Table 1).

2 | DESCRIPTION

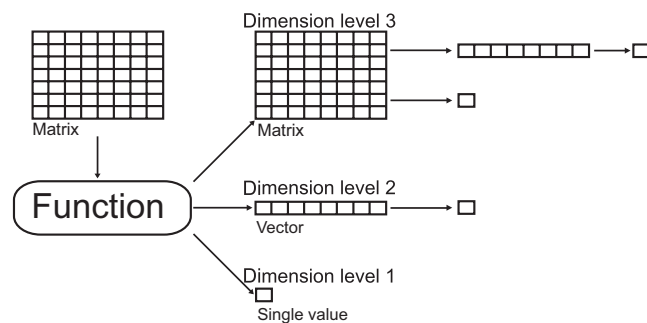
In brief, the package takes a `matrix` object (the multidimensional space), calculates a disparity metric from the space and analyse the resulting `disparity` object through hypothesis testing and visualisation. Some additional functions modify the space, for example by dividing it by groups or through time and/or bootstrapping it (see Figure 2). Note that the input matrix is not restricted to an ordinated matrix, but can be any kind of matrix as long as its rows represent elements (e.g. the space can be a distance matrix: Close et al., 2015). The matrix is always considered as the final multidimensional space to analyse and no correction is applied to it (e.g. potential corrections should be applied prior to using the *disparity* package) (Figure 2).

2.1 | Measuring disparity

The `disparity` function measures disparity from a `matrix` where the columns correspond to the dimensions and the rows correspond to the elements present in the space. The disparity metric is passed

TABLE 1 Glossary and equivalences between this manuscript, the `disPRity` package and terms commonly used in palaeobiology or ecology.

In this manuscript	In <code>disPRity</code>	In palaeobiology	In ecology
Multidimensional space	Matrix ($n \times k$)	Morphospace, traitspace, etc.	Ecospace, function-space, etc.
Elements	Rows (n)	Taxa, specimen, etc.	Taxa, field sites, environments, etc.
Dimensions	Columns (k)	Ordination scores, distances, etc.	Ordination scores, distances, etc.
Subsets	matrix ($m \times k$, with $m \leq n$)	e.g. every element in a stratum or sharing the same ancestor	e.g. elements living in the same environment
Disparity	A dimension-level 1 or 2 function ^a	Disparity: e.g. the sum of variances (Wills, 2001), the average pairwise distances between taxa (Foote, 1994), etc.	Dissimilarity: e.g. ellipsoid volume (Donohue et al., 2013), convex hull volume (Cornwell et al., 2006), etc.

^aSee Figure 1.**FIGURE 1** Illustration of the different metric dimension-levels in the `disPRity` package. In this example, each cell corresponds to a single value (e.g. a 8×7 matrix or a vector of eight elements). A dimension-level 3 matrix would be a metric output a matrix (e.g. the function `stats::cor` to calculate the correlation between each dimension), a dimension-level 2 metric would output a vector (i.e. a distribution, e.g. `disPRity::variances` which calculates the variance within each dimension) and a dimension-level 1 metric would output a single value (e.g. `stats::sd` which calculates the standard deviation of the input matrix)

through the `metric` argument and is defined by the user as one or more function(s) that can either transform the matrix into:

- Another matrix (a dimension-level 3 function—e.g. a variance-covariance matrix; `stats::cor`)
- A vector (a dimension-level 2 function—e.g. the variance of each dimension; `disPRity::variances`—see below)
- A single value (a dimension-level 1 function—e.g. the overall standard deviation; `stats::sd`)

The disparity metrics can be any R function (see Table 2 for metrics implemented in the package). When multiple functions are passed to the `metric` argument, they are sorted by dimension-level and applied in decreasing order to the data. For example, if the metric is defined as `metric = c(prod, ranges)` (the hypercube volume), the `ranges` function (dimension-level 2) is first applied to data and the function `prod` is then applied to the results (`prod(ranges(data))`).

One can also directly pass a function description to the `metric` argument (e.g. `metric = function(x) mean(dist(x)^2)` for the average squared pairwise distance). Note that this function also allows to work on only a subset of dimensions via the `dimensions` argument (e.g. if only the m first dimensions must be considered).

2.2 | Splitting the multidimensional space into subsets

Prior to calculating disparity, the space can be subdivided into subsets, typically to be compared to each other. For example, one may compare the disparity of a specific subset of the space to another or, how different subsets change sequentially (e.g. through time). The original space corresponds to the overall space (e.g. a morphospace contains all the observed morphologies). Subsets correspond to parts of the space with pooled characteristics.

This splitting can be done using the `custom.subsets` or `chrono.subsets` functions. The first function takes a matrix defining the space and a list of elements defining the subsets. The second also takes a matrix and arguments giving the age of the taxa (a dated phylogeny of the elements present in the morphospace—see below) and which subsets to create: (1) discrete time subsets (or time-binning) or (2) continuous time subsets (or time-slicing).

The time-binning method groups elements by specific age range. The time-slicing method works by using a phylogeny and looking at which taxa are present at any specific point in time. This method thus requires the nodes to be part of the space, a dated phylogeny (chronogram) and which model to use when slicing through branches rather than tips and nodes. When a slice occurs not on a tip or a node, six methods are available to select either the descendent or the ancestor's node/tip as an element for this time slice: “acctran”, “deltran”, “random” and “proximity” as proxy for punctuated evolution models; and “equal.split”, “gradual.split” as a proxy for gradual evolution. See Guillerme and Cooper (2018) for full description of the method. Note that there is a trade-off between precision and accuracy when using the time-slicing method: a higher number of slices increases the precision of the disparity analysis but also decreases accuracy.

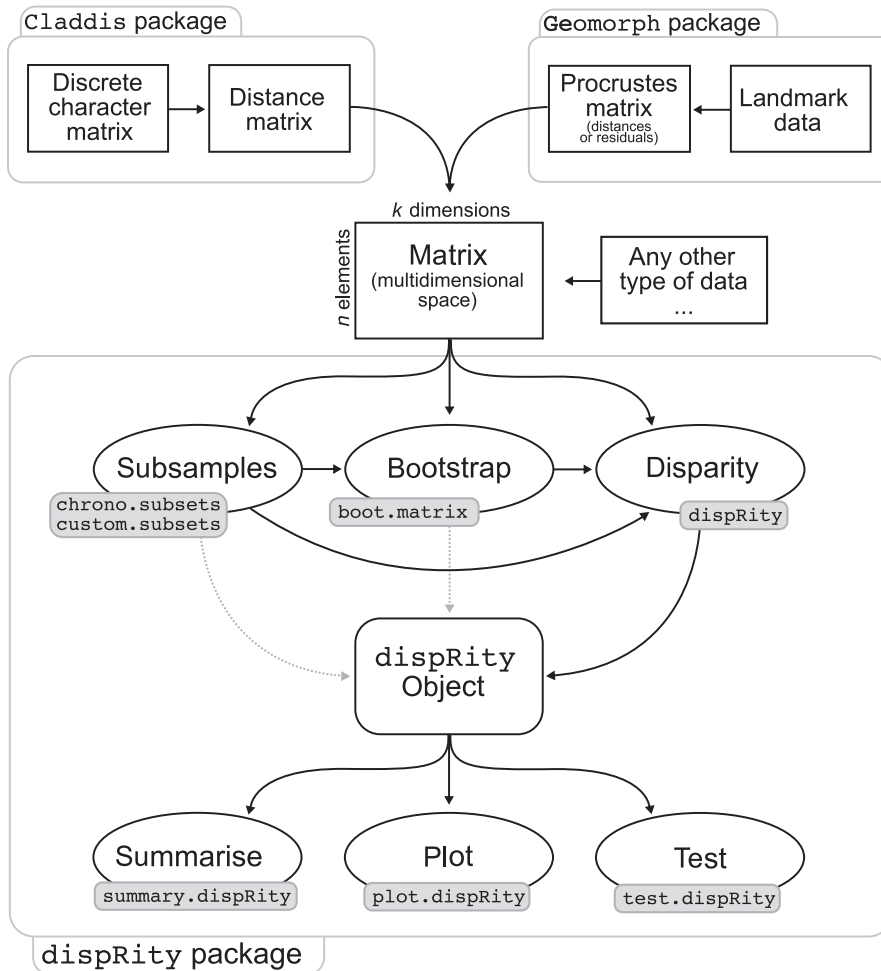


FIGURE 2 `dispRity` package workflow: rectangles represent matrices; ellipses represent functions; plain black arrows indicate input/output; dashed grey arrows indicate output (though the summary, plot, and test function cannot be applied if no disparity has been calculated)

2.3 | Bootstrapping and rarefying

Disparity measurement can be influenced by sampling (Butler, Brusatte, Andres, & Benson, 2012). To take this source of bias into account, one can bootstrap the multidimensional space or/and rarefy the data. Additionally, if disparity is defined as a dimension-level 1 metric, it can be useful to measure it on bootstrapped data to obtain a distribution on which to perform statistical analyses.

Bootstrapping can be achieved by using the `boot.matrix` function which pseudo-replicates the space following two algorithms: (1) the “full” algorithm where the bootstrapping is entirely stochastic (n elements are replaced by any m elements drawn from the data); and (2) the “single” algorithm where $n = 1$ (similar to jackknife).

Similarly, rarefaction can be achieved through the same `boot.matrix` function. In practice, rarefaction limits the number of elements to be drawn for each bootstrap replication: only $n-x$ elements are selected at each bootstrap replicate (where x is the number of non-sampled elements).

2.4 | Interpreting results

The functions above all generate a `dispRity` object that can be summarised or plotted using the S3 method functions `summary`.

`dispRity` and `plot.dispRity`. These results can also be analysed using the `test.dispRity` function for comparing subsets or testing hypotheses.

2.4.1. | Summarising and plotting

The `summary.dispRity` and `plot.dispRity` functions allow users to set which central tendency and which quantiles should be represented. The `plot.dispRity` function graphically represents the summarised results using different representations: (1) “continuous” for displaying continuous disparity curves and (2) “box”, “lines”, or “polygons” to display them using boxplots, confidence interval lines or polygons, respectively. Additional arguments specific to `dispRity` objects can also be used such as `observed` to display the observed disparity (i.e. non-bootstrapped) or `rarefaction` to only plot the disparity for a certain number of elements (i.e. the rarefaction level). The function can also take any additional graphic arguments (`main`, `xlab`, `col`, etc...) from base R.

2.4.2. | Testing hypotheses

The `test.dispRity` function allows users to test hypotheses on the disparity data. Similarly to the `dispRity` function described

TABLE 2 Definition of the disparity metrics currently implemented in the `dispRity` package. k is the number of dimensions, n the number of elements, Γ is the Gamma distribution, λ_i is the eigenvalue of each dimensions, σ^2 is their variance and $Centroid_k$ is their mean, $Ancestor_n$ is the coordinates of the ancestor of element n , $f(vk)$ is a function to select one value from the vector v of the dimension k (e.g. its maximum, minimum and mean, etc.), R is the radius of the sphere or the product of the radii of each dimensions ($\prod_{i=1}^k R_i$ —for a hyper-ellipsoid).

Name	Description	Dim	Definition	Source
<code>ancestral.dist</code>	The distance between an element and its ancestor	2	$\sqrt{\sum_{i=1}^n (k_n - Ancestor_n)^2}$	This package
<code>centroids</code>	The distance between each element and a fixed point ^a of the space	2	$\sqrt{\sum_{i=1}^n (k_n - Centroid_k)^2}$	This package
<code>convhull.surface</code>	The surface of the convex hull	1	NA	<code>geometry::convhulln</code> (Barber, Dobkin & Huhdanpaa, 1996; Habel, Grasman, Gramacy, Stahel, & Sterratt, 2015)
<code>convhull.volume</code>	The volume of the convex hull	1	NA	<code>geometry::convhulln</code> (Barber et al., 1996; Habel et al., 2015)
<code>diagonal</code>	The greatest Euclidean distance	1	$\sqrt{\sum_{i=1}^k (\max(k_i) - \min(k_i))^2}$	This package
<code>ellipse.volume^b</code>	The volume of the ellipsoid	1	$\frac{\pi^{k/2}}{\Gamma(\frac{k}{2}+1)} \prod_{i=1}^k (\lambda_i^{0.5})$	This package; based on Donohue et al. (2013)
<code>mode.val</code>	The modal value	1	NA	This package
<code>n.ball.volume</code>	The hyper-spherical (n -ball) volume	1	$\frac{\pi^{k/2}}{\Gamma(\frac{k}{2}+1)} \prod_{i=1}^k R$	This package
<code>pairwise.dist</code>	The pairwise distances between elements	2	NA	<code>vegan::vegdist</code> (Oksanen et al., 2007)
<code>radius</code>	The radius of each dimensions	2	$ \frac{\sum_{i=1}^n k_i}{n} - f(vk) $	This package
<code>ranges</code>	The absolute ranges of each dimension	2	$ \max(k_i) - \min(k_i) $	This package
<code>span.tree.length</code>	The minimal spanning tree length	1	$\sum(\text{branch length})$	<code>vegan::spantree</code> (Oksanen et al., 2007)
<code>variances</code>	The variance of each dimension	2	$\sigma^2 k_i$	This package

^aBy default that point is the centroid of the elements

^bThis function uses a fast estimation of the eigenvalue that only works in an ordinated space based on MDS or PCO/PCoA (not PCA)

above, this function can take any test defined by the user or from other R packages. The `comparison` arguments indicates in which order (if any) the tests should be applied to the subsets: (1) “pairwise” for pairwise comparisons; (2) “referential” for comparing the first subset to all the others; (3) “sequential” for comparing subsets sequentially (e.g. first against second, second against third, etc.); (4) “all” for comparing all the subsets simultaneously (i.e. `disparity ~ subsets`) or (5) any list of pairs of subsets to compare.

Some tests are implemented within the package such as the Bhattacharyya Coefficient (`bhatt.coef`; Bhattacharyya, 1943; Guillarme & Cooper, 2016), a permutation test based on null hypothesised multidimensional space following (`null.test`; Díaz et al. 2016; Manly 1997) as well as a wrapper for the `vegan::adonis` (Oksanen et al., 2007) and `geiger::dtt` (Harmon et al., 2008) functions (respectively `adonis.dispRity` and `dtt.dispRity`). This function also allows additional arguments such as `rarefaction` (as described above) or `correction` to adjust p -values when using multiple parametric tests.

3 | EXAMPLES

Multivariate analysis can be really useful for looking at multiple aspects of organisms' diversity together. For example, one can also look the diversity of morphologies (or disparity; Foote, 1991). Using disparity, it is then also possible to assess whether one ecosystem or/and time period display more morphological variation. The following example is based on a classical morphological disparity analysis. Note that more examples are available in the package manual (https://rawgit.com/TGuillarme/dispRity/master/inst/gitbook/_book/index.html).

3.1 | dispRity data

The package contains a dataset that is a subset from Beck and Lee (2014) and includes the following:

- `BeckLee_mat50`: an ordinated matrix for 50 mammals based on the distance between discrete morphological characters.

TABLE 3 Summarising a `disprity` object (disparity per groups). *n* is the number of elements per subsets, *obs* the observed disparity (not bootstrapped), *bs.median* is the median bootstrapped disparity (here the median of the sum of variances) and the 2.5%, 25%, 75% and 97.5% are the confidence intervals

	Subsets	<i>n</i>	<i>Obs</i>	<i>bs.median</i>	2.5%	25%	75%	97.5%
1	Crown	30	2.00	1.93	1.87	1.92	1.95	1.98
2	Stem	20	1.72	1.63	1.53	1.60	1.66	1.69

- `BeckLee_mat99`: the same matrix `BeckLee_mat50` with the reconstruction of their 49 ancestors.
- `BeckLee_tree`: a chronogram with the 50 mammal species present in `BeckLee_mat50` and `BeckLee_mat99`.
- `BeckLee_ages`: the first and last occurrence data for 14 of the mammal species present in `BeckLee_mat50` and `BeckLee_mat99`.
- `disparity`: a pre-analysed `disprity` object based on the data above.

In this example, the space is defined as a morphospace: the ordination of the distances among discrete morphological characters for 50 mammal species (Beck & Lee, 2014). Additionally, we can define disparity as the sum of the variances on each dimension (Foote, 1991; Wills, Briggs, & Fortey, 1994) that will represent an aspect of the volume of the morphospace.

3.2 | Typical disparity among groups analysis

One typical question with such analysis would be to test whether two groups of species have a different disparity. For example, using the data described above, we can test whether the crown mammals are more diverse in term of morphology than the stem ones. In other words, whether the approximation of the volume within the morphospace is different in crown or stem mammals. These two groups can be defined using one of the package's utility functions, `crown.stem` that separate the crown and stem species given a phylogeny (allowing to ignore the nodes or not):

```
> mammal_groups <- crown.stem(BeckLee_tree, inc.nodes = FALSE)
```

It is then possible to measure the disparity between the two groups as follows:

```
> disparity <- disprity.per.group(data = BeckLee_mat50,
  group = mammal_groups, metric = c(sum, variances))
```

Note that this function is a wrapper function that is the equivalent to:

```
> subsets <- custom.subsets(data = BeckLee_mat50, group = mammal_groups)
> bootstraps <- boot.matrix(subsets)
> disparity2 <- disprity(bootstraps, metric = c(sum, variances))
```

Which allows fine tuning of the optional arguments in each function. The three arguments here are defined as follows: `data = BeckLee_mat50` is our space, `group = mammal_groups` indicates which mammals belong to which group and `metric =`

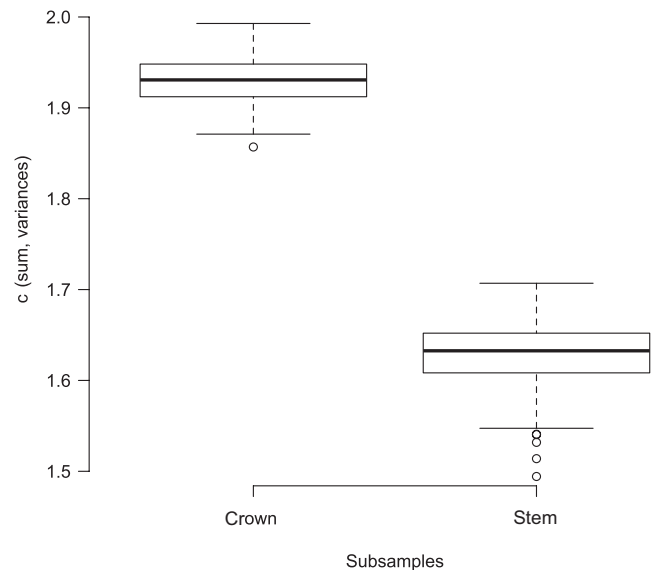


FIGURE 3 `disprity` plot of disparity differences between groups

`c(sum, variances)` is our definition of disparity (Ciampaglio et al., 2001; Foote, 1991; Wills et al., 1994).

This function returns a `disprity` object that summarises the disparity analysis:

```
> disparity2
```

```
-- disprity object --
```

2 customised subsets for 50 elements with 48 dimensions:

```
crown, stem.
```

Data was bootstrapped 100 times (method:"full").

Disparity was calculated as: `c(sum, variances)`.

As indicated, the `disprity` object contains two customised subsets from a morphospace made of 50 elements for 48 dimensions. The `disprity` object also displays information on the number and method of the bootstrap replicates as well as the definition of disparity. To visualise the actual disparity values, one can use the `summary` or/and `plot` functions (Table 3 and Figure 3):

```
> summary(disparity2)
```

```
> plot(disparity2)
```

As we can see from the summary table (Table 3) and the plot (Figure 3), there seems to be a significant difference in morphospace volume occupied between the two groups. It is possible to test this

TABLE 4 Summarising a `dispRity` object (disparity through time). *n* is the number of elements per subsets, *obs* the observed disparity (not bootstrapped), *bs.median* is the median bootstrapped disparity (here the median of the sum of variances) and the 2.5%, 25%, 75% and 97.5% are the confidence intervals

	Subsets	<i>n</i>	<i>Obs</i>	<i>bs.median</i>	2.5%	25%	75%	97.5%
1	100.5-66	15	1.67	1.55	1.40	1.51	1.58	1.65
2	66-56	9	1.88	1.69	1.43	1.63	1.77	1.83
3	56-33.9	13	1.96	1.83	1.62	1.77	1.86	1.90

hypothesis by using, for example, a nonparametric Wilcoxon test (`stats::wilcox.test`):

```
> test.dispRity(disparity2, test = wilcox.test)
```

```
[[1]]
```

```
statistic: W
```

```
crown : stem 10000
```

```
[[2]]
```

```
p.value
```

```
crown : stem 2.562144e-34
```

As indicated by the *p* value, there is a significant difference in disparity between the groups. Note that by default the function only outputs the test's statistic, parameter (if parametric) and the *p* value. However, the raw test results can also be output using the option `details = TRUE` in the function above. Additionally, the test is here performed on the pooled bootstrapped pseudo-replications which can increase the type I error. It is possible to compare each bootstrap in a pairwise way without pooling the data by using the `concatenate = FALSE` argument. The results will then be a distribution of statistics and *p* values. Relating back to our question: yes, crown mammals display a higher diversity in morphologies than their stem counterparts (in this example and dataset).

3.3 | Typical disparity-through-time analysis

A subsequent question to this observation could be to test whether this difference is due to an overall change in disparity through time or not. Using the same definition of the multidimensional space and disparity as in the previous example, we can measure, for example, changes in disparity through time between the Late Cretaceous (100.5–66.0 million years ago—Mya), the Paleocene (66.0–56.0 Mya) and the Eocene (56.0–33.9 Mya). Note that stratigraphic times can be generated automatically, using the `get.bin.ages` utility function.

```
> time_bins <- c(100.5, 66, 56, 33.9)
```

It is then possible to measure disparity-through-time using the following function:

```
> disparity3 <- dispRity.through.time(data = BeckLee_mat50,
  tree = BeckLee_tree, time = time_bins, metric = c(sum, variances))
```

Note that this function is a wrapper function that is the equivalent to:

```
> time_subsets <- chrono.subsets(data = BeckLee_mat50, tree = BeckLee_tree,
  time = time_bins, method = "discrete")
> bootstraps2 <- boot.matrix(time_subsets)
> disparity4 <- dispRity(bootstraps2, metric = c(sum, variances))
```

The arguments `data = BeckLee_mat50` and `metric = c(sum, variances)` are the same as in the example above. However, in this type of analysis, we also need to have additional arguments: the `time = time_bins` indicates the boundaries of the different time bins, the `tree = BeckLee_tree` argument provides information on the age of each element and `method = "discrete"` indicates that the data is time-binned. The resulting `dispRity` object can be summarised and plotted (Table 4 and Figure 4):

```
> summary(disparity4)
```

```
> plot(disparity4, type = "continuous")
```

Note that many plot options specific to `dispRity` objects are available such as plotting disparity in a "continuous" fashion (inferring disparity between the time bins).

Similarly to the example above, it is also possible to statistically test this hypothesis using, for example, multivariate permutation ANOVA (PERMANOVA; Anderson, 2001) through the `adonis`.

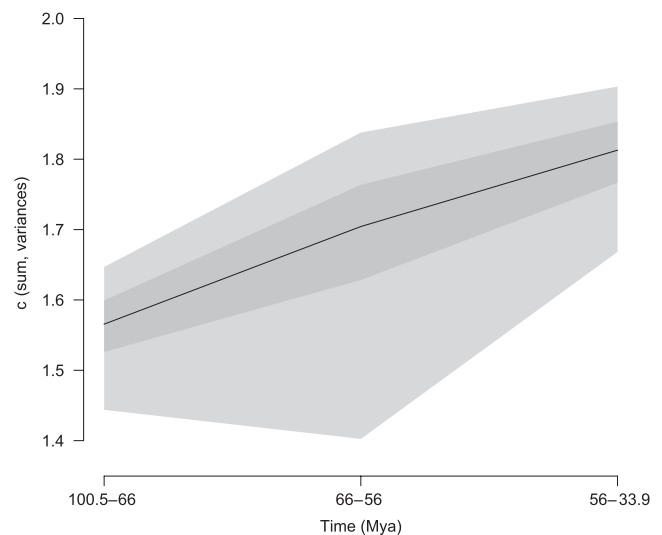


FIGURE 4 `dispRity` plot of disparity-through-time. The black line represents the median disparity (median sum of variances), the dark grey and light surfaces represent respectively the 50% and 95% confidence intervals

TABLE 5 Raw PERMANOVA output from the `adonis.dispRity` function: Call: `vegan::adonis(formula = dist(matrix) ~ time, data = disparity2, method = "euclidean")`; Permutation: free; Number of permutations: 999; Terms added sequentially (first to last). Signif. codes: 0 '***'

	df	Sum Sq	Mean Sq	F model	R ²	Pr(>F)
Time	2	7.50	3.75	2.06	0.11	<0.01***
Residuals	34	61.82	1.81		0.89	
Total	36	69.32			1.00	

`dispRity` function that is a wrapper of the `vegan::adonis` function (Oksanen et al., 2007) for `dispRity` objects:

```
> test.dispRity(disparity4, test = adonis.dispRity)
```

To answer our specific question above: yes, there is an effect of time on morphological disparity (an increase) in this dataset (Table 5). Note that in this case, the function outputs different warnings on the usage of such test and the eventual data not used in the test. Additionally, the test is not applied to the bootstrapped data and thus might be sensitive to outliers and sampling size.

4 | ADDITIONAL INFORMATION

4.1 | Manuals and vignette

Supplementary information concerning the package and each function can be found in R, on the project page (<https://github.com/tguillerme/disprity>) or in the online manual (https://rawgit.com/TGuillerme/disprity/master/inst/gitbook/_book/index.html). This manual contains substantially more information and detailed examples including a tutorial for a "classic" disparity analysis in palaeobiology as well as an introduction to the use of this package in ecology or other disciplines.

4.2 | Data simulations

This package also contains functions for simulating random discrete morphological matrices (`sim.morpho`) or random multidimensional spaces (`space.maker`). These functions are based on a similar modular architecture as that used by the `dispRity` functions, allowing users to provide their own distribution parameters for the simulations. For example, `stats::rnorm` can be provided as an argument for drawing normal characters rates with `sim.morpho` or normally distributed spaces with `space.maker`. The discrete morphological data simulations are based on protocols from Guillaume and Cooper (2016), O'Reilly et al. (2016) and Puttick et al. (2017). The space simulations are based on the methods from Díaz et al. (2016). Both functionalities are described in more details in the package manual.

5 | CONCLUSION

The `dispRity` package is based on a modular architecture allowing researchers to simply define both their multidimensional space and

their disparity metric to efficiently analyse multivariate data. The `dispRity` object allows users to pipeline disparity analysis from the data input (the matrix) to publication standard results (tables, plots, hypothesis testing).

6 | PACKAGE LOCATION

The `dispRity` package is available on the CRAN at <https://cran.r-project.org/web/packages/disprity/index.html> or on GitHub at <https://github.com/TGuillerme/disprity> with more associated information. All the versions of the package are archived on ZENODO with associated DOI <https://zenodo.org/record/1186467#.WtfbGsi-kW8>.

ACKNOWLEDGEMENTS

Many thanks to Natalie Cooper for encouraging and helping with the writing of this paper and the package manuals. Thanks to David Bapst, Martin Brazeau, Rompy Chompee, Andrew Jackson, Graeme Lloyd and Emma Sherratt and to Michael Collyer, Gavin Simpson and two other anonymous reviewers for comments on the package and manuscript. I acknowledge support from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement number 311092 awarded to Martin Brazeau and from the Australian Research Council Discovery Project Grant DP170103227 awarded to Vera Weisbecker.

DATA ACCESSIBILITY

The `dispRity` package is available on the CRAN at <https://cran.r-project.org/web/packages/disprity/index.html> or on GitHub at <https://github.com/TGuillerme/disprity> with more associated information. All the versions of the package are archived on ZENODO with associated DOI <https://zenodo.org/record/1186467#.WtfbGsi-kW8>.

REFERENCES

- Adams, D. C., Collyer, M. L., & Kaliontzopoulou, A. (2018). Geomorph: Software for geometric morphometric analyses. R package version 3.0.6. Retrieved from <https://cran.r-project.org/package=geomorph>
- Adams, D. C., & Otárola-Castillo, E. (2013). geomorph: An R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, 4, 393–399. <https://doi.org/10.1111/2041-210x.12035>

- Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26, 32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>
- Barber, C. B., Dobkin, D. P., & Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)*, 22, 469–483. <https://doi.org/10.1145/235815.235821>
- Beck, R. M., & Lee, M. S. (2014). Ancient dates or accelerated rates? Morphological clocks and the antiquity of placental mammals. *Proceedings of the Royal Society B: Biological Sciences*, 281, 1–10. <https://doi.org/10.1098/rspb.2014.1278>
- Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35, 99–109.
- Bouxin, G. (2005). Ginkgo, a multivariate analysis package. *Journal of Vegetation Science*, 16, 355–359. <https://doi.org/10.1111/j.1654-1103.2005.tb02374.x>
- Brusatte, S. L., Benton, M. J., Ruta, M., & Lloyd, G. T. (2008). Superiority, competition, and opportunism in the evolutionary radiation of dinosaurs. *Science* 321, 1485–1488. <https://doi.org/10.1126/science.1161833>
- Butler, R. J., Brusatte, S. L., Andres, B., & Benson, R. B. J. (2012). How do geological sampling biases affect studies of morphological evolution in deep time? A case study of pterosaur (Reptilia: Archosauria) disparity. *Evolution* 66, 147–162. <https://doi.org/10.1111/j.1558-5646.2011.01415.x>
- Ciampaglio, C. N., Kemp, M., & McShea, D. W. (2001). Detecting changes in morphospace occupation patterns in the fossil record: Characterization and analysis of measures of disparity. *Paleobiology*, 71, 695–715. [https://doi.org/10.1666/0094-8373\(2001\)027<0695:dcimop>2.0.co;2](https://doi.org/10.1666/0094-8373(2001)027<0695:dcimop>2.0.co;2)
- Close, R., Friedman, M., Lloyd, G., & Benson, R. (2015). Evidence for a mid-Jurassic adaptive radiation in mammals. *Current Biology*, 25, 2137–2142. <https://doi.org/10.1016/j.cub.2015.06.047>
- Cornwell, W. K., Schwikl, D. W., & Ackerly, D. D. (2006). A trait-based test for habitat filtering: Convex hull volume. *Ecology*, 87, 1465–1471. [https://doi.org/10.1890/0012-9658\(2006\)87\[1465:atthf\]2.0.co;2](https://doi.org/10.1890/0012-9658(2006)87[1465:atthf]2.0.co;2)
- De Caceres, M., Oliva F., Font, X., & Vives, S. (2007). Ginkgo, a program for non-standard multivariate fuzzy analysis. *Advances in Fuzzy Sets and Systems*, 2, 41–56. <https://doi.org/10.1111/j.1654-1103.2005.tb02374.x>
- Díaz, S., Kattge, J., Cornelissen, J. H., Wright, I. J., Lavorel, S., Dray, S., ... Gorné, L. D.. (2016). The global spectrum of plant form and function. *Nature*, 529, 167–171. <https://doi.org/10.1038/nature16489>
- Donohue, I., Petchey, O. L., Montoya, J. M., Jackson, A. L., McNally, L., ... Emmerson, M. C. (2013). On the dimensionality of ecological stability. *Ecology Letters*, 16, 421–429.
- Foote, M. (1991). Morphological and taxonomic diversity in clade's history: The blastoid record and stochastic simulations. Contributions from the Museum of Paleontology, University of Michigan. Vol. 28, pp. 101–140.
- Foote, M. (1993). Contributions of individual taxa to overall morphological disparity. *Paleobiology*, 19, 403–419. <https://doi.org/10.1017/s0094837300014056>
- Foote, M. (1994). Morphological disparity in Ordovician–Devonian Crinoids and the early saturation of morphological space. *Paleobiology*, 20, 320–344. <https://doi.org/10.1017/s009483730001280x>
- Foote, M. (1996). Ecological controls on the evolutionary recovery of post-Paleozoic Crinoids. *Science*, 274, 1492–1495. <https://doi.org/10.1126/science.274.5292.1492>
- Guillerme, T., & Cooper, N. (2016). Effects of missing data on topological inference using a Total Evidence approach. *Molecular Phylogenetics and Evolution*, 94 (Part A), 146–158. <https://doi.org/10.1016/j.ympev.2015.08.023>
- Guillerme, T., & Cooper, N. (2018). Time for a rethink: Time sub-sampling methods in disparity-through-time analyses. *Palaeontology*. <https://doi.org/10.1111/pala.12364>
- Habel, K., Grasman, R., Gramacy, R. B., Stahel, A., & Sterratt, D. C. (2015). geometry: Mesh Generation and Surface Tesselation. R package version 0.3-6. Retrieved from <https://CRAN.R-project.org/package=geometry>
- Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., & Challenger, W. (2008). Geiger: Investigating evolutionary radiations. *Bioinformatics*, 24, 129–131. <https://doi.org/10.1093/bioinformatics/btm538>
- Hopkins, M. J., & Gerber, S. (2017). *Morphological disparity*. Cham, Switzerland: Springer International Publishing.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417. <https://doi.org/10.1037/h0070888>
- Liow, L. H. (2004). A test of Simpson's "rule of the survival of the relatively unspecialized" using fossil crinoids. *The American Naturalist*, 164, 431–443. <https://doi.org/10.1086/423673>
- Lloyd, G. T. (2016). Estimating morphological diversity and tempo with discrete character-taxon matrices: Implementation, challenges, progress, and future directions. *Biological Journal of the Linnean Society* 118, 131–151. <https://doi.org/10.1111/bij.12746>
- Manly, B. F. J. (1997). *Randomization, bootstrap and Monte Carlo methods in biology*. London, UK: Chapman and Hall.
- Navarro, N. (2003). MDA: A matlab-based program for morphospace-disparity analysis. *Computers & Geosciences*, 29, 655–664. [https://doi.org/10.1016/s0098-3004\(03\)00043-8](https://doi.org/10.1016/s0098-3004(03)00043-8)
- Oksanen, J., Kindt, R., Legendre, P., O'ÁZ'Hara, B., Stevens, M. H. H., Oksanen, M. J., & Suggests, M. (2007). The vegan package. *Community Ecology Package*, 10, 631–637.
- O'Reilly, J. E., Puttick, M. N., Parry, L., Tanner, A. R., Tarver, J. E., Fleming, J., ... Donoghue, P. C. J. (2016). Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biology Letters*, 12, pii: 20160081. <https://doi.org/10.1098/rsbl.2016.0081>
- Price, S., Friedman, S., & Wainwright, P. (2015). How predation shaped fish: The impact of fin spines on body form evolution across teleosts. *Proceedings of the Royal Society B*, 282, 20151428. <https://doi.org/10.1098/rspb.2015.1428>
- Puttick, M. N., O'Reilly, J. E., Tanner, A. R., Fleming, J. F., Clark, J., Holloway, L., ... Donoghue, P. C. (2017). Uncertain-tree: Discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proceedings of the Royal Society B*, 284, pii: 20162290. <https://doi.org/10.1098/rspb.2016.2290>
- Raup, D. M. (1966). Geometric analysis of shell coiling: General problems. *Journal of Paleontology*, 40, 1178–1190.
- Shepard, R. N. (1962). The analysis of proximities: multidimensional scaling with an unknown distance function. *Psychometrika*, 27, 125–140. <https://doi.org/10.1007/bf02289621>
- Torgerson, W. S. (1958). *Theory and methods of scaling*. Oxford, UK: Wiley.
- Wills, M. (2001). Morphological disparity: A primer. In J. M. Adrain, G. D. Edgecombe, & B. S. Lieberman (Eds.), *Fossils, phylogeny, and form: An analytical approach*. Boston, MA: Springer US.
- Wills, M. A., Briggs, D. E. G., & Fortey, R. A. (1994). Disparity as an evolutionary index: A comparison of Cambrian and recent arthropods. *Paleobiology*, 20, 93–130. <https://doi.org/10.1017/s009483730001263x>
- Zelditch, M. L., Swiderski, D. L., & Sheets, H. D. (2012). *Geometric morphometrics for biologists: A primer*. Waltham, MA: Academic Press.

How to cite this article: Guillerme T. dispRity: A modular R package for measuring disparity. *Methods Ecol Evol*. 2018;00:1–9. <https://doi.org/10.1111/2041-210X.13022>