APPLICATION

CODYN: An R package of community dynamics metrics

Lauren M. Hallett[1,2]*, Sydney K. Jones[3], A. Andrew M. MacDonald[4], Matthew B. Jones[5], Dan F. B. Flynn[6], Julie Ripplinger[7], Peter Slaughter[5], Corinna Gries[8] and Scott L. Collins[3]

[1]*Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA;* [2]*Institute of Arctic and Alpine Research, University of Colorado Boulder, Boulder, CO 80303, USA;* [3]*Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA;* [4]*Department of Zoology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada;* [5]*National Center for Ecological Analysis and Synthesis, University of California Santa Barbara, 735 State St., Suite 300, Santa Barbara, CA 93101, USA;* [6]*The Arnold Arboretum of Harvard University, 1300 Centre Street, Boston, MA 02131, USA;* [7]*School of Life Sciences, Arizona State University, PO Box 874601, Tempe, AZ 85287, USA; and* [8]*University of Wisconsin, 223A Center for Limnology, 680 North Park Street, Madison, WI 53706, USA*

Summary

1. New analytical tools applied to long-term data demonstrate that ecological communities are highly dynamic over time. We developed an R package, `library("codyn")`, to help ecologists easily implement these metrics and gain broader insights into ecological community dynamics.

2. `library("codyn")` provides temporal diversity indices and community stability metrics. All functions are designed to be easily implemented over multiple replicates.

3. Temporal diversity indices include species turnover, mean rank shifts and rate of community change over time. Community stability metrics calculate overall stability and patterns of species covariance and synchrony over time, and include a null-modelling method to test significance.

4. Finally, `library("codyn")` contains vignettes that describe methods and reproduce figures from published papers to help users contextualize and apply functions to their own data.

Key-words: biodiversity analysis, cyclic shift, long-term data, mean rank shift, R package, rank clock, synchrony, temporal analysis, turnover, variance ratio

Introduction

As long-term data sets increase in scope and length, new analytical tools are being developed to capture the patterns of community dynamics over time. This has the potential to deepen our understanding of ecological patterns and processes. For example, many traditional measurements of community structure, such as diversity indices and rank-abundance curves, represent 'snapshots in time' that poorly capture community dynamics (Collins *et al.* 2008). The development of analogous temporal metrics, such as species turnover and rank shifts, has highlighted how much communities can vary over time (Collins, Micheli & Hartt 2000; Collins *et al.* 2008; Cleland *et al.* 2013). Similarly, species dynamics can affect the stability of community properties such as total cover and net primary production. A growing number of synthesis studies have applied new metrics to long-term data sets to test what factors influence the stability of ecosystems (Houlahan *et al.* 2007; Grman *et al.* 2010; Hector *et al.* 2010; Hallett *et al.* 2014). As ecological records continue to grow and data storage and sharing barriers diminish, it is likely that these types of long-term studies will drive new advances in ecology.

However, there are several impediments to the broad adoption of temporal analyses by the ecological community. First, many metrics of community dynamics have not yet been published as functions in R. R is becoming the programming language of choice for most ecologists; however, new users often rely on published instead of self-written functions. The primary R package for vegetation analyses, `library("vegan")`, provides a well-stocked toolbox of diversity indices and ordination techniques that may be applied over time (Oksanen *et al.* 2015), and other packages such as `library("FD")` provide indices of functional diversity (Laliberté & Legendre 2010; Laliberté, Legendre & Shipley 2014), but neither include the metrics provided in `library("codyn")`. Thus, an ecologist seeking to measure temporal dynamics needs to write much of the code from scratch. Second, most temporal metrics are applied to data sets that include multiple replicates as well as time points. This adds an additional layer of complexity that may stall their general use and application. Third, many metrics of temporal dynamics are actively being developed, and so ecologists may be unfamiliar with the metrics available and their relative merits.

Here, we present a new R package, `library("codyn")`, that fills this niche for published functions that characterize communities over time (Hallett *et al.* 2015). Short

*Correspondence author. E-mail: lauren.m.hallett@gmail.com

for 'community dynamics', `library("codyn")` functions are explicitly temporal and can be easily implemented over multiple replicates or experimental treatments. Functions in `library("codyn")` fall into two categories: temporal diversity indices and community stability metrics. Below, we describe the metrics by category and demonstrate their application with examples from the Konza Prairie Long-Term Ecological Research (LTER) Site, in Kansas, USA.

Temporal diversity indices

The diversity indices in `library("codyn")` are temporal analogues to traditional diversity indices such as richness and rank-abundance curves (Table 1). They include species turnover (`turnover()`), mean shifts in rank abundance (`rank_shift()`) (Collins *et al.* 2008) and rate of directional change in community composition over time (`rate_change ()`) (Collins, Micheli & Hartt 2000). Functions in `library ("codyn")` require a data frame (`df`) with abundance data and grouping variables for time, species and, optionally, replicate. This format reflects a 'tidy' data structure; see Wickham (2014), `library("tidyr")` (Wickham 2015) and `library ("dplyr")` (Wickham & Francois 2015) for guidelines and code for tidying and manipulating data. Variables in `library ("codyn")` are consistently specified with the `time.var`, `species.var`, `abundance.var` and `replicate.var` parameters (Table 1).

SPECIES TURNOVER

Species turnover represents a temporal analogue to species richness (MacArthur & Wilson 1963; Diamond 1969; Collins *et al.* 2008; Cleland *et al.* 2013). The function `turnover` calculates three metrics of species turnover: total turnover, appearances and disappearances. The default metric `total` refers to total turnover, which calculates the proportion of species that differ between time points as:

$$\text{Total turnover} = \frac{\text{Species gained} + \text{Species lost}}{\text{Total species observed in both timepoints}}$$

The `turnover` function includes the option to calculate only the proportion of species that appear or only those that

Table 1. Temporal diversity and community stability functions in `library("codyn")`.

CODYN function
turnover (df, time.var, species.var, abundance.var, replicate.var, metric)
rank_shift(df, time.var, species.var, abundance.var, replicate.var)
rate_change (df, time.var, species.var, abundance.var, replicate.var)
variance_ratio(df, time.var, species.var, abundance.var, bootnumber, replicate.var, li, ui, average.replicates)
synchrony (df, time.var, species.var, abundance.var, metric, replicate.var)

disappear. This allows the detection of differences in the time points in which many species appear vs. when species drop from the system, even while the total turnover value in both scenarios may be similar.

MEAN RANK SHIFTS

Mean rank shifts represent a temporal analogue of species rank-abundance distributions and indicate the degree of species reordering between two time points (Collins *et al.* 2008). This metric is calculated by the `rank_shift()` function as:

$$\text{MRS} = \sum_{i=1}^{N} (|R_{i,t+1} - R_{i,t}|)/N$$

where N is the number of species in common in both time points, t is the time pointand $R_{i,t}$ is the relative rank of species i at time t.

RATE OF COMMUNITY CHANGE

The `rate_change()` function provides a measure of the rate and direction of change in a community. Community compositional changes are described by Euclidean distances, which are calculated on pair-wise communities across the entire time series. For example, a data set with six time intervals will have distance values for five-one-interval time lags (e.g. t_1 vs. t_2, t_2 vs. t_3...), four-two-interval time lags (e.g. t_1 vs. t_3, t_2 vs. t_4...) and so forth. These distance values are regressed against the time lag interval. The slope of this relationship indicates the rate and direction of community change (Collins, Micheli & Hartt 2000). For use in graphing the data underlying this metric, the full data set of lagged distances is returned using the function `rate_change_interval()`.

WORKED EXAMPLE: TEMPORAL DIVERSITY PATTERNS AT KONZA

The Konza Prairie LTER is a tallgrass prairie research site in which different watersheds have experienced different experimental fire frequencies. Collins *et al.* (2008) compared the temporal dynamics of grassland plant species in areas of Konza that were either burned annually or unburned. Species abundance data were collected over 18 years in 20 permanently marked 10 m^2 circular subplots at an annually burned and an unburned location. To demonstrate `library ("codyn")`, we utilize the Collins *et al.* (2008) data set, presented in two forms: `collins08`, which averages species abundances across subplots within the burn treatment; and `knz_001d`, which includes all subplots from the annually burned watershed.

`library("codyn")` examples of temporal diversity indices utilize the `collins08` data set:

```
library(codyn)
data("collins08")
head(collins08[1:3,])
##       replicate year species abundance
```

```
## 1 annually burned 1984 achimill  0.05
## 2 annually burned 1984 ambrpsil  1.35
## 3 annually burned 1984 amorcane  1.32
```

Visualizing temporal dynamics with rank clocks

As one method to visualize temporal community change `library("codyn")` includes a vignette to produce 'rank clocks' using `library("ggplot2")`, which plot the rank order of abundance of each species over time in a circle, starting with a vertical axis at 12 o'clock (Collins *et al.* 2008). Rank clocks highlight that there has been tremendous reordering in the relative abundance of dominant species in the annually burned but not the unburned location at Konza. For example, big bluestem (*Andropogon gerardii*) decreased substantially in the annually burned plot over time but remained stable and dominant in the unburned plot (Fig. 1). Code to reproduce this and subsequent figures is in Data S1 (Supporting Information).

Quantifying temporal dynamics

Turnover. Average annual species richness was higher in the unburned site at Konza, although richness varied a fair amount over time in both sites (Fig. 2a). Total turnover indicates that, in addition to higher overall richness, there were also greater fluctuations in the species present in the unburned than burned location at Konza (Fig. 2b). This insight helps to contextualize the richness patterns; even between consecutive years with similar species richness, the identity of the composite species could differ by 30%. Higher species richness in the unburned plots can also be related to the higher overall proportional species appearances in the unburned plots (Fig. 2b). Below is an example of the code to calculate turnover; the remaining functions in `library`
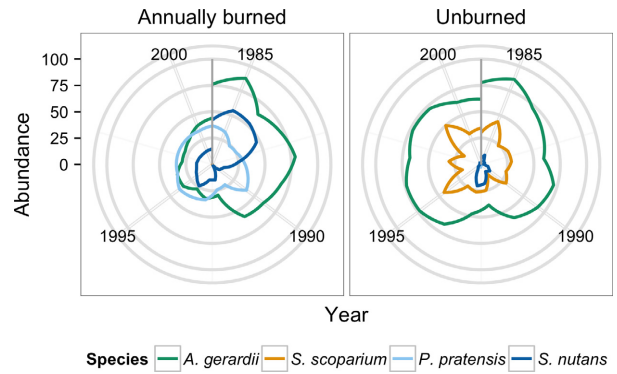
Fig. 1. Rank clock plots of the four dominant grassland plant species in the annually burned vs. unburned locations at Konza Prairie, a long-term ecological research site in Kansas, USA. Vertical black bars show the starting '12 o'clock' position on the rank clock.

(`"codyn"`) follow a similar syntax. Note that `year` in the table below refers to the second year in a 2-year comparison; thus, `1985` refers to the turnover from 1984 to 1985.

```
KNZ_turnover <- turnover(df = collins08,
         time.var = "year",
         species.var = "species",
         abundance.var = "abundance",
         replicate.var = "replicate",
         metric = "total")
KNZ_turnover <- KNZ_turnover[order(KNZ_turnover
$year),]
head(KNZ_turnover[1:3,])
## total year replicate
## 0.237 1985 annually burned
## 0.217 1985 unburned
## 0.302 1986 annually burned
```

```
library ("codyn") syntax

turnover(df = collins08,
  time.var = "year",
  species.var = "species",
  abundance.var = "abundance",
  replicate.var = "replicate",
  metric="total")

Variations: metric = "appearance"'
      metric = "disappearance"

rank_shift(df = collins08,
  time.var = "year",
  species.var = "species",
  abundance.var = "abundance",
  replicate.var = "replicate")

rate_change(df = collins08,
  time.var = "year",
  species.var = "species",
  abundance.var = "abundance",
  replicate.var = "replicate")

rate_change_interval(df = collins08,
  time.var = "year",
  species.var = "species",
  abundance.var = "abundance",
  replicate.var = "replicate")
```
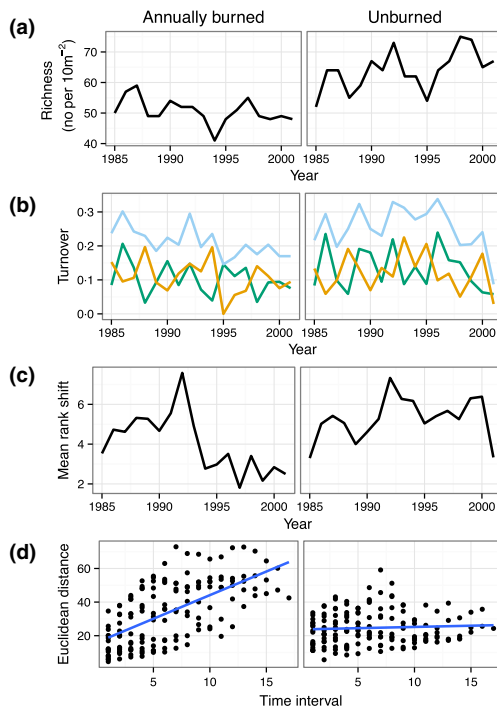

Fig. 2. (a) Species richness, (b) turnover (total = light blue, appearances = dark green, disappearances = yellow), (c) mean rank shifts and (d) community change over time for plant species at the annually burned and unburned location at Konza Prairie, a long-term ecological research site in Kansas, USA.

Mean rank shifts. The results of the mean rank shift calculation indicate that the degree of species reordering at the burned and unburned communities began to differ around 1992 (Fig. 2c). In particular, rank shifts decreased and stabilized in the annually burned site following a substantial spike in species reordering in 1992. This spike corresponds with the marked decline in *Andropogon gerardii* highlighted by the rank clocks.

Rate of community change. Rank shifts could reflect continual reshuffling of species within the same community type, or successional change from one community type to another. Calculating the rate of community change at Konza indicates that rank shifts in the annually burned site reflect directional change (i.e. the community was increasingly dissimilar over time), whereas rank shifts in the unburned site occur despite a lack of directional change (Fig. 2d). Taken together, the three `library("codyn")` diversity indices tell a story of two communities with high species turnover but different degrees of directional change – both elements which are masked by non-temporal indices such as richness.

Community stability metrics

Ecologists have long debated the relationship between species diversity and stability (MacArthur 1955; Tilman 1996; Cottingham, Brown & Lennon 2001). Unstable species populations may stabilize aggregate community properties if a decrease in one species is compensated for by an increase in another. That is, abundance of one species may decline while another increases, resulting in no net change in the aggregate variable net primary production. In a time series, this should be reflected by a pattern in which species negatively covary or fluctuate asynchronously while total community abundance remains relatively stable (Yachi & Loreau 1999). `library("codyn")` includes a function to characterize community stability (i.e. the stability of summed species abundances), `community_stability()`, and three metrics to characterize species covariance (`variance_ratio()`) (Schluter 1984; Hallett *et al.* 2014) and asynchrony (`synchrony()` with two options) (Loreau & de Mazancourt 2008; Gross *et al.* 2014). See another recent R package, `library("synchrony")`, for additional metrics such as Kendall's W (Gouhier & Guichard 2014).

COMMUNITY STABILITY

The `community_stability()` function aggregates species abundances within replicate and time period, and uses these values to calculate community stability as the temporal mean divided by the temporal standard deviation (Tilman 1999).

VARIANCE RATIO

The variance ratio was one of the first metrics to characterize patterns of species covariance (Schluter 1984) and was used in an early synthesis paper of species covariance in long time series (Houlahan *et al.* 2007). The metric compares the variance of the community (C) as a whole relative to the sum of the individual population (x_i) variances:

$$VR = \frac{\text{Var (C)}}{\sum_i^N \text{Var}(x_i)}$$

where:

$$\text{Var (C)} = \sum_{i=1}^{N} \text{Var}(x_i) + 2\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \text{Cov}(x_i, x_j)\right)$$

If species vary independently, then the variance ratio will be close to 1. A variance ratio <1 indicates predominately negative species covariance, whereas a variance ratio >1 indicates that species generally positively covary.

Significance testing

The variance ratio remains widely used but has been subject to a number of criticisms. Importantly, early uses of the variance ratio either did not include significance tests, or tested significance by comparing observed values to those returned by scrambling each species' time series by randomizing the order of values in the time series. Null models using such a fully scrambled species time series can generate spurious null expectations of covariance because the process disrupts within-species autocorrelation. Phase-scrambling (Grman *et al.* 2010) and cyclic shift permutations (adapted from Harms *et al.* 2001; Hallett *et al.* 2014) have been used to address this issue.

The `variance_ratio()` function uses a cyclic shift permutation to conduct null modelling for significance tests. In this method, a starting time point is randomly selected for each species' time series. This generates a null community matrix in which species abundances vary independently, but within-species autocorrelation is maintained (for each species, the time series is disrupted only once). `library("codyn")` also includes the option to apply a cyclic shift permutation for other functions using the `cyclic_shift()` function, which returns an S3 object that includes null test statistics from a user-specified function. The user-specified function should operate on a community matrix. The function `confint.cyclic_shift()` returns confidence intervals for the S3 object returned by `cyclic_shift()`.

SYNCHRONY OPTION 1: 'LOREAU'

Loreau & de Mazancourt (2008) developed a metric of species synchrony that compares the variance of aggregated species abundances with the summed variances of individual species:

$$\text{Synchrony} = \frac{\sigma(x_T)^2}{(\Sigma_i \sigma_{x_i})^2}$$

where:

$$x_T(t) = \sum_{i=1}^{N} x_i(t)$$

This measure of synchrony is standardized between 0 (perfect asynchrony) and 1 (perfect synchrony) and is the default calculation in the `synchrony()` function. A virtue of this metric is that it can be applied across communities of variable species richness. It can also be applied not only to species abundance but also population size and per capita growth rate. However, unlike the variance ratio, it does not lend itself to significance testing.

SYNCHRONY OPTION 2: 'GROSS'

Gross *et al.* (2014) developed a metric of synchrony that compares the average correlation of each individual species with the rest of the aggregated community:

$$\text{Synchrony} = (1/N) \sum_i Corr\left(x_i, \sum_{i \neq j} x_j\right)$$

This measure of synchrony is standardized from −1 (perfect asynchrony) to 1 (perfect synchrony) and is centred at 0 when species fluctuate independently. It can be specified in the `synchrony()` function. This metric is not sensitive to richness and has the potential for null-model significance testing. It may underperform on short time series because it is based on correlation, and care should be taken when applying it to communities that contain very stable species (i.e. whose abundances do not change throughout the time series).

WORKED EXAMPLE: COMMUNITY STABILITY AT KONZA

`library("codyn")` examples of community stability indices utilize the `knz_001d` data set:

```
library(codyn)
data("knz_001d")
head(knz_001d[1:3,])
##              species year subplot abundance
## 1 achillea millefolium 1986  A_1      0.5
## 2 ambrosia psilostachya 1988  A_1      0.5
## 3 ambrosia psilostachya 1990  A_1      3.0
```

Community stability (i.e. the stability of aggregate cover over time) varied considerably in the annually burned watershed (Fig. 3). There was a strong negative relationship between community stability and both the variance ratio and synchrony as calculated by `Loreau` (Fig. 3), indicating that greater negative covariance/higher asynchrony was associated with increased stability. This relationship was much weaker for synchrony as calculated by `Gross` (Fig. 3). This is because both the `Loreau` method and the variance ratio are based on variances, and are therefore more heavily influenced by abundant species (which also comprise most of the aggregate cover). In contrast, the `Gross` method is based on correlation and consequently weighs all species equally. Differences between the metrics should decrease in more even communities.

Acknowledgements

Data accessibility

`library("codyn")` is available on CRAN (https://cran.r-project.org/web/packages/codyn/index.html) and GitHub at (https://github.com/laurenmh/codyn).


```
variance_ratio(df = knz_001d,
  time.var = "year",
  species.var = "species",
  abundance.var = "abundance",
  replicate.var = "subplot",
  bootnumber = 1,
  average.replicates = F)
```

```
synchrony(df = knz_001d,
  time.var = "year",
  species.var = "species",
  abundance.var = "abundance",
  replicate.var = "subplot",
  metric = "Loreau")
```

```
synchrony(df = knz_001d,
  time.var = "year",
  species.var = "species",
  abundance.var = "abundance",
  replicate.var = "subplot",
  metric = "Gross")
```
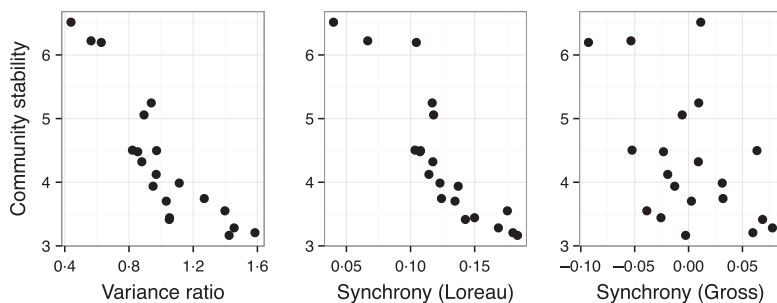
```
community_stability(df = knz_001d,
  time.var = "year",
  species.var = "species",
  abundance.var = "abundance",
  replicate.var = "subplot")
```

Fig. 3. Community stability in relation to the variance ratio and two measures of synchrony for subplots in the annually burned location at Konza Prairie, a long-term ecological research site in Kansas, USA.

Data used in the examples presented here are included in the package. Bug reports and feature requests can be sent using the GitHub issue tracker.

References

Cleland, E.E., Collins, S.L., Dickson, T.L., Farrer, E.C., Gross, K.L., Gherardi, L.A. *et al.* (2013) Sensitivity of grassland plant community composition to spatial vs. temporal variation in precipitation. *Ecology*, **94**, 1687–1696.

Collins, S.L., Micheli, F. & Hartt, L. (2000) A method to determine rates and patterns of variability in ecological communities. *Oikos*, **91**, 285–293.

Collins, S.L., Suding, K.N., Cleland, E.E., Batty, M., Pennings, S.C., Gross, K.L. *et al.* (2008) Rank clocks and plant community dynamics. *Ecology*, **89**, 3534–3541.

Cottingham, K.L., Brown, B.L. & Lennon, J.T. (2001) Biodiversity may regulate the temporal variability of ecological systems. *Ecology Letters*, **4**, 72–85.

Diamond, J.M. (1969) Avifaunal equilibria and species turnover rates on the Channel Islands of California. *Proceedings of the National Academy of Sciences of the USA*, **64**, 57–63.

Gouhier, T.C. & Guichard, F. (2014) Synchrony: quantifying variability in space and time. *Methods in Ecology and Evolution*, **5**, 524–533.

Grman, E., Lau, J.A., Schoolmaster, D.R. & Gross, K.L. (2010) Mechanisms contributing to stability in ecosystem function depend on the environmental context. *Ecology letters*, **13**, 1400–1410.

Gross, K., Cardinale, B.J., Fox, J.W., Gonzalez, A., Loreau, M., Wayne Polley, H., Reich, P.B. & van Ruijven, J. (2014) Species richness and the temporal stability of biomass production: a new analysis of recent biodiversity experiments. *The American Naturalist*, **183**, 1–12.

Hallett, L.M., Hsu, J.S., Cleland, E.E., Collins, S.L., Dickson, T.L., Farrer, E.C. *et al.* (2014) Biotic mechanisms of community stability shift along a precipitation gradient. *Ecology*, **95**, 1693–1700.

Hallett, L.M., Jones, S.K., MacDonald, A.M., Flynn, D.F.B., Slaughter, P., Ripplinger, J., Collins, S.L., Gries, C. & Jones, M.B. (2015). codyn: Community dynamics metrics. R package version 1.0.1. https://github.com/laurenmh/codyn/

Harms, K.E., Condit, R., Hubbell, S.P. & Foster, R.B. (2001) Habitat associations of trees and shrubs in a 50-ha Neotropical forest plot. *Journal of Ecology*, **89**, 947–959.

Hector, A., Hautier, Y., Saner, P., Wacker, L., Bagchi, R., Joshi, J. *et al.* & others. (2010) General stabilizing effects of plant diversity on grassland productivity through population asynchrony and overyielding. *Ecology*, **91**, 2213–2220.

Houlahan, J.E., Currie, D.J., Cottenie, K., Cumming, G.S., Ernest, S.K.M., Findlay, C.S. *et al.* & others. (2007) Compensatory dynamics are rare in natural ecological communities. *Proceedings of the National Academy of Sciences of the USA*, **104**, 3273–3277.

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

Laliberté, E., Legendre, P. & Shipley, B. (2014). FD: Measuring functional diversity from multiple traits, and other tools for functional ecology. R package version 1.0-12. http://CRAN.R-project.org/package=FD

Loreau, M. & de Mazancourt, C. (2008) Species synchrony and its drivers: neutral and nonneutral community dynamics in fluctuating environments. *The American Naturalist*, **172**, E48–E66.

MacArthur, R. (1955) Fluctuations of animal populations and a measure of community stability. *Ecology*, **36**, 533.

MacArthur, R.H. & Wilson, E.O. (1963) An equilibrium theory of insular zoogeography. *Evolution*, **17**, 373–387.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B. *et al.* (2015). Vegan: Community ecology package. R package version 2.3-1. http://CRAN.R-project.org/package=vegan

Schluter, D. (1984) A variance test for detecting species associations, with some example applications. *Ecology*, **65**, 998–1005.

Tilman, D. (1996) Biodiversity: population versus ecosystem stability. *Ecology*, **77**, 350–363.

Tilman, D. (1999) The ecological consequences of changes in biodiversity: a search for general principles. *Ecology*, **80**, 1455–1474.

Wickham, H. (2014) Tidy data. *Journal of Statistical Software*, **59**, 1–23.

Wickham, H. (2015): tidyr: Easily tidy data with 'spread()' and 'gather()' functions. R package version 0.3.1. http://CRAN.R-project.org/package=tidyr

Wickham, H. & Francois, R. (2015). dplyr: A grammar of data manipulation. R package version 0.4.3. http://CRAN.R-project.org/package=dplyr

Yachi, S. & Loreau, M. (1999) Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proceedings of the National Academy of Sciences of the USA*, **96**, 1463–1468.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Data S1. Supplemental ʀ code to run the analyses and create the figures in this paper.