

windex: Analyzing Convergent Evolution Using the Wheatsheaf Index in R

Kevin Arbuckle and Amanda Minter

Institute of Integrative Biology, University of Liverpool, Liverpool, Merseyside, UK.

ABSTRACT

SUMMARY: *windex* is a package developed for the R statistical environment to provide novel tools for the analysis of convergent evolution. The recently described Wheatsheaf index provides quantitative measures of the strength of convergence and opens up new possibilities for exploring this evolutionary phenomenon. The *windex* package allows implementation of this method with additional functions that can be used to create plots and perform statistical tests. R provides compatibility with other packages, and the R environment is familiar to many researchers.

AVAILABILITY: The *windex* package is freely available from CRAN: <http://cran.r-project.org/web/packages/windex/>. Consequently, *windex* can be installed directly from R and is distributed under the GNU General Public License 2.0.

KEYWORDS: convergence, phenotypic evolution, phylogenetic comparative methods, software

CITATION: Arbuckle and Minter. *windex*: Analyzing Convergent Evolution Using the Wheatsheaf Index in R. *Evolutionary Bioinformatics* 2015;11 11–14 doi: 10.4137/EBO.S20968.

RECEIVED: October 13, 2014. **RESUBMITTED:** December 03, 2014. **ACCEPTED FOR PUBLICATION:** December 08, 2014.

ACADEMIC EDITOR: Jake Cui, Associate Editor

TYPE: Short Report

FUNDING: KA was funded by a NERC Doctoral Training grant. The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

CORRESPONDENCE: k.arbuckle@liverpool.ac.uk

COPYRIGHT: © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY 4.0 International Public License.

Paper subject to independent expert blind peer review by minimum of two reviewers. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE).

Published by Libertas Academica. Learn more about this journal

The use of phylogenetic comparative methods in evolutionary biology has seen a remarkable increase in recent years.^{1–4} Much of this growth has resulted from the proliferation of newly developed methods (eg, see Refs. 5–7) and a shift toward implementation of these methods in R, which has enhanced the flexibility and between-method compatibility of their implementation.

Convergent evolution, or the independent evolution of similar phenotypes, is a commonly observed phenomenon across the tree of life.⁸ Nevertheless, methods designed to study convergence have traditionally been limited to identifying its presence (eg, see Refs. 7 and 9), ie, whether convergence has or has not occurred in a given case. Recently, Arbuckle et al.¹⁰ developed a new method that aims to provide a quantitative measure of the strength of convergent evolution – the Wheatsheaf index. By quantifying convergence, this method allows an expanded range of questions we can ask, such as “Do life history traits show greater convergence than morphological traits?” or “Do limbs or eyes show stronger convergence in burrowing animals?” This more detailed understanding of how convergent evolution operates as an evolutionary mechanism can only be achieved once a suitable measure is available that can be used to analyze a wide range of traits.

Briefly, the Wheatsheaf index generates phenotypic (Euclidean) distances from any number of traits across species and penalizes these by phylogenetic distance before investigating similarity (in order to weight close phenotypic

similarity higher for distantly related species). It also takes an a priori designation of focal species, which are defined as species belonging to a niche for which the traits are hypothesized to converge. The method then calculates a ratio of the mean (penalized) distances between all species to the mean (penalized) distances between focal species. In effect, the Wheatsheaf index detects stronger convergence as the focal species diverge more in phenotypic space from the non-focal species and also as the focal species show a tighter clustering to each other.

Upon describing the Wheatsheaf index, Arbuckle et al.¹⁰ made available a MatLab script to implement the method, although this was quite inflexible and many potential users are not familiar with MatLab. Therefore, in this paper, we introduce a user friendly R package (*windex*) with which researchers can use the Wheatsheaf index to analyze convergent evolution.

To illustrate the use of the R package *windex*, we analyze morphological convergence for burrowing in monitor lizards (*Varanus*) using the Wheatsheaf index applied to data taken from Thompson et al.¹¹ *windex* contains three functions: `plotTrait`, `windex`, and `test.windex` (Table 1). These functions require up to two inputs, which we will herein refer to as the traits and the tree for convenience. The tree is a phylogenetic tree of the class `phylo`, as is standard for most phylogenetic packages in R. `Traits` is a data frame consisting of a few necessary columns. The first column must

**Table 1.** Brief summary of functions in the `windex` package.

FUNCTION	INPUT	OUTPUT
<code>plotTrait</code>	Traits	A plot of phenotypic space for visualisation of raw (not phylogenetically corrected) data.
<code>windex</code>	Traits and tree	Wheatsheaf index along with 95% confidence intervals obtained by jackkniving the data.
<code>test.windex</code>	Traits and tree	<i>P</i> -value for a test of particularly strong convergence, including a graphical display of the result.

be named species and contains species names, which match the tip labels in the tree. One column must designate the focal taxa (see Ref. 10 for details of the method itself and further understanding of these terms) as 1 and non-focal taxa as 0. The focal taxa are those species for which you are interested in testing convergence (eg, burrowing species in our *Varanus* example). Other columns contain the trait values, typically (semi-)continuous traits, but if there are a large number of binary traits in the dataset, then they can also be used, as they would similarly allow the calculation of meaningful phenotypic distances. By (semi-)continuous, we mean ordinal or count data in addition to truly continuous measurements, as all of these types would generate a meaningful phenotypic distance. The data frame is then read into the R environment. The data frame for our example herein consists of three traits of interest: head depth (`headD`), upper fore limb length (`UforelimbL`), and upper hind limb length (`UhindlimbL`). The first six rows are shown below to illustrate the format of the data frame (titled `dat`), called as follows:

```
> head(dat)
  species      focal  headD  Ufore-  Uhind-
                headD  limbL   limbL
1 acanthurus      1    10.9    14.2    20.5
2 breviceuda      1     6.4     5.9     7.4
3 caudolineatus  0     6.8     7.8    11.3
4 eremius         1     9.6    11.3    16.8
5 giganteus       1    25.7    43.6    55.9
6 gilleni         0     7.3     8.7    11.7
```

The `plotTrait` function only requires the traits (not the tree), and is intended as a tool for data exploration. It produces a plot that represents a phenotypic space with one to three dimensions (traits) with focal taxa highlighted to visualize where they appear relative to non-focals, although this plot does not take into account phylogenetic relationships. Nevertheless, it may often be a useful preliminary step for understanding how the data are structured.

For our monitor lizard example, we plot one-, two- and three-dimensional plots with the traits head depth (`headD`), upper fore limb length (`UforelimbL`), and

upper hind limb length (`UhindlimbL`) (Fig. 1) using the following code:

```
> par(mfrow=c(1,3))
> plotTrait(dat,traits="headD")
> plotTrait(dat,traits=c("headD","UforelimbL"))
> plotTrait(dat,traits=c("headD","UforelimbL","UhindlimbL"))
```

The core function of the package is `windex`, which takes both the tree and traits as input and calculates the Wheatsheaf index. This function also performs jackknife resampling of the traits as per Ref. 10 and uses these samples to return 95% confidence intervals alongside the calculated index. The method requires that measurements for each trait are standardized by the standard error for the trait across species. Although this can be done as pre-treatment of the data file, the `windex` function includes an option that allows this step to be included as part of implementation, removing the need for any such pre-treatment of the data and, hence, increasing the method's ease of use.

Here we use the function `windex` to calculate the Wheatsheaf index for a combination of three traits that are likely to be important in burrowing in our *Varanus* example: head depth (`headD`), upper fore limb length (`UforelimbL`), and upper hind limb length (`UhindlimbL`):

```
> windex(dat,tree,traits=c("headD","UforelimbL","UhindlimbL"), SE=T)
```

```
$`Wheatsheaf Index`
```

```
[1] 1.322459
```

```
$`Lower 95% CI`
```

```
[1] 1.243233
```

```
$`Upper 95% CI`
```

```
[1] 1.389691
```

The final function in the package is `test.windex`, which implements the statistical test for exceptionally strong convergence, given the topological constraints of the tree (see Ref. 10 for more details). The function takes the same arguments as the `windex` function (as this is called internally by `test.windex`) plus two additional arguments. The first (`reps`) specifies the number of bootstrap replicates from which the *P*-value is derived. The number of replicates is, of course, case dependent, but we have chosen 2000 in the example below as a compromise between computation time and accuracy. The user may wish to increase or decrease the number of replicates, though too few replicates may lead to unreliable results. The second additional argument (`plot`) is an option to plot a visualization of the result in addition

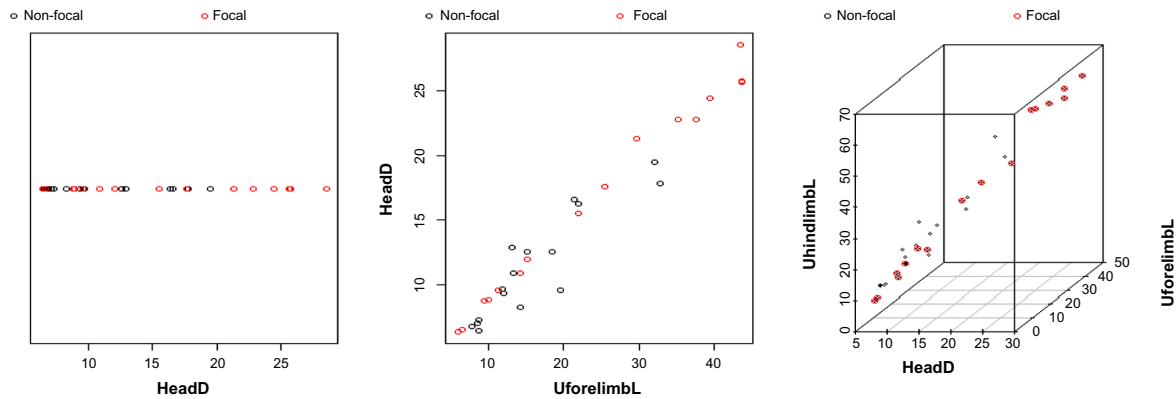


Figure 1. An illustrative example of the `plotTrait` function for one-dimensional, two-dimensional and three-dimensional plots using the traits head depth (`headD`), upper fore limb length (`UforelimbL`), and upper hind limb length (`UhindlimbL`).

to returning the P -value for the test, and the default for this argument is set as `plot = TRUE`. The plot consists of a histogram of the distribution of the Wheatsheaf index from bootstrap replicates, with the calculated value and its 95% confidence interval marked on the plot. Additional arguments are passed to the basic `hist` function in R to allow the histogram to be customized. Since this function can take several minutes or longer on large datasets, we have incorporated a simple status bar to allow the user to monitor the progress of the function. To return to our monitor lizard example, we now illustrate the `test.windex` function on the same set of traits as were used above for the `windex` function:

```
> test.windex(dat,tree,traits=c("headD",
  "UforelimbL","UhindlimbL"),SE=TRUE,
  reps=2000, col="light grey")
```

```
$`P-value`=
[1] 0.097
```

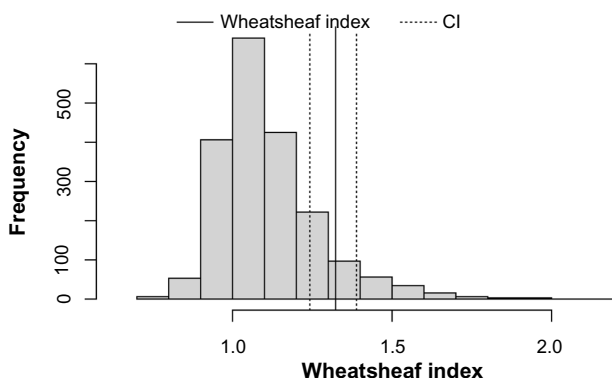


Figure 2. Histogram of the distribution of bootstrapped Wheatshaeaf index values from our example of morphological convergence for burrowing in monitor lizards. The calculated Wheatshaeaf index observed in the dataset is shown along with its jackknived 95% confidence interval.

The P -value obtained here is 0.097, marginally non-significant and therefore indicating that although convergence is present in *Varanus*,¹¹ it is not exceptionally strong in the selected traits for burrowing. The plot generated by the code above can be seen in Figure 2.

We hope that `windex` will greatly increase the ease of using the Wheatshaeaf index to analyze convergent evolution. The method is not intended to overshadow currently existing analyses for studying convergence but rather to complement them. Indeed, methods designed to test for the presence of convergence are strongly advised before using the Wheatshaeaf index because it makes little sense to quantify the strength of something that does not exist in a given dataset. As such, we have developed `windex` as a component of the analytical toolbox available to investigators of convergent evolution that provides an easy to use and useful extension to the suite of methods available in R (eg, Refs. 7, 12, and 13).

Acknowledgment

The `windex` package is an R implementation (and extension) of MatLab code written by M.P. Speed.

Author Contributions

Conceived the project: KA. Wrote the software package: KA, AM. Analysed example data: KA. Wrote the first draft of the manuscript: KA. Contributed to the writing of the manuscript: KA, AM. Agree with manuscript results and conclusions: KA, AM. Jointly developed the structure and arguments for the paper: KA, AM. Made critical revisions and approved final manuscript: KA, AM. Both authors reviewed and approved of the final manuscript.

REFERENCES

1. Freckleton RP. The seven deadly sins of comparative analysis. *J Evol Biol.* 2009;22:1367–75.
2. Harvey PH, Rambaut A. Comparative analyses for adaptive radiations. *Philos Trans R Soc Lond B Biol Sci.* 2000;355:1599–605.
3. Morlon H. Phylogenetic approaches for studying diversification. *Ecol Lett.* 2014;17:508–25.



4. Münkemüller T, Lavergne S, Bzeznik B, et al. How to measure and test phylogenetic signal. *Methods Ecol Evol.* 2012;3:743–56.
5. Alfaro ME, Santini F, Brock C, et al. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc. Nat. Acad. Sci. U.S.A.* 2009;106:13410–4.
6. FitzJohn RG. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods Ecol Evol.* 2012;3:1084–92.
7. Ingram T, Mahler DL. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike information criterion. *Methods Ecol Evol.* 2013;4:416–25.
8. McGhee GR. *Convergent Evolution: Limited Forms Most Beautiful.* Cambridge, MA: MIT Press; 2011.
9. Muschick M, Indermaur A, Salzburger W. Convergent evolution within an adaptive radiation of cichlid fishes. *Curr Biol.* 2012;22:2362–8.
10. Arbuckle K, Bennett CM, Speed MP. A simple measure of the strength of convergent evolution. *Methods Ecol Evol.* 2014;5:685–93.
11. Thompson GG, Clemente CJ, Withers PC, Fry BG, Norman JA. Is body shape of varanid lizards linked with retreat choice? *Aust J Zool.* 2009;56:351–62.
12. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics.* 2004;20:289–90.
13. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol.* 2012;3:217–23.