A simple measure of the strength of convergent evolution

Kevin Arbuckle*, Cheryl M. Bennett and Michael P. Speed

Institute of Integrative Biology, Biosciences Building, University of Liverpool, Crown Street, Liverpool, Merseyside L69 7ZB, UK

Summary

1. Convergent evolution, the independent occurrence of phenotypic similarity, is a widespread and common phenomenon. Methods have been developed to identify instances of convergence, but there is a lack of techniques for quantifying the strength of convergence. We therefore investigated whether convergent evolution can be quantified in a meaningful way.

2. We have developed a simple metric (the Wheatsheaf index) that provides an index of the strength of convergent evolution incorporating both phenotypic similarity and phylogenetic relatedness. The index is comparable across any quantitative or semiquantitative traits and thus will enable the testing of various hypotheses relating to convergence.

3. The index performs well over a range of conditions. We apply it to an empirical example using Anolis lizard ecomorphs to demonstrate how it can be used.

4. The Wheatsheaf index provides an additional tool that complements methods aimed at identifying cases of convergent evolution. It will enable cases of convergence to be analysed in more detail, test hypotheses about its mechanics as an evolutionary process and, more generally, the predictability of evolution (how often do we see strong convergence and does this mean evolutionary solutions are limited?).

Key-words: Convergence, phylogenetic comparative methods, phenotypic evolution, evolutionary processes, functional ecomorphology, adaptation

Introduction

The independent evolution of similar phenotypic traits in multiple organisms, or convergence, has been recognized as a key evolutionary process since Darwin (1859). Convergent evolution is often a consequence of adaptation to a similar niche (although not always, see Stayton 2008) and has therefore been recognized and studied in cases of replicated adaptive radiations such as Anolis ecomorphs (Losos 1992, 2009; Beuttell & Losos 1999) and African cichlids (Kocher et al. 1993; Muschick, Indermaur & Salzburger 2012). In addition, convergence may be seen when organs have similar uses and converge on a similar form, as in the camera eye which has evolved in both vertebrates and invertebrates. Convergence between organisms for a particular niche can promote speciation by causing divergent selection within a lineage inhabiting two niches (Rosenblum 2006), limit the suite of phenotypic adaptive optima (Mahler 2013) and drive distantly related organisms towards the same niche (although not always, see Stayton 2008) and limit biological diversity in this trait. If correct, Conway Morris (2003) has argued that convergence of traits towards a limited number of ‘engineering optima’ is a central guiding force in phenotypic evolution. For example, there are only a small number of ways to construct an effective, functioning eye; hence, engineering constraints cause convergence and limit biological diversity in this trait. If correct, Conway Morris’s view is profoundly important for our understanding of biological variation. Therefore, an understanding of convergent evolution is important to understanding the generation of biodiversity, constraints on adaptation, and how natural selection optimizes an organism for a particular niche. For the purposes of this study, we use ‘niche’ to refer to an aspect (or aspects) of the biotic and/or abiotic environment of an organism that is of interest for a hypothesis under study.

There have been several approaches and methods developed to identify instances of convergent evolution, and these have enabled a large number of cases to be described and recognized as such. At its simplest, convergence may be identified by carefully cataloguing traits across many species. McGhee’s recent text (2011) is an excellent example of this.

More formally, perhaps the most commonly used and simplest method for identifying convergence is ancestral state reconstruction of the (purportedly) convergent trait. For example, this method has provided support for convergent evolution of plumage coloration in Icterus orioles (Omland & Lanyon 2000) and in the chemically defended Pitohui birds (Dumbacher & Fleischer 2001). In such an analysis, the phenotype is reconstructed in some way over the phylogeny, and independent origins (multiple shifts to the same state) are taken as evidence of convergence.

Muschick, Indermaur and Salzburger (2012) used an alternative approach to test for convergence in cichlid fishes by considering that convergence should result in a pattern of reduced phenotypic differentiation when compared with phylogenetic distance. These authors thus calculated Euclidean distances

*Correspondence author. E-mail: k.arbuckle@liverpool.ac.uk

© 2014 The Authors. Methods in Ecology and Evolution published by John Wiley & Sons Ltd on behalf of British Ecological Society. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
between species in the morphological traits of interest and plotted them against the phylogenetic distances. They then used simulations to identify instances where phenotypic divergence was significantly lower than expected based on phylogenetic distance. As this method involves a straightforward comparison of phylogenetic and phenotypic distances, Muschick, Indermaur and Salzburger (2012) included both convergence and slower-than-expected divergence within their measure, as the two would produce the same signature.

A third approach was described by Ingram and Mahler (2013) which explicitly models trait evolution onto a phylogeny to identify convergent evolution. Their ‘SURFACE’ method takes a continuous trait and fits Ornstein–Uhlenbeck models with varying numbers of selective regimes and with shifts at varying points on the tree. Akaike’s information criterion is then used to select the best fitting model. Convergence is identified by the independent adoption of the same selective regime at multiple points on the phylogeny.

Each of these methods represents a technique to identify when convergence has occurred. Statistical recognition of convergence is, of course, fundamental. However, once convergence is established, a number of important questions can be explored. For example, we may be interested in whether there are general rules in the way convergence operates. Do some traits show stronger convergence than others? Do different types of traits converge more easily than others (e.g. morphological vs. biochemical traits), and if so, is evolution more predictable for some kinds of traits than for others? Do particular ‘levels’ of convergence (e.g. functional, structural, developmental, genetic) vary in their contribution to adaptive evolution? Why might such differences exist (e.g. what might drive stronger convergence in protein sequences than limb anatomy)? It is perhaps notable that most analyses of convergent evolution have focussed on morphological traits, which limits our knowledge base on how different types of traits may differ in aspects of convergence; however, some exceptions do exist (e.g. Miretta et al. 2013).

To answer such questions, we need a way of quantifying the strength of convergence. When we have a suitable measure of convergent evolution, we can start to test hypotheses about the nature of convergence, rather than simply recognizing it. Specifically, we require a metric that is comparable across many types of traits, incorporates both phylogenetic relatedness and the extent of phenotypic similarity, and is quantitative.

In this study, we describe a simple measure of the strength of convergent evolution, which we call the ‘Wheatsheaf index’. For the purposes of our method and this study, we consider convergence to be the pattern that results from the process of convergent evolution, rather than the process itself. Furthermore, because we use a pattern-based description of convergence, parallelism is indistinguishable from ‘true’ convergence using our method and so comes under the concept of convergence for the purposes of this study. The index was designed to meet the requirements outlined above and with the underlying assumption that we can define a set of species as convergent or have a working hypothesis as to the niche upon which the organisms are adapted (or adapting towards).

The Wheatsheaf index

To calculate the Wheatsheaf index we take a set of organisms, and within that identify a subset that we treat as convergent (we call this the subset of ‘focal’ taxa), and the residual species as members of the ‘non-focal’ subset. The index measures the similarity of focal species to each other and the isolation in phenotypic space of the focal group from non-focal species, all penalized for phylogenetic relatedness. To understand this in a conceptual way, we can consider convergence to be movement in phenotypic space over a fitness landscape towards an elevated position (such as an adaptive peak) which characterizes a particular environment or niche. The distance between non-focal and focal species represents the distance across such a landscape that focals have had to move to reach the peak, with movement over larger distances representing more evolution and therefore a stronger signature of convergence. In addition, the more tightly clustered the focal species are in this phenotypic space (the more similar they are to each other), the stronger are the selective forces pulling converging species towards the peak, or the narrower the peak itself, which in either case would indicate a more intense pull towards a particular point in phenotypic space.

Both of these aspects seem to be good foundations for a conceptual view of the strength of convergence, providing phylogenetic relatedness is accounted for, as is the case with the Wheatsheaf index. Thus, we consider convergence to be stronger when focal species are more phenotypically similar to each other, and when the focal species are more dissimilar to the non-focal species – in other words, when they have had to evolve further from the baseline of non-focal species to reach the convergent state. We note that some patterns of convergence may leave convergent species still more similar to their close relatives than each other in many phenotypic attributes (Stayton 2006), but we view this as a manifestation of differing strengths of convergence rather than a challenge to our definition. This phenotypic aspect of the index is penalized for close phylogenetic affinities and generates a quantitative measure which can subsequently be used to test hypotheses about the strength of convergence across traits.

Before we can apply the Wheatsheaf index, we require a clade to work with in which some members have been demonstrated to exhibit convergent evolution. In other words, we would use other methods (e.g. ancestral state reconstruction or SURFACE) which identify convergence so that we can start with a supported assumption that there is convergence in our group of interest. We then need to assign (a priori) species within that group as either ‘focal’ or ‘non-focal’ species. This is often related to a working hypothesis on the niche the organisms are expected to be converging on such that focals are those species occupying that niche (expected to show convergent adaptations) and non-focals are those species not occupying that niche. To give two examples, we might be interested in measuring convergence in body form for burrowing in lizards; in this case, burrowing species would be assigned to the focal group. Or we might look at convergence in salinity tolerance for brackish habitats, in which case species inhabiting estuaries
and other such environments would form the focal group. Alternatively, we could consider the species already identified as convergent as the focal group, which would allow us to measure how strong the convergence is in selected phenotypes of these taxa, regardless of any adaptive reason for it.

Other information required for the Wheatsheaf index is a phylogeny for the clade of interest and trait information. How we choose traits will depend on the purposes of the study. If we are interested in whether a particular set of traits are important for a given niche, then the selection of traits should be hypothesis-driven such that traits are chosen so that they may be convergent for that niche. This approach has the benefit that specific adaptive hypotheses of convergence for a given niche are examined. If, on the other hand, we are interested in an unguided investigation of which traits might be convergent for a given niche (if we have no working hypothesis with which to make a priori predictions), then we could use a large number of traits spanning the range of those we can measure, run the index on all of them and therefore obtain estimates of which ones are most convergent. However, an important stipulation is that the traits must be (semi)quantitative (e.g. continuous, count or ordinal data; see Discussion for further details).

Calculation of the Wheatsheaf index requires the data (both phylogenetic and phenotypic) to be represented in pairwise distance matrices. For the phylogeny, a matrix of proportion shared distances between species is used, such that the total tree height is scaled to one and distances are given as the proportion of the tree shared between two species. In other words, bigger distances represent more closely related species. For phenotypic traits (which are first standardized for variance by dividing by the standard error of the trait across species), a matrix of Euclidean distances between species is used, which enables any number of traits to be incorporated, and bigger shared distances represent more dissimilar species for the included traits. This allows us to look at single traits individually or grouped traits as appropriate for the hypothesis being tested, for example, we could obtain a distance matrix for a set of morphological traits and a second one for a set of physiological traits. Again, the selection of traits to include in the study as a whole and in a given distance matrix will be driven by the hypothesis in question.

To calculate the Wheatsheaf index, we first obtain a corrected (for phylogenetic relatedness) phenotypic distance matrix as follows:

$$d_{ij} = \frac{d_{ij}}{1 - \log(p_{ij} + 0.01)}$$

eqn 1

where $d_{ij}$ is the phenotypic (Euclidean) distance between species $i$ and $j$, $p_{ij}$ is the shared proportional distance between species $i$ and $j$ obtained from the phylogeny, and $d_{ij}$ is therefore the phenotypic distance between species $i$ and $j$ corrected for phylogeny. Note that $p_{ij}$ is transformed by adding a small (and arbitrary) value and logging; this is so that $p_{ij}$ scales approximately linearly with $d_{ij}$. If a pair of species are closely related, and therefore, $p_{ij}$ is close to 1, then $d_{ij}$ will be much larger than $d_{ij}$. As species become more distantly related, then $p_{ij}$ will decrease and $d_{ij}$ will become progressively smaller and approach $d_{ij}$. This is an intuitive way of correcting for phylogeny as more weight (i.e. a smaller distance) is assigned to more distantly related taxa being similar, therefore penalizing the phenotypic similarity of closely related species. As $p_{ij}$ and $d_{ij}$ are approximately linearly related in the equation, this is in effect assuming that the phenotype diverges in proportion to time (phylogenetic history). Note that as we consider convergence to be a pattern in this paper, no model is fitted and so no parameterization is conducted, and thus, eqn 1 should be robust to the particular evolutionary model that best fits the trait data, providing that we can expect more phenotypic divergence when species pairs are more distantly related. Nevertheless, it might be possible to extend this method in the future to incorporate specific evolutionary models in the penalizing term, should this become necessary.

Using the corrected phenotypic distances (pairwise matrix of $d_{ij}$ between each pair of species), we can now calculate the Wheatsheaf index ($w$) as follows:

$$w = \frac{\overline{d}_j}{\overline{d}_f}$$

eqn 2

where $\overline{d}_j$ is the mean $d_{ij}$ for pairwise comparisons between all species, and $\overline{d}_f$ is the mean $d_{ij}$ for pairwise comparisons between focal species only. As $\overline{d}_j$ increases and $\overline{d}_f$ decreases, then $w$ will increase, showing stronger convergence and vice versa. A visual representation of this is provided in Fig. 1. Because a greater separation in phenotypic space between the focal and non-focal groups will result in larger distances between focal taxa and non-focal taxa, $\overline{d}_j$ will be larger and so $\overline{d}_f$ will be relatively smaller, therefore showing stronger convergence (a larger $w$). Similarly, a tighter clustering of focal species will decrease $\overline{d}_f$, relatively increasing $\overline{d}_j$ and so again showing a signature of stronger convergence. We note that our method shows some similarity to that of Stayton (2006) in that both use ratios of phenotypic and phylogenetic measures to generate a corrected phenotypic distance and compare convergence species to the set as a whole. However, the Wheatsheaf index differs in a number of ways including calculating pairwise phylogenetic and phenotypic distances between all species in the phylogeny, rather than using information only from sister groups (or similar comparisons).

As the calculation of $w$ is not amenable to multiple, independent sampling (it uses information from the entire sample – all species in the clade), 95% confidence intervals are generated by jackknifing the data set and using the resulting distribution of values to calculate the intervals.

Because the topology of the tree may constrain the possible values of $w$, we used a bootstrapping approach to resample the tips of the tree along with their trait values and thus obtain a distribution of possible $w$ indices given the phylogeny and the trait values for each species. Using this distribution and the calculated value of $w$, we can generate a ‘$P$-value’ by taking the proportion of bootstrap samples that are greater than or equal to the value of $w$ calculated from the original data set (see Fig. 2). We stress that this $P$-value is not a test for the presence
of convergent evolution; as described earlier, we begin an analysis with the Wheatsheaf index with the knowledge that convergence has occurred in our clade of interest. Rather, it represents a test of whether convergence is significantly stronger than we would expect compared to a random distribution of trait values across the specified tree. A further advantage of this is that comparisons of the P-values provide a measure of convergence that accounts for the given tree structure and so, in effect, standardizes for this. In other words, we can potentially use the P-value to compare the strength of convergence across trees, which is not possible using our value of w alone. However, we would add that as P-values are bound between zero and one, comparisons using this part of the method may be limited in extreme cases by floor and ceiling effects.

Materials and methods

We evaluated our index in two separate ways: simulations and empirical data. Using data simulated with specific parameters means we can investigate how particular attributes of a given data set influence the calculation of the Wheatsheaf index and therefore whether there are any particular conditions that warrant caution. We should also ensure our method is appropriate for use on empirical data, and so we present an example to show how it can be used on an ecomorphological data set of Anolis lizards.

SIMULATIONS

To assess the general performance of the Wheatsheaf index under various conditions, we simulated a range of phylogenies and continuous traits in R version 2.15.2 (R Development Core Team 2012). All data manipulation, such as generating the Euclidean distance matrices, prior to calculation of the index was also conducted in R. The matrices of shared proportional distances from the phylogenies were extracted using the packages ape (Paradis, Claude & Strimmer 2004) and geiger (Harmon et al. 2008).

Ten trees were simulated using a birth-death model in geiger with a birth rate of 0.5 and a death rate of 0.1 resulting in 100 species each (except when number of species was the parameter being varied, in which case 10 trees with each number of species were simulated). Trait data were simulated over each tree in two ways. First, to assess type 1 error, trait data were simulated under a Brownian motion (BM) model across the tree, such that convergence would be very unlikely to occur amongst focal species (Stayton 2008). Secondly, to assess type 2 error, trait data were simulated under a Brownian motion (BM) model for non-focal species but under an Ornstein–Uhlenbeck (OU) model for focal species. In each of these simulations, focal and non-focals were present in equal numbers, except where the proportion of focal species was the parameter being varied. Trait simulation was conducted in diversitree (FitzJohn 2012), with parameters as follows (except when a particular parameter was the one being varied, as detailed below): $\sigma^2 = 10$ for BM models, and $\alpha = 5$, $\theta = 20$, $\sigma^2 = 10$ for OU models. All analyses were conducted on Euclidean distances over one, two and three traits for each tree to check sensitivity to number of traits involved in the calculations.

We varied three parameters (in turn) to assess what influence they had on the performance of the Wheatsheaf index: the number of species in the tree; the proportion of focal species in the tree; and the ‘strength of selection’ (variation around the optima, or $\alpha$ in the OU model). We recognize that ‘strength of selection’ is perhaps an overly simplistic interpretation of $\alpha$ in an OU model (Hanssen 2012; Ingram & Mahler 2013), but we use it here for ease of intuitive discussion (as in Hansen & Orzack 2005; Beaulieu et al. 2012) while acknowledging that factors other than the strength of selection can influence $\alpha$. The number of species in the phylogeny, reflecting sample size, was varied with the following values: 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 1000. The proportion of focal (cf. non-focal) species was varied with the following values: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9. The ‘strength of selection’ was varied by changing $\alpha$ in the OU model to the following values: 0.1, 0.5, 1, 2, 3, 4, 5, 10, 20, 50.
Methods in Ecology and Evolution

than b) both in absolute terms (calculated value is higher) and with respect to the topological constraints of the tree (further to the right of the arrow) are used to generate the distributions greater than or equal to the calculated value (to the right of the arrow). Vertical arrows represent calculated values, and the proportion of values obtained from bootstrap resampling of the tips of the tree. We can see that the example in a) is more strongly convergent than b) both in absolute terms (calculated value is higher) and with respect to the topological constraints of the tree (further to the right of the distribution).

We used the P-values to assess how the Wheatsheaf index performs across these parameter values. Specifically, we expected $P > 0.05$ when all traits were simulated under BM and $P \leq 0.05$ when focal species were simulated under an OU process. We were also able to determine the power of our method as 1-type 2 error rate.

**Empirical Example**

To examine how the index performs on a real data set, we performed analyses using an empirical example consisting of ecomorphological traits in anole lizards, a model system for studies of convergent evolution (Harmon et al. 2005) for which morphological data, phylogenetic information and a good literature base to assess our results are available. Caribbean *Anolis* lizards have repeatedly and independently evolved six ‘syndromes’ consisting of linked morphological, behavioural and ecological traits; these forms are termed ‘ecomorphs’ (Williams 1972; Losos 2009). The six *Anolis* ecomorphs are named after the microhabitat they inhabit as follows: crown-giant, trunk-crown, twig, trunk, trunk-ground and grass-bush (Losos 2009). We therefore decided to apply the Wheatsheaf index to investigate the strength of morphological convergence in ecomorphs as an empirical demonstration of the utility of the method.

Morphological data were extracted from the literature (Losos 1990a, 1992, 2009; Thomas, Meiri & Phillimore 2009). Data were obtained for six traits (snout-vent length, tail length, body mass, forelimb length, hindlimb length and number of toe lamellae) in 28 species, and a phylogeny for *Anolis* was taken from Thomas, Meiri and Phillimore (2009). Species were coded for ecomorph, but the trunk ecomorph and one unique species (i.e. not falling within any of the ecomorph classes) were represented by one species each, precluding analysis of convergence in these two ecomorphs. The tree was pruned in Mesquite v2.75 (Maddison & Maddison 2011) in order that the final tree contained only the 28 species for which we had data. These species were *A. allisoni*, *A. bahorocrensis*, *A. baleatus*, *A. chlorocyamus*, *A. christopheli*, *A. coelestinus*, *A. cybotes*, *A. distichus*, *A. insulatus*, *A. longitibiadis*, *A. olssonii*, *A. sembliebus*, *A. singularis*, *A. cristatellus*, *A. cuvieri*, *A. evermanni*, *A. gandachi*, *A. kraji*, *A. occultus*, *A. ponceensis*, *A. pulchellus*, *A. stratus*, *A. garmani*, *A. grahami*, *A. lineatus*, *A. palpus*, *A. sagrei* and *A. valencienni*.

A datafile for analysis was created for each ecomorph, such that each file had one ecomorph coded as the focal group. The index was first calculated for each datafile using 6-dimensional phenotypic distances consisting of an aggregate of all our traits (‘total morphology’). Next, the traits were analysed as functionally related aggregates to provide a more detailed, and biologically meaningful, look at morphological convergence amongst ecomorphs. These aggregate traits were as follows: body size (snout-vent length and body mass combined), limb length [forelimb and hindlimb length combined as together they are indicative of locomotor adaptations (Losos 1990b)], tail length (on its own due to a potentially separate role from body size in balancing ability or other adaptations to arboreal habits) and number of lamellae (on its own due to its functionally independent potential role in climbing ability). P-values were generated from 100 000 bootstrap replications.

**Results**

**Simulations**

Using general linear mixed models (accounting for the particular tree and parameter values as random effects), we found no effect of the number of phenotypic traits used to generate the Euclidean distance matrix on the value of $w$ for any of our six data sets (one to assess type one and type two error each for number of species, ‘strength of selection’, and proportion of focal species; all $P > 0.05$), bearing in mind that each of the three traits were simulated using the same parameter values. This suggests that the incorporation of distances between species in ‘combined’ traits does not, in itself, influence the method and that it appears to perform adequately across this variation. As such, all of the following results are given on analyses conducted on one-dimensional Euclidean distances only.

Figure 3 shows the results of our simulations. When all traits were simulated under BM, most of the estimated P-values were $>0.05$, giving an overall type 1 error rate of 0.053 (across all simulations), and no obvious relationship with any of our parameters is evident. When traits were simulated under
OU for focal species, almost all of the estimated $P$-values were <0.05, giving an overall type 2 error rate of 0.003 (across all simulations). Although the index performed well across all parameter estimates, it did so slightly worse when the ‘strength of selection’ in the OU model was very low and when the total number of species in the tree was low (although even in our 10 species trees all simulations gave $P < 0.05$) (Fig. 3). The Wheatsheaf index, when used with the $P$-value as a test, has good statistical power (0.997) to detect the presence of particularly strong convergence.

**EMPIRICAL EXAMPLE**

Table 1 presents the calculated values of $w$ for all analyses on the *Anolis* data sets, along with their 95% confidence intervals and $P$-values. The convergence within most ecomorphs (although present based on previous work) was not significantly stronger than expected given the tree. However, grass-bush anoles consistently showed very strong convergent evolution in all traits tested, as did trunk-ground anoles in overall (total) morphology and number of lamellae (Table 1).
Furthermore, despite not being significantly stronger than expected, based on the $P$-values, a number of other instances of relatively strong convergence ($P < 0.1$) were also observed (Table 1).

**Discussion**

An important question in evolutionary biology is whether convergence can be quantified. To begin to examine this question, we have described a new method (the Wheatsheaf index) for measuring the strength of convergent evolution. The index provides a simple quantification of convergence and achieves a number of desirable qualities: comparability, intuitive interpretation and phylogenetically informed.

The basis of the index is the relative phenotypic distances rather than absolute distances (and particularly as the traits are standardized to account for the degree of variation) and consequently is comparable between a wide variety of traits. It therefore provides a useful measure which can be compared directly between, for example, behavioural, morphology and molecular traits, or between functional and developmental traits, for species within the same overall set. This provides a high level of flexibility in how the method can be used and opens up a range of questions which can now be explicitly tested. Because $w$ increases as convergence becomes stronger, it has an intuitive interpretation.

Although the interpretation of a particular value is made more difficult by the possible influence of topological constraints, the $P$-value incorporates this aspect and can also be used to compare across trees – further assisting with interpretation. The index provides a measure that incorporates both the similarity of focal species to each other and the differentiation from non-focal species, which we regard as two key aspects of convergence. However, we must note that a high (or low) Wheatsheaf index can result from either of these aspects, for example, from close similarity in phenotypic values or from less phenotypically similar species that are more phylogenetically distant. Therefore, if we are interested in how a given value arose, we must look back at the tree to further inform our interpretations of the underlying patterns. In most or all cases, it is probable that both of these elements will be responsible in part.

**Limitations to the application of the index**

As mentioned earlier, the Wheatsheaf index requires (semi) continuous rather than discrete traits, unless there are multiple discrete traits to be included in the same analysis. This restriction is imposed on logical grounds. If a trait is either present or absent, then organisms cannot be more or less convergent for that trait: they either are convergent (share the trait) or not. Therefore, in the case of single discrete traits, it is meaningless to give a measure of the strength of convergence and the best we can do is to identify whether or not convergence has occurred and look for correlates with any hypothesized focal niche. If, however, there are multiple discrete traits, then we

---

**Table 1.** Wheatsheaf indices ($w$) with associated 95% confidence intervals (given as lower and upper bounds as they are not necessarily symmetric) for each group of traits in each Anolis ecomorph. $P$-values from analyses are also provided, and significant ($P < 0.05$) values are highlighted in bold. Of the 28 species in total, the number of focal species for each ecomorph used as a focal in the analyses was as follows: twig (3), crown-giant (3), grass-bush (6), trunk-crown (8), trunk-ground (6)

| Phenotype          | Ecomorph         | $w$   | Lower bound | Upper bound | $P$-value |
|--------------------|------------------|-------|-------------|-------------|-----------|
| Total morphology   | Twig             | 2.9213 | 2.8609      | 2.983       | 0.494     |
|                    | Crown-giant      | 1.2232 | 1.2127      | 1.2343      | 0.678     |
|                    | Grass-bush       | 4.4692 | 4.4156      | 4.5188      | $<0.001$  |
|                    | Trunk-crown      | 1.8444 | 1.8257      | 1.8654      | 0.129     |
|                    | Trunk-ground     | 2.5807 | 2.5533      | 2.6038      | 0.023     |
| Body size          | Twig             | 2.3238 | 2.2748      | 2.3829      | 0.686     |
|                    | Crown-giant      | 0.72149| 0.71219     | 0.73258     | 0.745     |
|                    | Grass-bush       | 9.0667 | 8.9076      | 9.2539      | $<0.001$  |
|                    | Trunk-crown      | 1.9864 | 1.9663      | 2.0138      | 0.253     |
|                    | Trunk-ground     | 1.3605 | 1.3357      | 1.3665      | 0.059     |
| Tail length        | Twig             | 3.0959 | 3.0426      | 3.1694      | 0.458     |
|                    | Crown-giant      | 2.0299 | 1.9878      | 2.092       | 0.384     |
|                    | Grass-bush       | 4.0083 | 3.9642      | 4.0651      | 0.01      |
|                    | Trunk-crown      | 1.7331 | 1.7158      | 1.7539      | 0.177     |
|                    | Trunk-ground     | 1.9416 | 1.9168      | 1.9514      | 0.063     |
| Limb length        | Twig             | 2.6358 | 2.5895      | 2.6855      | 0.51      |
|                    | Crown-giant      | 1.4204 | 1.4085      | 1.4335      | 0.508     |
|                    | Grass-bush       | 5.5203 | 5.4029      | 5.6252      | 0.002     |
|                    | Trunk-crown      | 1.9208 | 1.8984      | 1.9416      | 0.089     |
|                    | Trunk-ground     | 1.4258 | 1.4074      | 1.4341      | 0.15      |
| Number of lamellae | Twig             | 1.7206 | 1.6943      | 1.7577      | 0.574     |
|                    | Crown-giant      | 3.5478 | 3.5182      | 3.578       | 0.129     |
|                    | Grass-bush       | 3.0619 | 3.0211      | 3.0976      | 0.026     |
|                    | Trunk-crown      | 1.2035 | 1.1957      | 1.2146      | 0.24      |
|                    | Trunk-ground     | 1.6949 | 1.6725      | 1.696       | 0.01      |
may sensibly ask questions about the strength of convergence providing we are concerned with a set of such traits rather than each one individually. In this case, we can measure the strength of convergence in a phenotypic space defined by a set of binary traits, as this essentially creates a quantitative scale of similarity across traits (i.e. species can be more similar by sharing a larger number of discrete traits).

We have not examined the impact of taxon sampling within a clade, but given that all distances are pairwise distances, we do not expect incomplete sampling to be a problem, at least for analyses on the same tree. If incomplete sampling does not pose a problem, we could potentially take a large taxonomic group (e.g. birds, insects, animals) and sample a number of species from this group, encompassing both focal and non-focal taxa, with which we can calculate the Wheatsheaf index. However, we recommend where possible using reasonably well-sampled clades for analysis as this will reduce any concerns over selection of species for inclusion and so avoid potential confirmation bias arising from non-random choice of species to include. In particular, and given that the index works well on small trees, we would recommend that such questions are addressed by taking a number of smaller trees and comparing results across them, rather than using a very large but very poorly sampled clade.

It is important to choose the focal group based on clear, objective criteria based on an a priori hypothesis for two reasons. First, if we assume that convergence is due to adaptation for a particular niche, then it must be considered in relation to that niche. In essence, this instils a biological context to studies of convergence and encourages hypothesis-driven research. Even if we do not assume that the observed convergence is adaptive, the analysis should still be hypothesis-driven in that foci may be defined based on a priori identification of convergent species using other methods (e.g. SURFACE). Secondly, where we consider convergence to be adaptive, it allows us to consider whether convergence has been driven by adaptation to the hypothesized niche. In the case of body shape in burrowing lizards, we might have three data sets with different classifications for the focal group: burrowing, sandy soils and dense ground vegetation. We could then compare the strength of convergence for each of these and examine whether one shows a stronger signal than the others.

A final limitation of our method is that in the current implementation it is problematic to include fossil taxa. Because phylogenetic relatedness is penalized based on the distance from the root of the tree till the point when the pair of species diverged, it assumes that the species continued along independent lineages until the present day. As an extinct taxa pair may have been closely related at the time of their extinction but would be penalized based only on the time of their divergence, they would be considered by our method to be more distantly related than they actually are. Therefore, the Wheatsheaf index can currently only be applied to trees of extant species, although this could potentially be addressed in a future development by using a cophenetic phylogenetic distance to penalize phenotypic similarity when extinct species are included in the study.

CONCORDANCE OF EMPIRICAL RESULTS WITH PREVIOUS LITERATURE

In our Anolis lizard data set, perhaps the most notable finding is that ecomorphs differ in the strength of their convergence – grass-bush and trunk-ground anoles stand out as having particularly strong convergence compared to others. Furthermore, some traits are more strongly convergent within some ecomorphs but not others. Therefore, patterns of convergence in particular traits are ecomorph specific. Given the different niches inhabited by each ecomorph, this is perhaps not surprising as different traits may be more or less needed for a given situation and so the divergence between ecomorphs drives the evolution of different combinations of traits. We will now discuss and highlight that many of our results are consistent with previous literature, which again indicates that the Wheatsheaf index is a useful and meaningful measure of convergent evolution.

Our analyses found the strongest convergence in limb length occurred in grass-bush anoles compared to the other ecomorphs, consistent with Losos’ (1990b, 2009) finding of relationships between limb length and jumping and sprinting (perhaps particularly important for grass-bush anoles). The strong convergence of lamellae number detected in trunk-ground anoles suggests that there is a notable degree of adaptation in this trait. This could be a consequence of opposing selection pressures favouring fewer lamellae than highly arboreal ecomorphs but still enough to permit adequate climbing ability, for example, for making quick dashes down tree trunks to capture prey (Losos 2009). Grass-bush anoles have a small body size to facilitate movement through their structurally complex microhabitat and have long hindlimbs, short forelimbs and an exceptionally elongated tail (Losos 2009). Consistent with this, we found that the Wheatsheaf index was very high for body size, limb length and tail length in grass-bush anoles.

EXTENDIBILITY AND FINAL COMMENTS

It should be noted that, in the current version of the index, the term used to penalize phenotypic similarity for phylogenetic relatedness includes a matrix of shared proportional distances. Consequently, penalized phenotypic distances increase with time since divergence of a given species pair. This implicitly assumes an evolutionary model similar to Brownian motion, wherein we expect greater phenotypic disparity with greater time since divergence. However, the method can be readily extended to explicitly incorporate other evolutionary models by generating the matrix of phylogenetic distances under these models, such as the various variance-covariance structures available in the R package ape (Paradis, Claude & Strimmer 2004). This is a simple extension that relates to the creation of the input files before the calculations of \( w \) are conducted, but may serve to increase the flexibility of the index further.

Another useful extension would be a ‘multi-focal-group’ implementation of the Wheatsheaf index. By this, I mean the ability to investigate many focal groups in the same analysis.
For instance, having several focal groups (e.g. ecomorphs) included in the same index value to assess the extent of convergence in the clade as a whole. However, care would need to be taken to ensure that differences between focal groups would not mask convergence within each focal group.

Finally, we would like to highlight once more that the Wheatsheaf index is not designed to test for the presence of convergence. There are many good methods available for this (see introduction), and we assume that the selection of a group to use our index on is based on the presence of convergent evolution in the clade and that species contained within it have desirable characteristics for the question being asked in a given study. When convergence has been demonstrated, our method then allows the strength of this convergence to be quantified. Also, and particularly, if the specific value of $w$ is to be interpreted, the $P$-values must be discussed in relation to any inference in order to account for topological constraints on $w$.

We have developed and herein presented a novel method for the quantification of convergent evolution. The Wheatsheaf index is intended as an addition to the methodological toolkit for the analysis of convergence (used along with other methods, e.g. those for identification of convergence), and is hoped that it will prove useful in elucidating details of this important and widespread evolutionary process.

**Acknowledgements**

The authors thank M. Berenbrink, S. Cornell and J. Hodgson for helpful discussions while the method was being developed, H.J. Nichols and two anonymous reviewers for comments on the manuscript, and J.B. Losos for providing additional literature for the Anolis example. M.S. thanks S. Conway Morris for a highly stimulating talk on convergent evolution at the University of Liverpool (November 2009) which led the authors to consider the question of whether convergence can be quantified. K.A. was funded by a NERC Doctoral Training Grant. The Wheatsheaf index was named in honour of the pub in which K.A. first devised the method.

**Data accessibility**

Datafiles from the Anolis example are publically hosted on Figshare, accessible via the following URL: http://dx.doi.org/10.6084/m9.figshare.987096. Matlab scripts for implementing the Wheatsheaf index are available on request from the authors.

**References**

Beaulieu, J.M., Jhweng, D.-C., Boettiger, C. & O’Meara, B.C. (2012) Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution*, 66, 2369–2383.

Beuttell, K. & Losos, J.B. (1999) Ecological morphology of Caribbean anoles. *Herpetological Monographs*, 13, 1–28.

Conway Morris, S. (2003) *Life’s Solution: Inevitable Humans in a Lonely Universe*. Cambridge University Press, Cambridge.

Darwin, C.R. (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray Publishers, London, UK.

Dumbacher, J.P. & Fleischer, R.C. (2001) Phylogenetic evidence for colour pattern convergence in toxic pitohuis: Mollérian mimicry in birds? *Proceedings of the Royal Society of London B: Biological Sciences*, 268, 1971–1976.

FitzJohn, R.G. (2012) Diversitree: comparative phylogenetic analyses of diversification in *R*. *Methods in Ecology and Evolution*, 3, 1084–1092.

Hansen, T.F. (2012) Adaptive landscapes and macroevolutionary dynamics. *The Adaptive Landscape in Evolutionary Biology* (eds. E. Svensson & R. Calbick), pp. 205–226. Oxford University Press, Oxford, UK.

Hansen, T.F. & Orzack, S.H. (2005) Assessing current adaptation and phylogenetic inertia as explanations of trait evolution: the need for controlled comparisons. *Evolution*, 59, 2063–2072.

Harmon, L.J., Kolbe, J.J., Cheverud, J.M. & Losos, J.B. (2005) Convergence and the multidimensional niche. *Evolution*, 59, 409–421.

Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E. & Challenger, W. (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics*, 24, 129–131.

Ingram, T. & Mahler, D.L. (2013) SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike information criterion. *Methods in Ecology and Evolution*, 4, 416–425.

Kocher, T.D., Conroy, J.A., McKay, K.R. & Stauffer, J.R. (1993) Similar morphologies of cichlid fish in Lakes Tanganyika and Malawi are due to convergence. *Molecular Phylogenetics and Evolution*, 2, 158–165.

Losos, J.B. (1990a) Ecomorphology, performance capability, and scaling of West Indian *Anolis* lizards: an evolutionary analysis. *Ecological Monographs*, 60, 569–588.

Losos, J.B. (1990b) The evolution of form and function: morphology and locomotor performance in West Indian *Anolis* lizards. *Evolution*, 44, 1189–1203.

Losos, J.B. (1992) The evolution of convergent structure in Caribbean *Anolis* communities. *Systematic Biology*, 41, 403–420.

Losos, J.B. (2009) Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of *Anoles*. University of California Press, Berkeley.

Maddison, W.P. & Maddison, D.R. (2011) Mesquite: a modular system for evolutionary analysis, version 2.75. http://mesquiteproject.org

Mahler, D.L., Ingram, T., Revel, I.J. & Losos, J.B. (2013) Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science*, 341, 292–295.

Martin, C.H. & Wainwright, P.C. (2013) Multiple fitness peaks on the adaptive landscape drive adaptive radiation in the wild. *Science*, 339, 208–211.

McGhee, G.R. (2011) Convergent Evolution: Limited Forms Most Beautiful. Massachusetts Institute of Technology Press, Cambridge.

Miretza, S., Signore, A.V., Burns, J.M., Cossins, A.R., Campbell, K.L. & Berenbrink, M. (2013) Evolution of mammalian diving capacity traced by myoglobin net surface charge. *Science*, 340, 1234192.

Muschick, M., Indenmaur, A. & Salzburger, W. (2012) Convergent evolution within an adaptive radiation of cichlid fishes. *Current Biology*, 22, 2362–2368.

Omland, K.E. & Lanyon, S.M. (2000) Reconstructing plumage evolution in orioles (*Icterus*): repeated convergence and reversal in patterns. *Evolution*, 54, 2119–2133.

Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290.

R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org

Rosenblum, E.B. (2006) Convergent evolution and divergent selection: lizards at the White Sands ecotone. *American Naturalist*, 167, 1–15.

Stayton, C.T. (2006) Testing hypotheses of convergence with multivariate data: morphological and functional convergence among herbivorous lizards. *Evolution*, 60, 824–841.

Stayton, C.T. (2008) Is convergence surprising? An examination of the frequency of convergence in simulated datasets. *Journal of Theoretical Biology*, 252, 1–14.

Turner, G.H., Meiri, S. & Phillimore, A.B. (2009) Body size diversification in *Anolis*: novel environment and island effects. *Evolution*, 63, 2017–2030.

Williams, E.E. (1972) The origin of faunas. Evolution of lizard congeners in a complex island fauna: a trial analysis. *Evolutionary Biology*, 5, 47–89.

© 2014 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society, *Methods in Ecology and Evolution*, 5, 685–693

*Measuring convergent evolution* 693