A null biogeographic test for assessing ecological niche evolution

Laura A. Nunes\textsuperscript{1,2*} and Richard G. Pearson\textsuperscript{1}

1. Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK

2. Vertebrate Division, Department of Life Sciences, Natural History Museum, London, SW7 5BD, UK

*Correspondence: Laura A. Nunes, Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK

Email: laura.nunes.13@ucl.ac.uk

Abstract:

Aims:

Quantification of the degree to which ecological niches change over evolutionary timescales is important for deepening our understanding of evolutionary and ecological processes. Phylogenetic niche conservatism (PNC) is when closely related species differ less ecologically than expected by chance, whereas Phylogenetic Niche Divergence (PND) is when closely related species differ more ecologically than expected by chance. We present a new null model to test for PNC and PND (the RTR significance test), which we combine with a novel metric for quantifying niche overlap.

Location:

Europe, North America and Madagascar

Wordcount: 8154  Pages for Figures and Tables: 5
Methods:

The RTR null model comprises many thousands of replicates generated by randomly translocating and rotating the set of occurrence records for two populations (e.g., sister species) while maintaining the spatial configuration between all occurrences within each replicate. For each replicate we calculate niche overlap as the proportion of the combined niche breadth that is shared by the two species, averaged over \( n \) environmental dimensions. This approach enables us to test whether the observed niche overlap is more or less than expected by chance given the environmental conditions present in the study area. We test the performance of our approach in comparison to other methods using both simulated and real case scenarios, including crested newts in Europe, pocket gophers in North America, and lemurs in Madagascar.

Results:

We find that our measure of niche overlap performs better than other metrics in an artificial simulation scenario, and we find evidence for both PNC and PND in our case studies for Europe, North America and Madagascar. Our results demonstrate that both the RTR significance test and the novel metric of niche overlap are consistent with evolutionary theory and are suitable methods to test for PNC and PND.

Main Conclusions:

We make available scripts to implement the RTR test and metric of niche overlap, and expect that the methods will prove useful for addressing a broad set of questions relating to ecological niche evolution and speciation, particularly for restricted-range species for which few known occurrence records are available.
Keywords: Ecological niche, niche conservatism, niche overlap, phylogeography, null biogeographic model, speciation
Introduction

Knowledge of how species respond to environmental conditions over evolutionary timescales is important for understanding the causes of biodiversity proliferation, change and persistence (Barraclough, 1998; Wiens & Donoghue, 2004). In particular, there is interest in whether tolerance to environmental stressors tends to be conserved across a phylogeny (Wiens et al., 2004) or divergent from species to species (Losos et al., 2008; Ogburn & Edwards, 2015). Phylogenetic niche conservatism (PNC) is the tendency for closely related species to differ less ecologically than expected by chance, and phylogenetic niche divergence (PND) is the tendency for closely related species to differ more ecologically than expected by chance (Pyron et al., 2015). These concepts are of particular importance for understanding the biogeography of speciation; for instance, allopatric speciation has been inferred in cases of PNC (Peterson et al., 1999; Wiens, 2004) whereas PND is expected in cases of parapatric speciation (whereby ecological divergence along an environmental gradient results in species with distinct ecological niches). Studies to-date have revealed evidence both for PNC and PND but methodological difficulties have made the search for general patterns difficult (Wiens & Graham, 2005; Peterson, 2011; Warren et al., 2008; Graham et al., 2004).

A combination of newly available phylogenies, growing databases of species occurrence records, new fine-resolution environmental variables derived from remote sensing, and recently developed GIS-based statistical and machine-learning tools (e.g., ecological niche models, ENMs; also termed Species Distribution Models) provide opportunities to substantially advance understanding of PND and PNC (McCormack et al., 2010; Soberón, 2007). A general methodology for testing for PNC and PND is to: (i) collect
georeferenced occurrence records for populations with known phylogenetic relationships (e.g., sister species); (ii) couple the occurrence records with a set of georeferenced environmental variables (e.g., temperature, precipitation, vegetation cover); and (iii) statistically compare the environments occupied by different populations. We note that this approach tests for differences in the niches currently occupied by the populations (the ‘occupied niche’, sensu Peterson et al., 2011) rather than the niches within which the populations could exist (either the ‘full fundamental niche’ or ‘existing fundamental niche’ sensu Peterson et al., 2011).

PNC and PND can be quantified in terms of niche overlap; that is, the proportion of an environmental niche that is shared between two species (Colwell & Futuyma, 1971). These measures of niche overlap range from 1 (niche equivalency) to 0 (niche divergence), with varying degrees of niche similarity in between (Warren et al., 2008). An important advance proposed by Warren et al. (2008) has been the use of null model tests to assess observed niche differences (or similarities) in the context of the environmental conditions available in the study area. Generating a suitable null model allows us to ask whether the observed similarity or difference between the niches of two populations is statistically meaningful given the available environments. Warren et al. (2008) proposed two null tests, and several other authors have proposed methods for testing niche similarity (see Appendix S3 in Supporting Information for review of methods and their limitations).

Here we present a new method that takes an alternative approach to testing for PNC and PND and addresses some of the problems with other approaches. Specifically, we have devised a method to test whether two populations are currently distributed in such a way that niche overlap is higher (PNC) or lower (PND) than would be expected by chance. We introduce a new null model (the RTR significance test) and a novel metric for quantifying
niche overlap. The new methods are suitable for the study of range-restricted species with few known occurrence records, and were designed to aid our understanding of the landscape-scale ecological processes involved in speciation. We highlight that the current RTR approach is designed to test for PNC among closely related species and is not well suited to applications outside of phylogeography; for example, studies of niche differentiation among invasive species require tests that cover two geographic areas (native and invaded ranges) but the RTR test is limited to a single study area. We make available R scripts to implement the methods (see Appendix S1).

**Materials and Methods**

We propose a methodology that involves four main steps: 1) Collection of georeferenced occurrence data for two populations (e.g., sister species) and environmental variables for the region of interest; 2) Measurement of observed niche overlap; 3) Production of a null reference frequency distribution of niche overlap values; and 4) Comparison of the observed niche overlap value to the null reference distribution to make the decision of rejecting or accepting the null hypothesis. The null hypothesis is that the observed niche overlap is no different to that expected at random, as defined by our null model (see below). Rejection of the null hypothesis would suggest that environmental conditions have played an active role in defining distributions (i.e., populations are adapted to particular ecological niches). Failure to reject the null hypothesis would suggest that spatial rather than ecological processes have been dominant in defining present day distributions.

The two methods that we present below – the new metric and the RTR null model – can be used together (as we do here) but may also be used separately with existing metrics
of niche overlap and significance tests (e.g., the RTR approach could form the basis for running ENMs and calculating the distribution of niche overlap using Schoener’s D or Hellinger’s I under the null hypothesis). All the analyses were conducted in R (R Core Team, 2014).

**A metric for measuring multidimensional niche overlap**

We quantify the overlap ($x$) along a given environmental axis ($e$) between two species ($i$ and $i'$) as:

$$x_e(i, i') = \frac{\min\left(y_{e,i}^{\max}, y_{e,i'}^{\max}\right) - \max\left(y_{e,i}^{\min}, y_{e,i'}^{\min}\right)}{\max\left(y_{e,i}^{\max}, y_{e,i'}^{\max}\right) - \min\left(y_{e,i}^{\min}, y_{e,i'}^{\min}\right)}$$

where $y_{e,i}^{\max}$ and $y_{e,i}^{\min}$ are the maximum and minimum values of $e$ for species $i$ or $i'$ respectively and the overlap ($x_e(i, i')$) is a real-value ranging between 0 and 1. In instances where the environmental breadths of the species do not overlap, our equation will return a negative value which we replace by a value of 0 to indicate no overlap between the axes.

We next average all the axes overlap values across all of the dimensions used to define the niches of the two species using a metric we term $MO$ (for Multidimensional Overlap):

$$MO_{i,i'} = \frac{1}{N} \sum_{e=1}^{N} x_e(i, i')$$

such that the overall overlap ($MO_{i,i'}$) is a real-value between 0 and 1 (see Appendix S4 for further details on the $MO$ metric).

The $MO$ metric is a presence-only approach that is has similarities with the BIOCLIM method (Busby, 1991) in that it constructs simple climate envelopes around the occurrence records (Booth et al., 2014). One limitation of such methods is that they are sensitive to
occurrence records that are ecological outliers, which can lead to overestimation of the ecological niche (Farber & Kadmon, 2003). Overestimation due to outliers can be addressed by selecting a ‘core niche region’, such as the 5-95% percentile of the niche (Carpenter et al., 2003). We therefore implemented a method to undertake an optional ‘trimming’ function to remove ecological outliers (see Appendix S1) and we have tested sensitivity of results to removal of outliers (see Appendix S4). A second limitation is that very limited sampling of the species’ distribution (e.g., fewer than five occurrence records) is likely to lead to underestimation of the species’ niche. However, all methods are limited by the availability of empirical data and our approach has the advantage of avoiding making unfounded extrapolations beyond the range of the available data. A third limitation of BIOCLIM-like methods is that all environmental variables are treated as equally important (there is no weighting of variable importance, unlike in methods such as MAXENT; Phillips et al. 2006). It is therefore important to apply a priori ecological knowledge (e.g., Blair et al. 2013) and/or statistical assessment of variable importance (e.g., by jackknifing, Wielstra et al., 2012; Soto-Centeno et al., 2013) to select relevant environmental variables for the taxa under consideration. In our three case studies (see below) we used the same variables that were used in the original studies, each of which applied a priori statistical analysis or expert knowledge to identify important variables. A fourth limitation is that our approach examines only elements of niche evolution that affect the minimum and maximum values along each niche dimension. Unlike methods that fit response curves in SDMs, our method does not attempt to identify more subtle differences in niches that are reflected in the shape of the functional response to the environment. The benefit of this is that we avoid the many assumptions that go into fitting response curves to limited data and uncertainty over what form the curves should take (Elith et al., 2009).
The RTR test is suitable for populations (e.g., of sister species) that have restricted ranges and few occurrence records, and has several advantages over alternative tests, including: (i) there is no need for reciprocal tests because an individual ‘background’ region does not need to be defined for each species; (ii) it maintains the spatial autocorrelation of the point data; (iii) the test is based on overlap in N-dimensional niche space; (iv) there is no limit to the number of ecological dimensions or occurrence records that can be included; and (v) the test returns a p-value to assess statistical significance rather than two p-values from a reciprocal test due to the need for individual background regions.

RTR null biogeographic model

We have developed a novel test that uses what we term the Random Translocation and Rotation (RTR) null model. The RTR null model comprises many thousands of replicates that are generated by randomly translocating and rotating the pooled set of occurrence records for two populations while maintaining the spatial configuration between all occurrences (i.e., of both compared species together) within each replicate. Niche overlap is calculated for each replicate and the observed niche overlap is then compared to the distribution of overlap values from the null model (Fig. 1). If the observed niche overlap falls outside a critical boundary, we reject the null hypothesis and infer that the niches are conserved (e.g., above 95% percentile of null distribution, PNC) or divergent (e.g., below 5% percentile of null distribution, PND). The significance threshold of this null model approach is not restricted to upper and lower 5% boundaries, thus the investigator is able to select the critical threshold of the model, as well as choose between a two-tailed or one-tailed test for PNC or PND (see Appendix S1).
The RTR null model maintains the spatial configuration, and thus spatial autocorrelation, between all occurrences within each replicate (note somewhat comparable methods by Beale et al., 2008 and Chapman, 2010). The method enables us to test whether the observed distributions are likely to be driven by environmental factors: we ask whether the niches occupied by two species are more or less similar than would be expected at random if the spatial configuration of the set of occurrences is maintained within a given landscape (background region). In effect, we keep the spatial configuration constant so we can ask whether there is something ecologically ‘special’ about the way that the two species are currently located on the landscape. Notice that here we are not assessing the present day spatial configuration of the two species (e.g., are the ranges adjacent?) but rather we are assessing whether the ecological niches currently occupied are more similar or different than expected by chance within the landscape.

As with other null models (e.g., Warren et al., 2008), results from the RTR approach are impacted by the extent of the landscape over which the replicates are run. We address sensitivity to selection of the study region below (see Testing sensitivity to extent of study region). However, an important difference to current methods is that we do not define separate background regions for each species; instead, we randomly translocate and rotate within a single region the set of occurrence records for the two species combined, thus maintaining the spatial configuration between species. We therefore make no assumptions about geographic constraints that might separate the distributions of the two populations. In some instances it may be appropriate to identify likely geographic constraints (e.g., the Isthmus of Tehuantepec in southern Mexico; Peterson et al., 1999, Warren et al., 2008); however, in most cases geographic constraints are less clear and we do not want to impose
a priori any geographic separation when we are testing whether a single population might have split into separate populations (e.g., sister species) due to ecological divergence.

Comparing niche overlap metrics using simulated species

To assess the performance of different measures of niche overlap, we compared our MO metric against three alternatives: (i) Warren et al.’s (2008) I statistic, which is based on a comparison of ENMs and has been widely used; (ii) Broennimann et al.’s (2012) PCA-env framework with the D statistic, a more recent methodology that has been applied to invasive species; and (iii) Blonder et al.’s (2014) n-dimensional technique, in which niches are built as multidimensional hypervolumes.

We compared the metrics using simulated (artificial) species. Simulated species were preferable here to real case studies from nature because the degree of PND and PNC could be precisely defined, providing a ‘known truth’ against which the different metrics could be compared. We simulated the environmental niche overlap of two simulated species in a two-dimensional environmental domain of 100 x 100 grid cells (following Broennimann et al., 2012, and Colwell et al., 2009). We generated two opposing gradients across the artificial landscape to represent two uncorrelated environmental variables. We represented each species’ distribution as a square of 30x30 grid cells (see Appendix S5 for further information on methods and expected outcomes). One species’ distribution was kept static in the bottom left corner of the environmental domain, while the other was initially placed in the bottom left corner but then moved one grid cell at a time in either direction (up or right) or diagonally away from the other species. The movement of one simulated species away from the other across the environmental gradient meant that there was increasing ecological divergence with increasing geographic separation (Fig. S5 in
To calculate the $I$ statistic, we needed to construct ENMs. We used the R package ‘DISMO’ 1.0-5 (Hijmans et al., 2011) to build ENMs and the package ‘SDMTools’ 1.1-221 (VanDerWal, et al., 2012) to calculate $I$. We opted for ENMs constructed using MAXENT 3.3.3x (Phillips et al., 2006) as this method performs well in comparison with other ENM approaches (Elith et al., 2006) and was used in Warren et al.’s original paper (2008). We followed Warren et al. (2008) in maintaining default values for all program settings, including regularization and feature selection. To calculate $D$ we used the R script for uncalibrated PCA-env functions provided by Broennimann et al. (2012). To build and measure the intersection of two hypervolumes following Blonder et al. (2014), we used their package ‘HYPERVOLUME’ 1.4.1 with 1,000 random numbers, a bandwidth of 0.1, a quantile of 0, and a reduction factor of 0.5, as recommended by Blonder et al. (2014).

Testing the null model using three case studies

We also tested our methodology (RTR null model combined with the new MO metric of niche overlap) using three real-world case studies. We selected case studies based on the following criteria: (i) a previous study has been published that includes estimates or hypotheses regarding PNC or PND; and (ii) occurrence records are available for the included sister species. Following each of the original studies, we used georeferenced environmental layers from Worldclim at 30 arc-seconds resolution (Hijmans et al., 2005), in each case selecting the same set of variables as the original study (see Appendix S6). We selected two case studies with a continental setting (the Balkan Peninsula and North America) and one island (Madagascar) since islands have more obvious natural boundaries and therefore raise different issues concerning the selection of a suitable study region:
i. Crested newts in the Balkans. The phylogeography of the crested newts has been extensively studied (Arntzen et al., 2007, Wielstra et al., 2010, Wielstra et al., 2012), which allows us to make some predictions about the potential for niche conservatism. There is evidence for vicariance among two sister pairs: (i) *Triturus karelinii* and *T. ivanbureschi*, which is thought to have split due to the uplift of the Armenian Plateau; and (ii) *T. carnifex* and *T. macedonicus*, which is thought to have split due to the formation of the Adriatic Sea (Wielstra et al., 2010). Wielstra et al. (2012) used ordination methods to assess niche evolution and found evidence for PNC. This finding is in line with the theoretical expectation that adaptation to new niches tends to be slower than extinction rates, leading to PNC in cases of allopatric speciation (Peterson et al. 1999). We aimed to test this using the RTR method.

Occurrence records for the crested newt species were obtained from Wielstra et al. (2012). We set the western and eastern boundaries of Europe (Fig. 2a) based on the overall distribution of the *Triturus* clade in Europe (Wielstra et al., 2012). The environmental layers used (see Appendix S6) were clipped to the same extent using the crop function from the ‘RASER’ 2.5-2 package (Hijmans, 2015), and all layers were converted to a Lambert Conformal Conic projection (at central meridian 18°E, standard parallels 42°N and 46°N), which represents low spatial distortion for the Balkan region (Zagmajster et al., 2008).

ii. Pocket gophers in North America. Though studies of ecological divergence are commonly conducted at the species level, there is also interest in looking at a population level in order to understand biogeographic patterns of within-species divergence (Graham et al., 2004; Glor & Warren, 2011). *Geomys pinetis* is the only species of pocket gopher found in south-eastern USA and within this species there is a geographic and genetic
subdivision between eastern (*G. pinetis*) and western populations (hereafter *G. mobilensis*) on either side of the Apalachicola River (Soto-Centeno et al., 2013). Soto-Centeno et al. (2013) used the ‘identity’ test and the ‘blob’ range-breaking test of ENMTools (Glor & Warren, 2011) and could not reject a hypothesis of niche similarity between the two populations across the river barrier. Here we aimed to test whether a signal for PNC or PND could be detected using our RTR method.

Occurrence records for *G. pinetis* and *G. mobilensis* were obtained from MaNIS (manisnet.org), following Soto-Centeno et al. (2013). We ran the RTR test for the south-eastern region of the country (Fig. 2b) which is equivalent to the extent used in Soto-Centeno et al. (2013). The Lambert conformal conic projection (central meridian 96°E, standard parallels 20°N and 60°N) was used to transform the environmental layers (see Appendix S6) because this has low spatial distortion for North America (Les et al., 2013).

iii. Lemurs in Madagascar. Blair et al. (2013) found that two sister pairs of *Eulemur* lemurs in Madagascar (*E. collaris*-*E. cinereisceps* and *E. rufus*-*E. rufifrons*) have clear riverine barriers; for one of these pairs (*E. rufus*-*E. rufifrons*) they found evidence of PNC, and for the other pair (*E. collaris*-*E. cinereisceps*) they found no significant signal, based on the null background tests of Warren et al. (2008). By contrast, two other sister pairs (*E. albifrons*-*E. sanfordi* and *E. flavifrons*-*E. macaco*) were found to have less well defined geographic barriers; for one of these pairs they found support for significant PND for one pair (*E. flavifrons*-*E. macaco*) and for the other pair (*E. albifrons*-*E. sanfordi*) they found no significant signal (Blair et al., 2013).

We obtained the *Eulemur* occurrence records from Blair et al. (2013) and restricted the analysis to the island of Madagascar (Fig. 2c). The environmental layers used (see...
Appendix S6) were transformed to an oblique Mercator projection for Madagascar (following Pearson et al. 2007).

**Testing sensitivity to extent of study region**

The RTR test requires selection of an overall study region within which the RTR replicates are generated. The selection of this overall background region will affect the model output and the interpretation of the findings (e.g., the observed niche overlap being unique in relation to a particular region but not to another). Selecting an extent that is too wide might include environments that are too different from what the species or clade are likely to experience thus biasing the null library to unrealistic measurements of niche overlap. Selecting a small extent will lead to a reduction in the available environmental heterogeneity being sampled which could result in the exclusion of such unrealistic environments, thus reducing irrelevancy in null library. However, a too narrow extent could mean that there is more similarity between the environments occupied by the observed distribution and the environments sampled within the study region due to higher likelihood of partial overlap within and between the simulated distributions and the observed distributions. Smaller study regions are therefore expected to return lower type I error rates (i.e., false rejection of the null hypothesis) than tests performed across larger areas. In general, a study region should be selected that bounds the landscape, and hence the set of environmental conditions, that the species could reasonably be expected to have had the opportunity to occur in. Thus, factors such as the dispersal capacity of the species, topographic features in the landscape (e.g., barriers to dispersal), and the distribution of major clades to which the species belong might be considered when selecting the study region.
We tested sensitivity of the RTR significance test to the extent of the study region by undertaking *Eulemur* analysis for the whole island of Madagascar and also for alternative smaller extents. We did this for the two most range-restricted sister pairs: *E. collaris* - *E. cinereiceps* and *E. falvifrons* - *E. macaco*. We divided Madagascar latitudinally, approximately cutting the island in half, to generate sub-island extents that take into account the island’s general north-south zonation (Blair *et al.*, 2013). We also generated further restricted study regions by approximating areas of lemur endemism proposed by Pastorini *et al.* (2003). Thus, *E. collaris* - *E. cinereiceps* was tested with a southern extent and a more restricted south-eastern extent (corresponding to region E2 in Pastorini *et al.* (2003); Fig. 2c) and *E. falvifrons* - *E. macaco* was tested with a northern extent and a smaller north-western extent (corresponding to region X in Pastorini *et al.*, 2003; Fig. 2c).

**Results**

**Performances of different niche overlap metrics in a common simulated scenario**

We found contrasting performance between niche overlap metrics when tested using simulated species, with some metrics having a tendency to overestimate niche overlap while others tend to underestimate niche overlap in relation to the ‘known truth’ scenario (Fig. 3b).

The novel metric presented in this study, *MO*, captures the expected results more closely than the other three metrics (Fig. 3), with the expected ranges for each region being matched precisely with the outcomes observed in the new metric (Fig. 3c).
Warren’s $I$ metric over-predicted niche overlap in instances where no overlap should be observed in region D (Fig. 3d). This metric also did not capture the expected gradient in niche overlap in region C (Fig. 3d). However, it captured the expected range of range of niche overlap in region B (Fig. 3d). Moreover, niche identity (complete niche overlap) was correctly predicted to occur only in Region A (Fig. 3d).

Broennimann’s $D$ statistic also tended to over-predict niche overlap in region D (Fig. 3e); however, this was to a lesser extent than Warren’s $I$ statistic, as it was able to identify some cases of complete niche divergence in that region. The expected gradient from 0 to 0.5 was observed in region C (Fig. 3e). The metric was the only one to overestimate niche overlap (ranging from 0.2-1) in region B (Fig. 3e). Broennimann et al.’s method also overpredicted the occurrence of identical niches, which was predicted beyond region A (Fig. 3e). Nevertheless, overall it was able to capture the expected range of 0 to 1 (Fig. 3e).

Blonder et al.’s (2014) hypervolume approach also captured the 0 to 1 range, though niche overlap was under-predicted in regions C and D, where only niche divergence was observed (Fig. 3f). However, the approach accurately captured the expected range from 0 to 1 in region B, and complete niche overlap was correctly detected only in Region A.

Performance of the RTR null biogeographic model in real case scenarios

The RTR test supports a finding of PNC for both sister pairs of crested newts (Table 1). This is in agreement with previous findings (Wielstra et al., 2012).

For the populations of pocket gophers, we found no significant signal for either PNC or PND, which is consistent with the results of Soto-Centeno et al. (2013). However, the observed niche overlap was close to significant for PNC (observed $MO = 0.59$ and 95%
threshold for significance = 0.61), indicating that further studies may be appropriate to test this finding (Table 1).

Results for lemurs show contrasting findings between the RTR null model and the null model of Warren et al. (2008), with agreement in the findings for only one out of four sister pairs. The RTR test finds support for PNC for the pair *E. collaris* - *E. cinereiceps* (Table 1). This is in contrast to the findings of Blair et al. (2013), who found no significant signal using the tests of Warren et al. (2008). For the pair *E. rufus* - *E. rufifrons*, we find evidence for PND. This is again in contrast to Blair et al. (2013), who found some support for PNC. For the other two *Eulemur* pairs, we find no significant signal using the RTR test. This is in agreement with Blair et al. (2013) for the *E. albifrons* - *E. sanfordi* pair, but in conflict for *E. flavifrons* - *E. macaco* (Blair et al. (2013) found evidence for PND).

Changing the extent of the study region for *Eulemur* pairs showed that the RTR test is sensitive to selection of the area over which the test is run (Table 2). For one sister pair (*E. collaris* - *E. cinereiceps*) we observed loss of significant signal as the extent of the study region was reduced; however this loss of signal was only observed when the background area was very close in extent to that of the two species. Thus, use of a small study extent produced a result consistent with the findings from Blair et al. (2013), who also used a small extent (‘background’ regions were selected by constructing minimum convex polygons around occurrence records). For the other sister pair, no significant signal was found regardless of the extent of background area used (Table 2).

**Discussion**

Performances of different niche overlap metrics in a common simulated scenario
Our analysis showed contrasting outcomes across niche overlap metrics for the same simulated scenario. These findings highlight the importance of choosing an appropriate metric for answering specific ecological and/or evolutionary questions. We found that the statistic introduced in this paper, the MO metric, performed better than other metrics by producing results closer to those expected from the ‘known truth’ scenario.

We found poorer performance of Warren’s I statistic and Broenniman’s PCA-env procedure with the D statistic for quantifying niche overlap when compared to other metrics. The two statistics tend to be correlated (Warren et al., 2008) and we have demonstrated a common tendency to overestimate the simulated niche overlap when no niche overlap is expected to be found and underestimated when one environmental variable overlaps but not the other. This tendency to overestimate the niche overlap has been reported previously in simulated environments (Broennimann et al., 2012) and is likely because the approaches measured the intersection of predictions from two ENMs, which are designed to estimate suitability and therefore tend to overestimate species’ distributions (because some suitable habitats will be unoccupied; Peterson et al., 2011).

We found a tendency for the n-dimensional hypervolume approach to underestimate niche overlap in our simulation. This is particularly noticeable in instances where the niches do not overlap on at least one axis. There are multiple aspects that could account for underestimation of niche overlap using the hypervolume approach (Blonder et al., 2014). In our simulations, the most likely explanation for underestimation is that the hypervolume method measures niche overlap as the intersection of two volumes such that the volumes will not intersect at all (niche overlap is measured as 0) if they have one or more non-overlapping variables. None of the other metrics we assessed make this strict assessment of
niche overlap, either because they calculate niche overlap by doing a cell-by-cell comparison (e.g., Warren’s $I$ and Broenniman’s $D$) or they calculate an averaged proportion of the niche breadth shared across all axis (our MO metric). However, we note that in cases when the volumes did overlap on both axes in our simulation, the performance of the hypervolume approach was comparable to that of the MO metric (which closely followed the ‘known truth’).

In comparison with the other metrics tested, we found that our new MO metric better captured niche overlap in the simple simulated scenario that we tested. There is, however, scope to refine this metric in light of limitations (see Methods). One limitation is the potential sensitivity to ecological outliers (see Appendix S4). We have implemented a function to remove ecological outliers (see Appendix 1) but removal of outliers is only advised when it is expected that some points may be erroneous (e.g., misidentifications) or there are likely sink populations that do not represent the niche. Removing true ecological extremes will lead to a misrepresentation of the species occupied niche. Although there are limitations, we have demonstrated in our comparisons that the conceptually simple MO metric is a useful approach for quantifying niche overlap.

**Performance of the RTR null biogeographic model in real case studies**

By applying the RTR test to real case studies, we have shown that the new test can provide comparable outcomes to expected and previously observed patterns from the literature. We have also shown differences in the outputs of the RTR method and Warren *et al.*’s (2008) background test, with agreement for only one out of four pairs, but congruency between the RTR method and ordination techniques and the ‘blob’ range-breaking test for the remaining pairs. Application of the MO metric and the RTR significance test supports a
hypothesis of PNC for two allopatric sister pairs of crested newts in the Balkans (*T. carinifex* - *T. macedonicus* and *T. karelinii* - *T. ivanbureschi*). This finding is in line with previous work that has used ordination techniques to identify PNC in this study system (Wielstra et al., 2012) and offers a way of further testing the theoretical expectation that PNC is a key pattern emerging from allopatric speciation (Cooper et al., 2010; Losos, 2011; Crisp & Cook, 2011). For pocket gophers in south-eastern North America, we found no significant signal for either PNC or PND, which implies that observed ecological differences between the populations are not likely due to selection for a particular set of conditions. Rather, our results suggest that niche differences are a coincidental result of different environments available on each side of a geographic barrier (the Apalachicola river).

The endemic lemurs of Madagascar have been assessed previously in tests of which, if any, mode of speciation may have been most important in driving local endemism and speciation across the island (Pearson & Raxworthy, 2009; Blair et al., 2013). Our measurements of niche overlap using the *MO* metric were congruent with Warren et al.’s *I* and *D* statistics, with our results consistently falling within the range of values presented by Blair et al., (2013). However, the RTR test identified significant statistical support for PNC in the pair *E. collaris*-*E. cinereiceps*, which is not congruent with the results using Warren et al.’s background test (Blair et al., 2013). Our finding, combined with strong genetic support for the Mananara river acting as a barrier to gene flow (Wyner et al., 2002), suggests an allopatric mode of speciation for this pair, with niches failing to evolve on either side of a geographic divide.

Our analyses find no significant support for either PNC or PND in two *Eulemur* sister pairs (*E. flavifrons*-*E. macaco* and *E. albifrons*-*E. sanfordi*). These results are in contrast to
those of Blair et al. (2013), who found support for PND for one of the pairs (E. flavifrons - E. macaco). Closer analysis of the differences between the null libraries produced by Blair et al. (2013) using Warren’s background test and our RTR significance test shows that the RTR test captured a broader null distribution for E. flavifrons - E. macaco. This explains the difference in statistical inference between the two null models for these sister pairs, despite comparable observed niche overlap values between the two studies. However, it is important to highlight that both E. flavifrons - E. macaco and E. albifrons - E. sanfordi have no complete geographical barriers and have potential hybrid zones (see Blair et al., 2013 and references therein). This suggests that other factors besides geographic or environmental separation may be responsible for speciation within these pairs (e.g., microhabitat selection, Rakotondranary & Ganzhorn, 2011) which would account for the lack of signal found in our analysis.

We found support for PND for E. rufus - E. rufifrons, which opposes the hypothesis of PNC of Blair et al. (2013), who found weak support for PNC. However, there is limited spatial overlap between the potential distributions of the pairs (Blair et al., 2013), which suggests that the species have different ecological preferences and is consistent with our finding of PND.

One explanation for differences in results between the RTR test and Warren et al.’s (2008) null models is that the RTR approach focuses on the ecological dimension of niche evolution between species while Warren et al. (2008) focuses on the geographical dimension (i.e., our RTR method measures overlap in niche space, whereas Warren et al. (2008) measures the spatial overlap of ENMs). This difference likely accounts for different biological inferences from the alternative methods. We contend here that our approach of
measuring overlap in niche space is most appropriate for robustly testing for PNC and PND, but further research is needed to decipher the different facets of niche evolution that are picked up by each of the approaches. Consequently, we expect that the new methods presented here will complement, rather than replace, existing methods.

We highlight that the RTR approach is best suited to study range-restricted species where many unique RTR replicates can be generated within a background region. Instances where a species is wide-ranging relative to the study region, or where the ‘shapes’ of the species’ distribution and study region are such that RTR replicates can be located in only a limited number of ways, will result in few replicates and potential spatial biases. We illustrate that spatial biases are case specific in Appendix S7 by showing the locations of 10,000 RTR replicates for a pair of newts and a pair of lemurs. Spatial bias in the RTR null model acts to curtail the background region, and therefore the range of environments considered (note that this does not negate the meaning of the significance test for PNC or PND, but it does mean that the range of conditions over which the calculation is performed is only a subset of those in the selected background region). As with selection of the background region, spatial bias within the RTR null model will be an important area for future research. To facilitate this we provide R code in Appendix S1 for replicating the analyses we have done in Appendix S7.

As with other tests for PNC and PND, the RTR test found instances with no significant signal. In addition to the actual absence of either PNC or PND, there are a number of methodological factors that can explain non-significant results, including: (i) the choice of predictor variables (some key variables may be excluded from the analysis); (ii) the coarse resolution of analysis (niche differentiation within the 1km$^2$ cells used will not be picked up);
(iii) incomplete knowledge of occurrences (modern-day loss and fragmentation of habitats means we have an incomplete picture of the niches of the species), and (iv) inappropriate extent of the study region (explored in more detail below).

**Testing sensitivity to extent of study region**

We have shown that the RTR test is sensitive to the extent of the study region. Our tests for Madagascar suggest that sensitivity to the selection of study region extent is low, unless the extent becomes very small (e.g., of similar size to the extent of the two species’ distributions). Further exploration of this sensitivity, and of different strategies for selecting the study region (e.g., based on dispersal capacity, following Anderson & Raza, 2010), is warranted.

Discordance between our results and those of Blair *et al.* (2013) may be due in part to the different sensitivities of the RTR test and Warren *et al.*’s (2008) test to the extent of the study region. Blair *et al.* (2013) defined the background area of each species based on a minimum convex polygon bound by the occurrence records of each species. A serious drawback of using minimum convex polygons around occurrence records for each species is the assumption that all locations within the polygon are suitable habitat for the species. This assumption is unlikely to hold in many cases (e.g., consider a species that occupies warm lowlands around the base of a mountain: drawing a minimum convex polygon around the species’ occurrence records will encompass both the lowlands and also the cooler mountain top).

A result that is non-significant is not expected to become significant if the extent is made smaller. This was observed in our analysis of *E. flavifrons* - *E. macaco*. We thus find
that a tight study region, such as those defined by Blair et al. (2013), is more prone to
result in falsely accepting the null hypothesis. A wider extent, by contrast, is more likely to
result in rejection of the null hypothesis.

Conclusions

Overall, we find that the new metric of niche overlap, the MO metric, and the new RTR
significance test are suitable methods for testing for PNC and PND, particularly when
applied to range-restricted species with few occurrence records. Given its novel approach to
the study of niche dynamics between populations, the RTR method holds great promise for
testing for PNC and PND across large phylogenies with many sister pairs, and thus shedding
new light on evolutionary processes, in particular speciation.

Acknowledgements

We thank Mark Wilkinson, Daniel Chapman, Town Peterson and two anonymous reviewers
for very useful comments, and Phillip Staniczenko for help with the mathematical notation.
L.A.N. is supported by a studentship co-funded by UCL and the Natural History Museum.

References

models of species geographic distributions and estimates of niche evolution: Preliminary
tests with montane rodents (genus Nephelomys) in Venezuela. Journal of Biogeography,
37,1378–1393.


The climate envelope may not be empty. *Proceedings of the National Academy of Sciences USA*, **106**, E47.


**Supporting Information**
Additional Supporting Information may be found in the online version of this article:

- **Appendix S1** R codes.
- **Appendix S2** Methods for R code of the RTR test.
- **Appendix S3** Review of existing methods.
- **Appendix S4** MO metric.
- **Appendix S5** Virtual scenarios for testing niche overlap metrics.
- **Appendix S6** Environmental variables for case studies.
- **Appendix S7** Assessing potential spatial bias.

**Biosketch**
Laura Nunes is currently doing a Ph.D. on ecological niche evolution and the biogeography of speciation, particularly among chameleons in Madagascar. Richard Pearson is interested in the ecological and evolutionary processes that determine species distributions.
Author contributions: R.G.P. and L.A.N. conceived the ideas and designed the analyses.
L.A.N. wrote the code and performed the analyses. L.A.N. wrote the paper with input from R.G.P.

Editor: Daniel Chapman
Figures and Tables

Table 1. Observed niche overlap measured with the MO metric and outputs from the RTR null biogeographic model as a significance test for the observed niche overlap value against a null library for two sister pairs of crested newts in the Balkans (T. karelinii – T. ivanbureschi and T. carnifex-T. macedonicus), a sister pair of pocket gophers (G. pinetis – G. mobilensis) in North America and four sister pairs of lemurs from in Madagascar (E. collaris – E. cinereiceps; E. flavifrons-E. macaco; E. albifrons – E. sanfordi and E. rufus-E. rufifrons). PNC refers to Phylogenetic Niche Conservatism when the observed niche overlap is higher than expected by chance, and PND refers to Phylogenetic Niche Divergence for cases where the observed niche overlap is lower than expected by chance. No significant signal refers to cases where the observed niche overlap does not occur less often than expected by random chance. Results from the RTR test are compared against expectations based on published papers.

<table>
<thead>
<tr>
<th>Sister pair</th>
<th>Original study</th>
<th>Observed Niche Overlap</th>
<th>5%, mean and 95% tails for Niche Overlap based on a null RTR distribution (MO statistic)</th>
<th>Hypothesis from original study</th>
<th>RTR test result</th>
</tr>
</thead>
<tbody>
<tr>
<td>T. karelinii - T. ivanbureschi</td>
<td>Wielstra et al., 2012</td>
<td>0.61</td>
<td>0.26;0.37;0.51</td>
<td>PNC</td>
<td>PNC</td>
</tr>
<tr>
<td>T. carnifex - T. macedonicus</td>
<td>Wielstra et al., 2012</td>
<td>0.69</td>
<td>0.23;0.42;0.65</td>
<td>PNC</td>
<td>PNC</td>
</tr>
<tr>
<td>Species Pairs</td>
<td>Authors and Year</td>
<td>MO</td>
<td>0.31; 0.47; 0.61</td>
<td>No significant signal</td>
<td>No significant signal</td>
</tr>
<tr>
<td>--------------------</td>
<td>------------------</td>
<td>-----</td>
<td>-----------------</td>
<td>-----------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>G. pinetis - Soto-Centeno et al., 2013</td>
<td>0.59</td>
<td>0.59</td>
<td>0.31; 0.47; 0.61</td>
<td>No significant signal</td>
<td>No significant signal</td>
</tr>
<tr>
<td>G. mobilensis</td>
<td>Blair et al., 2013</td>
<td>0.56</td>
<td>0.05; 0.24; 0.51</td>
<td>No significant signal</td>
<td>PNC</td>
</tr>
<tr>
<td>E. collaris - E. cinereiceps</td>
<td>0.34</td>
<td>0.09; 0.31; 0.57</td>
<td>PND</td>
<td>No significant signal</td>
<td></td>
</tr>
<tr>
<td>E. flavifrons - E. macaco</td>
<td>0.24</td>
<td>0.09; 0.25; 0.45</td>
<td>No significant signal</td>
<td>No significant signal</td>
<td></td>
</tr>
<tr>
<td>E. albifrons - E. sanfordi</td>
<td>0.24</td>
<td>0.09; 0.25; 0.45</td>
<td>No significant signal</td>
<td>No significant signal</td>
<td></td>
</tr>
<tr>
<td>E. rufus – E. rufifrons</td>
<td>0.19</td>
<td>0.27; 0.48; 0.67</td>
<td>PNC/No</td>
<td>PND</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Observed niche overlap measured with the MO metric and outputs from the RTR null biogeographic model as a significance test for the observed niche overlap value against a null library for two Eulemur sister pairs in Madagascar using alternative study region extents. Regions E2 and X refer to proposed areas of endemism for lemurs according to Pastorini et al. (2003) which coincide with the range-restricted sister pairs (see main text). The results from the RTR null model are compared against hypotheses based on published papers. Warren et al.’s (2008) background similarity test is based on results from Blair et al. (2013). PNC refers to Phylogenetic Niche Conservatism when the observed niche overlap is higher than expected by chance, and PND refers to Phylogenetic Niche Divergence for cases where the observed niche overlap is lower than expected by chance. No significant signal refers to cases where the observed niche overlap does not occur less often than expected.
by random chance.

<table>
<thead>
<tr>
<th>Sister pair</th>
<th>Extent of study</th>
<th>Niche Overlap</th>
<th>P-value</th>
<th>Background similarity (obtained from RTR test)</th>
<th>RTR test result</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. collaris - E. cinereiceps</td>
<td>Region E2</td>
<td>0.10 ; 0.40 ; 0.70</td>
<td>0.26</td>
<td>No significant signal</td>
<td>PNC</td>
</tr>
<tr>
<td>E. cinereiceps</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southern Region</td>
<td>0.56</td>
<td>0.06 ; 0.23 ; 0.48</td>
<td>0.02</td>
<td>Significant signal</td>
<td>PNC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole island</td>
<td></td>
<td>0.05 ; 0.24 ; 0.52</td>
<td>0.03</td>
<td></td>
<td>PNC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. flavifrons - E. macaco</td>
<td>Region X</td>
<td>0.20; 0.37; 0.56</td>
<td>0.41</td>
<td>No significant signal</td>
<td>PND</td>
</tr>
<tr>
<td>E. macaco</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northern Region</td>
<td>0.36</td>
<td>0.10; 0.32; 0.60</td>
<td>0.42</td>
<td>Significant signal</td>
<td>No significant signal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PND</td>
</tr>
<tr>
<td>Whole island</td>
<td></td>
<td>0.10; 0.31; 0.57</td>
<td>0.40</td>
<td>No significant signal</td>
<td>No significant signal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Random translation and rotation (RTR) null model test for phylogenetic niche divergence (PND) and phylogenetic niche conservatism (PNC), illustrated for a pair of sister lemur species in Madagascar (see Blair et al., 2013). (a) Occurrence records for the sister species, visualized on an example environmental layer (multiple environmental layers e.g., temperature, precipitation) are used in the test to characterize the n-dimensional ecological niche space. (b) Thousands of null replicates are generated by randomly rotating and translating the set of occurrence records for the two species. (c) For the observed distributions and for all null replicates, the niche overlap is calculated and the observed overlap is compared against the null model. In this instance the observed niche overlap is in the highest 5% of the null distribution, so we infer PNC.
Figure 2. Species’ occurrence records and extents of the study regions used in the three case studies. (a) Crested newts in the (T. karelinii – T. ivanbureschi and T. carnifex-T. macedonicus); (b) Populations of pocket gophers in south-eastern USA (G. pinetis – G. mobilensis); (c) Lemurs in Madagascar (E. collaris – E. cinereiceps; E. flavifrons-E. macaco; E. albifrons – E. sanfordi and E. rufus-E. rufifrons). Regions E2 and X refer to proposed areas of endemism for lemurs according to Pastorini et al. (2003) which coincide with the range-restricted sister pairs (see main text). The outlines (dashed lines in (c)) represent the alternative extents used to test sensitivity of RTR null model to differences in the extent of the study region.
Figure 3. Niche overlap values generated by four alternative metrics for a simulated scenario. (a) We used the classification of niche overlap established by Rödder & Engler (2011) but also distinguished between completely divergent niches (niche overlap [NO]=0) and identical niches (NO=1). (b) The ‘known truth’ represents the expected ranges of niche overlap in each section of a 2-dimensional 70x70 grid (see Appendix S4 for more details). (c) The new niche overlap statistic (MO) introduced in this paper. (d) Warren et al.’s (2008) background similarity test with the I statistic. (e) Broenniman et al.’s (2012) PCA uncalibrated technique with the D statistic. (f) Blonder et al.’s (2014) intersection of hypervolumes between two species.