



HAL
open science

Putting phylogeny into the analysis of biological traits: A methodological approach

Thibaut Jombart, Sandrine Pavoine, Sébastien Devillard, Dominique Pontier

► To cite this version:

Thibaut Jombart, Sandrine Pavoine, Sébastien Devillard, Dominique Pontier. Putting phylogeny into the analysis of biological traits: A methodological approach. *Journal of Theoretical Biology*, 2010, 264 (3), pp.693. 10.1016/j.jtbi.2010.03.038 . hal-00591239

HAL Id: hal-00591239

<https://hal.science/hal-00591239>

Submitted on 8 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

Putting phylogeny into the analysis of biological traits: A methodological approach

Thibaut Jombart, Sandrine Pavoine, Sébastien Devillard, Dominique Pontier

PII: S0022-5193(10)00173-6
DOI: doi:10.1016/j.jtbi.2010.03.038
Reference: YJTBI5939



www.elsevier.com/locate/jtbi

To appear in: *Journal of Theoretical Biology*

Received date: 20 January 2010
Revised date: 25 March 2010
Accepted date: 25 March 2010

Cite this article as: Thibaut Jombart, Sandrine Pavoine, Sébastien Devillard and Dominique Pontier, Putting phylogeny into the analysis of biological traits: A methodological approach, *Journal of Theoretical Biology*, doi:[10.1016/j.jtbi.2010.03.038](https://doi.org/10.1016/j.jtbi.2010.03.038)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Putting phylogeny into the analysis of biological traits: a methodological approach

Thibaut Jombart^a, Sandrine Pavoine^b, Sébastien Devillard^c, Dominique Pontier^c

^a*MRC Centre for Outbreak Analysis & Modelling, Department of Infectious Disease Epidemiology, Imperial College London, Faculty of Medicine, Norfolk Place, London W2 1PG, UK.*

^b*Museum National d'Histoire Naturelle; Département Ecologie et Gestion de la Biodiversité; UMR 7204 MNHN-CNRS-UPMC; CRBPO, 61 rue Buffon, 75005, Paris, France*

^c*Université de Lyon; Université Lyon 1; CNRS; UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, 43 boulevard du 11 novembre 1918, Villeurbanne F-69622, France.*

Abstract

Phylogenetic comparative methods have long considered phylogenetic signal as a source of statistical bias in the correlative analysis of biological traits. However, the main life-history strategies existing in a set of taxa are often combinations of life history traits that are inherently phylogenetically structured. In this paper, we present a method for identifying evolutionary strategies from large sets of biological traits, using phylogeny as a source of meaningful historical and ecological information. Our methodology extends a multivariate method developed for the analysis of spatial patterns, and relies on finding combinations of traits that are phylogenetically autocorrelated. Using extensive simulations, we show that our method efficiently uncovers phylogenetic structures with respect to various tree topologies, and remains powerful in cases where a large majority of traits are not phylogenetically structured. Our methodology is illustrated using empirical data, and implemented in the free software R.

Keywords:

phylogenetic principal component analysis, pPCA, autocorrelation, multivariate, comparative method, phylogenetic signal

Email address: tjombart@imperial.ac.uk (Thibaut Jombart)

1 1. Introduction

2 Phylogeny has long been recognised as a major source of biological variation.
3 For instance, Gregory (1913) and Osborn (1917) considered that species' vari-
4 ability should be partitioned between *heritage* (*i.e.*, phylogenetic inertia) and
5 *habitus* (*i.e.*, adaptation). In their well-known criticism of the adaptationist
6 paradigm, Gould and Lewontin (1979) underlined the importance of the con-
7 straints imposed by the phylogeny to the variability observed among organisms.
8 In comparative studies, the effect of phylogeny has merely been perceived as
9 a source of nuisance, since it reveals non-independence among trait values ob-
10 served in taxa (Dobson, 1985; Felsenstein, 1985), and thus violates one of the
11 basic assumptions required by most statistical tools (Harvey and Pagel, 1991).

12
13 Phylogenetic comparative methods (PCM) were especially designed to solve
14 this problem. Various methods have been developed that transform quantita-
15 tive traits into new variables that are not correlated to phylogeny, according to
16 a given model of evolution. For instance, phylogenetic independent contrasts
17 (PIC, Felsenstein, 1985) transform values observed at the n tips of a phylogeny
18 into $(n - 1)$ node values that are not phylogenetically autocorrelated under a
19 Brownian motion model. Generalised least squares (GLS, Grafen, 1989; Rohlf,
20 2001) is a more general technique that allows specifying the autocorrelation of
21 observations as a component of a linear model. This approach can therefore
22 account for the non-independence among observations using a wide variety of
23 models of evolution (Hansen and Martins, 1996). As stressed by Rohlf (2006),
24 these approaches do not actually remove phylogenetic autocorrelation from the
25 data, but merely take it into account to provide more accurate estimates of
26 model parameters. In fact, PIC, GLS, along with other existing PCM all aim
27 towards the same goal: 'correcting for phylogeny' in the correlative analysis of
28 biological traits at the species level (Harvey and Purvis, 1991; Martins, 2000;
29 Martins et al., 2002; Garland et al., 2005).

30

31 Nonetheless, studying the phylogenetic patterns of trait variation allows for
32 formation of hypotheses about the evolutionary pathways that led to the trait
33 values of extant species. It also allows shedding light onto the influence of
34 historical and ecological processes on community assembly (Webb et al., 2002).
35 Many biologically meaningful patterns are inherently structured with phylogeny.
36 Indeed, many life-history and ecological strategies are likely to be phylogenet-
37 ically structured (Webb et al., 2002). Inheritance from a common ancestor and
38 phylogenetic inertia (*i.e.*, constraints to evolution) may cause phylogenetic sig-
39 nal (similar trait values across closely related species) to occur. Other factors
40 leading to phylogenetic signals in traits act at the population level rather than
41 at the species level such as high gene flow, lack of genetic variation, stabilising
42 selection if changes in trait states reduce fitness, or population growth if traits
43 are pleiotropically linked to other traits that reduce fitness (Wiens and Gra-
44 ham, 2005). However, traits might also be affected by variations unrelated to
45 the phylogeny, but relating to ecological conditions experienced by the species.
46 For instance, biotic interactions might drive character displacement and abiotic
47 interactions might lead to trait convergence. From this perspective, phyloge-
48 netic signal becomes a source of precious biological information that can be
49 used to identify historical as well as recent evolutionary strategies. Interest-
50 ingly, a similar paradigm shift occurred in spatial ecology (Legendre, 1993)
51 when it was pointed out that spatial patterns in species' distribution were not
52 only sources of spurious correlations, but also indicators of critical ecological
53 structures such as localised species assemblages and species-environment asso-
54 ciations. This paradigm shift proved particularly fecund and still motivates
55 innovative developments in statistical ecology (*e.g.*, Dray et al., 2006; Griffith
56 and Peres-Neto, 2006).

57

58 In this paper, we present a method which uses phylogenetic information
59 to uncover the main phylogenetic structures observable in multivariate data
60 associated with a phylogeny. Our approach, *phylogenetic principal component*
61 *analysis* (pPCA), extends a methodology developed in spatial ecology (Dray

62 et al., 2008) and in spatial genetics (Jombart et al., 2008) to the analysis of
 63 phylogenetic structures in biological features of taxa such as life-history traits.
 64 We emphasise that phylogenetic structures can be measured and quantified in
 65 the same way as spatial structures, as they are both associated with the concept
 66 of autocorrelation. We then define different kinds of phylogenetic structures,
 67 and show how pPCA can be used to identify them. After evaluating the ability
 68 of pPCA to uncover phylogenetic patterns through extensive simulations, we
 69 illustrate our method using an empirical example. pPCA is implemented in
 70 the *adephylo* package (Jombart and Dray, 2009) for the free software R (R
 71 Development Core Team, 2009).

72 2. Methods

73 2.1. Measuring phylogenetic autocorrelation

74 Phylogenetic autocorrelation is said to occur whenever the values taken by a
 75 set of taxa for a given biological trait are not independent of the phylogeny. Fre-
 76 quently, closely related taxa exhibit more similar traits than randomly-chosen
 77 taxa. Moran's I , an index originally used to measure spatial autocorrelation
 78 (Moran, 1948, 1950), has been proposed for measuring phylogenetic autocor-
 79 relation (Gittleman and Kot, 1990). Adapting the former definition (Cliff and
 80 Ord, 1973, p13) to the phylogenetic context, I is defined as:

$$I_{\mathbf{W}}(\mathbf{x}) = \frac{\mathbf{x}^T \mathbf{W} \mathbf{x}}{n} \frac{1}{\text{var}(\mathbf{x})} \quad (1)$$

81 where \mathbf{x} is the centred vector of a trait observed on n taxa, $\text{var}(\mathbf{x})$ is the
 82 usual variance of \mathbf{x} , and \mathbf{W} is a matrix of phylogenetic proximities among taxa
 83 ($\mathbf{W} = [w_{ij}]$ with $i, j = 1, \dots, n$), whose diagonal terms are zero ($w_{ii} = 0$),
 84 and rows sum to one ($\sum_{j=1}^n w_{ij} = 1$). The null value, *i.e.* the expected value
 85 when no phylogenetic autocorrelation arises, is $I_0 = -1/(n-1)$ (Cliff and Ord,
 86 1973). In its initial formulation (Gittleman and Kot, 1990), *i.e.* before row
 87 standardisation so that $\sum_{j=1}^n w_{ij} = 1$, \mathbf{W} contained binary weights. Before this
 88 standardisation, the entry at row i and column j was set to 1 if taxon i shared

89 a common ancestor with taxon j at a given taxonomic level, and to 0 otherwise.
 90 Hence, taxa were considered as either phylogenetically related or not. Moran's
 91 I then compared the trait value of a taxon to the mean trait value in related
 92 taxa to detect phylogenetic autocorrelation.

93
 94 Such binary relationships are clearly not sufficient to model the possibly
 95 complex structure of proximities among taxa induced by the phylogeny. To
 96 achieve better resolution in these comparisons, we propose using as entries of
 97 \mathbf{W} any measurement of phylogenetic proximity valued in \mathbb{R}^+ verifying:

$$\begin{cases} w_{ij} \geq 0 & \forall i, j = 1, \dots, n \\ w_{ii} = 0 & \forall i = 1, \dots, n \\ \sum_{j=1}^n w_{ij} = 1 & \forall i = 1, \dots, n \end{cases} \quad (2)$$

98 Then, Moran's I compares the value of a trait in one taxon (terms of \mathbf{x})
 99 to a weighted mean of other taxa states (terms of $\mathbf{W}\mathbf{x}$) in which phylogenet-
 100 ically closer taxa are given stronger weights. This extension gives the index
 101 considerable flexibility for quantifying phylogenetic autocorrelation, as phyloge-
 102 netic proximities can be derived from any model of evolution (including or not
 103 branch lengths). For instance, one interesting possibility would be using the
 104 covariance matrix estimated in a GLS model (Grafen, 1989) to define phyloge-
 105 netic proximities. This could be achieved by setting diagonal terms (variances)
 106 of the covariance matrix to zero, adding the smallest constant ensuring that all
 107 terms are positive, and row-standardizing the resulting matrix.

108
 109 This formulation of Moran's I also relates the index to other PCM. For
 110 instance, the test proposed by Abouheif (1999), initially based on the many
 111 possible planar representations of a tree, turned out to be a Moran's I test
 112 using a particular measure of phylogenetic proximity for \mathbf{W} (Pavoine et al.,
 113 2008).

114 Moran's I is also related to autoregressive models. In their simplest form,

115 these models are written as (Cheverud and Dow, 1985; Cheverud et al., 1985):

$$\mathbf{x} = \rho \mathbf{W}\mathbf{x} + \mathbf{Z}\boldsymbol{\beta} + \mathbf{e} \quad (3)$$

116 where ρ is the autocorrelation coefficient, \mathbf{Z} is a matrix of explanatory variables,
 117 $\boldsymbol{\beta}$ is the vector of coefficients, and \mathbf{e} is a vector of residuals. The matrix of phy-
 118 logenetic relatedness \mathbf{W} (Cheverud and Dow, 1985; Cheverud et al., 1985) is
 119 exactly the weight matrix of our definition of Moran's I (equation 1). The essen-
 120 tial difference between the two approaches is that autoregressive models perform
 121 the regression of \mathbf{x} onto $\mathbf{W}\mathbf{x}$, while I computes the inner product between both
 122 vectors (numerator of equation 1) to measure phylogenetic autocorrelation.

123 Lastly, the weighting matrix \mathbf{W} is also the core of another approach pro-
 124 ducing variables that model phylogenetic structures (Peres-Neto, 2006). Like
 125 Moran's I , this approach was initially developed in spatial statistics (Griffith,
 126 1996), and consisted in finding eigenvectors of a doubly centered spatial weight-
 127 ing matrix (Dray et al., 2006). Applied to a matrix of phylogenetic proxim-
 128 ity \mathbf{W} , this method yields uncorrelated variables modeling different observable
 129 phylogenetic patterns, each related to a value of Moran's I . Peres-Neto (2006)
 130 performed the regression of a variable \mathbf{x} onto these eigenvectors to partial-out
 131 the phylogenetic autocorrelation from \mathbf{x} . Alternatively, we suggest using these
 132 eigenvectors to simulate what we further call 'global' and 'local' phylogenetic
 133 structures.

134 2.2. Global and local phylogenetic structures

135 Phylogenetic autocorrelation relates to the non-independence of trait values
 136 observed in taxa given their phylogenetic proximity. There are two ways in
 137 which this non-independence can arise, depending on whether closely related
 138 taxa tend to have more similar, or more dissimilar trait values than expected at
 139 random, resulting in *positive* and *negative autocorrelation*, respectively. Positive
 140 phylogenetic autocorrelation most often results in global patterns of similarity
 141 in related taxa; we thus refer to these patterns as *global structures*. Global pat-
 142 terns reflect the general idea of phylogenetic signal: trait values observed in a set

143 of taxa are not independent, but tend to be more similar in closely related taxa
 144 (*e.g.*, Figure 1A). Most common explanations for this phenomenon are inheri-
 145 tance from a common ancestor, or the conservation of ecological niches (Harvey
 146 and Pagel, 1991). Traits whose evolution can be modeled by a Brownian or by
 147 an OU process with low stabilising constraint generally display global patterns
 148 (Abouheif, 1999; Pavoine et al., 2008). Such phenomenon typically results in
 149 close-to-the-root divergence in evolutionary strategies.

150 Conversely, negative phylogenetic autocorrelation corresponds to dissimilarities
 151 among tips localised in specific parts of the tree, which we call *local struc-*
 152 *tures*. A local structure would be observed whenever closely related taxa tend
 153 to be more different with respect to a given trait than randomly chosen taxa
 154 (*e.g.*, Figure 1E). Local structures correspond to relatively recent events that
 155 induced divergence of the evolutionary strategies close to the tips of the phy-
 156 logenetic tree, such as convergence and character displacement (following past
 157 or present biotic interactions). This also occurs when the trait under study has
 158 been selected towards different optimal values, resulting in opposed evolution-
 159 ary strategies being observed in sister taxa.

160 Both structures can be identified using Moran's index (equation 1). The sign of
 161 I depends on how values of a trait (\mathbf{x}_i) relate to the values observed on closely
 162 related taxa ($\mathbf{W}\mathbf{x}_i$). Moran's I will be greater than (respectively less than) I_0
 163 (value of I in the absence of autocorrelation) when closely related taxa tend
 164 to have similar (respectively dissimilar) values for the studied trait. Obviously,
 165 the definition of phylogenetic proximities in \mathbf{W} will condition the measurement
 166 of global and local structures. As shown by Pavoine et al. (2008), not all phy-
 167 logenetic proximities are equal in detecting phylogenetic structures. Especially,
 168 the phylogenetic proximities underlying Abouheif's test (matrix $\mathbf{A} = [a_{ij}]$ in
 169 Pavoine et al. 2008) proved superior to several common phylogenetic proxim-
 170 ities for testing phylogenetic inertia in traits simulated under Brownian and
 171 Ornstein-Uhlenbeck (OU) processes. More generally, the matrix \mathbf{W} can be de-
 172 rived from any model of evolution which seems appropriate to the data, taking
 173 branch lengths into account whenever these are accurately estimated, and rely-

174 ing only on the topology in other cases.

175 Ecological and life-history strategies of species require not one, but several traits
 176 to be adequately described. Accordingly, ecological and life-history strategies
 177 are likely to involve combinations of traits with both global and local phyloge-
 178 netic structures. In the following, we describe a methodology which explicitly
 179 investigates multivariate phylogenetic structures that have barely been consid-
 180 ered so far.

181 2.3. The phylogenetic principal component analysis

182 Dray et al. (2008) and Jombart et al. (2008) developed a multivariate ap-
 183 proach for identifying spatial structures in multivariate data. Essentially, this
 184 approach consists in constraining the principal components of a multivariate
 185 method to exhibit spatial autocorrelation, as measured by Moran's index. This
 186 methodology proved better at detecting autocorrelated patterns than usual mul-
 187 tivariate methods such as principal component analysis (Dray et al., 2008; Jom-
 188 bart et al., 2008). Here, we use the same rationale to define the *phylogenetic*
 189 *principal component analysis* (pPCA), a method designed to summarise a set
 190 of traits into a few synthetic variables exhibiting global or local phylogenetic
 191 structures. Note that while we presented pPCA for the analysis of quantitative
 192 traits for the sake of simplicity, this approach can be extended to qualitative
 193 traits, or even to mixtures of quantitative and qualitative variables (Dray et al.,
 194 2008).

195
 196 We denote $\mathbf{X} = [x_{ij}]$ ($\mathbf{X} \in \mathbb{R}^{n \times p}$) a matrix containing p quantitative traits
 197 measured on n taxa, and \mathbf{W} a matrix of phylogenetic weights used in the com-
 198 putation of Moran's I (equation 1). As in classical PCA, missing data can be
 199 set to the mean of the corresponding trait, which does not add artefactual struc-
 200 tures to the analyzed traits. Without loss of generality, we assume that traits
 201 are centered (*i.e.*, $\sum_i x_{ij} = 0$ with $j = 1, \dots, p$). The purpose of pPCA is to
 202 find linear combinations of traits (columns of \mathbf{X}) containing a large variance and
 203 displaying global or local phylogenetic structures. Mathematically, this problem

204 translates into finding the appropriate loadings $\mathbf{u} \in \mathbb{R}^p$ (with $\|\mathbf{u}\|^2 = 1$) that
 205 minimise and maximise, respectively, the function:

$$f: \mathbb{R}^{n \times p} \times \mathbb{R}^{n \times n} \times \mathbb{R}^p \longrightarrow \mathbb{R}$$

$$(\mathbf{X}, \mathbf{W}, \mathbf{u}) \longmapsto \text{var}(\mathbf{X}\mathbf{u})I_{\mathbf{W}}(\mathbf{X}\mathbf{u}) \quad (4)$$

206 The solution to this problem is given by the diagonalisation of the matrix
 207 $\frac{1}{2n}\mathbf{X}^T(\mathbf{W} + \mathbf{W}^T)\mathbf{X}$ (Dray et al., 2008; Jombart et al., 2008). It results in a
 208 set of loadings $\{\mathbf{u}_1, \dots, \mathbf{u}_k, \dots, \mathbf{u}_r\}$ with $\mathbf{u}_k \in \mathbb{R}^p$ forming linear combinations
 209 of traits ($\mathbf{X}\mathbf{u}_k$, the so-called principal components) associated with decreasing
 210 eigenvalues λ_k , so that:

$$\text{var}(\mathbf{X}\mathbf{u}_k)I_{\mathbf{W}}(\mathbf{X}\mathbf{u}_k) = \lambda_k \quad (5)$$

211 The largest eigenvalues likely correspond to a large variance and a strong posi-
 212 tive I , indicating global structures (close-to-root variation in trait states). Con-
 213 versely, the lowest (*i.e.*, most negative) eigenvalues correspond to a high vari-
 214 ance and a large negative I , indicating local structures (close-to-tips variation
 215 in trait states). As in other reduced space ordinations, the eigenvalues indicate
 216 the amount of structure expressed by each synthetic variable. A sharp decrease
 217 in the screeplot is likely to indicate a shift between strong and weak structures.
 218 The amount of variance ($\text{var}(\mathbf{X}\mathbf{u}_k)$) and phylogenetic autocorrelation ($I(\mathbf{X}\mathbf{u}_k)$)
 219 in each principal component ($\mathbf{X}\mathbf{u}_k$) can be computed for a better interpretation
 220 of each structure. Moreover, the loadings \mathbf{u}_k can be used to assess how traits
 221 contribute to a given principal component, and thus understand the nature of
 222 the corresponding biological structure.

223
 224 One important choice is that of the phylogenetic weights (\mathbf{W}) used in the
 225 analysis. Here, we use the measure of phylogenetic proximity underlying the test
 226 of Abouheif (1999) to define \mathbf{W} , because of its good performances at detecting
 227 phylogenetic structures (Pavoine et al., 2008). The phylogenetic proximity a_{ij}
 228 among tips i and j is defined as:

$$a_{ij} = \frac{1}{\prod_{p \in P_{ij}} dd_p} \text{ for } i \neq j \quad (6)$$

229 where P_{ij} is the set of internal nodes on the shortest path from tips i to j ,
 230 and dd_p is the number of direct descendants from the internal node p . The
 231 phylogenetic proximity a_{ij} defines the entries of the off-diagonal terms of \mathbf{W} ,
 232 all diagonal entries being set to 0. As \mathbf{W} is row-standardised, it is defined as:

$$w_{ij} = \frac{a_{ij}}{\sum_{j=1, i \neq j}^n a_{ij}} \quad (7)$$

233 2.4. Sensitivity study

234 Extensive simulations were conducted to evaluate the sensitivity of the pPCA
 235 to various parameters. Datasets were simulated with different characteristics
 236 concerning the type of tree, the tree size, the type, strength and numbers of
 237 phylogenetically structured traits, and the total number of traits (including
 238 structured and unstructured traits). These parameters are summarised in Ta-
 239 ble 1. Five types of trees of 16, 32, or 128 tips were simulated to encompass
 240 a wide range of tree topologies and sizes: completely symmetric trees (Figure
 241 1A), trees obtained by random clustering of tips (as implemented by `rtree`
 242 function of the `ape` package, Paradis et al., 2004, Figure 1B), the Yule model
 243 (Yule, 1924, Figure 1C), the biased model (Kirkpatrick and Slatkin, 1993, Figure
 244 1D), and completely asymmetric trees (Figure 1E). Datasets including random
 245 traits and phylogenetically structured traits (*i.e.*, displaying global and/or lo-
 246 cal structures) were obtained for each tree. Random traits were drawn from a
 247 normal distribution ($\mathcal{N}(0, 1)$), while ‘structured traits’ were obtained by adding
 248 normally-distributed random noise to phylogenetic eigenvectors of \mathbf{W} (Peres-
 249 Neto, 2006). Whenever several structures of the same type (global or local)
 250 were simulated in a given dataset, these were derived from the same eigenvec-
 251 tor, so that we could evaluate the performance of pPCA when a ‘consensus’
 252 phylogenetic signal exists in a set of traits (*e.g.*, Figure 1B). This was consis-
 253 tent with the fact that several phylogenetically structured traits are expected
 254 to exhibit the same patterns, either because these structures are caused by the
 255 same evolutionary process, or because all traits are correlated to another phy-
 256 logenetically structured trait. 200 datasets were simulated for each of the 810

257 different combinations of parameters, resulting in a total of 162 000 datasets.

258

259 All simulations were performed in the R software (R Development Core
260 Team, 2009). Trees were simulated using the R packages *ape* (Paradis et al.,
261 2004) and *apTreeshape* (Bortolussi et al., 2006), and scripts developed by TJ for
262 the symmetric model. Structured traits were simulated using the *ade4* package
263 (Chessel et al., 2004; Dray et al., 2007), and data were handled using the *phy-*
264 *lobase* package (Bolker et al., 2007).

265

266 Each dataset was analysed by a pPCA using Equation 7 to define phyloge-
267 netic proximities. In each analysis, the structured traits were compared to the
268 first relevant (global and/or local) principal component of pPCA, to assess how
269 the method performed. The strength of the link between the original simulated
270 structures and patterns identified by pPCA was measured using the absolute
271 value of Spearman's rank correlation, $|\rho|$. Whenever the dataset included sev-
272 eral distinct structured traits, $|\rho|$ values were averaged by type of structure (*i.e.*,
273 global or local). Hence, we obtained one or two $|\rho|$ per simulated dataset, used
274 as indicator of the performance of pPCA ($|\rho|$ close to one = high performance,
275 $|\rho|$ close to zero = low performance).

276

277 Variations in $|\rho|$ according to the different simulation parameters were in-
278 vestigated using a linear model. The relationship between $|\rho|$ and the predic-
279 tors was linearised using a logit link, *i.e.* using $\text{logit}(|\rho|) = \log \frac{|\rho|}{1-|\rho|}$ as the
280 response variable. When interpreting coefficients of the model, predictions $\hat{\mu}$
281 were re-transposed onto the $|\rho|$ scale, that is, replacing $\hat{\mu}$ by $\frac{1}{1+e^{-\hat{\mu}}}$. Qualita-
282 tive variables were modeled using treatment-coded contrasts (Faraway, 2004, p.
283 173). The explanatory variables were the type of tree (factor 'tree', the biased
284 model being the intercept), the type of structuring (factor 'strutype', with level
285 'global' at the intercept), the number of tips ('ntips', intercept=16), the total
286 number of traits ('ntraits', intercept=10), the standard deviation of the random
287 noise added to structured variables ('noise', intercept=0.5), and the number of

288 structures (factor 'nstruc', intercept=1).

289 2.5. Empirical data analysis

290 We illustrate pPCA with the data on reproductive and morphometric traits
291 within a small group of species and subspecies from the lizard family Lacer-
292 tidae published by Bauwens and Diaz-Uriarte (1997). These data are currently
293 available as the dataset *lizards* in the *adephylo* package (Jombart and Dray,
294 2009) for R (R Development Core Team, 2009). They consist in a molecular
295 phylogeny and 8 life-history traits measured for 16 taxa: mean adult length
296 (in mm, abbreviated *mean.L*), length at maturity (in mm, *matur.L*), maximum
297 length (in mm, *max.L*), hatchling length (in mm, *hatch.L*), hatchling mass (in
298 g, *hatch.m*), clutch size (in number of descendents, *clutch.S*), clutch frequency
299 (in number of events per year, *clutch.F*), and age at maturity (in number of
300 months of activity, *age.mat*). All traits were measured on females. Adult life
301 span and egg size were discarded from the analysis because data were missing
302 for several taxa.

303 The analyses were conducted in the R software (R Development Core Team,
304 2009), using the *ade4* package (Chessel et al., 2004; Dray et al., 2007) for factor-
305 ial analyses and *adephylo* (Jombart and Dray, 2009) to perform the pPCA. As
306 in Bauwens and Diaz-Uriarte (1997), data were log-transformed and regressed
307 onto mean adult female length to partial out the body size effect. As a con-
308 sequence, mean adult female length was removed from the analysis. We then
309 investigated phylogenetic structures of the transformed life-history traits using
310 pPCA.

311 3. Results

312 3.1. Sensitivity study

313 All explanatory variables had a very significant effect on the response vari-
314 able (Appendix A, Table A.1), which was trivial because even very low effects
315 might be significant with a large number of observations. All coefficients of

316 the model (Appendix A, Table A.2) have therefore been interpreted quantita-
 317 tively to determine the level of effects of the explanatory variables. The model
 318 explained satisfyingly 57% of the total variance. Overall, the average $|\rho|$ was
 319 relatively high ($CI_{99\%} = [0.667; 0.669]$), showing that phylogenetic structures
 320 were well retrieved by pPCA. The strongest effect was by far that of the type of
 321 structure: global patterns were more easily retrieved than local structures, with
 322 a difference of 0.31 in predicted $|\rho|$ (later denoted $\Delta_{|\hat{\rho}|}$). pPCA performed bet-
 323 ter in larger trees ($\Delta_{|\hat{\rho}|} = 0.11$ between trees with 16 and 128 tips), suggesting
 324 that phylogenetic signal is more easily captured when a large number of taxa
 325 is available, which is in line with previous findings for a different PCM (Mar-
 326 tins and Hansen, 1997). For a given number of structured traits, the number
 327 of random traits slightly lowered the method's ability to retrieve phylogenetic
 328 patterns ($\Delta_{|\hat{\rho}|} = 0.10$ between 10 and 50 traits). Phylogenetic structures in-
 329 corporating larger amounts of random noise were also more difficult to retrieve
 330 ($\Delta_{|\hat{\rho}|} = 0.10$ between noise of 0.5 and 1). Lastly, the number of structured traits
 331 and the type of tree only marginally affected the ability of pPCA to identify
 332 phylogenetic patterns.

333 3.2. Empirical data analysis

334 Both global and local phylogenetic structures were found by pPCA in the lac-
 335 ertid lizards data (Figure 2). The first global principal component of pPCA first
 336 opposed a lineage with three species (*Lacerta schreiberi*, *L. agilis*, *L. vivipara*)
 337 having the largest negative scores to the rest of the tree (taxa with positive
 338 scores, or scores closer to zero), with the subspecies *Podarci h. h.* exhibiting
 339 the most opposite life histories (Figure 2A). Among the remaining species or
 340 subspecies, *Lacerta monticola cantabrica* and *L. m. cyreni* were distinctive by
 341 their negative scores. The loadings of the analysis (Figure 2B) provided further
 342 insights on the corresponding evolutionary strategies, and showed a trade-off
 343 between the frequency of reproductive events per year (clutch.F) and the clutch
 344 size (clutch.S). *L. schreiberi*, *L. agilis*, *L. vivipara*, and to a lesser extent *L. m.*
 345 *cantabrica* and *L. m. cyreni*, reproduce less often but produce a larger number

346 of eggs per reproductive event than other populations. A possible explanation
347 for this structure is that environmental conditions only allow for a few repro-
348 duction events in these populations, which then deliver lots of eggs with poor
349 individual survival rates. In contrast, the first local principal component of
350 pPCA (Figure 2A) highlighted a strong opposition between related taxa, es-
351 pecially between *Takydromus tachydromoides* and *Acanthodactylus erythrurus*,
352 but also to a lesser extent between *L. schreiberi* and *L. agilis*. This opposition
353 was also apparent, although weak, within two additional lineages (first *Podarcis*
354 *muralis*, *P. bocagei* versus *P. h. atrata*, *P. h. hispanica* Asturias; second *L.*
355 *m. cyreni* versus *L. m. cantabrica*). Figure 2B (vertical axis) shows the mean-
356 ing of these local variations. The species with positive scores on the first local
357 principal component (especially, *T. tachydromoides* and *L. schreiberi*) produce
358 a large number of small eggs while species with negative scores (especially *A.*
359 *erythrurus* and *L. agilis*) produce fewer, but larger descendants.

360 4. Discussion

361 Phylogenetic autocorrelation has so far been considered as a mere nuisance
362 to the correlative analysis of comparative biological data, when exploring trade-
363 offs as well as in allometric studies. In this paper, we advocate that phylogenetic
364 autocorrelation is a source of relevant biological information for the exploratory
365 analysis of such data. To accomplish this task, we introduced the *phylogenetic*
366 *principal component analysis* (pPCA), a method that we adapted from existing
367 multivariate spatial statistics (Dray et al., 2008; Jombart et al., 2008) to anal-
368 yse phylogenetic structures in multivariate sets of traits. Based on the results
369 obtained from simulated and empirical data, we discuss the ability of the pPCA
370 to retrieve phylogenetic signals in a multivariate set of traits, and the impact
371 that this approach could have in evolutionary ecology.

372

373 *4.1. Performance of the approach - methodological discussion*

374 Preliminary results stemming from our sensitivity study were very promis-
375 ing, and provided guidelines for applications of pPCA. Overall, pPCA performed
376 well to retrieve simulated phylogenetic structures, even in some cases where only
377 1 out of 50 traits was phylogenetically structured. pPCA seemed to retrieve
378 global phylogenetic structures more easily than local structures. This may be
379 due to the asymmetry of Moran's I distribution, which often has a smaller range
380 of variation in negative values (local structures) than in positive values (global
381 structures) (de Jong et al., 1984). As pPCA seeks principal components with
382 extreme values of I , global structures (associated with large positive I) would
383 be more easily detected than local structures (associated with large negative I).
384 Therefore, local structures may be interpreted even though the corresponding
385 eigenvalue seems negligible compared to global structures, provided it is bio-
386 logically significant. Other results of our sensitivity study suggest that pPCA
387 performs better in larger trees, although performances on small phylogenies were
388 satisfying. Interestingly, pPCA seemed rather insensitive to the shape of the,
389 indicating that the method can be used with virtually any kind of phylogeny.

390
391 Although pPCA will be best appreciated using empirical datasets, further
392 simulation studies may be considered. In this study, we used eigenvectors of
393 a phylogenetic proximity matrix (Peres-Neto, 2006) to simulate phylogenetic
394 structures. This method allows simulation of complex phylogenetic patterns in
395 negligible computational time, which permitted examination of the influence a
396 large number of parameters on pPCA results. The drawback of this approach is
397 that eigenvectors of phylogenetic proximity matrices are not directly related to
398 an model of evolution such as the Brownian motion or the OU models. While
399 current procedures implementing trait simulation under these models are more
400 computer-intensive, it could be possible to study how pPCA behaves under these
401 models, given variation in a few parameters.

402
403 The main parameter that should be investigated in further detail is the

404 phylogenetic proximity used in pPCA. A previous study demonstrated that
405 some phylogenetic proximities were better than others at detecting phyloge-
406 netic structures using Moran's index (Pavoine et al., 2008). However, it is likely
407 that the most appropriate measurement of phylogenetic proximity depends on
408 the dataset under scrutiny. To assess whether a given phylogenetic proximity is
409 adapted to a particular dataset, we advocate to perform Moran's I test using
410 this proximity matrix. Whenever significant structures are detected, one can
411 input this phylogenetic proximity in pPCA to uncover the nature of the under-
412 lying phylogenetic structures.

413

414 4.2. Potential impacts of the approach in evolutionary and ecological studies

415 This novel approach should complement nicely the usual PCM toolbox,
416 bringing a new perspective to the analysis of comparative biological data. Con-
417 trary to usual PCM, our approach does not attempt to improve estimates of
418 correlations among traits by 'correcting' for phylogenetic dependence among
419 species. Instead, it seeks biologically meaningful combinations of traits that are
420 globally or locally phylogenetically structured, thus allowing us to uncover fun-
421 damental evolutionary patterns. As noted by Bauwens and Diaz-Uriarte (1997),
422 *theories of life-history evolution are explicitly micro-evolutionary [...] whereas*
423 *patterns of life-history covariation are most evident when comparisons are made*
424 *among higher taxonomic levels.* pPCA covers both of these aspects, by provid-
425 ing insights about broad macro-evolutionary patterns (global structures) and
426 more recent, even micro-evolutionary patterns (local structures).

427

428 Life histories, for example, are likely to be phylogenetically structured (Gail-
429 lard et al., 1989; Pontier et al., 1993; Rochet et al., 2000). In our case study, the
430 pPCA identified phylogenetic patterns in the main life-history tactics adopted
431 by a set of taxa. Our results suggest that the trade-off between clutch fre-
432 quency and size may have resulted in the ancient divergence of evolutionary
433 strategies. In contrast, the trade-off between hatchling mass on the one hand

434 and clutch size and frequency on the other hand appears to be more labile, in-
435 volving more recent character changes in most of the lineages and especially be-
436 tween *T. tachydromoides* and *A. erythrurus*. The pPCA thus allows description
437 of global (close-to-root, phylogenetic signal) versus local (close-to-tips) phylo-
438 genetic structures in a multivariate set of traits, and highlights which lineages
439 and which taxa are involved in these structures. Overall, this illustration using
440 an empirical dataset showed that pPCA can bring new insights about evolu-
441 tionary strategies of a set of taxa. Moreover, whenever a molecular clock is
442 available for the considered phylogeny, it would be possible to estimate the age
443 of the involved lineages and taxa, by dating their most recent common ances-
444 tors. This would allow assessing how and when different evolutionary strategies
445 might have appeared and evolved along the history of the considered taxa. Local
446 structures uncovered by pPCA point out more recent evolutionary events, such
447 as speciation caused by diversifying selection or niche separation, and are thus
448 also of fundamental interest. Dating these recent events would be even more
449 interesting as historical information about the considered taxa is more likely
450 to be available for recent speciation events. For instance, we could investigate
451 whether a recent speciation highlighted as local structure would have been pre-
452 ceded by significant modifications of the environment.

453
454 A further strength of pPCA lies in its ability to analyse very large sets of
455 traits (*i.e.*, hundreds or thousands of traits) simultaneously. Usual PCM typ-
456 ically rely on pairwise comparisons among traits, which becomes cumbersome
457 when lots of variables are under scrutiny, and often requires discarding traits
458 from the analysis. This issue will be increasingly concerning in the near future
459 as new and large databases of life-history traits will become available. pPCA
460 can be used to explore such data, to unveil evolutionary trade-offs among a large
461 number of traits, without having to make a prior selection of analysed traits.
462 Previous methods have already attempted to analyse phylogenetic signals in a
463 series of traits. These methods determine the proportion of variation in a set
464 of traits correlated with the phylogenetic relatedness among species (Giannini,

2003; Desdevises et al., 2003). They involve factorial analyses, but with the aim of partitioning variation in traits instead of depicting phylogenetic structures. Nevertheless, one step of the Giannini (2003) variation partitioning approach consists of selecting the nodes of the phylogeny that better explain the variation in the trait values. Accordingly, the selected nodes could be used to determine at which depth in the phylogeny the taxa differ in their trait values. However, this selection results from a long series of tests to determine if the differences between the lineages that descend from a given node are significantly responsible for trait variation. The number of tests depends on the size of the phylogenetic tree, with an increasing risk of erroneously significant tests. The pPCA approach thus brings a new optimised way of disentangling the phylogenetic patterns in a set of traits by identifying the lineages and also the combination of traits responsible for global versus local trait variation.

A potential application of the pPCA concerns phylogenetic community ecology (Webb et al., 2002). Phylogenetic clustering in a community (lower phylogenetic diversity than expected by chance in a regional pool of species) merely reflects the simultaneous action of environmental filtering and phylogenetic conservatism. In contrast, distinct, even opposed processes can lead to phylogenetic overdispersion (higher phylogenetic diversity within a community than expected by chance in a regional pool of species). For example, phylogenetically overdispersed communities can arise from (i) limiting similarity and conservative traits or (ii) environmental filters with convergent traits (Kraft et al., 2007). Consequently, knowledge of the evolution of traits is necessary to interpret observed structures in phylogenetic diversity. A difficulty is that the traits involved in environmental filters and those involved in limiting similarity might follow different evolutionary pathways (Emerson and Gillepsie, 2008; Ackerly et al., 2006). Although pPCA does not provide a formal test of phylogenetic conservatism or over-dispersion, it can be used to describe the phylogenetic signal induced by these processes. Therefore, our methodology could be applied to describe the level of phylogenetic signal in sets of traits from labile traits with local close-

496 to-tips variations to conserved traits with global close-to-root variations. This
497 approach could provide even more details by highlighting lineage-dependent sig-
498 nals. For example, a single trait might exhibit a general pattern of phylogenetic
499 signal (global structure) and also strong localised trait variations in a single
500 lineage (local structure). Moreover, pPCA can be used together with other
501 multivariate methods to relate the combination of traits identified with a given
502 phylogenetic structure (either global or local) to explanatory factors, such as
503 environmental variables. For instance, co-inertia analysis (Dolédéc and Chessel,
504 1994; Dray et al., 2003a,b) could be used to link phylogenetic structures iden-
505 tified by pPCA to descriptors of the ecological niche, so as to assess potential
506 patterns of adaptation. pPCA could therefore complement both existing ecolog-
507 ical methods (*e.g.*, co-inertia) and evolutionary approaches (*e.g.*, phylogenetic
508 overdispersion/clustering), providing a link between trait evolution, patterns in
509 phylogenetic diversity, and biotic or abiotic interactions, and giving insights into
510 the historical and ecological processes that underpin community assembles.

511 To conclude, we illustrate the intersection between issues in spatial and phy-
512 logenetic methods. Spatial and phylogenetic patterns are generated by very
513 different processes, but the mathematical tools that can be used to measure
514 and model these patterns may be similar. This is because both rely on the con-
515 cept of autocorrelation, which can be defined as the non-independence among
516 observations with respect to a set of underlying proximities. Several spatial
517 methods developed in ecology have already been successfully adapted to PCM
518 (Cheverud et al., 1985; Gittleman and Kot, 1990; Diniz-Filho et al., 1998; Des-
519 devises et al., 2003; Giannini, 2003). Originally, spatial autocorrelation was
520 perceived by ecologists as a nuisance that precluded the use of standard statisti-
521 cal tools in correlative studies. However, the study of spatially autocorrelated
522 patterns turned out to be a fecund paradigm, as ecologists realised that these
523 structures were mere indicators of considerable underlying ecological processes.
524 The same may be true of phylogenetically autocorrelated patterns. Rewording
525 Legendre (1993), we can now also ask the question: *is phylogenetic autocorrela-*
526 *tion trouble, or a new paradigm?*

527 **5. Acknowledgements**

528 We are grateful to the CCIN2P3 for providing access to their computers,
529 and particularly to Simon Penel for his help. We thank Anne-Béatrice Dufour
530 for useful discussions on a former version of the manuscript. We address many
531 thanks to F. Stephen Dobson and Nigel G. Yoccoz for their insightful com-
532 ments and thorough review of our manuscript. We finally wish to thank the
533 team developing the phylobase package for easing the handling of phylogenetic
534 comparative data.

Accepted manuscript

- 535 Abouheif, E., 1999. A method for testing the assumption of phylogenetic inde-
536 pendence in comparative data. *Evolutionary Ecology Research* 1, 895–909.
- 537 Ackerly, D. D., Schwilk, D. W., Webb, C. O., 2006. Niche evolution and adaptive
538 radiation: testing the order of trait divergence. *Ecology* 87, S50–S61.
- 539 Bauwens, D., Diaz-Uriarte, R., 1997. Covariation of life-history traits in lacertid
540 lizards: a comparative study. *The American Naturalist* 149, 91–111.
- 541 Bolker, B., Butler, M., Cowan, P., de Vienne, D., Jombart, T., Kembel, S.,
542 Orme, D., Paradis, E., Zwickl, D., 2007. phylobase: base package for phylo-
543 genetic structures and comparative data. R package version 0.3.
544 URL <http://phylobase.R-forge.R-project.org>
- 545 Bortolussi, N., Durand, E., Blum, M., François, O., 2006. apTreshape: statis-
546 tical analysis of phylogenetic tree shape. *Bioinformatics* 22, 363–364.
- 547 Chessel, D., Dufour, A.-B., Thioulouse, J., 2004. The ade4 package-I- one-table
548 methods. *R News* 4, 5–10.
- 549 Cheverud, J. M., Dow, M. M., 1985. An autocorrelation analysis of genetic
550 variation due to lineal fission in social groups of Rhesus macaques. *American*
551 *Journal of Physical Anthropology* 67, 113–121.
- 552 Cheverud, J. M., Dow, M. M., Leutenegger, W., 1985. The quantitative assess-
553 ment of phylogentic constraints in comparative analyses: sexual dimorphism
554 in body weights among primates. *Evolution* 39, 1335–1351.
- 555 Cliff, A. D., Ord, J. K., 1973. *Spatial autocorrelation*. Pion, London.
- 556 de Jong, P., Sprenger, C., van Veen, F., 1984. On extreme values of Moran's *I*
557 and Geary's *c*. *Geographical Analysis* 16, 17–24.
- 558 Desdevises, Y., Legendre, L., Azouzi, L., Morand, S., 2003. Quantifying phylo-
559 genetically structured environmental variation. *Evolution* 57 (11), 2647–2652.

- 560 Diniz-Filho, J. A. F., de Sant'Ana, C. E. R., Bini, L. M., 1998. An eigenvector
561 method for estimating phylogenetic inertia. *Evolution* 52, 1247–1262.
- 562 Dobson, F. S., 1985. The use of phylogeny in behavior and ecology. *Evolution*
563 39, 1384–1388.
- 564 Dolédec, S., Chessel, D., 1994. Co-inertia analysis: an alternative method for
565 studying species-environment relationships. *Freshwater Biology* 31, 277–294.
- 566 Dray, S., Chessel, D., Thioulouse, J., 2003a. Co-inertia analysis and the linking
567 of ecological data tables. *Ecology* 84 (11), 3078–3089.
- 568 Dray, S., Chessel, D., Thioulouse, J., 2003b. Procrustean co-inertia analysis for
569 the linking of multivariate datasets. *Ecoscience* 10, 110–119.
- 570 Dray, S., Dufour, A.-B., Chessel, D., 2007. The *ade4* package - II: Two-table
571 and *K*-table methods. *R News* 7, 47–54.
- 572 Dray, S., Legendre, P., Peres-Neto, P., 2006. Spatial modelling: a comprehensive
573 framework for principal coordinate analysis of neighbour matrices (PCNM).
574 *Ecological Modelling* 196, 483–493.
- 575 Dray, S., Saïd, S., Debias, F., 2008. Spatial ordination of vegetation data using
576 a generalization of Wartenberg's multivariate spatial correlation. *Journal of*
577 *Vegetation Science* 19, 45–56.
- 578 Emerson, B., Gillepsie, R., 2008. Phylogenetic analysis of community assembly
579 and structure over space and time. *Trends in Ecology and Evolution* 23, 619–
580 630.
- 581 Faraway, J. J., 2004. *Linear Models with R*. Chapman et Hall.
- 582 Felsenstein, J., 1985. Phylogenies and the comparative method. *The American*
583 *Naturalist* 125, 1–15.
- 584 Gaillard, J.-M., Pontier, D., Allainé, D., Lebreton, J. D., Trouvilliez, J., Clobert,
585 J., 1989. An analysis of demographic tactics in birds and mammals. *Oikos* 56,
586 59–76.

- 587 Garland, T., Bennett, A. F., Rezende, E. L., Aug 2005. Phylogenetic approaches
588 in comparative physiology. *J Exp Biol* 208 (Pt 16), 3015–3035.
589 URL <http://dx.doi.org/10.1242/jeb.01745>
- 590 Giannini, N. P., 2003. Canonical phylogenetic ordination. *Systematic Biology*
591 52, 684–695.
- 592 Gittleman, J. L., Kot, M., 1990. Adaptation: statistics and a null model for
593 estimating phylogenetic effects. *Systematic Zoology* 39, 227–241.
- 594 Gould, S. J., Lewontin, R. C., 1979. The spandrels of san marco and the pan-
595 glossian paradigm: a critique of the adaptationist program. *Proceedings of*
596 *the Royal Society of London, Series B* 205, 581–598.
- 597 Grafen, A., 1989. The phylogenetic regression. *Philosophical Transactions of the*
598 *Royal Society of London Series B - Biology* 326, 119–157.
- 599 Gregory, W. K., 1913. Convergence and applied phenomena in the mammalia.
600 Report of the British Association for the Advancement of Science IV, 525–526.
- 601 Griffith, D. A., 1996. Spatial autocorrelation and eigenfunctions of the geo-
602 graphic weights matrix accompanying geo-referenced data. *The Canadian Ge-*
603 *ographer* 40, 351–367.
- 604 Griffith, D. A., Peres-Neto, P., 2006. Spatial modeling in ecology: the flexibility
605 of eigenfunction spatial analyses. *Ecology* 87, 2603–2613.
- 606 Hansen, T. F., Martins, E. P., 1996. Translating between microevolutionary pro-
607 cess and macroevolutionary patterns: the correlation structure of interspecific
608 data. *Evolution* 50 (4), 1404–1417.
- 609 Harvey, P. H., Pagel, M., 1991. *The Comparative Method in Evolutionary Bi-*
610 *ology*. Oxford University Press.
- 611 Harvey, P. H., Purvis, A., Jun 1991. Comparative methods for explaining adap-
612 tations. *Nature* 351 (6328), 619–624.
613 URL <http://dx.doi.org/10.1038/351619a0>

- 614 Jombart, T., Devillard, S., Dufour, A.-B., Pontier, D., 2008. Revealing cryptic
615 spatial patterns in genetic variability by a new multivariate method. *Heredity*
616 101, 92–103.
- 617 Jombart, T., Dray, S., 2009. adephylo: exploratory analyses for the phylogenetic
618 comparative method.
619 URL <http://r-forge.r-project.org/projects/adephylo/>
- 620 Kirkpatrick, M., Slatkin, M., 1993. Searching for evolutionary patterns in the
621 shape of a phylogenetic tree. *Evolution* 47, 1171–1181.
- 622 Kraft, N. J. B., Cornwell, W. K., Webb, C. O., Ackerly, D. D., 2007. Trait
623 evolution, community assembly, and the phylogenetic structure of ecological
624 communities. *American Naturalist* 170, 271–283.
- 625 Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology*
626 74, 1659–1673.
- 627 Martins, E. P., 2000. Adaptation and the comparative method. *Trends in Ecology & Evolution* 15 (7), 296–299.
- 629 Martins, E. P., Diniz-Filho, J. A. F., Housworth, E. A., Jan 2002. Adaptive
630 constraints and the phylogenetic comparative method: a computer simulation
631 test. *Evolution* 56 (1), 1–13.
- 632 Martins, E. P., Hansen, T. F., 1997. Phylogenies and the comparative method: a
633 general approach to incorporating phylogenetic information into the analysis
634 of interspecific data. *The American Naturalist* 149 (4), 646–667.
- 635 Moran, P. A. P., 1948. The interpretation of statistical maps. *Journal of the*
636 *Royal Statistical Society, B* 10, 243–251.
- 637 Moran, P. A. P., 1950. Notes on continuous stochastic phenomena. *Biometrika*
638 37, 17–23.
- 639 Osborn, H. F., 1917. Heritage and habitus. *Science* 45, 660–661.

- 640 Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and
641 evolution in R language. *Bioinformatics* 20, 289–290.
- 642 Pavoine, S., Ollier, S., Pontier, D., Chessel, D., 2008. Testing for phylogenetic
643 signal in life history variable: Abouheif's test revisited. *Theoretical Popula-*
644 *tion Biology* 73, 79–91.
- 645 Peres-Neto, P., 2006. A unified strategy for estimating and controlling spatial,
646 temporal and phylogenetic autocorrelation in ecological models. *Oecologica*
647 *Brasiliensis* 10, 105–119.
- 648 Pontier, D., Gaillard, J.-M., Allainé, D., 1993. Maternal investment per offspring
649 and demographic tactics in placental mammals. *Oikos* 66, 424–430.
- 650 R Development Core Team, 2009. R: A Language and Environment for Statis-
651 tical Computing. R Foundation for Statistical Computing, Vienna, Austria,
652 ISBN 3-900051-07-0.
653 URL <http://www.R-project.org>
- 654 Rochet, M. J., Cornillon, P. A., Sabatier, R., Pontier, D., 2000. Comparative
655 analysis of phylogenetic and fishing effects in life history patterns of teleost
656 fishes. *Oikos* 91, 255–270.
- 657 Rohlf, F. J., 2001. Comparative methods for the analysis of continuous variables:
658 geometric interpretations. *Evolution* 55 (11), 2143–2160.
- 659 Rohlf, F. J., 2006. A comment on phylogenetic correction. *Evolution* 60, 1509–
660 1515.
- 661 Webb, C. O., Ackerly, D. D., McPeck, M. A., Donoghue, M. J., 2002. Phyloge-
662 nies and community ecology. *Annual Review of Ecology and Systematics* 33,
663 475–505.
- 664 Wiens, J., Graham, C., 2005. Niche conservatism: integrating evolution, ecology,
665 and conservation biology. *Annual Review of Ecology Evolution and system-*
666 *atics* 36, 519–539.

⁶⁶⁷ Yule, G. U., 1924. A mathematical theory of evolution based on the conclusions
⁶⁶⁸ of Dr J. C. Willis, F.R.S. Philosophical Transactions of the Royal Society of
⁶⁶⁹ London, A 213, 21–87.

Accepted manuscript

670 **6. Figure legends**

671 **Figure 1: pPCA of simulated data.** Example of simulated traits (light
 672 yellow) for different tree structures (A-E), and structures identified by pPCA.
 673 Global components and corresponding eigenvalues are indicated in red, while
 674 local components and their eigenvalues are displayed in blue. Positive and
 675 negative values of traits and PCs are indicated by black and white circles, re-
 676 spectively. Symbol size is proportional to absolute values. Simulated traits are
 677 labelled as G_i : i^{th} global structure, L_i : i^{th} local structure, and R_i : i^{th} ran-
 678 dom (*i.e.*, non-phylogenetically structured) trait. Principal components (PC)
 679 of pPCA are labelled as GPC1: first global PC (*i.e.*, associated with the largest
 680 positive eigenvalue). LPC1: first local PC (*i.e.*, associated with the largest neg-
 681 ative eigenvalue). (A) Symmetric tree; random noise added structures ('noise')
 682 equaled 0.5. (B) Random clustering of tips; noise=1. (C) Yule model; noise=0.5.
 683 (D) biased model; noise=0.75. (E) Assymmetric tree; noise=1.

684

685

686 **Figure 2: pPCA of lizards data.** (A) First global (red) and local (blue)
 687 principal components of the pPCA of lacertid lizards data, after removal of size
 688 effect. Inset barplot displays the corresponding eigenvalues. Positive and nega-
 689 tive scores are indicated by black and white circles, respectively. Symbol size is
 690 proportional to absolute values. Taxa are labelled as: *Podarcis h. atrata* ('Pa'),
 691 *P. h. hispanica* ('Ph'), *Lacerta lepida* ('Ll'), *L. monticola cantabrica* ('Lmca'),
 692 *L. m. cyreni* ('Lmcy'), *Podarcis h. hispanica* Asturias ('Phha'), *P. h. h. Sala-*
 693 *manca* ('Pha'), *P. bocagei* ('Pb'), *P. muralis* ('Pm'), *Acanthodactylus erythru-*
 694 *rus* ('Ae'), *Takydromus tachydromoides* ('Tt'), *T. septentrionalis* ('Ts'), *Lacerta*
 695 *vivipara* ('Lviv'), *L. agilis* ('La'), *L. schreiberi* ('Ls'), and *L. viridis* ('Lviv').
 696 (B) Loadings of the traits for the first global (red) and local (blue) principal
 697 components. Inset barplot displays the corresponding eigenvalues. $d=0.5$ in-
 698 dicates the mesh of the grid. Analysed traits are hatchling length (hatch.L)
 699 and mass (hatch.m), clutch frequency (clutch.F) and size (clutch.S), mean and

700 maximum female length (mean.L, max.L), mean female length and age at sex-
701 ual maturity (matur.L, age.mat). See text for a more detailed description of
702 analysed traits.

Accepted manuscript

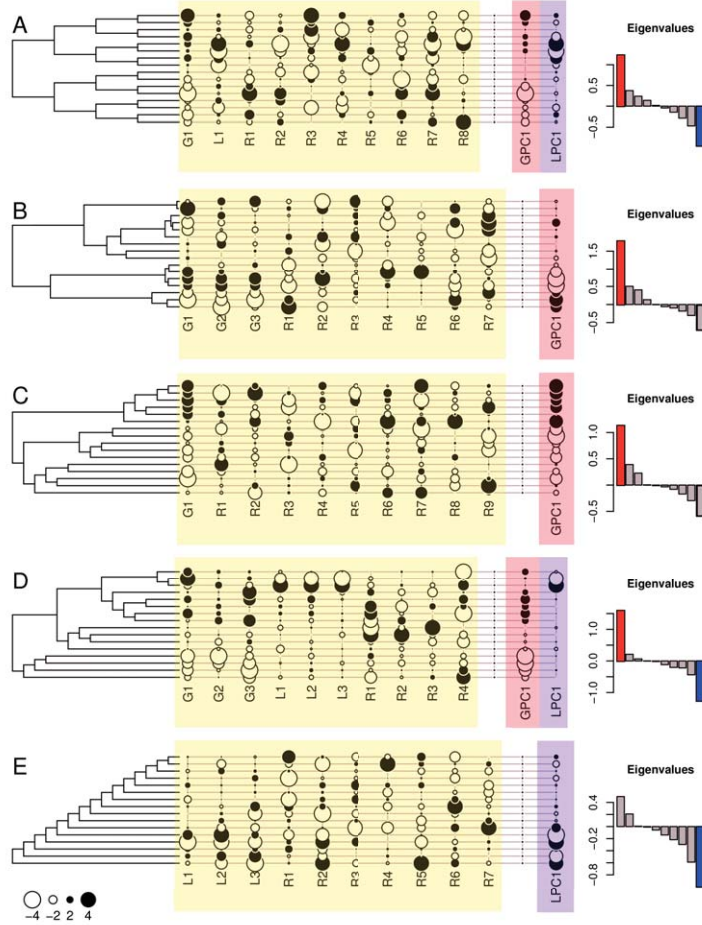
703 **7. Table legends**

704 **Table 1: parameters of the simulated data.** 200 datasets were simu-
705 lated for all combinations of these parameters. (1) expressed in number of tips.
706 (2) number of phylogenetically structured traits (global/local). (3) standard de-
707 viation of normal variates added to phylogenetically structured traits. (4) total
708 number of traits in the dataset, including phylogenetically structured traits.

Accepted manuscript

709

Figure 1:



710

713

Table 1:

Parameter	Values
Tree model	Symmetric; Random clustering; Yule; Biased; Asymmetric
Tree size ¹	16; 32; 128
Structures ²	1/0; 0/1; 3/0; 0/3; 1/1; 3/3
Random noise ³	0.5; 0.75; 1
Number of traits ⁴	12; 20; 50

Accepted manuscript

714 **Appendix A. Tables of the analysis of simulations**

715 This appendix presents two tables corresponding to the analysis of simulated
 716 data by the linear model described in sections 2.4 and 3.1.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
fac.tree ¹	4	6632	1658	2757.68	$< 2.2e^{-16}$
fac.strutype ¹	1	117621	117621	195635.60	$< 2.2e^{-16}$
ntips ³	1	11412	11412	18981.57	$< 2.2e^{-16}$
ntraits ⁴	1	31102.96	31102.96	51732.57	$< 2.2e^{-16}$
noise ⁵	1	8156.31	8156.31	13566.13	$< 2.2e^{-16}$
fac.nstruc ⁶	1	385.43	385.43	641.08	$< 2.2e^{-16}$
Residuals	215990	129858.80	0.60		

Table A.1: Analysis of variance of the model. Factors are preceded by 'fac'. (1) type of tree. (2) type of structure (global or local). (3) number of tips. (4) total number of traits. (5) number of structured traits (1 or 3). (6) standard deviation of the random noise added to the structured traits.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.8474	0.0053	158.99	$< 2e^{-16}$
fac.treeclust ¹	0.0431	0.0053	8.16	$< 2e^{-16}$
fac.treecomb ²	0.2980	0.0053	56.48	$< 2e^{-16}$
fac.treesym ³	0.4450	0.0053	84.34	$< 2e^{-16}$
fac.treeyule ⁴	0.0410	0.0053	7.78	$< 2e^{-16}$
fac.strutypelocal ⁵	-1.4759	0.0033	-442.31	$< 2e^{-16}$
ntips ⁶	0.0046	0.0000	137.77	$< 2e^{-16}$
ntraits ⁷	-0.0223	0.0001	-227.45	$< 2e^{-16}$
fac.nstruc3 ⁸	0.0845	0.0033	25.32	$< 2e^{-16}$
noise ⁹	-0.9520	0.0082	-116.47	$< 2e^{-16}$

Table A.2: Coefficients of the model. Factors are preceded by 'fac', followed by the levels. (1) trees obtained by random clustering of tips. (2) comb-like model (completely asymmetric trees). (3) completely symmetric trees. (4) Yule model. (5) local phylogenetic structure. (6) number of tips. (7) total number of traits. (8) number of structured traits (1 or 3). (9) standard deviation of the random noise added to the structured traits.