

László Zsolt Garamszegi
Editor

Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology

Concepts and Practice

Editor

László Zsolt Garamszegi
Department of Evolutionary Ecology
Estación Biológica de Doñana-CSIC
Seville
Spain

ISBN 978-3-662-43549-6 ISBN 978-3-662-43550-2 (eBook)

DOI 10.1007/978-3-662-43550-2

Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014944334

© Springer-Verlag Berlin Heidelberg 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Cover figure: Miklós Laczi

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

An Online Practical Material (OPM) for
this book is available at:
<http://www.mpcm-evolution.org>

Foreword

Humans and chimpanzees share around 99.8 % of their evolutionary histories, and nearly that much identity in the sequences of their genes, and yet most people are happy to lock the one up behind bars, throw it bits of food and watch it have sex, while protecting the other's rights with an exhausting array of laws, and granting it unique access to God's goodwill.

The surprise and even disorientation this image elicits goes to the heart of why comparative biology is fundamental to studying evolution and adaptation. Any two species that share that much of their histories are bound to be similar in many respects, as indeed humans and chimpanzees are. But we can also see in these two lineages the magisterial workings of natural selection, which has sculpted them in strikingly different ways since they parted company somewhere in Africa, perhaps 6–10 million years ago.

And that is the task that comparative methods must confront: how to separate what is shared from what is newly evolved, because it is the twists and turns of the latter that reveal how an evolving lineage has responded—and in most cases adapted—to its circumstances when compared to another evolving lineage whose circumstances differ. Remarkably, this simple truth has only become widely appreciated among evolutionary biologists in the last 25 years or so, and even today some recalcitrant troglodytes refuse to acknowledge it, or perhaps more charitably, the word has yet not filtered as far as their caves, or in one or two cases I know of, desert hideouts.

But happily, these people are an ever-shrinking minority whose eventual extinction no one will record on a Red List, and comparative methods are now widely used, having jumped from their historical origins in studying morphological and behavioural evolution, to virology, biogeography, ecology, gene and genome evolution, acoustics and perception, and to ever-finer analyses of the genetic basis of variation in organismic traits. Anthropologists use it, as do sociologists, and comparative methods have even been used to track down a deadly dentist who infected his patients with HIV.

But touting the hegemony of comparative methods is to adopt a narrow view. Their real power lies in combining information on the phylogenetic relationships among a group of species with data on the traits and behaviours of those species. Then, in combination with some hunches about the way evolution proceeds,

hunches that take the form of statistical models of evolution, comparative methods grant their users the power to reconstruct what the past was like, and then to test competing ideas about how that past gave rise to the present. When used properly, and with imagination, comparative methods make it possible to re-play the tape of evolution under different scenarios, giving credence to some scenarios over the others based on how well they explain the present.

It is to giving researchers the tools to do just this that László Zsolt Garamszegi's edited volume is devoted. It is a timely work as the past 10 years or so have witnessed useful ferment in the field as ever more fine-grained, large and detailed datasets and phylogenies become available, and statistically minded researchers have responded with ever more flexible methods. Garamszegi has organized the chapters so that readers are first prepared for and then gently eased into these various approaches. *Modern Phylogenetic Comparative Methods* will also have an Internet presence, allowing Garamszegi to keep users apprised of new developments and interpretations.

These are both useful strategies because different methods carry implicit controversies concerning the ways that evolution is assumed to proceed and what its patterns mean, and sometimes divide their followers. For instance, Brownian motion is still the dominant null-model of most investigations of continuously evolving traits, but where one camp adopts a catholic view, treating departures from a Brownian background as a statistical problem with multiple possible evolutionary causes, including varying rates of evolution and episodic bursts of change, others adopt a more Calvinist tone, insisting on 'stabilising selection' or attraction to a mysterious niche-optimum, as the cause—and their models admit no other view.

But these opposing views are—or should be—minor irritants in what is otherwise a scientific field in rude health, and one that in many respects should become the subject of young Ph.D. students in the philosophy of science. The old-guard Feyerabend-esque naysayers who cling to the desperate belief that science is just the province of who can shout loudest, and most effectively corrupt and coerce others, all in pursuit of their favourite myths, should take stock of the field of comparative biology: combative, and yes, often petty and self-serving, it has in these past 25 years or produced a steady, even if sometimes stumbling, triumph of the scientific method applied to this particular outpost of the field of evolution.

Preface

Many evolutionary biologists are concerned with the tremendous amount of diversity of species, and their phenotypes, that we can currently observe in nature, and one of the most challenging tasks is to find evolutionary explanations for such interspecific differences. The philosophy of comparing attributes of different species that have undergone different selection regimes has heavily dominated the way we think about evolution since the days of Darwin. Subsequently, a vast number of studies have interpreted inter-specific variations in species-specific traits in the light of parallel variations in certain environmental or social factors.

However, a significant paradigm shift in the application of comparative methods occurred in the 80s, mostly from the influential work of Felsenstein in 1985, when it was recognized that patterns of inter-specific variations cannot be interpreted without taking into account the underlying common descent of species that delineates certain evolutionary constraints and also leads to similarity between species' phenotypes. Statistically, the effect of phylogeny can be regarded as a confounding factor that violates assumptions about non-independence of the unit of analysis, and that potentially introduces spurious correlations across traits. The development of the independent contrast method caused a flourish in the literature, in which most comparative papers adopted the method as an efficient way to get rid of these unwanted effects of phylogeny.

Modern phylogenetic methods go well beyond of this simple task of achieving a simple statistical control for phylogeny. They rather treat the evolutionary history of species as an interesting phenomenon on its own, that allows tracing character states back through time along series of ancestral states. The real value of modern phylogenetic methods is, therefore, the capacity to examine biological diversity in the light of the phylogeny, a perspective that opens up horizons for making inferences about where, when, and how traits have changed over an evolutionary time scale. The phylogenetic comparative framework by today has grown to address a large number of fascinating questions about the correlated evolution of traits, phylogenetic signals in interspecific data, ancestral states, the mode of evolution, evolutionary rates, alternative evolutionary mechanisms, speciation and diversification and between-species interactions. Most of the methods are available for both discrete (like mating system) and continuous (like body size) characters, and have already started to spread into disciplines other than evolutionary biology (e.g., anthropology, genomics, linguistics, law, and sociology). Furthermore, given

that the phylogenetic comparative approach essentially offers a general framework for studying hierarchically structured data of any kind, one must assume that the exploitation of the underlying toolbox it provides has lagged behind its inherent potential.

In the current state of evolutionary biology, when phylogenies and interspecific data are accumulating at an enormous speed, it is becoming crucial that practitioners are armed with a diversity of comparative methods that help them make inferences from such data. However, overseeing these statistical tools becomes particularly challenging because of the richness and mathematical complexity of the relevant literature. In that situation, a secondary source, in which the primarily literature is brought into the attention of the user community in a consumable way may enhance the statistical integration of the discipline. This book has been assembled under this motivation, and aims at providing descriptions about the most recent phylogenetic methods for evolutionary biologist.

There are several other, very useful books on similar topics. Our book is not only novel because it revolves around the most recent developments, but also because this is an edited volume comprising contributions from several authors, each being expert on his/her respective subfield. This extensive collaborative project may, hence, offer a broad focus on a diverse array of topics and perspectives, which could not be covered efficiently by a handful of specialized authors only. The contributors to this book have been working in different fields of evolutionary biology for many years, under the ultimate aim of solving important biological questions. While fulfilling this enthralling mission we are often confronted with the task of broadening existing approaches and exploring new advances in the comparative study of diverse taxonomic systems and communities. In the light of this scientific background and our experiences obtained during teaching and in various statistical courses, we felt that a textbook was needed to provide a broad overview in the field of phylogeny-based evolutionary biology to our students and less experienced colleagues. Therefore, we wanted to compile a wide range of different perspectives and practices in the phylogenetic comparative method: from an introduction to the topic, through the diversity of statistical designs that can powerfully incorporate phylogenetic information, to more enhanced applications that offer studying evolutionary mechanisms. We must note, however, that due to various constraints, we were not able to review the entire literature that might be relevant. We hope that we will be able to adjust for such a shortcoming in the future.

Another extra value we offer is the accompanying online resource (available at <http://www.mpcm-evolution.org>), where we wish to post and permanently update practical materials to help embed methods into practice. As Online Practical Material (OPM), we will provide tutorials, example files and the underlying statistical scripts, with which the users will be able to apply the presented methods to their own data and scientific questions. New approaches appear like mushrooms in the forest, with some of them being implemented in practice at a rapid pace. It is also becoming very common that more than one statistical approach is available for the same evolutionary problem. Being able to select from alternative

approaches also allows researchers to gain better comprehension of biological diversity present in natural systems. Therefore, it is vital to keep these methodologies updated and accessible for the broad user community, and the attributes of an online surface can be fruitfully exploited in that direction.

Statistics is not without mathematics, and the derivation of certain formulas is unavoidable for the appropriate argumentations. In this book, we constrain ourselves to present only the key mathematical formulas that are necessary for understanding the philosophy of different approaches. Doing so should make the reasoning accessible to a broad readership. For those who are interested in the mathematical details, we provide pointers to the primary literature. And for those not interested in the mathematical details, do not fear. The equations herein should be treated as resources if you decide to incorporate these methods into your research.

The quality of this book in large part resides on inputs from several expert reviewers, who provided valuable comments on earlier versions of chapter manuscripts during their review phase. Along that line, we are extremely grateful to David Bapst, Roger Benson, Kierstin K. Catlett, Natalie Cooper, Thomas Currie, Pierre de Villemereuil, Joe Felsenstein, Sive Finlay, Rob Freckleton, Jesualdo Fuentes-Gonzalez, Alejandro Gonzalez-Voyer, Tatiana Giraud, Alan Grafen, Randi Griffin, Jarrod Hadfield, Thomas Hansen, Lam Si Tung Ho, Elisabeth Housworth, Antony Ives, Jason Kamilar, Tom Kraft, Oriol Lapiedra, Jessica Light, Graeme Lloyd, Rafael Maia, Emilia Martins, Nick Matzke, Mark A. McPeck, Magdalena N. Muchlinski, Shinichi Nakagawa, Chris Nasrallah, Charles L. Nunn, Christopher E. Oufiero, Emmanuel Paradis, Sandrine Pavoine, Matt Pennell, Pedro Peres-Neto, Samantha Price, Liam Revell, Aaron Sandel, Holger Schielzeth, William Shipley, Graham Slater, Jeet Sukumaran, Matthew Symonds, Gavin Thomas, Chris Venditti, Jamie Caroline Winternitz and Derrick Zwickl for their constructive assistance in the evaluation process.

Although the list of authors appears male biased, this is completely unintentional. Female scientists were also invited to write chapters, but regretfully they could not make contribution due to various and completely understandable reasons.

Seville, Spain

László Zsolt Garamszegi

Contents

Part I Introduction

- 1 An Introduction to the Phylogenetic Comparative Method** 3
Emmanuel Paradis
- 2 Working with the Tree of Life in Comparative Studies:
How to Build and Tailor Phylogenies to Interspecific Datasets . . .** 19
László Zsolt Garamszegi and Alejandro Gonzalez-Voyer
- 3 An Introduction to Supertree Construction (and Partitioned
Phylogenetic Analyses) with a View Toward the Distinction
Between Gene Trees and Species Trees** 49
Olaf R. P. Bininda-Emonds
- 4 Graphical Methods for Visualizing Comparative Data
on Phylogenies** 77
Liam J. Revell
- 5 A Primer on Phylogenetic Generalised Least Squares** 105
Matthew R. E. Symonds and Simon P. Blomberg
- 6 Statistical Issues and Assumptions of Phylogenetic
Generalized Least Squares** 131
Roger Mundry

Part II Handling Phylogenies in Different Statistical Designs

- 7 Uncertainties Due to Within-Species Variation
in Comparative Studies: Measurement Errors
and Statistical Weights** 157
László Zsolt Garamszegi

8	An Introduction to Phylogenetic Path Analysis	201
	Alejandro Gonzalez-Voyer and Achaz von Hardenberg	
9	Phylogenetic Regression for Binary Dependent Variables.	231
	Anthony R. Ives and Theodore Garland Jr.	
10	Keeping Yourself Updated: Bayesian Approaches in Phylogenetic Comparative Methods with a Focus on Markov Chain Models of Discrete Character Evolution	263
	Thomas E. Currie and Andrew Meade	
11	General Quantitative Genetic Methods for Comparative Biology.	287
	Pierre de Villemereuil and Shinichi Nakagawa	
12	Multimodel-Inference in Comparative Analyses.	305
	László Zsolt Garamszegi and Roger Mundry	
 Part III Specific Models for Studying Evolutionary Mechanisms		
13	Simulation of Phylogenetic Data	335
	Emmanuel Paradis	
14	Use and Misuse of Comparative Methods in the Study of Adaptation	351
	Thomas F. Hansen	
15	Modelling Stabilizing Selection: The Attraction of Ornstein–Uhlenbeck Models	381
	Brian C. O’Meara and Jeremy M. Beaulieu	
16	Hidden Markov Models for Studying the Evolution of Binary Morphological Characters	395
	Jeremy M. Beaulieu and Brian C. O’Meara	
17	Detecting Phenotypic Selection by Approximate Bayesian Computation in Phylogenetic Comparative Methods	409
	Nobuyuki Kutsukake and Hideki Innan	
18	Phylogenetic Comparative Methods for Studying Clade-Wide Convergence.	425
	D. Luke Mahler and Travis Ingram	

19	Metrics and Models of Community Phylogenetics	451
	William D. Pearse, Andy Purvis, Jeannine Cavender-Bares and Matthew R. Helmus	
20	Event-Based Cophylogenetic Comparative Analysis	465
	Michael Charleston and Ran Libeskind-Hadas	
21	Phylogenetic Prediction to Identify “Evolutionary Singularities”	481
	Charles L. Nunn and Li Zhu	
22	Preparing Paleontological Datasets for Phylogenetic Comparative Methods	515
	David W. Bapst	
	Index	545

Part I

Introduction

Chapter 1

An Introduction to the Phylogenetic Comparative Method

Emmanuel Paradis

Abstract The phylogenetic comparative method (PCM) has an important place in evolutionary biology. This chapter aims at giving an overview on some selected topics. We first review briefly some important historical milestones including some early contributions and the relationships of comparative methods with phylogenetics. Some fundamental points on statistical inference, adaptation, and causality are then discussed. We also discuss briefly the application of the PCM to anthropology and conclude with some perspectives on its future development and applications.

1.1 Introduction

A comparison of apples and oranges occurs when two items or groups of items are compared that cannot be practically compared. ... However, apples are actually more closely related to pears (both are rosaceae) than to oranges.

—Wikipedia¹

The phylogenetic comparative method has undoubtedly been one of the most important phenomena of evolutionary biology during the last few decades. Comparative methods exist in many fields such as anthropology (Bock 1966), law (Kieckbaev 2003), linguistics (Forster et al. 1998), and evolutionary biology (Harvey and Pagel 1991). The concepts and uses of these different comparative methods vary widely. Since the present book is specifically concerned with biological evolution, it is thus useful to define our subject.

¹ http://en.wikipedia.org/wiki/Apples_and_oranges

E. Paradis (✉)

Institut de Recherche pour le Développement, Montpellier, France
e-mail: emmanuel.paradis@ird.fr

We may define the *comparative method* as an analytical approach based on the comparison of different objects with the aim to elucidate the mechanisms at the origin of their diversity. From this, we can define the *phylogenetic comparative method* as the analytical study of species, populations, and individuals in a historical framework with the aim to elucidate the mechanisms at the origin of the diversity of life.

It is important to note that the phylogenetic comparative method (PCM) is distinct from but not independent of *phylogenetics*, the study and reconstruction of the historical relationships among species. For instance, in linguistics or in anthropology, the goal of the comparative method is the historical reconstruction of spoken languages or of human cultures (e.g., Forster et al. 1998, with some historical references therein; see Sect. 1.7 below).

The goal of this chapter is to give a general introduction to the PCM by examining some topics. The next section presents the main historical milestones of phylogenetics and the comparative method—since both have been tightly linked through their history. The following sections give some essential elements on statistical inference of evolutionary processes with comparative data. The last two sections aim to put the PCM in a broader perspective by looking at its relationships with anthropology and speculating about some of its current advances and its future.

1.2 History of Phylogenetics and the Comparative Method

1.2.1 Early Developments

In the nineteenth century, trees were essential graphical tools for the development of evolutionary ideas. Lamarck (1809) used a downward-growing tree to represent the relationships among the main groups of animals with a caption indicating that this “table displays the origin of the different animals.” History was central in Lamarck’s argumentation: “A strong reason prevents us to identify the changes that have successively diversified the animals as we know them today: we have never witnessed these changes.” This could be taken as a manifesto of today’s comparative method in evolutionary biology.

Cuvier was Lamarck’s great rival and strong opponent to the idea of evolution. However, Cuvier acknowledged that species are more or less closely related so that they can be classified in a hierarchical system and that different characters of these species are relevant at different levels, especially through his gradual characters (*caractères gradués*, Cuvier 1798). In spite of his backward ideas on fixism (Laurent 1986), Cuvier had a profound impact on comparative anatomy through the numerous illustrations and drawings included in his books—which could appear in a modern textbook on evolution after updating the captions.

It has been widely appreciated that Darwin (1859) used a phylogenetic tree as the only figure in the *Origin of Species*. He also used comparative data to support several of his points; for instance, “Genera which are polymorphic in one country seem to be, with some few exceptions, polymorphic in other countries, and likewise, judging from Brachiopod shells, at former periods of time.” Thus, Darwin characterized the patterns of diversity in space and in time and also used these facts to infer the processes of diversification of species: “...the larger genera also tend to break up into smaller genera. And thus, the forms of life throughout the universe become divided into groups subordinate to groups.”

The end of the nineteenth century has witnessed the wide acceptance of the idea of evolution, particularly with the contributions of Haeckel, the father of phylogenetics: “For the purpose of constructing a hypothetical genealogical tree of the Radiolaria, as of all other organisms, three sources of information are open to us, viz., palæontology, comparative ontogeny, and comparative anatomy.” (Haeckel 1887).

During the first half of the twentieth century, phylogenetics and biological comparative studies have followed separate paths. The discovery of the physical support of heredity (genes, chromosomes, and later DNA) led scientists to focus their interest on the genetic mechanisms of evolution. Fisher (1930) certainly best illustrates this change of paradigm where history was less important than previously thought: “For mutations to dominate the trend of evolution it is thus necessary to postulate mutation rates immensely greater than those which are known to occur...” For Fisher, mutations could not explain evolutionary novelties and we should rather look at other evolutionary forces such as selection or population structure to explain the diversity of life. At the same time, phylogenetics made fundamental contributions to evolutionary thinking. Paleontologists integrated phylogenetic ideas, mainly because of the historical nature of their data (Simpson 1944). Phylogenetic trees became the analytical tool of a school of systematists (cladistics), leading to the first numerical treatments of phylogenies (see a historical account in Felsenstein 2004).

1.2.2 Modern Developments

The late 1960s have witnessed some crucial turns. The development of statistical methods to reconstruct phylogenies from genetic data was a major step accomplished by Cavalli-Sforza and Edwards (1967). Because the approach they developed was statistical, it was possible to extend it to other kinds of data such as continuous characters. This next step was achieved by Felsenstein (1973) who proposed a method to calculate the likelihood of a tree for a set of continuous traits. The significance of this work was not obvious until the same author published a related method to calculate the phylogenetically independent contrasts (PICs), a major difference being that the calculations under this new method could be done with a hand calculator (Felsenstein 1985).

Until the 1970s, comparative biology developed its statistical tools independently of phylogenetic or historical ideas. Comparative data from n species used to be analyzed with standard statistical methods, assuming that they were n independent observations. This separation between comparative biology and the historical dimension of biological evolution seems surprising when considering that the idea of evolution, and particularly adaptation, was at the heart of most comparative studies (Clutton-Brock and Harvey 1979).

The 1980s can be seen as the golden age of PCMs when a great variety of methods were published (reviewed by Pagel and Harvey 1988). Two important papers were published in the same year: Cheverud et al. (1985), who proposed an approach based on auto-regression including the possibility to account for intra-specific variation, and Felsenstein (1985), already cited. A few years later, Grafen (1989), in a very rich and dense paper, proposed the use of generalized least squares (GLS) to derive a method now widely known as the phylogenetic generalized least squares (PGLS). Gittleman and Kot (1990) further developed the use of auto-correlation functions to assess phylogenetic signal in diverse settings, including using taxonomic levels when a phylogeny is not available. Nevertheless, the power of these methods to infer evolutionary models and parameters was not yet fully acknowledged, and the view that phylogeny was a confounding effect in comparative analyses still prevailed: “Confounding effects of phylogeny and other variables may lurk behind any comparative relationship, and they must be removed or controlled prior to considering adaptive arguments.” (Pagel and Harvey 1988).

During the 1990s, the developments of the previous decade were confirmed and strengthened. Lynch (1991) made a link between models of quantitative genetics and phylogenetics. He developed a method to partition the variance of a trait into an environmental and a phylogenetic component. Importantly, the same decomposition can be done for the covariance between two traits, thus providing a formal way to quantify the historical component of the link between two characters. New methods were proposed for the analysis of discrete traits (Pagel 1994; Grafen and Ridley 1997). In two important papers, Hansen and Martins (1996) and Martins and Hansen (1997) showed how GLS can be used to address evolutionary questions beyond the basic Brownian motion model.

In the 2000s, some efforts were given to issues left temporarily aside such as fitting more complicated models combining continuous and discrete traits (Paradis and Claude 2002; Felsenstein 2005; Hadfield and Nakagawa 2010) or combining interspecific and intraspecific data (Felsenstein 2008; Garamszegi and Møller 2010; Stone et al. 2011; see Chap. 7). At the end of the decade, three papers by Revell (2009), Jombart et al. (2010), and Pavoine et al. (2010) defined a general framework for multivariate statistical analyses in a phylogenetic context. The concept of phylogenetic signal has also attracted significant interest with the aim of clarifying previous ideas on phylogenetic confounding effect (Blomberg et al. 2003; Ollier et al. 2006; Pavoine et al. 2008; Münkemüller et al. 2012). At the same time, PCMs have achieved maturity with some generalizations such as the development of a Brownian model with variable parameters (O’Meara et al. 2006)

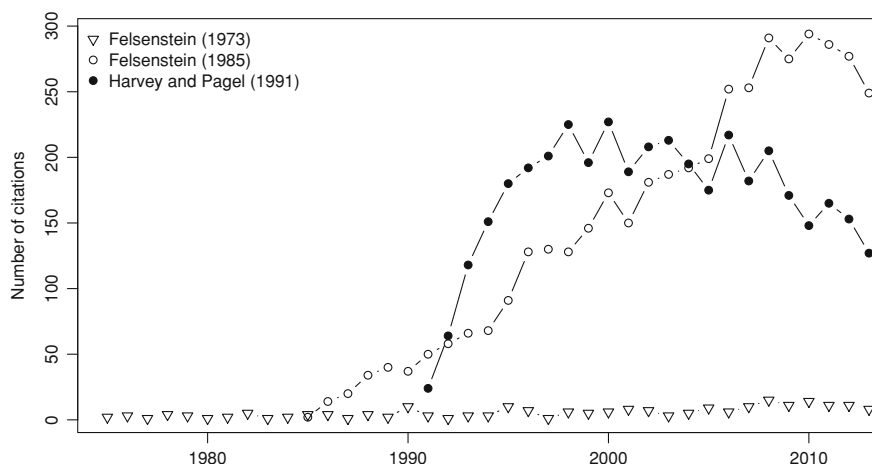


Fig. 1.1 The annual number of citations of three major contributions to the phylogenetic comparative method (*Source* Web of Science)

or the use of sophisticated model-fitting techniques such as Monte Carlo Markov chains (MCMC) to fit complicated models (Pagel et al. 2004; Pagel and Meade 2006; Hadfield and Nakagawa 2010).

Remarkable progress in phylogenetics also contributed to PCMs, particularly with the publication of more and more phylogenies, some of them being complete over a very large number of species (Bininda-Emonds et al. 2007; Smith et al. 2011; Jetz et al. 2012; see Chap. 3). Some methods have been developed to combine different sources of phylogenetic information in order to build trees for comparative analyses (Kuhn et al. 2011; Eastman et al. 2013; Thomas et al. 2013; see Chap. 2).

Figure 1.1 gives the number of citations of two earlier papers over the years together with another major contribution to the development of PCMs. After almost three decades, the range of applications of PCMs has grown to reach all branches of biological science: 6,533 citations of Felsenstein (1985) or Harvey and Pagel (1991) are found in 771 periodical titles. The PCM, through the development of a wide range of analytical tools, has contributed insights into many questions on evolution and the diversity of life.

1.3 The Covariance Structure of Comparative Data

A central issue with comparative data is the non-independence of observations. A similar problem is found in other fields such as geography (Cliff and Ord 1981), climatology (Tiao et al. 1990), ecology (Legendre 1993), or medical research (Houwing-Duistermaat et al. 1998).

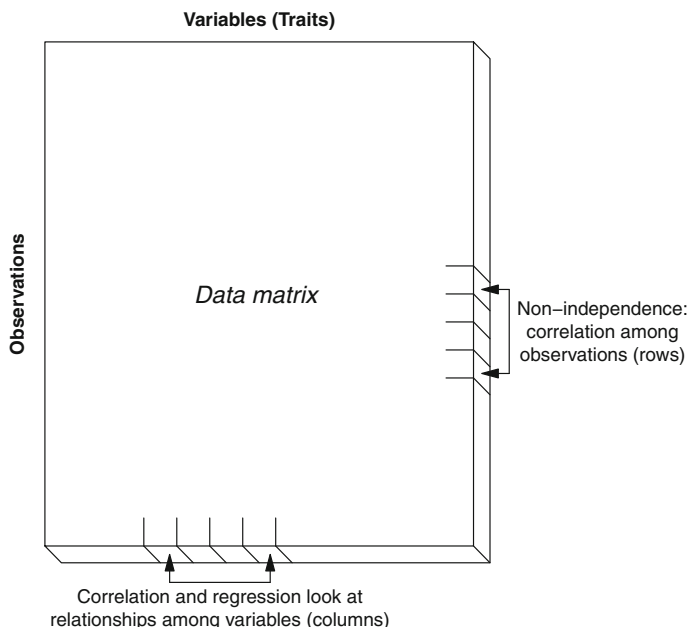


Fig. 1.2 A general depiction of a data set

In a very general way, a data set can be arranged in a matrix where the rows are the observations (individuals, populations, species, sequences, cells, etc.) and the columns are variables (size, area, nucleotide sites, RNA transcription levels, etc.). Data analyses seek for relationships among the columns of this matrix, and common statistical methods assume that the rows are independent observations; in other words, the values observed on a given row are not affected by the values at others (Fig. 1.2).

To statistically handle non-independence of observations, a general approach is to assume that two observations (rows), say i and j , are related through a covariance parameter denoted as σ_{ij}^2 . This parameter specifies the strength of the relation between the values of the same variable (column) observed for these two observations. The way this parameter enters in the analyses depends on the method used, the kinds of variables, and the question asked. The covariance parameters are usually arranged in a symmetric matrix with the diagonal elements equal to the variances and the off-diagonal elements being the covariances (see Chap. 5). This matrix has n rows and n columns and so contains $n(n-1)/2$ off-diagonal elements.

There are many ways to define the values of σ_{ij}^2 : they may be all equal or not, they may follow a specific distribution or may be related to another variable, they may be fixed or estimated from the data, etc. For instance, with spatial data, it is common to use a covariance function related to geographical distance.

In the case of comparative data from several species, it is possible to calculate a priori the covariances among species traits if we know how these traits have evolved. For instance, if we assume that a trait has evolved under a Brownian motion (BM) model, these covariances can be calculated from the phylogenetic tree linking these species without observing the trait itself. The PIC and PGLS methods were directly derived from this assumption. Both methods are identical though they are computationally very different (Blomberg et al. 2012). PGLSs directly use the covariances by calculating a correlation matrix among observations, which is simply a covariance matrix scaled to have values between -1 and 1 (see Chap. 5).

1.4 Statistical Inference

The covariance matrix is at the core of most PCMs. Fortunately, it is possible to calculate it for other models of trait evolution, in particular for the Ornstein–Uhlenbeck (OU) model which is appropriate to model evolution of traits under constraints (see Chaps. 13, 14 and 15). Using this and other models, it is thus possible to relax the assumption underlying the BM model. This feature of PCMs allows us to go beyond the paradigm that phylogeny is a confounding effect that must simply be corrected (see Rohlf 2006).

We can illustrate this point with a small simulation exercise. Taking the phylogeny in Fig. 1.3, we simulate two independent traits that evolve according to either a BM or an OU model. In this second model, the traits are constrained to evolve toward an optimal value with a strength controlled by the parameter denoted as α . The simulated traits were analyzed with two methods: a standard regression (assuming the species are independent) and a regression using the PICs calculated with the original phylogeny (which was thus assumed to be perfectly known). Table 1.1 shows the estimated rejection rates for both methods. Since both traits are independent, we expect these rates to be close to 5 %. The PIC-based analyses gave the correct answer with the data simulated from a BM model or from an OU model with a small value of α . On the other hand, for the large values of α , the PIC-based analyses had a high type I error rate, whereas the standard regression had a rejection rate close to 5 %.

Figure 1.4 shows the correlation matrices among the 20 leaves of the tree under different models of trait evolution. These matrices would be used in PGLS analyses (see Chap. 5). This shows clearly that an OU model with small α is close to a Brownian motion one, whereas when α is large, the observations are expected to be almost independent.

The critical point in a PCM-based analysis is to use the correct correlation structure among observations. The more distant the assumed correlation structure from the real one, the more biased the analysis will be. This property explains the statement that “in a comparative analysis a wrong phylogeny is better than no phylogeny at all” (Losos 1994; Martins 1996). Indeed, if the traits evolved on a

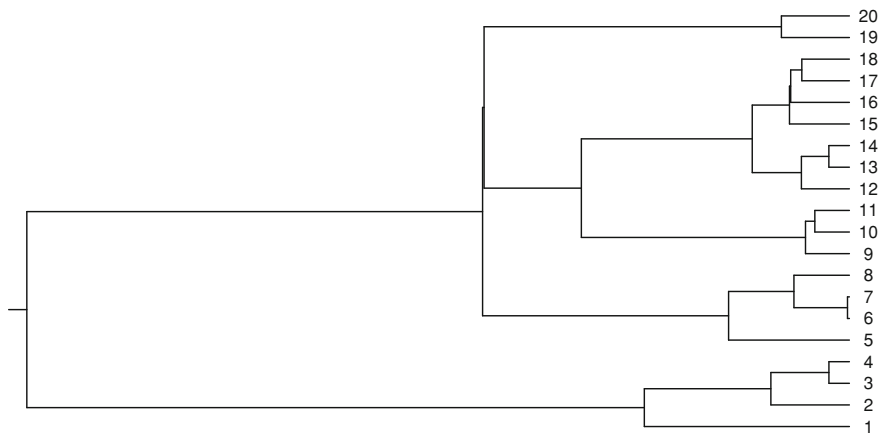


Fig. 1.3 A simulated phylogeny with $n = 20$

Table 1.1 Rejection rate of the test of correlation between two independent traits simulated on the phylogeny in Fig. 1.3 using the model indicated in the table

Simulated model	α	Standard regression	PIC regression
Brownian motion		0.396	0.051
Ornstein–Uhlenbeck	1	0.223	0.056
	10	0.065	0.120
	100	0.051	0.343

Simulations were replicated 10,000 times

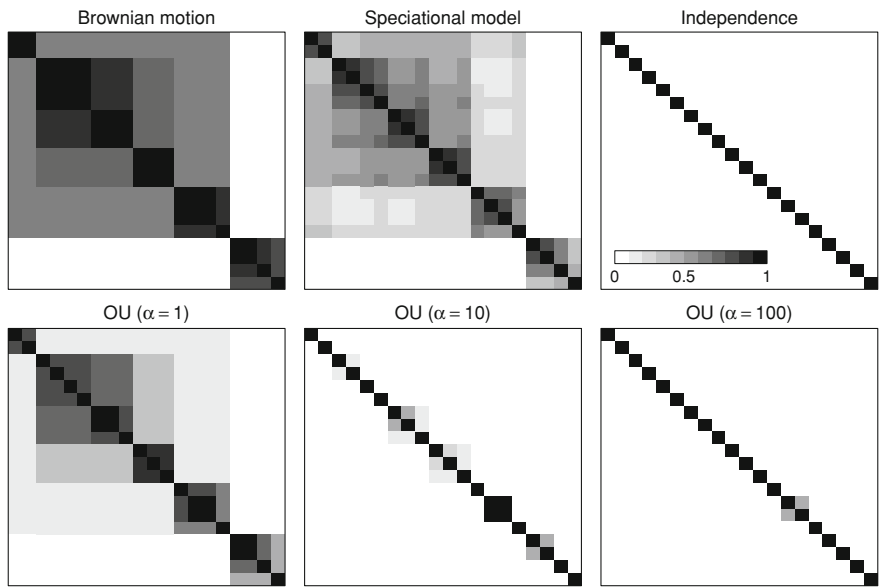


Fig. 1.4 Correlation matrices among the 20 tips of the tree in Fig. 1.3. The speciational model is one where change occurs only after a speciation event

phylogeny and another phylogeny is used for data analysis, the latter will result in a correlation structure closer to the correct one than assuming no correlation at all (i.e., independence of observations).

From the point of view of data analysis, one problem often encountered in published studies is that the phylogenetic correlation structure of the data is usually not assessed. This certainly comes from the view traditionally defended by most authors that only the phylogenetically controlled analyses are relevant. We now know that this can lead to wrong inference. This seems relatively easy to fix with tests of phylogenetic signal and model selection with information criteria such as the AIC (see Chap. 5). Remarkably, when different correlation structures are compared with real data, the Brownian motion model is rejected against more complex models such as the OU one (Whitney et al. 2011; Lapedra et al. 2013).

1.5 Inferring Adaptation

Perhaps because of its history, the comparative method is most often used to infer adaptation. However, the scope and power of the PCM to reveal adaptation have been criticized several times over the years (e.g., Leroi et al. 1994; Martins 2000; Grandcolas et al. 2011). Such criticism is not really surprising: It has been discussed since a long time ago that characterizing the adaptative nature of a trait is a complicated endeavor (Bock 1959). Even the characterization of adaptation in viruses, which are far simpler than the organisms studied by most evolutionists, appears to be an arduous task (Pepin et al. 2010). The use of traits such as “habitat use” or “environment” with PCMs has been questioned because the analysis of such variables in a phylogenetic framework is meaningless (Grandcolas et al. 2011). On the other hand, it is hard to not consider these variables in evolutionary models since the assessment of the adaptive value of a trait cannot be separated from extrinsic variables such as habitat, resources, or climate (Bock 2003; Losos 2011; Watt 2013).

Some recent developments in PCMs provide a solution to the limitations underlined by the critiques cited above. As we have seen in the previous sections, the PCM does not simply aim at correcting for phylogenetic dependence or inferring repeated evolution of the same trait in different lineages, but rather to provide tools to analyze comparative data in a historical framework, and this includes fitting complex models of trait evolution that can handle various complications of the study design. For instance, some methods make possible to analyze several traits that evolve under different models: Bartoszek et al. (2012) developed a multi-trait model where traits can evolve following different processes of BM or OU.

One situation illustrated by Losos (2011) is the one of “incomplete convergence.” Convergence toward similar phenotypes among distantly related species is often viewed as evidence for adaptation. However, adaptive evolution can proceed in different ways in different groups, and the patterns thus produced are likely to be

masked or obscured by other variables (such as the taxonomical background). Losos (2011) gives the example of the head shape of lizards which is mainly related to phylogenetic relatedness; however, within distinct clades, some species evolved independently toward herbivory and share some similarities, but the convergence is incomplete as they retain their respective phylogenetic background. In this case, a standard comparative analysis will likely fail to characterize the limited convergence among species affected by similar selective forces. On the other hand, statistical methods in a historical framework, including models of trait evolution, are helpful to characterize such patterns of adaptation. The use of these and other recently developed models of trait evolution may help to solve this “paradox” of using non-heritable traits in comparative analyses.

1.6 Inferring Causality

Correlation is not causality, and PCMs do not escape this reality. In spite of the importance of causality in evolutionary theory (see Watt 2013, for a recent view), the application of PCMs is generally oblivious of this point. This has led to some debate about the applicability of the PCM in order to identify evolutionary mechanisms. The vast majority of publications do not elaborate much on correlation and causality in their predictions: A simple linear correlation is usually derived from the hypotheses under test.

In general statistical inference, the causal relationship between two variables (say x and y) can be assessed if one of them can be controlled and then used as a predictor in data analyses. With comparative data, x and y cannot be controlled: They are evolving traits (or intrinsic variables) which are measured “on the species” (or they are extrinsic variables, like habitat, which cannot generally be controlled). Therefore, in the situation of a PCM with two variables, it is not possible to determine which regression (x on y or y on x) best describes the data. In other words, we cannot infer the causal relationship between these two variables.

When three or more uncontrolled variables are analyzed, it is possible to assess alternative causal relationships among them with a method known as path analysis (Freedman 2009). This method considers explicitly the causal relationships among variables under alternative hypotheses. A causal relationship can be expressed as “the variation in y is caused by the variation in the value of x ” and has the statistical consequence that the regression of y on x is meaningful. Under a given hypothesis of the causal relationships among variables, some regressions are meaningful, while others are not. Using a procedure called the d-sep, it is possible to test which hypothesis best describes the data (Shipley 2013). Santos and Cannatella (2011) and von Hardenberg and Gonzalez-Voyer (2013) proposed to extend the framework of path analysis to PCMs (see Chap. 8). In a traditional path analysis, the regressions are done assuming independence of observations. Therefore, it is straightforward to generalize this method to cases where the observations are not independent, using tools such as (P)GLS.

Causality in general, and in evolution in particular, is fundamental, but this is a difficult concept to apply in practice. Hopefully, future applications of PCMs will help to progress on this issue.

1.7 Phylogenetic Comparative Method and Anthropology

As mentioned above, the present book focuses on the uses of the comparative method in evolutionary biology. One reason for this restriction is that other scientists do not see the comparative approach in the same way than evolutionists do. For instance, Bock (1966) described the comparative method as follows (*italics as original*):

It should be recalled at the outset that the primary objective of users of the comparative method is historical reconstruction. *What* history or *whose* history is by no means clear in the nineteenth century literature, and this question has hardly been resolved in recent controversy.

Thus, what anthropologists call “comparative method” seems close to what evolutionists call “phylogenetics.” Mace and Pagel (1994) revisited this issue by introducing a phylogenetic approach to anthropology directly inspired from evolutionary biology. Considering the links between comparative biology and phylogenetics, their message does not differ radically from the one formulated 28 years before by Bock.

In practice, the application of PCMs in anthropology differs substantially compared to evolutionary biology. A remarkable difference is that with anthropological data, the historical sequence of changes in traits (cultures, political systems, etc.) is often recorded—at least more often than in biology. For instance, Lindenfors et al. (2011) studied changes in political systems in the world between 1800 and 2008. Using several variables, they built a score ranging between -10 (total autocracy) and $+10$ (full democracy) and measured transitions among these different scores. They showed that most political changes occurred from autocratic systems (with a peak around -6) toward democratic ones (with a peak around $+8$). However, one interesting point about this study is their comment with respect to the historical dimension of the problem:

A reconstruction of democracy as a political system on a language phylogeny would almost certainly indicate democracy as the ancestral state for large sections of the phylogeny. However, since we have exact information of all transitions, we know this not to be true.

In biological terms, there is a trend (or directional evolution) from autocracy toward democracy so that the second system is the most widespread today among countries. If we ignore this historical trend, we would make wrong inference. Here also, we see that using the wrong model of evolution can be misleading. Similar situations can be found with evolutionary data; for instance, if a trend exists in the evolution of a trait (say, increase in body size), then ancestral inference will likely

be misleading if this trend is not taken into account (Grafen 1989). Anthropological data have other peculiarities, like the ubiquity of horizontal transfers (Borgerhoff Mulder et al. 2006), so that a full comparison between cultural and biological evolution would require a much longer discussion.

1.8 The Future of the Phylogenetic Comparative Method

The comparative method in biology has evolved through several centuries to reach its present status. Today, PCMs have attained a level of maturity and sophistication that the readers can appreciate in the chapters of this book. The directions of future progress are certainly multiple.

We have seen the importance of the evolutionary models in statistical inference with the PCM. Some researchers currently explore the possibility to analyze complex models with several variables and an explicit formulation of the relationships among them. Hadjipantelis et al. (2013) analyzed an evolutionary model of “function-valued traits” by combining dimensionality reduction and a “bagging” (bootstrap aggregating) procedure. Complex relational models fitted with structural equation models seem also a very promising approach for future comparative analyses (Chap. 8).

It has not been widely appreciated that some PCMs link a model of microevolution (random evolution through genetic drift or stabilizing selection) with the patterns of interspecific variation in a trait and thus with macroevolution. Further works in this direction will likely lead to some interesting investigations on evolutionary mechanisms.

During many decades, the fossil record has been considered as the only source of information about evolutionary change. This paradigm has been broken by two steps forward in evolutionary biology: phylogenetic methods which try to reconstruct the past from the present and observations of real evolutionary changes over recent years such as the spread of resistance alleles in pathogens. However, the divorce between paleobiology and PCMs does not seem natural, and several researchers try to reconcile them (Pennell and Harmon 2013). This task will surely be very difficult but considering the many contributions of fossils to evolutionary biology, this is worth the effort (Chap. 22).

A scientific discipline is sometimes judged by how it contributes to everyone's well-being. The PCM may well move very positively in this direction as comparative analyses could have concrete applications. Spreitzer et al. (2005) used simple comparisons between different plant and alga species to create new enzymes involved in photosynthesis by targeted mutations on “phylogenetic residues.” The resulting enzyme is one with original characteristics, a kind of “phylogenetic chimaera.” Such an approach of “phylogenetic engineering” may be promising to design new proteins or even new organisms based on predictions from evolutionary phylogenetic models.

Yan et al. (2012) used a phylogenetic analysis of a number of bacteria in order to propose cocktails of probiotic bacteria to reduce pathogens in food. Their approach is based on an investigation of a protein, MazF, which has an antimicrobial activity, and for which they propose an engineered variant. The phylogeny of the studied bacteria was instrumental in designing this new protein. The combination of molecular structure approaches with phylogenetic comparative analyses seems a promising venue to develop a variety of new molecules with desired properties.

Rich of its long history, the PCM seems to have a bright future both for addressing fundamental questions and for delivering applications.

Acknowledgments I am grateful to László Zsolt Garamszegi for inviting me to write this chapter. Many thanks to two anonymous reviewers for their positive comments.

References

- Bartoszek K, Pienaar J, Mostad P, Andersson S, Hansen TF (2012) A phylogenetic comparative method for studying multivariate adaptation. *J Theor Biol* 314:204–215
- Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A (2007) The delayed rise of present-day mammals. *Nature* 446:507–512
- Blomberg SP, Garland T Jr, Ives AR (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57:717–745
- Blomberg SP, Lefevre JG, Wells JA, Waterhouse M (2012) Independent contrasts and PGLS regression estimators are equivalent. *Syst Biol* 61:382–391
- Bock KE (1966) The comparative method of anthropology. *Comp Stud Soc Hist* 8:269–280
- Bock WJ (1959) Preadaptation and multiple evolutionary pathways. *Evolution* 13:194–211
- Bock WJ (2003) Ecological aspects of the evolutionary processes. *Zool Sci* 20:279–289
- Borgerhoff Mulder M, Nunn CL, Towner MC (2006) Cultural macroevolution and the transmission of traits. *Evol Anthropol* 15:52–64
- Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis models and estimation procedures. *Am J Hum Genet* 19:233–257
- Cheverud JM, Dow MM, Leutenegger W (1985) The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. *Evolution* 39:1335–1351
- Cliff AD, Ord JK (1981) Spatial and temporal analysis: autocorrelation in space and time. In: Wrigley EN, Bennett RJ (eds) *Quantitative geography: a british view*. Routledge & Kegan Paul, London, pp 104–110
- Clutton-Brock TH, Harvey PH (1979) Comparison and adaptation. *Proc R Soc Lond B* 205:547–565
- Cuvier G (1798) *Tableau élémentaire de l’histoire naturelle des animaux*. Baudouin, Paris
- Darwin C (1859) *On the origin of species by means of natural selection*. John Murray, London
- Eastman JM, Harmon LJ, Tank DC (2013) Congruification: support for time scaling large phylogenetic trees. *Meth Ecol Evol* 4:688–691
- Felsenstein J (1973) Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am J Hum Genet* 25:471–492
- Felsenstein J (1985) Phylogenies and the comparative method. *Am Nat* 125:1–15
- Felsenstein J (2004) *Inferring phylogenies*. Sinauer Associates, Sunderland

- Felsenstein J (2005) Using the quantitative genetic threshold model for inferences between and within species. *Phil Trans R Soc Lond B* 360:1427–1434
- Felsenstein J (2008) Comparative methods with sampling error and within-species variation: contrasts revisited and revised. *Am Nat* 171:713–725
- Fisher RA (1930) The genetical theory of natural selection (a complete variorum edition, 1999). Oxford University Press, Oxford
- Forster P, Toth A, Bandelt HJ (1998) Evolutionary network analysis of word lists: visualising the relationships between Alpine romance languages. *J Quant Linguist* 5:174–187
- Freedman DA (2009) Statistical models: theory and practice (revised edition). Cambridge University Press, Cambridge
- Garamszegi LZ, Møller AP (2010) Effects of sample size and intraspecific variation in phylogenetic comparative studies: a meta-analytic review. *Biol Rev* 85:797–805
- Gittleman JL, Kot M (1990) Adaptation: statistics and a null model for estimating phylogenetic effects. *Syst Zool* 39:227–241
- Grafen A (1989) The phylogenetic regression. *Phil Trans R Soc Lond B* 326:119–157
- Grafen A, Ridley M (1997) A new model for discrete character evolution. *J Theor Biol* 184:7–14
- Grandcolas P, Nattier R, Legendre F, Pellens R (2011) Mapping extrinsic traits such as extinction risks or modelled bioclimatic niches on phylogenies: does it make sense at all? *Cladistics* 27:181–185
- Hadfield JD, Nakagawa S (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *J Evol Biol* 23:494–508
- Hadjipantelis PZ, Jones NS, Moriarty J, Springate DA, Knight CG (2013) Function-valued traits in evolution. *J R Soc Interface* 10(20121):032
- Haeckel E (1887) Report on the Radiolaria collected by H.M.S. Challenger during the years 1873–1876. Her Majesty's Stationery Office, London
- Hansen TF, Martins EP (1996) Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* 50:1404–1417
- Harvey PH, Pagel MD (1991) The comparative method in evolutionary biology. Oxford University Press, Oxford
- Houwling-Duistermaat JJ, van Houwelingen HC, Terhell A (1998) Modelling the cause of dependency with application to filaria infection. *Statist Med* 17:2939–2954
- Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO (2012) The global diversity of birds in space and time. *Nature* 491:444–448
- Jombart T, Pavoine S, Devillard S, Pontier D (2010) Putting phylogeny into the analysis of biological traits: a methodological approach. *J Theor Biol* 264:693–701
- Kieckbaev DI (2003) Comparative law: method, science or educational discipline? *Electronic journal of comparative law* 7.3. url <http://www.ejcl.org/73/art73-2.html>
- Kuhn TS, Mooers AO, Thomas GH (2011) A simple polytomy resolver for dated phylogenies. *Meth Ecol Evol* 2:427–436
- Lamarck JB (1809) Philosophie zoologique. Flammarion (1994 edition), Paris
- Lapiedra O, Sol D, Carranza S, Beaulieu JM (2013) Behavioural changes and the adaptive diversification of pigeons and doves. *Proc R Soc Lond B* 280(20122):893
- Laurent G (1986) Cuvier et Lamarck: la querelle du catastrophisme. *La Recherche* 17:1510–1518
- Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology* 74:1659–1673
- Leroi AM, Rose MR, Lauder GV (1994) What does the comparative method reveal about adaptation? *Am Nat* 143:381–402
- Lindenfors P, Jansson F, Sandberg M (2011) The cultural evolution of democracy: saltational changes in a political regime landscape. *PLoS ONE* 6:e28270
- Losos JB (1994) An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. *Syst Biol* 43:117–123
- Losos JB (2011) Convergence, adaptation, and constraint. *Evolution* 65:1827–1840
- Lynch M (1991) Methods for the analysis of comparative data in evolutionary biology. *Evolution* 45:1065–1080

- Mace R, Pagel M (1994) The comparative method in anthropology (with discussion). *Curr Anthropol* 35:549–564
- Martins EP (1996) Conducting phylogenetic comparative studies when the phylogeny is not known. *Evolution* 50:12–22
- Martins EP (2000) Adaptation and the comparative method. *Trends Ecol Evol* 15:296–299
- Martins EP, Hansen TF (1997) Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am Nat* 149:646–667 erratum vol 153, p 488
- Münkemüller T, Lavergne S, Bzeznik B, Dray S, Jombart T, Schiffrers K, Thuiller W (2012) How to measure and test phylogenetic signal. *Meth Ecol Evol* 3:743–756
- Ollier S, Couteron P, Chessel D (2006) Orthonormal transform to decompose the variance of a life-history trait across a phylogenetic tree. *Biometrics* 62:471–477
- O'Meara BC, Ané C, Sanderson MJ, Wainwright PC (2006) Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60:922–933
- Pagel M (1994) Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc R Soc Lond B* 255:37–45
- Pagel M, Meade A (2006) Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am Nat* 167:808–825
- Pagel M, Meade A, Barker D (2004) Bayesian estimation of ancestral character states on phylogenies. *Syst Biol* 53:673–684
- Pagel MD, Harvey PH (1988) Recent developments in the analysis of comparative data. *Quart Rev Biol* 63:413–440
- Paradis E, Claude J (2002) Analysis of comparative data using generalized estimating equations. *J Theor Biol* 218:175–185
- Pavoine S, Ollier S, Pontier D, Chessel D (2008) Testing for phylogenetic signal in phenotypic traits: new matrices of phylogenetic proximities. *Theor Pop Biol* 73:79–91
- Pavoine S, Baguette M, Bonsall MB (2010) Decomposition of trait diversity among the nodes of a phylogenetic tree. *Ecol Monogr* 80:485–507
- Pennell MW, Harmon LJ (2013) An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. *Ann NY Acad Sci* 1289:90–105
- Pepin KM, Lass S, Pulliam JRC, Read AF, Lloyd-Smith JO (2010) Identifying genetic markers of adaptation for surveillance of viral host jumps. *Nat Rev Microbiol* 8:802–813
- Revell LJ (2009) Size-correction and principal components for interspecific comparative studies. *Evolution* 63:3258–3268
- Rohlf FJ (2006) A comment on phylogenetic correction. *Evolution* 60:1509–1515
- Santos JC, Cannatella DC (2011) Phenotypic integration emerges from aposematism and scale in poison frogs. *Proc Natl Acad Sci USA* 108:6175–6180
- Shipley B (2013) The AIC model selection method applied to path analytic models compared using a d-separation test. *Ecology* 94:560–564
- Simpson GG (1944) *Tempo and mode in evolution*. Columbia University Press, New York
- Smith SA, Beaulieu JM, Stamatakis A, Donoghue MJ (2011) Understanding angiosperm diversification using small and large phylogenetic trees. *Am J Bot* 98:404–414
- Spreitzer RJ, Peddi SR, Satagopan S (2005) Phylogenetic engineering at an interface between large and small subunits imparts land-plant kinetic properties to algal Rubisco. *Proc Natl Acad Sci USA* 102:17225–17230
- Stone GN, Nee S, Felsenstein J (2011) Controlling for non-independence in comparative analysis of patterns across populations within species. *Phil Trans R Soc Lond B* 366:1410–1424
- Thomas GH, Hartmann K, Jetz W, Joy JB, Mimoto A, Mooers AO (2013) PASTIS: an R package to facilitate phylogenetic assembly with soft taxonomic inferences. *Meth Ecol Evol* 4:1011–1017
- Tiao GC, Reinsel GC, Xu DM, Pedrick JH, Zhu XD, Miller AJ, DeLuisi JJ, Mateer CL, Wuebbles DJ (1990) Effects of autocorrelation and temporal sampling schemes on estimates of trend and spatial correlation. *J Geophys Res-Atmos* 95:20507–20517

- von Hardenberg A, Gonzalez-Voyer A (2013) Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis. *Evolution* 67:378–387
- Watt WB (2013) Causal mechanisms of evolution and the capacity for niche construction. *Biol Philos* 28:757–766
- Whitney KD, Boussau B, Baack EJ, Garland T (2011) Drift and genome complexity revisited. *PLoS Genet* 7(e1002):092
- Yan XH, Gurtler JB, Fratamico PM, Hu J, Juneja VK (2012) Phylogenetic identification of bacterial MazF toxin protein motifs among probiotic strains and foodborne pathogens and potential implications of engineered probiotic intervention in food. *Cell Biosci* 2:39