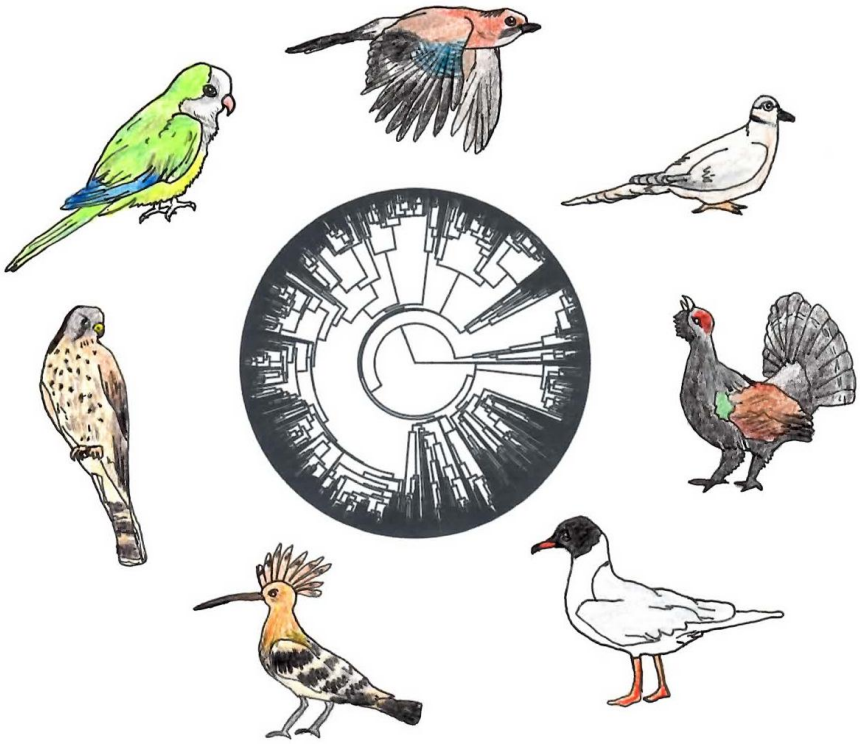
BRAIN SIZE AND EVOLUTIONARY DIVERSIFICATION IN BIRDS



Ferran Sayol i Altarriba

Supervisor: Daniel Sol

Center for Ecological Research and Applied Forestries (CREAF); **12th September 2013**

Master of Terrestrial Ecology and Biodiversity management. Modality of Terrestrial Ecology.

Starting of the project:

The decision to work on brain evolution was chosen in March 2013, after an extensive process of two months reading articles from different topics (behavioral ecology, brain evolution and life-history) to get ideas to work on. In April I started to work with the project exclusively.

Partial contribution of the student to the project:

During three months, me and Mar Unzeta searched articles in the literature with data from different species to construct three general databases on i) life-history traits, ii) ecology information and iii) behavioral traits that were used to get the data for our respective work and also will be of great utility for future projects. We searched the sources and we cleaned and ordered data, compared different variables from different authors to provide that they were comparable and we built the database with the help of Joan Maspons in joining all the data with previous datasets that the Sol Lab group had, looking for synonyms of species and calculate one single value from each species.

Complete contribution of the student to the project:

In parallel with this process, I searched for antecedents on the topic of my project to get familiarized with the terms and methodologies in comparative studies of evolutionary diversification and understand the state of discussion of the topic. Then I wrote the introduction to set the framework of the project, planned the basic questions and launched the predictions from the theory. After that, I merged the data in Brain and Body size from 1326 species (This data was facilitated by Andrew Iwaniuk) with the general datasets previously built and complete the information for those variables that had controversies from two different authors or by which I do not had information. Once the project dataset was built, I talked with Daniel Sol and Oriol Lapiedra to get a basic introduction in the phylogenetic methods currently available. Then I started to do the analysis detailed in the Methods, summarized the results and finally wrote the manuscript of the project.

BRAIN SIZE AND EVOLUTIONARY DIVERSIFICATION IN BIRDS

Ferran Sayol

The role of behavior in evolution remains controversial, despite that some ideas are over 100 years old. Changes in behavior are generally believed to enhance evolution by exposing individuals to new selective pressures and by facilitating range expansions. However, this hypothesis lacks firm empirical evidence. Moreover, behavioral changes can also inhibit evolution by hiding heritable variation from natural selection. Taking advantage of the complete phylogeny of extant birds, a new species-level measure of past diversification rate and the best existing measures of brain size (n = 1326 species), I show here that relative brain size is associated (albeit weakly) with diversification rates. Assuming that brain relative size reflects behavioral flexibility, an assumption well-supported by evidence, this finding supports the idea that behavior can enhance evolutionary diversification. This view is further supported by the discovery that the most important factor influencing diversification rates is ecological generalism, which is believed to require behavioral flexibility. Thus, behavioral changes that expose animals to a variety of environments can have played an important role in the evolution of birds.

Behavior has long been proposed to be a major driver of evolutionary diversification in animals. Over fifty years ago, Mayr (1963) argued that any shift into a new adaptive zone is initiated by a change in behavior, other traits acquired secondarily. Twenty years later, Wyles et al. 1983 further proposed that by adopting a new behavior, the species faces a new set of selection pressures favoring those mutations that improve the individual's proficiency at living in the new way. This idea was called the "Behavioral Drive" hypothesis and makes two main predictions. The first is that changes in behavior generally drive changes in the phenotype, a prediction supported by experimental (Losos et al. 2004) and phylogenetic-based analyses (Lapiedra et al. 2013). The second prediction is that animals with high propensity for behavioral innovation should evolve at a faster rate than those with less propensity, whether in terms of phenotype disparity or taxonomic diversity. A high propensity for innovation is expected in animals with large brains (Lefebvre et al. 2004, 2013), which have greater learning skills to find new ways to interact with the environment and to use new opportunities and resources in new environments (Sol et al. 2005a). By exposing themselves to new selection pressures more frequently, large-brained lineages should experience enhanced rates of evolutionary diversification (Wyles et al. 1983).

Behavior can also promote diversification by mechanisms other than the behavioral drive effect. Mayr (1963), for example, suggested that behavioral adjustments may help species invade new regions and expand their range, which may increase the chances of population divergence through genetic drift, subdivision across geographical barriers, and/or increased persistence over evolutionary time (see also Rosenweig, 1995). The idea that behavior promotes diversification such as favoring range expansion that can cause allopatric speciation

is known as the geographic model of diversification (Sol & Price 2008), and also predicts that large-brained lineages should experience enhanced rates of evolutionary diversification.

Nevertheless, the role of behavior on evolution remains controversial because changes in behavior are a main mechanism through which animals confront new environmental pressures; behavioral changes can thus hide genetic variation from natural selection, retarding evolution (Huey et al. 2003). This is called the "Bogert effect", and has been proposed as an alternative of the Behavioral Drive Hypothesis. It has also been argued that when behavioral shifts are an adaptation to the current environment, they should maintain evolutionary stasis; instead, when behavioral shifts produce novel behavior they are more likely to drive evolution (Duckworth 2008). Consequently, the question of whether behavioral plasticity inhibits or promotes evolution remains unresolved, despite the fact that the first hypothesis was launched over half a century ago.

The behavioral drive and geographic hypotheses predict thus a positive association between brain size and evolutionary diversification rate, whereas the Bogert effect and related hypotheses predict a negative relationship or no relation at all. Previous studies have found that brain size is related to morphological diversification in major taxonomic groups (Wyles et al. 1983), species diversification in birds (Nicolakakis et al. 2003), subspecies diversification in Holartic passerines (Sol et al. 2005b) and body size diversification in bird families (Sol & Price 2008). Although these studies are consistent with the "Behavioral drive" idea (but see Lynch, 1990), they use indirect measures of evolutionary diversification and the comparisons are at the family or higher taxonomic levels. In this study, I take advantage of the first set of complete phylogenies of extant bird species, a new species-level measure of past diversification rate (Jetz et al. 2012a) and precise measures of brain size to address the question of whether large-brained species (i.e. with increased behavioral flexibility) have manifested accelerated diversification rates.

The prediction is tested in the light of well-established intrinsic and extrinsic factors known to affect evolutionary diversification in birds. These include: i) Generation time, that is invertible proportional to rates of DNA substitution and hence can retard evolutionary change (Martin & Palumbi 1993); ii) Ecological generalism, that is proved to increase evolutionary diversification under the ecological theory of speciation (Phillimore et al. 2006); iii) Degree of geographic isolation or insularity (Emerson 2002; Pinto et al. 2008), that may increase diversification because islands often combine ecological opportunities with geographic isolation; iv) Biogeographical region, as diversification rate vary between world regions (Ericson 2012; Jetz et al. 2012b); and v) Migratory behavior, because factors explaining species richness may depend on functional groups as migratory/resident (Carnicer & Díaz-Delgado 2008). Also Migratory Behavior can serve as a confound variable as Brain size can change between resident and migrant species (Sol et al. 2010). All this, in conjunction with the recently available data on Brain size and phylogeny of birds, provides a rare opportunity to address the long questioned implication of behavior on evolutionary diversification.

METHODS

Calculating diversification rate of species

Using the first complete phylogeny of all 9,993 extant species of birds (Jetz et al. 2012b), I calculated a diversification rate measure (DR) for each species based on the inverse of the Equal Splits (ES) metric of evolutionary isolation (Redding & Mooers 2006).

The ES metric distributes the evolutionary history represented by branches lengths among all the species. The method consists in dividing the evolutionary time represented by a branch equally among its daughter branches. The ES measure for a single species is the sum of the edge lengths from the species to the root, with each consecutive edge discounted by a factor of ½.

$$ES_i = \sum_{j=1}^{N_i} L_j \frac{1}{2^{j-1}}$$

$\boldsymbol{N_i}$ is the number of internal nodes on the path from species ***i*** to the root, and $\boldsymbol{L_j}$ is the length of the branch **j**, with **j=1** being the pendant branch leading to the species and **j=$\boldsymbol{N_i}$** being the branch nearest the root.

The ES measure represents the phylogenetic distinctiveness of a species relative to the other species. The function *evol.distinct* from the *R*-package "picante" (see all *R* functions in Appendix A) was used to calculate the evolutionary distinctiveness for all species by equal splits. For this purpose, I used a set of 100 trees of 9993 species (Jetz et al. 2012b), half of each built using two different backbones from two independent phylogenetic studies (Hackett et al. 2008; Ericson 2012) as a way to integrate phylogenic uncertainty in the analysis.

The inverse of ES measure can be seen as the splitting rate of species from the root to the edge, and is termed diversification rate or DR (Jetz et al. 2012b). Species in rapidly-diversifying clades will have short branch lengths shared among many species (High DR), while species in slowly-diversifying clades will have long branches (Low DR).

As the distribution of DR values over the 100 trees had some extreme values on the right tail for the majority of the species (Appendix B, Fig. B1), I took the median of the DR value of each species for 100 trees rather than the mean value to avoid overestimate DR. The median values for Ericson and Hackett backbones were highly congruent (Appendix B, Fig. B2), indicating that the measure was robust to the effect of phylogenetic uncertainties. In the analyses, DR was log-transform to improve normality (Appendix B, Fig. B3).

Brain data and estimates of brain residuals

The use of brain size as surrogate of behavioral flexibility is based on firm evidence that birds with larger brains, relative to their body size, have a higher propensity to learn new behaviors (Lefebvre et al. 2004, 2013) and that these changes in behavior facilitate the response to novel environments (Sol et al. 2005a).

Brain size was provided by Andrew Iwaniuk, who measured the endocranial volume in skulls of museum specimens in both males and females (5319 specimens from 1326 species). The

endocast method has previously been found to be a highly reliable estimate of the whole brain size (Iwaniuk & Nelson 2002). Because large birds also tend to have large brains, I estimated a relative measure of brain size independent of body size to obtain a brain size metric that is biologically meaningful (Lefebvre et al. 2004). This was facilitated because, unlike previous studies, brain size and body size data came from the same specimens. There are different forms of obtaining this relative measure, all consisting in constructing a model between the logarithm of body size and logarithm of brain size, and then extract the residuals. The residual give us information about how much bigger or smaller is a brain compared from the value expected from body size. The residuals were estimated as follows. Firstly, I calculated a species value of Brain size and Body size as the mean between the two sexes of each species. Then, I used three different methods to estimate the residuals between Log(Body Size) and Log(Brain Size): i) An ordinary linear regression, ii) a quadratic regression, and iii) a phylogenetic-corrected least-squares regression (Appendix B, Fig. B4). All three methods yielded similar results, and I decided to use the phylogenetic-corrected least-squares regression because it reduced possible biases derived from the non-random sampling of species. To this purpose, I used the function *phyl.resid* of the *R*-package "Phytools", which conducts phylogenetic size-corrections based on least-squares regression (Revell 2012). I used a set of 100 phylogenetic trees of the 1326 species for which I had information from Brain and body size, so the residuals of the model were obtained 100 times. Then, I calculated the median of the 100 measures of residuals obtained for each species. This constituted the relative brain size variable, which is quite an intuitive measure: Values greater than 0 correspond to species that have brains relative large compared to their body size and values lower than 0 correspond to species with small brains relative to their body size.

Intrinsic and extrinsic traits affecting diversification rate

As DR can be affected by many factors, I also included in my analyses some variables considered in the literature to be important in explaining diversification rate. For the 1326 species from which I had information about brain and body size, I looked for data on six key variables:

i) Biogeography, defined by seven World regions (Apendix A, Fig. 5): *Palearctic*, *Nearctic*, *Neotropical*, *Africa*, *Australia*, *Indomalaysia*, *Antartida*. An eighth category (*Multi-region*) was defined when the distribution range of a species overlapped more than one world region. The information was obtained from published sources (mostly from (Cockburn 2006) and Del Hoyo et al.), complemented with maps from BirdLife International (http://www.birdlife.org/).

ii) Insularity (Strict island endemic vs. Mainland species), taken from published sources (Figuerola & Green 2005; McNab 2009; Wasser & Sherman 2010; Covas 2012) and complemented with information extracted from the distribution maps of BirdLife International. I considered a species as island endemic if the species distribution was restricted to an island or group of islands of less than 500.000 km^2. When a species was encountered in both island and mainland, this was considered as a mainland species.

iii) Migratory behavior, described in three categories (migrant, resident or nomadic), considering altitudinal migrants as residents and partial migrants as migrants. The information was obtained from published sources (Meiri et al. 2003; Cockburn 2003; Scheuerlein & Ricklefs 2004; Hanowski et al. 2005; McNab 2009; Sol et al. 2010; Wasser & Sherman 2010; Reif et al. 2010; Sibly et al. 2012) . When information for a species that was not available in literature or when there were discrepancies between two authors, I obtained the information directly from the distribution maps of BirdLife International.

iv) Generation time, calculated as AFB+[1/m] where *AFB* is the Age at first breeding in years and *m* is mortality. Information was obtained from the previous cited papers, and complemented with information from BirdLife International.

v) Habitat breadth, estimated as the number of breeding habitats used by the species during the breeding season. The habitats considered were forest, wooded, shrubs, tundra, grassland, marsheswetland, cliffs, urban, and rural (Phillimore et al. 2006). Information was obtained from a variety of published sources (Jones, 2009; McNab, 2009; Wasser, 2010; Poysa, 2012).

vi) Taxonomic assignation (Order, and *Non-passerines*, *Passeri* and *Tyranii*), obtained from Jetz et al. (Jetz et al., 2012) as a variable to control for the fact that some groups have diversified more than others for reasons other than those included here. Because of analytical difficulties, order containing less than four species were merged with sister orders. Using the set of trees of 1326 species, I joined some Orders to create monophyletic new groups, by taking a set of 100 trees with one species of each order from Jetz et al., 2012 and building a consensus tree with *consensus* function from *R*-package "Ape". After examining the tree, I merged: a) Coraciiformes, Piciformes, Bucerotiformes, Coliiformes; b) Sheniciformes, Procelaniformes, Pelecaniformes; c) Gaviiformes, Suliformes, Ciconiiformes; d)Phaethoniformes, Musophagiformes; e) Podicipediformes, Phoenicopteriformes (Appendix B, Fig. B6).

Analysis

To avoid problems of co-linearity, before the analysis I examined the correlation between explanatory variables and relative brain size. It is known that habitat breadth, generation time and migratory behavior are associated with brain size. Big brained species have greater lifespan and delayed sexual maturity (González-Lagos et al. 2010) and hence higher generation times, and also tend to be generalist (Lefebvre & Sol 2008) so they use the habitat more widely. On the other hand, migratory birds have smaller brains than resident birds (Sol et al. 2010). For this reason, I used function *vif* from *R*-package "car" to calculate the variance inflator factor between relative brain size and the rest of explanatory variables. There was no reason to omit any variable in our data as the VIF was lower than five in all cases (Appendix B, Table B1).

Closely related species tend to be more similar than distantly related ones in both their phenotype and ecology (Adams 2008). Thus, treating species as independent in the analyses can violate the assumption of independence of the data. I dealt with this problem by first use the phylogeny to estimate the parameter λ, which measures the degree of phylogenetic

autocorrelation of the data. To estimate the parameter ʎ for phylogenetic dependence of data (Pagel 1999) I used the function *phylosig* from *R*-package "phytools" (Revell 2012) to compute phylogenetic signal using maximum likelihood, resulting in a ʎ parameter for a continuous variable as a scale transformation of branches of the tree. Thus, as Lambda moves from 1 to 0, the internal branches become smaller so the tree reflects less phylogenetic structure (Nunn 2011). I estimated the phylogenetic signal of the dependent variable (Diversification Rate) and also of the main independent variable (relative brain size) using a set of 100 trees with 1326 species (Appendix B, Fig. B7).

Because I found important phylogenetic autocorrelation in the data, I explored the relationship between brain size and diversification rate with a Phylogenetic least square regression (PGLS) approach that take into account this non-independence of the data. PGLS is a useful comparative methods proposed by Pagel in 1999 to account for phylogenetic relationships, which fits a linear model accounting for the strength and type of the phylogenetic signal in the data matrix by adjusting the branch length of the phylogeny (Nunn 2011). For this, I used the two sets of trees from BirdTree (Jetz et al. 2012b) of the 1326 species from my dataset differing in the backbone used (Hackett et al. 2008; Ericson 2012).

I first tested the relation between brain size and diversification rate with an univariate PGLS, with Log(DR) as dependent variable. Next, using the dataset of species with information for all variables (relative brain size, Habitat breadth, Migratory, Insularity, Generation Time and Biogeography; N=603 species), I ran a model selection process using the *dredge* function from the "MuMIn" *R*-package to examine the importance of the different variables in the best models. To ensure that the conclusions were not contingent of the phylogeny used, I repeated the model selection process using different random trees from each of the two phylogeny backbones (Hackett et al. 2008; Ericson 2012). The most important variables were chosen by examining their sum of weights using the function *importance* from "MuMIn". In all the models I examined diagnostic plots in order to check for outliers, heteroscedasticity, and non-normal errors.

RESULTS

The lambda (ʎ) estimated for a set of 100 trees of 1326 species is almost invariable significantly higher than zero both for Diversification Rate (0.405 ± 0.029, N=1326) and Relative Brain Size (0.943±0.007, N=1326) (**Fig. 1**). This indicates that both variables are phylogenetic autocorrelated, which highlights the need of using a phylogeny corrected methods (here PGLS) instead of ordinary regressions.

In the univariant PGLS, relative brain size is a significant predictor of DR, although the correlation is admittedly marginally significant (**Fig. 2**). Apart from brain residual, only habitat breadth is a significant predictor of DR (**Table 1**). Factors that are considered critical in diversification process, such as geographic isolation or generation time, are not associated with DR in the univariate analyses (**Table 1**).

In the model selection process with the all the factors tested simultaneously, habitat breadth and generation time are the variables with greater support (higher weights), followed by insularity, migratory behavior and relative brain size and. Order and Biogeography are absent in most models (**Table 1**). In some of the best models, relative brain size is not even significant. However, this does not necessarily mean that the effect of relative brain size on diversification rate is indirectly caused by other factors like habitat breadth or generation time. Rather, this reflects a reduction in the power of the test as complete information of all the variables was only available for 603 out of the 1326 species. Indeed, when the effect of relative brain size on diversification rate is tested with each confounding factor in turn (**Table 2**), the estimate of brain residual slope is very similar to that observed when brain residual is tested alone. The lack of a significant effect of relative brain size on DR seems thus to indicate that the strength of the effect is weak, as it is also for the other factors studied, which makes significance to disappear when sample size decreases.

DISCUSSION

My analyses suggest that having a large brain, relative to body size, can sometimes favor higher evolutionary diversification rates. Relative brain size, after habitat breadth, generation time and insularity and migratory behavior, is the one of the most important factors in the model selection. Assuming that a large relative brain reflects enhanced behavioral flexibility, an assumption well supported in previous studies (Lefebvre et al. 2004; Sol et al. 2007), my results are thus consistent with the view that behavior enhances (rather than inhibits) evolutionary rates.

However, relative brain size alone can only explain a reduced percentage of variation in the diversification rate of birds. This can reflect that there are a number of factor that can affect diversification rate. In my analysis, the strongest factor influencing diversification rate is habitat breadth, consistent with the ecological theory of evolution (Schluter 2000). This factor appears to be a significant predictor for diversification rate both in the univariate analysis and when adding brain size. This is congruent with the results from other studies (Phillimore et al., 2006), where ecological generalism was an important predictor of diversification, but in those case ecological generalism was measured as diet breadth and here I have measured as habitat breadth. Nevertheless, results suggest that species that use more habitat types, have experienced higher diversification rates, which may be explained by the niche variation hypothesis, which proposes that some species of ecological generalists are in fact groups of relatively specialized individuals (Bolnick, Svanbäck, Araújo, & Persson, 2007), and this specialized individuals can diverge into different species, promoting evolutionary diversification. When including in the same model habitat breadth and relative brain size, this later variable turns out non-significant. This could be caused by the correlation between these variables, as suggested in previous studies (Lefebvre & Sol 2008) where big brained animals are found to be ecological generalists. However, this is not the cause in our case as there is no correlation between both variables. A more likely explanation is a reduction in the power of

the test, as data for habitat breadth was only available for over 652 species. Relative brain size is non-significant when we use only the information for those species, perhaps reflecting that the signal is so weak than it requires large samples to be detected. Future studies should try to enlarge sample size in habitat breadth to further disentangle the direct and habitat mediated effects of relative brain size on diversification rates.

The relationship between brain size and diversification rate could also be obscured by other factors influencing evolution, which are more difficult to quantify. These include range expansions and geographical barriers, which are difficult to estimate in large-scale analysis. For instance, my analyses did not find any effect of isolation on diversification rates even when there is strong empirical evidence that geographic isolation is critical in both speciation processes and adaptive radiations(Price 2008). Future studies should scale-down the level of analysis, focusing on regions where the geographic component can be more easily assessed.

It is also worth noting that my analyses are restricted to diversification rates, which is not always associated with phenotypic change rates (i.e. disparity). Darwin finches, for example, are only 14 species yet the group exhibits enormous morphological disparity in bill shape and body size (Grant 1999). While diversification rates are appropriate to test the role of behavior in the geographical model of evolution, it is not so obvious that is the best metric to test the behavioral drive. This hypothesis assumes that changes in behavior lead the shifts to a new phenotypic optima by changing the selection forces that cause adaptive change (Wyles et al. 1983). However, if a high divergence in phenotype does not always translate to higher evolutionary diversification, the behavioral drive hypothesis cannot be accurately inferred from diversification rates calculated from branch lengths of phylogeny. This issue can be clarified by incorporating phenotypic disparity in the analyses, allowing in turn to disentangle the geographic and behavioral drive effects on evolutionary rates.

Finally, it could be also possible that the factors influencing evolutinary diversification are different between groups, with some exhibiting a positive relation between relative brain size and diversification rate and others a negative association in others or no association at all (behavioral inhibition or Bogert effect). When conducting a general analysis with all the groups, positive and negative correlation counter balance each other, reducing the strength of the association. This possibility can be investigated by additional analyses where the relationship between brain size and diversification is compared within more homogeneous taxonomic groups that still exhibit variation in brain size. Alternatively, the analyses could be conducted within well-defined regions, which would allow to control for some of the factor that may affect diversification in larger scales such as biogeographical barriers.

In sum, my analyses provide some support for a role of behavioral flexibility, as described by relative brain size, and diversification rates. However, the relationship between brain size and diversification rates is weak, which call into question the general importance of behavior in shaping diversification rates. This nonetheless does not deny the importance of behavioral changes in the adaptive diversification of animals.

AKNOWLEDGEMENTS

I am grateful to the whole team of Sol Lab for helpful discussions and methodological advice, and specially to Daniel Sol for supervising the work. I also thank Andrew Iwaniuk for facilitating all brain information.

REFERENCES

Adams, D. C. 2008. Phylogenetic meta-analysis. *Evolution*, **62**, 567–72.

Carnicer, J. & Díaz-Delgado, R. 2008. Geographic differences between functional groups in patterns of bird species richness in North America. *Acta Oecologica*, **33**, 253–264.

Cockburn, A. 2003. Cooperative breeding in oscine passerines: does sociality inhibit speciation? *Proceedings of the Royal Society of London, Series B*, **270**, 2207–14.

Cockburn, A. 2006. Prevalence of different modes of parental care in birds. *Proceedings of the Royal Society of London, Series B*, **273**, 1375–83.

Covas, R. 2012. Evolution of reproductive life histories in island birds worldwide. *Proceedings. Biological sciences / The Royal Society*, **279**, 1531–7.

Duckworth, R. a. 2008. The role of behavior in evolution: a search for mechanism. *Evolutionary Ecology*, **23**, 513–531.

Emerson, B. C. 2002. Evolution on oceanic islands : molecular phylogenetic. 951–966.

Ericson, P. G. P. 2012. Evolution of terrestrial birds in three continents: biogeography and parallel radiations. *Journal of Biogeography*, **39**, 813–824.

Figuerola, J. & Green, A. J. 2005. A comparative study of egg mass and clutch size in the Anseriformes. *Journal of Ornithology*, **147**, 57–68.

González-Lagos, C., Sol, D. & Reader, S. M. 2010. Large-brained mammals live longer. *Journal of evolutionary biology*, **23**, 1064–74.

Grant, P. R. 1999. *Ecology and Evolution of Darwin's Finches.* Princeton University Press.

Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C. K., Braun, E. L., Braun, M. J., Chojnowski, J. L., Cox, W. A., Han, K.-L., Harshman, J., Huddleston, C. J., Marks, B. D., Miglia, K. J., Moore, W. S., Sheldon, F. H., Steadman, D. W., Witt, C. C. & Yuri, T. 2008. A phylogenomic study of birds reveals their evolutionary history. *Science (New York, N.Y.)*, **320**, 1763–8.

Hanowski, J., Lind, J., Danz, N., Niemi, G. & Jones, T. 2005. Regional Breeding Bird Monitoring in Western Great Lakes National Forests. *USDA forest service general technical report*, 974–981.

Huey, R. B., Hertz, P. E. & Sinervo, B. 2003. Behavioral drive versus behavioral inertia in evolution: a null model approach. *The American naturalist*, **161**, 357–66.

Iwaniuk, A. N. & Nelson, J. E. 2002. Can endocranial volume be used as an estimate of brain size in birds? *Journal of Zoology*, **80**, 16–23.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, a O. 2012a. The global diversity of birds in space and time. *Nature*, **491**, 444–8.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, a O. 2012b. The global diversity of birds in space and time. *Nature*, **491**, 444–8.

Lapiedra, O., Sol, D., Carranza, S. & Beaulieu, J. M. 2013. Behavioural changes and the adaptive diversification of pigeons and doves. *Proceedings of the Royal Society of London, Series B*, **280**, 20122893.

Lefebvre, L. & Sol, D. 2008. Brains, lifestyles and cognition: are there general trends? *Brain, behavior and evolution*, **72**, 135–44.

Lefebvre, L., Reader, S. M. & Sol, D. 2004. Brains , Innovations and Evolution in Birds and Primates. *Brain, behavior and evolution*, **63**, 233–246.

Lefebvre, L., Reader, S. M. & Sol, D. 2013. Innovating Innovation Rate and Its Relationship with Brains, Ecology and General Intelligence. *Brain, behavior and evolution*, **1**, 1–3.

Losos, J., Schoener, T. & Spiller, D. 2004. Predator-induced behaviour shifts and natural selection in field-experimental lizard populations. *Nature*, **432**, 505–508.

Lynch, M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. *The American naturalist*, **136**, 727–741.

Martin, a P. & Palumbi, S. R. 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 4087–91.

Mayr, E. 1963. *Animal species and evolution.* Cambridge, MA: Harvard University press.

McNab, B. K. 2009. Ecological factors affect the level and scaling of avian BMR. *Comparative biochemistry and physiology. Part A, Molecular & integrative physiology*, **152**, 22–45.

Meiri, S., Dayan, T. & Aviv, T. 2003. On the validity of Bergmann ' s rule. *Journal of Biogeography*, **30**, 331–351.

Nicolakakis, N., Sol, D. & Lefebvre, L. 2003. Behavioural flexibility predicts species richness in birds, but not extinction risk. *Animal Behaviour*, **65**, 445–452.

Nunn, C. L. 2011. *The comparative approach in evolutionary anthropology and biology.* University of Chicago Press.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature*, **401**, 877–84.

Phillimore, A. B., Freckleton, R. P., Orme, C. D. L. & Owens, I. P. F. 2006. Ecology predicts large-scale patterns of phylogenetic diversification in birds. *The American naturalist*, **168**, 220–9.

Pinto, G., Mahler, D. L., Harmon, L. J. & Losos, J. B. 2008. Testing the island effect in adaptive radiation: rates and patterns of morphological diversification in Caribbean and mainland Anolis lizards. *Proceedings of the Royal Society of London, Series B*, **275**, 2749–57.

Price, T. D. 2008. *Speciation in Birds.* Greenwood Village: Roberts and Company Publishers.

Redding, D. W. & Mooers, A. Ø. 2006. Incorporating evolutionary measures into conservation prioritization. *Conservation biology : the journal of the Society for Conservation Biology*, **20**, 1670–8.

Reif, J., Vermouzek, Z., Voříšek, P., Šťastný, K., Bejček, V. & Flousek, J. 2010. Population changes in Czech passerines are predicted by their life-history and ecological traits. *Ibis*, **152**, 610–621.

Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.

Rosenweig, M. . 1995. *Species Diversity in Space and Time. Cambridge.*

Scheuerlein, A. & Ricklefs, R. E. 2004. Prevalence of blood parasites in European passeriform birds. *Proceedings. Biological sciences / The Royal Society*, **271**, 1363–70.

Schluter, D. 2000. *The Ecology of Adaptive Radiation.* Oxford, U.K.: Oxford University Press.

Sibly, R. M., Witt, C. C., Wright, N. a, Venditti, C., Jetz, W. & Brown, J. H. 2012. Energetics, lifestyle, and reproduction in birds. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 10937–41.

Sol, D. & Price, T. D. 2008. Brain size and the diversification of body size in birds. *The American naturalist*, **172**, 170–7.

Sol, D., Duncan, R. P., Blackburn, T. M., Cassey, P. & Lefebvre, L. 2005a. Big brains, enhanced cognition, and response of birds to novel environments. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 5460–5.

Sol, D., Stirling, D. G. & Lefebvre, L. 2005b. Behavioral drive or behavioral inhibition in evolution: subspecific diversification in Holarctic passerines. *Evolution*, **59**, 2669–77.

Sol, D., Székely, T., Liker, A. & Lefebvre, L. 2007. Big-brained birds survive better in nature. *Proceedings of the Royal Society of London, Series B*, **274**, 763–9.

Sol, D., Garcia, N., Iwaniuk, A., Davis, K., Meade, A., Boyle, W. A. & Székely, T. 2010. Evolutionary divergence in brain size between migratory and resident birds. *PloS one*, **5**, e9617.

Wasser, D. E. & Sherman, P. W. 2010. Avian longevities and their interpretation under evolutionary theories of senescence. *Journal of Zoology*, **280**, 103–155.

Wyles, J. S., Kunkel, J. G. & Wilson, a C. 1983. Birds, behavior, and anatomical evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **80**, 4394–7.

TABLES

Table 1: PGLS analysis for Intrinsic and Extrinsic Factors influencing DR. All parameters are based on models in which each variable has been tested alone, except in the case of the weights which correspond to the weight of each variable in all models with the best AIC values (range 0-4).

Model	**Estimate**	**SD**	**AIC Weigth**	**N**	**λ**	**p-value Model**	**p-value of variable**
Habitat breath	0.039	0.016	0.613	652	0.901	0.032	0.016
Generation time	0.014	0.010	0.542	1247	0.944	0.150	0.168
Insularity	0.021	0.040	0.320	1326	0.945	0.771	0.610
Migratory behavior*			0.318	1326	0.945	0.151	
Nomadic	-0.141	0.079					0.075
Resident	$4 \cdot 10^{-4}$	0.036					0.991
Relative brain size	0.178	0.090	0.287	1326	0.944	0.021	0.048

*The category migrant has been set to zero and is used as baseline for comparison.

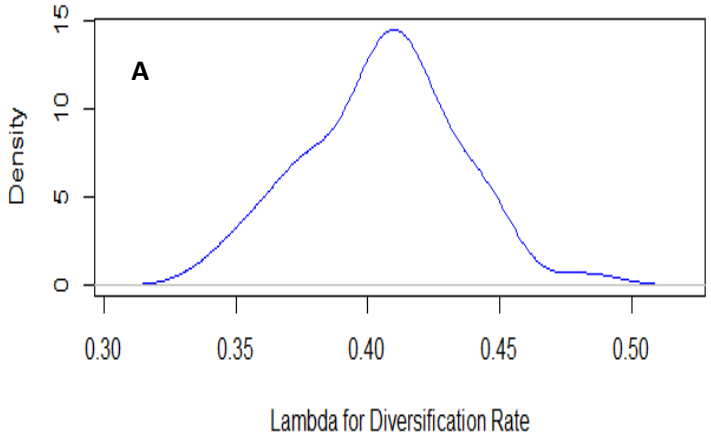
Table 2: Comparison of univariate PGLS with bivariant PGLS on relative brain size together with each of the other factors found to affect RD in the model selection procedure. Each pair-wise comparison is made with the same sample size to make the models comparable (generation time, N=1247; habitat breadth, N=652; rest of models, N = 1326).

Model	**Estimate**	**SD**	**N**	**λ**	**p-value Model**	**p-value of variable**
Relative brain size	0.06	0.13	652	0.90	0.83	0.67
Relative brain size	0.06	0.13	652	0.90	0.03	0.66
Habitat breadth	0.03	0.02				0.02
Relative brain size	0.16	0.09	1247	0.94	0.06	0.09
Relative brain size	0.12	0.09	1247	0.94	0.16	0.22
Generation Time	0.01	0.01				0.26
Relative brain size	0.18	0.09	1326	0.94	0.11	0.05
Insularity	0.01	0.04				0.74
Relative brain size	0.18	0.09	1326	0.94	0.15	0.05
Migratory behavior*						
Nomadic	-0.14	0.08				0.07
Resident	$-8 \cdot 10^{-3}$	0.04				0.82

*The category migrant has been set to zero and is used as baseline for comparison.

FIGURES

Figure 1. Phylogenetic Signal (ʎ) distribution of diversification rate **(A)** and Relative brain size **(B)** over 100 phylogenetic trees.



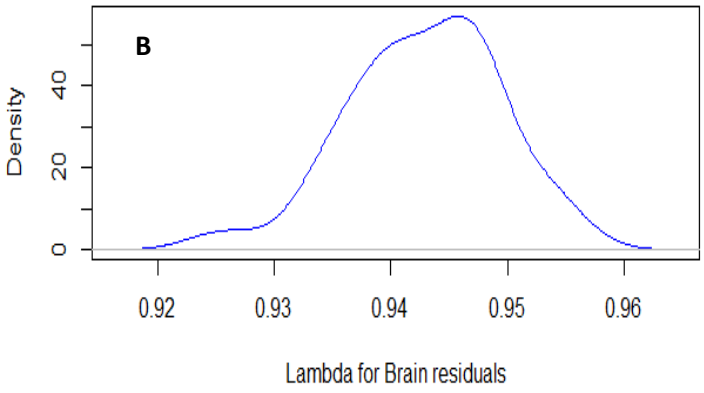
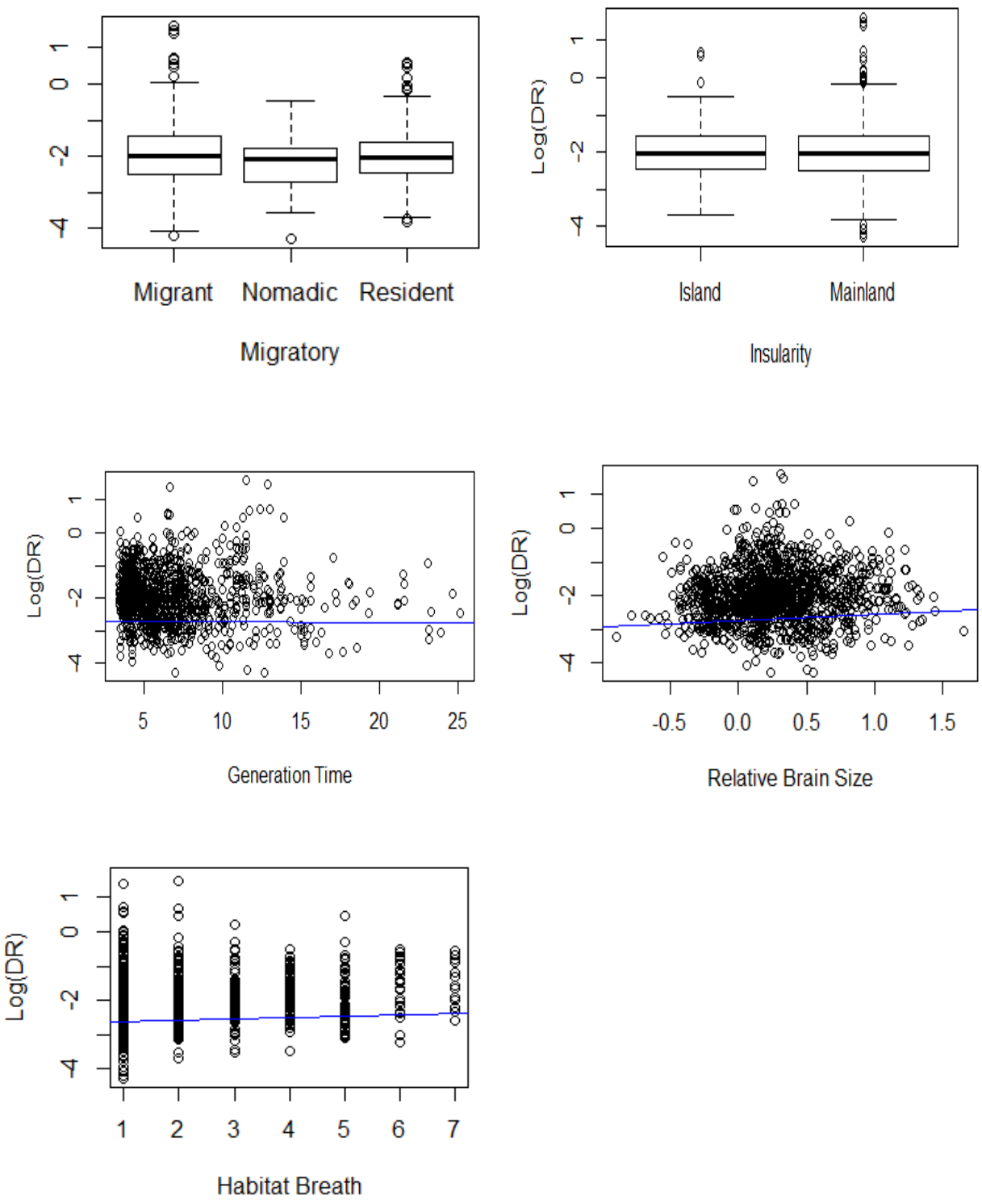


Figure 2. Relationship between the most important factors in the model selection and Log(diversification rate). The blue line describes the model fitted with a PGLS.



APPENDIX A: R functions used

a) *consensus* {Ape}

Given a series of trees, this function returns the consensus tree. By default, the strict-consensus tree is computed. To get the majority-rule consensus tree, use p = 0.5. Any value between 0.5 and 1 can be used.

Usage consensus(..., p = 1, check.labels = TRUE)

Author(s) Emmanuel Paradis

References

Felsenstein, J. (2004) Inferring Phylogenies. Sunderland: Sinauer Associates.

b) *dredge* {MuMIn}

Generate a set of models with combinations (subsets) of the terms in the global model, with optional rules for model inclusion.

Usage dredge(global.model, beta = FALSE, evaluate = TRUE, rank = "AICc", fixed = NULL, m.max = NA, m.min = 0, subset, marg.ex = NULL, trace = FALSE, varying, extra, ct.args = NULL, ...)

Author(s) Kamil Bartón

c) *evol.distinct* {picante}

Calculates evolutionary distinctiveness measures for a suite of species by: a) equal splits (Redding and Mooers 2006).

Usage: evol.distinct(tree, type = c("equal.splits", "fair.proportion"), scale = FALSE, use.branch.lengths = TRUE)

Author(s): Karen Magnuson-Ford,Will Cornwell, Arne Mooers, Mark Vellend

References:

Redding, D.W. and Mooers, A.O. (2006). Incorporating evolutionary measures into conservation prioritisation. Conservation Biology, 20, 1670-1678.

Isaac, N.J.B., Turvey, S.T., Collen, B., Waterman, C. and Baillie, J.E.M. (2007). Mammals on the EDGE: conservation priorities based on threat and phylogeny. PLoS ONE, 2, e296.

Mark Vellend,William K. Cornwell, Karen Magnuson-Ford, and Arne Mooers. In press. Measuring phylogenetic biodiversity. In: Biological diversity: frontiers in measurement and assessment. Edited by Anne Magurran and Brian McGill.

d) *phylosig* {phytools}

This function computes phylogenetic signal using two different methods. It can also conduct the hypothesis tests for significant phylogenetic signal, and estimate phylogenetic signal incorporating sampling error following Ives et al. (2007).

Usage phylosig(tree, x, method="K", test=FALSE, nsim=1000, se=NULL, start=NULL, control=list())

Author(s) Liam Revell <liam.revell@umb.edu>

References

Pagel, M. (1999) Inferring the historical patterns of biological evolution. *Nature*, **401**, 877--884.

Blomberg, S. P., T. Garland Jr., A. R. Ives (2003) Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, **57**, 717--745.

Ives, A. R., P. E. Midford, T. Garland Jr. (2007) Within-species variation and measurement error in phylogenetic comparative biology. *Systematic Biology*, **56**, 252-270.

e) *phyl.resid* {phytools}

This function fits one or multiple phylogenetic regressions (depending on the number of columns in Y) and computes the residuals. Designed for phylogenetic size correction using GLS regression (e.g., Revell 2009; Evolution).

Usage phyl.resid(tree, x, Y, method="BM")

Author(s) Liam Revell <liam.revell@umb.edu>

References

Revell, L. J. (2009) Size-correction and principal components for interspecific comparative studies. Evolution, 63,3258–3268.

Revell, L. J. (2010) Phylogenetic signal and linear regression on species data. Methods in Ecology and Evolution, 1, 319–329.

f) *pgls* {caper}

Fits a linear model, taking into account phylogenetic non-independence between data points. The strength and type of the phylogenetic signal in the data matrix can also be accounted for by adjust- ing branch length transformations (lambda, delta and kappa). These transformations can also be optimised to find the maximum likelihood transformation given the data and the model.

Usage pgls(formula, data, lambda = 1.0, kappa = 1.0, delta= 1.0, param.CI = 0.95, control = list(fnscale= pgls.likelihood(optimPar, fixedPar, y, x, V, optim.output=TRUE, names.optim=NULL) pgls.blenTransform(V, fixedPar)

Author(s) Rob Freckleton; David Orme

References

R. P. Freckleton, P. H. Harvey, and M. Pagel. Phylogenetic analysis and comparative data: A test and review of evidence. American Naturalist, 160:712-726, 2002.

g) *read.nexus* {ape}

Description This function reads one or several trees in a NEXUS file.

Usage read.nexus(file, tree.names = NULL)

Author(s) Emmanuel Paradis References

References

Maddison, D. R., Swofford, D. L. and Maddison, W. P. (1997) NEXUS: an extensible file format for systematic information. Systematic Biology, 46, 590–621.

h) *vif* {car}

Calculates variance-inflation and generalized variance-inflation factors for linear and generalized linear models.

Usage vif(mod, ...)

Author(s) Henric Nilsson and John Fox <jfox@mcmaster.ca>

References

Fox, J. and Monette, G. (1992) Generalized collinearity diagnostics. *JASA*, 87, 178–183.

Fox, J. (2008) *Applied Regression Analysis and Generalized Linear Models*, Second Edition. Sage.

Fox, J. and Weisberg, S. (2011) *An R Companion to Applied Regression*, Second Edition, Sage.

APPENDIX B: FIGURES FROM MATERIALS AND METHODS

Figure B1: Histograms of measures of DR for the 100 trees in different species (out of 9993)

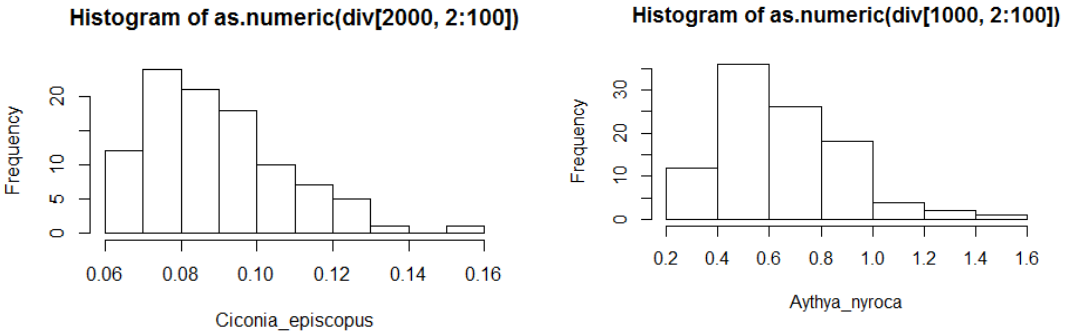


Figure B2. Lineal correlation between the measure of Diversity Rate (DR) doing the median of 50 trees from Ericson backbone compared to the 50 trees from Hackett backbone.

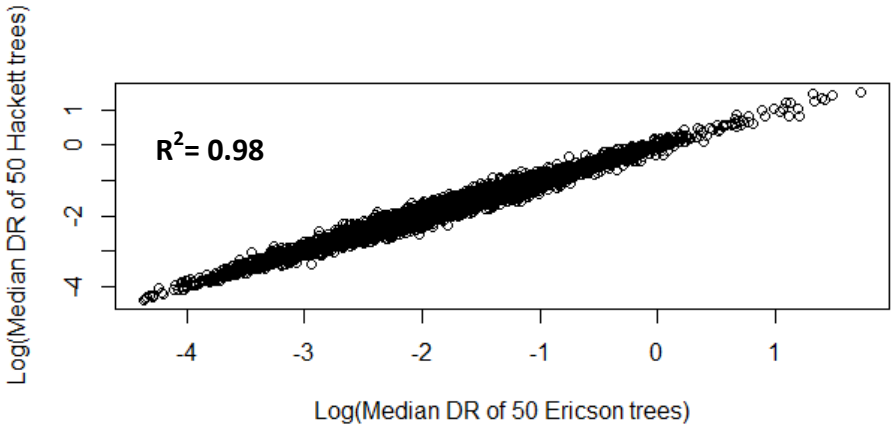


Figure B3. Histograms for Diversification rate **(A)** and Log(Diversification rate) with an adjusted Normal curve **(B)** for 9993 species of birds.

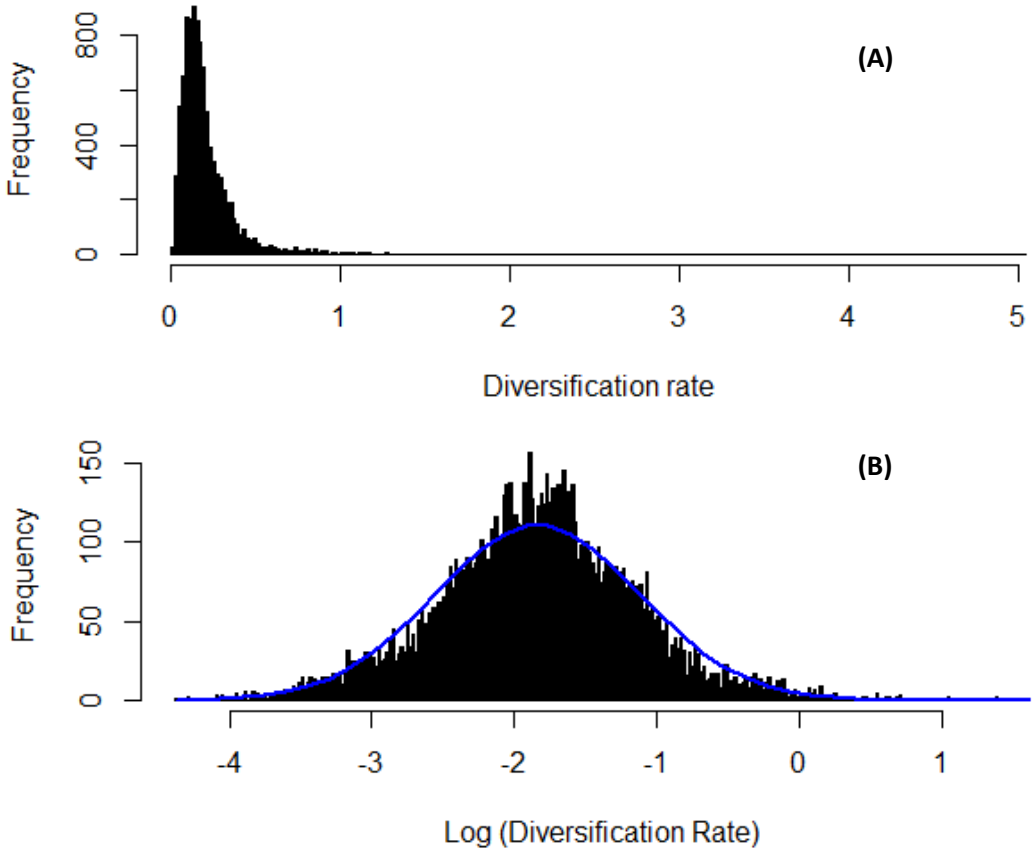


Figure B4: Model between the Log(body size) and Log(brain size) using an ordinary linear regression (red), a quadratic regression (blue) and using Revell's approach (green) [by one of the 100 trees].

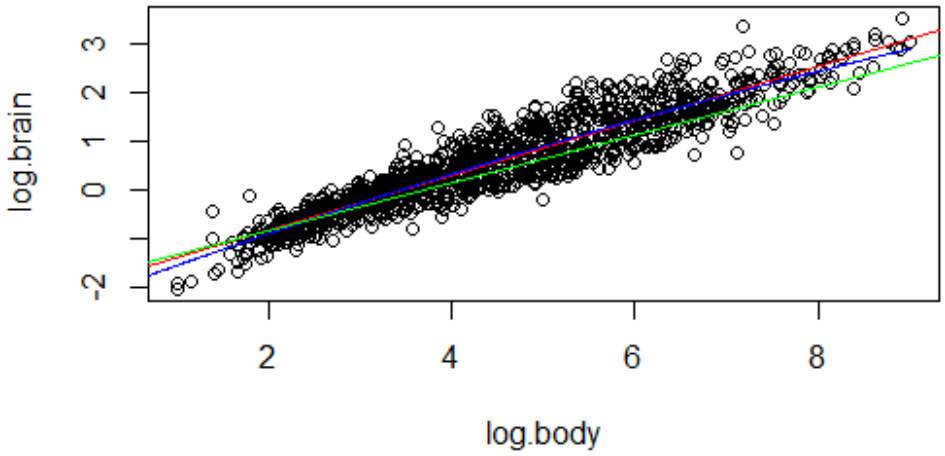
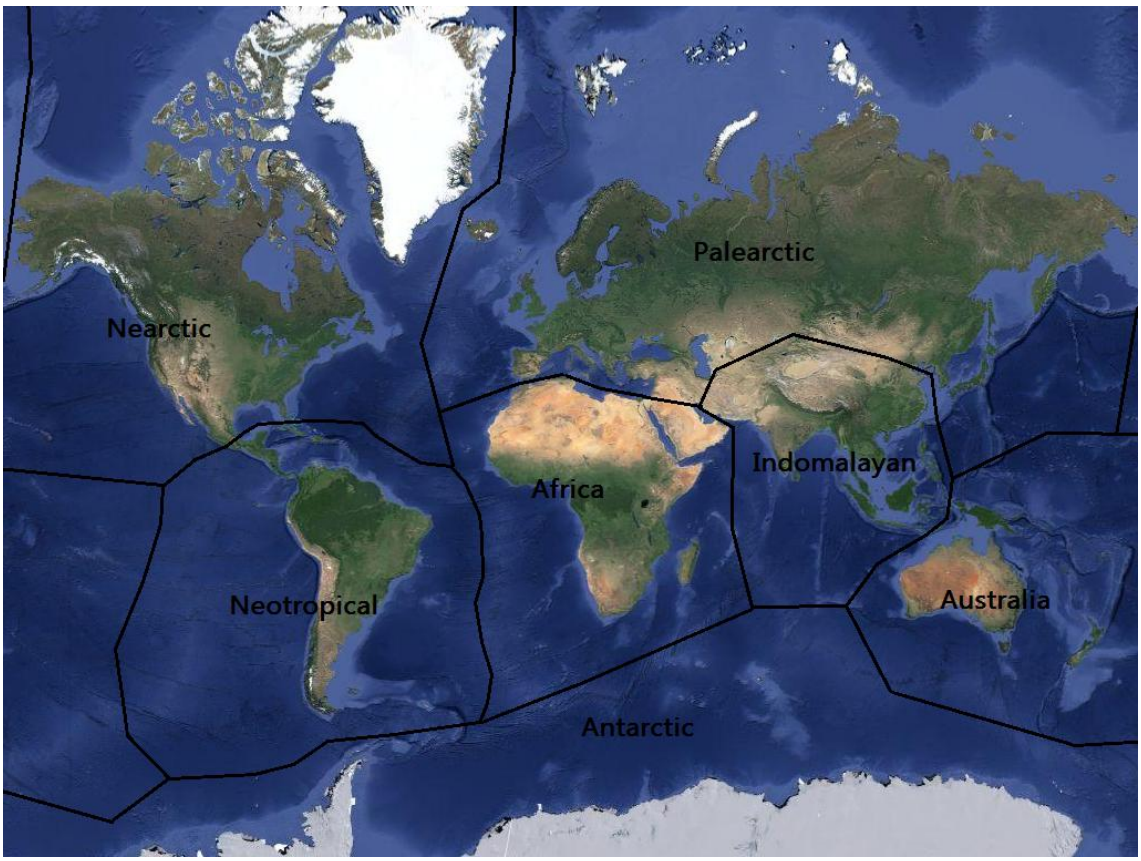


Figure B5. Biogeographical zones used to define variable *Biogeography*. If the home range of a species was between more than one region, it was categorized as *Multi-region*.



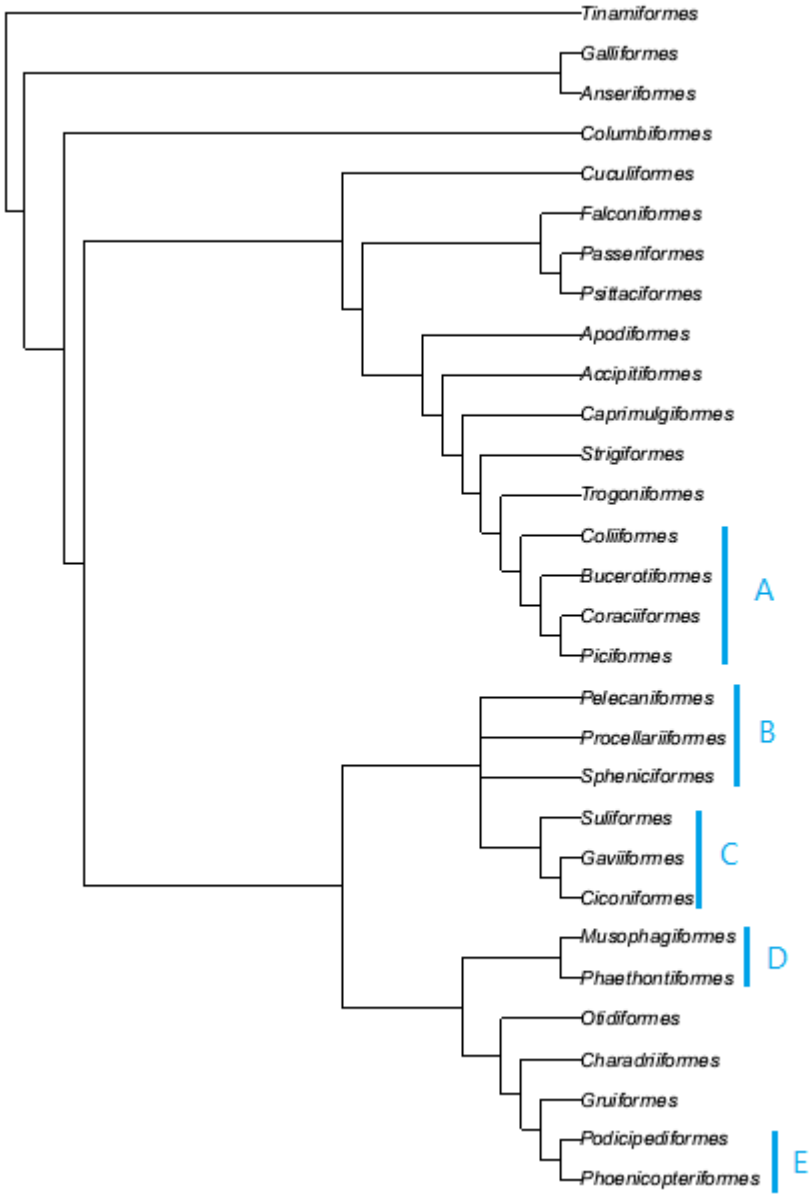


Figure B6: Phylogeny of Orders of Birds (70% Consensus Tree from 100 trees)

Figure B7. One of the 100 trees of 1326 species used in the analyses.

Table B1: Analysis of tolerance for the explanatory factors that affect brain residuals.

VIF (variance inflator factor)

	GVIF	Df	GVIF^(1/(2*Df))
migratory	1.591763	2	1.123232
insularity	1.167547	1	1.080531
gt.birdlife	1.304526	1	1.142158
biogeography	2.070686	8	1.046543
Bhabitat	1.140598	1	1.067988

Table B2. Model selection for **PGLS (package caper)** with dredge function from package (MuMIn) with N= 603 species and all variables using one random tree. Only models with difference in AICc (delta) < 4 are shown.

Model selection table

	(Int)	Habitat	bgg	brain	gener	ins	mgr	Ord	df	logLik	AICc	delta	weight
10	-2.606	0.03187			0.02232				3	-583.656	1173.4	0.00	0.122
2	-2.421	0.03105							2	-584.823	1173.7	0.31	0.105
9	-2.539				0.02147				2	-585.332	1174.7	1.33	0.063
1	-2.362								1	-586.403	1174.8	1.46	0.059
26	-2.610	0.03290			0.02219	0.07510			4	-583.412	1174.9	1.54	0.057
18	-2.427	0.03212				0.07735			3	-584.567	1175.2	1.82	0.049
14	-2.610	0.03183		-0.052490	0.02392				4	-583.589	1175.2	1.89	0.048
42	-2.550	0.02979			0.02223		+		5	-582.674	1175.4	2.10	0.043
6	-2.424	0.03107		0.012170					3	-584.819	1175.7	2.33	0.038
34	-2.366	0.02898					+		4	-583.839	1175.7	2.39	0.037
41	-2.482				0.02147		+		4	-584.123	1176.3	2.96	0.028
25	-2.540				0.02136	0.05799			3	-585.186	1176.4	3.06	0.027
58	-2.547	0.03088			0.02204	0.11580	+		6	-582.137	1176.4	3.06	0.026
33	-2.306						+		3	-585.203	1176.4	3.09	0.026
17	-2.365					0.06043			2	-586.246	1176.5	3.16	0.025
13	-2.542			-0.054000	0.02312				3	-585.262	1176.6	3.21	0.025
50	-2.365	0.03012				0.11820	+		5	-583.285	1176.7	3.32	0.023
30	-2.615	0.03290		-0.061170	0.02406	0.07888			5	-583.322	1176.7	3.39	0.022
5	-2.364			0.008431					2	-586.401	1176.8	3.47	0.022
22	-2.428	0.03212		0.004146		0.07708			4	-584.566	1177.2	3.85	0.018

APPENDIX C: Code used for the analyses

1 ## DIVERSITY RATE FOR ALL SPECIES

```
setwd("C:/Users/Ferran/Desktop/Treball_MASTER/2_Materials&Methods/Statistical_Analysis")

## 1.1 ## DIVERSIFICATION RATE MEASURE (To obtain DR from phylogeny)

library(picante) #For function evol.distinct

## DR for ERICSON Trees
First.tree<-1
Last.tree<-50
Form<-"Eri"  ## "Eri" or "Hac"

# First tree
tree<-read.tree(paste(Form,"_tree",First.tree,".tre",sep=""))
ES<-evol.distinct(tree, type = "equal.splits", scale = FALSE, use.branch.lengths = TRUE)
cat("Tree",First.tree,"/",Last.tree,"\n")
div.rate <- data.frame("Species"=ES$Species,"dr1"=(1/ES$w))
names(div.rate)[2]<-paste("dr",First.tree,sep="") #change name of column 2

# The other trees

for (i in (First.tree+1):Last.tree) {
 tree<-read.tree(paste("Eri_tree",i,".tre",sep=""))
 ES<-evol.distinct(tree, type = c("equal.splits", "fair.proportion"), scale = FALSE, use.branch.lengths = TRUE)
 DR<-data.frame("Species"=ES$Species,"dr"=(1/ES$w)) #Create a data.frame (Species, Inverse of Equal.Splits)
 names(DR)[2]<-paste("dr",i,sep="") #change name of column 2 in DR
 cat("Tree",i,"/",Last.tree,"\n")
 div.rate<-merge(x=div.rate,y=DR,by="Species") #merge two datasets by "Species" column
} # end for i
cat ("JA HE ACABAT")

write.table(x=div.rate,file=paste("DiversityRate_",Form,First.tree,"_",Last.tree,".txt",sep=""))
cat("\n","JA POTS TANCAR")

### Doing Mean or median for the 100 trees

## Diversity Rate for all species (100 trees)   (Bellow are scripts for calculating DR)
div<-read.csv("DiversificationRate/DiversityRate_100trees.csv",header=T,sep=";")
head(div)
ls(div)

# Histogram of DR for 100 trees of each species
br<-20
hist(as.numeric(div[1,2:100]),xlab=div[1,1],breaks=br) # [rows,columns]
hist(as.numeric(div[1000,2:100]),xlab=div[1000,1],breaks=br) # [rows,columns]
hist(as.numeric(div[2000,2:100]),xlab=div[2000,1],breaks=br) # [rows,columns]
hist(as.numeric(div[3000,2:100]),xlab=div[3000,1],breaks=br) # [rows,columns]

min(div$Median.eri)
max(div$Median.eri)
min(div$Median.hac)
max(div$Median.hac)
hist(div$Median.eri,breaks=300)
hist(div$Median.hac,breaks=300)

plot(log(div$Median.eri),log(div$Median.hac),xlab="Log(Median DR of 50 Ericson trees)",ylab="Log(Median DR of
50 Hackett trees)") #Eri and Hac are equivalent
m<-lm(log(div$Median.eri)~log(div$Median.hac))
```

```
summary(m)

hist(div$Median,breaks=300,col="black",xlab="Diversification rate",main="")

x<-log(div$Median)
h<-hist(x,breaks=300,col="black",xlab="Log (Diversification Rate)",main="")

xfit<-seq(min(x),max(x),length=40)
yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))
yfit <- yfit*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col="blue", lwd=2)

hist(fdata$log.dr,breaks=50,xlab="Log (Diversity Rate)",main="Histogram of Log(DR) for 1326 sps")


# Phylogenetic signal of Diversity Rate

library (phytools)
library (geiger)
library(ape)

#tree1<-read.tree("Data/tree_9993.tre") # One tree for all birds species

### FOR FDATA
fdata<-read.table("Data_1326sps.csv",sep=";") # Full data for 1326 sps

#tree<-read.nexus("Data/tree_288.tre") # ndata tree
tree<-read.nexus("Data/tree_1326.tre") #fdata tree
ls(fdata)
#div.species<-ndata$median.dr
div.species<-fdata$median.dr
#names(div.species)<-ndata$species #div rate data with species label
names(div.species)<-fdata$species #div rate data with species label

lambdas<-numeric(100)
for (i in 1:100) {
  signal<-phylosig(tree=tree[[i]],x=div.species,"lambda")
  lambdas[i]<-signal$lambda
  print(format(Sys.time(), "%H:%M:%S"))
  cat("tree",i,"/",100,"lambda=",signal$lambda,"\n")
} # end for i

#save("lambdas",file="Results/Lambda_1326sps.Rdata")
load("Results/Lambda_1326sps.Rdata")
mean(lambdas)
sd(lambdas)
hist(lambdas,xlab="Phylogenetic signal (Lambda)",main="",breaks=10)

dens<-density(lambdas)
plot(dens,xlab="Lambda for Diversification Rate",main="",col="blue")
mean(lambdas)
sd(lambdas)
```

2 ## ESTIMATING BRAIN RESIDUALS

```
## INDEX
###### 2.1 Ordinary Regressions (linial and quadratic)
###### 2.2 Revell's approach
###### 2.3 Using a PGLS
###### 2.4 Different slopes for each group (smatr)
```

```
###### 2.5 Phylogenetic Independent Contrast

# Opening Data (fdata is data of 1326 species)

setwd("C:/Users/Ferran/Desktop/Treball_MASTER/2_Materials&Methods/Statistical_Analysis")

fdata<-read.table("Data/Brains_Iwaniuk.txt",header=T)
head(fdata)

lbrain<-log(fdata$brain)
lbody<-log(fdata$body)

fdata<-cbind(fdata,lbody,lbrain)
head(fdata)
## 2.1 # Using ordinary regressions

plot(fdata$lbody,fdata$lbrain)

model1.reg <- lm(fdata$lbrain ~ fdata$lbody)
model2.reg <- lm(fdata$lbrain ~ fdata$lbody+ I(fdata$lbody**2))
res.reg1<- resid(model1.reg)
res.reg2<- resid(model2.reg)
AIC(model1.reg,model2.reg)
summary(model1.reg)

abline(a=-1.971428, b=0.562818, col = "red") # ordinary regression

summary(model2.reg)

# ndata<-cbind(ndata,res.reg1,res.reg2)

## we cannot use abline to plot the non-linear regression line (because it is not straight);
## the easiest and most general solution is to obtain model predictions for different values of X

z <- seq(0.01,9,0.001)
z <- sample(z,1326)
z <- z[order(z)]

myfun <- function(newdist, model2.reg) {
  coefs <- coef(model2.reg)
  res <- coefs[1] + (coefs[2] * newdist) + (coefs[3] * newdist^2)
  return(res)
}

w <- myfun(z, model2.reg)
points(z, w, type="l", col="blue", lwd=1)


## 2.2 # REVELL'S APPROACH
# We can finally calculate the residuals using Revell's approach
plot(tree[[1]])
library(phytools)
tree<-read.nexus("Data/tree_1326.tre")
ls(fdata)
log.brain<-fdata$lbrain
names(log.brain)<-fdata$species
log.body<-fdata$lbody
names(log.body)<-fdata$species

### Calculate Brain residuals for the 100 trees
```

```
for (i in 1:100) {
  res.bm<-phyl.resid(tree[[i]],x=log.body,Y=log.brain) #Function phyl.resid
  species<-row.names(res.bm$resid)
  residuals<-res.bm$resid[1:1326]
  table<-data.frame("species"=species,"res"=as.numeric(residuals))
  names(table)[2]<-paste("res",i,sep="")
  fdata<-merge(fdata,table,by="species")
  cat("Tree",i,"out of 100","/n")
} # end for i

ls(fdata)
hist(as.numeric(fdata[1,6:105]),n=20)
hist(as.numeric(fdata[100,6:105]),n=20)
hist(as.numeric(fdata[200,6:105]),n=20)
hist(as.numeric(fdata[300,6:105]),n=20)
hist(as.numeric(fdata[400,6:105]),n=20)
hist(as.numeric(fdata[500,6:105]),n=20)
hist(as.numeric(fdata[800,6:105]),n=20)

write.table(fdata,"Revell_Residuals.csv",sep=";")

plot(log.body,log.brain)
abline(-1.9827243,0.5110845,col="red")

tree<-read.nexus("Data/tree_1326.tre")

nrow(fdata)
brain.residual<-fdata$brain.res
names(brain.residual)<-fdata$species #Brain.res data with species label

lambdas<-numeric(100)
for (i in 1:100) {
  signal<-phylosig(tree=tree[[i]],x=brain.residual,"lambda")
  lambdas[i]<-signal$lambda
  print(format(Sys.time(), "%H:%M:%S"))
  cat("tree",i,"/",100,"lambda=",signal$lambda,"\n")
} # end for i

save("lambdas",file="Results/LambdaBRAIN_1326sps.Rdata")
#load("Results/LambdaBRAIN_1326sps.Rdata")
mean(lambdas)
sd(lambdas)
hist(lambdas,xlab="Phylogenetic signal (Lambda)",main="Histogram of lambda for 100 trees & 1326
sps",breaks=10)

dens<-density(lambdas)
plot(dens,xlab="Lambda value",main="Lambda distribution of Brain Residual over 100 trees",col="blue")
mean(lambdas)
sd(lambdas)

## 2.3 # USING A PGLS

library(ape)
library(caper)

tree<-read.nexus("Data/tree_1326.tre") # 100 phylogenies from Hackett & Ericson
brain.pgls <- comparative.data(phy=tree[[10]],data=fdata,names.col=species, vcv=TRUE, na.omit = FALSE,
warn.dropped = TRUE, vcv.dim=3)

model.pgls <- pgls(lbrain ~ lbody, data=brain.pgls, lambda="ML")
summary(model.pgls)
```

```
res.pgls<-resid(model.pgls)

#brain.residuals<-data.frame("species"=fdata$species,res.reg1,res.reg2,res.bm,res.pgls)
#save("brain.residuals",file="Data/Brain_residuals.Rdata")
```

3 ## MERGING DATA SOURCES AND EXPLORING THE DATA

```
# fdata<-read.table("Data/Ecology_1326sps.csv",header=T,sep=";")
fdata<-read.table("Data_1326sps.csv",sep=";",header=T)

plot(log(fdata$gt.birdlife),fdata$brain.res)
mod<- lm(fdata$brain.res~log(fdata$gt.birdlife))
summary(mod)
abline(-0.11334,0.21135,col="blue")

### Phylogeny Orders (Building Consensus Tree)

library(ape)
tree<-read.nexus("Data/tree_orders.tre")
plot(tree[[15]],cex=0.5)

tree1<-consensus(tree,p=0.5,check.labels=T)
plot(tree1,cex=0.7)

## Plotting trees

plot.phylo(tree1,"phylogram",cex=0.8,direction="upwards")
plot.phylo(tree1,"fan",cex=0.5)


### Merging Brain Residuals
brain<-read.table("Data/Revell_Residuals.csv",header=T,sep=";")
brain.residuals<-data.frame("species"=brain$species,"brain.res"=brain$brain.res)
fdata<-merge(fdata,brain.residuals,by="species")
ls(fdata)

write.table(fdata,"Data_1326sps.csv",sep=";")

### Merging with Ecology and Behavior to see the sources

sdata<-data.frame("species"=fdata$species)

load(file="Data/Original_data.Rdata")
ls()
dEco$species<-gsub(" ","_",dEco$species) #put a "_" between species names
dEco<-merge(sdata,dEco,by="species",all.x=T,)
nrow(dEco)
write.table(dEco,file="Ecology_1326.csv")

unique(dEco$variable)
Insul<-subset(dEco,variable=="Insularity")
unique(Insul$source) # Sources Insularity

dBeh$species<-gsub(" ","_",dBeh$species) #put a "_" between species names
View(dBeh)
dBeh<-merge(sdata,dBeh,by="species",all.x=T,)
nrow(dBeh)

unique(dBeh$variable)
Migrat<-subset(dBeh,variable=="MigratoryBehavior")
```

```
unique(Migrat$source) # Sources Migratory

## Plotting Multivariate Data
head(fdata)
library(car)
scatterplotMatrix(fdata[6:10])

attach(fdata)

plot(migratory,log.dr)
plot(insularity,log.dr)
plot(log(distribution.size),log.dr)
plot(log(gt.birdlife),log.dr)
plot(brain.res,log.dr)

m<-lm(log.dr~migratory+insularity+log(distribution.size)+gt.birdlife+brain.res)
m1<-lm(log.dr~migratory+log(distribution.size)+brain.res)
m2<-lm(log.dr~migratory+brain.res)
m3<-lm(log.dr~brain.res+I(brain.res**2))
anova(m3)

## PROFILE PLOT (Function below!!!!!!!!)
detach(fdata)
fdata<-na.omit(fdata)
nrow(fdata)
attach(fdata)
library(RColorBrewer)
names <- c("Brain.res","Log.dr")
mylist <- list(brain.res,log.dr)
makeProfilePlot(mylist,names)

#### Outliers of Log.dr
hist(fdata$log.dr,n=25) # There is a big rigth cue
fdata<-subset(fdata,log.dr<0)
nrow(fdata)

fdata<-read.table(file="data_1309sps.csv",sep=";")

outliers<-subset(fdata,log.dr>0)
outliers$species

plot(out.dr$brain.res,out.dr$log.dr)
plot(fdata$brain.res,fdata$log.dr)

hist(fdata$brain.res,n=20)
hist(log(fdata$gt.birdlife),n=20)
hist(log(fdata$distribution.size))

## DIVERSITY RATE BY GROUPS

fdata<-read.csv("Data_1326sps.csv",header=T,sep=";")
ls(fdata)
fdata<-merge(fdata,median.dr,by="species")

Mgroup.dr<-aggregate(fdata$log.dr,list(group=fdata$group), FUN=mean)
Mgroup.dr #Means of the hole group
Morder.dr<-aggregate(fdata$log.dr,list(order=fdata$order), FUN=mean)
Morder.dr #Means of the hole order

##### INSULAR SPECIES ONLY
fdata<-read.table(file="data_1309sps.csv",sep=";",header=T)
```

```
island<-subset(fdata,fdata$insularity==1)
nrow(island)
passer<-subset(fdata,fdata$order=="PASSERIFORMES")

plot(island$brain.res,island$log.dr)
plot(passer$brain.res,passer$log.dr)

## PCA
fdata$lbody<-as.numeric(fdata$lbody)
fdata$lbrain<-as.numeric(fdata$lbrain)
fdata$insularity<-as.numeric(fdata$insularity)
fdata$migratory<-as.numeric(fdata$migratory)
fdata$Bhabitat<-as.numeric(fdata$Bhabitat)
fdata$Bdiet<-as.numeric(fdata$Bdiet)
fdata$gt.birdlife<-as.numeric(fdata$gt.birdlife)
fdata$distribution.size<-as.numeric(fdata$distribution.size)
fdata$res.bm<-as.numeric(fdata$res.bm)

head(fdata)
fdata<-na.omit(fdata)
nrow(fdata)
stn.fdata <- as.data.frame(scale(fdata[c(7,9,11,20)])) # standardise the variables
fdata.pca <- prcomp(stn.fdata)  # do a PCA
summary(fdata.pca)
biplot(fdata.pca)


### Function to Profile Plot

makeProfilePlot <- function(mylist,names)
{
 require(RColorBrewer)
 # find out how many variables we want to include
 numvariables <- length(mylist)
 # choose 'numvariables' random colours
 colours <- brewer.pal(numvariables,"Set1")
 # find out the minimum and maximum values of the variables:
 mymin <- 1e+20
 mymax <- 1e-20
 for (i in 1:numvariables)
 {
  vectori <- mylist[[i]]
  mini <- min(vectori)
  maxi <- max(vectori)
  if (mini < mymin) { mymin <- mini }
  if (maxi > mymax) { mymax <- maxi }
 }
 # plot the variables
 for (i in 1:numvariables)
 {
  vectori <- mylist[[i]]
  namei <- names[i]
  colouri <- colours[i]
  if (i == 1) { plot(vectori,col=colouri,type="l",ylim=c(mymin,mymax)) }
  else        { points(vectori, col=colouri,type="l")                  }
  lastxval <- length(vectori)
  lastyval <- vectori[length(vectori)]
  text((lastxval-10),(lastyval),namei,col="black",cex=0.6)
 }
}
```

4 ## ORDINARY AND PGLS ANALYSIS

4.1 ## LINEAL MODEL APPROACH

```
library(MuMIn)

setwd("C:/Users/Ferran/Desktop/Treball_MASTER/2_Materials&Methods/Statistical_Analysis")

fdata<-read.table("Data_1326sps.csv",sep=";",header=T) # Full data for 1326 sps
ls(fdata)

## Variability in the data
table(fdata$insularity)
table(fdata$migratory)
table(fdata$group)
table(fdata$order)

ndata<-na.omit(fdata)
nrow(ndata)

attach(fdata)
model.selection<-dredge(lm(log.dr ~
brain.res+I(brain.res**2)+Bhabitat+migratory+insularity+gt.birdlife+biogeography+order, data=fdata))
model.selection
m1<-lm(log.dr ~ brain.res+I(brain.res**2)+order+migratory)
m1<-lm(log.dr ~ brain.res*order+I(brain.res**2)*order+migratory)
m2<-lm(log.dr ~ brain.res*order+I(brain.res**2)+migratory)
AIC(m,m1,m2)
summary(m1)
anova(m1)
influence(m1)
plot(m1)

## Plot Quadratic
plot(fdata$brain.res,fdata$log.dr,xlab="Brain residuals",ylab="Diversity Rate")

#### Abline Quadratic line

min(fdata$brain.res)

z <- seq(-1,2,0.001)
z <- sample(z,1326)
z <- z[order(z)]

myfun <- function(newdist, m1) {
 coefs <- coef(m1)
 res <- coefs[1] + (coefs[2] * newdist) + (coefs[3] * newdist^2)
 return(res)
}

w <- myfun(z, m1)
points(z, w, type="l", col="blue", lwd=1)

##### The best lineal model (The most simple)
attach(fdata)
m1<-lm(log.dr~res.bm+I(res.bm**2)+migratory+order)
anova(m1)

m2<-lm(log.dr~res.bm*order+I(res.bm**2)+migratory)
m3<-lm(log.dr~res.bm+I(res.bm**2)+group)
```

```
anova(m2)
plot(res.bm,log.dr)
AIC(m2,m3)
summary(m3)
anova(m3)

## 4.2 # MODEL SELECTION PROCESS FOR PGLS ANALYSES

library(caper)
library(MuMIn)

fdata<-read.table("Data_603sps.csv",sep=";",header=T) # Full data for 1326 sps
ls(fdata)

## Variability in the data
table(fdata$insularity)
table(fdata$migratory)
table(fdata$group)
table(fdata$order)
table(fdata$Morder)

# Model Selection
library(ape)
library(caper)
library(MuMIn)

nrow(fdata)
fdata<-na.omit(fdata)
rm(brain,brain.residuals)
tree<-read.nexus("Data/tree_1326.tre")
print(format(Sys.time(), "%H:%M:%S"))
comp.data <- comparative.data(phy=tree[[71]],data=fdata,names.col=species, vcv=TRUE, na.omit = FALSE,
warn.dropped = TRUE, vcv.dim=3)

print(format(Sys.time(), "%H:%M:%S"))
model.selection<-dredge(pgls(log.dr ~
brain.res+I(brain.res**2)+Bhabitat+migratory+insularity+gt.birdlife+biogeography+Morder, data=comp.data,
lambda="ML"),fixed="brain.res")
print(format(Sys.time(), "%H:%M:%S"))
model.selection

## 4.3 ## MOST IMPORTANT VARIABLES. FINAL MODELS

library(caper)
tree<-read.nexus("Data/tree_1326.tre")
comp.data <- comparative.data(phy=tree[[5]],data=fdata,names.col=species, vcv=TRUE, na.omit = FALSE,
warn.dropped = TRUE, vcv.dim=3)
rm(tree,fdata)

## HABITAT
pgls1<-pgls(log.dr ~ brain.res+Bhabitat, data=comp.data, lambda="ML")
pgls2<-pgls(log.dr ~ Bhabitat, data=comp.data, lambda="ML")
AIC(pgls1,pgls2)
summary(pgls1)
summary(pgls2)

## GEN TIME
pgls1<-pgls(log.dr ~ brain.res+gt.birdlife, data=comp.data, lambda="ML")
pgls2<-pgls(log.dr ~ gt.birdlife, data=comp.data, lambda="ML")
AIC(pgls1,pgls2)
summary(pgls1)
```

```
summary(pgls2)
rm(pgls1,pgls2)


## INSULARITY
pgls1<-pgls(log.dr ~ brain.res+insularity, data=comp.data, lambda="ML")
pgls2<-pgls(log.dr ~ insularity, data=comp.data, lambda="ML")
AIC(pgls1,pgls2)
summary(pgls1)
summary(pgls2)
rm(pgls1,pgls2)

## MIGRATORY
pgls1<-pgls(log.dr ~ brain.res+migratory, data=comp.data, lambda="ML")
pgls2<-pgls(log.dr ~ migratory, data=comp.data, lambda="ML")
AIC(pgls1,pgls2)
summary(pgls1)
summary(pgls2)
rm(pgls1,pgls2)

## BRAIN RESIDUALS
pgls1<-pgls(log.dr ~ brain.res, data=comp.data, lambda="ML")
pgls2<-pgls(log.dr ~ I(brain.res^2), data=comp.data, lambda="ML")
AIC(pgls1,pgls2)
summary(pgls1)
anova(pgls1)
summary(pgls2)
rm(pgls1,pgls2)

pgls3<-pgls(log.dr ~ brain.res+I(brain.res^2), data=comp.data, lambda="ML")
AIC(pgls1,pgls3)
summary(pgls3)
rm(pgls3)

## ALL

pgls3<-pgls(log.dr ~ Bhabitat+gt.birdlife+insularity+migratory, data=comp.data, lambda="ML")
summary(pgls3)
AIC(pgls3)
rm(pgls3)
pgls3<-pgls(log.dr ~ Bhabitat+gt.birdlife+migratory, data=comp.data, lambda="ML")
summary(pgls3)
AIC(pgls3)
rm(pgls3)
```