SUPPORTING INFORMATION

A: Analytical Justification: Alternate Methods for Second-Moment Exchangeability

The method of phylogenetic transformation was described by Garland and Ives (2000), based on the well-known OLS method of coefficient estimation for GLS models (Judge et al. 1985; Johnston and DiNardo 1997; Rencher 2000; Kariya and Kurata 2004). For an $n \times n$ phylogenetic covariance matrix $\mathbf{C}$, there exists a solution, $\mathbf{PCP}^{\boldsymbol{T}} = \mathbf{I}$, where the superscript, $^{\boldsymbol{T}}$, means matrix transposition, and $\mathbf{P}$ is an $n \times n$ orthogonal phylogenetic transformation matrix, which can be calculated as $\mathbf{P} = \mathbf{UW}^{\mathbf{1/2}}\mathbf{U}^{\boldsymbol{T}}$, based on the eigenvectors ($\mathbf{U}$) and eigenvalues ($\mathbf{W}$) from eigenanalysis of $\mathbf{C}$. Thus, an OLS model can be described as: $\mathbf{Y}_{\boldsymbol{phy}} = \mathbf{X}_{\boldsymbol{phy}}\mathbf{B} + \mathbf{E}_{\boldsymbol{phy}}$, with $\mathbf{Y}_{\boldsymbol{phy}} = \mathbf{PY}$, $\mathbf{X}_{\boldsymbol{phy}} = \mathbf{PX}$, and $\mathbf{E}_{\boldsymbol{phy}} = \mathbf{PE}$; where $\mathbf{Y}$ is an $n \times p$ matrix of observations for n taxa with p variables (each taxon represented by a $1 \times p$ vector of values); $\mathbf{X}$ is a $n \times k$ matrix, where k is the number of model parameters, equal to the matrix rank in the absence of parameter redundancy; and $\mathbf{E}$ is an $n \times p$ matrix of residuals. Phylogenetic transformation does not alter the dimensions of these matrices. The $k \times p$ matrix of coefficients, $\mathbf{B}$, is estimated as, $\widehat{\mathbf{B}} = \left(\mathbf{X}_{\boldsymbol{phy}}^{\boldsymbol{T}}\mathbf{X}_{\boldsymbol{phy}}\right)^{\mathbf{-1}}\mathbf{X}_{\boldsymbol{phy}}^{\boldsymbol{T}}\mathbf{Y}_{\boldsymbol{phy}}$, which is the same as the computationally more intensive solution, $\widehat{\mathbf{B}} = (\mathbf{X}^{\boldsymbol{T}}\mathbf{C}^{\mathbf{-1}}\mathbf{X})^{\mathbf{-1}}\mathbf{X}^{\boldsymbol{T}}\mathbf{C}^{\mathbf{-1}}\mathbf{Y}$, where the superscript, $^{\mathbf{-1}}$, means matrix inversion.

Garland and Ives (2000) demonstrated that phylogenetic transformation of residuals produces uncorrelated error, as expected with a linear model; namely, $E\{\mathbf{E}_{\boldsymbol{phy}}\boldsymbol{E}_{\boldsymbol{phy}}^{\boldsymbol{T}}\} = E\{\mathbf{PE}(\mathbf{PE})^{\boldsymbol{T}}\} = \mathbf{P}E\{\mathbf{EE}^{\boldsymbol{T}}\}\mathbf{P}^{\boldsymbol{T}} = \mathbf{P}\boldsymbol{\sigma}^{\mathbf{2}}\mathbf{CP}^{\boldsymbol{T}} = \boldsymbol{\sigma}^{\mathbf{2}}\mathbf{I}$. Adams and Collyer (2015) reasoned that the $\mathbf{P}E\{\mathbf{EE}^{\boldsymbol{T}}\}\mathbf{P}^{\boldsymbol{T}}$ state of this equation was important for recognizing how a randomized residual permutation procedure (RRPP) should be performed. By randomizing the row vectors of $\mathbf{E}$, to a shuffled matrix of residuals, $\mathbf{E}^{*}$, $E\{\mathbf{EE}^{\boldsymbol{T}}\}$ remains constant because $\mathbf{EE}^{\boldsymbol{T}} = \mathbf{E}^{*}(\mathbf{E}^{*})^{\boldsymbol{T}}$. This philosophy for RRPP attempts to maintain second-moment exchangeability (constant variance), as global (first-moment, or constant mean) exchangeability is not possible unless a linear model contains only an intercept and observations are independent (such that $\mathbf{C} = \mathbf{I}$; Commenges 2003). (For further details, see Section C.)

However, the focus on preserving $E\{\mathbf{EE}^{\boldsymbol{T}}\}$ in this case fails to preserve second-moment exchangeability of transformed residuals; specifically, $E\left\{\mathbf{E}_{\boldsymbol{phy}}^{*}\left(\mathbf{E}_{\boldsymbol{phy}}^{*}\right)^{\boldsymbol{T}}\right\} \neq \mathbf{P}E\{\mathbf{EE}^{\boldsymbol{T}}\}\mathbf{P}^{\boldsymbol{T}}$. If residuals are inherently correlated, treating transformed residuals as exchangeable units rather than the observed residuals has been argued to have better type I error rates (Commenges 2003). In the case of linear models, especially where GLS solutions are appropriate, producing non-correlated error and preserving second-moment exchangeability via transformation of residuals appears to be the better strategy for RRPP, particularly if distributions of statistics are robust against departures from global exchangeability. For PGLS, there does not exist a RRPP method that preserves global exchangeability (even if the linear model contains only an intercept), but using an argument from consilience, we demonstrate below that empirical evidence suggests that randomizing transformed residuals yields asymptotically exact (approximate) F distributions over more cases than randomizing untransformed residuals, consistent with the suggestion of Commenges (2003). Further, as shown in the main text and as shown below, such an RRPP approach is insensitive to aggregated groups relative to the phylogeny.

An alternative and important way to appreciate residual transformation prior to RRPP is as a GLS generalization of the OLS solution. For any generalization to be appropriate, the simpler case must hold when the generalization is applied. For example, solving linear model coefficients via GLS is, $\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\boldsymbol{T}}\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\boldsymbol{T}}\mathbf{C}^{-1}\mathbf{Y}$. If $\mathbf{C} = \mathbf{I}$, then this GLS solution is the OLS solution, $\hat{\boldsymbol{\beta}} =$

$(\mathbf{X}^{\boldsymbol{T}}\mathbf{IX})^{-1}\mathbf{X}^{\boldsymbol{T}}\mathbf{IY} = (\mathbf{X}^{\boldsymbol{T}}\mathbf{X})^{-1}\mathbf{X}^{\boldsymbol{T}}\mathbf{Y}$. Likewise, if $\mathbf{C} = \mathbf{I}, \mathbf{P} = \mathbf{I}$, and therefore, $E\left\{\mathbf{E}^*_{\boldsymbol{phy}}\left(\mathbf{E}^*_{\boldsymbol{phy}}\right)^{\boldsymbol{T}}\right\} = E\{(\mathbf{PE})^*((\mathbf{PE})^*)^{\boldsymbol{T}}\} = E\{\mathbf{E}^*(\mathbf{E}^*)^{\boldsymbol{T}}\} = E\{\mathbf{EE}^{\boldsymbol{T}}\} = \mathbf{P}E\{\mathbf{EE}^{\boldsymbol{T}}\}\mathbf{P}^{\boldsymbol{T}} = \boldsymbol{\sigma}^{\mathbf{2}}\mathbf{I}$. Thus, transforming residuals prior to RRPP is commensurate with RRPP, as performed on OLS models, meaning this GLS generalization retains the OLS solution for cases where $\mathbf{C} = \mathbf{I}$. As an appropriate generalization, and because transformation of residuals offers better statistical behavior under more circumstances, it is a heuristically preferred step in RRPP.

Because Adams and Collyer (2015) focused on linear regression models and found appropriate type I error rates, this point was not appreciated until now. Furthermore, Goolsby (2016) identified inflated type I error rates for PGLS with RRPP using pure birth trees, which Adams and Collyer (2018) suggested was an issue associated more so with the method of random tree generation than the empirical *F*-distributions generated by RRPP. However, the results in this study suggest that randomizing untransformed residuals might have either caused or exacerbated type I error rate issues, but maintaining second-moment exchangeability via transformation, appears to alleviate this concern. These issues, along with the inflated type I error rates we found with aggregated indicator values emphasizes a problem that cannot be ignored; despite the best methods available, there is no panacea for the strong interplay between phylogenetic covariance and ecological variables in these models.

However, because RRPP with transformed residuals retains appropriate second-moment exchangeability, displays appropriate type I error, and yields asymptotically exact *F* distributions (results below), it offers the most generality for the statistical designs of interest to the evolutionary biologist when conditioning data on the phylogeny.

B: Statistical Properties of the Refined Permutation Procedure

Here we use computer-generated simulation experiments to evaluate the statistical properties of RRPP with residual transformation (henceforth, simply "RRPP") for phylogenetic GLS models. A number of statistical properties (criteria) are important to consider. First, parameter estimates (F_{obs}, β_{obs}) should be identical to those found from other GLS implementations. Additionally, there should be minimal bias in estimates of β_{obs} across a range of "true" β simulated ($\beta_{\mathrm{input})}$. With respect to permutation, RRPP should generate empirical sampling distributions of F that match the theoretical expectation under ideal conditions (normal, independent, homoscedastic error). Finally, the method should be rotation-invariant, be insensitive to levels of trait covariation in the response variables, and statistical tests based on the approach should have appropriate type I error and statistical power. All of these criteria are evaluated individually, below.

1: Summary of simulation results

In the simulations to follow, we demonstrate that RRPP yields identical parameter estimates (F_{obs}, β_{obs}) compared to alternative GLS implementations. Additionally, the method displays minimal bias in its estimates of β_{obs} across a range of β_{input}, and for phylogenetic regression, the procedure displays significance levels that are highly correlated with earlier methods. Further, the empirical sampling distribution generated from RRPP matches the theoretical F distribution under a wide range of conditions. Finally, the procedure is rotation-invariant, insensitive to levels of covariation in the response variables, and displays appropriate type I error rates and high power across a range of ANOVA conditions. Combined with the analytical justification above (Section A), these empirical findings demonstrate that RRPP represents a more general permutation procedure for the analysis of empirical data for models of phylogenetic ANOVA and regression.

2. Comparison of parameter estimates: Phylogenetic regression

First we examine the case of phylogenetic regression, and compare parameter estimates from the currently proposed method of RRPP to those from previous permutation procedures (Adams 2014; Adams and Collyer 2015). (The chief difference in RRPP methods was that previous methods either randomized observed values or residuals prior to phylogenetic transformation.) The simulation protocol was as follows. First, a pure-birth tree containing N = 32 taxa was generated using the ‘pbtree’ function in the R-package *phytools* 0.6-20 (Revell 2012), and 1000 datasets of an independent (X) variable were simulated on that tree under a Brownian motion model of evolution using the ‘fastBM’ function (*phytools*: Revell 2012). Likewise, 1000 datasets of a dependent (Y) variable were simulated under Brownian motion using the same procedure. Parameter estimates (F_{obs}, β_{obs}) were obtained, and significance (P-value < 0.05) was determined using 999 random permutations (that along with the observed case totaled 1,000 random permutations). All estimates were obtained using both the existing permutation procedure in ‘procD.pgls’ of *geomorph* 3.0.5 (Adams et al. 2017) and RRPP, as described above.

__Results:__ Results (Fig. S1) reveal that RRPP yields identical parameter estimates (F_{obs}, β_{obs}) to those found with earlier implementations. Additionally, for regression, significance levels were similar to, and were highly correlated with, earlier estimates (r = 0.976 for results in Fig. S1). Notably, the type I error rate for RRPP was closer to the nominal rate of α = 0.05 than was the type I error rate of earlier implementations (Type I = 0.048 vs. Type I = 0.068, respectively). Taken together these results demonstrate that RRPP correctly estimates model parameters, and for the case of regression yields virtually identical significance estimates as compared with previous implementations, with slightly improved type I error rates.

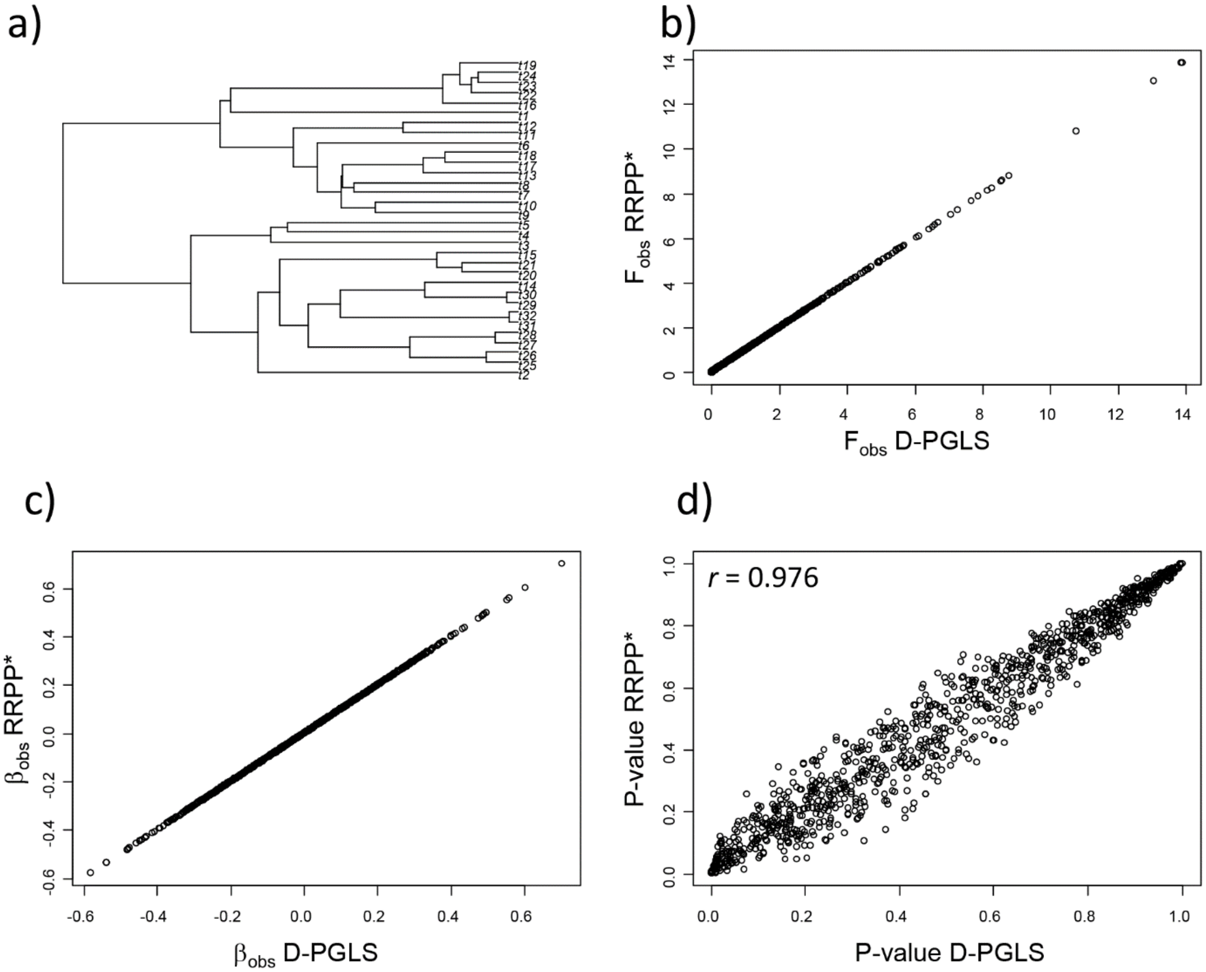

Fig. S1. Results from computer simulations comparing parameter estimates for phylogenetic regression for existing procedures and RRPP. Results based on 1,000 BM datasets evolved on: a) a pure-birth phylogeny, b) comparison of *F*-values, c) comparison of regression coefficients (β), and d) comparison of *P*-values.

3. Bias in parameter estimates: Phylogenetic ANOVA

Net we examined bias in parameter estimates from RRPP for the case of single-factor ANOVA under two scenarios: the first where group designation (a categorical value of 0 or 1) was randomly distributed across the phylogeny and the second where group designation was aggregated and clumped relative to sub-clades. These scenarios represent extremes of how groups may be dispersed across the phylogeny (see main text). Our simulation protocol was as follows.

First, we generated 1,000 pure-birth trees containing N = 32 taxa using the 'pbtree' function in the R-package *phytools* 0.6-20 (Revell 2012). Next we generated grouping factors for our two scenarios such that the 32 species on each phylogeny were randomly distributed (X_1) or were aggregated and clumped (X_2). We then selected input β_{input} to represent the known group differences: β_{input} = 0.0, 0.5, 1.0, 3.0, 5.0, and simulated datasets representing the dependent (Y) variable under Brownian motion for each phylogeny, but with known differences in groups incorporated. This was accomplished using the 'fastBM' function (*phytools*: Revell 2012), combined with the β_{input} and X information as: $Y = \beta X$ + fastBM(tree). The process was repeated for each level of β_{input} and the mean $\beta_{\mathrm{est.}}$ was treated as a measure of bias.

Results: For randomly distributed groups on the phylogeny, the mean $\beta_{\mathrm{est.}}$ from RRPP was virtually identical to the input value (Fig. S2), implying minimal bias in parameter estimates from the procedure. Likewise for aggregated groups (Fig. S2), the mean $\beta_{\mathrm{est.}}$ was virtually identical to the input value, though there was additional variation in parameter estimates across simulated datasets. However, even in this extreme grouping scenario, no obvious bias in parameter estimation was observed, implying that RRPP was capable of estimating model parameters even in the face of strong group-phylogeny covariation.

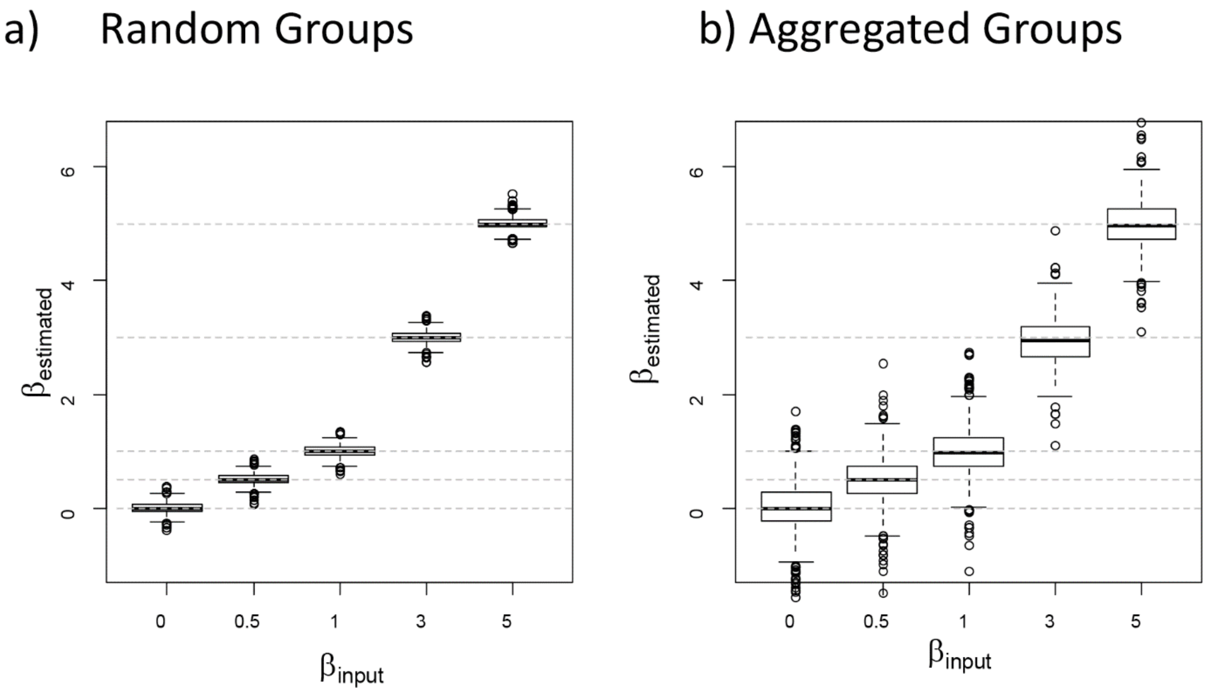

Fig. S2. Results from computer simulations evaluating parameter bias in scenarios with a) random b) and aggregated groups on the phylogeny. Dashed lines correspond to simulated "true" values. Mean β across 1, 000 simulated datasets are:

Random groups: β_{mean} = 0.0036, 0.500, 0.996, 3.002, 5.006
Aggregated: β_{mean} = 0.0026, 0.485, 0.997, 2.934, 4.977

In addition to the above test of bias we examined variation in the empirical sampling distribution of $\beta_{\mathrm{est.}}$ across a wider set of phylogenies. For this simulation we generated 100 pure-birth phylogenies and on each simulated 100 datasets as above. For each simulation, we then obtained the sampling distribution for $\beta_{\mathrm{est.}}$ as obtained from RRPP for each simulation, and calculated the standard deviation of the sampling distribution of $\beta_{\mathrm{est.}}$ for each of 100 datasets across the 100 phylogenies.

Results: For randomly distributed groups, the variation in $\beta_{\mathrm{est.}}$ from RRPP was lower than that observed when groups were aggregated on the phylogeny (Fig. S3), implying under extreme grouping scenarios, variation in the empirical sampling distribution of $\beta_{\mathrm{est.}}$ is increased.

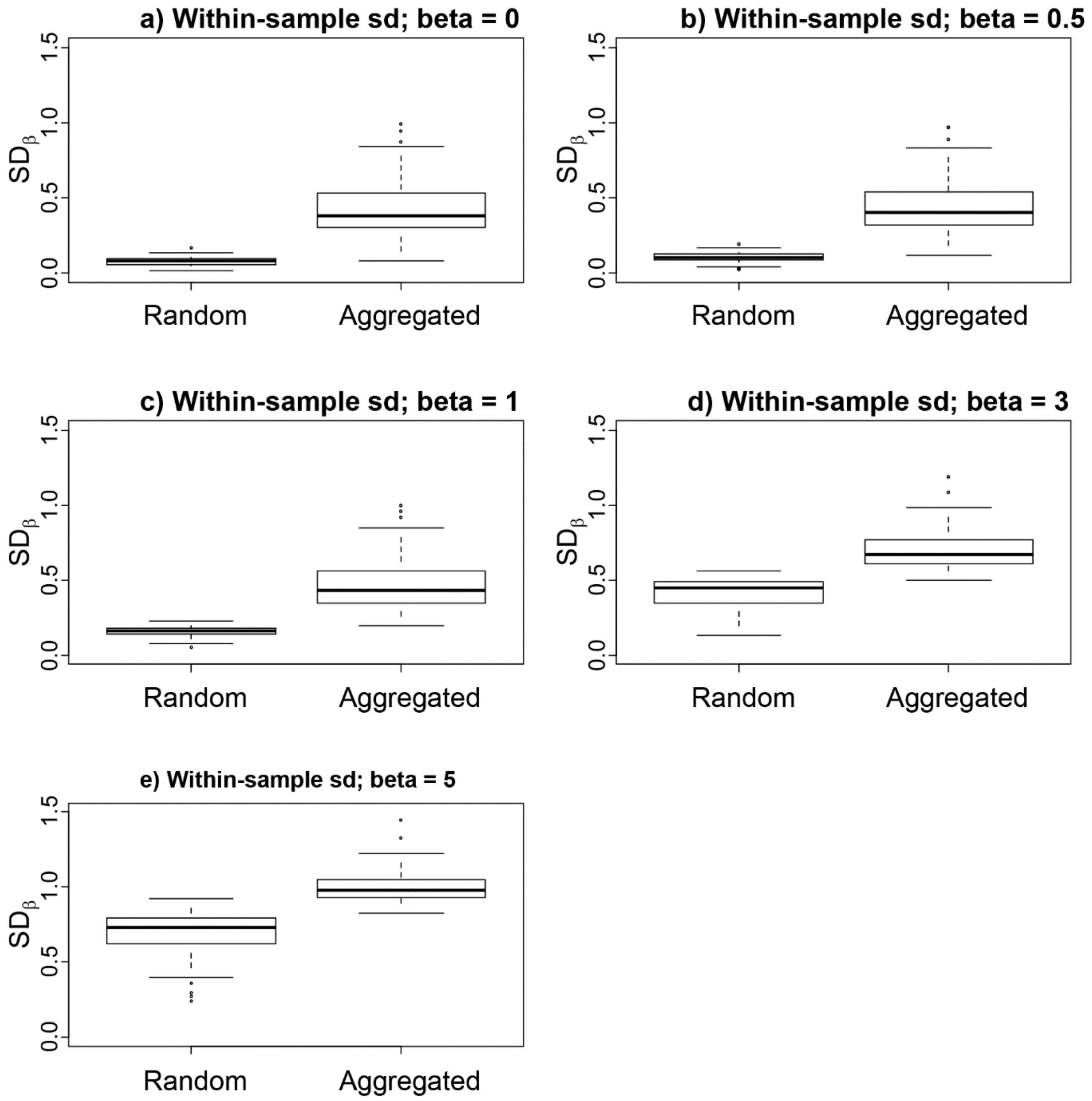

Fig. S3. Results from computer simulations evaluating the variance in sampling distributions of $\beta_{est.}$ for random and aggregated groups on the phylogeny.

4. Empirical sampling F distributions: Phylogenetic ANOVA

One important characteristic of proper permutation procedures is that their empirical sampling distributions match theoretical expectations under ideal conditions. To determine this for RRPP we performed a series of simulations for three different sample sizes (N = 32, 64, and 128) and for both univariate and multivariate data (p = 1, 5, and 10). We also evaluated the empirical sampling distribution for two grouping scenarios: the first where groups were randomly distributed along the phylogeny and the second where groups were aggregated (see above).

For each simulation, we generated a pure-birth phylogeny containing N taxa, upon which we evolved a dependent trait or set (matrix) of traits (**Y**) following a Brownian motion model of evolution. Next, we generated grouping factors for our two scenarios such that the N species on the phylogeny were either randomly distributed or aggregated and clumped. We then performed phylogenetic ANOVA using RRPP and 199,999 iterations for each of the two *X* variables. Finally, we compared the empirical sampling distributions obtained from RRPP to the theoretical *F*-distribution.

Results: In all cases, the empirical sampling distribution closely matched the theoretical *F* distribution Fig S4-S6. As expected, the fit of the empirical and theoretical distributions improved as the number of taxa increased. Additionally, this pattern remained consistent as the number of variables (*p*) increased, and regardless of how groups were dispersed across the phylogeny. These results demonstrate that RRPP is insensitive to the pattern of group dispersion with respect to generating its sampling distribution, and that method yield empirical sampling distributions as expected from statistical theory.

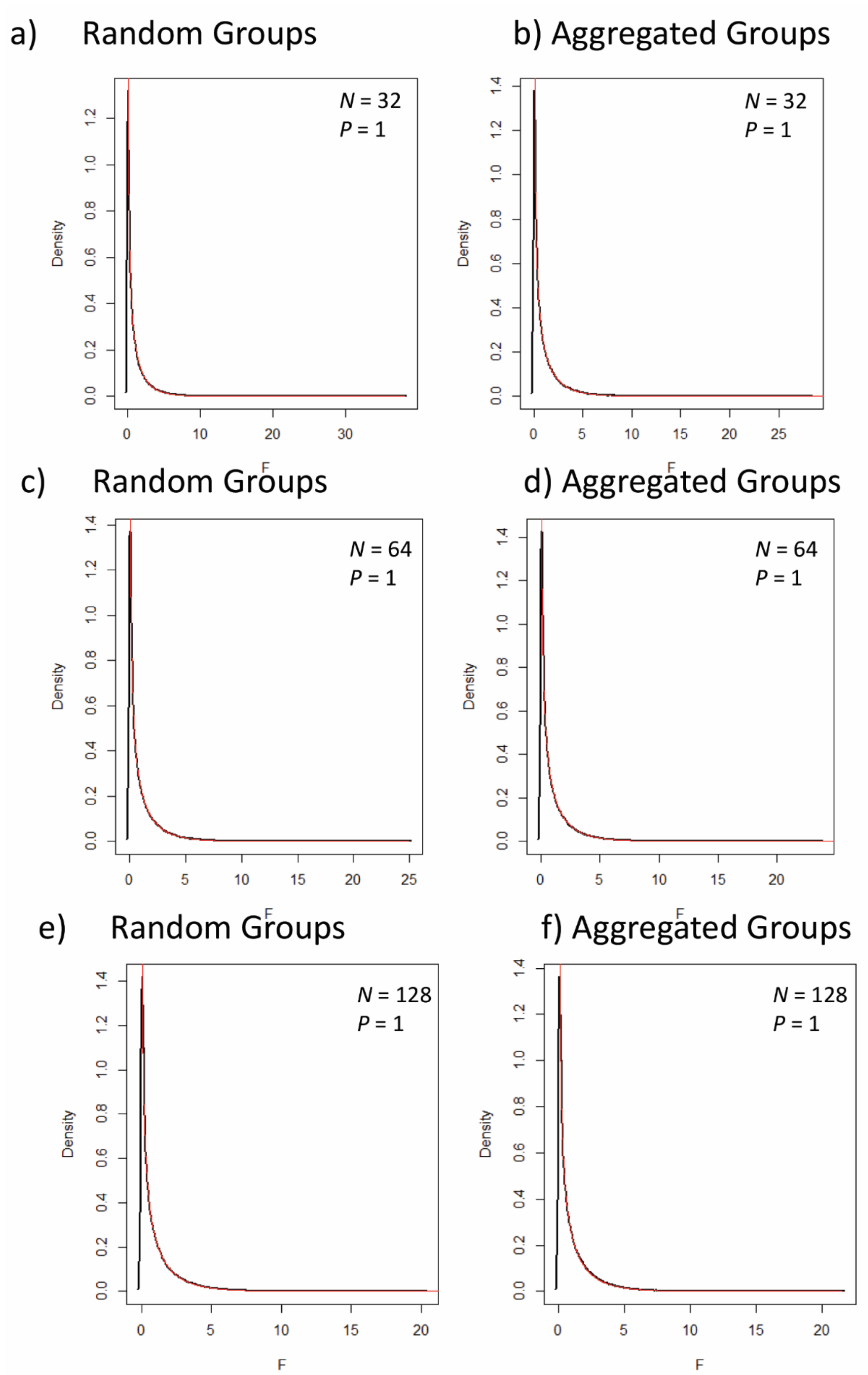

Fig. S4. Empirical sampling distributions (black) and theoretical *F* distribution (red) for alternating and aggregated groups (*X*) for differing numbers of taxa (*N* = 32, 64, 128. Dependent (**Y**) variables are: *p* = 1.

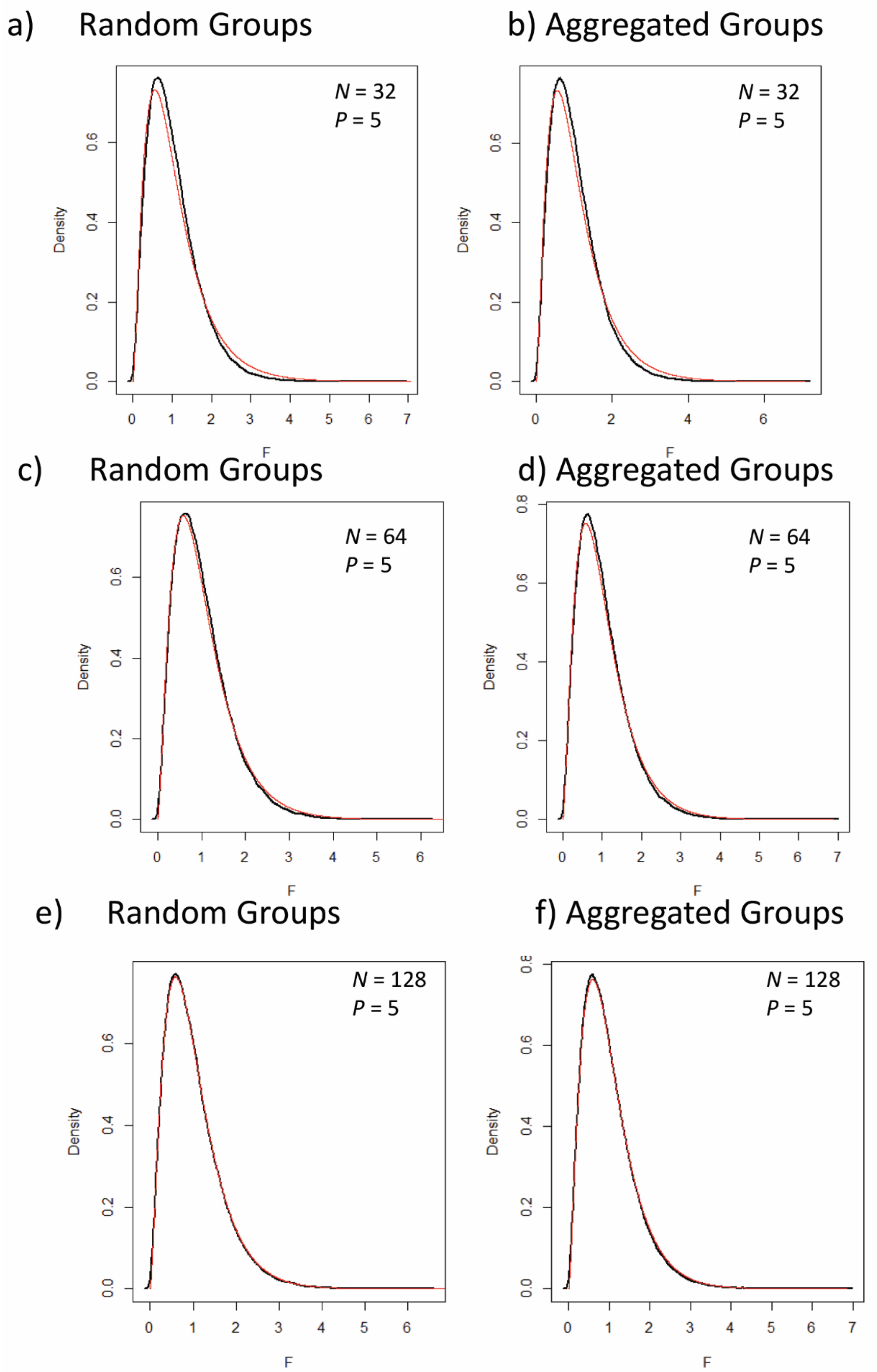

Fig. S5. Empirical sampling distributions (black) and theoretical *F* distribution (red) for alternating and aggregated groups (*X*) for differing numbers of taxa (*N* = 32, 64, 128. Dependent (**Y**) variables are: *p* = 5.

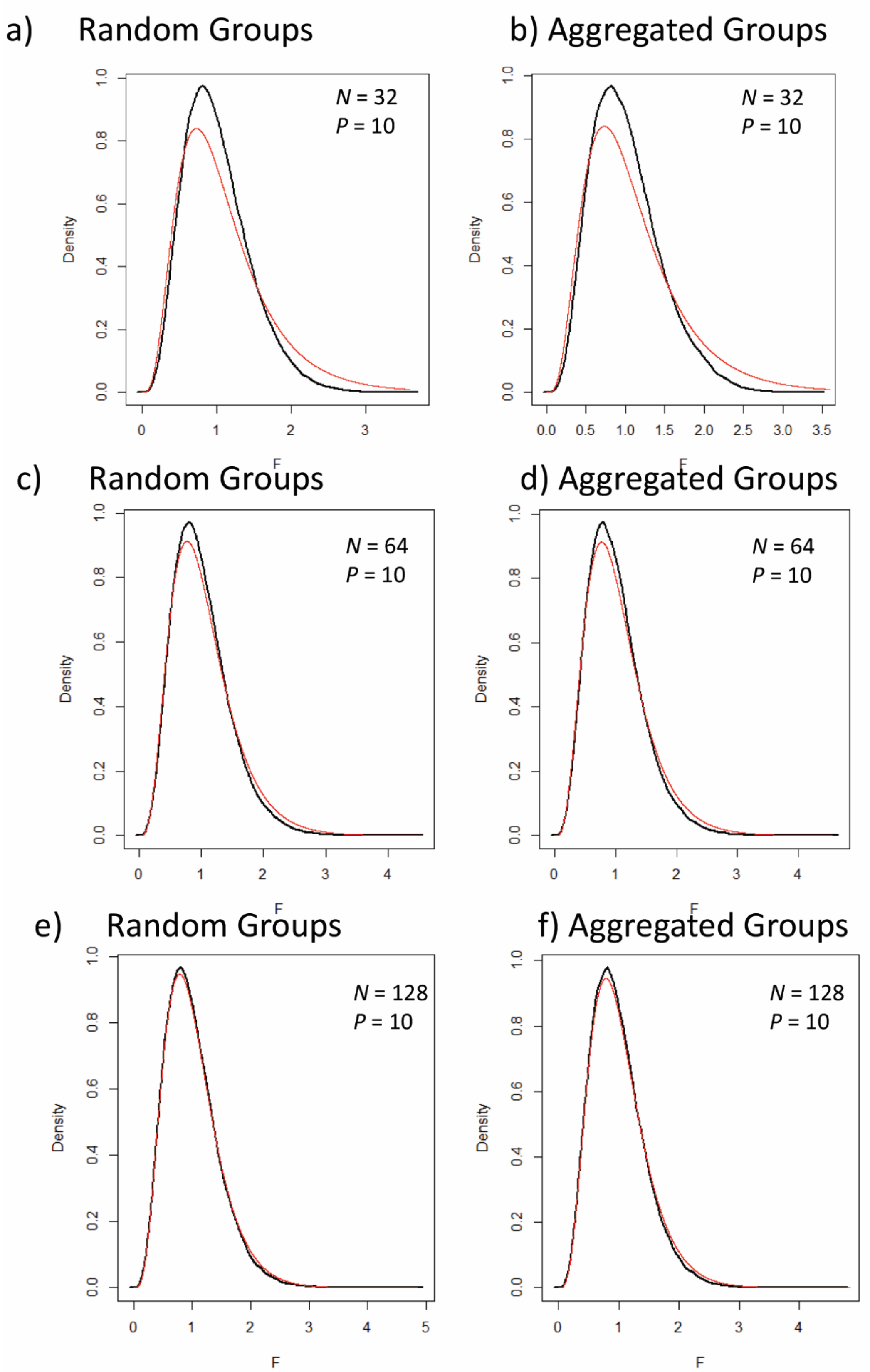

Fig. S6. Empirical sampling distributions (black) and theoretical *F* distribution (red) for alternating and aggregated groups (*X*) for differing numbers of taxa (*N* = 32, 64, 128. Dependent (**Y**) variables are: *p* = 10.

5. Rotation-invariance and insensitivity to Y_{cov}

We also examined the degree to which the approach is rotation-invariant and insensitive to levels of trait covariation in the dependent variables. As above we generated 1,000 pure-birth phylogenies (*N* = 32) and on each simulated multivariate response (**Y**) data under a Brownian motion model of evolution. Three datasets were generated on each phylogeny with each displaying a differing degree of covariation between trait dimensions (*r* = 0.0, 0.5, 0.9). Next, each dataset was rotated to its principle axes via a rigid rotation ($\mathbf{Y_{rot}}$). Species were then assigned to levels of grouping factors under two scenarios: random and aggregated, as above. We then performed phylogenetic ANOVA using RRPP for both **Y** and $\mathbf{Y_{rot}}$, obtained the significance value, and correlated these between datasets.

Results: In all cases, there was a perfect correlation between statistical results of the dataset in its original orientation with the same dataset rotated to its principle axes (Fig. S7). Further, this pattern was the same regardless of the amount of correlation between traits. As such these results confirmed that the method is both rotation-invariant and insensitive to differing levels of trait covariation.

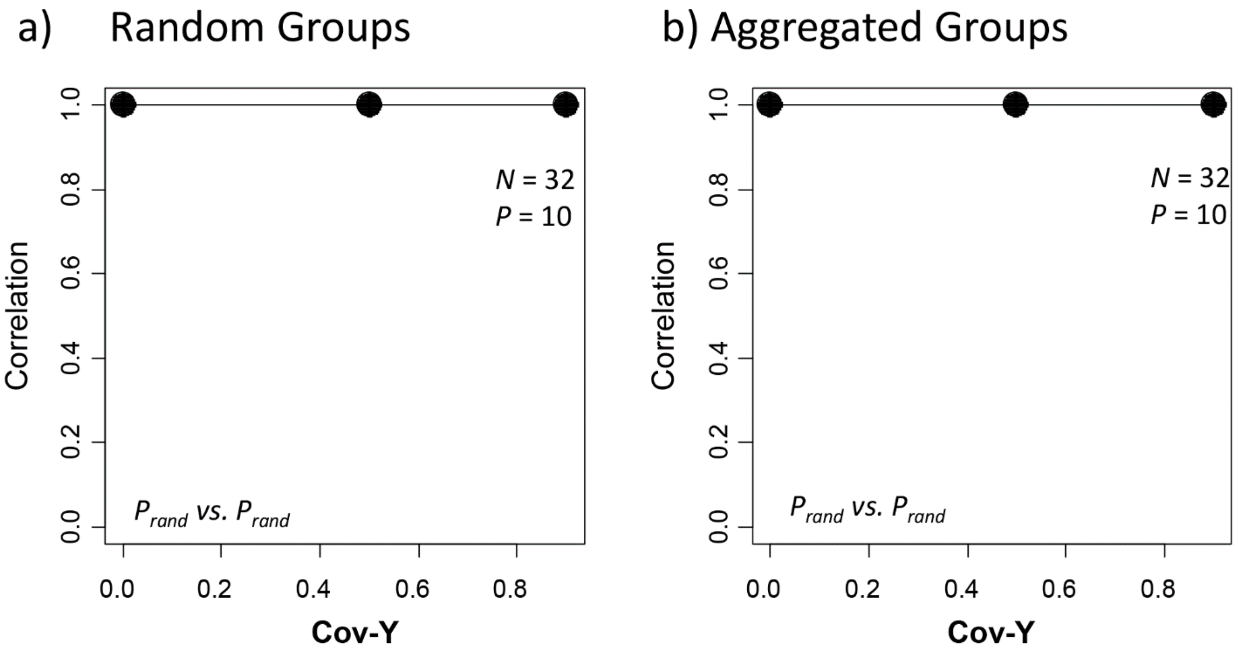

Fig. S7. Correlation between significance levels from phylogenetic ANOVA for simulated datasets versus the same datasets rotated to their principal axes for: a) random groups and b) aggregated groups.

6. Type I error and power: Phylogenetic ANOVA

Finally, we examined the type I error and power of RRPP under two differing ANOVA scenarios: one where groups were randomly distributed across the phylogeny and a second where group designation was aggregated and clumped relative to phylogenetic sub-clades. We performed these simulations for three different sample sizes (*N* = 32, 64, and 128) and for both univariate and multivariate data (*p* = 1, 5, and 10).

For each simulation, we generated 1,000 pure-birth phylogenies containing *N* taxa. Next, we generated grouping factors for our two scenarios such that the *N* species on the phylogeny were either randomly distributed, or were aggregated and clumped. We then selected input β_{input} to represent the known group differences (β_{input} = 0.0, 0.1, 0.25, 0.5, 0.75, 1.0), and simulated 1,000 datasets (1 per phylogeny) of the dependent (**Y**) variable under Brownian motion using the function 'sim.char' in the R-package Geiger 2.0.6 (Pennell et al. 2014), with known differences in groups incorporated (see above). For each simulated dataset the significance was determined with RRPP, and the proportion of simulated datasets whose significance level is less than the nominal $\alpha = 0.05$ was treated as an estimate of Type I error ($\beta_{\text{input}} = 0.0$) or power ($\beta_{\text{input}} > 0.0$).

Results: Simulations revealed that RRPP displayed appropriate type I error rates for both scenarios of group dispersion relative to phylogenetic clades (Figs S8). Further, power increased as the known difference (β_{input}) between groups increased and as the dimensionality of the data increased. This is a pattern observed in other permutation-based procedures for phylogenetic regression (Adams 2014), and demonstrates that the method is robust to increasing trait dimensionality. Finally, power was slightly lower when groups were aggregated on the phylogeny as compared to when they were randomly distributed, suggesting that the extreme aggregation did have an effect on the ability to detect group differences in response variables (**Y**). Together, these results revealed that RRPP displayed appropriate type I error under different dispersions of grouping (X) variables, and displayed appropriate statistical power to detect group differences when they were present.

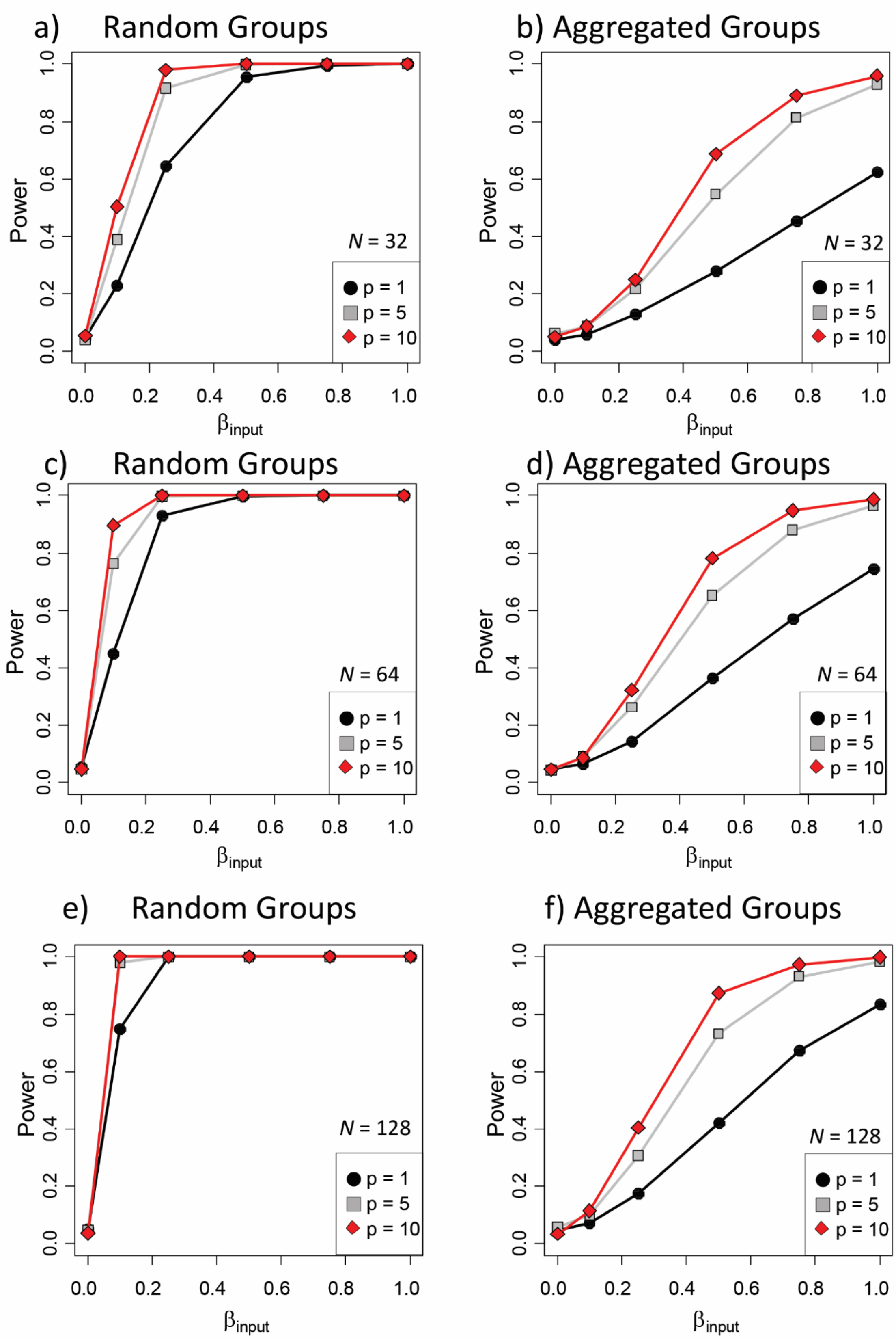

Fig. S8. Power curves for RRPP for differing numbers of response (**Y**) variables (*p* = 1, 5,10), for differing numbers of taxa (*N* = 32, 64, 128).

C: Further Details about First-moment, Second-moment, and Global Exchangeability

These details largely follow the definitions of Commenges (2003). Exchangeability is the condition of applying a permutation procedure to observations that retain constancy of moments through permutations. The type of exchangeability that is simplest to appreciate is first-moment exchangeability. For example, a vector of observations, Y, has first-moment exchangeability if randomized to form the vector, Y^*, because the mean of Y is equal to the mean of Y^*. Randomizing Y to form Y^* also has second-moment exchangeability, as the variance of Y is equal to the variance of Y^*. Simply randomizing Y, say to perform a two-sample *t*-test, where the *t*-statistic is estimated in each random permutation to generate an empirical *t*-distribution, is a procedure that has global exchangeability (maintaining first- and second-moment exchangeability). Statistical tests based on permutation procedures with global exchangeability are exact tests (Anderson and Braak 2003).

It is helpful to visualize the previous example as a linear model. In this case, the linear model contains only an intercept and the residuals from this model are the exchangeable units under the null hypothesis. Through every permutation, the intercept is constant, as is the variance of the residuals. When a linear model becomes more complicated - contains one or more slopes - global exchangeability is violated but second-moment exchangeability is somewhat preserved as long as the variance of residuals is rather constant across permutations. For OLS estimation of coefficients, for uncorrelated residuals with mean of 0, this is the case (or at least, the Central Limit Theorem holds). Statistical tests based on permutation procedures with second-moment exchangeability are approximate tests (and in some cases, exact tests can be performed - see Anderson & ter Braak 2003 for further details).

Difficulty arises when residuals are correlated (non-independent), perhaps because of repeated measures of subjects or inherent relatedness, due to spatial, temporal, genetic, or phylogenetic similarity. In such cases, GLS estimation of coefficients is preferred, and as described above, can be achieved through linear transformation of data and design matrices for the purpose of OLS estimation of coefficients, which is often more efficient. As such, residuals are also linearly transformed. This linear transformation has the benefit of producing uncorrelated error with mean of 0 (if the transformation is appropriate). Although heuristic in nature, it was suggested by Commenges (2003) and demonstrated here that preserving second-moment exchangeability in transformed residuals rather than untransformed residuals results in appropriate type I error rates and approximate test distributions that more closely resemble exact distributions.

As a final point, one must consider the case where a linear model contains only an intercept but observations are inherently correlated, for example, because of phylogenetic relatedness. There is no method that preserves either global or second-moment exchangeability of the observed data when employing a linear transformation. However, after transforming the data and estimating the GLS coefficients via OLS, randomizing the residuals preserves both second-moment and global exchangeability of phylogenetically transformed data. Thus, randomizing the transformed residuals of any phylogenetic linear model comes closest to preserving second-moment exchangeability, which perhaps explains why distributions and type I error rates tend to be more appropriate.

References

Adams, D. C. 2014. A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. Evolution 68:2675-2688.

Adams, D. C. and M. L. Collyer. 2015. Permutation tests for phylogenetic comparative analyses of high-dimensional shape data: what you shuffle matters. Evolution 69:823-829.

Adams, D. C. and M. L. Collyer. 2018. Multivariate phylogenetic comparative methods: Evalutions, comparisons, and recommendations. Syst. Biol. 67:14-31.

Adams, D. C., M. L. Collyer, A. Kaliontzopoulou, and E. Sherratt. 2017. Geomorph: Software for geometric morphometric analyses. R package version 3.0.6. http://CRAN.R-project.org/package=geomorph.

Anderson, M. J. and C. J. F. t. Braak. 2003. Permutation tests for multi-factorial analysis of variance. Journal of Statistical Computation and Simulation 73:85-113.

Commenges, D. 2003. Transformations which preserve exchangeability and application to permutation tests. J. Nonparam. Stat. 15:171-185.

Garland, T. J. and A. R. Ives. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. Am. Nat. 155:346-364.

Goolsby, E. W. 2016. Likelihood-based parameter estimation for high-dimensional phylogenetic comparative models: overcoming the limitations of “distance-based” methods. Syst. Biol. 65:852-870.

Johnston, J. and J. DiNardo. 1997. Econometric methods. McGraw Hill PUblishing, New York.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lutkepohl, and T.-C. Lee. 1985. The theory and practice of econometrics. John Wiley & Sons, New York.

Kariya, T. and H. Kurata. 2004. Generalized least squares. John Wiley & Sons, New York.

Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, M. E. Alfaro, and L. J. Harmon. 2014. Geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics 30:2216-2218.

Rencher, A. C. 2000. Linear models in statistics. John Wiley & Sons, New York.

Revell, L. J. 2012. Phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217-223.