Controls for phylogeny and robust analysis in Pareto Task Inference

Miri Adler* 1, Avichai Tendler*2, Jean Hausser*3, Yael Korem2, Pablo Szekely 4, Noa Bossel6, Yuval Hart6, Omer Karin2, Avi Mayo2, Uri Alon§
1 Eli and Edythe L. Broad Institute of Harvard and MIT Cambridge, MA, USA
2 Weizmann Institute of Science Molecular Cell Biology Rehovot, IL
3 Karolinska Institute, Department of Cellular and Molecular Biology Stockholm, Stockholm, SE
4 Duke University Durham, NC, USA
5 Weizmann Institute of Science, Biological Regulation Rehovot, IL
6 The Hebrew University of Jerusalem, Department of Psychology Jerusalem, Jerusalem, IL
* equal contribution
§ Corresponding Author: urialonw@gmail.com

Abstract:

Understanding the tradeoffs faced by organisms is a major goal of evolutionary biology. One of the main approaches for identifying these tradeoffs is Pareto task inference (ParTI). Two recent papers claim that results obtained in ParTI studies are spurious due to phylogenetic dependence (Mikami and Iwasaki, 2021) or hypothetical p-hacking and population-structure concerns (Sun and Zhang, 2020). Here we show that these claims are baseless. We present a new method to control for phylogenetic dependence, called SibSwap, and show that published ParTI inference is robust to phylogenetic dependence. We show how researchers avoided p-hacking by testing for robustness of preprocessing choices. We also provide new methods to control for population structure and detail the extensive experimental tests of ParTI in systems ranging from ammonites to cancer gene expression. The methods presented here may help to improve future ParTI studies.

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
When organisms perform multiple tasks, they face tradeoffs; understanding these tradeoffs is important for understanding evolution. A widely used approach for identifying evolutionary tradeoffs is Pareto optimality theory (Shoval et al., 2012). This theory predicts that under certain assumptions, traits fill a pointed shape in trait space called a polytope (triangle, tetrahedron, etc). At the vertices are phenotypes optimal for a certain task, and the number of vertices equals the number of tasks. To detect polytopes and find features that are enriched near the archetypes, our lab developed the ParTI algorithm (Hart et al., 2015). ParTI has been used in different contexts including morphology (Tendler, Mayo and Alon, 2015), gene expression (Friedman et al., 2020; Hausser and Alon, 2020) and life-history traits (Szekely et al., 2015).

Recent papers (Sun and Zhang, 2020) (Mikami and Iwasaki, 2021) claim that many of the results obtained with ParTI are spurious. It is of significant interest to understand whether these claims have merit, because if they do, one may conclude that the ParTI approach is not useful. Here we show that these claims are baseless, and present new approaches to control for caveats in future ParTI studies.

**New SibSwap algorithm to test for phylogenetic dependence in ParTI**

Phylogenetic dependence is widely studied in comparative biology (Felsenstein, 1985; Grafen, 1989; Pagel and Harvey, 1989; Freckleton, Harvey and Pagel, 2003). In the context of ParTI, phylogenetic inheritance simulations can sometimes generate triangle-like shapes that do not stem from adaptation, as noted by (Edelaar, 2013). The ParTI approach for assessing the significance of polytopes is based on swapping traits between species as if they were independent, which ignores phylogenetic correlations and breaks phylogenetically independent contrasts (Felsenstein, 1985). It can thus lead to inflated p-values. This caveat has therefore been addressed in the two relevant ParTI papers, on ammonite shells (Tendler, Mayo and Alon, 2015) and on mammalian life-history traits (Szekely et al., 2015).

The study on ammonites specifically aimed to address phylogenetic concerns (Tendler, Mayo and Alon, 2015). ParTI showed that ammonite shell traits fill a triangle with three shell archetypes. After mass extinctions, in which only a few genera survive, the ammonites refilled statistically the same triangle (Fig 1a). This convergent evolution is evidence for the adaptive nature of the archetypes.

Sun and Zhang revisit the concern of phylogenetic dependence by using simulations of Brownian motion on a tree, which can create triangle-like shapes. They do not analyze any specific ParTI...
dataset, and dismiss the controls used in ParTI studies without offering an alternative phylogenetic test.

To address phylogeny, it would be important to have a phylogenetic test made specifically for ParTI. Such a test, called the flipping t-ratio test, was recently proposed by (Mikami and Iwasaki, 2021). The authors concluded that the ammonite and life-history triangles are not significant when controlled for phylogeny using the flipping t-ratio test.

First, we analyzed the flipping t-ratio algorithm. It elegantly preserves the phylogenetically independent contrasts of the original dataset. However, it does not preserve the distribution of each trait: it generates new trait values that are far from the range of the original data, sometimes exceeding the range by a factor of ten or more (Fig 1 b,c). The flipping-t triangle area is on average 13 times larger than the original triangle area (Fig 1d). Thus, the triangles produced by the flipping t-ratio algorithm are spurious. Due to the same reason, this algorithm gives false negatives in control datasets with a star phylogeny (SI Fig S1). The flipping t-ratio method should therefore not be used in practice unless it is somehow modified to properly handle outliers.

A more appropriate phylogenetic test would not create outliers by preserving the marginal distribution of each of the traits. Here we present a new algorithm for testing the phylogenetic significance of polytopes, which preserves both the phylogenetic constraints and the marginal distribution of all traits. The algorithm, called SibSwap, is simple (Fig 2a): for each set of terminal nodes with a shared parental node (sibling tips, Fig S2), permute each of the traits independently. This mixes traits between sibling tips (whether in polytomies or not), but not between non-sibling tips. Next, compute the significance of the triangle or polytope using the standard t-ratio test of ParTI (Hart et al., 2015). The t-ratio is the ratio between the area of the polytope and the area of the convex hull of the data. The closer the t-ratio is to 1, the better the polytope fits the data. Significance is assessed by the probability that the polytope inferred for SibSwap-shuffled data has a t-ratio closer to 1 than the real data. A low p-value indicates that the polytope is not caused by phylogenetic constraints. Conversely, high p-values indicate that phylogenetic constraints cannot be rejected as a cause for the polytope. SibSwap rejects phylogeny appropriately in control datasets with a star phylogeny (SI Fig S1) and performs as well as the flipping t-ratio test on simulated Brownian evolution, Fig 2b and SI Fig S3.

Importantly, SibSwap preserves the phylogenetically independent contrasts (PICs, defined in (Felsenstein, 1985)), as shown in Fig 2a, in the standard case where terminal branch lengths are equal (as in ultrametric trees, see SI). SibSwap also preserves any other single-trait statistic, such as Pagel's $\lambda$, a common measure for phylogenetic signature (Pagel and Harvey, 1989), Fig 2a.
The PIC distributions for ammonite and life-history datasets are indistinguishable in the original and SibSwapped datasets. SibSwap thus improves on the original ‘naive’ ParTI algorithm which swaps traits between any two tips (not only sibling tips) and thus breaks the PIC distribution. More elaborate versions of SibSwap in which traits are permuted among species closer than a given phylogenetic distance are discussed in the SI.

For both ammonite and life-history datasets, the real triangle has a t-ratio significantly closer to 1 than the SibSwap-shuffled data (p=0.024 life-history, p=0.012 ammonite). The reason that phylogenetic effects are not of major importance in these datasets is that ammonite shells and mass-longevity of mammals can evolve rapidly on the timescale of speciation (Szekely et al., 2015). We conclude that the ParTI inference for these datasets is well-controlled for phylogenetic inheritance effects.

Cancer archetypes are not due to genomic population structure

Sun and Zhang raised the possibility, noted previously (Edelaar, 2013; Hart et al., 2015), that population structures such as different ethnic groups can produce polytopes. To do so they simulated mutations on a chromosome and assumed that simulated traits are binary combinations of mutations. Data falls in three well-separated clusters due to the three simulated “ethnic groups” (Fig 3a), which can cause false positives in ParTI. This simulation is of doubtful relevance to data used by ParTI papers.

The ParTI papers dealing with human populations analyzed cancer gene-expression datasets (Hausser et al., 2019; Hausser and Alon, 2020). Here, we tested the association between ancestry and the cancer tasks detected by ParTI, using a recent approach that allows ancestry to be inferred directly from the genomic sequences inherent in the gene-expression dataset (Carrot-Zhang et al., 2020). We find no significant association between ancestry and the ParTI cancer tasks. An example showing ancestry on the inferred triangle for low-grade glioma is shown in (Fig. 3b). These observations challenge the hypothesis that population structure is a major factor for ParTI in these cancer datasets.

More generally, the SibSwap approach can be adapted to help reject polytopes arising exclusively from ancestry groups or other data identifiers. One permutes traits within each ancestry group. The “ethnic group” simulation of Sun and Zhang yields a poor p-value (p=0.09), because shuffling within the clusters leaves the dataset essentially the same (Fig 3a,c), whereas the same analysis for the cancer data yields p<0.001, because shuffling within ancestry groups ruins the triangle (Fig 3b,d). Similar results are obtained in cancer data that is down-sampled so that each ancestry group has the same number of datapoints (10 data points per group, p=0.006).
A similar test can reject cases where the polytope is due to a few discrete data clusters, even if ancestry is unknown. One classifies the data points into $n$ clusters, where $n$ is the number of ParTI archetypes, by using a standard algorithm such as k-means, and then shuffles traits within each cluster. The “ethnic group” simulation fails this test, whereas the cancer, ammonite and life-history datasets pass it because their polytopes are continuously filled and are not due to discrete clusters.

We note, however, that there may be other types of data structures that do not yield clusters, but still produce polytopes, emphasizing the need to test for data structure as extensively as possible when using ParTI.

**Best practices to avoid p-hacking**

We next address the claim by Sun and Zhang that the need to pre-process data for ParTI promotes p-hacking. They do not provide evidence from any particular paper. Instead, the proposed evidence is a simulation of random data in which one tries many processing choices (thresholds) and picks the ones that give a good p-value. Preprocessing is a standard and necessary step in the analysis of biological data. Therefore, such a simulation would ‘prove’ p-hacking in any algorithm (clustering, etc.).

The simulation of Sun and Zhang does not resemble what researchers in ParTI papers actually did. Instead, ParTI researchers used standard processing methods (e.g., taking the log of gene expression). When there were several possible choices (e.g., thresholds), they tested whether the results were robust to processing choices. Results were only published if they were robust. Table S1 (SI) lists processing choices in papers published by our group using ParTI. We advocate the following best practices for ParTI analyses: 1) use biologically reasonable preprocessing steps, and 2) be transparent and include all steps in the paper or supplementary information.

**Alternative explanation for yeast deletion triangle**

Sun and Zhang analyze what they state is a negative control: a biological dataset which did not undergo evolutionary optimization. Their proposed negative control is a gene-expression database of 1484 yeast deletion strains (Kemmeren et al., 2014). The argument is that deletion strains did not have time to evolve after the deletion, and thus cannot be optimal. They find that ParTI detects a triangle with enriched gene functions and concludes that this is a false-positive result.

As Sun and Zhang note, the deletion dataset they used is nearly identical to the control wildtype dataset. Since the inferred triangle is essentially that of biological repeats of the wild-type strain
(Fig 3e), one should ask whether biological repeats are truly a negative control for adaptive responses. Biological repeats are grown and handled in slightly different conditions. These conditions can trigger adaptive gene expression changes, which evolved to handle natural environmental changes. The archetypes shown in Tables 1 and 2 in Sun and Zhang are related to mitochondrial function, carbohydrate metabolism, and protein synthesis. These processes are consistent with the possibility that the biological repeats largely reflect batch-to-batch variation in growth conditions.

Before publishing such conclusions, however, we would recommend doing additional experimental tests with independent data, as detailed in the ParTI manual. We note that when applied to the 703 deletion strains that significantly differed from their wild-type controls, namely the ‘responsive mutant’ dataset of (Kemmeren et al., 2014), ParTI finds no significant polytope (p=0.64), Fig 3f.

**ParTI studies perform experimental and theoretical tests of the archetypes**

Sun and Zhang give an incomplete account of ParTI studies by failing to mention experimental tests. ParTI papers consider the inferred archetypes to be hypotheses and tested these hypotheses using calculations, independent experimental data and/or new, specially conducted experiments. For example, in (Friedman et al., 2020), a fibroblast archetype showed an unexpected antigen-presenting function. To test this, the authors conducted new experiments showing that these fibroblasts indeed express the antigen-presentation complex MHC-class-II in vivo and in vitro. Table S1 (SI) provides examples of experimental tests in ParTI studies.

In sum, we presented the SibSwap method to control for phylogeny and population-structure caveats and find that published ParTI archetypes are not due to such caveats. Well-conducted ParTI studies avoid p-hacking by using transparent and reasonable preprocessing methods and treat archetypes as hypotheses which they test with independent experiments.
Fig 1. Convergent evolution in ammonites and spurious triangles in the flipping t-ratio test
(a) Ammonites refill statistically the same triangle after mass extinctions. Each point is a genus. W and D are dimensionless shell-shape parameters, the whorl expansion rate and internal/external shell ratio. (b) The flipping t-ratio test creates outliers in ammonite data. (c) The test does not preserve the marginal trait distributions (original data in orange, after the flipping t-ratio algorithm in blue), and (d) creates much larger triangles than the original data triangle as shown by the ratio of their areas (see also (b)). Settings are as described in (Mikami and Iwasaki, 2021).
Figure 2. The SibSwap algorithm preserves trait distributions as well as phylogenetically independent contrasts. (a) SibSwap-shuffled result (right) of original data (left) preserves the trait distributions. Here each terminal node has two traits represented by numbers in curly brackets. Branch lengths are in gray. SibSwap also preserves absolute phylogenetically independent contrasts (PICs) and Pagel’s $\lambda$, both calculated using the Mathematica package “Phylogenetics for Mathematica (Ver. 2.1)” (Polly, 2012). (b) Simulations of Brownian diffusion on a phylogenetic tree can create false-positive triangles in the original Naive ParTI shuffling. These triangles are rejected by SibSwap, which makes only slight changes to the triangle.
Fig 3. Controls for ancestry (a) Sun and Zhang “ethnic group” simulation from their Fig 3c. (b) Low-grade glioma triangle (Hausser et al., 2019) with ancestry indicated. (c) Permuting traits within the three “ethnic group” clusters results in a nearly identical triangle. (d) Low-grade glioma triangle is disrupted upon trait permutation within ancestry groups. (e) Full deletion strain dataset of (Kemmeren et al., 2014) analyzed by Sun and Zhang’s is indistinguishable from the wild-type biological repeats grown with each strain. (f) The responsive mutant dataset of (Kemmeren et al., 2014) differs from their wild-type repeats and shows no significant ParTI polytope.
Data Availability Statements

Ammonite data (Fig. 1) is as reported in (Tendler, Mayo and Alon, 2015)
Life-history data is as reported in (Szekely et al., 2015)
“Ethnic group” simulation data (Fig. 3a) is as reported in (Sun and Zhang, 2020)
Low-grade glioma data (Fig. 3b) is as reported in (Hausser et al., 2019)
Yeast deletion data (Fig. 3e,3f) is as reported in (Kemmeren et al., 2014)

All algorithms used will be posted in a public repository, GitHub.
Carrot-Zhang, J. et al. (2020) ‘Comprehensive Analysis of Genetic Ancestry and Its Molecular Correlates in Cancer’, *Cancer Cell*. Cell Press, 37(5), pp. 639-654.e6. doi: 10.1016/j.ccell.2020.04.012.

Edelaar, P. (2013) ‘Comment on “evolutionary trade-offs, pareto optimality, and the geometry of phenotype space”’, *Science*. American Association for the Advancement of Science, pp. 757–757. doi: 10.1126/science.1228281.

Felsenstein, J. (1985) ‘Phylogenies and the comparative method.’, *American Naturalist*. doi: 10.1086/284325.

Freckleton, R. P., Harvey, P. H. and Pagel, M. (2003) ‘Bergmann’s rule and body size in mammals’, *American Naturalist*. doi: 10.1086/374346.

Friedman, G. et al. (2020) ‘Cancer-associated fibroblast compositions change with breast cancer progression linking the ratio of S100A4+ and PDPN+ CAFs to clinical outcome’, *Nature Cancer*. doi: 10.1038/s43018-020-0082-y.

Grafen, A. (1989) ‘The phylogenetic regression.’, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. doi: 10.1098/rstb.1989.0106.

Hart, Y. et al. (2015) ‘Inferring biological tasks using Pareto analysis of high-dimensional data’, *Nature Methods*, 12(3). doi: 10.1038/nmeth.3254.

Haussler, J. et al. (2019) ‘Tumor diversity and the trade-off between universal cancer tasks’, *Nature Communications*. doi: 10.1038/s41467-019-13195-1.

Haussler, J. and Alon, U. (2020) ‘Tumour heterogeneity and the evolutionary trade-offs of cancer’, *Nature Reviews Cancer*. Nature Research, pp. 247–257. doi: 10.1038/s41568-020-0241-6.

Kemmeren, P. et al. (2014) ‘Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors’, *Cell*. Cell Press, 157(3), pp. 740–752. doi: 10.1016/j.cell.2014.02.054.

Mikami, T. and Iwasaki, W. (2021) ‘The flipping t-ratio test: Phylogenetically informed assessment of the Pareto theory for phenotypic evolution’, *Methods in Ecology and Evolution*. doi: 10.1111/2041-210X.13553.

Pagel, M. D. and Harvey, P. H. (1989) ‘Comparative methods for examining adaptation depend on evolutionary models.’, *Folia primatologica; international journal of primatology*. doi: 10.1159/000156417.

Polly, P. D. (2012) ‘Phylogenetics for Mathematica (Ver. 2.1)’. P. David Polly. Available at: https://scholarworks.iu.edu/dspace/handle/2022/14614 (Accessed: 24 June 2021).

Shoval, O. et al. (2012) ‘Evolutionary trade-offs, pareto optimality, and the geometry of phenotype space’, *Science*, 336(6085). doi: 10.1126/science.1217405.

Sun, M. and Zhang, J. (2020) ‘Rampant False Detection of Adaptive Phenotypic Optimization by ParTI-Based Pareto Front Inference’, *Molecular Biology and Evolution*. doi: 10.1093/molbev/msaa330.
Szekely, P. et al. (2015) ‘The Mass-Longevity Triangle: Pareto Optimality and the Geometry of Life-History Trait Space’, *PLoS Computational Biology*. doi: 10.1371/journal.pcbi.1004524.

Tendler, A., Mayo, A. and Alon, U. (2015) ‘Evolutionary tradeoffs, Pareto optimality and the morphology of ammonite shells’, *BMC Systems Biology*, 9(1). doi: 10.1186/s12918-015-0149-z.