Controls for Phylogeny and Robust Analysis in Pareto Task Inference

Miri Adler,[†,1] Avichai Tendler,[†,2] Jean Hausser,[†,3] Yael Korem,[2] Pablo Szekely,[4] Noa Bossel,[5] Yuval Hart,[6] Omer Karin,[2] Avi Mayo,[2] and Uri Alon*,[2]

[1]Klarman Cell Observatory, Eli and Edythe L. Broad Institute of Harvard and MIT, Cambridge, MA, USA
[2]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel
[3]Department of Cellular and Molecular Biology Stockholm, Karolinska Institute and SciLifeLab, Stockholm, Sweden
[4]Department of Biology, Duke University, Durham, NC, USA
[5]Department of Biological Regulation, Weizmann Institute of Science, Rehovot, Israel
[6]Department of Psychology Jerusalem, The Hebrew University of Jerusalem, Jerusalem, Israel
[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: urialonw@gmail.com.
Associate editor: Aida Ouangraoua

Letter

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 T, Iwasaki W. 2021. The flipping t-ratio test: phylogenetically informed assessment of the Pareto theory for phenotypic evolution. Methods Ecol Evol. 12(4):696–706) or hypothetical p-hacking and population-structure concerns (Sun M, Zhang J. 2021. Rampant false detection of adaptive phenotypic optimization by ParTI-based Pareto front inference. Mol Biol Evol. 38(4):1653–1664). 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 the robustness of preprocessing choices. We also provide new methods to control for population structure and detail the experimental tests of ParTI in systems ranging from ammonites to cancer gene expression. The methods presented here may help to improve future ParTI studies.

Key words: phenotypic selection, statistics, systems biology, ecology.

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 task inference 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 et al. 2015), gene expression (Friedman et al. 2020; Hausser and Alon 2020), and life-history traits (Szekely et al. 2015).

Recent papers (Sun and Zhang 2021; 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 Sibling Swap (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 et al. 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

Open Access

ParTI papers, on ammonite shells (Tendler et al. 2015) and on mammalian life-history traits (Szekely et al. 2015).

The study on ammonites specifically aimed to address phylogenetic concerns (Tendler et al. 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 data set 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 data set. 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. 1b and 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 data sets with a star phylogeny (supplementary fig. S1, Supplementary

Material online). 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 Sibling Swap (SibSwap), is simple (fig. 2a): for each set of terminal nodes with a shared parental node (sibling tips, supplementary fig. S2, Supplementary Material online), permute each of the traits independently. This mixes traits between sibling tips (whether in polytomies or not), but not between nonsibling 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 data sets with a star phylogeny (supplementary fig. S1, Supplementary Material online) and performs as well as the flipping t-ratio test on simulated Brownian evolution, figure 2b and supplementary figure S3, Supplementary Material online.
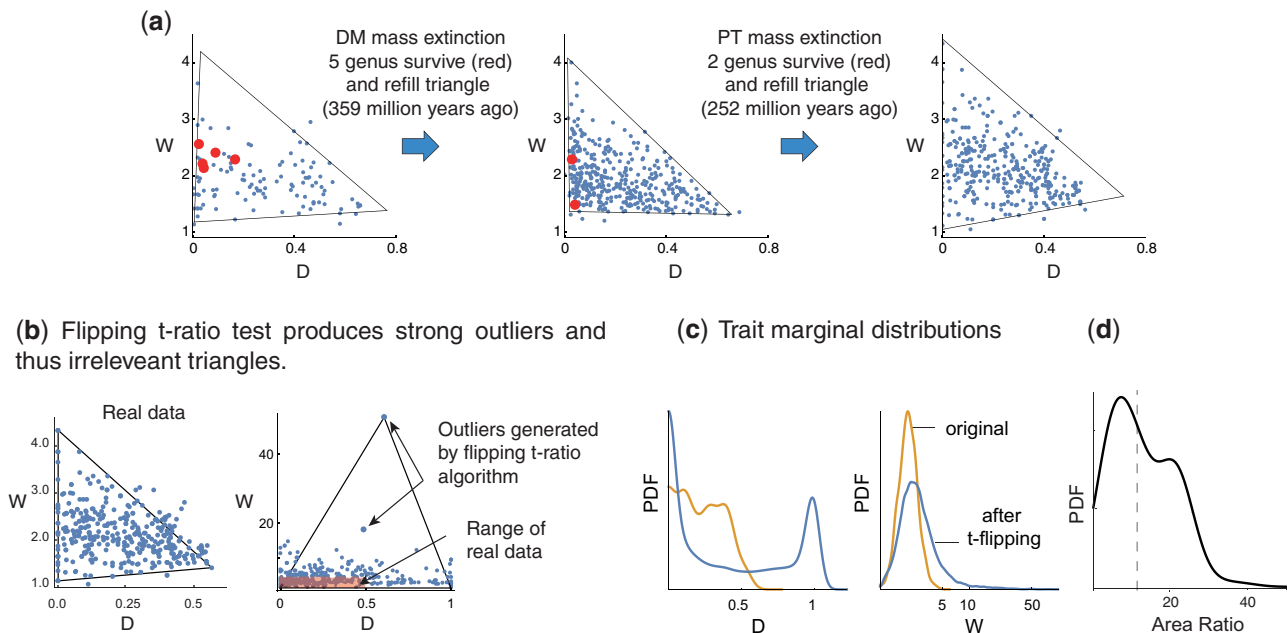
FIG. 1. Convergent evolution in ammonites and spurious triangles in the flipping t-ratio test (a) Ammonites refill statistically the same triangle after mass extinctions. Each point is a genus. W and D are dimensionless shell-shape parameters, the whorl expansion rate and internal/external shell ratio. (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).
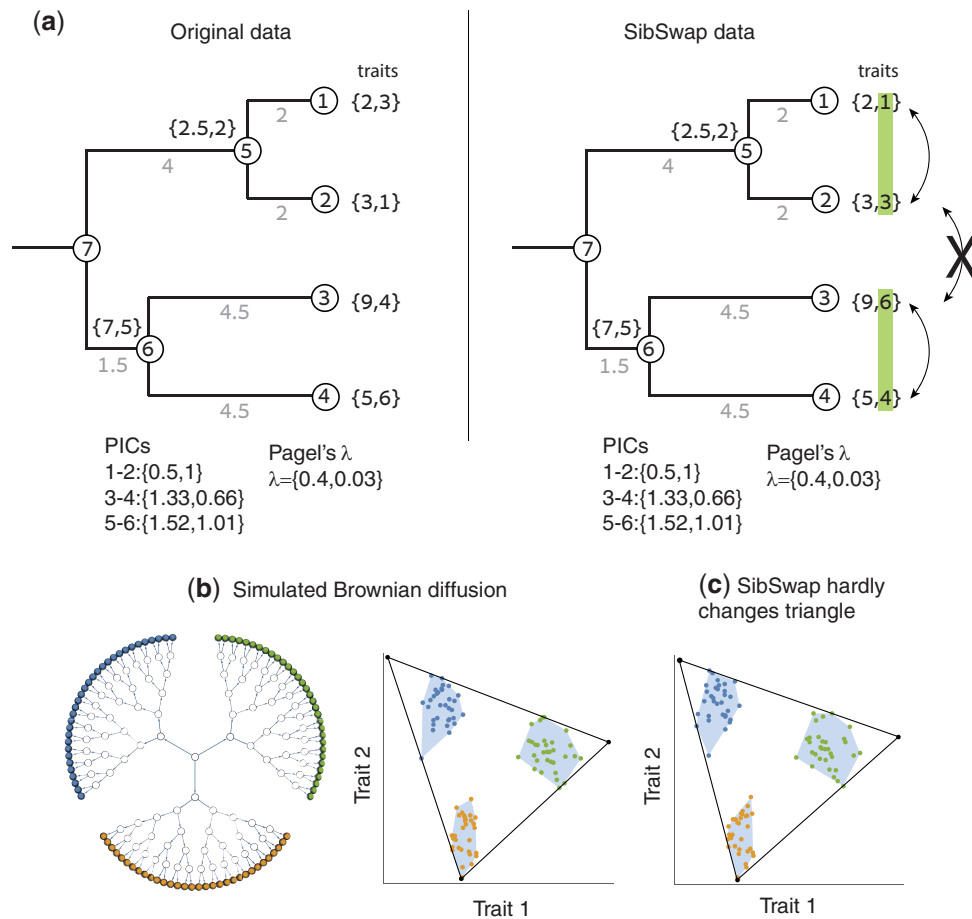
FIG. 2. The SibSwap algorithm preserves trait distributions as well as phylogenetically independent contrasts. (*a*) SibSwap-shuffled result (right) of original data (left) preserves the trait distributions. Here, each terminal node has two traits represented by numbers in curly brackets. Branch lengths are in gray. SibSwap also preserves absolute phylogenetically independent contrasts (PICs) and Pagel's λ, 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.

Importantly, SibSwap preserves the phylogenetically independent contrasts (PICs), defined in Felsenstein (1985), as shown in figure 2*a*, in the standard case where terminal branch lengths are equal (as in ultrametric trees, see supplementary material, Supplementary Material online). SibSwap also preserves any other single-trait statistic, such as Pagel's λ, a common measure for phylogenetic signature (Pagel and Harvey 1989), figure 2*a*. The PIC distributions for ammonite and life-history data sets are indistinguishable in the original and SibSwapped data sets. 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 supplementary material, Supplementary Material online.
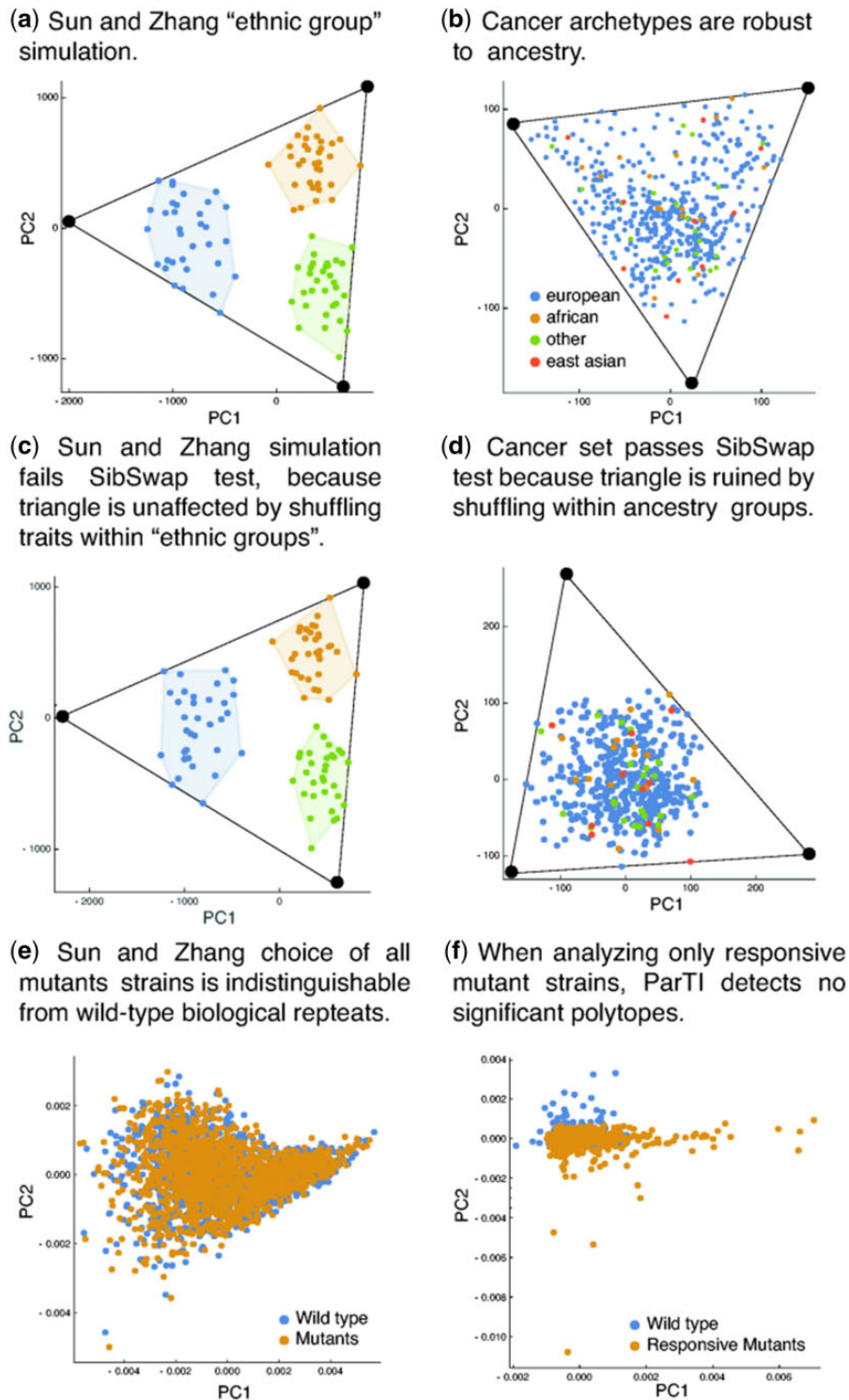
For both ammonite and life-history data sets, the real triangle has a t-ratio significantly closer to 1 than the SibSwap-shuffled data ($P = 0.024$ life-history, $P = 0.012$ ammonite). The reason that phylogenetic effects are not of major importance in these data sets is that ammonite shells and mass-longevity of mammals can evolve rapidly on the timescale of

speciation (Szekely et al. 2015). We conclude that the ParTI inference for these data sets 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 fall in three well-separated clusters due to the three simulated "ethnic groups" (fig. 3*a*), 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 data sets (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 sequences in the gene-expression data set (Carrot-

(a) Sun and Zhang "ethnic group" simulation.

(b) Cancer archetypes are robust to ancestry.

(c) Sun and Zhang simulation fails SibSwap test, because triangle is unaffected by shuffling traits within "ethnic groups".

(d) Cancer set passes SibSwap test because triangle is ruined by shuffling within ancestry groups.

(e) Sun and Zhang choice of all mutants strains is indistinguishable from wild-type biological repeats.

(f) When analyzing only responsive mutant strains, ParTI detects no significant polytopes.

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 data set of Kemmeren et al. (2014) analyzed by Sun and Zhang is indistinguishable from the wild-type biological repeats grown with each strain. (*f*) The responsive mutant data set of Kemmeren et al. (2014) differs from their wild-type repeats and shows no significant ParTI polytope.

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 figure 3*b*. These observations challenge the

hypothesis that population structure is a major factor for ParTI in these cancer data sets.

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 data set essentially the same (fig. 3a and c), whereas the same analysis for the cancer data yields $P < 0.001$, because shuffling within ancestry groups ruins the triangle (fig. 3b and 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 data sets 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 preprocess data for ParTI promotes p-hacking. They do not provide evidence from any particular publication. 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. Supplementary table S1, Supplementary Material online 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 data set that 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 data set they used is nearly identical to the control wild-type data set. 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" data set of Kemmeren et al. (2014), ParTI detects no significant triangle ($P = 0.64$), figure 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 considered 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. Supplementary table S1, Supplementary Material online 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.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Data Availability

Ammonite data (fig. 1) is as reported in Tendler et al. (2015). Life-history data are 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 and f) are as reported in Kemmeren et al. (2014). All algorithms used will be posted in a public repository, GitHub: https://github.com/orgs/AlonLabWIS/repositories.

Acknowledgments

References

Carrot-Zhang J, Chambwe N, Damrauer JS, Knijnenburg TA, Robertson AG, Yau C, Zhou W, Berger AC, Huang K-L, Newberg JY, et al.; Cancer Genome Atlas Analysis Network. 2020. Comprehensive analysis of genetic ancestry and its molecular correlates in cancer. *Cancer Cell.* 37(5):639–654.

Edelaar P. 2013. Comment on "evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space" *Science* 339(6121):757–757.

Felsenstein J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125(1):1.

Freckleton RP, Harvey PH, Pagel M. 2003. Bergmann's rule and body size in mammals. *Am Nat.* 161(5):821–825.

Friedman G, Levi-Galibov O, David E, Bornstein C, Giladi A, Dadiani M, Mayo A, Halperin C, Pevsner-Fischer M, Lavon H, et al. 2020. Cancer-associated fibroblast compositions change with breast cancer progression linking the ratio of S100A4+ and PDPN+CAFs to clinical outcome'. *Nat Cancer.* 1(7):692–708.

Grafen A. 1989. The phylogenetic regression. *Philos Trans R Soc Lond B.* 326(1233):119–157. doi: 10.1098/rstb.1989.0106.

Hart Y, Sheftel H, Hausser J, Szekely P, Ben-Moshe NB, Korem Y, Tendler A, Mayo AE, Alon U. 2015. Inferring biological tasks using Pareto analysis of high-dimensional data. *Nat Methods.* 12(3):233–235.

Hausser J, Szekely P, Bar N, Zimmer A, Sheftel H, Caldas C, Alon U. 2019. Tumor diversity and the trade-off between universal cancer tasks. *Nat Commun.* 10:5423.

Hausser J, Alon U. 2020. Tumour heterogeneity and the evolutionary trade-offs of cancer. *Nat Rev Cancer.* 20(4):247–257.

Kemmeren P, Sameith K, van de Pasch LAL, Benschop JJ, Lenstra TL, Margaritis T, O'Duibhir E, Apweiler E, van Wageningen S, Ko CW, et al. 2014. Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* 157(3):740–752.

Mikami T, Iwasaki W. 2021. The flipping t-ratio test: phylogenetically informed assessment of the Pareto theory for phenotypic evolution. *Methods Ecol Evol.* 12(4):696–706.

Pagel MD, Harvey PH. 1989. Comparative methods for examining adaptation depend on evolutionary models. *Folia Primatol (Basel).* 53(1-4):203–220.

Polly PD. 2012. Phylogenetics for mathematica (Ver. 2.1). Available from: https://scholarworks.iu.edu/dspace/handle/2022/14614 (accessed Oct 17, 2021).

Shoval O, Sheftel H, Shinar G, Hart Y, Ramote O, Mayo A, Dekel E, Kavanagh K, Alon U. 2012. Evolutionary trade-offs, pareto optimality, and the geometry of phenotype space. *Science* 336(6085):1157–1160.

Sun M, Zhang J. 2021. Rampant false detection of adaptive phenotypic optimization by parTI-based Pareto front inference. *Mol Biol Evol.* 38(4):1653–1664.

Szekely P, Korem Y, Moran U, Mayo A, Alon U. 2015. The mass-longevity triangle: pareto optimality and the geometry of life-history trait space. *PLoS Comput Biol.* 11(10):e1004524.

Tendler A, Mayo A, Alon U. 2015. Evolutionary tradeoffs, Pareto optimality and the morphology of ammonite shells. *BMC Syst Biol.* 9:12.