EVOLUTION OF HEAVY METAL HYPERACCUMULATION IN WILD SUNFLOWERS (*HELIANTHUS*)

by

ERIC WAYNE GOOLSBY

(Under the Direction of Lisa A. Donovan)

ABSTRACT

This dissertation proposes novel methods phylogenetic comparative methods for studying complex biological traits and investigates the evolutionary history of heavy metal hyperaccumulation in a comparative framework. Current phylogenetic comparative methods, which limit trait observations to univariate means with normally distributed errors, are unable to incorporate several biologically important trait types, including phenotypic plasticity and dose-response curves. The ability to study such phenotypes, known as function-valued traits, in a comparative framework is critical to unifying the historically disparate fields of evolutionary biology and toxicology. Two methods-based chapters lay out a framework for explicitly accounting for function-valued traits in in the context of macroevolution, representing a substantial improvement in comparative methods which are currently limited regarding the capacity to incorporate complex multivariate traits. The first of these chapters extends phylogenetic generalized least squares methods to allow for function-valued traits assuming a Brownian motion model of evolution. The next chapter expands function-valued and other high-dimensional comparative analyses to allow for the incorporation of alternative evolutionary models,

fixed effects, within-species variation, and missing data, and also addresses issues of statistical power, model flexibility, and computational tractability.

The next two chapters explore the evolution of metal hyperaccumulation in the *Helianthus* genus (sunflowers). First, the elemental defense hypothesis, an adaptive hypothesis for metal hyperaccumulation which predicts a deterrent effect of leaf metals on herbivory, is evaluated in *Helianthus* and the generalist herbivore *Vanessa cardui* (the Painted Lady butterfly) using a comparative approach. Mixed support for the elemental defense hypothesis is found, with herbivores exhibiting a preference for leaves from non-metal-treated plants over metal-treated plants in certain species. However, in the absence of a choice, *V. cardui* were not deterred by leaf metals. Next, the evolutionary history of As, Cd, Cr, Cu, Ni, Pb, Se, and Zn accumulation was investigated across *Helianthus*. Hyperaccumulation of Cd, Ni, and Zn is widespread throughout *Helianthus*. Cd and Zn hyperaccumulation, as well as elevated Ni accumulation, likely evolved in wild *Helianthus* prior to sunflower domestication. These results, in conjunction with the comparative methods developed in this dissertation, provide a framework for further investigation of metal hyperaccumulation as a function-valued trait.

INDEX WORDS: Elemental defense hypothesis, Function-valued traits, *Helianthus*, Herbivory, High-dimensional phenotypes, Macroevolution, Metal hyperaccumulation, Phylogenetic comparative methods, Sunflower.

EVOLUTION OF HEAVY METAL HYPERACCUMULATION IN WILD SUNFLOWERS (*HELIANTHUS*)

by

ERIC WAYNE GOOLSBY

B.S., University of Georgia, 2012

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2016

EVOLUTION OF HEAVY METAL HYPERACCUMULATION IN WILD SUNFLOWERS (*HELIANTHUS*)

by

ERIC WAYNE GOOLSBY

Major Professor:	Lisa A. Donovan
Committee:	Richard P. Shefferson
	Marsha C. Black
	Stacey L. Lance

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2016

DEDICATION

To my Mother, Kay Goolsby, for 28 years of unconditional love and support, encouragement, laughter, and the happiest of memories.

To my Father, Louis Goolsby, who introduced me to the wondrous world of science, for helping me to pursue my ever-evolving interests from the very beginning, from seismology to space.

To my husband, principal research collaborator, and best friend, Chase Mason, for being a constant source of inspiration and enthusiasm.

ACKNOWLEDGEMENTS

I would like to thank my advisors Lisa Donovan and Richard Shefferson, for continued support and guidance throughout my graduate career. I would also like to thank my undergraduate research mentor and graduate committee member Marsha Black for fostering my initial and continued development as a scientist, and Stacey Lance for helping me to integrate my research in an evolutionary ecotoxicology framework. I also wish to thank my undergraduate major advisor Anne Marie Zimeri, who first inspired me to pursue a graduate degree and introduced me to the fascinating plant trait metal hyperaccumulation. I would also like to acknowledge the staff of the Department of Plant Biology, the Odum School of Ecology, and the Interdisciplinary Toxicology Program.

I would like to thank the Donovan lab as well as the greater Sunflower Mafia family, in particular co-authors Chase Mason, Alan Bowsher, Ethan Milton, Caitlin Ishibashi, and Devon Humphreys. I would also like to acknowledge the hard work and enthusiasm of my undergraduate researchers and collaborators.

Finally, I would like to thank the funding sources that made this work possible, including the National Science Foundation Graduate Research Fellowship (DGE-1443117), Doctoral Dissertation Improvement Grant (DEB-1501215), and IOS-1122842, as well as the Graduate School Innovative and Interdisciplinary Research Grant, departmental research grants from the Odum School of Ecology and the Interdisciplinary Toxicology Program, and the Sigma Xi Grants-in-Aid of Research.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Hyperaccumulation is a rare plant trait by which plants accumulate extremely high concentrations of metals (e.g., ≥1,000 ppm) in leaf tissues without suffering toxicity (Reeves and Baker 2000). Although the physiology and genetics of this phenomenon have been extensively studied, very little information exists regarding the complex evolutionary dynamics of tolerance and hyperaccumulation, its adaptive significance, and potential tradeoffs with key functional plant traits. Studying the evolution of hyperaccumulation requires the integration of two disparate fields – evolutionary biology and toxicology, the study of which are both intrinsically linked to understanding the physiological responses of organisms to their environment through time (Bickham 2011). Despite their common objectives, theoretical obstacles and practical limitations have historically restricted the integrated study of evolution and toxicology (Bickham 2011, Goolsby 2015).

The study of phenotypic evolution is itself complicated by the non-independence of species phenotypes due to shared ancestry. Three decades ago, the basis of modern phylogenetic comparative methods emerged with the development of Felsenstein's phylogenetically independent contrasts (Felsenstein 1985). Since then, comparative methods have seen an explosion in capabilities – ancestral state reconstruction (Garland et al. 1999), estimation of phylogenetic signal (Pagel 1999, Blomberg et al. 2003), comparing evolutionary rates (O'Meara et al. 2006), estimation of non-Brownian

evolutionary models (Hansen 1997, Harmon et al. 2010), and the ability to account for within-species variation (Ives et al. 2007, Felsenstein 2008, Hansen and Bartoszek 2012) and missing data (Bruggeman et al. 2009). Although these methods are primarily focused on the evolution of univariate traits, more recent methods have sought to incorporate a multivariate framework, including phylogenetic principal components analysis (PCA) (Revell 2009), phylogenetic path analysis (Hardenberg and Gonzalez-Voyer 2013), and comparative analysis of high-dimensional morphometric traits (Adams 2014c, b, a, Adams and Felice 2014). Despite the numerous and diverse capabilities offered by univariate phylogenetic comparative methods, these approaches are seldom integrated for multivariate analyses. For instance, phylogenetic PCA and high-dimensional morphometric analyses offer no approaches for dealing with non-Brownian evolutionary models or within-species variation. These limitations are problematic, as several simulated and empirical analyses have demonstrated that substantial bias is introduced when intraspecific variation is ignored (Hansen and Bartoszek 2012, Silvestro et al. 2015) or when a Brownian motion model of evolution is inappropriately applied (Pennell et al. 2015, Uyeda et al. 2015).

In approaching toxicology in an evolutionary framework, current phylogenetic comparative methods fall short as they are unable to incorporate phenotypically plastic traits, such as developmental trajectories, environmental variation, and exposure to toxic substances. This is because comparative methods operate under the assumption that species trait values are fixed with normally distributed error. Although appropriate for some traits, many traits relevant to evolutionary biology (particularly those relating to toxicology) are phenotypically plastic along environmental and temporal gradients (e.g.,

phenotypic responses to abiotic stress, developmental trajectories, etc.). Such traits are more appropriately described by mathematical functions rather than univariate trait means.

This dissertation seeks to bridge the current barriers separating evolutionary biology and toxicology, drawing from the rapidly emerging theoretical, computational, and empirical resources relevant to their study. The second chapter proposes statistical tools for expanding phylogenetic comparative methods for the incorporation of function-valued traits. The ability to directly study the evolutionary interactions of function-valued traits is critical for the study of many biologically important traits, as well as for integrating toxicological data, such as dose-response curve, in a macroevolutionary framework. These tools are fully compatible with several recently developed high-dimensional comparative methods, which can be used to test for phylogenetic signal (Adams 2014a), assess correlated trait evolution (Adams 2014b, Adams and Felice 2014), perform phylogenetic ANOVA (Adams 2014b), and compare evolutionary rates (Adams 2014c), all in a function-valued context.

With the flexibility of high-dimensional comparative methods also comes its associated limitations – an apparent inability to incorporate intraspecific function-valued variation and alternative evolutionary models (Adams 2014c, b, a, Adams and Felice 2014). In addition to function-valued phenotypes, many important biological traits are inherently multidimensional, such as shape/morphometric data, physiological suites/syndromes, and tolerance to toxic substances. For such traits, the number of dimensions for a given trait may be large, often exceeding the number of observed individuals or species. Conventional phylogenetic comparative statistical methods are

unable to handle such high-dimensional phenotypes, as the number of parameters to estimate increases polynomially with increases in trait dimensionality (Ho and Ane 2014). Consequently, dramatic reductions in statistical power, intractable log-likelihoods, and lack of parameter identifiability inhibit meaningful inference for such traits. Recent advances in high-dimensional comparative methods have enabled simple investigations but are limited to a small class of model specifications. The third chapter of this dissertation addresses these limitations using a likelihood-based extension of the methods proposed in the second chapter that eliminates these shortcomings for both function-valued and other high-dimensional morphometric traits. This chapter develops a flexible framework for analyzing high-dimensional traits in a phylogenetic comparative context which allows for complex model specification while maintaining high statistical power and parameter identifiability. These novel methods allow for ultra-high dimensional phenotypes to be studied using both frequentist and Bayesian approaches.

Unlike most hyperaccumulator species, domesticated sunflower (*Helianthus annuus*) is a hyperaccumulator of multiple metals, including As, Cd, Cr, and Ni. *H. annuus* is also capable of accumulating high concentrations of several other metals, such as Cu, Fe, Zn, Mn, U, Cs, and Sr (Blamey et al. 1986, Lin et al. 2003, Prasad and Freitas 2003, Solhi et al. 2005, Cutright et al. 2010, Walliwalagedara et al. 2010). Wild *Helianthus* occurs across North America in a variety of habitats, including serpentine soils, beaches, coastal prairies, deserts, salt marshes, mountain meadows, and open woodlands, and are adapted to a range of strong environmental stressors, such as drought, salinity, and low nutrient availability (Heiser et al. 1969, Stephens et al. 2015). A wealth of genetic, phylogenetic, and physiological data and resources are available for wild and

domesticated sunflowers (Heiser et al. 1969, Timme et al. 2007, Mandel et al. 2011, Mandel et al. 2013, Stephens et al. 2015), making *Helianthus* ideal study system for investigating metal hyperaccumulation and placing it in an integrated evolutionary toxicology framework.

The fourth chapter of this dissertation tests the most commonly cited explanation regarding selective benefits of metal hyperaccumulation, the elemental defense hypothesis, which predicts an anti-herbivore or anti-pathogen effect of high leaf metal concentrations. Specifically, the anti-herbivore aspect of the elemental defense hypothesis is evaluated in a phylogenetic context using a small clade within wild *Helianthus* closely related to domesticated sunflower using control, nickel, and cadmium treatments. The fifth chapter investigates the evolutionary histories of eight different metals and metalloids (As, Cd, Cr, Cu, Ni, Pb, Se, and Zn) in the entire *Helianthus* genus. Together, these chapters develop a framework for the study of metal hyperaccumulation in the context of evolutionary toxicology.

REFERENCES

Adams, D. C. 2014a. A Generalized K Statistic for Estimating Phylogenetic Signal from Shape and Other High-Dimensional Multivariate Data. *Systematic Biology* 63:685-697.

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

Adams, D. C. 2014c. Quantifying and Comparing Phylogenetic Evolutionary Rates for Shape and Other High-Dimensional Phenotypic Data. *Systematic Biology* 63:166-177.

Adams, D. C., and R. N. Felice. 2014. Assessing Trait Covariation and Morphological Integration on Phylogenies Using Evolutionary Covariance Matrices. *Plos One* 9.

Bickham, J. W. 2011. The four cornerstones of Evolutionary Toxicology. *Ecotoxicology* 20:497-502.

Blamey, F. P. C., D. C. Joyce, D. G. Edwards, and C. J. Asher. 1986. Role of trichomes in sunflower tolerance to manganese toxicity. *Plant and Soil* 91:171-180.

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

Bruggeman, J., J. Heringa, and B. W. Brandt. 2009. PhyloPars: estimation of missing parameter values using phylogeny. *Nucleic Acids Research* 37:W179-184.

Cutright, T., N. Gunda, and F. Kurt. 2010. Simultaneous Hyperaccumulation of Multiple Heavy Metals by *Helianthus Annuus* Grown in a Contaminated Sandy-Loam Soil. *International Journal of Phytoremediation* 12:562-573.

Felsenstein, J. 1985. Phylogenies and the comparative method. *American Naturalist* 125:1-15.

Felsenstein, J. 2008. Comparative methods with sampling error and within-species variation: Contrasts revisited and revised. *American Naturalist* 171:713-725.

Garland, T., P. E. Midford, and A. R. Ives. 1999. An Introduction to Phylogenetically Based Statistical Methods, with a New Method for Confidence Intervals on Ancestral Values. *American Zoologist* 39:374-388.

Goolsby, E. W. 2015. Phylogenetic Comparative Methods for Evaluating the Evolutionary History of Function-Valued Traits. *Systematic Biology* 64:568-578.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51:1341-1351.

Hansen, T. F., and K. Bartoszek. 2012. Interpreting the Evolutionary Regression: The Interplay Between Observational and Biological Errors in Phylogenetic Comparative Studies. *Systematic Biology* 61:413-425.

Hardenberg, A. v., and A. Gonzalez-Voyer. 2013. Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis. *Evolution* 67:378-387.

Harmon, L. J., J. B. Losos, T. J. Davies, R. G. Gillespie, J. L. Gittleman, W. B. Jennings, K. H. Kozak, M. A. McPeek, F. Moreno-Roark, T. J. Near, A. Purvis, R. E. Ricklefs, D. Schlüter, J. A. S. Ii, O. Seehausen, B. L. Sidlauskas, O. Torres-Carvajal, T. W. Jason, and A. Ø. Mooers. 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64:2385-2396.

Heiser, C. B. J., D. M. Smith, S. B. Clevenger, and W. C. J. Martin. 1969. *The North American Sunflowers-D Helianthus-D*. Memoirs of the Torrey Botanical Club 22:1-218.

Ho, L. S. T., and C. Ane. 2014. A Linear-Time Algorithm for Gaussian and Non-Gaussian Trait Evolution Models. *Systematic Biology* 63:397-408.

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

Lin, J., W. Jiang, and D. Liu. 2003. Accumulation of copper by roots, hypocotyls, cotyledons and leaves of sunflower (Helianthus annuus L.). *Bioresource Technology* 86:151-155.

Mandel, J. R., J. M. Dechaine, L. F. Marek, and J. M. Burke. 2011. Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus* L. *Theoretical and Applied Genetics* 123:693-704.

Mandel, J. R., S. Nambeesan, J. E. Bowers, L. F. Marek, D. Ebert, L. H. Rieseberg, S. J. Knapp, and J. M. Burke. 2013. Association Mapping and the Genomic Consequences of Selection in Sunflower. *Plos Genetics* 9.

O'Meara, B. C., C. Ané, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60:922-933.

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

Pennell, M., R. G. FitzJohn, W. K. Cornwell, and L. J. Harmon. 2015. Model adequacy and the macroevolution of angiosperm functional traits. *American Naturalist* 186:E33-E50.

Prasad, M. N. V., and H. M. D. Freitas. 2003. Metal hyperaccumulation in plants - Biodiversity prospecting for phytoremediation technology. *Electronic Journal of Biotechnology* 6:285-321.

Reeves, R. D., and A. J. M. Baker. 2000. Metal-accumulating plants. *Phytoremediation of Toxic Metals*:193-229.

Revell, L. J. 2009. Size-correction and Principal Components for Interspecific Comparative Studies. *Evolution* 63:3258-3268.

Silvestro, D., A. Kostikova, G. Litsios, P. B. Pearman, and N. Salamin. 2015. Measurement errors should always be incorporated in phylogenetic comparative analysis. *Methods in Ecology and Evolution* 6:340-346.

Solhi, M., H. Shareatmadari, and M. Hajabbasi. 2005. Lead and Zinc Extraction Potential of Two Common Crop Plants, *Helianthus Annuus* and *Brassica Napus*. *Water, Air, and Soil Pollution* 167:59-71.

Stephens, J. D., W. L. Rogers, C. M. Mason, L. A. Donovan, and R. L. Malmberg. 2015. Species tree estimation of diploid Helianthus (Asteraceae) using target enrichment. *American Journal of Botany* 102:910-920.

Timme, R. E., B. B. Simpson, and C. R. Linder. 2007. High-resolution phylogeny for Helianthus (Asteraceae) using the 18s-26s ribosomal DNA external transcribed spacer. *American Journal of Botany* 94:1837-1852.

Uyeda, J. C., D. S. Caetano, and M. W. Pennell. 2015. Comparative analysis of principal components can be misleading. *Systematic Biology*.

Walliwalagedara, C., I. Atkinson, H. van Keulen, T. Cutright, and R. Wei. 2010. Differential expression of proteins induced by lead in the Dwarf Sunflower *Helianthus annuus*. *Phytochemistry* 71:1460-1465.

CHAPTER 2

PHYLOGENETIC COMPARATIVE METHODS FOR EVALUATING THE EVOLUTIONARY HISTORY OF FUNCTION-VALUED TRAITS[1]

[1]**Goolsby EW**. 2015. Phylogenetic comparative methods for evaluating the evolutionary history of function-valued traits. *Systematic Biology* 64:568-578. Reprinted here with permission of the publisher.

ABSTRACT

Phylogenetic comparative methods offer a suite of tools for studying trait evolution. However, most models inherently assume fixed trait values within species. Although some methods can incorporate error around species means, few are capable of accounting for variation driven by environmental or temporal gradients, such as trait responses to abiotic stress or ontogenetic trajectories. Such traits, often referred to as function-valued or infinite-dimensional, are typically expressed as reaction norms, dose-response curves, or time plots and are described by mathematical functions linking independent predictor variables to the trait of interest. Here, I introduce a method for extending ancestral state reconstruction to incorporate function-valued traits in a phylogenetic generalized least squares (PGLS) framework, as well as extensions of this method for testing phylogenetic signal, performing phylogenetic ANOVA, and testing for correlated trait evolution using recently proposed multivariate PGLS methods. Statistical power of function-valued comparative methods is compared to univariate approaches using data simulations, and the assumptions and challenges of each are discussed in detail.

INTRODUCTION

Phylogenetic comparative methods such as ancestral state reconstruction, phylogenetically independent contrasts (PICs), and phylogenetic generalized least squares (PGLS) comprise a suite of tools for inferring the evolutionary history of traits and for testing hypotheses of trait evolution (Felsenstein 1985; Grafen 1989; Martins and Hansen 1997). Although intraspecific differences due to measurement error, sampling error, or natural variation can be incorporated into error structures, these models generally operate under the assumption that individual species possess a fixed mean for any given trait (Martins and Hansen 1997; Ives et al. 2007). Many traits relevant to ecology and evolutionary biology are inherently function-valued, including developmentally varying traits, niche preference, thermal performance, life history patterns, plant ecophysiological strategies, responses to environmental hazards such as heavy metals and anthropogenic contaminants, and tolerance to abiotic stressors such as salinity, drought, and nutrient limitation. Rather than a single mean value, function-valued traits are best represented by curves defined by mathematical functions. Accordingly, univariate analytical approaches to such traits, which fail to account for trait dependence on independent exogenous variables such as time or environment, are statistically inappropriate (Kingsolver et al. 2001; Rocha and Klaczko 2012; Stinchcombe et al. 2012; Donovan et al. 2014).

Until recently, standard phylogenetic comparative methods have lacked tools for direct incorporation of function-valued analyses. Instead, many researchers have approached evolutionary questions regarding function-valued traits by summarizing trait curves with univariate mean values and applying these values directly to comparative

methods. This approach typically utilizes one of two metrics: mean trait values evaluated at specific environmental levels (Pollard et al. 2001; Pigliucci et al. 2003), or inverse functions evaluated at a standardized biological response (Guenard et al. 2011; Strachan et al. 2011; Larras et al. 2014). The inverse of a function $f(x) = y$ is the function $f^{-1}(y) = x$ that gives the value of *x* that is expected to produce a response *y*. For example, in toxicological studies, statistical responses to chemicals are typically expressed as inverse functions rather than response at a specific dose (the 50% lethal response dose (LD50) of a chemical is the dose expected to cause mortality in 50% of exposed subjects). For linear functions with similar slopes, both approaches typically yield consistent results; for nonlinear functions, approaches using mean trait values across treatment levels are unreliable. Conversely, inverse evaluation of monotonic functions at analogous reference trait values across species (e.g., LD50) is a valid approach regardless of linearity or nonlinearity. This is because inverse function evaluation at a fixed biological endpoint returns the value of the independent variable that produces said response, and is thus not subject to weighting by curve thresholds or other nonlinear features. However, summary values such as LD50 only capture a single dimension of function-valued traits and are likely to mask important trait variation across species. Consider two aquatic species exhibiting identical LD50 values for salinity toxicity, but the slope of the dose-response curve for one species is much steeper than the dose-response curve of the other species. In this example, interpretation of the LD50s in isolation might lead to the incorrect conclusion that salinity tolerance in the two species is identical. Similarly, other univariate summaries of function-valued traits, such as area under the curve, fail to distinguish biological differences such as generalist-specialist

shifts in which the shape of the curve changes but the total area under the curve remains constant (Izem and Kingsolver 2005; Stinchcombe et al. 2012). For these reasons, evolution of function-valued traits should be studied using methods that incorporate entire species trait curves rather than univariate summaries.

One particularly important challenge related to function-valued traits is the study of plasticity, which is the focus of the next section. However, it should be noted that the principles of function-valued traits discussed in the next section, in addition to trait plasticity, apply broadly to other classes of function-valued traits as well (e.g., environmental tolerance, ontogenetic variation, etc.).

Plastic traits: motivation for study, common approaches, and challenges. A central question implicit to the study of phenotypic plasticity is how plasticity affects evolutionary processes, such as gene flow, speciation, diversification, and persistence in novel environments (Wund 2012). Questions concerning the evolutionary dynamics of phenotypic plasticity, plasticity costs, and tradeoffs are also of particular interest (Callahan et al. 2008). Such questions may be applied to exploring evolvability and rates of evolutionary change of both plastic and fixed traits to better understand or anticipate species responses to processes such as environmental pollution (Guenard et al. 2011; Guenard et al. 2013; Larras et al. 2014), climate change (Bradshaw and Holzapfel 2006), and species invasions (Richards et al. 2006). By inferring the evolutionary history of trait responses to environmental conditions, evolutionary hypotheses relating to phenotypic plasticity, such as adaptive plasticity, plasticity costs, and tradeoffs in plastically varying traits, may be tested. Additionally, by taking into account phylogenetic variation in reaction norms, certain patterns previously obscured or falsely detected, either due to

sampling traits in too few or too narrowly varying environments, assuming linearity of nonlinear reaction norms, or failing to consider interspecific differences in phenotypic minimum and maximum values, may be clarified.

Traits are often sampled under natural field conditions and assumed to be fixed (non-plastic), despite clear evidence in many traits of substantial variation driven by environment and development (Whitman and Agrawal 2008; Mason et al. 2013; Donovan et al. 2014). By failing to account for environmental or ontogenetic variation, or by assuming that environmental differences can be treated as covariates that uniformly affect all species of interest, statistical and biological inferences may be unreliable (Fig. 2.1a) (Rocha and Klaczko 2012). Additionally, environmental variation may obscure evolutionary signatures, leading to the incorrect conclusion that a trait is too labile for phylogenetic analysis (Revell 2010).

As an alternative to field measurements, studies may be performed in homogenous laboratory settings to minimize plastic intraspecific variation, with the intention of "removing" the effect of environment and allowing "baseline" phenotypic expression. However, uniform environments fail to capture the full extent of trait combinations possible within a species; instead, only a snapshot of possible trait combinations is identified, which is potentially problematic when making evolutionary inferences on putatively adaptively plastic traits. Another strategy involves estimating the difference in field trait values with controlled common environment trait values as a measure of total trait plasticity (Knight and Ackerly 2003). Although an improvement, this method has several limitations: 1) the ability to isolate specific drivers of trait plasticity is limited; 2) providing a single ideal environment for multiple species native to

diverse habitats may not be possible, so some species might be stressed under common environment conditions; and 3) it is unlikely that the full range of trait plasticity is captured by sampling traits under field and single-treatment lab conditions (Fig. 2.1b) (Jacobs and Latimer 2012).

Researchers can capture a more comprehensive representation of the capacity for plasticity by exposing species to a controlled common environment while manipulating a single environmental variable. Such environmental manipulations often involve treating continuous variables as arbitrarily categorized treatments (e.g., "low" and "high"), which inherently impose the assumption of linearity and normality of data. Accordingly, problems are likely to arise if the slope, strength, or shape of reaction norms or plastically induced shifts in correlated traits differs across species (Fig. 2.1c) (Rocha and Klaczko 2012). Furthermore, categorical environmental manipulations pose several analytical challenges. In particular, trait measurements must be made under comparable treatment levels, and sampling differences and missing data are incompatible with many statistical methods. Additionally, the magnitude of arbitrarily categorized or binned continuous treatment values is ignored, thus sacrificing statistical power and limiting the ability to predict trait values at unobserved levels (Griswold et al. 2008; Stinchcombe et al. 2012).

Evaluating traits at multiple levels along a continuous environmental gradient offers the opportunity to more fully assess the function-valued relationship between environment and plastically varying traits (Fig. 2.1d) (Kingsolver et al. 2001; Rocha and Klaczko 2012; Stinchcombe et al. 2012; Murren et al. 2014). In general, function-valued analyses offer several advantages to other approaches: 1) they are robust to sampling differences; 2) there is no need for measurements to be conducted at identical treatment

levels (provided measurements provide sufficient coverage of the range of interest to accurately estimate function parameters); 3) they have higher statistical power; 4) they incorporate the magnitude and trend from continuous predictor variables; and 5) they allow prediction of trait responses in extant and ancestral taxa at unobserved levels (Kingsolver et al. 2001).

A novel approach for function-valued phylogenetic comparative methods. Although several studies have implemented function-valued analyses for studying and predicting the evolution of function-valued traits within a single species (Stinchcombe et al. 2012), methods for analyzing function-valued traits in an explicitly phylogenetic context have only recently begun to emerge. Recently, a method called phylogenetic Gaussian process regression (PGPR), an extremely flexible method for reconstructing the evolutionary history of function-valued traits, was proposed (Aston et al. 2012; Hadjipantelis et al. 2013; Jones and Moriarty 2013). In PGPR, curves of any shape may be analyzed without *a priori* assumptions of function structure, multiple sources of uncertainty may be incorporated, and a variety of evolutionary models may be tested (Hadjipantelis et al 2013). Despite its flexibility, PGPR currently lacks compatibility with several commonly used comparative methods such as phylogenetic regression and estimation of phylogenetic signal, as such methods inherently rely on phylogenetic generalized least squares (PGLS)-based reconstruction of ancestral states (Martins and Hansen 1997; Blomberg et al. 2003).

Here, I present a function-valued extension of ancestral state reconstruction in a PGLS framework. This method inherits the flexibility of PGLS and is fully compatible with several recently proposed multivariate PGLS-based methods which can be used to

test for phylogenetic signal (Adams 2014a), perform phylogenetic ANOVA (Adams 2014b), evaluate correlated trait evolution between functions and univariate traits (Adams 2014b) or between multiple functions (Adams and Felice 2014), and test for shifts in evolutionary rates of function-valued traits (Adams 2014c).

METHODS

Conceptually, a function-valued trait is composed of infinite discrete landmarks along a curve that evolve as a single multivariate trait (Adams 2014a). Representation of a function for a given species can be approximated with a finite-length sequence of landmarks described by *x*,*y* coordinates. However, prior to performing PGLS-based comparative analyses, functions must first be aligned to make landmarks analogous across species (Adams 2014a). Depending on specific function-valued attributes, landmark alignment may be accomplished using a variety of techniques, such as inverse function alignment, dynamic time warping (Myers et al. 1980; Giorgino 2009), and generalized time warping (Zhou and De la Torre 2012), as described in the following two sections.

Ancestral curve reconstruction of sigmoidal curves. Monotonic functions with constant minimum and maximum values, such as sigmoidal functions (e.g., as with binomial, proportional, or probabilistic data), are a special case of function-valued traits in which the *y*-axis represents absolute endpoints that are comparable across species. In such cases, the *y*-axis serves as a pre-aligned point of reference, and function-valued evolution of these types of curves need only be expressed along the *x*-axis using inverse functions evaluated at a constant vector of *y*-values spanning minimum and maximum

function values (e.g., between 0 and 1) for each species. In this way, the trait being studied is the vector of *x*-values that corresponds to a fixed vector of reference *y*-values for a given species. For example, in dose-response curves, the LD50 is a comparable landmark across species that can be appropriately analyzed using conventional univariate phylogenetic comparative methods. Conversely, the response at a specific dose is not comparable across species due to the non-linearity of sigmoidal curves, and inferences based on such approaches are incorrect.

Univariate ancestral state reconstruction can be easily extended to incorporate function-valued traits using a PGLS framework (Martins and Hansen 1997). First, parameters for function-valued traits $f(x)$ are estimated for each of *N* species. The reconstructed vector of function-valued landmarks at the root of the tree, $\boldsymbol{\mu}_{root}$, is then calculated using the generalized least squares formula

$$\boldsymbol{\mu}_{root} = (\mathbf{1}^t\mathbf{C}^{-1}\mathbf{1})^{-1}\mathbf{1}^t\mathbf{C}^{-1}\mathbf{Y} \quad (1)$$

where $\mathbf{C}$ is the matrix of expected trait variance-covariance given by the phylogenetic tree and assuming a specific model of trait evolution (e.g., Brownian motion), $\mathbf{1}$ is an $N \times 1$ matrix of ones, y is a vector of n evenly spaced *y*-axis landmarks between a and b (0 and 1 for binomial data), and $\mathbf{Y}$ is the $N \times n$ matrix of inverse functions f_{tips}^{-1} evaluated at *y*. Root vector $\boldsymbol{\mu}_{root}$ is extended into an $N \times n$ matrix $\mathbf{R}$, and each row is filled with $\boldsymbol{\mu}_{root}$. Next, the matrix of M internal node landmark vectors is calculated

$$\boldsymbol{\mu}_{nodes} = \left(\mathbf{D}\mathbf{C}^{-1}(\mathbf{Y} - \mathbf{R})\right)^t + \mathbf{R}^{\mathrm{t}}{}_{n\times M} \quad (2)$$

where $\mathbf{D}$ is the $M \times N$ matrix of expected covariance between internal nodes and tips as specified by the phylogenetic tree and model of evolution, $\mathbf{R}^{\mathrm{t}}{}_{n\times M}$ is an $n \times M$ matrix consisting of *M* columns each filled with $\boldsymbol{\mu}_{root}$. $\boldsymbol{\mu}_{nodes}$ represents $f_{nodes}^{-1}(y)$ in the set of

landmark coordinates $(f_{nodes}^{-1}(y), y)$, which can then be used to estimate parameters for curves at each node. A hypothetical example of ancestral curve reconstruction of a sigmoidal trait is visualized in figure 2.2.

Variance and 95% confidence intervals may also be estimated for reconstructed ancestral curves (Rohlf 2001). First, the Brownian motion parameter σ^2 is estimated for each landmark *i* from 1 to *n*

$$\sigma_i^2 = \frac{\left(\left(\mathbf{Y}_i - \boldsymbol{\mu}_{root_i}\right)^t \mathbf{C}^{-1}\left(\mathbf{Y}_i - \boldsymbol{\mu}_{root_i}\right)\right)^t}{N-1} \tag{3}$$

Next, variance is calculated for each reconstructed landmark at each internal node

$$\sigma_i^2{}_{nodes} = \sigma_i^2 \mathrm{diag}\big(\mathbf{D}^* - (\mathbf{D}\mathbf{C}^{-1}\mathbf{D}^{\mathrm{t}}) + (\mathbf{1}^* - \mathbf{D}\mathbf{C}^{-1}\mathbf{1})(\mathbf{1}^t\mathbf{C}^{-1}\mathbf{1})^{-1}(\mathbf{1}^* - \mathbf{D}\mathbf{C}^{-1}\mathbf{1})\big) \tag{4}$$

where $\mathbf{D}^*$ is the $M \times M$ matrix of expected covariance among internal nodes due to phylogeny and the specified evolutionary model, and $\mathbf{1}^*$ is an M-length vector of ones (Cressie 1993). Finally, 95% confidence intervals for each curve can be calculated as $\boldsymbol{\mu}_{nodes} \pm 1.96(\sigma_{nodes})$.

Ancestral curve reconstruction of other function types. Many function-valued traits cannot be appropriately expressed with *y*-bounded sigmoidal functions. In such cases, inverse function alignment does not properly align curves across species. For instance, unimodal and multimodal functions lack a unique inverse solution for any given *y*-value. Additionally, certain function-valued traits require complex curve specification such as with Gaussian process regression. For such traits, ancestral curve reconstruction must be performed on both the *x*-axis and the *y*-axis, both of which must first be aligned across species. This can be accomplished for most function-valued relationships using a

non-linear alignment method called generalized time warping (Zhou and De la Torre 2012), a multi-sequence extension of dynamic time warping (Myers et al. 1980; Giorgino 2009). Although applicable to a wide variety of complex curve shapes, the appropriateness of time warping methods to specific function types should be assessed, and individual time warping parameters may require specification. For a more in-depth discussion, see Giorgino (2009).

For ancestral state reconstruction of aligned curves, Equations (1) and (2) must be applied individually to both the *x*-axis and the *y*-axis. For *x*-axis ancestral state reconstruction, Y is the $N \times n$ matrix of aligned *x*-values; for *y*-axis ancestral state reconstruction, Y is the $N \times n$ matrix of aligned *y*-values. The resulting set of μ_{nodes} forms coordinates (x_{nodes}, y_{nodes}) which are then used to estimate parameters for f_{nodes}. Similarly, variance and 95% confidence intervals for reconstructed curves must be applied on both the *x* and *y* axes for curves aligned with time warping. For monotonic sigmoidal function-valued traits, inferences from inverse function approaches and time warping alignment are equivalent (R code provided in supplementary material).

Compatibility with distance-based multivariate phylogenetic comparative methods. Recently, several multivariate extensions of phylogenetic comparative methods have been proposed for high-dimensional traits (Adams 2014a; Adams 2014b; Adams 2014c; Adams and Felice 2014). These methods allow quantification of metrics such as multivariate phylogenetic signal, correlated trait evolution, phylogenetic ANOVA, and evolutionary rates. Distance-based comparative methods rely on removal of phylogenetic covariance from data by projecting phenotypic data onto the phylogenetic matrix **E** (Garland and Ives 2000):

$$\mathbf{E} = \left(\mathbf{U}\mathbf{W}^{\mathbf{1/2}}\mathbf{U}^{\boldsymbol{t}}\right)^{-\mathbf{1}} \quad (5)$$

where $\mathbf{U}$ and $\mathbf{W}$ are eigenvectors and eigenvalues of the phylogenetic covariance matrix $\mathbf{C}$. Although primarily presented as tools for analyzing Procrustes-aligned morphometric data (Rohlf and Slice 1990), these methods are not limited to the number of supplied trait dimensions and can handle hundreds of non-isotropic covarying dimensions. In particular, unlike covariance-based approaches, distance-based multivariate comparative methods maintain statistical power as the number of trait dimensions increases, and uniquely allow the number of dimensions to far exceed the number of species in the phylogeny. Thus, the evolution of function-valued traits can be appropriately analyzed using distance-based multivariate comparative methods (Adams 2014a).

As with Procrustes-aligned landmark coordinates (Adams 2014a; Adams 2014b), function-valued traits are supplied to multivariate methods in the form of coordinate sequences from generalized time warping alignment (for curves of any form), or in the case of inverse function evaluations (for sigmoidal curves), only $f_{nodes}^{-1}(y)$ is supplied. Estimation of the phylogenetic mean is inherent to distance-based multivariate phylogenetic methods. Importantly, the estimated phylogenetic mean for distance-based methods is equivalent to the root estimate $\boldsymbol{\mu}_{root}$ obtained from Equation (1) when applied to properly aligned landmarks of function-valued data. Specific applications and methodological details of applying distance-based multivariate comparative methods to function-valued traits are discussed in the following sections.

Phylogenetic signal of function-valued traits. Phylogenetic signal is a measure of the extent to which species exhibit phenotypic similarity due to phylogenetic relatedness. Various methods of signal quantification have been proposed, including

Blomberg's *K* (Blomberg et al 2003), Pagel's λ (Pagel 1999), autocorrelation techniques (Pavoine and Ricotta 2012), and correlation-based methods (Zheng et al. 2009). Blomberg's *K* is based on the ratio of observed variation relative to the variation expected under Brownian motion:

$$K = \frac{\left(\mathbf{Y} - E(\mathbf{Y})\right)^t \left(\mathbf{Y} - E(\mathbf{Y})\right)}{\left(\mathbf{Y} - E(\mathbf{Y})\right)^t \mathbf{C}^{-1} \left(\mathbf{Y} - E(\mathbf{Y})\right)} \Big/ \frac{\mathrm{tr}(\mathbf{C}) - N(\mathbf{1}^t \mathbf{C}^{-1} \mathbf{1})^{-1}}{N - 1} \quad (6)$$

where $\mathbf{Y}$ is an $N \times 1$ vector of phenotypic values for a univariate trait and $E(\mathbf{Y})$ is the phylogenetic mean (estimated phenotypic value at the root of the phylogeny), Thus, the expectation of *K* under Brownian motion is 1. Phylogenetic permutation can be used to statistically test if *K* represents significant phylogenetic signal (Blomberg et al. 2003).

Adams (2014a) proposed a multivariate generalization of Blomberg's *K* (K_{mult}) which allows phylogenetic signal to be estimated in high-dimensional traits. K_{mult} utilizes a distance-based approach which produces identical estimates of *K* when applied to univariate traits but also allows for calculations of K_{mult} for multivariate traits. K_{mult} is calculated as follows:

$$K_{\mathrm{mult}} = \frac{\mathbf{D}_{Y,\hat{a}}^t \mathbf{D}_{Y,\hat{a}}}{\mathbf{PD}_{U,0}^t \mathbf{PD}_{U,0}} \Big/ \frac{\mathrm{tr}(\mathbf{C}) - N(\mathbf{1}^t \mathbf{C}^{-1} \mathbf{1})^{-1}}{N - 1} \quad (7)$$

where $\mathbf{D}_{Y,\hat{a}}$ is an $N \times 1$ vector of Euclidian distances between species means and the root of the phylogeny, as expressed by

$$\mathbf{D}_{i,\hat{a}} = \sqrt{\left(\mathbf{Y_i} - E(\mathbf{Y})\right)\left(\mathbf{Y_i} - E(\mathbf{Y})\right)^t} \quad (8)$$

and $\mathbf{PD}_{U,0}$ is an $N \times 1$ vector of Euclidean distances of species means to the origin projected onto the phylogenetic transformation matrix $\mathbf{E}$ from Equation (6). Current software implementations of K_{mult} in the R (R Core Team 2014) library geomorph 2.0

(physignal function) (Adams and Otárola-Castillo 2013; Adams et al. 2014) can readily incorporate the PGLS-based function-valued approaches presented here. For function-valued trait analysis utilizing inverse functions for alignment, $\mathbf{Y}$ is the $N \times n$ matrix of inverse functions f_{tips}^{-1} evaluated at y, and $E(\mathbf{Y})$ calculated by the physignal function is equivalent to the $1 \times n$ vector $\boldsymbol{\mu}_{root}$ obtained from Equation (1). For function-valued landmarks aligned using generalized time-warping, x,y coordinates are analyzed as an $N \times 2n$ matrix of function-valued landmark coordinates. Again, the resulting $E(\mathbf{Y})$ is equivalent to results obtained from $\boldsymbol{\mu}_{root}$ in Equation (1).

Correlated evolution between function-valued traits and categorical (ANOVA) or continuous (PGLS) univariate traits. Distance-based comparative approaches can be implemented to apply phylogenetic ANOVA or assess for correlated evolution between univariate and function-valued traits using a multivariate extension of PGLS called *D*-PGLS (Adams 2014b). In *D*-PGLS, independent variable $\mathbf{X}$ (a column of ones followed by one or more columns of values for independent variables) and dependent variable $\mathbf{Y}$ (an $N \times n$ or $N \times 2n$ matrix, depending on whether inverse functions or generalized time warping alignment is used) are both projected onto $\mathbf{E}$ to remove phylogenetic covariance from trait data, resulting in $\mathbf{X}_{phy}$ and $\mathbf{Y}_{phy}$. $\mathbf{Y}_{phy}$ is regressed on $\mathbf{X}_{phy}$ to obtain predicted values $\widehat{\mathbf{Y}_{\mathrm{X}}}$, and $\mathbf{Y}_{phy}$ is regressed on $\mathbf{1}_{phy}$ to obtain predicted values for $\widehat{\mathbf{Y}_1}$. Next, the trace of the outer-product of predicted values is used to calculate variation explained by the model:

$$SS_X = \mathrm{tr}\left(\left(\widehat{\mathbf{Y}_X} - \widehat{\mathbf{Y}_1}\right)\left(\widehat{\mathbf{Y}_X} - \widehat{\mathbf{Y}_1}\right)^t\right) \quad (9)$$

SS_X and residuals of $(\mathbf{Y}_{phy} \sim \mathbf{X}_{phy})$ are then used to calculate F-ratios and R^2, and phylogenetic permutation is used to determine the significance of the correlation. *D*-PGLS is implemented in the procD.pgls function in the R package geomorph (Adams and Otárola-Castillo 2013; Adams 2014b; Adams et al. 2014).

Additional distance-based multivariate phylogenetic comparative methods. Although not discussed in in detail here, similar methods are available to address questions such as correlated evolution between two separate function-valued traits (or between a function-valued trait and some other multivariate trait) using phylogenetic partial least squares (Adams and Felice 2014), as implemented in the phylo.pls function in geomorph (Adams and Otárola-Castillo 2013; Adams et al. 2014). Additionally, multivariate rates of evolution may be quantified, and evolutionary rates of function-valued traits for subgroups within a phylogeny may be statistically compared (Adams 2014c), as implemented in the compare.evol.rates function in geomorph (Adams and Otárola-Castillo 2013; Adams et al. 2014). For example, one could test the hypothesis that a function-valued plastic response to a particular environmental variable evolves more rapidly in a clade exposed to high levels of the environmental variable than a non-exposed clade. As distance-based multivariate methods continue to be developed, additional phylogenetic comparative methods, such as testing alternative models of evolution (e.g., Ornstein-Uhlenbeck) are expected to be available for further study of function-valued trait evolution (Adams 2014b).

Statistical performance. Function-valued trait evolution was simulated under various conditions to evaluate the performance of function-valued ancestral curve reconstruction and to assess the statistical power of K_{mult} and *D*-PGLS when applied to

function-valued traits. All curve simulations were based on evolution of a generalized linear model using a logit link function:

$$f(x|\theta) = \frac{\exp(\theta_1 + \theta_2)}{1 + \exp(\theta_1 + \theta_2)} \quad (10)$$

Curve evolution was simulated by evolving a root curve defined by two independently simulated function features, the median response level ($-\theta_1/\theta_2$) and the slope at the median response level ($\theta_2/4$), using the fastBM function in the R library phytools (Revell 2012). Internal node values were preserved and "true" regression parameters for each node were subsequently back-calculated. To simulate correlated evolution between curves and univariate traits, sim.corrs was used to generate correlated evolution of a univariate trait with the curve median response level. Root curves parameters were set to $\theta_1 = -10.0$ and $\theta_2 = 2.0$, and to maintain positively-sloped curves in the first quadrant (as with a typical dose-response curve), median response slopes of evolved curves were bounded between 0.2 and 1.0, and median response levels were bounded between 2.0 and 20.0.

For ancestral curve reconstruction, 1000 simulations were performed on randomly generated 128-taxa trees as described above. Ancestral curve reconstruction was performed on inverse tip functions (Equations (1) and (2)) and regression parameters were estimated for each node. For comparison with non-function-valued approaches (in which a single arbitrary *x*-value is chosen for evaluation of $f(x|\theta)$ and the function-valued relationship is ignored), univariate ancestral state reconstruction was conducted individually at 100 evenly spaced tip evaluations of $f(x|\theta)$ where x ranges from 0.0 to 15.0 (spanning the entire sigmoid portion of curves for all species). Coefficients of the known root state and the root state estimated from ancestral curve reconstruction were

evaluated at $f(x|\theta)$ at the same 100 points for which univariate ancestral state reconstruction was conducted. The difference between the true value of $f(x|\theta_{root})$ and values obtained from both ancestral curve reconstruction and individual univariate ancestral state reconstruction were averaged across 1000 simulations and compared. PGPR ancestral curve reconstruction was also conducted for these simulated datasets using the R implementation provided in the supplement of Hadjipantelis (2013).

All simulations related to K_{mult} and *D*-PGLS were carried out on randomly generated pure-birth trees with 128 taxa, and all simulations were conducted 1000 times with randomly generated data. For K_{mult} and *D*-PGLS hypothesis testing, 999 iterations of phylogenetic permutation were conducted. Varying numbers of trait dimensions (p = 2, 5, 10, 25, and 50) were simulated in order to estimate the optimal number of dimensions for representing logit-type functions. Dimensions were input as sequences evaluated at $f^{-1}(y|\theta)$ where y is a sequence of evenly divided numbers from 0.001 to 0.999 of length p. Type I error and power were quantified for K_{mult} using the proportion of significant K_{mult} values for data simulated under λ tree transformations (0, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, and 1.0). For *D*-PGLS Type I error and power analysis, functions and a continuous univariate trait were simulated with covariances of σ =0.0, 0.1, 0.3, 0.5, 0.7, and 0.9 and the proportion of significant correlations was recorded. For comparison, all multivariate simulations were performed alongside analogous univariate approaches (Blomberg's K and PGLS) which were performed at 100 individual evaluations of $f(x|\theta)$ where x ranges from 0.0 to 15.0. R code for all simulations is available in the online supplementary material.

RESULTS AND DISCUSSION

Performance of ancestral curve reconstruction. Root estimations of ancestral curve reconstruction were compared to simulated logit curve evolution using the inverse function approach described in previous sections. Root coefficients were estimated from reconstructed inverse functions (Equation (1)), and both the known and estimated functions were evaluated along 100-length evenly spaced *x*-values ranging from 0.0 to 15.0. $f(x|\theta_{estimated})$ and $f(x|\theta_{true})$ were compared at each point. The average predicted curve for each method is plotted in figure 2.3a. The same approach was applied to reconstructed curves obtained from PGPR. For comparison with univariate methods in which the function-valued nature of the trait is ignored, ancestral state reconstruction of each observed tip *y*-value at each *x*-value is also plotted. 95% confidence intervals were calculated for PGLS (Fig. 2.3b), PGPR (Fig. 2.3c), and univariate (Fig. 2.3d) ancestral reconstruction methods using a single representative simulation (Rohlf 2001; Hadjipantelis et al. 2013). On average, ancestral curve reconstruction using PGLS-based and PGPR-based methods yielded similar results (Fig. 2.3a). The small discrepancies that are present are likely attributable to differences such as the underlying assumptions of the Brownian motion simulations performed here and the assumption of an Ornstein-Uhlenbeck model of evolution inherent to PGPR. Although univariate ancestral state reconstruction visually appears to give similar results to both PGLS and PGPR ancestral curve reconstruction, in practice, the entire curve would not be reconstructed with this approach; rather, univariate ancestral state reconstruction is typically performed in isolation at a single *x*-level, and no knowledge of surrounding parts of the curve would be represented. Therefore, only instantaneous deviation from the true *y*-value at an

individual *x*-value should be considered for univariate ancestral state reconstruction (i.e., not the sum of the resulting curve as a whole).

Even for *x*-values where the estimated univariate response is identical to the true response, a function-valued trait cannot be properly represented by a single value. Such attempts are likely to produce misleading inferences and potentially contradictory conclusions based on the arbitrary *x*-value choice at which the trait is evaluated. These issues are further highlighted in simulations of phylogenetic signal and correlated trait evolution.

In contrast to univariate ancestral state reconstruction, PGPR ancestral curve reconstruction *should* be taken in context of the entire reconstructed curve. Accordingly, PGPR ancestral curve reconstruction provided similar results to PGLS-based curve reconstruction overall (Fig. 2.3a). Therefore, both methods provide vastly superior representations of ancestral function-valued traits than univariate attempts at ancestral state reconstruction.

Statistical power of K_{mult} and D-PGLS. Type I error and statistical power of K_{mult} and *D*-PGLS for function-valued traits were assessed using the methods described above. Type 1 error rates for all simulations of K_{mult} and *D*-PGLS where $p \geq 5$ (as determined by data simulations on star trees and input covariance of zero, respectively) were approximately 0.05. Similarly, Type I error rates were approximately 0.05 for univariate Blomberg's *K* and PGLS evaluated at arbitrary *x*-levels. As expected, statistical power increased for K_{mult} and *D*-PGLS as the number of dimensions increased (Adams 2014a; Adams 2014b) (Fig. 2.4). Dimensionality beyond $p = 25$ resulted in diminishing gains in statistical power, coupled with a substantially increased

computational burden. K_{mult} power scaled approximately linearly with λ transformations up to 1.0, whereas *D*-PGLS power reached 1.0 with input covariation of 0.5 and higher (Fig. 2.5). For univariate methods, statistical power varied widely depending on the *x*-level at which the trait was assessed, although statistical power approached that of K_{mult} and *D*-PGLS when *x*-level was near the median response level of the root function and phylogenetic signal or input covariation was high. It should be noted that these results do *not* suggest that univariate approximations of phylogenetic signal or correlated trait evolution can be applied to function-valued traits. Indeed, phylogenetic comparative experiments evaluating species traits at a single level are blinded to true function-valued relationships, so the potential for erroneously drawn conclusions is likely to be overlooked, and such attempts provide no means for validation of results.

Overall, Type I error rates and statistical power of K_{mult} and *D*-PGLS suggest that distance-based multivariate phylogenetic comparative methods can be appropriately applied to the study of function-valued trait evolution, and further highlight that univariate approaches to function-valued traits are inappropriate.

Limitations. Function-valued phylogenetic comparative methods offer promising improvements over conventional methods for many types of analyses. However, several assumptions and limitations must be accounted for in order to properly take advantage of these methods. First, a substantial amount of data is required for every species at multiple levels to accurately infer function parameters (Stinchcombe et al. 2012). Second, even with large amounts of data, function parameters are estimated with error, but the methods described here assume species curves are estimated without error. Third, a thorough understanding of function-valued relationships is necessary to appropriately align

function-valued landmarks and to express function-valued trait evolution appropriately. Finally, no methods for testing alternative models (e.g., Ornstein-Uhlenbeck) of high-dimensional multivariate trait evolution are currently available. Further development of distance-based multivariate phylogenetic comparative methods offers promising expansions of the possibilities of approaching function-valued trait evolution.

CONCLUSIONS

Phylogenetic comparative methods have transformed the field of comparative biology. While extensive work has contributed to the integration of evolution with inherently function-valued applications in developmental biology, toxicology, gene expression, and trait plasticity (particularly the prediction of species short-term evolutionary trajectories in responses to selective gradients), few methods exist to unite function-valued approaches with phylogenetic comparative methods. Building on the methods presented here, future work should seek to address issues such as testing alternative evolutionary models, developing methods for dealing with sparse data, incorporating parameter estimation error into models, accounting for phylogenetic uncertainty, and applying the methods described here in alternative statistical frameworks, such as Bayesian approaches.

SUPPLEMENTARY MATERIAL – Data available from the Dryad Digital Repository. doi:10.5061/dryad.5nd50.

REFERENCES

Adams D.C., Otárola-Castillo E. 2013. geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution* 4:393–399.

Adams D.C. 2014a. A Generalized K Statistic for Estimating Phylogenetic Signal from Shape and Other High-Dimensional Multivariate Data. *Systematic Biology* 63:685-697.

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

Adams D.C. 2014c. Quantifying and Comparing Phylogenetic Evolutionary Rates for Shape and Other High-Dimensional Phenotypic Data. *Systematic Biology* 63:166-177.

Adams D.C., Felice R.N. 2014. Assessing Trait Covariation and Morphological Integration on Phylogenies Using Evolutionary Covariance Matrices. *PLOS ONE* 9.

Adams D.C., Otárola-Castillo E., Sherratt E. 2014. geomorph: Software for geometric morphometric analyses. R package version 2.0.

Aston J.A.D, Buck D., Coleman J., Cotter C.J., Jones N.S., Macaulay V., MacLeod N., Moriarty J.M., Nevins A., Functional Phylogenies Group. 2012. Phylogenetic inference for function-valued traits: speech sound evolution. *Trends in Ecology and Evolution* 27:160-166.

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

Bradshaw W.E., Holzapfel C.M. 2006. Climate change – Evolutionary response to rapid climate change. *Science* 312:1477-1478.

Callahan H.S., Maughan H., Steiner U.K.. 2008. *Phenotypic plasticity, costs of phenotypes, and costs of plasticity toward an integrative view*. In: Schlichting C.D., Mousseau T.A. editors. Ann. N. Y. Acad. Sci. p. 44-66.

Cressie N.A.C. 1993. *Statistics for Spatial Data*. New rev. ed. New York: Wiley-Interscience.

Donovan L.A., Mason C.M., Bowsher A.W., Goolsby E.W., Ishibashi C.D.A. 2014. Ecological and evolutionary lability of plant traits affecting carbon and nutrient cycling. *Journal of Ecology* 102:302-314.

Felsenstein J. 1985. Phylogenies and the comparative method. American Naturalist 125:1-15.

Garland T., Ives A.R. 2000. Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. *American Naturalist* 155:346-364.

Garland T., Midford P.E., Ives A.R. 1999. An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral values. *American Zoologist* 39:374-388.

Giorgino T. 2009. Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. *Journal of Statistical Software* 31:1-24.

Grafen A. 1989. The phylogenetic regression. *Proceedings of the Royal Society B* Biol. Sci. 326:119-157.

Griswold C.K., Gomulkiewicz R., Heckman N. 2008. Hypothesis testing in comparative and experimental studies of function-valued traits. *Evolution* 62:1229-1242.

Guenard G., Legendre P., Peres-Neto P. 2013. Phylogenetic eigenvector maps: a framework to model and predict species traits. *Methods in Ecology and Evolution* 4:1120-1131.

Guenard G., von der Ohe P.C., de Zwart D., Legendre P., Lek S. 2011. Using phylogenetic information to predict species tolerances to toxic chemicals. *Ecological Applications* 21:3178-3190.

Hadjipantelis P.Z., Jones N.S., Moriarty J., Springate D.A., Knight C.G. 2013. Function-valued traits in evolution. *Journal of the Royal Society Interface* 10.

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

Izem R., Kingsolver J.G. 2005. Variation in continuous reaction norms: Quantifying directions of biological interest. *American Naturalist* 166:277-289.

Jacobs B.S., Latimer A.M. 2012. Analyzing reaction norm variation in the field vs. greenhouse: Comparing studies of plasticity and its adaptive value in two species of Erodium. *Perspecitves in Plant Ecology* 14:325-334.

Jones N.S., Moriarty J. 2013. Evolutionary inference for function-valued traits: Gaussian process regression on phylogenies. *Journal of the Royal Society Interface* 10.

Kingsolver J.G., Gomulkiewicz R., Carter P.A. 2001. Variation, selection and evolution of function-valued traits. *Genetica* 112:87-104.

Knight C.A., Ackerly D.D. 2003. Evolution and plasticity of photosynthetic thermal tolerance, specific leaf area and leaf size: congeneric species from desert and coastal environments. *New Phytologist* 160:337-347.

Larras F., Keck F., Montuelle B., Rimet F., Bouchez A. 2014. Linking diatom sensitivity to herbicides to phylogeny: a step forward for biomonitoring? *Environmental Science and Technology* 48:1921-1930.

Martins E.P., Hansen T.F. 1997. Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist* 149:646-667.

Mason C.M., McGaughey S.E., Donovan L.A. 2013. Ontogeny strongly and differentially alters leaf economic and other key traits in three diverse Helianthus species. *Journal of Experimental Botany* 64:4089-4099.

Murren C.J., Maclean H.J., Diamond S.E., Steiner U.K., Heskel M.A., Handelsman C.A., Ghalambor C.K., Auld J.R., Callahan H.S., Pfennig D.W., Relyea R.A., Schlichting C.D., Kingsolver J. 2014. Evolutionary change in continuous reaction norms. *American Naturalist* 183:453-467.

Myers C., Rabiner L.R., Rosenberg A.E. 1980. Performancy tradeoffs in dynamic time warping algorithms for isolated word recognition. *IEEE Transactions on Acoustic Speech and Signal Processing*. 28:623-635.

Pagel M.D. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877-844.

Pavoine S., Ollier S., Pontier D., Chessel D. 2008. Testing for phylogenetic signal in phenotypic traits: new matrices of phylogenetic proximities. *Theoretical Population Biology* 73:79-91.

Pigliucci M., Pollard H., Cruzan M.B. 2003. Comparative studies of evolutionary responses to light environments in Arabidopsis. *American Naturalist* 161:68-82.

Pollard H., Cruzan M., Pigliucci M. 2001. Comparative studies of reaction norms in Arabidopsis. I. Evolution of response to daylength. *Evolutionary Ecology Research* 3:129-155.

R Core Team 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Revell L.J. 2010. Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution* 1:319-329.

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

Richards C.L., Bossdorf O., Muth N.Z., Gurevitch J., Pigliucci M. 2006. Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. *Ecology Letters* 9:981-993.

Rocha F.B., Klaczko L.B. 2012. Connecting the dots of nonlinear reaction norms unravels the threads of genotype-environment interaction in Drosophila. *Evolution* 66:3404-3416.

Rohlf F.J., Slice D.E. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology* 39:40-59.

Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution* 55:2143–2160.

Stinchcombe J.R., Kirkpatrick M., Function-valued Traits Working Group. 2012. Genetics and evolution of function-valued traits: understanding environmentally responsive phenotypes. *Trends in Ecology and Evolution* 27:637-647.

Strachan L.A., Tarnowski-Garner H.E., Marshall K.E., Sinclair B.J. 2011. The Evolution of cold tolerance in drosophila larvae. *Physiological and Biochemical Zoology* 84:43-53.

Whitman D.W., Agrawal A.A. 2009. What is phenotypic plasticity and why is it important? In: Whitman D.W., Ananthakrishnan T.N., editors. *Phenoytpic plasticity of insects*. Enfield (NH): Science Publishers. p. 1-63.

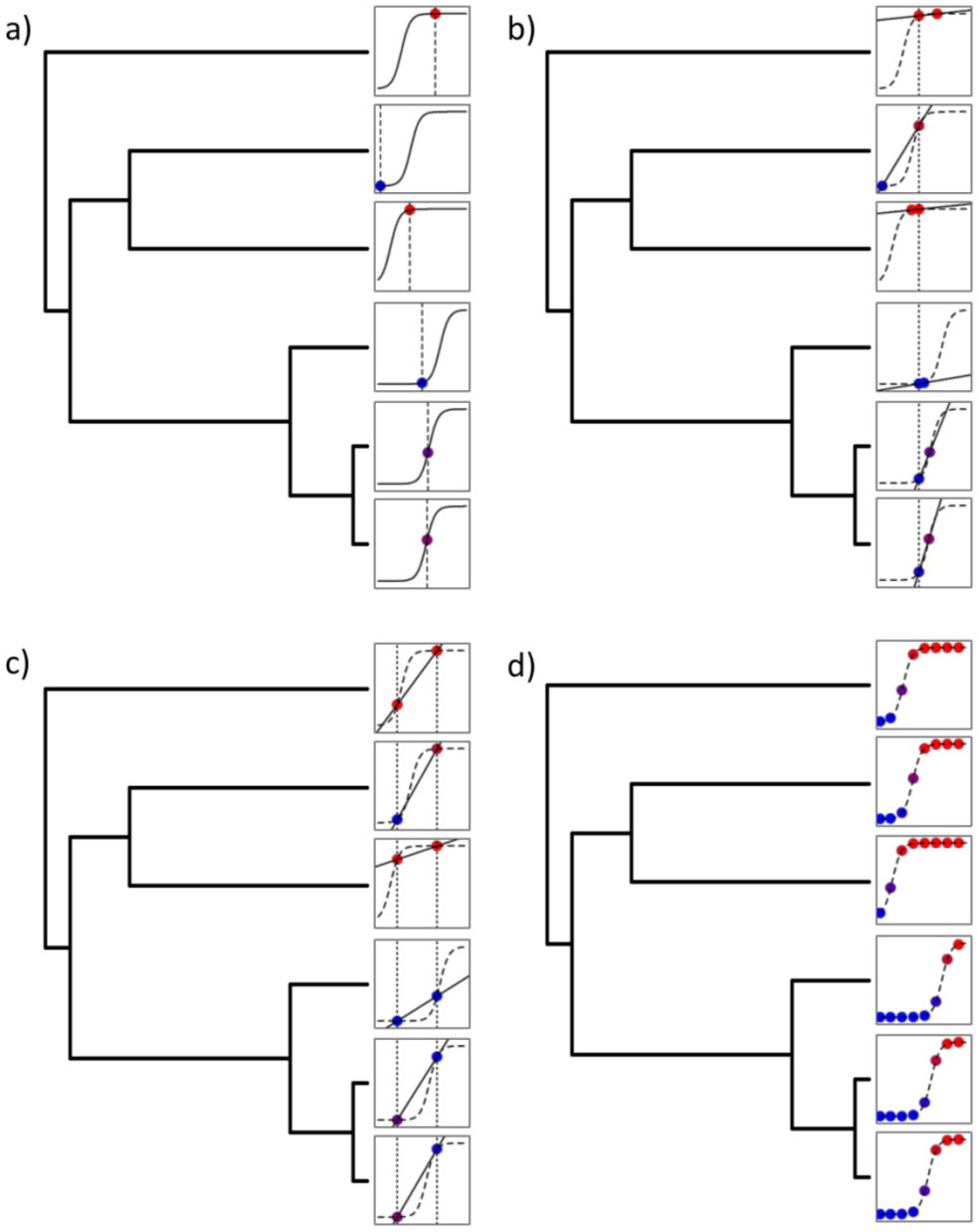
Wund M.A. 2012. Assessing the Impacts of Phenotypic Plasticity on Evolution. *Integrative and Comparative Biology* 52:5-15.

Zheng L., Ives A.R., Garland T., Larget B.R., Yu Y., Cao K. 2009. New multivariate tests for phylogenetic signal and trait correlations applied to ecophsyiological phenotypes of nine *Manglietia* species. *Functional Ecology* 23:1059-1069.

Zhou, F., De la Torre, F. 2012. Generalized Time Warping for Multi-modal Alignment of Human Motion. 2012 *IEEE Conference on Computer Vision and Pattern Recognition (Cvpr)*: 1282-1289.

FIGURES

Figure 2.1 Hypothetical result of performing phylogenetic comparative analysis on plastic traits from species traits a) measured in nature; b) measured in the field and in a controlled common environment; c) measured in a controlled common environment at two different treatment levels; and d) measured in a controlled common environment at multiple treatment levels. The *x*-axis represents an environmental gradient; the *y*-axis represents the trait value. In 1a, the vertical dotted line represents the value of the environmental variable for each species collected in nature; in 1b, the vertical dotted line represents the value of the environmental variable for the controlled common environment, and the second point represents the trait value measured in nature; and in 1c, the vertical lines represent the two treatment levels of the environmental variable in the controlled common environment. Solid lines in 1b and 1c represent inferred reaction norms, several of which deviate considerably from true reaction norms. Depending on specific assumptions regarding the method used, contrasting conclusions regarding trait phylogenetic signal, model of evolution, and evolutionary history may be drawn.


a)
b)
c)
d)

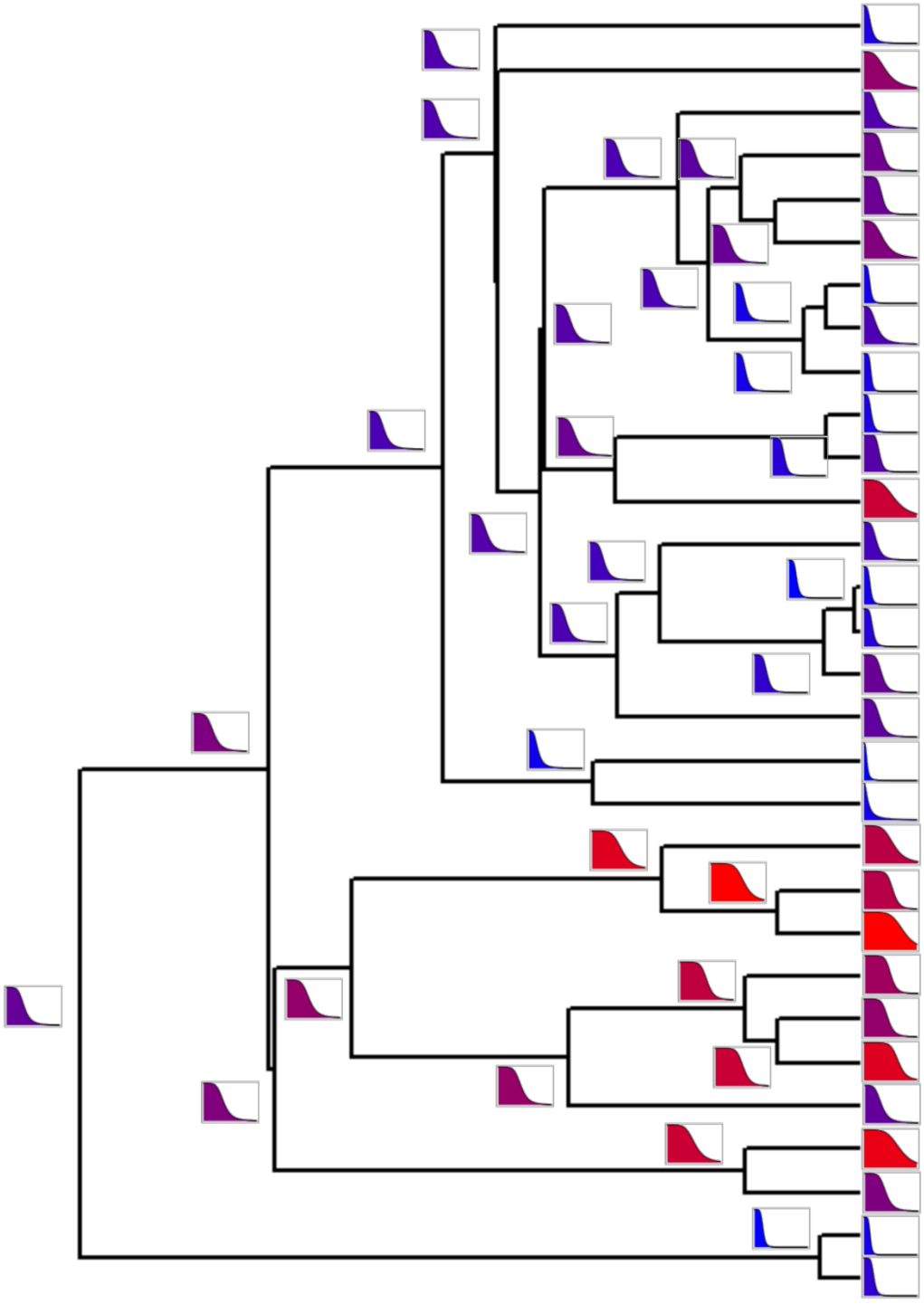
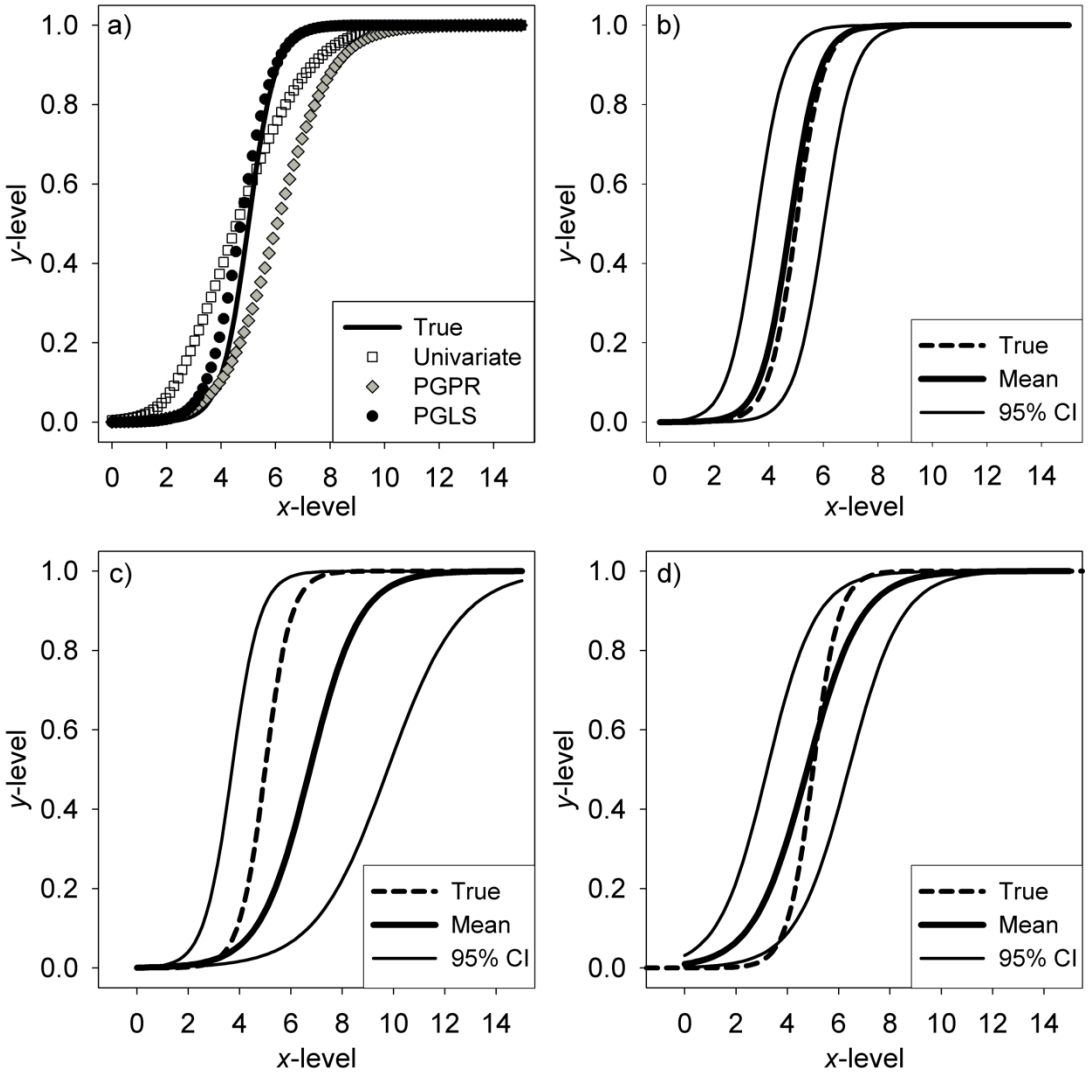


Figure 2.2 Ancestral curve reconstruction of a simulated dataset for a proportion-based trait response (*y*-axis) in response to an environmental gradient (*x*-axis).

Figure 2.3 Average result of ancestral root curve reconstruction (a) from 1000 simulated datasets (N=128) for PGLS-based ancestral curve reconstruction (circles), PGPR (diamonds), and univariate ancestral state reconstruction (squares). 95% confidence intervals of root reconstructions are presented from a single representative simulation for b) PGLS, c) PGPR, and d) univariate curve reconstruction. All three methods produced resulting curves similar to the actual root curve (solid line). However, univariate methods are only evaluated at a single point (not in the context of an entire curve), so only deviation from the true curve at individual x-levels should be considered for the univariate approach. Univariate representations of function-valued traits are inappropriate regardless of the instantaneous deviation from the actual curve, and such approaches are likely to result in incorrect interpretations. Therefore, despite greater absolute deviation from the root at various x-levels than univariate approaches, the overall performance of PGPR was comparable to that of PGLS, with both methods providing root curve estimates close to actual root curve.


a)
b)
c)
d)
y-level
x-level
True
Univariate
PGPR
PGLS
Mean
95% CI

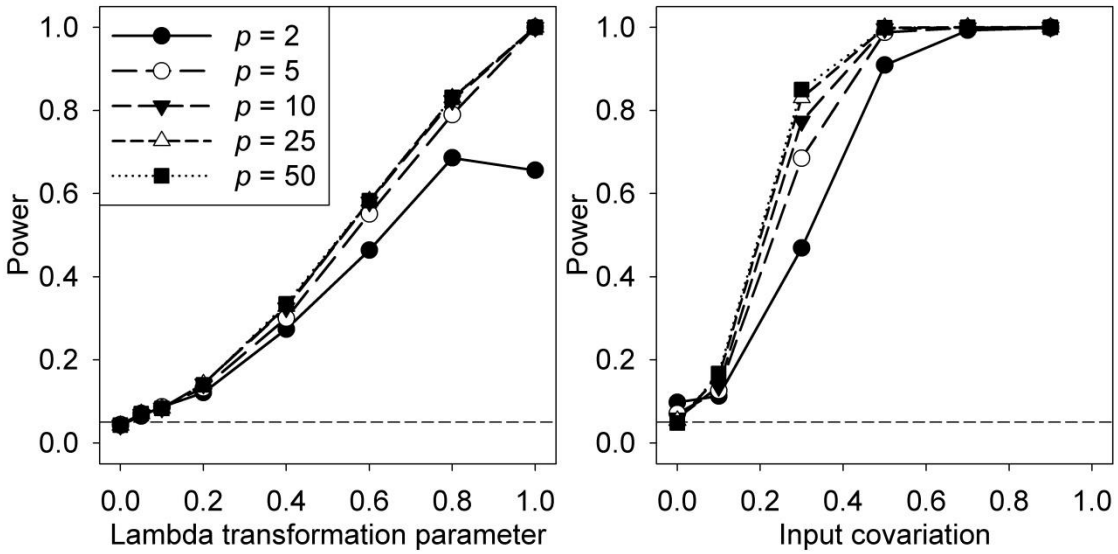


Figure 2.4 Results from K_{mult} (left) and *D*-PGLS (right) type I error and power from 1000 data simulations (*N*=128) with increasing trait dimensionality (*p*=2, 5, 10, 25, 50).

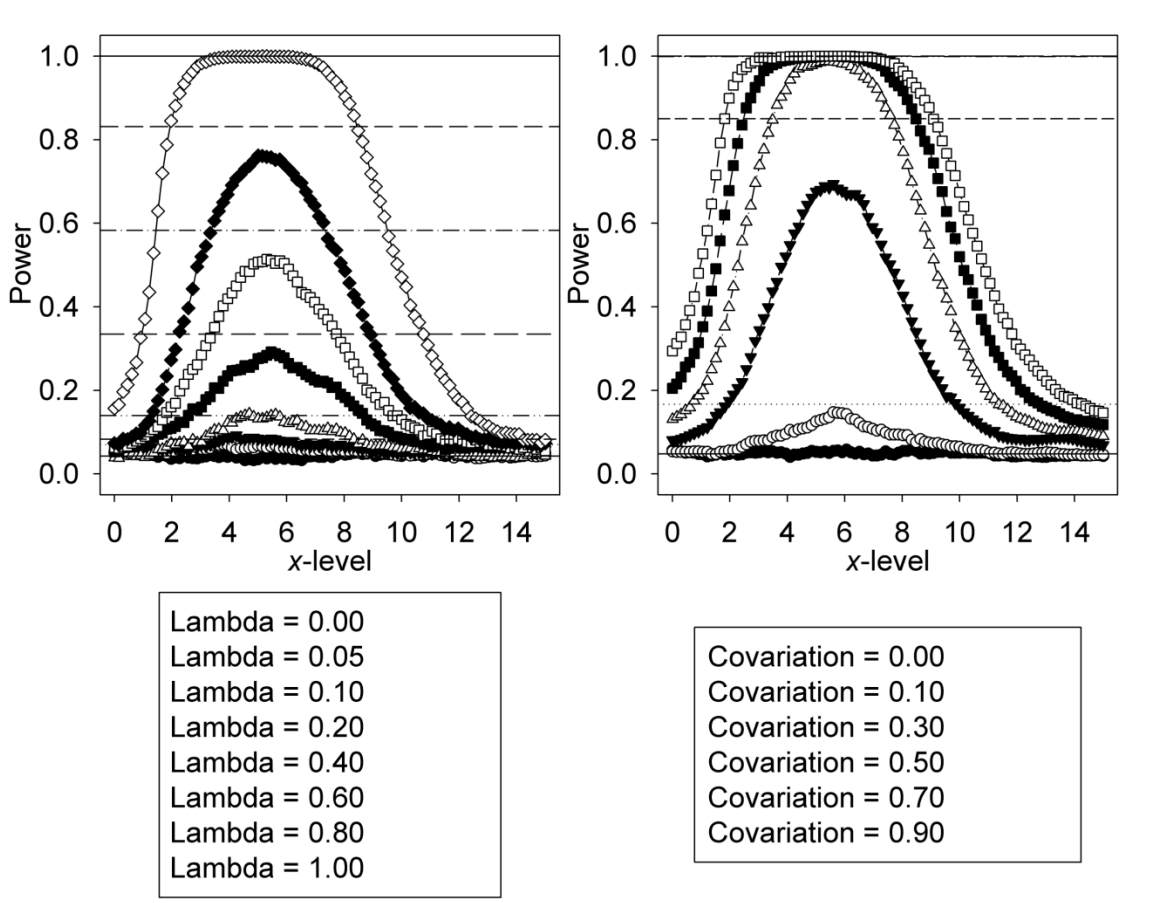


Figure 2.5 Statistical power of univariate Blomberg's K (left) and PGLS (right) at individual *x*-levels ranging from 0.00 to 15.00 (N=128). K_{mult} and D-PGLS power results (p=50) are plotted as horizontal lines for comparison. Statistical power for univariate approaches to function-valued traits varies widely for these tests.

CHAPTER 3

LIKELIHOOD-BASED PARAMETER ESTIMATION FOR HIGH-DIMENSIONAL PHYLOGENETIC COMPARATIVE MODELS: OVERCOMING THE LIMITATIONS OF 'DISTANCE-BASED' METHODS [1]

[1]**Goolsby EW**. Submitted to *Systematic Biology* on November 2, 2015.

ABSTRACT

Recently proposed distance-based (Q-mode) multivariate phylogenetic comparative methods can be expressed in a covariance-based (R-mode) framework. It is shown that a properly specified covariance-based approach performs identically to distance-based approaches and can be performed using fast linear-time algorithms. Additionally, a composite likelihood approach is introduced for maximum pseudolikelihood parameter estimation, opening up the ability to estimate alternative evolutionary models, allow missing data, and incorporate within-species variation. Simulations reveal low statistical power and high Type I error for distance-based methods under various scenarios, whereas composite likelihood approaches demonstrate appropriate Type I error and high statistical power while substantially expanding model flexibility. These methods are implemented in the R package *phylocurve*.

INTRODUCTION

Phylogenetic comparative methods provide a framework for testing hypotheses in comparative biology while accounting for statistical non-independence of hierarchically related species. In recent years, multivariate traits in the of context comparative data have been of increasing interest, leading to the development of multivariate extensions of phylogenetic comparative methods (Revell 2009; Bartoszek et al. 2012; Adams 2013). Of particular interest are high-dimensional traits, such as with morphometric data (Adams 2014a-c; Adams and Felice 2014; Adams and Collyer 2015; Denton and Adams 2015) and function-valued traits (Goolsby 2015), which pose computational and statistical challenges as the number of trait dimensions increases. Namely, as the number of parameters to be estimated increases for a given sample size, statistical power decreases substantially (Adams 2014b). Additionally, when the number of trait dimensions equals or exceeds the number of species in a study, maximum likelihood trait covariance matrices are non-invertible and thus cannot be used for calculations central to most phylogenetic comparative methods (Adams 2014b). Finally, even assuming computational feasibility and adequate statistical power, it may be difficult to draw unified conclusions from multiple (and potentially conflicting) individual dimension-specific metrics.

Accordingly, it is potentially beneficial to approach comparative analyses of high-dimensional traits with single generalized multivariate metrics rather than many dimension-specific metrics. Adams (2014a-c) proposed a suite of multivariate phylogenetic comparative methods for studying high-dimensional traits while maintaining statistical power and providing generalized test statistics for multivariate

traits as a whole. The methods, which include multivariate extensions of Blomberg's *K* (Adams 2014a), phylogenetic generalized least squares (Adams 2014b; Adams and Collyer 2015), comparisons of evolutionary rates (Adams 2014c; Denton and Adams 2015), and phylogenetic partial least squares (Adams and Felice 2014), avoid the problem of dealing with non-invertible covariance matrices and can even handle traits in which the number of dimensions far exceeds the number of species in the study. This is accomplished by phylogenetic transformation of the data and subsequent distance-based (Q-mode), rather than conventional covariance-based (R-mode), analyses. When applied to a single univariate trait, distance-based methods and conventional phylogenetic comparative methods provide identical results. When applied to higher-dimensional traits, such as landmark coordinates of morphometric shape data (e.g., leaf shape coordinates obtained from Procrustes analysis (Chitwood et al. 2014)) or function-valued traits (e.g., species reaction norms of a phenotypically plastic trait (Goolsby 2015)), distance-based comparative methods fit a single consensus metric that attempts to capture the variation of the entire high-dimensional trait as a whole.

Limitations of distance-based comparative methods. Despite potential advantages, the distance-based comparative framework involves considerable shortcomings. Specifically, the inability to calculate log-likelihoods restricts distance-based models to Brownian motion. Additionally, model selection is limited to relatively simple hypothesis tests, lacking any clear way to compare complex combinations of model specifications (e.g., simultaneous modeling of fixed effects, phenotypic integration, rate heterogeneity, etc.). Certain distance-based methods also suffer from inappropriately high Type I error and low statistical power under a variety of

evolutionary scenarios. Distance-based methods also require eigendecomposition and inversion of the phylogenetic covariance matrix, which is extremely time-consuming and inefficient for large phylogenies.

Although the methods described above are explicitly distance-based, an equivalent covariance-based (R-mode) formulation can be expressed. Like the Q-mode, the corresponding R-mode approach retains identical Type I error and statistical power as trait dimensionality increases when correctly specified and avoids inversion of singular matrices. Fast linear-time computational methods for performing relevant calculations can also be used to perform covariance-based calculations (Felsenstein 1973; Freckleton 2012; Ho and Ané 2014), thus reducing the computational challenges posed by extremely large phylogenies and high-dimensional data. Finally, a composite likelihood approach for parameter estimation and model selection is introduced, providing a flexible and statistically powerful framework for expanding high-dimensional comparative methods.

A Covariance-Based (R-Mode) Re-Expression of Multivariate Distance-Based (Q-Mode) Comparative Methods

To address issues of statistical power and non-invertible matrices, and to provide a framework for estimating generalized statistics for high-dimensional multivariate traits, Adams (2014 a-c) proposed several phylogenetic comparative methods based on a distance-based (Q-mode) approach. Consider a phylogeny with N extant species on which M traits are observed, given as an $N \times M$ data matrix ($\mathbf{Y}$). The $N \times N$ phylogenetic covariance matrix ($\mathbf{C}$) is parameterized by branch lengths and the specified evolutionary

model (e.g., Brownian motion). The eigenvectors (**U**) and diagonal matrix of eigenvalue square roots ($\mathbf{V}^{1/2}$) of **C** are used to construct the phylogenetic transformation matrix (**T**)

$$\mathbf{T} = \left(\mathbf{U}\mathbf{V}^{1/2}\mathbf{U}^{t}\right)^{-1} \quad (1)$$

which is then matrix multiplied by relevant matrices of interest (e.g., **TX**, **TY**, $\mathbf{T}\left(\mathbf{Y} - E(\mathbf{Y})\right)$) to remove phylogenetic covariance from the data (Garland and Ives 2000). Next, the Euclidean distances of phylogenetically transformed data from the origin are calculated (**PD**), resulting in an N-length vector which is then used for multivariate comparative calculations.

Evolutionary rates: distance-based (Q-mode) methods. Various methods have been proposed to quantify and compare evolutionary rates for univariate traits, including contrast-based (Garland 1992), generalized least squares-based (Martins and Hansen 1997), and likelihood-based (O'Meara et. al 2006; Thomas et al. 2006), as well as Bayesian methods (Rabosky et al. 2014). For high-dimensional data, Adams (2014c) proposed an estimate called σ^2_{mult}, which is a single consensus evolutionary rate for the entire multivariate trait. σ^2_{mult} is estimated by dividing the Euclidean distance of the cross-product of phylogenetically transformed residuals from the origin is by the total number of observations (NM):

$$\sigma^2_{\text{mult}} = \frac{\mathbf{PD}^{t}_{\mathbf{Y}-E(\mathbf{Y}),0}\mathbf{PD}_{\mathbf{Y}-E(\mathbf{Y}),0}}{NM} \quad (2)$$

The estimate of σ^2_{mult} can be used to statistically compare evolutionary rates among groups of species (Adams 2014c) or among groups of traits (Denton and Adams 2015). To test whether σ^2_{mult} differs among species groups (as implemented in the *geomorph* function *compare.evol.rates*), subset groups of size $N_{\text{species.sub}_j}$ are used to

estimate $\sigma^2_{\text{species.sub}}$ for each species group by replacing $\mathbf{PD}_{\mathbf{Y}-E(\mathbf{Y}),0}$ with $\left(\mathbf{PD}_{\mathbf{Y}-E(\mathbf{Y}),0}\right)_{\text{species.sub}_j}$ and N with $N_{\text{species.sub}_j}$ in equation (2). The observed ratio of regime-specific $\sigma^2_{\text{species.sub}}$ values is compared to the null distribution of $\sigma^2_{\text{species.sub}}$ ratios via phylogenetic simulation (Adams 2014c). Similarly, to test for differences in σ^2_{mult} among groups of traits (as implemented in the *geomorph* function *compare.multi.evol.rates*), $\sigma^2_{\text{trait.sub}}$ is calculated for each trait group by replacing M with $M_{\text{traits.sub}_k}$ and $\mathbf{PD}_{\mathbf{Y}-E(\mathbf{Y}),0}$ with $\mathbf{PD}_{\mathbf{Y}_k-E(\mathbf{Y}_k),0}$ in equation (2) (where $\mathbf{Y}_k$ contains $M_{\text{trait.sub}_k}$ variables), and the observed ratio of group-specific $\sigma^2_{\text{trait.sub}}$ values is compared to the null distribution of $\sigma^2_{\text{trait.sub}}$ via phylogenetic simulation (Denton and Adams 2015).

Evolutionary rates: covariance-based (R-mode) methods. The calculation of σ^2_{mult} can be simplified considerably using a covariance-based approach, as σ^2_{mult} is simply the arithmetic mean of maximum likelihood evolutionary rates for each trait considered individually:

$$\sigma^2_{\text{mult}} = \frac{\sum_{i=1}^{M}\left(\left(\mathbf{y}_i - E(\mathbf{y}_i)\right)^t \mathbf{C}^{-1}\left(\mathbf{y}_i - E(\mathbf{y}_i)\right)\right)}{NM} \quad (3)$$

Similarly, $\sigma^2_{\text{trait.sub}}$ for a trait group subset k is simply the arithmetic mean of evolutionary rates for individual traits in $\mathbf{Y}_k$.

The calculation of regime-specific $\sigma^2_{\text{species.sub}}$ can be calculated as the mean maximum likelihood variance of the phylogenetically transformed residuals $\left(\mathbf{T}\left(\mathbf{Y} - E(\mathbf{Y})\right)_{\text{species.sub}}\right)$. Alternatively, the computation of $\mathbf{T}$ can be avoided using either the 'noncensored' or 'censored' approach described by O'Meara et. al (2006), the latter of

which provides an efficient closed-form estimate of $\sigma^2_{\text{species.sub}}$, in which equation (3) is applied to a phylogeny and dataset pruned to only contain the species represented in regime *j*. The censored estimate is not identical to the distance-based estimate of $\sigma^2_{\text{species.sub}}$ (the censored approach estimates a separate $E(\mathbf{Y})$ for each pruned tree). However, the censored approach is a close approximation, retains appropriate Type I error and statistical power, and is far more efficient than the eigendecomposition of $\mathbf{C}$ and subsequent inversion of an $N \times N$ matrix (which is required in equation (1) in order to construct the phylogenetic transformation matrix **T**). If phylogenetic transformation cannot be avoided, an alternative to equation (1) which avoids matrix inversion can be used to construct the phylogenetic transformation matrix:

$$\mathbf{T} = \mathbf{U}\mathbf{V}^{*}\mathbf{U}^{t} \quad (4)$$

where $\mathbf{V}^{*}$ is an $N \times N$ diagonal matrix with $\sqrt{1/\mathbf{v}}$ along the diagonal, where $\mathbf{v}$ contains the eigenvalues of $\mathbf{C}$ (Li 2007). If $\mathbf{C}$ is singular, infinite or undefined values in $\mathbf{V}^{*}$ may be replaced with zero (although the consequences of proceeding with comparative analyses on singular matrices are largely untested). Functions for comparing evolutionary rates using the described covariance-based approach are implemented in the *phylocurve* functions *fast.geomorph.compare.evol.rates* and *fast.geomorph.compare.multi.evol.rates*. These functions (and all other *phylocurve* functions that begin with '*fast.geomorph.*') are implemented to demonstrate the equivalence and application of fast linear-time computations for analogous distance-based functions in *geomorph* (see below for a discussion of fast covariance-based approaches).

Phylogenetic signal (Blomberg's K). Blomberg's *K* (Blomberg 2003), which in univariate form is calculated as

$$K = \frac{\left(\mathbf{y} - E(\mathbf{y})\right)^t \left(\mathbf{y} - E(\mathbf{y})\right)}{\left(\mathbf{y} - E(\mathbf{y})\right)^t \mathbf{C}^{-1} \left(\mathbf{y} - E(\mathbf{y})\right)} \Big/ \frac{\mathrm{tr}(\mathbf{C}) - N(\mathbf{1}^t \mathbf{C}^{-1} \mathbf{1})^{-1}}{N - 1} \quad (5)$$

can be extended to multivariate form K_{mult} (Adams 2014a, as implemented in the *geomorph* function *physignal*) using the distance-based formula

$$K_{\mathrm{mult}} = \frac{\mathbf{D}^t_{\mathbf{Y}-E(\mathbf{Y}),0} \mathbf{D}_{\mathbf{Y}-E(\mathbf{Y}),0}}{\mathbf{P}\mathbf{D}^t_{\mathbf{Y}-E(\mathbf{Y}),0} \mathbf{P}\mathbf{D}_{\mathbf{Y}-E(\mathbf{Y}),0}} \Big/ \frac{\mathrm{tr}(\mathbf{C}) - N(\mathbf{1}^t \mathbf{C}^{-1} \mathbf{1})^{-1}}{N - 1} \quad (6)$$

where $\mathbf{D}_{\boldsymbol{X},0}$ is the non-phylogenetically transformed Euclidean distance between some matrix and the origin. As with σ^2_{mult}, the value K_{mult} can be calculated using a covariance-based approach (implemented in the *phylocurve* function *fast.geomorph.physignal*) by considering the sums of squared residuals for each individual trait:

$$K_{\mathrm{mult}} = \frac{\sum_{i=1}^{M} \left(\left(\mathbf{y}_i - E(\mathbf{y}_i)\right)^t \left(\mathbf{y}_i - E(\mathbf{y}_i)\right) \right)}{\sum_{i=1}^{M} \left(\left(\mathbf{y}_i - E(\mathbf{y}_i)\right)^t \mathbf{C}^{-1} \left(\mathbf{y}_i - E(\mathbf{y}_i)\right) \right)} \Big/ \frac{\mathrm{tr}(\mathbf{C}) - N(\mathbf{1}^t \mathbf{C}^{-1} \mathbf{1})^{-1}}{N - 1} \quad (7)$$

As with univariate Blomberg's *K*, significance of K_{mult} is determined by phylogenetic permutation (Blomberg 2003; Adams 2014a).

Phylogenetic generalized least squares. Distance-based phylogenetic generalized least squares (*D*-PGLS, as implemented in the *geomorph* function *procD.pgls*) regression can be performed by regressing $\mathbf{EY}{\sim}\mathbf{EX}$ and $\mathbf{EY}{\sim}\mathbf{E1}$ to obtain predicted values $\widehat{\mathbf{Y}}_{\mathbf{X}}$ and $\widehat{\mathbf{Y}}_{\mathbf{1}}$, which is then used to calculate summary statistics including sums of squares, F-ratios, and R^2, and significance is determined by phylogenetic permutation (Adams 2014b). To perform the covariance-based equivalent of *D*-PGLS regression (*phylocurve* function *fast.geomorph.procD.pgls*), first, a $p \times M$ (where p is the

number of regression coefficients to be estimated for each trait dimension) matrix of dimension-specific regression coefficients is calculated as

$$\boldsymbol{\beta} = (\mathbf{X}^t\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{C}^{-1}\mathbf{Y} \quad (8)$$

where $\mathbf{X}$ is the $M \times p$ model matrix for PGLS regression, typically consisting of a column of ones and one or more columns of univariate predictor variables (Martins and Hansen 1997; Revell 2010). Note that $\mathbf{Y}$ is an $N \times M$ matrix in equation (8), unlike univariate PGLS in which $\mathbf{Y}$ is an $N \times 1$ vector. Next, $N \times M$ matrices of predicted species values from the regression $\widehat{\mathbf{Y}}_{\mathbf{X}} = \mathbf{X}\boldsymbol{\beta}$ and null model $\widehat{\mathbf{Y}}_{\mathbf{1}} = \mathbf{1}\, E(\mathbf{Y})$ are used to calculate residuals, and sums of squares are obtained from $SS_{\hat{\varepsilon}_{\mathbf{X}}} = \sum_{i=1}^{M}\left(\left(\mathbf{y}_i - \widehat{\mathbf{y}}_{i_{\mathbf{X}}}\right)^t\mathbf{C}^{-1}\left(\mathbf{y}_i - \widehat{\mathbf{y}}_{i_{\mathbf{X}}}\right)\right)$ and $SS_{\hat{\varepsilon}_{\mathbf{1}}} = \sum_{i=1}^{M}\left(\left(\mathbf{y}_i - \widehat{\mathbf{y}}_{i_{\mathbf{1}}}\right)^t\mathbf{C}^{-1}\left(\mathbf{y}_i - \widehat{\mathbf{y}}_{i_{\mathbf{1}}}\right)\right)$ to calculate mean squared error, F-ratios, and R^2. Phylogenetic permutation is then performed to determine significance, and results are identical to *D*-PGLS regression (Adams 2014b; Adams and Collyer 2015). It should be noted that although phylogenetically independent contrasts can be used to calculate these quantities (Felsenstein 1985), phylogenetic permutation must be performed on raw permuted values (*not* contrasts), and then independent contrasts must be recalculated for each permutation (Adams and Collyer 2015). However, in contrast to the findings of Adams and Collyer (2015), an appropriately and efficiently implemented phylogenetically independent contrasts-based approach is indeed faster than *D*-PGLS (Fig. 3.2 c-d).

Phylogenetic partial least squares. In addition to the methods discussed above, Adams and Felice (2014) proposed a distance-based method for evaluating covariation between two multivariate traits $\mathbf{Y}_1$ and $\mathbf{Y}_2$ called phylogenetic partial least squares (PLS).

First, the evolutionary rate matrix $\mathbf{R}$ for $\mathbf{Y}$ is calculated using the generalized least squares restricted maximum likelihood estimator:

$$\mathbf{R} = \frac{\big(\mathbf{Y} - \mathbf{X}\,E(\mathbf{Y})\big)^{t}\mathbf{C}^{-1}\big(\mathbf{Y} - \mathbf{X}\,E(\mathbf{Y})\big)}{N-1} \tag{9}$$

The covariance of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ is partitioned into four blocks

$$\mathbf{R} = \begin{pmatrix}\mathbf{R}_{11} & \mathbf{R}_{12}\\ \mathbf{R}_{21} & \mathbf{R}_{22}\end{pmatrix} \tag{10}$$

and singular-value decomposition is subsequently performed on $\mathbf{R}_{12}$. Next, the values $\mathbf{Y} - E(\mathbf{Y})$ are projected onto the phylogenetic transformation matrix $\mathbf{T}$, and $\mathbf{T}\big(\mathbf{Y}_1 - E(\mathbf{Y}_1)\big)$ and $\mathbf{T}\big(\mathbf{Y}_2 - E(\mathbf{Y}_2)\big)$ are matrix multiplied by the left (**U**) and right (**V**) singular vectors of $\mathbf{R}_{12}$, respectively. The first two columns of the resulting scores are regressed to determine the evolutionary correlation between $\mathbf{Y}_1$ and $\mathbf{Y}_2$, and phylogenetic permutation is used to assess significance of the PLS regression (Adams and Felice 2014; *geomorph* function *phylo.pls*). To perform the covariance-based equivalent of phylogenetic PLS (as implemented in the *phylocurve* function *fast.geomorph.phylo.pls*), $\big(\mathbf{Y}_1 - E(\mathbf{Y}_1)\big)\mathbf{U}_1$ and $\big(\mathbf{Y}_2 - E(\mathbf{Y}_2)\big)\mathbf{V}_1$ are regressed using phylogenetically independent contrasts (Felsenstein 1985) regressed through the origin, or equivalently using PGLS. The resulting regression correlation is equivalent to the PLS correlation obtained from distance-based phylogenetic PLS, and significance is assessed using phylogenetic permutation (Adams and Felice 2014).

Pairwise Composite Likelihood for High-Dimensional Comparative Models

Distance-based methods (and covariance-based equivalents) offer an algorithmic solution to high-dimensional comparative problems. However, parameters for many types

of models (e.g., non-Brownian evolution, missing data, within-species variation) lack closed-form solutions, and parameters must be estimated by maximizing the likelihood function. Because there is no likelihood function for distance-based methods, a covariance-based framework is vital to expanding the capabilities of high-dimensional comparative methods. For multivariate phylogenetic comparative models, the log-likelihood function is defined as

$$\log L(\boldsymbol{\theta}|\mathbf{Y}) = -\frac{1}{2}\left(\left(\left(\mathbf{Y} - E(\mathbf{Y})\right)^t(\mathbf{R}\otimes\mathbf{C})^{-1}\left(\mathbf{Y} - E(\mathbf{Y})\right)\right) + \log|\mathbf{R}\otimes\mathbf{C}| + NM\log(2\pi)\right) \tag{11}$$

and the restricted log-likelihood is defined as

$$\log L(\boldsymbol{\theta}|\mathbf{Y}) = -\frac{1}{2}\left(\left(\left(\mathbf{Y} - E(\mathbf{Y})\right)^t(\mathbf{R}\otimes\mathbf{C})^{-1}\left(\mathbf{Y} - E(\mathbf{Y})\right)\right) + \log|\mathbf{R}\otimes\mathbf{C}| + \log|\mathbf{X}^t(\mathbf{R}\otimes\mathbf{C})^{-1}\mathbf{X}| + (NM - M)\log(2\pi)\right) \tag{12}$$

where $\mathbf{Y}$ contains species values for each trait stacked into a single $NM \times 1$ column vector and $\mathbf{X}$ is an $NM \times M$ matrix consisting of ones and zeros describing which rows and columns correspond to elements of $\mathbf{Y}$ ($\mathbf{X}_{ij} = 0$ when $i \neq j$ and $\mathbf{X}_{ij} = 1$ when $i = j$). Unfortunately, the likelihood function become unstable as the number of traits approaches the number of species, and is undefined when $M \geq N$. Additionally, the number of parameters to estimate for an $M \times M$ symmetric matrix is $(M^2 - M)/2 + M$, which may fail to converge on the maximum likelihood parameters even with moderate trait dimensionality.

A potential solution to this dilemma is the substitution of a pseudolikelihood metric into existing likelihood-based estimation methods. It has been shown that the

product of the likelihoods (i.e., the sum of the log-likelihoods) for all possible pairwise combinations of variables, termed *pairwise composite likelihood*, shares many desirable properties with the full likelihood function. In particular, maximum pairwise composite likelihood estimates are consistent, unbiased, and asymptotically normal (Cox and Reid 2004; Varin and Vidoni 2005; Fieuws and Verbeke 2006). It is straightforward then to reduce high-dimensional problems into multiple small maximum likelihood estimation problems.

Specifically, for M traits, $(M^2 - M)/2$ pairwise models must be estimated. Although this is a large number of models, each individual estimation problem contains an extremely small number of parameters which can be estimated for each pairwise trait combination using efficient linear-time computations (Felsenstein 1973; Freckleton 2012; Ho and Ané 2014). Additionally, the log-likelihood function for each pairwise trait combination is computationally stable because $N \gg 2$. For ultra-high dimensional traits (e.g., $M > 1000$), pairwise composite log-likelihood can be approximated using Monte Carlo sampling, as the total number of pairwise likelihood combinations may be computationally prohibitive.

A general framework for hypothesis testing and model selection using pairwise composite likelihoods. Despite the reliability of maximum pairwise composite likelihood estimates, pairwise likelihoods represent overlapping information (and are therefore non-independent from one another), so pairwise composite likelihoods cannot be used for conventional model selection criteria such as AIC, BIC, and likelihood ratio tests (Varin and Vidoni 2005). Similarly, standard errors of parameter estimates based on

Fisher information matrices (which are also non-independent among pairwise trait combinations) are uninterpretable (Fieuws and Verbeke 2006).

Instead, a parametric bootstrapping procedure is adapted from the methods developed by Boettiger et al. (2012). To compare two models, the pairwise composite log-likelihood is calculated for the null model (e.g., simple Brownian motion) and for the alternative model (e.g., Brownian motion with different rates for two groups of species, as in Adams 2014c). Next, random data is simulated from the parameters of the null model 1,000 (or more) times, and pairwise composite log-likelihoods are estimated for both the null and alternative models (refit to the simulated data). The likelihood ratio test statistic $\delta = -2(\log L_{null} - \log L_{alt})$ is computed for the observed data (δ_{obs}) and for the data simulated under the null hypothesis ($\delta_{\text{sim.null}}$). For a nominal significance level of $P \leq 0.05$, the critical value for the test statistic (δ_*) is set so that 95% of $\delta_{\text{sim.null}}$ values fall under δ_*. The proportion of times $\delta_{\text{obs}} \geq \delta_{\text{sim.null}}$ provides an approximation of the *P*-value for comparing the null model to the alternative model. A similar procedure can be applied to assess the statistical power of a model comparison by simulating the alternative model many times (1,000 or more) and calculating $\delta_{\text{sim.alt}}$. The proportion of $\delta_{\text{sim.alt}}$ values greater than or equal to δ_* provides an approximation of the statistical power of the test (Boettiger et al. 2012).

This procedure is extremely flexible and maintains appropriate Type I error and high statistical power for the comparative methods described here. In particular, because the likelihood ratio statistic can be calculated for any model, complex combinations of multiple evolutionary hypotheses (e.g., fixed effects, multiple evolutionary rates, phylogenetic signal, etc.) may be incorporated simultaneously, whereas hypothesis tests

for various distance-based models (which rely on model-specific metrics, such as ratios of *F*-statistics or evolutionary rates) cannot be combined.

The approach described above (as well as the methods described in the following sections) is implemented in *phylocurve*, in which null and alternative models can be fit separately using the *evo.model* function and subsequently compared via parametric bootstrapping (Boettiger et al. 2012) using the *compare.models* function.

New hypothesis tests for existing high-dimensional methods. To simulate the null hypothesis for comparing rates (σ^2_{mult}) among regimes ($\sigma^2_{\text{species.sub}}$), traits are simulated (for example, by using the *sim.char* function in *geiger* (Pennell et al. 2014)) using the restricted maximum likelihood evolutionary rate matrix (equation 9; for the maximum likelihood rate matrix, $N - 1$ is simply replaced with N). To simulate the alternative hypothesis (distinct evolutionary rates among regimes), the following procedure is used: 1) estimate the evolutionary rate matrix for each regime using either the censored method (O'Meara et al. 2006), the noncensored approach (O'Meara et al. 2006), or by subsetting transformed residuals (Adams 2014c); 2) determine the proportion of each tree edge to be assigned to each respective regime rate matrix, for instance by assigning entire clades values of either zero or one (for known discrete regime shifts), or by reconstructing the probabilities of ancestral regime states (Yang et al. 1995; Pupko et al. 2000; Paradis et al. 2004; Revell 2012); 3) simulate phenotypic evolution under the evolutionary rate matrix for each regime by scalar multiplication of $\mathbf{R}_j$ by the branch lengths of the phylogeny and by the regime-specific proportions determined in (2); and 4) add the resulting simulated phenotypic values together. Note that this procedure differs substantially from the null hypothesis described in Adams

(2014c), which assumes a diagonal evolutionary rate matrix and results in unacceptably high Type I error for correlated traits (see below). In contrast, the procedure described here results in appropriate Type I error and statistical power to compare evolutionary rates among regimes.

To compare evolutionary rates among traits (Denton and Adams 2015), the null hypothesis is simulated under a modified evolutionary rate matrix in which the diagonal of **R** is constrained to equal σ^2_{mult} (the mean of the diagonal of **R**). For the alternative hypothesis, the diagonal of **R** is divided into trait groups subsets, in which subset k is set to equal $\sigma^2_{\text{trait.sub}_k}$ (the mean of the diagonal of the subset of **R** corresponding to traits represents in trait group k). In many cases, the resulting matrix is not positive semidefinite. Following Denton and Adams (2015), the nearest positive definite matrix to the constrained rate matrix is found using the *nearPD* function in the *Matrix* package for trait simulations (Bates and Maechler 2015).

To assess correlations between two multivariate traits, the full evolutionary rate matrix is partitioned into four blocks ($\mathbf{R}_{11}$, $\mathbf{R}_{21}$, $\mathbf{R}_{12}$, $\mathbf{R}_{22}$ – see above) in an approach comparable to phylogenetic partial least squares (Adams and Felice 2014). However, rather than performing singular value decomposition on the evolutionary rate matrix, the null hypothesis is simulated by setting all elements of blocks $\mathbf{R}_{12}$ and $\mathbf{R}_{21}$ to zero. For the alternative hypothesis, data are simulated under the unconstrained rate matrix (**R**). In a similar manner, the presence of significant evolutionary covariation among traits can be tested by setting non-diagonal elements of the rate matrix to zero for null hypothesis simulations.

To test the significance of fixed effects (as in *D*-PGLS (Adams 2014b)), both the null and alternative hypotheses are simulated under **R**, and **Xβ** (equation 8) is added to the simulated **Y** under the alternative hypothesis. It should be noted that restricted likelihood (or restricted pairwise composite likelihood) cannot be used for model comparisons in which fixed effects differ between the null and alternative hypotheses. Instead, comparisons must be made using maximum pairwise composite likelihood estimates.

Finally, an alternative method is proposed to test the significance of Blomberg's *K* and K_{mult}. Rather than phylogenetic permutation (Blomberg et al. 2003; Adams 2014a), phylogenetic simulation of both the null and alternative hypotheses is proposed. Under Blomberg's *K*, the null hypothesis is an absence of phylogenetic signal, so data are simulated under **R** on a star phylogeny; for the alternative hypothesis (Brownian motion), data are simulated under **R** on the original phylogeny. For this procedure, *K* is used as the summary statistic (rather than the likelihood ratio δ).

The null distribution of *K* is used to calculate the critical value (K_*), and the proportion of $K_{\text{obs}} \geq K_{\text{sim.null}}$ is the *P*-value for the test of phylogenetic signal. The mean of $K_{\text{sim.alt}}$ provides the expectation of *K* under Brownian motion (which should be approximately 1.0 if the model is correctly specified), and the proportion of $K_{\text{sim.alt}} \geq K_*$ provides an estimate of the statistical power to detect significant phylogenetic signal. To simplify calculations under complex evolutionary models, the expectation of the ratio of raw to phylogenetic mean squared error (the denominator of K_{mult}) can be approximated by simulation under the alternative hypothesis (see below). This implementation of K_{mult} is implemented in the *phylocurve* function *K.mult*.

Estimation of alternative evolutionary models. Parameters for alternative evolutionary models, such as Early-Burst (Harmon et al. 2010) or Ornstein-Uhlenbeck (Hansen 1997), or tree transformations such as Pagel's λ (Pagel 1999) can be fit to high-dimensional comparative models by transforming the branch lengths such that a Brownian motion-like process of trait evolution applies on the transformed tree. Estimation of maximum pairwise composite likelihood tree transformation parameters proceeds as follows: 1) transform the phylogeny according to an initial guess for a tree transformation parameter (e.g., α for an Ornstein-Uhlenbeck process); 2) estimate maximum likelihood parameters for each pairwise combination of traits using closed-form solutions (if available) or by numerical optimization; 3) sum the pairwise log-likelihoods for each combination of traits; 4) repeat steps 1-3 with a new guess for the tree transformation parameter until convergence on the maximum pairwise composite likelihood estimate is achieved. Given a particular tree transformation, the resulting estimates of evolutionary rate and phylogenetic mean for any given trait will be identical across individual pairwise models, assuming no missing data and only a single observation per species.

Combining multiple evolutionary hypotheses. The hypothesis testing framework described above (based on Boettiger et al. 2012) allows for straightforward combinations of multiple evolutionary hypotheses by imposing appropriate alterations or constraints to the evolutionary rate matrix or by adding predicted values based on fixed effects to simulated data ($\mathbf{Y}_{sim} + \mathbf{X\beta}$). To incorporate multiple models into tests of

phylogenetic signal, K_{mult}, the expectation of the ratio of raw mean squared error to phylogenetically corrected mean squared error

$$E(\text{MSE}_0/\text{MSE}) = E\left(\frac{\sum_{i=1}^{M}\left(\left(\mathbf{y}_i - E(\mathbf{y}_i)\right)^t\left(\mathbf{y}_i - E(\mathbf{y}_i)\right)\right)}{\sum_{i=1}^{M}\left(\left(\mathbf{y}_i - E(\mathbf{y}_i)\right)^t\mathbf{C}^{-1}\left(\mathbf{y}_i - E(\mathbf{y}_i)\right)\right)}\right) \tag{13}$$

is estimated by simulation under the hypothesized evolutionary model. Under simple Brownian motion, this should yield an estimate of approximately $(\text{tr}(\mathbf{C}) - N(\mathbf{1}^t\mathbf{C}^{-1}\mathbf{1})^{-1})/(N-1)$ (Blomberg et al. 2003). The observed ratio MSE_0/MSE is then scaled by $E(\text{MSE}_0/\text{MSE})$ to calculate K_{mult}:

$$K_{\text{mult}} = \frac{\dfrac{\sum_{i=1}^{M}\left(\left(\mathbf{y}_i - E(\mathbf{y}_i)\right)^t\left(\mathbf{y}_i - E(\mathbf{y}_i)\right)\right)}{\sum_{i=1}^{M}\left(\left(\mathbf{y}_i - E(\mathbf{y}_i)\right)^t\mathbf{C}^{-1}\left(\mathbf{y}_i - E(\mathbf{y}_i)\right)\right)}}{E(\text{MSE}_0/\text{MSE})} \tag{14}$$

Next, the hypothesized model is simulated on a star phylogeny to obtain the null distribution of K_{mult}, which is used to calculate the critical value (K_*). As before, the proportion of $K_{\text{obs}} \geq K_{\text{sim.null}}$ is the *P*-value for testing phylogenetic signal, and the proportion of $K_{\text{sim.alt}} \geq K_*$ provides an estimate of the statistical power of the test. Under simple Brownian motion, equations 6 and 7 should be nearly identical to 14, whereas under deviations from simple Brownian motion, equation 14 yields a generalization of K_{mult} which can incorporate fixed effects (by setting $E(\mathbf{Y}) = \mathbf{X\beta}$), multiple evolutionary rate regimes, non-Brownian evolutionary models, and other model specifications. These features are implemented in the *phylocurve* function *K.mult*.

Incorporation of missing data and within-species variation. Several methods have been developed to estimate evolutionary trait covariance in the presence of missing data and within-species variation (Ives et al. 2007; Felsenstein 2008; Bruggeman et al.

2009; Hansen and Bartoszek 2012). These methods can all be estimated in a maximum likelihood (or restricted likelihood) framework and can be readily incorporated into pairwise composite likelihood estimation. However, for such methods, there is no algorithmic solution for estimating evolutionary covariances, so step 2 of the outlined method for parameter estimation (see above) requires numerical optimization of covariance parameters. Problematically, estimated evolutionary rates for datasets with missing data or within-species variation will be different if parameters for pairwise trait combinations are estimated separately. One solution is to maximize the pairwise log-likelihood for all parameters in a single optimization routine, which constrains variance parameters for individual traits to be identical across pairwise combinations. This is a potentially large optimization problem which may have difficulty converging. A potential simplification involves separate maximum likelihood estimation for each pairwise trait combination (as described above), followed by averaging each estimate of the evolutionary rate for a given trait. The resulting estimates are not equivalent to the maximum pairwise composite likelihood estimates but should provide reasonable starting parameters for a single optimization routine. Alternatively, the averaged estimates could be used as is to approximate the maximum pairwise composite likelihood parameters. Optimal strategies for incorporating within-species variation and missing data for high-dimensional models require further investigation that is beyond the scope of this paper (but see Denton and Adams (2015), in which a bootstrapping approach is performed on randomly sampled within-species individual observations). Approaches for dealing with within-species variation and missing data are not yet (as of the time of publication) implemented in *phylocurve*.

For univariate data, both generalized least squares (GLS) and distance-based comparative approaches require the inversion of $N \times N$ matrices. For multivariate data, likelihood calculations require inversion of $MN \times MN$-dimension matrices. Because the memory and computation time required for individual matrix inversions increases polynomially with increases in species number, individual matrix inversions for extremely large phylogenies may take several hours or fail entirely (Ho and Ané 2014). Additionally, maximum likelihood and Bayesian approaches may require thousands or millions of matrix inversions. In recent years, several approaches have been developed that avoid large matrix inversions entirely and operate with linear time and memory requirements, thus substantially reducing the burden of working with large comparative datasets. For the covariance-based and pairwise composite likelihood-based high-dimensional multivariate approaches discussed in this paper, a variety of fast linear-time algorithm are implemented in the *phylocurve* R package using methods adapted from Felsenstein (1973; 1985), Freckleton (2012), and Ho and Ané (2014) (see also FitzJohn 2012). As described in previous sections, all *phylocurve* functions that begin with '*fast.geomorph.*' provide fast covariance-based implementations of analogous *geomorph* functions. These functions, which produce identical results to distance-based methods, should *not* be confused with the pairwise composite likelihood framework described in this paper, which is implemented in the *phylocurve* functions *evo.model*, *compare.models,* and *K.mult*.

Efficient calculations for high-dimensional data. For repeated calculations on a given phylogeny (e.g., as is associated with bootstrapping procedures), a substantial

portion of the computational burden of phylogenetically independent contrasts (a linear-time algorithm) is associated with redundant operations (e.g., reordering internal edges, conversion of data types, etc.). In some cases, repeating these steps can account for over 90% of computational time. To minimize redundant operations, all preparation steps can be performed in a single function call, and then relevant quantities (e.g., phenotypic data or branch lengths) can be updated as needed prior to calling the independent contrasts algorithm. Additionally, phylogenetically independent contrasts can be performed simultaneously on all traits of interest by performing relevant calculations on the full matrix $\mathbf{Y}$ rather than on individual vectors. Refer to Appendix 1 for details on implementing these methods.

The quantity $\big(\mathbf{Y} - E(\mathbf{Y})\big)^t \mathbf{C}^{-1} \big(\mathbf{Y} - E(\mathbf{Y})\big)$, which is simply the cross-product of the matrix of independent contrasts of $\mathbf{Y}$, can be calculated efficiently in linear time using the modified phylogenetically independent contrasts algorithm described above. The independent contrasts algorithm also automatically calculates $E(\mathbf{Y})$, so a separate ancestral reconstruction is unnecessary (see Appendix A). Contrast variances for each node ($\boldsymbol{v}$) and the length of the two edges extending from the root of the independent contrast-transformed tree ($\boldsymbol{x}$) can be used to calculate $\mathbf{1}^t \mathbf{C}^{-1} \mathbf{1} = \prod \boldsymbol{x} / \sum \boldsymbol{x}$ and $\log|\mathbf{C}| = \log(\prod \boldsymbol{x} / \sum \boldsymbol{x}) + \sum \log \boldsymbol{v}$ in linear time (Felsenstein 1985; Freckleton 2012). Using these formulas, any quantity of the form $\mathbf{L}^t \mathbf{C}^{-1} \mathbf{R}$ (corresponding to left and right matrices of compatible dimensions, such as $\mathbf{X}$ and $\mathbf{Y}$) can be computed in linear time as $\mathbf{L}^t \mathbf{C}^{-1} \mathbf{R} = \mathbf{L}^t{}_{pic} \mathbf{R}_{pic} + E(\mathbf{L}) E(\mathbf{R})^t (\mathbf{1}' \mathbf{C}^{-1} \mathbf{1})$ (for a related linear-time algorithm, see Ho and Ané 2014). Quantities required for the multivariate log-likelihood may also be calculated efficiently: $\log|\mathbf{R} \otimes \mathbf{C}| = n \log|\mathbf{R}| + m \log|\mathbf{C}|$ and

$\log|\mathbf{X}'(\mathbf{R}\otimes\mathbf{C})^{-1}\mathbf{X}| = m\log\left|\left(\mathbf{X}'_{pic}\mathbf{X}_{pic} + \mathbf{X}_{anc}\mathbf{X}'_{anc}(\mathbf{1}'\mathbf{C}^{-1}\mathbf{1})\right)\right| - p\log|\mathbf{R}|$ (where p is the number of columns of **X**).

A more complex approach is required to accommodate multiple regimes. The following algorithm, modified from the linear-time algorithm described by Ho and Ané (2014), calculates quantities of the form $\mathbf{Q} = \mathbf{L}^t\mathbf{W}^{-1}\mathbf{R}$ and $\log|\mathbf{W}|$ in linear time relative to the number of species, where **W** is the $NM \times NM$ matrix describing species-trait covariance for multiple regimes (under a single regime, $\mathbf{W} = \mathbf{R}\otimes\mathbf{C}$), and **L** and **R** are matrices of compatible dimensions (e.g., **X** and **Y**). These quantities can then be used for log-likelihood calculations (refer to Ho and Ané (2014) for details of the original approach on which the following is based).

1. For a tree with a single tip, let $\mathbf{Z}^{(e)} = \sum_{a=1}^{r}\left(t_e p_{a,e}\mathbf{R}_a\right)$ be an $M \times M$ matrix representing the length of edge *e* scaled by regime-specific rate matrices and their respective proportions assigned to each edge. Then $\mathbf{p} = {\mathbf{Z}^{(e)}}^{-1}$, $\mathbf{U}_l^t = \mathbf{L}^t\mathbf{p}$, $\mathbf{V}_r = (\mathbf{R}^t\mathbf{p})^t$, $\mathbf{Q} = \mathbf{L}^t\mathbf{p}\mathbf{R}$, and $\log|\mathbf{W}| = \log\left|\mathbf{Z}^{(e)}\right|$.
2. For a tree with two or more tips and root edge *e*, again let $\mathbf{Z}^{(e)} = \sum_{a=1}^{r}\left(t_e p_{a,e}\mathbf{R}_a\right)$. Then $\mathbf{p}_A = \sum_s \mathbf{p}_s$, $\log|\mathbf{W}| = \sum_s \log|\mathbf{W}_s| + \log\left|\mathbf{I} + \mathbf{Z}^{(e)}\mathbf{p}_A\right|$, $\mathbf{p} = \mathbf{p}_A\left(\mathbf{I} + \mathbf{Z}^{(e)}\mathbf{p}_A\right)^{-1}$, $\mathbf{Q} = \left(\sum_s \mathbf{Q}_s\right) - \mathbf{U}_{l,s}^t\left(\mathbf{I} + \mathbf{Z}^{(e)}\mathbf{p}_A\right)^{-1}\mathbf{Z}^{(e)}\mathbf{V}_{r,s}$, $\mathbf{U}_l^t = \left(\sum_s \mathbf{U}_{l,s}^t\right)\left(\mathbf{I} + \mathbf{Z}^{(e)}\mathbf{p}_A\right)^{-1}$, and $\mathbf{V}_r^t = \left(\left(\sum_s \mathbf{V}_{r,s}^t\right)\left(\mathbf{I} + \mathbf{Z}^{(e)}\mathbf{p}_A\right)^{-1}\right)^t$.
3. At the root of the full tree, return **Q** and $\log|\mathbf{W}|$.

Fast covariance-based implementations of distance-based methods. The computational performance of distance-based methods implemented in *geomorph*

(*compare.evol.rates*, *compare.multi.evol.rates*, *physignal*, *phylo.pls*, and *procD.pgls*) were compared to fast covariance-based *phylocurve* functions (*fast.geomorph.compare.evol.rates*, *fast.geomorph.compare.multi.evol.rates*, *fast.geomorph.physignal*, *fast.geomorph.phylo.pls*, and *fast.geomorph.procD.pgls*). Computational times for each method were compared using 10, 24, 50, 100, 250, 500, and 1,000 species and 10, 24, 50, 100, and 250 traits (Figs. 3.1-3.2). For datasets with a low number of species (<50), the speed of distance-based methods was generally comparable to covariance-based methods. For datasets with higher numbers of species, covariance-based methods were consistently faster (in some cases up to ~1,000 times faster). The most time-consuming steps of distance-based methods in *geomorph* include inverting the phylogenetic covariance matrix and performing an eigendecomposition of the phylogenetic covariance matrix, both of which are avoided by covariance-based methods (with the exception of *fast.geomorph.compare.evol.rates*, which requires an eigendecomposition for computing the phylogenetic transformation matrix, unless the censored approach is used). Covariance-based functions that rely on phylogenetic permutation for hypothesis tests (*fast.geomorph.procD.pgls*, *fast.geomorph.physignal*, and *fast.geomorph.phylo.pls*) operate in linear time relative to the number of species. Although *fast.geomorph.compare.evol.rates* and *fast.geomorph.compare.multi.evol.rates* are substantially faster than *compare.evol.rates* and *compare.multi.evol.rates* (respectively) for datasets with a large numbers of species, all four functions rely on the *sim.char* function in *geiger* (Pennell et al. 2014) for hypothesis testing, which operates with polynomial increases in time as the number of species increases and is thus a nonlinear rate-limiting step for *fast.geomorph.compare.multi*.

Ultra-high dimensional traits. The pairwise composite log-likelihood approach may become excessively cumbersome for extremely high-dimensional traits. For example, for a 1,024-dimensional trait there are 523,776 pairwise trait combinations, a prohibitively large number for parameter estimation, as the pairwise log-likelihood must be calculated several times for numerical optimization. For simulation-based hypothesis testing, this optimization procedure must be repeated a large number of times. For instance, supposing numerical optimization of a tree transformation parameter (e.g., Early-Burst rate) requires 50 log-likelihood calculations, and 1,000 null hypothesis simulations are performed, a total of $50 \times 1{,}000 \times 523{,}776 = 26{,}188{,}800{,}000$ pairwise log-likelihood evaluations would be required. For such computationally infeasible problems, a simple Monte Carlo-based approach is proposed. Rather than computing every possible pairwise log-likelihood, a random subset of pairwise combinations is sampled, and the subsetted pairwise composite log-likelihood is divided by the number of random samples and then multiplied by the total number of possible pairwise combinations. As the number of random samples increases, the estimated composite log-likelihood will approach the true pairwise composite log-likelihood. The composite log-likelihood surface for the tree transformation parameter is approximated by estimating the composite log-likelihood for several parameter values spanning the range of feasible parameters. Regression (e.g., polynomial or Gaussian process regression) is then performed on the approximated log-likelihood surface, and the parameter value corresponding to the maximum predicted value of the regression is used as the parameter estimate.

SIMULATION METHODS

To compare the Type I error and statistical power of distance-based methods with the pairwise composite log-likelihood approaches described here, simulations were performed under a variety of evolutionary scenarios. For each scenario, 1,000 random datasets were generated on randomly generated 32-species pure-birth phylogenies using the *pbtree* function in *phytools* (Revell 2012), and subsequently using the *sim.traits* and *sim.groups* functions in *phylocurve* (both of which call the *sim.char* function in *geiger* (Pennell et al. 2014)). Hypothesis tests were performed using both distance-based and pairwise composite log-likelihood-based parametric bootstrapping simulation (Boettiger et al. 2012), with 1,000 bootstrap simulations per simulated dataset.

To evaluate the statistical performance of comparing evolutionary rates among regimes, two regimes were assigned 16 species each. For each simulation, a random evolutionary rate matrix was simulated by parameterizing the upper triangle of an $M \times M$ with a randomly generated number drawn from the standard normal distribution. The transpose of the matrix was then matrix multiplied by itself, resulting in a random positive-definite covariance matrix ($\mathbf{R}_{\text{rand}}$). For one of the simulated regimes, traits were simulated under $\mathbf{R}_{\text{rand}}$; the other regime was simulated under $\mathbf{R}_{\text{rand}}$ scalar multiplied by either 1.0 (for Type I error), 1.5, 2.0, 3.0, or 4.0. Simulations were conducted for 2, 16, 32, and 64 traits. Similarly, the statistical performance of comparing evolutionary rates among traits was evaluated by generating $\mathbf{R}_{\text{rand}}$, subdividing the diagonal of $\mathbf{R}_{\text{rand}}$ into two groups, and setting one group to the mean of the diagonal of $\mathbf{R}_{\text{rand}}$ and the other group to the mean of the diagonal of $\mathbf{R}_{\text{rand}}$ multiplied by either 1.0 (for Type I error), 1.5, 2.0, 3.0, or 4.0. Simulations were conducted for 4, 8, 16, 32, and 64 traits. For

simulations involving fixed effects (for comparison with *D*-PGLS), the diagonal of **R** (of dimension $(M + 1) \times (M + 1)$) was set to 1.0 and off-diagonal elements were set to either 0.0 (for Type I error), 0.1, 0.3, 0.5, 0.7, or 0.9, and data were simulated for 2, 5, 10, 25, and 32 traits.

To assess the statistical performance of testing evolutionary covariance between two groups of multivariate traits (for comparison with phylogenetic partial least squares), the diagonal of **R** (of dimension $M \times M$) was set to 1.0 and off-diagonal elements were set to either 0.0 (for Type I error), 0.1, 0.3, 0.5, 0.7, or 0.9. To test the power of multivariate phylogenetic signal (K_{mult}), trait data was simulated under $\mathbf{R}_{\text{rand}}$ on a transformed phylogeny with Pagel's λ set to either 0.0 (for Type I error), 0.05, 0.1, 0.25, 0.5, 0.75, and 1.0 for 2, 5, 10, 32, and 64 traits. For simulations testing evolutionary covariance, traits were simulated under **R** with the diagonal set to 1.0 and off-diagonal elements set to either 0.0 (for Type I error), 0.1, 0.3, 0.5, 0.7, or 0.9 for 2, 5, 10, 32, and 64 traits.

Additionally, alternative evolutionary models were simulated under a randomly $\mathbf{R}_{\text{rand}}$ for 2, 5, 10, 25, 31, and 50 traits, and simulations were performed on phylogenies with the following tree transformations: Early-Burst (rate = 0.0 (for Type I error), -0.25, -0.5, -0.75, -1.0), Ornstein-Uhlenbeck (α = 0.0 (for Type I error), 0.25, 0.5, 1.0, 2.0), and Pagel's λ (λ = 1.0 (for Type I error), 0.75, 0.5, 0.25, 0.0). R code for performing the simulations used to generate Figures 3.3, 3.5, and 3.6 is provided in the Dryad supplement.

STATISTICAL PERFORMANCE OF PAIRWISE COMPOSITE LOG-LIKELIHOOD: COMPARISONS WITH DISTANCE-BASED METHODS

Comparing evolutionary rates. For comparisons of evolutionary rates among regimes (Fig. 3.3 b,d) and among groups of traits (Fig. 3.5 a-b) using both distance-based and pairwise composite likelihood-based hypothesis testing, all simulated scenarios displayed appropriate Type I error (approximately 0.05). As expected, as trait dimensionality was increased, statistical power also increased (as in Adams (2014c) and Denton and Adams (2015)), although distance-based approaches exhibited somewhat higher statistical power for rate ratios of 1.5 and 2.0 (both methods had statistical power of approximately 1.0 at higher rate ratios). Regime-specific evolutionary rates were also compared using the censored method (Fig. 3.3c) (O'Meara et al. 2006), which displayed appropriate Type I error and statistical power approximately identical to that of the distance-based procedure.

It should be noted that a modification to the null hypothesis described in Adams (2014c) was implemented for comparing rates among regimes. Adams (2014c) simulated the null distribution of rate ratios by setting the null hypothesis to be a diagonal evolutionary rate matrix, with each trait independently evolving under a Brownian motion rate of σ^2_{mult}. Because this approach fails to account for trait covariation, this procedure results in unacceptably high Type I error rates (Fig. 3.3a), as trait covariance artificially inflates the observed differences among regime-specific rates relative to independently simulated traits. This error-prone procedure was implemented in the supplemental code provided in Adams (2014c) as well as in the *compare.evol.rates* function in *geomorph* up to version 2.1.5 (Adams and Otárola-Castillo 2013). As of

version 2.1.6 of *geomorph*, the null hypothesis has been updated to account for trait covariance (Denton and Adams 2015). Likewise, earlier versions of *phylocurve* (1.0.0-1.3.0), which implemented fast covariance-based analogues of distance-based approaches, also used the original null hypothesis specified in Adams (2014c). *phylocurve* versions 2.0.0 and higher implement the modified null hypothesis incorporating trait covariance (however, the *fast.geomorph.compare.evol.rates* can be used to perform hypothesis tests under the old null hypothesis, as described in Adams (2014c), by setting the *force.diag* option to *TRUE*).

As an example of the potential consequences of failing to account for trait covariation, data from Adams (2014c) consisting of 11 cranial landmark coordinates from nine *Plethodon* salamanders was analyzed using the *phylocurve* function *fast.geomorph.compare.evol.rates* while assuming zero trait covariance (*force.diag* set to *TRUE*). When failing to account for trait covariance, the rate ratio between the two hypothesized regimes (1.84) was found to be significantly different from 1.0 (Fig. 3.4a; P=0.005), identical to the result reported in Adams (2014c). Next, the *phylocurve* function *compare.models* was used to test for significant trait covariation (where the null model was fit using *evo.model* with *diag.phylocov* set to *TRUE* and the alternative model was fit using *evo.model* with *diag.phylocov* set to *FALSE*). The null hypothesis of zero trait covariation was unequivocally rejected (P<0.001). Next, *fast.geomorph.compare.evol.rates* was used while accounting for trait covariance (*force.diag* set to *FALSE*), and the rate ratio (1.84) was found to be non-significant (Fig. 3.4a; P=0.261), in conflict with Adams (2014c).

To simultaneously assess the statistical significance and statistical power of the test using pairwise composite likelihood methods, the *phylocurve* function *compare.models* was used to compare the null hypothesis of a single rate regime and the alternative hypothesis of two regimes. This test, analogous to the *fast.geomorph.compare.evol.rates* test accounting for trait covariance, was also non-significant (P=0.64) and revealed essentially no statistical power (0.097) to detect a significant rate difference given the phylogeny and dataset (Fig. 3.4c).

Comparing evolutionary rates under violations of distance-based assumptions. Distance-based comparisons of evolutionary rates among regimes compare rates under the assumption that rate ratios between regimes are consistent among individual traits. To assess the consequences of deviations from this model, statistical power was assessed under a scenario in which the evolutionary rates among two clades differed substantially but in conflicting ways. Specifically, a 100-dimensional trait was simulated on a 32-species phylogeny divided into two equally sized regimes. Traits 1-50 were simulated with an evolutionary rate of 1.0 for the first regime (species 1-16) and an evolutionary rate of 4.0 for the second regime (species 17-32). Traits 51-100 were simulated with an evolutionary rate of 4.0 for species 1-16 and an evolutionary rate of 1.0 for species 17-32. Thus, the average simulated evolutionary rate for both regimes was 2.5, but substantial rate differences existed between regimes for each trait (where half the traits had a rate ratio of 4.0 and the other half had a rate ratio of 0.25). Simulations were repeated 1,000 times, and the proportion of significant results (corresponding to the statistical power of the test) was compared for distance-based and pairwise composite likelihood-based hypothesis testing.

The statistical power for distance-based methods was approximately 0.05, suggesting that distance-based rate comparisons are poorly equipped to detect rate differences for the biologically realistic scenario of heterogeneous rate ratios. In contrast, the statistical power for pairwise composite likelihood-based regime rates comparisons under the described scenario was exactly 1.0.

Phylogenetic signal: $K_{\mathbf{mult}}$. For testing the significance of K_{mult}, distance-based testing (Adams 2014a) follows the permutation procedure described in Blomberg et al. (2003). For the approach described here, the null hypothesis is simulated on a star phylogeny (see above). Results for both approaches yielded appropriate statistical power (Fig. 3.5 c-d), although the simulation-based approach exhibited decreasing Type I error as trait dimensionality increased (0.008 for simulations with 64 traits), whereas the distance-based maintained a Type I error of approximately 0.05. Consistent with Type I error findings, statistical power was very similar between the two methods but with slightly higher statistical power for distance-based phylogenetic permutation under high trait dimensionality (although both methods exhibited higher power as the number of traits increased).

Fixed effects: comparisons with *D*-PGLS. The statistical performance of testing for fixed effects (described above) in which the null hypothesis (without fixed effects) is compared to the alternative hypothesis (with fixed effects) was compared to the performance of *D*-PGLS. As with phylogenetic signal, *D*-PGLS relies on a phylogenetic permutation testing procedure. The simulation-based pairwise composite likelihood method described here exhibited similar statistical power to *D*-PGLS (Fig. 3.5 e-f). However, *D*-PGLS exhibited high Type I error, which increased as trait dimensionality

was increased (with Type I errors as high as 0.158). In contrast, the simulation-based pairwise composite likelihood method exhibited appropriate Type I error which *decreased* with increasing trait dimensionality (0.018 for simulations with 32 traits).

The reason for the unacceptably high Type I error for *D*-PGLS in these simulations is unknown, as *D*-PGLS exhibited appropriate Type I error in Adams (2014b). One possible explanation is the type of random phylogenies generated for simulations: here, pure-birth phylogenies were generated using the *pbtree* function in *phytools*, but simulations using phylogenies with random splits (using the *rtree* function in *ape*) and branch lengths computed using Grafen's method (1989) (implemented in the *compute.brlen* function in *ape*) appear to yield appropriate Type I error for *D*-PGLS, suggesting a possible sensitivity of *D*-PGLS to tree topology and branch lengths. It should also be noted that the methods described in the original published version of *D*-PGLS were used to assess statistical performance (Adams 2014b) rather than the current *procD.pgls* implementation in *geomorph* (version 2.1.7-1), the latter of which for unknown reasons reported *z*-scores that conflicted with the original published *D*-PGLS implementation of Adams (2014b) and earlier versions of *geomorph*, which in turn resulted in a Type I error rate of 0.0 and extremely low statistical power under all simulated conditions.

Testing for covariance among two multivariate traits: comparisons with phylogenetic partial least squares. Simulations were performed to assess the statistical performance of pairwise composite likelihood-based tests for evolutionary covariance among multiple traits. This test is analogous to phylogenetic partial least squares, as the null hypothesis for both methods is zero evolutionary covariance among two groups of

traits ($\mathbf{R}_{12} = \mathbf{R}_{21} = 0$) and the alternative hypothesis is $\mathbf{R}_{12} \neq \mathbf{R}_{21} \neq 0$. As with tests for phylogenetic signal and fixed effects, phylogenetic partial least squares relies on phylogenetic permutation, whereas pairwise composite likelihood-based tests use Monte Carlo simulations. However, the results of Adams and Felice (2014) could not be replicated for the simulations performed here, as phylogenetic partial least squares suffered from extremely high Type I error that was exacerbated by increasing trait dimensionality (for simulations with 64 traits, Type I error exceeded 0.90) (Fig. 3.5 g-h). In contrast, the Type I error and statistical power for pairwise composite likelihood-based tests was appropriate under all simulated scenarios, and statistical power increased with increases in trait dimensionality (which reflected the statistical performance originally reported for phylogenetic partial least squares by Adams and Felice (2014)).

In light of these findings, simulations were performed under a variety of scenarios in order to identify conditions under which distance-based phylogenetic least squares display appropriate Type I error, and it was determined that the statistical performance of phylogenetic partial least squares depends at least partially on the input evolutionary rate matrix (**R**) for simulations under the null hypothesis. Whereas the simulations for Type I error used to generate Figure 3.5 g-h assumed independently evolving traits (diagonal **R**), a fully parameterized **R** with high covariance within blocks $\mathbf{R}_{11}$ and $\mathbf{R}_{22}$ (and $\mathbf{R}_{12}$ and $\mathbf{R}_{21}$ subsequently set to zero) yielded appropriate Type I error and high statistical power for both distance-based and pairwise composite likelihood-based methods, suggesting that the robustness of distance-based phylogenetic partial least squares is highly sensitive to low or absent within-group covariation.

Testing for evolutionary covariation. Simulations were performed to assess the statistical performance of testing for the presence of evolutionary covariance in a multivariate trait. For this test, the null hypothesis is simulated by constraining the off-diagonal elements of the evolutionary rate matrix to zero, whereas the alternative hypothesis uses an unconstrained evolutionary rate matrix. As expected, this test exhibited statistical performance similar to that of tests for evolutionary covariance between trait groups (Fig. 3.5h), displaying appropriate Type I error rates and increasing statistical power as trait dimensionality increased.

Alternative evolutionary models: Ornstein-Uhlenbeck, Early-Burst, and Pagel's λ. Type I error and statistical power of maximum pairwise composite likelihood estimation was evaluated for three common tree transformations: Ornstein-Uhlenbeck, Early-Burst, and Pagel's λ (Fig. 3.6). Bias and error of parameter estimates was also assessed (Fig. 3.6). Estimates of Pagel's λ and the Early-Burst rate parameter were generally unbiased, whereas the Ornstein-Uhlenbeck parameter α was strongly biased toward 0.5. It should be noted that parameter estimation bias for α was also present in conventional (full) maximum likelihood parameter estimation (as the pairwise and full log-likelihoods for a dataset with only two traits are identical), so parameter bias should not be interpreted as a shortcoming of maximum pairwise composite log-likelihood estimation. Estimates of Pagel's λ were consistent regardless of the number of simulated traits, Early-Burst estimates were moderately consistent, and estimates of α were highly variable. Under all scenarios, model tests exhibited appropriate Type I error and

increasing statistical power with increases in trait dimensionality, but statistical power for α was substantially lower than tests for significant λ and rate.

Ultra-high dimensional traits. A random 32-species phylogeny was generated and multivariate data with 2, 32, 64, and 1,024 traits with random covariance were randomly generated under an Early-Burst model with a rate parameter of -0.5. The pairwise composite log-likelihood surface was assessed for 50 values (spanning -1 to 0) of the Early-Burst rate parameter for the simulated 2, 32, and 64-dimensional traits. For the 1,024-dimensional trait, the Monte Carlo-based approach (described above) was performed using 10,000 random pairwise combinations to approximate the composite log-likelihood surface. Polynomial regression was then performed to obtain a smoothed estimate of the pairwise composite log-likelihood surface. The rate value corresponding to the maximum predicted value of the resulting polynomial regression was -0.49, consistent with simulated conditions and the rates recovered for simulated datasets of dimension 2, 32, and 64 (rate=-0.5) (Fig. 3.7).

CONCLUSION

Distance-based multivariate comparative methods offer a framework for testing evolutionary hypotheses for high-dimensional traits while avoiding several problems associated with high dimensionality (e.g., singular covariance matrices and low statistical power). However, extremely large datasets may be computationally infeasible using a distance-based approach. An equivalent covariance-based approach is described that allows fast linear-time implementation (implemented in the R package *phylocurve*), thus avoiding most of the computational challenges associated with distance-based methods.

A novel approach based on pairwise composite likelihood is also proposed. These methods (also implemented in *phylocurve*) allow for greater flexibility than distance-based methods, such as the ability to incorporate multiple evolutionary hypotheses and alternative evolutionary models. An approach for incorporating within-species variation and missing data is also discussed. Simulations revealed high statistical power and appropriate Type I error for these methods, whereas some distance-based approaches (*D*-PGLS and phylogenetic partial least squares) exhibited elevated Type I error under various scenarios, and regime-specific rate comparisons exhibited extremely low statistical power when rate ratios differed among traits.

It is important to note that despite promising results from statistical simulations, both distance-based methods and the new methods described here make several critical assumptions regarding the evolution of high-dimensional traits. Most notably, all models were simulated such that model parameters applied equally to all traits of interest. However, the consequences of violating these assumptions are unknown. For instance, consider phylogenetic regression on a high-dimensional trait in which a handful of traits are strongly driven by a predictor variable but the remaining traits are independent of the predictor variable. Regardless of the overall *P*-value obtained, the biological relevance of the result is unclear: a significant finding suggests correlation where none exists for the majority of traits, whereas a non-significant result fails to detect potentially important correlations for the handful of affected traits. Under such circumstances, generalized high-dimensional metrics may not be ideal for testing evolutionary hypotheses, although methods to detect this type of situation are currently lacking. While high-dimensional comparative methods provide a potentially powerful framework for approaching the

study of complex phenotypic evolution, this framework will benefit substantially from the development of model diagnostic techniques, such as high-dimensional multivariate methods for detecting violations of model assumptions and for assessing model adequacy (Pennell et al. 2015).

SUPPLEMENTARY MATERIAL – Data available from the Dryad Digital Repository.

REFERENCES

Adams D.C. 2013. Comparing evolutionary rates for different phenotypic traits on a phylogeny using likelihood. *Systematic Biology* 62:181-192.

Adams D.C., Otárola-Castillo E. 2013. geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution* 4:393-399.

Adams D.C. 2014a. A generalized K statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. *Systematic Biology* 63:685-697.

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

Adams D.C. 2014c. Quantifying and comparing phylogenetic evolutionary rates for shape and other high-dimensional phenotypic data. *Systematic Biology* 63:166-177.

Adams D.C., Felice R. 2014. Assessing phylogenetic morphological integration and trait covariation in morphometric data using evolutionary covariance matrices. PLOS ONE 9(4):e94335.

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

Bartoszek K., Pienaar J., Mostad P., Andersson S., Hansen T.F. 2012. A phylogenetic comparative method for studying multivariate adaptation. *Journal of Theoretical Biology* 314:204-215.

Bates D., Maechler M. 2015. Matrix: Sparse and Dense Matrix Classes and Methods. *CRAN*. R package version 1.2-2.

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

Boettiger C., Coop G., Ralph P. 2012. Is your phylogeny informative? Measuring the power of comparative methods. *Evolution*. 66:2240-2251.

Bruggeman J., Heringa J., Brandt B.W. 2009. PhyloPars: estimation of missing parameter values using phylogeny. *Nucleic Acids Research*. 37:W179-W184.

Chitwood, D.H., Ranjan, A., Martinez, C.C., Headland, L.R., Thiem, T., Kumar, R., Covington M.F., Hatcher T., Naylor D.T., Zimmerman S., Downs N., Raymundo N., Buckler E.S., Maloof J.N., Aradhya M., Prins B., Li L., Myles S., Sinha, N.R. 2014. A modern ampelography: a genetic basis for leaf shape and venation patterning in grape. *Plant Physiology*. 164:259-272.

Cox D.R., Reid N. 2004. A note on pseudolikelihood constructed from marginal densities. *Biometrika*. 91:729-737.

Denton J.J., Adams D.C. A new phylogenetic test for comparing multiple high-dimensional evolutionary rates suggests interplay of evolutionary rates and modularity in lanternfishes (Myctophiformes; Myctophidae). *Evolution* 69:2425-2440.

Felsenstein J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics* 25:471-492.

Felsenstein J. 1985. Phylogenies and the comparative method. *American Naturalist* 125:1-15.

Felsenstein J. 2008. Comparative methods with sampling error and within-species variation: Contrasts revisited and revised. *American Naturalist* 171:713-725.

Fieuws S., Verbeke G. 2006. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*. 62:424-431.

FitzJohn R.G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution* 3:1084-1092.

Freckleton R.P. 2012. Fast likelihood calculations for comparative analyses. *Methods in Ecology and Evolution* 3:940-947.

Garland T.J. 1992. Rate tests for phenotypic evolution using phylogenetically independent contrasts. *American Naturalist* 171:713-725.

Garland T., Ives A.R. 2000. Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. *American Naturalist* 155:346-364.

Goolsby E.W. 2015. Phylogenetic comparative methods for evaluating the evolutionary history of function-valued traits. Systematic Biology 64:568-578.

Hansen T.F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51:1341:1351.

Hansen T.F., Bartoszek K. 2012. Interpreting the Evolutionary Regression: The Interplay between Observational and Biological Errors in Phylogenetic Comparative Studies. *Systematic Biology* 61:413-425.

Harmon L.J., Losos J.B., Jonathan Davies T., Gillespie R.G., Gittleman J.L., Bryan Jennings W., Kozak K.H., McPeek M.A., Moreno-Roark F., Near T.J., Purvis A., Ricklefs R.E., Schluter D., Schulte J.A. II, Seehausen O., Sidlauskas B.L., Torres-

Carvajal O., Weir J.T., Mooers A. 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64:2385-2396.

Ho L.S.T., Ané C. 2014. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systematic Biology* 63:397-408.

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

Li, L. 2007. Sparse sufficient dimension reduction. *Biometrika* 94:603-613.

Martins E.P., Hansen T.F. 1997. Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist* 149:646-667.

O'Meara B.C., Ané C., Sanderson M.J., Wainwright P.C. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60:922-933.

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

Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289-290.

Pennell M.W., Eastman J.M., Slater G.J., Brown J.W., Uyeda J.C., FitzJohn R.G., Alfaro M.E., Harmon L.J. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 15:2216-2218.

Pennell M.W., FitzJohn W.K., Cornwell W.K., Harmon L.J. 2015. Model adequacy and the macroevolution of angiosperm functional traits. *American Naturalist* 186:E33-E50.

Pupko T., Pe'er I., Shamir R., Graur D. 2000. A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Molecular Biology and Evolution* 17:890–896.

Rabosky D.L., Donnellan S.C., Grundler M., Lovette I.J. 2014. Analysis and Visualization of Complex Macroevolutionary Dynamics: An Example from Australian Scincid Lizards. *Systematic Biology* 63:610-627.

Revell L.J., Collar D.C. 2009. Phylogenetic analysis of the evolutionary correlation using likelihood. *Evolution* 63:1090-1100.

Revell L.J. 2010. Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution* 1:319-329.

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

Thomas G.H., Freckleton R.P., Székely T. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. *Proceedings of the Royal Society B* 273:1619-1624.

Varin C., Vidoni P. 2005. A note on composite likelihood inference and model selection. *Biometrika* 92:519-528.

Yang Z., Kumar S., Nei M. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* 141:1641-1650.

FIGURES

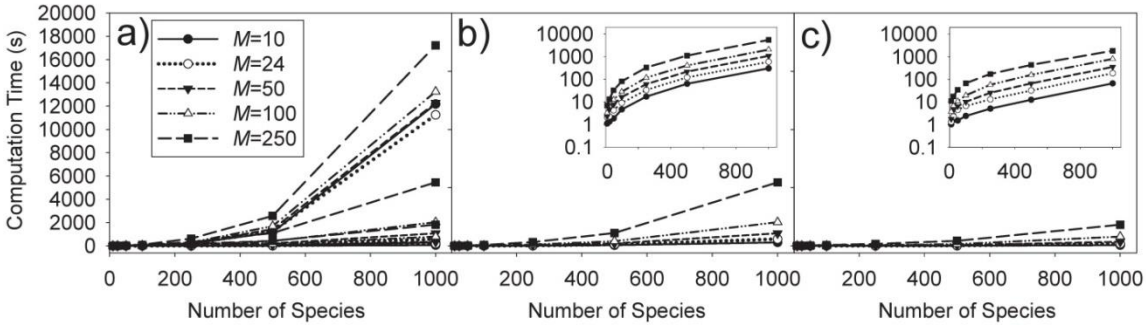


Figure 3.1. Speed comparisons for rate comparisons among species rate regimes using a) the distance-based *geomorph* function *compare.evol.rates*, b) the covariance-based *phylocurve* function *fast.geomorph.compare.evol.rates* (based on phylogenetic transformation of residuals; *censored=FALSE*), and c) and the covariance-based *phylocurve* function *fast.geomorph.compare.evol.rates* using the censored method of O'Meara et al. (2006) (*censored=TRUE*). Speed comparisons were performed on randomly generated pure-birth phylogenies containing 10, 24, 50, 100, 250, 500, and 1,000 species with simulated data containing 10, 24, 50, 100, and 250 traits. Computation times on a $\log_{10}$-scaled axis are also plotted for *phylocurve* functions.

Figure 3.2. Speed comparisons for rate comparisons among trait groups using a) the distance-based geomorph function *compare.multi.evol.rates* and b) the covariance-based phylocurve function *fast.geomorph.compare.multi.evol.rates*; tests of K_{mult} using c) the distance-based *geomorph* function *physignal* and d) the covariance-based *phylocurve* function *fast.geomorph.physignal*; multivariate phylogenetic regression using e) the distance-based *geomorph* function *procD.pgls* and f) the covariance-based *phylocurve* function *fast.geomorph.procD.pgls*; and phylogenetic partial least squares using h) the distance-based *geomorph* function *phylo.pls* and i) the covariance-based *phylocurve* function *fast.geomorph.phylo.pls*. Speed comparisons were performed on randomly generated pure-birth phylogenies containing 10, 24, 50, 100, 250, 500, and 1,000 species with simulated data containing 10, 24, 50, 100, and 250 traits. Computation times on a $\log_{10}$-scaled axis are also plotted for *phylocurve* functions.

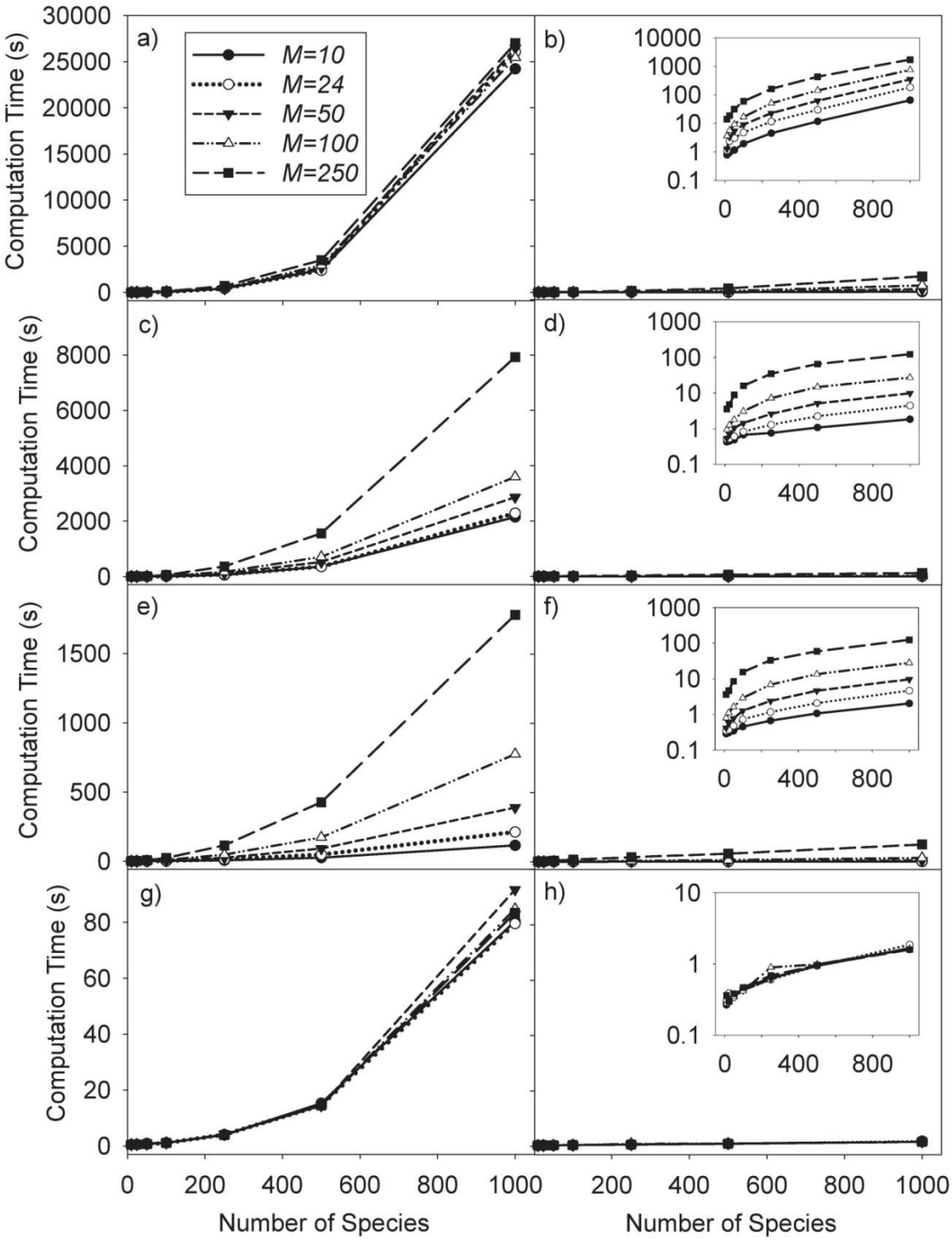

a)
b)
c)
d)
e)
f)
g)
h)
M=10
M=24
M=50
M=100
M=250
Computation Time (s)
Number of Species
0
200
400
600
800
1000
5000
10000
15000
20000
25000
30000
2000
4000
6000
8000
500
1500
20
40
60
80
0.1
1
10
100
1000

Figure 3.3. Statistical performance of evolutionary rate comparisons among regimes using a) distance-based hypothesis tests assuming a diagonal evolutionary rate matrix as was assumed in Adams (2014c) and implemented in the geomorph function compare.evol.rates in versions 1.1.5-2.1.5 (performed using the fast.geomorph.compare.evol.rates function in phylocurve with force.diag=TRUE), b) distance-based hypothesis tests assuming a fully parameterized evolutionary rate matrix is implemented in geomorph versions 2.1.6 and higher (performed using the fast.geomorph.compare.evol.rates function in phylocurve with force.diag=FALSE), c) distance-based hypothesis tests using a fully parameterized evolutionary rate matrix with regime-specific rates using the censored method of O'Meara et al. (2006) (performed using the fast.geomorph.compare.evol.rates function in phylocurve with censored=TRUE), and d) pairwise composite likelihood methods performed using the compare.models function in phylocurve. Failing to account for trait covariance (a) results in unacceptably high Type I error. Simulations were performed on 1,000 randomly generated 32-species pure-birth phylogenies with 2, 16, 32, and 64 traits simulated under two different regimes of equal size with a rate ratio of 1.0 (for Type I error), 1.5, 2.0, 3.0, and 4.0.

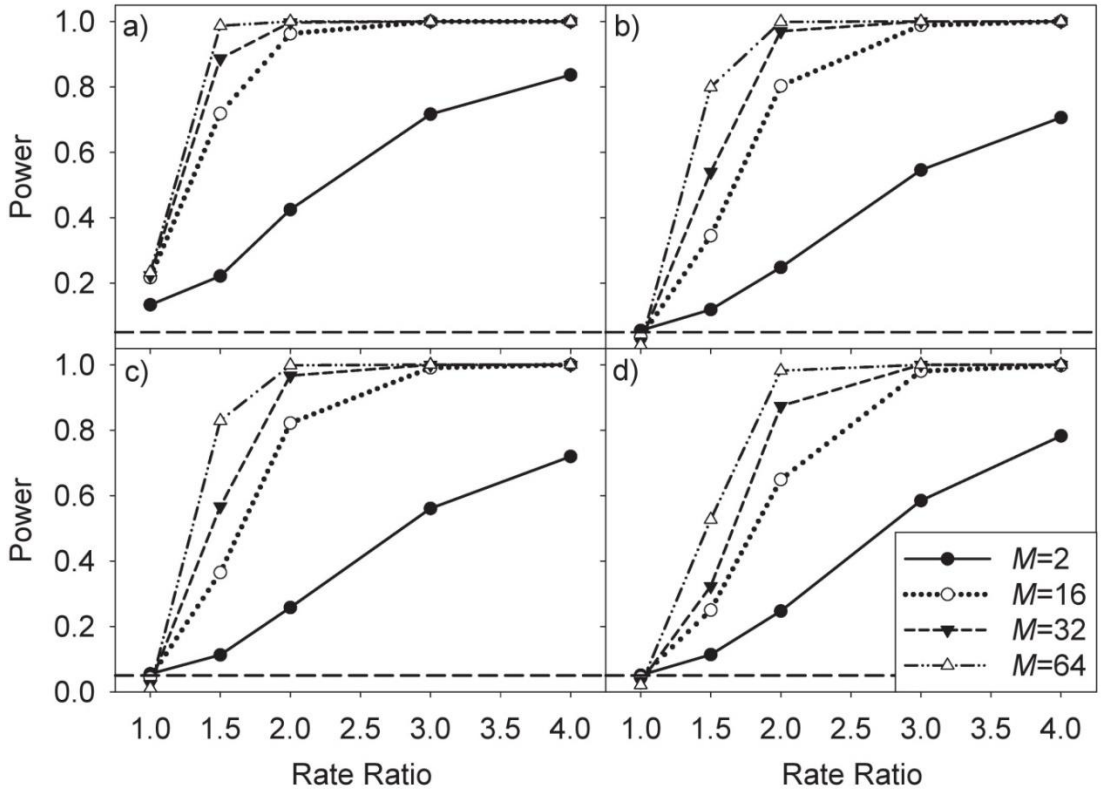

a)
b)
c)
d)
Power
Power
1.0
0.8
0.6
0.4
0.2
0.0
1.0 1.5 2.0 2.5 3.0 3.5 4.0
Rate Ratio
Rate Ratio
M=2
M=16
M=32
M=64

Figure 3.4. Comparison of rate differences among two regimes from nine *Plethodon* salamanders for 11 cranial landmark coordinates (Adams 2014c) using a) distance-based hypothesis testing while failing to account for trait covariance, as in Adams (2014c) (performed using the *phylocurve* function *fast.geomorph.compare.evol.rates* with *force.diag=FALSE*), b) distance-based hypothesis testing while accounting for trait covariance (performed using the *phylocurve* function *fast.geomorph.compare.evol.rates* with *force.diag=FALSE*), and c) pairwise composite likelihood-based hypothesis testing and statistical power assessment (performed the *phylocurve* function *compare.models* on null and alternative models both fit using the *evo.model* function), where the left distribution is the null distribution (no difference in rate ratio) and the right distribution is the alternative distribution (significant difference in rate ratio); the left vertical dotted line is the observed test statistic and the right vertical line is the critical value for the test statistic (indicating that the observed test statistic falls within the null distribution). While failing to account for trait covariance, the rate ratio between the two hypothesized regimes (1.84) was found to be significantly different from 1.0 (P=0.005), identical to the result reported in Adams (2014c). While accounting for trait covariance, the rate ratio (1.84) was found to be non-significant (P=0.261). Pairwise composite likelihood-based hypothesis testing was also non-significant (P=0.64) and revealed very little statistical power (0.097, the proportion of simulated alternative models greater than the critical value) to detect a significant rate difference for the given the phylogeny and dataset.

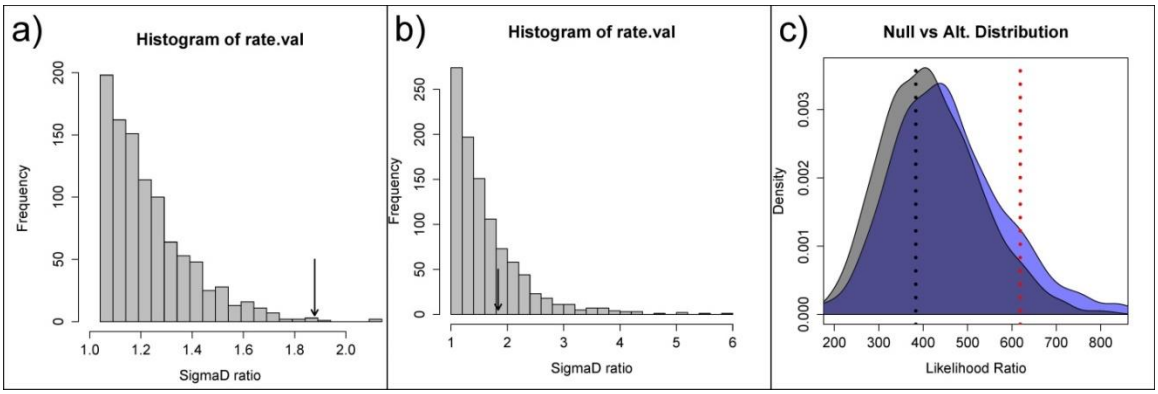

a)
Histogram of rate.val
Frequency
SigmaD ratio
b)
Histogram of rate.val
Frequency
SigmaD ratio
c)
Null vs Alt. Distribution
Density
Likelihood Ratio

Figure 3.5. Comparisons of the statistical performance of distance-based methods (left) and pairwise composite likelihood methods (right) for comparisons of evolutionary rates among trait groups (a-b, using the *geomorph* function *compare.multi.evol.rates* and the *phylocurve* function *compare.models*), tests of significant K_{mult} (c-d, using the *geomorph* function *physignal* and the *phylocurve* function *K.mult*), tests for significant covariation with fixed effects (e-f, using the *geomorph* function *procD.pgls* and the *phylocurve* function *compare.models*), and tests for significant covariation among two multivariate traits (g-h, using the *geomorph* function *phylo.pls* and the *phylocurve* function *compare.models*). Comparisons of evolutionary rates (a-b) among trait groups were simulated with rate ratios 1.0 (for Type I error), 1.5, 2.0, 3.0, and 4.0 for simulated data with 4, 8, 16, 32, and 64 traits (where subset trait groups were of size 2, 4, 8, 16, and 32, respectively). Both methods exhibited similar statistical performance. Tests of significant phylogenetic signal (c-d) were performed using Pagel's λ tree transformations of 1.0 (for Type I error), 0.75, 0.5, 0.25, and 0.0 on simulated data with 2, 5, 10, 32, and 64 traits. Statistical performance was similar for both methods, although distance-based tests of phylogenetic signal (c) had higher statistical power at higher simulated levels of Pagel's λ. Multivariate phylogenetic regression (e-f) was simulated with an input covariation among simulated data and fixed effects of 0.0 (for Type I error), 0.1, 0.3, 0.5, 0.7, and 0.9 for data of trait dimension 10, 24, 50, 100, and 250. Distance-based methods (e) exhibited elevated Type I error. Tests for significant covariation among multivariate traits (g-h) were simulated using an input covariation of 0.0 (for Type I error), 0.1, 0.3, 0.5, 0.7, and 0.9 for traits of dimension 2, 4, 8, 16, 32, and 64. Distance-based phylogenetic partial least squares (g) exhibited extremely high Type I error. All simulations were performed

on 1,000 randomly generated 32-species pure-birth phylogenies.

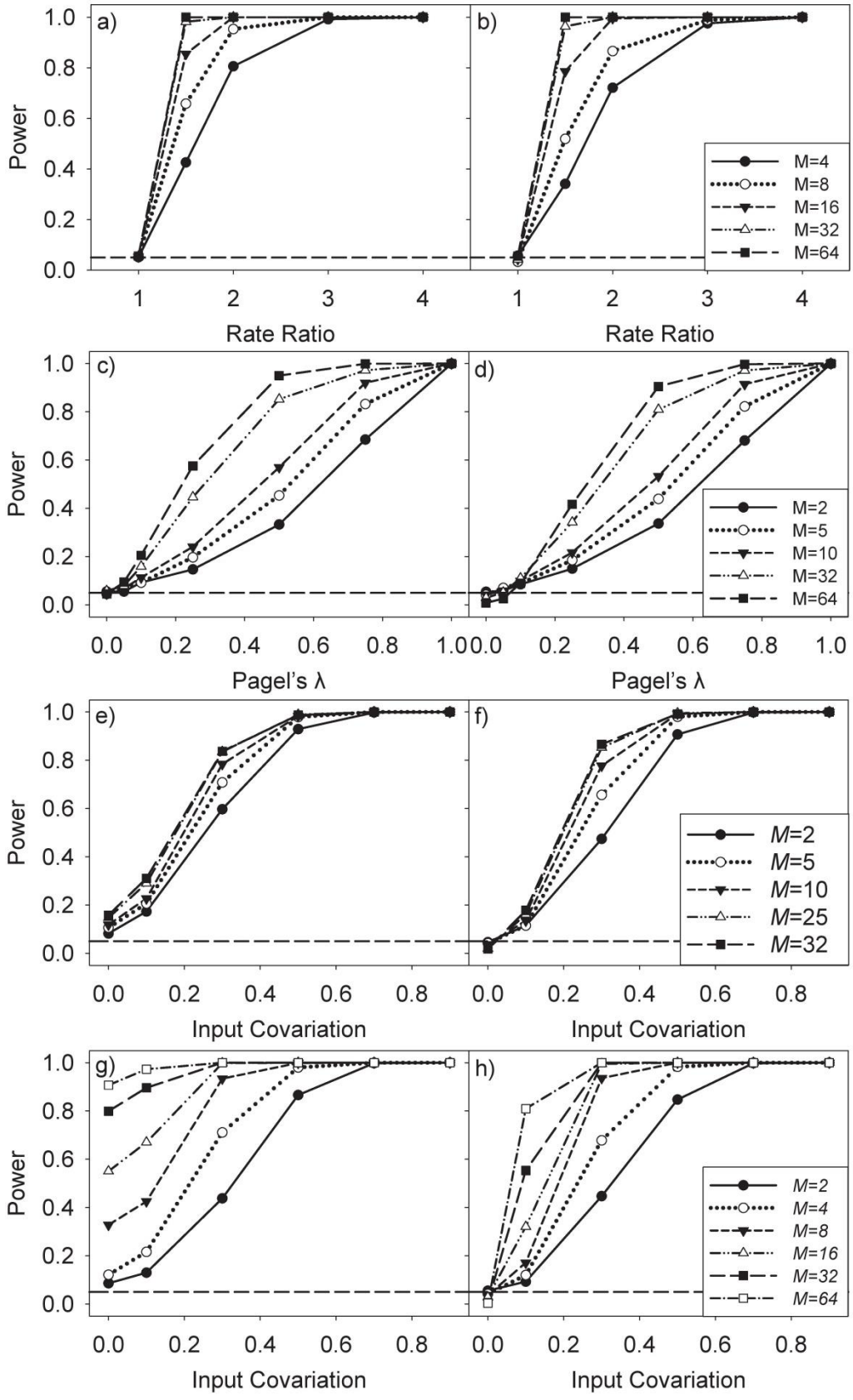
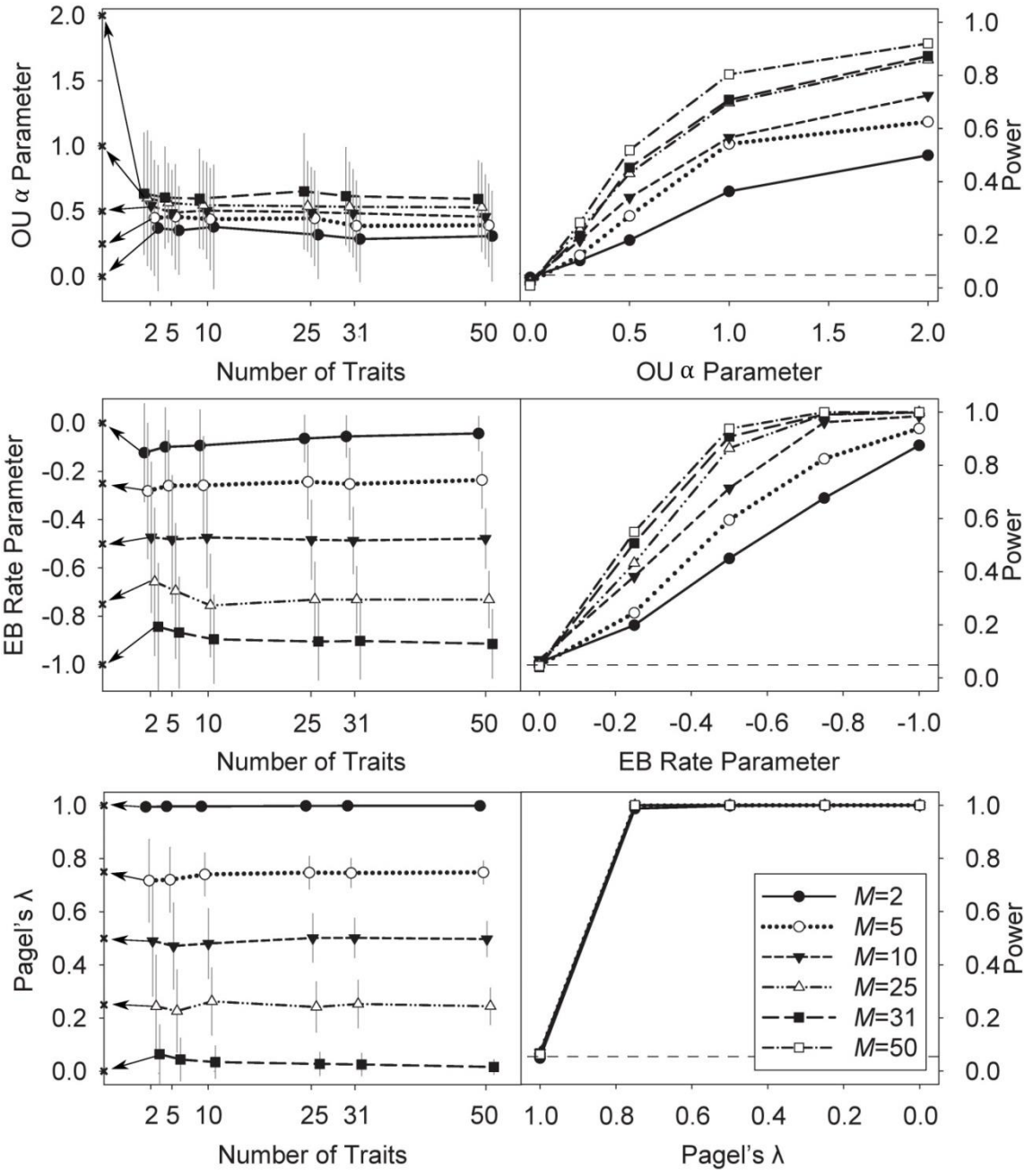

a)
b)
c)
d)
e)
f)
g)
h)
Power
Rate Ratio
Pagel's λ
Input Covariation
M=4
M=8
M=16
M=32
M=64
M=2
M=5
M=10
M=25
M=32
M=2
M=4
M=8
M=16
M=32
M=64

Figure 3.6. Results of maximum pairwise composite likelihood tree transformation parameters for the Ornstein-Uhlenbeck (a-b), Early-Burst (c-d), and Pagel's lambda (e-f) models. Simulations were performed on 1,000 randomly generated pure-birth phylogenies using randomly generated data of trait dimension 2, 5, 10, 25, 31, and 50. Simulated tree transformation parameters were as follows: Ornstein-Uhlenbeck α=0.0 (for Type I error), 0.25, 0.5, 1.0, and 2.0; Early-Burst rate=0.0 (for Type I error), -0.25, -0.5, -0.75, and -1.0; Pagel's λ=1.0 (for Type I error), 0.75, 0.5, 0.25, and 0.0. Left panels show mean parameter estimates of 1,000 simulations (grey bars represent standard deviation) corresponding to the tree transformation parameter indicated by arrows. Right panels indicate Type I error and statistical power under the range of simulated conditions.


OU α Parameter
Number of Traits
OU α Parameter
Power
EB Rate Parameter
Number of Traits
EB Rate Parameter
Power
Pagel's λ
Number of Traits
Pagel's λ
Power
M=2
M=5
M=10
M=25
M=31
M=50

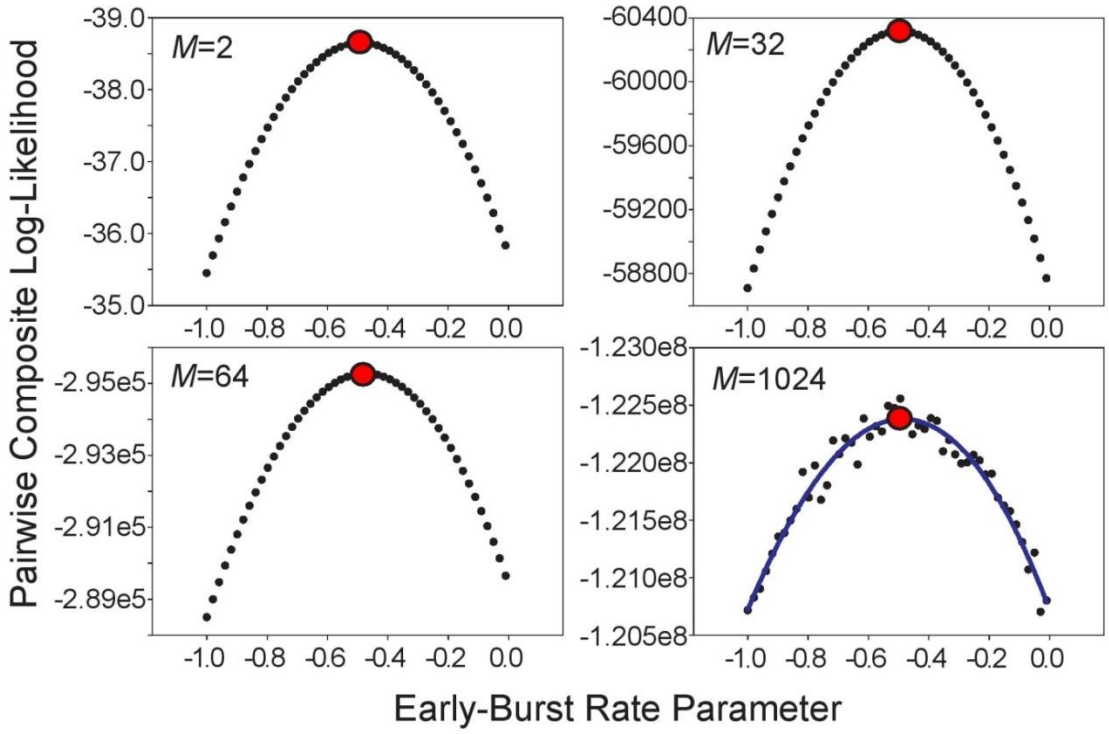


Figure 3.7. The pairwise composite log-likelihood surface for the Early-Burst rate transformation for trait data of dimension 2, 32, 64, and 1,024, which were simulated on a randomly generated 32-species phylogeny under an Early-Burst model of phenotypic evolution with a rate parameter of -0.5. For the 1,024-trait dataset, a Monte Carlo approach was performed using 10,000 random pairwise combinations to approximate the pairwise composite log-likelihood surface, and polynomial regression was performed to obtain a smoothed estimate of the likelihood surface. Regardless of the number of traits, the simulated Early-Burst rate of -0.5 was strongly identifiable and was recovered via maximum pairwise composite likelihood estimation.

CHAPTER 4

EVOLUTION OF METAL ACCUMULATION ACROSS WILD SUNFLOWERS AND IMPLICATIONS FOR ELEMENTAL DEFENSE [1]

[1]**Goolsby EW, Humphreys DP, Baxter MR, Shefferson RP, Donovan LA.** To be submitted to *Planta*.

ABSTRACT

Several adaptive hypotheses concerning the evolution of heavy metal hyperaccumulation have been proposed. One of the most popular hypotheses, the elemental defense hypothesis, predicts a protective effect against herbivores of elevated leaf metal concentrations. Here we assess the evolution of nickel and cadmium accumulation in 15 wild sunflower species (*Helianthus*). We find support for an evolutionary origin of elevated nickel and cadmium accumulation ancestral to *H. annuus* (the wild progenitor of domesticated sunflower, which is a nickel and cadmium hyperaccumulator), *H. exilis* (serpentine sunflower, which is known to be tolerant of extremely high soil nickel concentrations), and *H. argophyllus*. We also performed non-choice and choice feeding trials with larvae of the generalist herbivore *Vanessa cardui* (the painted lady butterfly). Using phylogenetic regression, we find no support for any pattern between leaf metal concentration and non-choice herbivory rates among species. However, within three focal sunflower species we find mixed results, with variable feeding deterrence and feeding stimulation in specific metal treatments and species.

INTRODUCTION

Hyperaccumulators are plants capable of accumulating extremely high foliar metal concentrations (>1,000 mg/kg dry leaf weight for most metals) (van der Ent et al. 2013). Over 75% of known hyperaccumulators (currently exceeding 500 taxa) are nickel hyperaccumulators, although some plants hyperaccumulate other elements, including arsenic, cadmium, chromium, lead, selenium, zinc, and manganese (Boyd 2012; Cappa and Pilon-Smits 2014). Recent evidence suggests that hyperaccumulation has evolved many times independently (Broadley et al. 2001; Cappa and Pilon-Smits 2014; Kraemer 2010; Moray et al. 2015). For example, there are likely at least twelve independent origins of this trait in the Brassicaceae family alone (Kraemer 2010). Because hyperaccumulation has repeatedly evolved across vascular plants, the evolutionary dynamics of metal hyperaccumulation is an area of strong interest in plant evolutionary ecology. Several adaptive benefits of hyperaccumulation have been proposed to explain why hyperaccumulation has arisen and persisted, including improved resistance to herbivores, pathogens, or drought, and decreased competition via metal-enrichment of soil from leaf litter (elemental allelopathy) (Cappa and Pilon-Smits 2014; Rascio and Navari-Izzo 2011). Hyperaccumulation also has many promising applications for humans, including environmental cleanup (phytoremediation), enhanced ability to extract precious elements from soils (phytomining), and targeted elemental enrichment of crops in malnourished regions (Rascio and Navari-Izzo 2011). However, at present, the underlying mechanisms and evolutionary dynamics of hyperaccumulation are poorly understood, and an improved understanding of these issues is necessary to maximize

benefits while minimizing the unintentional consequences of the applied use of hyperaccumulators.

Several hypotheses regarding the adaptive benefits of metal hyperaccumulation have been proposed. Of these, the elemental defense hypothesis, which postulates that high foliar metal concentrations help deter herbivores and inhibit microbial growth, has received the most attention and empirical support (Boyd 2007, 2012; Rascio and Navari-Izzo 2011). The elemental defense hypothesis can be extended into two secondary hypotheses: the defense trade-off hypothesis and the joint effects hypothesis. The defense trade-off hypothesis predicts a negative relationship between leaf metal concentration and both physical (e.g.. leaf toughness, trichomes) and chemical (e.g., tannins, glucosinolates) plant defenses (Boyd 2012; Moles et al. 2013). Because the use of each defensive strategy is putatively energetically costly (costs of metal uptake and tolerance, physical structure investment, and chemical biosynthesis, respectively) with potentially diminishing marginal returns, a high level of one defense mechanism should lead to reduced expression of other defenses. The joint effects hypothesis predicts that metal accumulation and chemical or physical defenses can work together in an additive or synergistic manner to prevent herbivory or pathogen infection (Boyd 2012). If various defenses provide additive or synergistic plant protection, each individual defense strategy is expected to be effective at lower levels, thus requiring a lower overall investment in each in order to maintain plant protection. The elemental allelopathy hypothesis predicts that plants concentrate high levels of metals into senescing leaves as a means of increasing soil metal concentrations, thereby making it less likely for less tolerant competitors to establish nearby. This hypothesis has received preliminary support in a

small number of systems (Cappa and Pilon-Smits 2014; Morris et al. 2009). The drought resistance hypothesis predicts that metal hyperaccumulation confers increased drought resistance, either by direct osmotic interactions or by priming via induction of plant stress response pathways. This hypothesis, the only one to focus on response to abiotic rather than biotic factors, has received some experimental support (Cappa and Pilon-Smits 2014; Rajkumar et al. 2013).

Although each of these adaptive hypotheses regarding metal hyperaccumulation evolution is accompanied by at least some level of empirical support, the overall evidence is mixed and sometimes contradictory. These data are also limited to a very small number of disparate taxa. Several recent reviews on hyperaccumulation have recommended a phylogenetically explicit approach to uncovering ecological dynamics of metal hyperaccumulation in an evolutionary framework (Cappa and Pilon-Smits 2014; Pollard et al. 2014; van der Ent et al. 2013). Phylogenetically explicit studies will help us to answer many important questions, particularly because the independent evolutionary origins of hyperaccumulation in multiple clades offer numerous opportunities for replicated assessment of its evolution. However, to our knowledge, no published study to date has yet examined adaptive hypotheses for metal hyperaccumulation in a phylogenetic context. Here, we explore the evolutionary history of nickel and cadmium hyperaccumulation in a clade of twelve annual *Helianthus* species and subspecies and two perennial outgroup *Helianthus* species. We also examine the elemental defense hypothesis using the painted lady butterfly *Vanessa cardui*, an Asteraceae specialist – a generalist within the family and known to feed on *Helianthus* – using phylogenetic comparative and experimental approaches.

MATERIALS AND METHODS

Study System. Domesticated sunflower (*Helianthus annuus*) is uniquely capable of hyperaccumulating at least four metals, including As, Cd, Cr, and Ni, and accumulating extraordinarily high levels of Cu, Fe, Zn, Mn, Se, U, Cs, Hg, Pb, and Sr (Blamey et al. 1986; Cutright et al. 2010; Lin et al. 2003; Prasad and Freitas 2003; Solhi et al. 2005; Walliwalagedara et al. 2010). Hyperaccumulation of each of these metals in sunflowers likely involves distinct evolutionary histories, genetic bases, and physiological mechanisms. *Helianthus* is an extremely diverse genus of plants, possessing substantial variation in many ecologically important traits, including climatic niche, abiotic stress tolerance, pathogen resistance, and growth rate (Donovan et al. 2014; Heiser et al. 1969). Additionally, wild *Helianthus* occur across North America in a wide variety of habitats, including serpentine soils, beaches, coastal prairies, deserts, salt marshes, mountain meadows, and open woodlands, and many species are adapted to major environmental stressors, such as drought, salinity, and low nutrient availability (Heiser et al. 1969). Furthermore, a recently constructed 170-gene phylogeny (Stephens et al. 2015) allows for the robust assessment of sunflower trait evolution using phylogenetic comparative methods (Mason et al. 2015; Mason and Donovan 2015).

Taxa and seed sources. This study focuses on the large annual clade of wild sunflowers, which includes the wild progenitor of domesticated sunflower and its diverse relatives (Heiser et al. 1969; Stephens et al. 2015; Timme et al. 2007). Twelve annual taxa were included: *H. annuus*, *H. argophyllus*, *H. debilis ssp. debilis*, *H. debilis ssp. silvestris*, *H. debilis ssp. tardiflorus*, *H. exilis*, *H. neglectus*, *H. niveus ssp. canescens*, *H.*

petiolaris, *H. praecox ssp. hirtus*, *H. praecox ssp. praecox*, and *H. praecox ssp. runyonii*. Two distantly-related perennial species from outside the annual clade (*H. angustifolius* and *H. cusickii*) were selected to serve as outgroups. Because *H. annuus* is the only confirmed hyperaccumulator species in the genus, several populations of *H. annuus* and its two sister species (*H. argophyllus* and *H. exilis*) were included in the study to capture the variation of metal accumulation within these closely-related species. Specifically, ten populations of *H. annuus*, twelve populations of *H. exilis*, and nine populations of *H. argophyllus* were chosen from across the range of each species. For all other species in the study, two to three populations or subspecies were selected, for a grand total of 51 populations. Seeds for each population were obtained from the USDA National Genetic Resources Program (Appendix B).

Plant growth. Seed germination began in May 2013. Achenes for each species were scarified, placed on moistened filter paper in Petri dishes, and stored in darkness for 24 hours at room temperature. Next, seed coats were removed, and seeds were transferred to new Petri dishes with moistened filter paper. Petri dishes were placed under fluorescent lights set to a twelve-hour photoperiod, until root and cotyledon development. Germination time ranged from three days to one week, depending on species. Germinated seedlings were transferred to the greenhouse (Department of Plant Biology, University of Georgia, Athens, GA) in seedling trays filled with Fafard 3B mix (Conrad Fafard, Aawam, MA, USA) under ambient light with temperature set to 27°C during the day and 20°C at night. Seedlings were grown in seedling trays for two weeks and then transferred to pots. Pots were arranged in a randomized split-plot design, consisting of three treatment types: control (no metals added to soil), nickel-amended, and cadmium-

amended. Each treatment was replicated four times, randomized across the greenhouse. Treatment plots consisted of plastic-lined reservoirs, each containing 51 pots (one for each represented *Helianthus* population) filled with Fafard 3B soil mix (Conrad Fafard, Agawam, MA, USA). Treatment plots were filled with 70 liters of water (approximately two inches of water), which was maintained throughout the experiment, providing water to the pots from beneath to ensure high water availability. Non-control reservoirs were amended with either nickel sulfate or cadmium sulfate to obtain elemental metal concentrations of 30 mg/L. Osmocote Plus 15-9-12 nine month slow-release fertilizer with micronutrients (Scotts, Marysville, OH, USA) was applied to the surface of each pot to ensure high nutrient availability.

Leaf measurements. Plants were harvested at a standardized juvenile ontogenetic stage, defined by 4-8 fully expanded leaf pairs (Mason et al. 2013). For each plant, the 3 most recently fully expanded leaf pairs were harvested immediately prior to sunrise: one leaf was used for leaf trait measurements, one for non-choice feeding trials, one (or two for control plants) for choice feeding trials (restricted to *H. annuus*, *H. argophyllus*, and *H. exilis*), and the remaining leaves were harvested and pooled for leaf metal quantification. Leaf traits measured included fresh and dry leaf mass, leaf chlorophyll content (SPAD-502 meter; Konica Minolta, Tokyo, Japan), and leaf trichome density (a putative anti-herbivore defense, as the number of trichomes per square centimeter, counted under a dissecting microscope). Fresh and dry masses for each leaf were used to calculate leaf water content (a measure of leaf succulence, defined as the ratio of fresh leaf water mass to dry mass). Fresh leaf area was quantified from scanned leaf images using ImageJ freeware (Schneider et al. 2012) from which leaf dry mass per unit fresh

area (LMA) was calculated. Leaves intended for herbivory experiments were placed in sealed plastic bags and refrigerated for approximately six hours, until feeding trials began. The remaining pooled leaves for metal analysis were dried in a forced-air drying oven and ground into a fine powder. Ground leaf tissue for each plant was weighed to 0.02g, dry-ashed at 450°C overnight, digested in aqua regia (20% nitric acid, 5% hydrochloric acid), and analyzed for leaf nickel and cadmium concentration via graphite furnace atomic absorption spectroscopy (Shimadzu AA6800 with GFA6500).

Caterpillar rearing. For herbivory experiments, *Vanessa cardui* eggs were obtained from Carolina Biological Supply Company (Burlington, NC, USA). *V. cardui* eggs were placed on lettuce leaves (*Lactuca sativa*) in 1200 cm^3 polyethylene trial chambers covered with non-airtight lids. Fluorescent lighting was supplied on a 16-hour photoperiod with a constant temperature of 23°C. Larvae emerged within 48 hours, and fresh lettuce leaves were continuously provided. Seven days after hatching, larvae were randomly selected for choice and non-choice feeding trials.

Herbivory trials. Feeding trials were conducted with harvested leaves and *V. cardui* larvae. For non-choice feeding trials, a single fresh leaf (one from each plant in the greenhouse experiment) was weighed and placed in a trial chamber with a single caterpillar. For choice feeding trials, a single control leaf and a single treatment leaf (either nickel or cadmium) from plants of the same population were weighed and placed in a trial chamber with a single caterpillar. Chambers were maintained under the same environmental conditions as caterpillar rearing. After one week, leaves were collected from feeding trial boxes and visually scored for percent area consumed (0, 25, 50, 75, or

100%). Fresh leaf mass consumed was estimated by multiplying fresh leaf mass and percent area consumed.

Data analysis. To evaluate the evolutionary history of metal accumulation in annual sunflowers, maximum likelihood ancestral state reconstruction was performed on average leaf metal concentration using the *fastAnc* function in the R package *phytools*, and phylogenetic signal (as Blomberg's *K*) was assessed using the *phytools* function *phylosig* (Blomberg et al. 2003; Revell 2012). Due to insufficient leaf mass, leaf cadmium concentrations could not be obtained for *H. praecox ssp. runyonii*, so an estimate for cadmium concentration was imputed for this taxon using the estimated phenotypic value for the common ancestor of *H. praecox ssp. runyonii* and *H. praecox ssp. hirtus* (Bruggeman et al. 2009). For non-choice herbivory experiments, leaf traits (leaf metal concentration, LMA, chlorophyll content, water content, and trichome density) were regressed against herbivory rates (fresh leaf mass consumed) using phylogenetic regression (Felsenstein 1985; Martins and Hansen 1997; Paradis et al. 2004).

For choice trials, relative consumption between control and metal-treatment leaves was calculated as treatment leaf consumption divided by total consumption of both leaves, such that a relative consumption less than 0.50 indicates preference for control leaves, a relative consumption greater than 0.50 indicates preference for treatment leaves, and a relative consumption equal to 0.50 indicates no preference for either. The four replicate choice trials for each metal in each population were averaged to calculate population point estimates of relative consumption for that metal treatment. Population point estimates were used to calculate species means of relative consumption in each

metal treatment, as well as the 95% confidence interval around those means based on the number of population point estimates (9-12).

RESULTS

Evolutionary history of metal accumulation. Nickel and cadmium leaf concentrations were, on average, elevated in *H. annuus*, *H. argophyllus*, and *H. exilis* relative to the other species in the study (Fig. 4.1). Nickel accumulation exhibited significant phylogenetic signal (Blomberg's K=1.27; p=0.02), and ancestral state reconstruction of nickel accumulation revealed a likely origin of elevated nickel accumulation ability in the common ancestor of *H. annuus*, *H. argophyllus*, and *H. exilis* (Fig. 4.1). The capacity for elevated cadmium accumulation also appears to have evolved in the common ancestor of *H. argophyllus*, *H. annuus*, and *H. exilis*. However, the capacity for cadmium accumulation is more evolutionarily labile (Blomberg's K=0.69; p=0.54) than nickel accumulation, as elevated cadmium accumulation appears to have also evolved independently in *H. debilis*, *H. neglectus*, and *H. niveus* (Fig. 4.1).

Analysis of herbivory experiments. In non-choice experiments, caterpillars were presented with a single leaf from control, nickel, or cadmium treatments. Phylogenetically independent contrasts regressed through the origin revealed no observable effect of leaf metal concentration on non-choice herbivory rates for either nickel-treated (R^2=0.186; P=0.11) or cadmium-treated (R^2=0.012; p=0.70) plants (Felsenstein 1985). Similarly, leaf trichome density (a putative herbivore deterrent) was uncorrelated with herbivory rates for all treatments. Conversely, species-specific herbivory rates in the control treatment strongly predicted herbivory rates in both nickel

(R^2=0.72; P<0.0001) and cadmium (R^2=0.81; P<0.0001) treatments. Leaf ecophysiological traits were also predictive of herbivory rates in both the control and nickel treatments (Table 4.1). Additionally, traits indicative of general plant vigor (height and leaf size) strongly predict herbivory rates in all treatments.

For choice feeding trials (which included *H. annuus*, *H. argophyllus*, and *H. exilis*), caterpillars were presented with a choice between a control leaf and a leaf from either a nickel or cadmium-treated plant. In contrast to non-choice experiments, choice feeding trials exhibited treatment-specific preferences in all three species (Fig. 4.2). Caterpillars preferentially consumed control leaves over cadmium leaves in both *H. annuus* and *H. argophyllus*. Caterpillars also displayed a preference for control *H. argophyllus* leaves over leaves from nickel-treated plants, whereas leaves from nickel-treated *H. exilis* were preferentially consumed over control leaves (Fig. 4.2).

DISCUSSION

Overall, we find mixed support for the elemental defense hypothesis. When faced with no alternatives in non-choice trials, herbivores readily consumed leaves from metal-treated plants without regard to leaf metal concentration. This result, although contrary to expectations under the elemental defense hypothesis, may simply suggest that *V. cardui* larvae would prefer to consume unpalatable or toxic leaves rather than starve. These results also suggest strong inherent species-level properties (e.g., leaf ecophysiological traits or phytochemistry) that may swamp out potential anti-herbivore effects of foliar metals. For instance, leaves with extremely high water content (i.e., succulence) are known to deter herbivory in other systems (Moles et al. 2013; Pérez-Harguindeguy et al.

2003), as was observed here. Negative correlations between herbivory and water content were strongest in the control treatment, whereas the relationships in nickel and cadmium treatments were weaker. Paradoxically, high LMA leaves, which are generally tougher and more physically defended against chewing insects (Moles et al. 2013), correlated positively with herbivory rates, with the correlations strongest in the control treatment. This result could be explained by the strongly negative correlation between leaf water content and LMA (R^2>0.55 in all three treatments), in which leaf succulence might have a stronger anti-herbivore effect than high LMA. Leaf chlorophyll content was also positively correlated with herbivory rates in control and nickel treatments, but this correlation was absent in the cadmium treatment. Although leaf metals did not appear to directly influence herbivory rates, correlations between herbivory rates and leaf ecophysiological traits were weaker in metal treatments relative to the control treatment. This could reflect an indirect effect of metals on leaf physiology, perhaps via induction or suppression of the production of certain secondary plant metabolites, as predicted under the defense trade-off and joint effects hypotheses (Boyd 2012). Future work explicitly considering the role of metals on chemical defense investment in a phylogenetic comparative framework will help to disentangle the evolutionary interplay of the defensive role of plant secondary compounds and leaf metal accumulation. The lack of a correlation between leaf trichome density and herbivory rates is consistent with previous findings in sunflowers, as the role of trichomes in sunflowers may be more tied to regulating leaf water balance, irradiance, and temperature rather than defense against leaf herbivores (Mason et al. 2015).

Herbivory rates also strongly correlated with vigor-related traits, including height and leaf area. If growth trades off with defense investment, this pattern may be driven by faster-growing species investing relatively less in plant defense. For example, across diploid *Helianthus* species, traits conducive to faster growth are negatively correlated with leaf toughness and leaf tannin activity (Mason et al. 2015). Growth may best predict herbivory rates because growth may function as an integrated measure of the degree of investment in total plant defense, as predicted by the growth-differentiation-balance hypothesis and the growth rate hypothesis of plant defense (Stamp 2003).

Despite the lack of evidence for a general deterrent effect of metals in non-choice herbivory trials, certain species-treatment combinations in choice experiments demonstrated some support for the elemental defense hypothesis, in which herbivores preferred control leaves over leaves from metal-treated plants. Similar findings have been found in other systems using Lepidopterans. For instance, elevated cadmium concentrations in *Arabidopsis halleri* reduce herbivory by the Brassicaceae specialist *Pieris rapae* (Plaza et al. 2015), and high leaf nickel content has a deterrent effect on *Evergestis rimosalis* herbivory in *Streptanthus polygaloides* (Jhee et al. 2005). Conversely, the reverse effect has also been observed, in which herbivores may actually prefer metal-rich plants, possibly due to a negative relationship between foliar metal concentration and secondary plant defense investment (Noret et al. 2007). The literature provides mixed evidence for the elemental defense hypothesis, as suggested by a 2009 meta-analysis of choice feeding trials with hyperaccumulator plants (Vesk and Reichman 2009) or potential coevolutionary dynamics (Boyd 2009; Pollard 2000). That analysis found a wide variety of foliar metal effects on herbivory, ranging from strong deterrence

to strong preference, with substantial variation depending on both the plant and herbivore species considered. Mixed support for the elemental defense hypothesis could also be the result of competing evolutionary forces in different environments, as would be expected if metal hyperaccumulation offers multiple distinct adaptive advantages (e.g., enhanced drought resistance, decreased competition, antimicrobial effects, etc.) (Boyd 2012; Cappa and Pilon-Smits 2014; Rascio and Navari-Izzo 2011).

Another possibility is that plant species in this study did not achieve sufficient quantities of metal to result in consistent herbivore deterrence due to limited metal availability in amended soils. Heavy metal hyperaccumulation is an environmentally dependent trait – in the absence of soil metals or under low bioavailability, the physiological capacity for metal uptake is not realized. Similarly, under limiting conditions, hyperaccumulation potential may be masked as metal concentrations are diluted by increasing plant biomass (Goolsby and Mason 2016), and it is possible that strong anti-herbivore effects only manifest in the presence of extremely high foliar metal concentrations. Nevertheless, evidence for the elemental defense hypothesis has been observed across a variety of herbivore and plant species at leaf metal concentrations comparable to those in this study (Coleman et al. 2005; Cheruiyot et al. 2013).

Although no plants in this study reached threshold concentrations to constitute metal hyperaccumulation, several individuals from *H. argophyllus*, *H. annuus*, and *H. exilis* achieved nickel concentrations approaching 100 ppm, 2-3 orders of magnitude higher than concentrations found in most eudicots (Broadley et al. 2001), suggesting these three species may indeed possess the capacity for heavy metal hyperaccumulation under conducive conditions. These results are consistent with previous findings in which

domesticated sunflower was demonstrated to accumulate leaf metal concentrations (including nickel and cadmium) exceeding established hyperaccumulator threshold concentrations (Cutright et al. 2010). Additionally, *H. exilis* (serpentine sunflower) is known to be tolerant of extremely high soil nickel concentrations (Sambatti and Rice 2006), which is also consistent with the finding of evolutionary conservation of nickel-related traits in this clade. Our study cannot definitively rule these species out as metal hyperaccumulators, nor can we rule out a consistent deterrent effect of metals on leaf herbivory. If accumulation and herbivore deterrence only manifest above certain species-specific thresholds of soil metal concentration, this effect will require investigation under multiple soil metal treatments in each species in order to be identified.

Ideally, future investigations of the evolutionary history of heavy metal hyperaccumulation capacity and associated adaptive hypotheses should assess plant metal uptake, tolerance, and adaptive dynamics as a function of soil metal concentration, in which concentrations are applied over a gradient ranging from zero soil metals to excess/non-limiting soil metal concentrations (Goolsby and Mason 2015). Such approaches can be applied at a fine scale within species and broadly across large clades of species.

Moreover, phylogenetic comparative methods incorporating an explicitly function-valued trait model can be used to explicitly test adaptive hypotheses while incorporating the complex evolutionary dynamics of metal accumulation (Goolsby 2015), such as the selective pressures involved in the colonization of serpentine habitats by *H. exilis* (Sambatti and Rice 2006). Such approaches, paired with a more thorough understanding of the biochemical and physiological processes underlying

hyperaccumulation, will provide powerful new insights into the evolutionary dynamics of this fascinating trait.

REFERENCES

Blamey, F. P. C., D. C. Joyce, D. G. Edwards, and C. J. Asher. 1986. Role of trichomes in sunflower tolerance to manganese toxicity. *Plant and Soil* 91:171-180.

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

Boyd, R. S. 2007. The defense hypothesis of elemental hyperaccumulation: status, challenges and new directions. *Plant and Soil* 293:153-176.

Boyd, R. S. 2009. High-nickel insects and nickel hyperaccumulator plants: A review. *Insect Science* 16:19-31.

Boyd, R. S. 2012. Plant defense using toxic inorganic ions: Conceptual models of the defensive enhancement and joint effects hypotheses. *Plant Science* 195:88-95.

Broadley, M. R., N. J. Willey, J. C. Wilkins, A. J. M. Baker, A. Mead, and P. J. White. 2001. Phylogenetic variation in heavy metal accumulation in angiosperms. *New Phytologist* 152:9-27.

Bruggeman, J., J. Heringa, and B. W. Brandt. 2009. PhyloPars: estimation of missing parameter values using phylogeny. *Nucleic Acids Research* 37:W179-W184.

Cappa, J. J., and E. A. H. Pilon-Smits. 2014. Evolutionary aspects of elemental hyperaccumulation. *Planta* 239:267-275.

Cheruiyot, D., R. Boyd, and W. Moar. 2013. Exploring Lower Limits of Plant Elemental Defense by Cobalt, Copper, Nickel, and Zinc. *Journal of Chemical Ecology* 39:666-674.

Coleman, C. M., R. S. Boyd, and M. D. Eubanks. 2005. Extending the elemental defense hypothesis: Dietary metal concentrations below hyperaccumulator levels could harm herbivores. *Journal of Chemical Ecology* 31:1669-1681.

Cutright, T., N. Gunda, and F. Kurt. 2010. Simultaneous Hyperaccumulation of Multiple Heavy Metals by *Helianthus annuus* Grown in a Contaminated Sandy-Loam Soil. *International Journal of Phytoremediation* 12:562-573.

Donovan, L. A., C. M. Mason, A. W. Bowsher, E. W. Goolsby, and C. D. A. Ishibashi. 2014. Ecological and evolutionary lability of plant traits affecting carbon and nutrient cycling. *Journal of Ecology* 102:302-314.

Felsenstein, J. 1985. Phylogenies and the comparative methods. *American Naturalist* 125:1-15.

Goolsby, E. W. 2015. Phylogenetic Comparative Methods for Evaluating the Evolutionary History of Function-Valued Traits. *Systematic Biology* 64:568-578.

Goolsby, E. W., and C. M. Mason. 2015. Toward a more physiologically and evolutionarily relevant definition of metal hyperaccumulation in plants. *Frontiers in Plant Science* 6:33.

Goolsby, E. W., and C. M. Mason. 2016. Response: "Commentary: Toward a more physiologically and evolutionarily relevant definition of metal hyperaccumulation in plants". *Frontiers in Plant Science* 6:1252.

Heiser, C. B. J., D. M. Smith, S. B. Clevenger, and W. C. J. Martin. 1969. *The North American Sunflowers-D Helianthus-D*. Memoirs of the Torrey Botanical Club 22:1-218.

Jhee, E. M., R. S. Boyd, and M. D. Eubanks. 2005. Nickel hyperaccumulation as an elemental defense of Streptanthus polygaloides (Brassicaceae): influence of herbivore feeding mode. *New Phytologist* 168:331-344.

Kraemer, U. 2010. Metal Hyperaccumulation in Plants. *Annual Review of Plant Biology*, Vol 61 61:517-534.

Lin, J., W. Jiang, and D. Liu. 2003. Accumulation of copper by roots, hypocotyls, cotyledons and leaves of sunflower (Helianthus annuus L.). Bior*esource Technology* 86:151-155.

Martins, E. P., and T. F. Hansen. 1997. Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist* 149:646-667.

Mason, C. M., A. W. Bowsher, B. L. Crowell, R. M. Celoy, C.-J. Tsai, and L. A. Donovan. 2015. Macroevolution of leaf defenses and secondary metabolites across the genus Helianthus. *New Phytologist*:209:1720-1733.

Mason, C. M., and L. A. Donovan. 2015. Evolution of the leaf economics spectrum in herbs: Evidence from environmental divergences in leaf physiology across Helianthus (Asteraceae). *Evolution* 69:2705-2720.

Mason, C. M., S. E. McGaughey, and L. A. Donovan. 2013. Ontogeny strongly and differentially alters leaf economic and other key traits in three diverse Helianthus species. *Journal of Experimental Botany* 64:4089-4099.

Moles, A. T., B. Peco, I. R. Wallis, W. J. Foley, A. G. B. Poore, E. W. Seabloom, P. A. Vesk, A. J. Bisigato, L. Cella-Pizarro, C. J. Clark, P. S. Cohen, W. K. Cornwell, W. Edwards, R. Ejrnaes, T. Gonzales-Ojeda, B. J. Graae, G. Hay, F. C. Lumbwe,

B. Magana-Rodriguez, B. D. Moore, P. L. Peri, J. R. Poulsen, J. C. Stegen, R. Veldtman, H. Zeipel, N. R. Andrew, S. L. Boulter, E. T. Borer, J. H. C. Cornelissen, A. G. Farji-Brener, J. L. DeGabriel, E. Jurado, L. A. Kyhn, B. Low, C. P. H. Mulder, K. Reardon-Smith, J. Rodriguez-Velazquez, A. De Fortier, Z. Zheng, P. G. Blendinger, B. J. Enquist, J. M. Facelli, T. Knight, J. D. Majer, M. Martinez-Ramos, P. McQuillan, and F. K. C. Hui. 2013. Correlations between physical and chemical defences in plants: tradeoffs, syndromes, or just many different ways to skin a herbivorous cat? *New Phytologist* 198:252-263.

Moray, C., E. Goolsby, and L. Bromham. 2015. The Phylogenetic Association Between Salt Tolerance and Heavy Metal Hyperaccumulation in Angiosperms. *Evolutionary Biology*:1-12.

Morris, C., P. R. Grossl, and C. A. Call. 2009. Elemental allelopathy: processes, progress, and pitfalls. *Plant Ecology* 202:1-11.

Noret, N., P. Meerts, M. Vanhaelen, A. Dos Santos, and J. Escarré. 2007. Do metal-rich plants deter herbivores? A field test of the defence hypothesis. *Oecologia* 152:92-100.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20:289-290.

Pérez-Harguindeguy, N., S. Díaz, F. Vendramini, J. H. C. Cornelissen, D. E. Gurvich, and M. Cabido. 2003. Leaf traits and herbivore selection in the field and in cafeteria experiments. *Australian Ecology* 28:642-650.

Plaza, S., J. Weber, S. Pajonk, J. Thomas, I. N. Talke, M. Schellenberg, S. Pradervand, B. Burla, M. Geisler, E. Martinoia, and U. Kraemer. 2015. Wounding of *Arabidopsis*

halleri leaves enhances cadmium accumulation that acts as a defense against herbivory. *Biometals* 28:521-528.

Pollard, A. J. 2000. Metal hyperaccumulation: a model system for coevolutionary studies. *New Phytologist* 146:179-181.

Pollard, A. J., R. D. Reeves, and A. J. M. Baker. 2014. Facultative hyperaccumulation of heavy metals and metalloids. *Plant Science* 217:8-17.

Prasad, M. N. V., and H. M. D. Freitas. 2003. Metal hyperaccumulation in plants - Biodiversity prospecting for phytoremediation technology. *Electronic Journal of Biotechnology* 6:285-321.

Rajkumar, M., M. N. V. Prasad, S. Swaminathan, and H. Freitas. 2013. Climate change driven plant-metal-microbe interactions. *Environment International* 53:74-86.

Rascio, N., and F. Navari-Izzo. 2011. Heavy metal hyperaccumulating plants: How and why do they do it? And what makes them so interesting? *Plant Science* 180:169-181.

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

Sambatti, J. B. M., and K. J. Rice. 2006. Local adaptation, patterns of selection, and gene flow in the Californian serpentine sunflower (*Helianthus exilis*). *Evolution* 60:696-710.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* 9:671-675.

Solhi, M., H. Shareatmadari, and M. Hajabbasi. 2005. Lead and Zinc Extraction Potential of Two Common Crop Plants, *Helianthus Annuus* and *Brassica Napus*. *Water, Air, and Soil* Pollution 167:59-71.

Stamp, N. 2003. Out of the quagmire of plant defense hypotheses. *Quarterly Review of Biology* 78:23-55.

Stephens, J. D., W. L. Rogers, C. M. Mason, L. A. Donovan, and R. L. Malmberg. 2015. Species tree estimation of diploid Helianthus (Asteraceae) using target enrichment. *American Journal of Botany* 102:910-920.

Timme, R. E., B. B. Simpson, and C. R. Linder. 2007. High-resolution phylogeny for Helianthus (Asteraceae) using the 18s-26s ribosomal DNA external transcribed spacer. *American Journal of Botany* 94:1837-1852.

van der Ent, A., A. J. M. Baker, R. D. Reeves, A. J. pollard, and H. Schat. 2013. Hyperaccumulators of metal and metalloid trace elements: Facts and fiction. *Plant and Soil* 362:319-334.

Vesk, P. A., and S. M. Reichman. 2009. Hyperaccumulators and Herbivores-A Bayesian Meta-Analysis of Feeding Choice Trials. *Journal of Chemical Ecology* 35:289-296.

Walliwalagedara, C., I. Atkinson, H. van Keulen, T. Cutright, and R. Wei. 2010. Differential expression of proteins induced by lead in the Dwarf Sunflower *Helianthus annuus*. *Phytochemistry* 71:1460-1465.

FIGURES

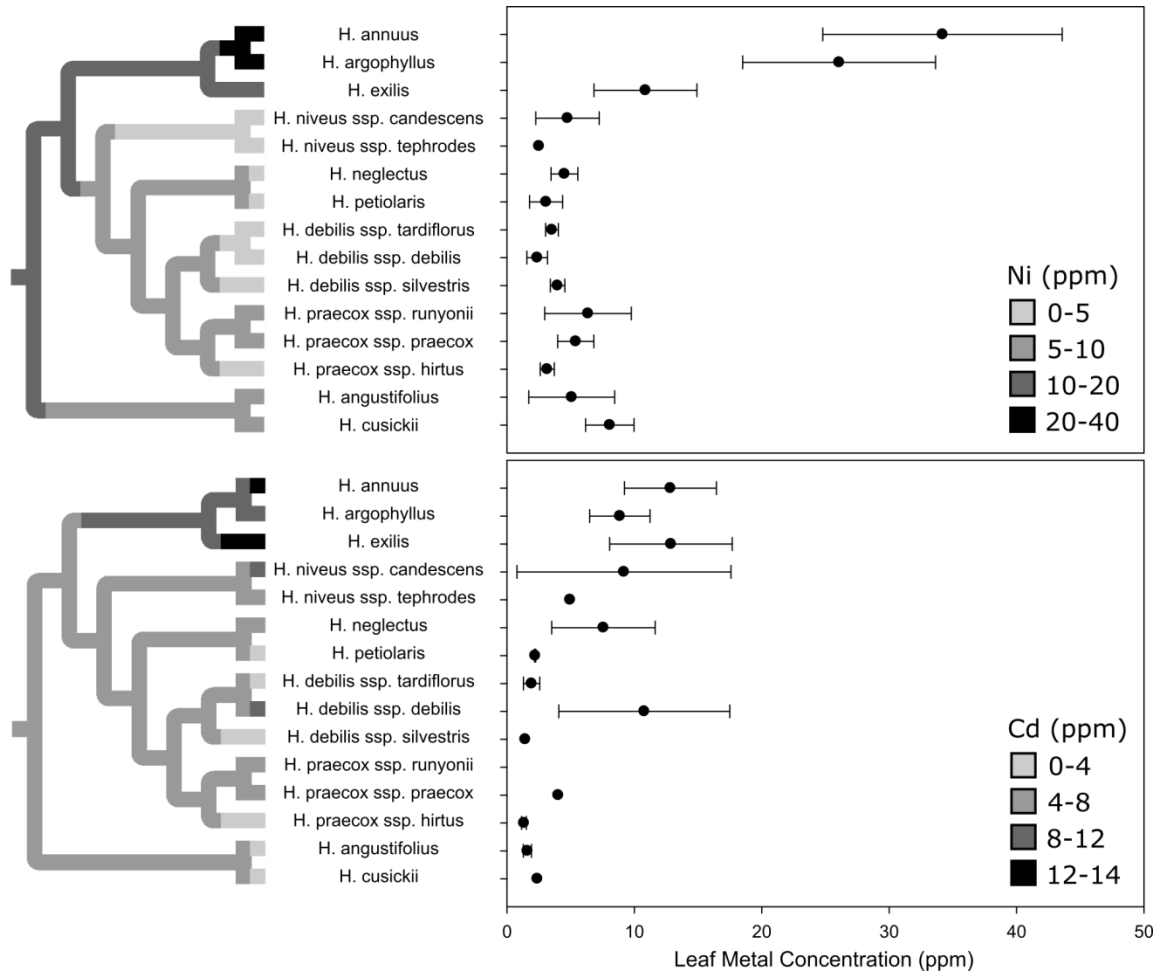


Figure 4.1. Ancestral state reconstructions of nickel (top left) and cadmium (bottom left) accumulation ability and mean metal concentrations (right) observed in each species. Error bars indicate standard errors.

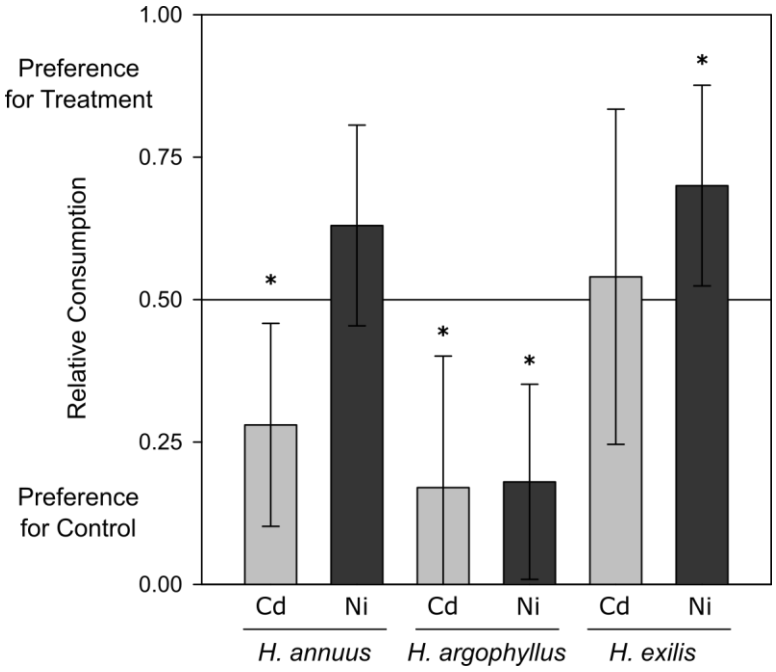


Figure 4.2. Mean relative consumption (treatment leaf consumption divided by total leaf consumption) of fresh leaf mass from choice feeding trials for *H. annuus*, *H. argophyllus*, and *H. exilis*. Asterisks denote groups in which 95% confidence intervals (represented by error bars) do not overlap 0.50, suggesting a significant preference.

TABLES

Table 4.1 Associations between traits and non-choice herbivory (grams of fresh mass consumed) by *Vanessa cardui* in control, nickel, and cadmium treatments. Associations were calculated using by regressing phylogenetically independent contrasts through the origin. R^2 values are presented, with directionality in parentheses. Bold relationships are significant at $p<0.05$.

Trait	*Control*	*Nickel*	*Cadmium*
Plant height (cm)	**(+) 0.68**	**(+) 0.42**	**(+) 0.51**
Leaf area (cm^2)	**(+) 0.86**	**(+) 0.50**	**(+) 0.66**
LMA ($g\ m^{-2}$)	**(+) 0.44**	**(+) 0.28**	(+) 0.22
Leaf chlorophyll content (SPAD)	**(+) 0.44**	**(+) 0.50**	(+) 0.03
Leaf water content ($g\ g^{-1}$)	**(-) 0.40**	(-) 0.26	(-) 0.23
Leaf trichome density (cm^{-2})	(+) 0.02	(+) 0.07	(-) 0.02
Leaf metal Content (ppm)	--	(+) 0.19	(+) 0.01

CHAPTER 5

EVOLUTION OF MULTIPLE METAL HYPERACCUMULATION IN WILD AND CULTIVATED SUNFLOWERS (HELIANTHUS)[1]

[1]**Goolsby EW, Donovan LA, Shefferson RP.** To be submitted to *New Phytologist*.

ABSTRACT

Plants capable of accumulating high concentrations of normally toxic metals are known as hyperaccumulators. The evolution of metal hyperaccumulation is poorly understood, and few studies have investigated the macroevolutionary dynamics of multiple metal hyperaccumulation at fine-scale taxonomic levels. Here we investigate the evolutionary history of metal accumulation for eight metals (As, Cd, Cr, Cu, Ni, Pb, Se, and Zn) in 33 wild sunflower (*Helianthus*) taxa and 12 accessions of cultivated *H. annuus* using a phylogenetic comparative approach in a common garden greenhouse study. We identify widespread Cd, Ni, and Zn hyperaccumulation across the entire *Helianthus* genus, as well as possible hyperaccumulation of Cu in two wild taxa. We also find potential support for correlated evolution of Cd and Zn hyperaccumulation as binary traits. Metal hyperaccumulation in cultivated sunflowers likely evolved in the wild prior to sunflower domestication, and the capacity for elevated Cd, Ni, and Zn accumulation may be ancestral to the entire genus.

INTRODUCTION

Hyperaccumulation is defined as the ability of a plant to accumulate leaf metal concentrations several orders of magnitude higher than what is typically found in most plants (Reeves and Baker 2000). For instance, Cd, which is normally toxic to plants at extremely low concentrations and plays no known physiological role, is said to be at hyperaccumulator concentration if plant leaves exceed 100 mg/kg Cd (mg elemental Cd per kg of dried leaf tissue) (Reeves and Baker 2000, van der Ent et al. 2013). Hyperaccumulator threshold concentrations for other metals, which tend to occur at higher concentrations, are currently set to 1,000 mg/kg for As, Cr, Ni, Pb, and Se, and 1,0000 mg/kg for Zn and Mn (Reeves and Baker 2000, van der Ent et al. 2013). Large-scale phylogenetic investigations have determined the approximate number of known hyperaccumulator species to include at least 500 taxa spanning over 60 families (Jansen et al. 2002, Kraemer 2010, van der Ent et al. 2013, Cappa and Pilon-Smits 2014). The broad distribution of hyperaccumulation across plants suggests that hyperaccumulation likely evolved independently many times (Broadley et al. 2001, Cappa and Pilon-Smits 2014, Pollard et al. 2014). However, the evolutionary history of metal hyperaccumulation at finer-scale taxonomic levels is poorly understood, as it is unknown what adaptive role, if any, hyperaccumulation may play in plant diversification (Boyd 2004, 2007, Rascio and Navari-Izzo 2011, Boyd 2012).

Recent work highlights that plant metal hyperaccumulation and tolerance are likely controlled by separate genetic and physiological processes and may possess distinct evolutionary trajectories (Hanikenne et al. 2008, O Lochlainn et al. 2011, Pollard et al. 2014, Goolsby and Mason 2015, Goolsby and Mason 2016). For example,

phylogenetic comparative investigation of Se tolerance and hyperaccumulation in *Stanleya* (Brassicaceae) revealed that Se tolerance likely evolved prior to the evolution of Se hyperaccumulation within the genus (Cappa et al. 2015). Phylogenetic comparative studies have also revealed that even within relatively narrow taxonomic groups, the capacity for hyperaccumulation in two or more species may be the result of multiple independent evolutionary origins (Cecchi et al. 2010).

Here we investigate the evolutionary history of metal hyperaccumulation in wild and cultivated *Helianthus* using a phylogenetic comparative approach. Cultivated *Helianthus annuus* (sunflower) has been shown to be a hyperaccumulator of As, Cd, Cr, Ni, and possibly other metals (Cutright et al. 2010). However, the evolutionary origins of multiple metal hyperaccumulation in sunflowers are unknown. For instance, it is possible that metal hyperaccumulation was inadvertently selected for during sunflower domestication or improvement, or metal hyperaccumulation may have already been present in wild species prior to domestication. Additionally, most hyperaccumulator species are only hyperaccumulators of a single metal. It is therefore unclear whether multiple metal hyperaccumulation in sunflowers is the result of a single evolutionary event or multiple separate developments.

In this study, we will address the following questions regarding the evolution of As, Cd, Cr, Cu, Ni, Pb, Se, and Zn accumulation in *Helianthus*: 1) What is the capacity for metal accumulation in cultivated *H. annuus*? 2) Is elevated metal accumulation confined to cultivated *H. annuus*, or is there an apparent ancestral origin? 3) In light of multiple metal hyperaccumulation in *H. annuus*, is the capacity for metal hyperaccumulation evolutionarily correlated among metals?

MATERIALS AND METHODS

Study system. Wild *Helianthus* consists of approximately 40 diploid and 10 polyploid species, including several perennials and a clade of annual species that includes *H. annuus* (Heiser et al. 1969, Stephens et al. 2015). We selected 34 wild sunflower taxa from a recent well-supported molecular phylogeny of *Helianthus* (Stephens et al. 2015). A total of 31 species were represented, with three infraspecific taxa of *H. debilis* (*ssp. debilis*, *ssp. silvestris*, and *ssp. tardiflorus*) and two of *H. niveus* (*ssp. canescens* and *ssp. tephrodes*). To capture a snapshot of the variation that exists among cultivated *H. annuus*, we also selected twelve cultivated sunflower (*H. annuus*) accessions which are known as the "core 12" of a sunflower association mapping population (Mandel et al. 2013). Together, these twelve lines represent approximately 50% of the allelic diversity known among cultivated sunflower germplasm (Mandel et al. 2011). Seeds for wild and cultivated accessions were obtained from the USDA National Genetic Resources Program (Appendix C). Plants were grown in a common garden greenhouse experiment at the University of Georgia Department of Plant Biology greenhouses in Athens, GA in the autumn of 2015.

Plant growth. Achenes were scarified and placed on moistened filter paper in Petri dishes and stored in darkness at room temperature for 24 hours. Seed coats were then removed and seeds were transferred to new Petri dishes, placed on greenhouse benches with ambient photoperiod, and kept moist until germination. Upon the production of root hairs, seedlings were transferred to seedling trays filled with

moistened sand. After two weeks, seedlings were transferred to 1.75-liter pots arranged in a randomized split-plot design.

The split-plot design consisted of ten treatments, with each treatment replicated twice. Each treatment consisted of 46 sand-filled pots (one for each taxa) in a plastic-lined reservoir, and each pot was filled with 2 kg of sand (dry mass). Reservoirs were filled with water to a depth of 5 cm, which was maintained throughout the experiment, and each pot was supplied with Osmocote Plus 15-9-12 nine month slow-release fertilizer with micronutrients (Scotts, Marysville, OH, USA). Treatments consisted of eight different soil metal amendments, including amending soils with As (sodium arsenate), Cd (cadmium sulfate), Cr (chromium (III) sulfate), Cu (copper (II) sulfate), Ni (nickel (II) sulfate), Pb (lead (II) nitrate), Se (sodium selenate), and Zn (zinc sulfate), as well as a control (non-metal-amended) treatment. Metal solutions were added to reservoirs and thoroughly mixed so that the ratio of elemental metal mass to dry sand mass was 60 mg/kg (elemental metal mass per kg sand) for As, Cd, Cr, Cu, Ni, Pb, and Se, and 600 mg/kg for Zn. Osmocote application and metal amendments were performed two weeks prior to plant transplants to allow for soil metal and nutrient equilibration.

Plant harvest and leaf metal analysis. Plants were harvested at a standardized pre-reproductive ontogenetic stage (Mason et al. 2013). All members of a given taxa were harvested simultaneously once all surviving plants had produced 4-8 fully-expanded leaf pairs. At harvest, plants were scored for survival and for visible signs of metal toxicity, including leaf chlorosis and leaf deformation. Plants were then completely defoliated, and leaves were dried at 60° C in a forced-air drying oven, and ground into a fine power. Samples were analyzed for metal concentrations at the Soil Testing and Plant

Analysis Laboratory of the Louisiana State University AgCenter (Baton Rouge, LA). Samples were digested in concentrated nitric acid and hydrogen peroxide, and metal concentrations were analyzed via inductively coupled plasma atomic emission spectrometry, and results were validated by standard reference material (National Institute of Standards and Technology, Standard Reference Material 1547, Peach Leaves).

Statistical analysis. Results for leaf metal concentrations were averaged among replicates. Due to metal-induced mortality, leaf metal concentrations could not be measured for certain wild species-treatment combinations. All plants in the Se treatment died shortly after the experiment began due to apparent toxicity, and thus Se was not considered further. For other metals, tissue was unavailable for one wild taxon each of As, Cr, and Ni treatments; four taxa for the Cd treatment; and five taxa for the Zn treatment. Given this missing data, the *Rphylopars* R package was used to perform phylogenetic comparative analyses and estimate trait covariance while accounting for incomplete observations (Bruggeman et al. 2009). Two models of metal accumulation evolution were fit: 1) a null model assuming the evolution of metal accumulation is uncorrelated among metals, and 2) an alternative model allowing for correlated evolution. The two models were compared using a likelihood ratio test, and the explanatory power of the correlated evolutionary model was not significantly better than the uncorrelated null model (D=15.29; df=21; p=1.00). Additionally, the correlated evolutionary model was rejected via AIC model comparison (AIC_{null}=2339.24; AIC_{alt}=2342.20). Therefore, the evolution of metal uptake was assumed to be independent for reconstructing ancestral states of metal accumulation of individual metals.

Ancestral state reconstructions were performed to calculate the evolutionary mean (root state) for accumulation of each metal in wild taxa, and the arithmetic mean was calculated for cultivated accessions. Phylogenetic signal was assessed separately for the accumulation of each metal using Blomberg's *K* using the *phylosig* function in the *phytools* R package (Blomberg et al. 2003, Revell 2012). Blomberg's *K* for each metal was not significantly different from zero (As *K*=0.66, p=0.66; Cd *K*=0.72 p=.64; Ni *K*=0.81, p=0.20; Cr *K*=.96, p=0.05; Pb *K*=0.85, p=0.13 ; Zn *K*=0.96, p=0.06 ; Cu K=0.83, p=0.17).

RESULTS

Elemental concentrations in the leaves of control plants fell within typical ranges for most plants and were as follows: As (0-5.90 mg/kg), Cd (0-2.60 mg/kg), Cr (0-7.22 mg/kg), Cu (1.11-38.01 mg/kg), Ni (1.18-17.76 mg/kg), Pb (0-6.48 mg/kg), and Zn (33.28-347.97 mg/kg) (Broadley et al. 2001, Kraemer 2010, van der Ent et al. 2013). Leaf Cr concentrations were similar among Cr and control treatments, with leaf concentrations in the Cr treatment ranging from 0-7.23 mg/kg in wild taxa and 0.52-8.28 mg/kg in cultivated *H. annuus*. Plants in the Cr treatment exhibited no visible signs of Cr toxicity. A similar result was observed in the Pb treatment, with wild taxa exposed to lead exhibiting leaf Pb concentrations of 0-9.74 mg/kg and cultivated *H. annuus* exhibiting Pb concentrations ranging between 0-2.60 mg/kg. Plants in the Pb treatment also exhibited no visible signs of metal toxicity.

In the As treatment, cultivated *H. annuus* accumulated an average of 36.04 mg/kg As (SE=7.47) in leaves, with no mortality and only one individual exhibiting signs of

metal toxicity (chlorotic leaves) (Fig. 5.1). Wild *H. annuus* accumulated 32.43 mg/kg As, and the sister species *H. argophyllus* accumulated 20.45 mg/kg As. The evolutionary mean of leaf As concentration was 11.94 mg/kg (SE=1.05), and the majority of remaining taxa accumulated less than 10 mg/kg. No signs of As toxicity were observed in wild species.

Leaf Cd concentrations ranged from 140.82-380.15 mg/kg in cultivated *H. annuus* (mean=243.13 mg/kg, SE=17.74), and a similar pattern was seen in the three most closely related wild taxa, *H. annuus* (216.78 mg/kg), *H. argophyllus* (266.95 mg/kg), and *H. exilis* (210.88 mg/kg), with no visible signs of toxicity observed in these taxa (Fig. 5. 1). The majority of wild taxa in the Cd treatment also exceeded the hyperaccumulation threshold of 100 mg/kg Cd, with an evolutionary mean leaf concentration of 209.21 mg/kg (SE=11.14). Mortality of both replicates was observed in four wild taxa: *H. petiolaris*, *H. niveus ssp. canescens*, *H. arizonensis*, and *H. angustifolius*.

The majority of cultivated *H. annuus* accessions exhibited leaf chlorosis and deformation in the Ni treatment, and leaf Ni concentrations ranged between 101.05-438.80 mg/kg (mean=280.40, SE=27.87) (Fig. 5.1). Similar leaf concentrations were observed in *H. annuus* (224.61 mg/kg), *H. argophyllus* (682.68 mg/kg), and both of these species also exhibited leaf chlorosis and deformation. Sister species *H. exilis* also achieved similar leaf concentrations (229.45 mg/kg) but exhibited no visible signs of Ni toxicity. Among wild taxa, Ni concentrations ranged between 89.31-2141.73 (evolutionary mean=911.10 mg/kg, SE=55.92). Six wild taxa exceeded the Ni hyperaccumulator threshold concentration of 1,000 mg/kg: *H. microcephalus*, *H. atrorubens*, *H. mollis*, *H. porteri*, *H. cusickii*, and *H. debilis ssp. silvestris*, the last of

which exhibited signs of Ni toxicity (leaf chlorosis). One of two replicates of *H. neglectus* exhibited leaf deformation, and mortality was observed in both replicates of *H. laciniatus*.

Zn accumulation varied widely among cultivated *H. annuus* (1125.27-53100.85 mg/kg, mean 11201.07 mg/kg, SE=4269.36), and all cultivated accessions exhibited severe Zn toxicity (Fig. 5.1). Similarly, Zn leaf concentrations among wild annual taxa were variable, and all wild annual taxa also exhibited severe leaf deformation and chlorosis. A total of four cultivated and 13 wild accessions achieved leaf Zn concentrations exceeding the hyperaccumulator threshold of 10,000 mg/kg, but only seven of the wild taxa achieved Zn hyperaccumulation without suffering visible signs of Zn toxicity: *H. gracilentus*, *H. occidentalis*, *H. heterophyllus*, *H. silphioides*, *H. atrorubens*, *H. radula*, and *H. microcephalus*.

Leaf Cu concentrations among cultivated *H. annuus* in the Cu treatment were slightly elevated relative to controls (34.07-54.54 mg/kg, mean=44.31, mg/kg SE=2.24), as were concentrations observed in the majority of wild taxa (evolutionary mean=55.19, SE=8.6) (Fig. 5.1). No taxa exceeded threshold concentrations for Cu hyperaccumulation (1.000 mg/kg), and seven wild taxa exhibited leaf chlorosis and deformation.

DISCUSSION

Here we identify several novel potential hyperaccumulator taxa of Cd (22 wild, all 12 cultivated), Ni (6 wild, 0 cultivated), and Zn (13 wild, 4 cultivated). Consistent with Cutright et al. (2010), we report Cd hyperaccumulation in cultivated *H. annuus*. However, despite sampling a broad variety of allelic diversity among cultivated *H.*

annuus (Mandel et al. 2011), we fail to replicate findings of As, Cr, and Ni hyperaccumulation in any Core 12 accessions of cultivated *H. annuus* (Cutright et al. 2010), finding instead relatively low levels of accumulation of these metals among cultivated accessions, and extremely low accumulation of Cr across the entire genus (relative to current hyperaccumulation threshold criteria). However, moderate (~30 mg/kg As) uptake of leaf As was observed in most cultivated accessions, and a single cultivated accession exhibited leaf As concentrations exceeding 100 mg/kg. Although far from the hyperaccumulation threshold criteria, these levels are between 1-2 orders of magnitude higher than concentrations typically found in land plants (van der Ent et al. 2013, Pollard et al. 2014). Similar As concentrations were also observed in wild *H. annuus* and its sister species *H. argophyllus*, suggesting a possible evolutionary origin of the capacity for elevated As accumulation in the most recent common ancestor of *H. annuus* and *H. argophyllus*. Other isolated instances of elevated As accumulation were observed in seven other taxa spread across the genus. These findings, coupled with a non-significant Blomberg's *K*, suggests that these instances of elevated As accumulation may represent additional independent evolutionary origins or repeated losses from an accumulating common ancestor.

We also find elevated Ni accumulation among cultivated *H. annuus*, exceeding leaf concentrations of 100 mg/kg which some authors regard to be the threshold criteria for Ni "hemi-accumulation," an intermediate classification between non-accumulators and hyperaccumulators (Boyd and Jaffre 2009, van der Ent et al. 2013). However, 10 out of 12 instances of Ni hemi-accumulation in cultivated *H. annuus* were also accompanied by visible signs of Ni toxicity, and similar results were observed in both wild *H. annuus*

and *H. argophyllus*. Conversely, the serpentine sunflower *H. exilis* (sister to the common ancestor of *H. annuus* and *H. argophyllus*), which occurs in serpentine soils with extremely high Ni concentrations, exhibited hemi-accumulator properties without apparent Ni toxicity. Outside of the annual clade, most species achieved either hemi-accumulator or hyperaccumulator status, but no taxa exhibited signs of Ni toxicity (with the exception of mortality in *H. arizonensis*). Together, these findings suggest that the capacity for elevated Ni accumulation may be ancestral to the genus, with one or potentially multiple losses of tolerance within the annual clade.

Recent studies have proposed lowering the threshold hyperaccumulator concentrations for Cu from 1,000 to 300 mg/kg and Zn from 10,000 to 3,000 mg/kg (Broadley et al. 2007, Kraemer 2010, van der Ent et al. 2013). Using these modified cutoffs, 25 wild taxa and 9 cultivated accessions would meet the threshold for Zn hyperaccumulation, and two wild taxa would meet the threshold for Cu hyperaccumulation. Additionally, under these criteria, 19 out of 25 observed Zn hyperaccumulators would also be Cd hyperaccumulators. However, no evolutionary correlation was supported between Cd and Zn accumulation, and the phylogenetic signal for both traits was non-significant. Despite this finding, the relative ubiquity of the capacity for Cd and Zn hyperaccumulation across the genus strongly suggests a possible shared evolutionary origin ancestral to *Helianthus*. The hyperaccumulation paradigm implies a bimodal distribution of metal accumulation in plants, with a clear distinction between non-accumulators and hyperaccumulators approximately defined by current hyperaccumulator criteria. If this is indeed the case, the study of metal accumulation as a continuous trait may fail to detect correlated evolution of presence/absence-type traits.

However, no consensus yet exists in phylogenetic comparative biology regarding appropriate methods for testing correlated evolution between categorical traits (Maddison and FitzJohn 2015). Nevertheless, previous work in other systems points toward a strong association between genetic and physiological processes controlling Zn and Cd root-to-shoot translocation and tolerance, which is consistent with the strong overlap in Zn and Cd hyperaccumulator taxa observe within *Helianthus* (Hanikenne et al. 2008, O Lochlainn et al. 2011).

To maximize the number of taxa and metals considered, this study uses a single soil treatment level per metal. Because of this, we cannot fully address here the evolutionary history of metal tolerance, a physiologically complex trait which would ideally be investigated using a dose-response curve approach over a gradient of soil metal treatment concentrations (Cappa et al. 2015). Thanks to recent statistical advances, such approaches can now be conducted in a phylogenetic comparative framework (Goolsby 2015, Goolsby and Mason 2016), providing a powerful means of investigating multiple metal tolerance and accumulation in the context of plant physiology.

CONCLUSION

Here we identify widespread metal hyperaccumulation of Cd and Zn and elevated metal accumulation ("hemi-accumulation") of Ni within cultivated sunflowers. The capacity for elevated accumulation of each of these elements appears to have evolved in wild *Helianthus* prior to sunflower domestication, although cultivated sunflowers lack the ability to tolerate high foliar concentrations of either Ni or Zn. Additionally, the capacity for hyperaccumulation of Cd, Ni, and Zn appears to be widespread throughout wild

Helianthus. We find possible support for correlated evolution of Cd and Zn hyperaccumulation as categorical presence/absence traits, but further work and statistical development is needed to further clarify the relationship. We also identify two instances of potential Cu hyperaccumulation (using the criterion of 300 mg/kg Cu) in wild *Helianthus*. No plants exposed to Se treatments in the present study survived, suggesting that sunflowers may be too sensitive to Se to properly investigate in the context of hyperaccumulation. We also find no evidence of hyperaccumulation of As, Cr, or Pb in any *Helianthus* species. Future work on multiple metal hyperaccumulation in *Helianthus* should further investigate the evolutionary interactions of Cd, Ni, and Zn hyperaccumulation and tolerance, as well as what potential adaptive benefits may favor metal hyperaccumulation.

REFERENCES

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

Boyd, R. S. 2004. Ecology of Metal Hyperaccumulation. *New Phytologist* 162:563-567.

Boyd, R. S. 2007. The defense hypothesis of elemental hyperaccumulation: status, challenges and new directions. *Plant and Soil* 293:153-176.

Boyd, R. S. 2012. Plant defense using toxic inorganic ions: Conceptual models of the defensive enhancement and joint effects hypotheses. *Plant Science* 195:88-95.

Boyd, R. S., and T. Jaffre. 2009. Elemental Concentrations of Eleven New Caledonian Plant Species from Serpentine Soils: Elemental Correlations and Leaf-age Effects. *Northeastern Naturalist* 16:93-110.

Broadley, M. R., P. J. White, J. P. Hammond, I. Zelko, and A. Lux. 2007. Zinc in plants. *New Phytologist* 173:677-702.

Broadley, M. R., N. J. Willey, J. C. Wilkins, A. J. M. Baker, A. Mead, and P. J. White. 2001. Phylogenetic variation in heavy metal accumulation in angiosperms. *New Phytologist* 152:9-27.

Bruggeman, J., J. Heringa, and B. W. Brandt. 2009. PhyloPars: estimation of missing parameter values using phylogeny. *Nucleic Acids Research* 37:W179-W184.

Cappa, J. J., and E. A. H. Pilon-Smits. 2014. Evolutionary aspects of elemental hyperaccumulation. *Planta* 239:267-275.

Cappa, J. J., C. Yetter, S. Fakra, P. J. Cappa, R. DeTar, C. Landes, E. A. H. Pilon-Smits, and M. P. Simmons. 2015. Evolution of selenium hyperaccumulation in Stanleya

(Brassicaceae) as inferred from phylogeny, physiology and X-ray microprobe analysis. *New Phytologist* 205:583-595.

Cecchi, L., R. Gabbrielli, M. Arnetoli, C. Gonnelli, A. Hasko, and F. Selvi. 2010. Evolutionary lineages of nickel hyperaccumulation and systematics in European Alysseae (Brassicaceae): evidence from nrDNA sequence data. *Annals of Botany* 106:751-767.

Cutright, T., N. Gunda, and F. Kurt. 2010. Simultaneous Hyperaccumulation of Multiple Heavy Metals by Helianthus Annuus Grown in a Contaminated Sandy-Loam Soil. *International Journal of Phytoremediation* 12:562-573.

Goolsby, E. W. 2015. Phylogenetic Comparative Methods for Evaluating the Evolutionary History of Function-Valued Traits. *Systematic Biology* 64:568-578.

Goolsby, E. W., and C. M. Mason. 2015. Toward a more physiologically and evolutionarily relevant definition of metal hyperaccumulation in plants. *Frontiers in Plant Science* 6:33.

Goolsby, E. W., and C. M. Mason. 2016. Response: "Commentary: Toward a more physiologically and evolutionarily relevant definition of metal hyperaccumulation in plants". *Frontiers in Plant Science* 6:1252.

Hanikenne, M., I. N. Talke, M. J. Haydon, C. Lanz, A. Nolte, P. Motte, J. Kroymann, D. Weigel, and U. Kraemer. 2008. Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of HMA4. *Nature* 453:391-U344.

Heiser, C. B. J., D. M. Smith, S. B. Clevenger, and W. C. J. Martin. 1969. *The North American Sunflowers-D Helianthus-D.* Memoirs of the Torrey Botanical Club 22:1-218.

Jansen, S., M. R. Broadley, E. Robbrecht, and E. Smets. 2002. Aluminum hyperaccumulation in angiosperms: A review of its phylogenetic significance. *Botanical Review* 68:235-269.

Kraemer, U. 2010. Metal Hyperaccumulation in Plants. *Annual Review of Plant Biology*, Vol 61 61:517-534.

Maddison, W. P., and R. G. FitzJohn. 2015. The Unsolved Challenge to Phylogenetic Correlation Tests for Categorical Characters. *Systematic Biology* 64:127-136.

Mandel, J. R., J. M. Dechaine, L. F. Marek, and J. M. Burke. 2011. Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, Helianthus annuus L. *Theoretical and Applied Genetics* 123:693-704.

Mandel, J. R., S. Nambeesan, J. E. Bowers, L. F. Marek, D. Ebert, L. H. Rieseberg, S. J. Knapp, and J. M. Burke. 2013. Association Mapping and the Genomic Consequences of Selection in Sunflower. *Plos Genetics* 9.

Mason, C. M., S. E. McGaughey, and L. A. Donovan. 2013. Ontogeny strongly and differentially alters leaf economic and other key traits in three diverse Helianthus species. *Journal of Experimental Botany* 64:4089-4099.

O Lochlainn, S., H. C. Bowen, R. G. Fray, J. P. Hammond, G. J. King, P. J. White, N. S. Graham, and M. R. Broadley. 2011. Tandem quadruplication of HMA4 in the zinc (Zn) and cadmium (Cd) hyperaccumulator Noccaea caerulescens. *Plos One* 6:e17814-e17814.

Pollard, A. J., R. D. Reeves, and A. J. M. Baker. 2014. *Facultative hyperaccumulation of heavy metals and metalloids*. Plant Science 217:8-17.

Rascio, N., and F. Navari-Izzo. 2011. Heavy metal hyperaccumulating plants: How and why do they do it? And what makes them so interesting? *Plant Science* 180:169-181.

Reeves, R. D., and A. J. M. Baker. 2000. Metal-accumulating plants. *Phytoremediation of Toxic Metals*:193-229.

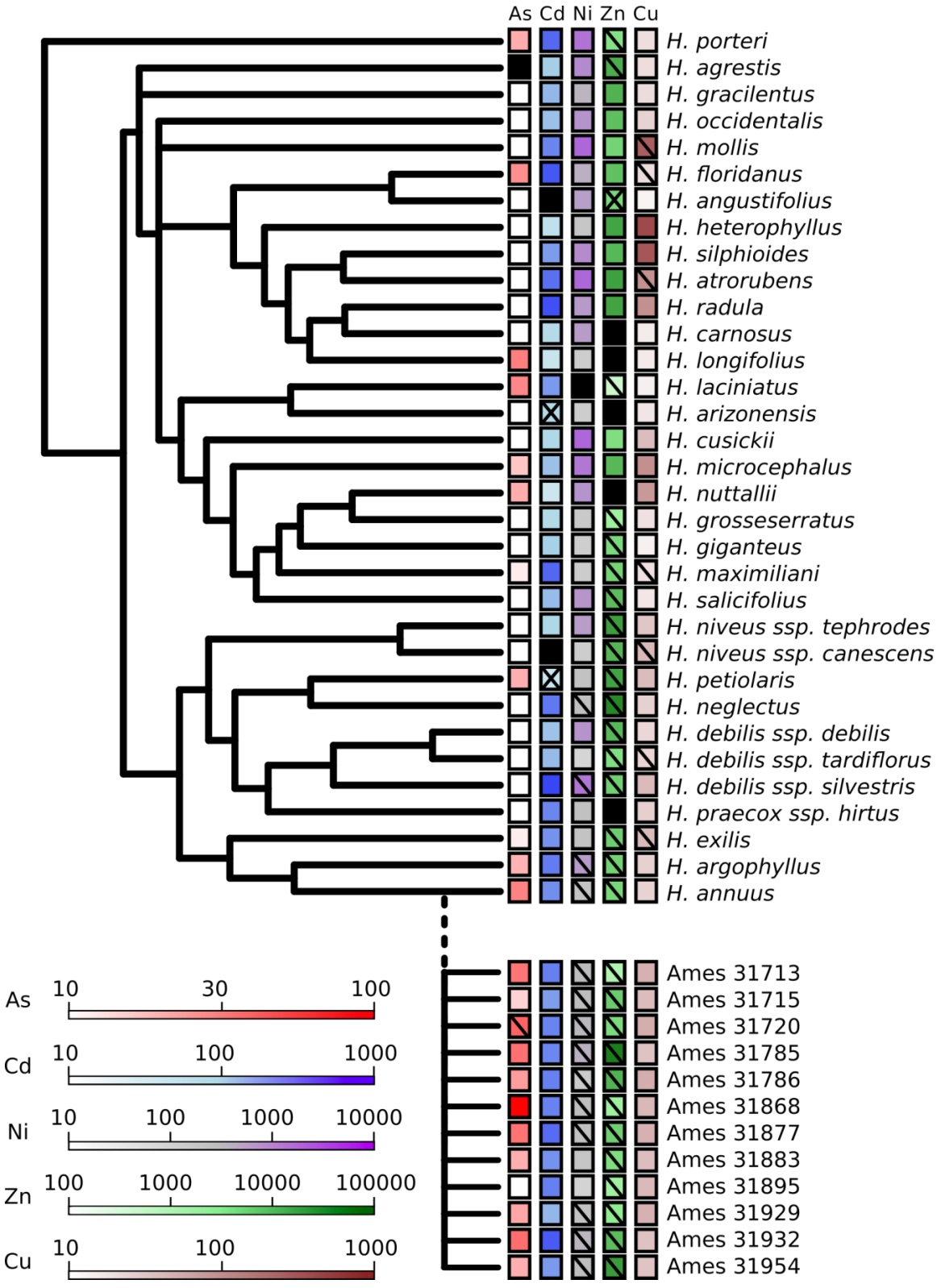
Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3:217-223.

Stephens, J. D., W. L. Rogers, C. M. Mason, L. A. Donovan, and R. L. Malmberg. 2015. Species tree estimation of diploid Helianthus (Asteraceae) using target enrichment. *American Journal of Botany* 102:910-920.

van der Ent, A., A. J. M. Baker, R. D. Reeves, A. J. pollard, and H. Schat. 2013. Hyperaccumulators of metal and metalloid trace elements: Facts and fiction. *Plant and Soil* 362:319-334.

FIGURES

Figure 5.1. Elemental concentrations of As, Cd, Ni, Zn, and Cu in dried sunflower leaves (mg/kg) for wild and cultivated *Helianthus* (cultivated *H. annuus* are represented by a polytomy and connected to wild *H. annuus* via a dotted line). Visible signs of metal toxicity (leaf chlorosis or deformation) is indicated by a diagonal slash over the box corresponding to a particular species-metal combination. Mortality in one of two replicates for a given species-metal combination is denoted by an (X), and mortality in both replicates is denoted by a solid black box.


As Cd Ni Zn Cu
H. porteri
H. agrestis
H. gracilentus
H. occidentalis
H. mollis
H. floridanus
H. angustifolius
H. heterophyllus
H. silphioides
H. atrorubens
H. radula
H. carnosus
H. longifolius
H. laciniatus
H. arizonensis
H. cusickii
H. microcephalus
H. nuttallii
H. grosseserratus
H. giganteus
H. maximiliani
H. salicifolius
H. niveus ssp. tephrodes
H. niveus ssp. canescens
H. petiolaris
H. neglectus
H. debilis ssp. debilis
H. debilis ssp. tardiflorus
H. debilis ssp. silvestris
H. praecox ssp. hirtus
H. exilis
H. argophyllus
H. annuus
Ames 31713
Ames 31715
Ames 31720
Ames 31785
Ames 31786
Ames 31868
Ames 31877
Ames 31883
Ames 31895
Ames 31929
Ames 31932
Ames 31954
As 10 30 100
Cd 10 100 1000
Ni 10 100 1000 10000
Zn 100 1000 10000 100000
Cu 10 100 1000

CHAPTER 6

CONCLUSIONS

This dissertation has put forward methods for studying complex phenotypes in phylogenetic comparative studies and explored the evolution of heavy metal hyperaccumulation in a comparative framework. In Chapter 2, tools were developed to extend phylogenetic comparative methods to incorporate function-valued traits. Function-valued comparative methods inherit the flexibility of many phylogenetic generalized least squares (PGLS) based analyses, and the statistical properties of these methods perform well assuming a Brownian motion model of evolution. However, the methods of Chapter 2 are limited in the types of evolutionary models that may be specified, as the PGLS-based approach lacks a log-likelihood function for estimating model parameters.

Chapter 3 further develops the methods of Chapter 2 with the introduction of a pseudolikelihood-based approach for parameter estimation. This framework allows for extremely flexible model specification, including alternatives to Brownian motion evolution, incorporation of fixed effects, combinations of multiple evolutionary hypotheses, within-species variation, and missing data. In addition to function-valued traits, Chapter 3 is applicable to a broad suite of high-dimensional traits, such as comparative analysis of morphometrics and expression data. The methods of Chapter 3 substantially outperform existing "distance-based" in terms of both Type I error and statistical power under a variety of evolutionary scenarios.

Heavy metal hyperaccumulation in sunflowers is then investigated in a phylogenetic context in Chapters 4 and 5. Chapter 4 investigated the elemental defense hypothesis, which predicts a negative relationship between foliar metal concentration and herbivory rates. Overall, mixed support was found for the elemental defense hypothesis as it relates to metal accumulation in sunflowers and the generalist insect herbivore *Vanessa cardui* (the painted lady butterfly): when faced with no choices, *V. cardui* caterpillars readily consumed leaves from plants grown in both Cd and Ni-amended soils at similar rates to leaves from plants grown in control (non-metal-amended) soils. In contrast, caterpillars tended to prefer control leaves over Cd leaves in both *H. annuus* and *H. argophyllus*. Similarly, caterpillars displayed a preference for control *H. argophyllus* leaves over nickel-treated plants. However, caterpillars exhibited a small but statistically significant preference for nickel-treated *H. exilis* leaves over control leaves. In Chapter 5, the evolutionary history of the capacity for metal accumulation was investigated across both wild and cultivated *Helianthus* for eight metals: As, Cd, Cr, Cu, Ni, Pb, Se, and Zn. Hyperaccumulation was identified for Cd and Zn and elevated metal accumulation ("hemi-accumulation") of Ni within cultivated sunflowers. Elevated accumulation of these metals appears to have evolved in wild *Helianthus* prior to sunflower domestication, although Ni and Zn tolerance appears to have been lost in cultivated sunflowers. Cd, Ni, and Zn hyperaccumulation is also widespread throughout wild *Helianthus*, and elevated Cu accumulation was also observed in two species. No evidence of hyperaccumulation was observed for As, Cr, Pb, or Se in any *Helianthus* species. The results from these experiments suggest further work on multiple metal hyperaccumulation in *Helianthus* should focus on the evolutionary interactions of Cd, Ni, and Zn

hyperaccumulation and tolerance, as well as what potential adaptive benefits may favor metal hyperaccumulation. The tools developed in Chapters 2 and 3 provide the framework to explore the evolutionary interactions of metal tolerance and hyperaccumulation as distinct function-valued traits, as well as for examining the adaptive hypotheses in an integrated comparative framework.

APPENDIX A

SUPPORTING INFORMATION FOR CHAPTER 3

R Code for Efficient Repeated Calculations of Phylogenetically Independent Contrasts.

```
require(phylocurve)
require(ape)

nspecies <- 1000
ntraits <- 500
tree <- rtree(n = nspecies) # Simulate a random 1000-species phylogeny
Y <- matrix(rnorm(nspecies*ntraits),ncol=ntraits) # Generate random
data
rownames(Y) <- tree$tip.label

# Call phylocurve:::prep_multipic (an internal phylocurve function)
prep.Y <- phylocurve:::prep_multipic(x = Y,phy = tree)

# Calculate phylogenetically independent contrasts
# using pic (ape package)
Y.pics.ape <- apply(X = Y,MARGIN = 2,FUN = pic,phy = tree)

# Calculate PICs using phylocurve:::multipic
# returns a list with contrasts, sum_invV, log_detV, root
# 'contrast' is a matrix of PICs
# 'sum_invV' is equal to sum(solve(vcv(tree)))
# 'log_detV' is equal to log(det(vcv(tree)))
# 'root' is the maximum likelihood phenotypic value at the root
Y.pics.phylocurve <- do.call(phylocurve:::multipic,prep.Y)
```

```
# Verify that results are identical
range(Y.pics.ape - Y.pics.phylocurve$contrasts)

# Generate 50 random datasets
niter <- 50
randomY <- vector(mode = "list",length = niter)
for(i in 1:niter)
{
  randomY[[i]] <- matrix(rnorm(nspecies*ntraits),ncol=ntraits)
  rownames(randomY[[i]]) <- tree$tip.label
}
#########################################################################
# Compare time to calculate PICs on 50 random datasets
# using pic vs multipic
#
##### pic function
system.time(for(i in 1:niter) apply(X = randomY[[i]],MARGIN = 2,FUN =
pic,phy = tree))
##### user  system elapsed
##### 18.35    0.23   18.61

##### multipic function
system.time(for(i in 1:niter)
{
  prep.Y$phe[1:nspecies,] <- randomY[[i]]
# update prep.Y$phe with new data
  do.call(phylocurve:::multipic,prep.Y)
})
##### user  system elapsed
##### 1.38    0.14    1.52
#########################################################################
```

APPENDIX B

SUPPORTING INFORMATION FOR CHAPTER 4

Dataset B1. USDA-GRIN accession information for all taxa included in this study.

Taxon	Country	State/Province	PI
Helianthus angustifolius	USA	Mississippi	673177
Helianthus angustifolius	USA	Tennessee	649936
Helianthus angustifolius	USA	North Carolina	664719
Helianthus annuus	USA	New Mexico	435473
Helianthus annuus	USA	California	435587
Helianthus annuus	USA	Oklahoma	468486
Helianthus annuus	USA	Idaho	531027
Helianthus annuus	USA	North Dakota	539903
Helianthus annuus	USA	Montana	586820
Helianthus annuus	USA	Arkansas	613727
Helianthus annuus	USA	California	649844
Helianthus annuus	USA	Ohio	649853
Helianthus annuus	USA	California	649858
Helianthus argophyllus	USA	Texas	435630
Helianthus argophyllus	USA	Texas	435635
Helianthus argophyllus	USA	Florida	468651
Helianthus argophyllus	USA	Texas	494572
Helianthus argophyllus	USA	Texas	494573
Helianthus argophyllus	USA	Texas	494576
Helianthus argophyllus	USA	Texas	494580
Helianthus argophyllus	USA	Texas	494582
Helianthus argophyllus	USA	North Carolina	664729
Helianthus cusickii	USA	California	649966
Helianthus cusickii	USA	Washington	664658
Helianthus cusickii	USA	Oregon	664662
Helianthus debilis subsp. debilis	USA	Florida	435669
Helianthus debilis subsp. silvestris	USA	Texas	435651
Helianthus debilis subsp. tardiflorus	USA	Florida	468689
Helianthus exilis	USA	California	649886

Helianthus exilis	USA	California	649893
Helianthus exilis	USA	California	649895
Helianthus exilis	USA	California	649897
Helianthus exilis	USA	California	649898
Helianthus exilis	USA	California	649899
Helianthus exilis	USA	California	649900
Helianthus exilis	USA	California	649901
Helianthus exilis	USA	California	649902
Helianthus exilis	USA	California	664626
Helianthus exilis	USA	California	664630
Helianthus exilis	USA	California	664631
Helianthus neglectus	USA	New Mexico	435769
Helianthus neglectus	USA	Texas	468769
Helianthus neglectus	USA	New Mexico	468781
Helianthus niveus subsp. canescens	USA	California	468788
Helianthus niveus subsp. canescens	USA	Arizona	649905
Helianthus niveus subsp. tephrodes	Mexico	Sonora	613758
Helianthus petiolaris subsp. fallax	USA	Utah	435838
Helianthus petiolaris subsp. petiolaris	Canada	Saskatchewan	592354
Helianthus praecox subsp. hirtus	USA	Texas	435855
Helianthus praecox subsp. praecox	USA	Texas	435847
Helianthus praecox subsp. runyonii	USA	Texas	435849

APPENDIX C

SUPPORTING INFORMfATION FOR CHAPTER 5

Dataset C1. USDA-GRIN accession information for all taxa included in this study.

Taxon	**State/Cultivar**	**PI**
Helianthus agrestis	Florida	673205
Helianthus angustifolius	Louisiana	673154
Helianthus annuus	HA 404	597368
Helianthus annuus	RHA 396	597373
Helianthus annuus	HA 314	599783
Helianthus annuus	VIR 847	386230
Helianthus annuus	Mammoth	476853
Helianthus annuus	HA 234	599778
Helianthus annuus	HA 821	599984
Helianthus annuus	RHA 408	603989
Helianthus annuus	RHA 426	617099
Helianthus annuus	RHA 328	664202
Helianthus annuus	HA 316	664225
Helianthus annuus	INRA line SF_230	SF_230
Helianthus annuus	Utah	673305
Helianthus argophyllus	Texas	494572
Helianthus arizonensis	Arizona	653549
Helianthus atrorubens	Alabama	649940
Helianthus carnosus	Florida	649956
Helianthus cusickii	California	649966
Helianthus debilis subsp. debilis	Florida	648666
Helianthus debilis subsp. silvestris	Texas	435651
Helianthus debilis subsp. tardiflorus	Florida	468689
Helianthus exilis	California	649895
Helianthus floridanus	Florida	673197
Helianthus giganteus	Ohio	664647
Helianthus gracilentus	California	435684
Helianthus grosseserratus	Iowa	613793
Helianthus heterophyllus	Louisiana	673162

Helianthus laciniatus	New Mexico	653562
Helianthus longifolius	Georgia	650000
Helianthus maximiliani	Iowa	613794
Helianthus microcephalus	South Carolina	664703
Helianthus mollis	Ohio	673147
Helianthus neglectus	Texas	468769
Helianthus niveus subsp. canescens	Arizona	649905
Helianthus niveus subsp. tephrodes	California	650019
Helianthus nuttallii	Montana	531053
Helianthus occidentalis	Ohio	664648
Helianthus petiolaris	Utah	435834
Helianthus porteri	Georgia	649914
Helianthus praecox subsp. hirtus	Texas	435855
Helianthus radula	Louisiana	673163
Helianthus salicifolius	Oklahoma	664781
Helianthus silphioides	Louisiana	673156