1 Venom gene sequence diversity and expression jointly shape diet adaptation in pitvipers

2 Andrew J. Mason[1*], Matthew L. Holding[2], Rhett M. Rautsaw[3], Darin R. Rokyta[4], Christopher L.
3 Parkinson[3,5], and H. Lisle Gibbs[1*]

4 [1] Department of Evolution, Ecology and Organismal Biology, The Ohio State University,
5 Columbus, OH, USA

6 [2] Life Sciences Institute, University of Michigan, Ann Arbor, MI, USA

7 [3] Department of Biological Sciences, Clemson University, Clemson, SC, USA

8 [4] Department of Biological Science, Florida State University, Tallahassee, FL, USA

9 [5] Department of Forestry and Environmental Conservation, Clemson University, Clemson, SC,
10 USA

11 *Corresponding Authors: Andrew J. Mason, H. Lisle Gibbs

12 **Email:** mason.501@osu.edu, gibbs.128@osu.edu

13

14

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

1 **Abstract**

2 Understanding the joint roles of amino acid sequences variation of proteins and differential
3 expression during adaptive evolution is a fundamental, yet largely unrealized, goal of
4 evolutionary biology. Here, we use phylogenetic path analysis to analyze a comprehensive
5 venom gland transcriptome dataset spanning three genera of pitvipers to identify the functional
6 genetic basis of a key adaptation (venom complexity) linked to diet breadth. Analysis of gene
7 family-specific patterns reveal that, for genes encoding two of the most important venom
8 proteins (SVMPs and SVSPs), there are direct, positive relationships between sequence
9 diversity, evenness of expression, and increased diet breadth. Further analysis of gene family
10 diversification for these proteins showed no constraint on how individual lineages achieved toxin
11 gene sequence diversity in terms of patterns of paralog diversification. In contrast, another
12 major venom protein family (PLA_2s) showed no relationship between venom molecular diversity
13 and diet breadth. Additional analyses suggest that other molecular mechanisms—such as
14 higher absolute levels of expression—are responsible for diet adaptation involving these venom
15 proteins. Broadly, our findings argue that functional diversity generated through sequence and
16 expression variation determine adaptation in key components of pitviper venoms, which mediate
17 complex molecular interactions between the snakes and their prey.

18

19

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

Introduction

Adaptation at the molecular level can occur through changes in protein coding sequence or patterns of gene expression and identifying the relative roles of these mechanisms is central to understanding trait evolution (Barrett and Hoekstra 2011; Rockman 2012; Rausher and Delph 2015; Smith et al. 2020). Although both mechanisms play important roles in evolution (Carroll 2005; Hoekstra and Coyne 2007; Carroll 2008), there are differing expectations for their relative contributions to complex traits. Protein coding mutations can produce novel functions, especially when coupled with gene duplications that reduce selective constraints (Ohno 1970; Hoekstra and Coyne 2007). Regulatory changes serve critical roles in morphological evolution, and the time and tissue-specific nature of gene expression is expected to reduce pleiotropic effects of regulatory variation facilitating the evolution of novel adaptations (Carroll 2008; Stern and Orgogozo 2008). Moreover, because there are more pathways for altering the expression of a gene compared to altering its sequence, regulatory mechanisms present larger mutational targets, which lead to differences in their evolutionary rates and lability compared to protein coding regions (Rokyta, Wray et al. 2015; Besnard et al. 2020). Understanding how protein coding and/or regulatory changes mediate realized adaptive function has significant implications for identifying general evolutionary processes linking genomic variation to adaptive phenotypes (Smith et al. 2020). This requires the development and use of detailed genotype-phenotype maps that are linked to realized ecological variation from diverse species groups.

Traditionally, genotype-to-phenotype maps for adaptive traits have been constructed using a “forward genetics” approach which focuses on experimental analyses of segregating genetic variation in model species (Barrett and Hoekstra 2011). Forward genetics has proved highly successful for identifying the molecular basis of many adaptations, but is limited by the need to work with model species amenable to either experimental manipulation or observational studies that link segregating genetic variants to phenotypes with statistical association methods (Tanksley 1993; Marigorta et al. 2018). These methods are incompatible with many adaptive phenotypes of interest to evolutionary biologists because such traits may occur in species that cannot be interbred or where phenotypic variation of interest may only occur between species (Smith et al. 2020). Studies to date are limited to small numbers of species in which the “forward genetics” paradigm can be applied, which raises questions about the generality of their results, especially at macroevolutionary scales.

A recently proposed approach to overcome these issues is to use comparative phylogenetic methods to analyze clade-wide genomic datasets to link phenotypic variation to its genetic underpinnings (Nagy et al. 2020; Smith et al. 2020). This approach builds on the

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

increasing availability of genomic datasets and uses long-standing comparative phylogenetic methods to identify associations between functionally relevant genetic and phenotypic variation while accounting for shared ancestry (Smith et al. 2020). While lacking the experimental certainty of forward genetic approaches, comparative phylogenetics methods broaden the scope of studies of adaptive phenotypes and can yield new insights into how evolutionary mechanisms mold the genetic basis of phenotypic variation (Pease et al. 2016; Hu et al. 2019; Sackton et al. 2019). Comparative methods like phylogenetic path analysis that test for causal structure amongst a suite of compared variables and have recently been used to understand genome-environment interactions in multiple groups (Hardenberg and Gonzalez-Voyer 2013; Voyer and Garamszegi 2014; Guignard et al. 2019; Chak et al. 2021). Phylogenetic path analysis therefore provides a useful method to apply to genome scale data for functional genetic variation from multiple species, especially when genetic and phenotypic variation are closely tied to ecological functions.

Animal venoms are a model system for investigating the molecular mechanisms that underlie adaptive traits because of the unusually direct connection between venom genes, phenotypes, and adaptive function that allows comprehensive investigation across levels of biological organization (Gibbs and Rossiter 2008; Casewell et al. 2012; Casewell et al. 2014; Rokyta, Margres, et al. 2015). Whole venoms are complex adaptive phenotypes that can be broken down into distinct components—individual proteins making up the venom—and linked to known molecular underpinnings, and their functional impacts (Casewell et al. 2013; Zancolli and Casewell 2020). Several of the major gene families that contribute to venom occur as tandemly arrayed gene islands in distinct genomic locations (Sanggaard et al. 2014; Gendreau et al. 2017; Casewell et al. 2019; Schield et al. 2019; Margres et al. 2021). This genomic architecture means the evolution of venom genes and the pathway from genotype to complex phenotype can be investigated in multiple gene families that underlie the same phenotype across a set of venomous species. These features make venom an exceptional system for examining how complex adaptive phenotypes are assembled and evolve, and for understanding the impact of phenotypic complexity on ecological function (Holding, Drabeck, et al. 2016; Sunagar et al. 2016; Arbuckle 2020; Giorgianni et al. 2020; Zancolli and Casewell 2020; Holding et al. 2021).

Studies of venomous species have yielded numerous important insights into how molecular adaptations arise. For example, molecular and ecological studies in cone snails have provided evidence for dynamic expansion of toxin gene families, evidence of pervasive positive selection, and correlations between venom compositions and diet (Duda and Palumbi 1999; Duda and Palumbi 2004; Duda and Remigio 2008; Remigio and Duda 2008; Chang and Duda

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

2012; Chang and Duda 2014; Phuong et al. 2016; Li et al. 2017). In spiders, venom complexity has been shown to vary based on feeding ecologies (Pekár et al. 2018). Several studies on individual snake species have also evaluated the roles of sequence and expression evolution in venom toxins and indicate that both mechanisms facilitate phenotypic evolution, possibly in different evolutionary or ecological contexts (Margres et al. 2016; Margres et al. 2017a; Margres et al. 2017b; Hofmann et al. 2018; Rautsaw et al. 2019; Zancolli et al. 2019).

At the macroevolutionary scale, a recent study by Holding et al. 2021 used k-mer based metrics from venom gland transcriptomes and whole venom RP-HPLC data from 68 primarily North American pitvipers (rattlesnakes and moccasins) to show a strong positive relationship between the molecular complexity of venom and phylogenetic diversity in diet. This work identified the molecular complexity of venom as an adaptive phenotype that is correlated with a key ecological trait (diet breadth) in these snakes, although their reliance on k-mers prevented the specific genetic mechanisms from being identified. Nonetheless, the availability of a comprehensive molecular dataset on venom variation for a phylogenetically diverse snake clade opens the door to using a comparative phylogenetics approach to identify the specific genetic mechanisms underlying this adaptive trait.

Here, we analyze fully assembled venom gland transcriptomes for the 68 lineages represented in Holding et al. 2021 using phylogenetic path analysis (Hardenberg and Gonzalez-Voyer 2013; Voyer and Garamszegi 2014) to dissect the relative roles of gene composition, protein sequence diversity, and expression diversity as they relate to diet breadth in these snakes. In addition, we capitalized on the nature of venom as a mixture of proteins from distinct multi-gene families to determine if separate or concerted evolutionary processes contribute to venom diversity from separate regions of the genome. Finally, for two families where toxin sequence diversity showed significant associations with dietary diversity, we tested whether lineages show evidence for similar or divergent evolutionary pathways for generating protein sequence diversity. Our results show that both sequence diversity and expression variation mediate adaptation in pitviper venoms, but the roles of sequence diversity and expression vary for different components of this complex phenotype. These results highlight how complex molecular traits can evolve via alternative routes to adaptation.

Results

Venom gland transcriptomes

We assembled and annotated venom-gland transcriptomes for the 214 individuals comprising 68 rattlesnake and moccasin lineages used in Holding et al. 2021, with specimen representation

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

ACCEPTED MANUSCRIPT

for each lineage varying from 1-10 individuals (Supplemental Table S1, Supplemental Table S2). Individual snakes expressed on average 78.4 transcripts encoding toxin proteins (range = 32-128). Using the annotated transcriptomes, we calculated gene content as the total number of toxins, toxin sequence diversity as the effective number of amino acid 20-mers (the number of unique k-mers that would represent equivalent diversity with uniform occurrence, see Methods), and toxin expression diversity as the effective number of expressed toxin transcripts (the number of expressed toxins that would represent equivalent diversity with uniform expression, see Methods). Lineage specific estimates of these measures were obtained by averaging across samples, though variation in these metrics was apparent within several lineages (Supplemental Figures S1, S2, S3).

To verify that technical variation in sample treatment (e.g., differences in sequencing depth and numbers of assembled transcripts) did not bias statistical inference, we tested for a relationship between these variables and the number of recovered toxins. Although we found some evidence of a significant correlation between the number of recovered toxins and numbers of merged reads among samples (p=0.063, Supplemental Figure S4), this relationship explained a relatively small amount of variation (R^2=0.016). Similarly, we found no significant relationship between the numbers of expressed transcripts and the number of recovered toxins (p=0.664, R^2 < 0.001, Supplemental Figure S5). Importantly, we found no evidence of an interaction between numbers of merged reads (p=0.369, Supplemental Table S3) or numbers of expressed transcripts with lineage assignment (p=0.618, Supplemental Table S4), indicating that inferences made among lineages are unbiased by technical variation.

We tested for evidence of phylogenetic signal among gene content, sequence diversity, and expression diversity metrics with Blomberg's K and lambda. Gene content, sequence diversity, and toxin expression diversity all showed evidence of significant phylogenetic signal based on estimates of Blomberg's K (gene content=0.47, expression diversity=0.38, sequence diversity=0.46), and both gene content and sequence diversity showed evidence of significant phylogenetic signal based on lambda (Supplemental Table S5, Supplemental Figure S6). Evidence of phylogenetic signal in these metrics indicates a moderate degree of predictability in the venom genotype-to-phenotype map based on degrees of evolutionary divergence among related snake lineages.

Path Analysis

To examine how expression and protein-coding sequence evolution affect the dynamics of venom and diet diversity we tested 10 path models defining hypothesized relationships among gene content, sequence diversity, expression diversity, and diet breadth (Supplemental Figure

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

S7) for 30 snake lineages for which we had reliable diet data. Here, diet breadth corresponded to the mean phylogenetic distance measure of diet used in Holding et al. (2021), who showed that snake diet breadth as a function of its phylogenetic diversity of prey species was a better predictor of venom complexity than prey species richness alone. Phylogenetic path models represented varying roles of sequence diversity and expression diversity as having direct or indirect effects on diet breadth, independently or in combination, while gene content was modelled as acting indirectly through these variables.

We found the highest support for Model 3 in which sequence diversity had a moderate, positive correlation with diet breadth and, surprisingly, expression diversity had a moderate negative correlation with diet breadth (Figure 1, Supplemental Figure S8, Supplemental Table S6). Hence, snakes with more diverse, but less evenly expressed sequences had broader diets. As expected, gene content was positively correlated with sequence diversity and expression diversity in this model, showing a strong indirect association with diet breadth mediated through sequence diversity and expression. However, support for Model 3 was not absolute. Model 1 was within 2 C statistic Information Criterion (CICc) of Model 3, indicating similar statistical support (Figure 1, Supplemental Figure S3). Unlike Model 3, Model 1 did not include a connection between expression diversity and diet breadth and showed a weaker relative relationship between sequence diversity and diet (Supplemental Figure S8). Because of the overall similarity of Model 3 and Model 1, the weighted average model we recovered was similar to Model 3 (Figure 1).

In both top-performing models sequence diversity and expression diversity predicted changes in diet. Importantly, although our path models modelled venom sequence and expression diversity as predictors of diet breadth, these relationships do not imply directional causality. Rather, the direct positive correlation between sequence diversity and diet breadth indicates that increased sequence variation is associated more diverse diets. Sequence variation, in turn, is heavily influenced by the underlying gene content. In contrast, more even and hence diverse toxin expression is associated with a narrower diet. Next, we sought to explore this initially counterintuitive result for expression diversity in more detail.

We suspected that the analysis of pooled data may obscure more subtle relationships between expression and diet breadth for individual toxin gene families which, because they are found at distinct genomic locations in these snakes (Schield et al. 2019), represent semi-independent replicates of how venom complexity evolves. To examine whether patterns of complexity detected for the whole venom phenotype are representative of patterns found in individual toxin families, we tested the possible path models in four tandemly arrayed toxin

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

1 families: C-type lectins (CTLs), phospholipase A_2s (PLA_2s), snake venom metalloproteases
2 (SVMPs), and snake venom serine proteases (SVSPs). These toxin families have previously
3 shown heterogeneous relationships between expressed transcript sequence complexity
4 (measured in k-mers) and diet breadth, with three of the families having positive relationships
5 while CTLs displayed no relationship (Holding et al. 2021).
6 Here, we report substantial differences in the optimal models for family-specific path
7 analyses. In particular, analysis of SVMP and SVSP families separately showed support for
8 models where both sequence diversity and expression variation had direct positive correlations
9 with diet breadth (Figure 1d, and 1e). Thus, in contrast with the overall analyses, within each of
10 these toxin families, more diverse patterns of expression were associated with increased diet
11 breadth. All competitive models for the SVSP family also supported a direct relationship
12 between sequence diversity and expression diversity. Models with opposing directions of the
13 relationship between sequence diversity and expression showed equivalent support, as
14 expected, but varied in effect estimates (Figure 1e). This finding indicates an interacting effect of
15 sequence and expression evolution in SVSPs where increased sequence diversity and more
16 even toxin expression are linked.
17 In contrast, for analyses of the CTL and PLA_2 gene families, the top ranked model set
18 included the null model, which did not include any direct connection between sequence and
19 expression diversity and diet breadth (Figure 1b, 1c). This result suggests that 'functional
20 diversity' in CTLs and PLA_2s does not influence the ability of these snakes to consume
21 phylogenetically diverse prey but that other characteristics, such as total expression or the
22 presence of paralogs with specific functions, may play more important roles for these toxin
23 family.
24 **Variation in Expression**
25 To explore how other aspects of venom composition are associated with diet breadth, we
26 compared how absolute expression patterns (rather than complexity in expression) varied
27 among and within major families and tested for correlations with diet breadth. As expected, the
28 number and mean expression of toxins varied significantly among families with PLA_2s exhibiting
29 the lowest number of toxins per lineage ($p < 0.001$, Supplemental Figure S9, Supplemental
30 Table S2), but the highest mean expression levels ($p < 0.001$, Supplemental Figure S9). PLA_2s
31 exhibited a positive correlation between mean expression and diet breadth (p=0.03, R^2=0.38)
32 (Figure 2). This relationship becomes even stronger when a single, high leverage outlier (the
33 South American Rattlesnake, *C. durissus)* is excluded from the analysis (p<0.001, R^2=0.68;
34 Figure 2).

ACCEPTED MANUSCRIPT
Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

These relationships explain why the global path analysis shows a negative relationship between expression diversity and diet breadth. The indices used for path analyses measure diversity as a function of richness and relative abundance. Expression diversity specifically is derived from the number of expressed transcripts and their relative expression (evenness), where we consider more even expression to be more complex. Because PLA_2s consist of only a few, often highly expressed transcripts, they exert a disproportionate effect on expression evenness. Thus, lineages with more complex diets with more highly expressed PLA_2s can show less diverse expression patterns overall. In sum, the strong positive relationship between mean PLA_2 expression and diet breadth suggest that abundance rather than compositional diversity of PLA_2s facilitates eating a broader range of prey.

Mechanisms of gene-family diversification

Our analysis showed that the SVMP and SVSP venom gene families both showed evidence of positive relationships between amino acid sequence diversity and diet breadth. In large gene families, gene sequence diversity is inextricably linked to gene duplications and divergence which collectively produce diverse paralogs. Most pitviper lineages express multiple SVMP and SVSP toxin paralogs and the diversity of these toxin assemblages can lend insight on patterns of gene diversification. Ancient duplications may be observed as highly divergent paralogs in modern taxa, but recent duplications also occur in many venom gene families (Wong and Belov 2012; Giorgianni et al. 2020). The assemblage of toxin paralogs in the venom of a given lineage may consist primarily of conserved ancient paralogs, less divergent recent paralogs, or a combination (Figure 3a). Each of these scenarios can generate sequence variation, but whether either is over-represented as an evolutionary pathway in venoms is not clear.

To assess what patterns of paralog diversification characterized venom gene diversity, we used a similar method to that of Chang and Duda (2014) to compare within-family toxin diversity of each individual against the within-family toxin diversity across *Agkistrodon, Crotalus,* and *Sistrurus.* Specifically, we calculated phylogenetically weighted, standardized mean genetic distance (MGD) for two toxin families where we expected paralog diversification could have an ecological impact acting through sequence diversity: SVMPs and SVSPs. The standardized values of MGD represent the diversity of toxins in a toxin family (i.e., SVMPs or SVSPs) expressed by an individual compared to the total diversity of the toxin family. In the context of a gene family, low estimates of assemblage MGD would occur through assemblages of highly similar (phylogenetically clustered) paralogs, while high estimates of MGD would result which assemblages were very diverse (phylogenetically dispersed) (Figure 3b). This approach therefore allowed us to infer whether diversity in these families arose primarily through

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

ACCEPTED MANUSCRIPT

expression/reliance on highly divergent genes such as ancient or highly derived paralogs versus clusters of more recently duplicated, less differentiated paralogs (Figure 3b).

We observed a range of negative and positive standardized MGD values for SVMPs and SVSPs, with slightly positive means for the overall distribution for both families (mean SVMP=0.29, median SVMP=0.39, mean SVSP=0.21, median SVSP=-0.03, Supplemental Figure S10 & S11). These results indicate that, on average, expressed genes tend to be more divergent than would be expected by chance alone. However, both the SVMP and SVSP distributions appeared multimodal (Figure 4) and Wilcoxon signed rank tests found the distribution of SVMP standardized MGD values to be different than 0 (p=0.005) although SVSPs were not (p=0.247). In the case of SVMPs, two clear peaks were visible centered at approximately -2 and 0.5, with some indication that the larger peak could be considered multimodal with peaks occurring at approximately 0, and slightly less than 1 (Figure 4). Interestingly, the lower peak (centered at approximately -2) in the SVMP distribution was composed exclusively of *Agkistrodon contortrix* and *A. piscivorus* lineages, suggesting that a reliance on a particular subset of SVMP paralogs may be characteristic of the *A. contortrix* + *A. piscivorus* lineage. In SVSPs the two apparent modes of the distribution appeared centered at approximately -0.5 and slightly less than 1 (Figure 4) though there was no apparent taxonomic pattern associated with either mode.

Under scenarios where SVMP and SVSP assemblages are evolutionarily constrained to emphasize either ancient orthologs or recent paralogs, we would expect distributions centered above or below zero, respectively. In contrast, the observed patterns suggest that SVMP and SVSP evolution occurs through a combination of gene duplication, divergence, and loss rather than either extreme mechanisms of high duplication or high divergence (Figure 4). Moreover, the multimodal patterns of each distribution indicate that while there is substantial variation in the diversity of assemblages, subsets of taxa exhibit especially similar or especially diverse SVMP and SVSP assemblages. Expression weighted MGD was highly correlated with standardized MGD for both metrics (Figure 4), demonstrating that lineages did not emphasize expression of more or less diverse paralogs in their total toxin assemblage.

Although, we found no evidence of constraint on the genetic mechanisms for generating sequence diversity, it is possible that different mechanisms of generating diversity could facilitate broader diets. For example, more genetically diverse toxin assemblages might affect a wider phylogenetic diversity of prey, increasing diet breadth. To test this possibility, we compared MGD estimates (which represented more and less diverse toxin assemblages) to diet breadth estimates for each lineage. However, we found no evidence for a relationship between

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022
ACCEPTED MANUSCRIPT

1 diet breadth and MGD (Supplemental Figure S12), indicating that the genetic diversity of toxin
2 assemblages (i.e., emphasis on highly diverged, versus recently diverged paralogs) did not
3 constrain the ecological function of venoms.

4 **Discussion**

5 Our results demonstrate that both sequence diversity and expression variation in toxin genes
6 jointly shape variation in venom, a crucial adaptive trait related to diet breadth in North American
7 pitvipers. Previous studies have provided evidence for positive selection acting on toxin genes
8 implicating the proteins they encode in trophic adaptions (Duda and Palumbi 1999; Li et al.
9 2005; Gibbs and Rossiter 2008; Sunagar and Moran 2015; Haney et al. 2016). Similarly, there
10 is substantial indirect evidence for a role of expression variation in single toxins mediating
11 trophic adaptations (Gibbs and Chiucchi 2011; Aird et al. 2015; Margres et al. 2016; Margres et
12 al. 2017b; Barua and Mikheyev 2019; Barua and Mikheyev 2020). Our work represents an
13 advance by using comparative methods to simultaneously link the contribution of each
14 molecular mechanism to phenotypic variation directly related to diet across diverse lineages.
15 For certain key venom proteins, sequence diversity and expression appear to act in a
16 hierarchical manner to generate the realized adaptive phenotype (whole venom composition).
17 Diversity in protein sequence defines the fundamental functional sequence space for toxin
18 proteins and expression variation brings about the realized toxin phenotype as a refined subset
19 of sequence space. Such a model has been proposed to explain diversity in other venomous
20 systems and variation in expression more broadly (Raser and O'Shea 2005; Lluisma et al.
21 2012). We suspect that a similar relationship will hold for other adaptive phenotypes whose
22 function is driven by additive effects among component proteins.

23 The positive relationship between toxin sequence diversity and diet breadth reinforces
24 the idea that target-mediated interactions at the protein sequence level are a fundamental
25 mechanism mediating predator-prey interactions through molecular phenotypes (Gibbs et al.
26 2020; Holding et al. 2021). Holding et al. (2021) demonstrated a correlation between overall
27 toxin diversity and divergence in homologous venom targets involved in interactions with a
28 single venom toxin (SVSPs). Our results build on this finding by demonstrating that both
29 increased sequence and expression diversity jointly underlie more diverse toxin compositions.
30 This may increase the number and type of physiological targets and, by extension, the number
31 of physiologically distinct prey taxa that a venom can affect (Davies and Arbuckle 2019). We
32 suggest that these same mechanisms underlie positive correlations between venom and diet
33 diversity that have been documented in other venomous animals such as snails and spiders
34 (Phuong et al. 2016; Pekár et al. 2018).

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

We have modeled the relationship between diet breadth, venom, and its genetic underpinning as a unidirectional genotype-phenotype relationship. This approach was effective for identifying how particular genetic mechanisms shape venom evolution but has limitations. In particular, path analyses cannot model bidirectional relationships as might be most appropriate in a feedback or co-evolutionary system. This is potentially important because venoms that function primarily for prey capture likely evolve in a complex, co-evolutionary arms races with their prey in a variety of ecological scenarios (Barlow et al. 2009; Holding, Biardi, et al. 2016; Davies and Arbuckle 2019; Gibbs et al. 2020). Deciphering if and how counter selection due to prey characteristics such as molecular resistance to venom (Holding et al. 2018; Gibbs et al. 2020) is a significant evolutionary force shaping snake venoms through coevolutionary interactions would be a valuable direction for future studies.

Our analysis of gene family evolution in SVMP and SVSP paralogs show no dominant mode of paralog duplication in achieving sequence diversity in toxin coding sequences. Instead, diverse toxin repertoires have emerged through the retention of deeply divergent paralogs, duplication and comparatively minor divergence of paralogs, or a combination of these processes with equal likelihood. These findings are consistent with previous work assessing expressed toxin assemblages in cone snails. Of the four species compared in snails (Chang and Duda 2014), two species expressed mostly similar paralogs (genetic under-dispersion), one species expressed mostly divergent paralogs (genetic over-dispersion), and one species fell between these extremes. Thus, in both snakes and cone snails there is little constraint on the evolutionary pathway to achieving high sequence diversity in toxin genes – rather all pathways seem equally likely. Moreover, we found no association between the genetic diversity of toxin assemblages (MGD) and diet breadth, indicating that having few, highly divergent paralogs versus many, less divergent paralogs did not have functional consequences for prey acquisition.

Given that venom targets basal physiological processes such as the coagulation cascade (Serrano 2013) and neurotransmission sites (Fry et al. 2009), it may be that relatively few amino acid substitutions can refine venom targeting for divergent prey tissues. The further divergence in more ancient paralogs may reflect combined effects of neutral evolution (Aird et al. 2017) and refinements to protein function not tied to prey specificity, such as structural stability of the protein (Sunagar et al. 2014), neofunctionalization for novel physiological targets (Whittington et al. 2018), and modifications during pairwise coevolution to avoid inhibitor molecules of resistant prey (Holding, Biardi, et al. 2016; Margres, Bigelow, et al. 2017). Broadly, diet expansion appears possible through sequence variation derived from multiple possible pathways rather than any specific type of variation.

ACCEPTED MANUSCRIPT
Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

1 Importantly, the variation in modes of adaptions that we observed among different toxin
2 families and the differences in their contribution to a complex phenotype demonstrate genomic
3 heterogeneity in response to selective pressures associated with prey acquisition. In our study
4 SVMP, and SVSP toxins appear to influence diet breadth by the maximization of toxin sequence
5 and expression diversity. However, we did find some evidence of non-independence of these
6 mechanisms in SVSPs, where phylogenetic path analyses suggested direct interactions
7 between sequence diversity and expression diversity. Such a case may reflect scenarios where
8 differentially expressed toxins are experiencing differential rates of sequence evolution, or cases
9 where selection to increase expression leads to increased gene duplication and differentiation
10 (Kondrashov and Kondrashov 2006; Kondrashov 2012; Aird et al. 2015; Margres, Bigelow, et al.
11 2017).

12 In contrast, path analysis of PLA_2s showed no support for a sequence diversity mediated
13 relationship with diet. Rather, PLA_2s showed a strong positive relationship between mean
14 expression and diet breadth, suggesting that an investment in PLA_2 expression is associated
15 with increased prey diversity. Why PLA_2s exhibit this distinct relationship between diet and
16 expression is not clear, but one possibility is that it reflects broad functional efficacy of the same
17 proteins across diverse taxa. PLA_2s exhibit a wide-range of functional effects including muscular
18 and nervous system targeted neurotoxicity and myotoxicity (Gutiérrez and Lomonte 2013),
19 which may be less specialized, but similarly effective among phylogenetically distinct prey
20 groups. Thus, the role of PLA_2s in shaping diet diversity might be better described by a
21 mechanism whereby a given toxin or toxin family is broadly effective in a variety of scenarios at
22 the cost of being less effective at targeting specific diet items. Alternatively, PLA_2s may be
23 especially effective against taxonomic groups that tend to be or are exclusively associated with
24 broader diets, although evidence for this hypothesis is mixed and in need of further investigation
25 (Lomonte et al. 2009).

26 The functions and effects of CTL diversity on diets remain unclear, as we found no
27 evidence of an association between genetic variation and diet breadth in this toxin family. The
28 deviation of CTLs from other snake venom families is consistent with earlier tests comparing the
29 relationship between diet breadth and mRNA k-mer diversity among toxin families (Holding et al.
30 2021). Notably, CTLs are unique among snake venom toxins for functioning as multimeric
31 heterodimers, which could impose unique restrictions on their evolvability or decouple a direct
32 relationship between genetic and functional variation. (Arlinghaus and Eble 2012; Eble 2019).

33 In conclusion, our study demonstrates the power of combining high-resolution
34 transcriptomic datasets with comparative approaches to identify the molecular underpinnings of

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

key adaptations in phylogenetically diverse non-model and emerging-model organisms. Our findings suggest both sequence diversity in protein coding genes and how this diversity is regulated and ultimately expressed play key roles in mediating functional variation in components of venom, but that the role of these mechanisms is not ubiquitous for all components. Molecular traits like animal venoms, phytochemicals, and immune gene products are at the interface of antagonistic interactions among much of the planet's biodiversity. Our study demonstrates that the genomic pathways to adaptive variation in these systems is as multifaceted and complex as the phenotypes themselves.

Materials and Methods

Bioinformatic processing of transcriptomes

We assembled and annotated venom-gland transcriptomes for transcriptomes for 214 individuals from 68 rattlesnake and moccasin lineages used in Holding et al. (2021). All data processing was conducted using the Owens computing cluster at the Ohio Supercomputing Center (Center 1987). Briefly, raw sequence data were trimmed using TrimGalore! v.0.6.4 (Krueger 2015) and merged using PEAR v0.9.6 (Zhang et al. 2014). Merged reads were used to generate three transcriptome assemblies for each individual following the recommendations of (Holding et al. 2018). We used Trinity v.2.9.1 (Grabherr et al. 2011) with default settings, Seqman NGen 14 with default settings, and Extender v1.03 (Rokyta et al. 2012) with an overlap value of 120, a minimum seed quality of 30, a replicates value of 20, and minimum of 20 passes. These three assemblies were combined into a single master assembly and annotated with ToxCodAn (Nachtigall et al. 2021).

Annotated transcriptomes were subjected to several filters to reduce the inclusion of erroneously recovered transcripts. First, a custom python script, ChimeraKiller v.0.7.3 (https://github.com/masonaj157/ChimeraKiller) was used to filter out likely chimeric sequences based on the distribution of reads across each site in the coding region. Second, transcripts were filtered for incomplete coding regions and putatively premature stop codons. Third, we filtered out sequences with unreliable read coverage. These were defined as sequences with < 10x coverage for >10 % of the sequence. Finally, we removed transcripts from the four largest snake toxin families (CTLs, PLA_2s, SVMPs, and SVSPs) with transcript per million (TPM) estimates < 300, which may have been assembled due to barcode misassignment during sequencing. All python scripts used in transcriptome filtering steps are available on GitHub at https://github.com/masonaj157/Statistical_Analyses_For_Phylogenetic_Comparisons_of_North_American_Pitviper_Transcriptomes

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

ACCEPTED MANUSCRIPT

After filtering, transcripts were clustered at 98% similarity using cd-hit-est v.4.8.1 (Fu et al. 2012) to cluster alleles or very recent paralogs (Hofmann et al. 2018; Strickland et al. 2018). This represented the final transcriptome assembly for each sample. To estimate transcript expression, merged reads for each individual were mapped to their final transcriptome using Bowtie2 (Langmead and Salzberg 2012) as implemented in RSEM v.1.3.3 (Li and Dewey 2011). At this stage, we excluded one sample, *C. durissus* SB0275, from downstream analysis because it had an unusually low number of raw reads which resulted in a low-quality transcriptome assembly.

Using the final transcriptome and estimated expression, we calculated three metrics characterizing genetic sources of complexity in venom toxins: 1) gene content (GC), 2) toxin amino acid sequence diversity (SD), and 3) expression diversity (ED). We calculated GC of the transcriptome as the total number of unique toxin transcripts recovered in the final transcriptomes. We use GC as an estimate of the number of distinct loci present in a given sample. Because the venom phenotype's interaction with prey is a function of protein composition, we characterized toxin sequence diversity (SD) through amino acid 20-mer content. For each individual, we translated toxins, counted all unique 20-mers (script available on the project GitHub), and summarized amino acid diversity with Shannon's diversity index (H) converted to effective numbers of k-mers. We assume this measure captures overall functional diversity in protein coding sequencies present in a transcriptome. Finally, to estimate expression diversity (ED), we calculated Shannon's H per specimen treating toxins as 'individuals' and transcripts per million (TPM) as 'counts', which were converted to effective numbers of transcripts. For this measure of expression diversity, higher values represent more even expression across transcripts and therefore greater functional diversity. These metrics were then averaged among specimens belonging to the same lineage to attain lineage-level estimates that were used in subsequent analyses. Further detail on the calculation of each index is provided in the Supplementary Material (Supplemental Methods).

We assessed the possible influence of technical variation such as variation in sequencing effort and transcriptome completeness on toxin transcript recovery by testing for correlations between gene content versus the number of reads and the total numbers of expressed transcripts with linear models implemented with the lm function in R. To further ensure that these technical sources of variation did not affect downstream analyses through phylogenetic biases, we also tested for an interaction between lineage and either read numbers or total numbers of expressed transcripts on toxin gene content with two linear models

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

implemented in R and summarized with the 'Anova' function of the car v.3.0-10 package (Fox and Weisberg 2019).

We tested whether our calculated variables for venom diversity exhibited evidence of phylogenetic signal as was found for the whole venom phenotype by testing for significance of Blomberg's K and lambda, two common metrics of phylogenetic signal. Blomberg's K assesses the variance among species compared to expected variance under Brownian motion while lambda is a tree scaling parameter with an expected value of 0 if there is no correlation among species and 1 if correlation matches Brownian motion. For each variable, we assessed phylogenetic signal and tested for significant phylogenetic signal using the 'phylosig' function of phytools (Revell 2012) specifying either 'method=K' or 'method=lambda' and 'test=TRUE'.

Phylogenetic path analysis

To test for possible causal relationships between diet breadth and molecular sources of venom variation, we evaluated a range of phylogenetic path models for the 30 pitvipers with reliable diet information (Holding et al. 2021) using the R package *phylopath* (van der Bijl 2018). We tested 10 models representing different hypotheses regarding direct and indirect influences of gene content (GC), sequence diversity (SD), expression diversity (ED) (defined as above), and diet breadth (DB; as measured by standardized mean phylogenetic distance of prey – see Holding et al. 2021) (Supplementary Figure S7). We used mean phylogenetic distance (MPD) of prey as our measure of diet breadth because Holding et al. (2021) found this estimate of diet showed the strongest positive relationship to different measures of venom complexity likely because it incorporates information on function diversity in venom targets in prey. Values for this index have a positive relationship with diet breadth with higher values indicating broader diets. Generally, these models incorporate varying roles of sequence diversity and expression diversity as directly or indirectly predicting diet breadth, independently or in combination, while gene content acted indirectly through these variables. This framework, where venom variables predict diet breath is consistent with a hierarchical "genotype → phenotype → ecological-outcome" framework (Barrett and Hoekstra 2011) which models how species adapt to their environments. Importantly, this model allows the cumulative variation of gene content, sequence diversity, and expression diversity cumulatively to predict diet breadth, but should not be taken to imply directionality in in the venom-diet association (Supplemental Methods).

Because cumulative sequence and expression diversity are partially a function of what genes are expressed, they covary with one another. To account for this covariance, we included direct effects of gene content on sequence and expression diversity in all tested models. A model which only included the effects of gene content on sequence diversity and expression

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

ACCEPTED MANUSCRIPT

diversity but no relationship between SD and ED on diet diversity was used as the null model to account for any consistent correlation that is otherwise unrelated to diet (Supplementary Figure S7). Likewise, because the effect of differential gene content can only be realized in the venom phenotype through changes in toxin sequence diversity and/or expression, no models included a direct relationship between gene content and diet specialization.

Models were estimated under a lambda model of evolution and compared using the *C* statistic Information Criteria (CICc). The framework for CIC was proposed by Cardon et al. 2011 and has recently been established for use in phylogenetic path analysis (Hardenberg and Gonzalez-Voyer 2013; Voyer and Garamszegi 2014). Briefly, CICc is calculated using a model's *C* statistic, number of parameters, and a correction for small sample size (Voyer and Garamszegi 2014). Under this framework, models with the same numbers of variable relationships but different directionalities are expected to show similar statistical support, but their differing effect estimates may still be informative. Because a single model was not statistically preferred over all other models, we also estimated a weighted average model with weights determined from model likelihoods. All paths within comparably performing models (i.e., those within 2 CICc) were averaged. We also obtained confidence intervals for path coefficient estimates (partial regression coefficients standardized to the other independent variables) with 500 bootstraps. The parameters provided to the 'phylo_path' function were the predefined model set, the data frame of venom and diet breadth variables, the calibrated phylogeny, and the model specification "model=lambda". All other parameters were left as defaults.

In addition to performing phylogenetic path analysis for the overall venom dataset (all toxin classes combined), we also examined variation among patterns of evolution within four major toxin families: CTLs, PLA_2s, SVMPs, and SVSPs which represent major components of venom in these snakes (Holding et al. 2021). For each family, we restricted the dataset to toxins assigned to that family based on ToxCodAn annotation and estimated gene content, sequence diversity, and expression diversity. Each family was subsequently tested with phylogenetic path analysis using the same methods that had been applied to the whole dataset.

Variation in expression

Phylogenetic path analyses found counter-intuitive and conflicting results for the role of expression diversity at the whole venom level compared to what was recovered for the SVMP and SVSP families. Because expression diversity can be decomposed into the roles of richness (number of transcripts) and relative expression of each transcript, we hypothesized that differences in the number and expression of toxins in highly expressed toxin families would explain the trends observed in the path analyses. To assess how transcript numbers and

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022
ACCEPTED MANUSCRIPT

1 expression varied among large, highly expressed toxin families, we compared transcript
2 numbers and mean toxin expression in CTLs, PLA_2s, SVMPs, and SVSPs. We then tested for a
3 correlation between expression and diet breadth in these families to identify disproportionate
4 drivers of expression diversity.

5 First, to account for the compositional constraints of expression estimates, we performed
6 a centered log-ratio (CLR) transformation of TPM data for each individual. The CLR transformed
7 TPM values were then used in all subsequent comparisons of expression. We then calculated
8 the mean expression of transcripts in the CTL, PLA_2, SVMP, and SVSP families. For a few
9 samples, no toxins were recovered for a particular gene family (i.e., CTLs, PLA_2s, SVMPs, or
10 SVSPs) and their toxin numbers and expression values were encoded as NA. As a failure to
11 recover a toxin could occur because of stochastic variation in transcriptome assembly or our
12 conservative approach to toxin filtering, such samples were excluded from the analysis of that
13 gene family. To attain lineage-specific estimates we averaged the number of expressed
14 transcripts and mean expression of individuals in a phylogenetic lineage. We tested for overall
15 differences in numbers of expressed toxins and mean toxin expression among toxin families
16 with an ANOVA in R treating toxin family as the independent variable and lineage as a block
17 variable. Differences among treatments were tested with Bonferroni corrected post-hoc t-tests.
18 Finally, to determine if any variation in expression was associated with diet breadth, we tested
19 for relationships between diet breadth and mean toxin expression within each toxin family with a
20 phylogenetic linear regression implemented with phylolm v.2.6 (Ho and Ane 2014).

21 **Evolution of genetic diversity of SVMP & SVSP paralogs**

22 Our path analyses showed a direct relationship between toxin sequence diversity and diet
23 breadth. To explore how sequence diversity was generated at the gene level for these toxins we
24 used an approach proposed by Chang & Duda (2014) which uses community phylogenetics
25 indices to characterize the diversity of a toxin assemblage against the total diversity of a gene
26 family – in this case the total diversity of SVMP or SVSP paralogs observed in *Agkistrodon,*
27 *Crotalus,* and *Sistrurus*. As individual snakes normally express several SVMP and SVSP
28 paralogs, metrics such as standardized mean genetic distance (MGD) can be calculated for
29 each gene family in each individual. These indices identify where on a continuum that ranges
30 from high divergence between distinct paralogs to limited divergence between related paralogs
31 a given set of expressed transcripts falls. This permits an indirect but quantitative inference of
32 the evolutionary processes in terms of gene family and sequence evolution.

33 To conduct these analyses on our data, we first compiled translated mRNA sequences
34 for all recovered toxins in each family and generated a gene-family alignment using MUSCLE

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

v3.8.1551 (Edgar 2004). We then generated separate maximum-likelihood gene-family phylogenies for the SVMP and SVSP gene families using iqtree (Nguyen et al. 2015). Evolutionary models were selected for each family using iqtree's ModelFinder feature and we recovered branch support estimates with 1000 ultrafast bootstraps. These full gene family phylogenies represented the full diversity of SVMPs and SVSPs observed among all *Agkistrodon, Crotalus,* and *Sistrurus.* Using these two trees, we calculated standardized MGD for the SVMP and SVSP gene families for each individual using the *ses.mpd* function in the 'picante' package in R (Kembel et al. 2010). The resultant standardized MGD values represented the relative diversity of SVMP or SVSP paralogs expressed by a given individual compared to the total diversity of SVMP or SVSP paralogs in *Agkistrodon, Crotalus,* and *Sistrurus.* To account for the possible role of expression variation in altering realized diversity of toxin assemblages, we also calculated expression weighted standardized MGD using TPM values of each toxin as abundance estimates. Standardized and expression-weighted MGD values were then averaged across individuals for lineages with multiple representatives for lineage-level estimates of standardized MGD. Additional details on the calculation of MGD and weighted MGD are provided in the Supplemental Material (Supplemental Methods).

Using the standardized MGD values, we estimated whether expression weighting had a strong effect on altering diversity and we tested for a relationship between standardized MGD, sequence diversity, and diet breadth. We tested for differences between standardized and expression weighted MGD with a standard linear regression and R^2 estimate using the 'lm' function in R. Because distributions appeared multi-modal, we also tested whether each distribution was significantly different than 0 with a one-sided Wilcoxon signed rank test with the 'wilcox.text' function in R. To determine if the genetic diversity of toxin assemblages was associated with venom evolution, we then tested for relationships between standardized MGD and sequence diversity with phylogenetic linear regression using the 'phylolm' package in R.

Data availability

Raw sequence data used in this work are available on NCBI's SRR database under accession numbers provided in Holding et al. (2021). Likewise, metrics of phylogenetic diet complexity were collected from and are available in Holding et al. (2021). Copies of the input data files and R script used for data analysis are available on GitHub at: https://github.com/masonaj157/Statistical_Analyses_For_Phylogenetic_Comparisons_of_North_American_Pitviper_Transcriptomes

Acknowledgments

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

1 This work was funded by the National Science Foundation (DEB 1638872 to H.L.G, DEB
2 1638879 and DEB 1822417 to C.L.P., and DEB 1638902 to D.R.R). We thank Matthew Hahn
3 and Samarth Mathur for comments on the manuscript. We gratefully acknowledge the Ohio
4 Supercomputing Center which provided high-performance computing resources used in this
5 work.
6
7

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

1 **References**

2 Aird SD, Aggarwal S, Villar-Briones A, Tin MM-Y, Terada K, Mikheyev AS. 2015. Snake venoms
3 are integrated systems, but abundant venom proteins evolve more rapidly. *BMC Genomics*
4 16:1–20.

5 Aird SD, Arora J, Barua A, Qiu L, Terada K, Mikheyev AS. 2017. Population genomic analysis of
6 a pitviper reveals microevolutionary forces underlying venom chemistry. *Genome Biol.*
7 *Evol.* 9:2640–2649.

8 Arbuckle K. 2020. From molecules to macroevolution: Venom as a model system for
9 evolutionary biology across levels of life. *Toxicon X* 6:100034.

10 Arlinghaus FT, Eble JA. 2012. C-type lectin-like proteins from snake venoms. *Toxicon* 60:512–
11 519.

12 Barlow A, Pook CE, Harrison R a, Wüster W. 2009. Coevolution of diet and prey-specific venom
13 activity supports the role of selection in snake venom evolution. *Proc. R. Soc. B* 276:2443–
14 2449.

15 Barrett RDH, Hoekstra HE. 2011. Molecular spandrels: Tests of adaptation at the genetic level.
16 *Nat. Rev. Genet.* 12:767–780.

17 Barua A, Mikheyev AS. 2019. Many options, few solutions: Over 60 million years snakes
18 converged on a few optimal venom formulations.Russo C, editor. *Mol. Biol. Evol.*

19 Barua A, Mikheyev AS. 2020. Toxin expression in snake venom evolves rapidly with constant
20 shifts in evolutionary rates. *Proc. R. Soc. B Biol. Sci.* 287:20200613.

21 Besnard F, Picao-Osorio J, Dubois C, Félix MA. 2020. A broad mutational target explains a fast
22 rate of phenotypic evolution. *Elife* 9:1–70.

23 van der Bijl W. 2018. Phylopath: Easy phylogenetic path analysis in R. *PeerJ* 2018:e4718.

24 Cardon M, Loot G, Grenouillet G, Blanchet S. 2011. Host characteristics and environmental
25 factors differentially drive the burden and pathogenicity of an ectoparasite: A multilevel
26 causal analysis. *J. Anim. Ecol.* 80:657–667.

27 Carroll SB. 2005. Evolution at two levels: On genes and form. *PLoS Biol.* 3:1159–1166.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* 134:25–36.

Casewell NR, Huttley GA, Wüster W. 2012. Dynamic evolution of venom proteins in squamate reptiles. *Nat. Commun.* 3:1–10.

Casewell NR, Petras D, Card DC, Suranse V, Mychajliw AM, Richards D, Koludarov I, Albulescu L-O, Slagboom J, Hempel B-F, et al. 2019. *Solenodon* genome reveals convergent evolution of venom in eulipotyphlan mammals. *Proc. Natl. Acad. Sci. U. S. A.* 116:25745–25755.

Casewell NR, Wagstaff SC, Wüster W, Cook DAN, Bolton FMS, King SI, Pla D, Sanz L, Calvete JJ, Harrison RA. 2014. Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* 111:9205–9210.

Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG. 2013. Complex cocktails: The evolutionary novelty of venoms. *Trends Ecol. Evol.* 28:219–229.

Center OS. 1987. Ohio Supercomputer Center. Available from: http://osc.edu/ark:/19495/f5s1ph73

Chak STC, Harris SE, Hultgren KM, Jeffery NW, Rubenstein DR. 2021. Eusociality in snapping shrimps is associated with larger genomes and an accumulation of transposable elements. *Proc. Natl. Acad. Sci.* 118:e2025051118.

Chang D, Duda TF. 2012. Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. *Mol. Biol. Evol.* 29:2019–2029.

Chang D, Duda TF. 2014. Application of community phylogenetic approaches to understand gene expression: Differential exploration of venom gene space in predatory marine gastropods. *BMC Evol. Biol.* 14:123.

Davies E-L, Arbuckle K. 2019. Coevolution of snake venom toxic activities and diet: Evidence that ecological generalism favours toxicological diversity. *Toxins.* 11:711.

Duda TF, Palumbi SR. 1999. Molecular genetics of ecological diversification: Duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc. Natl. Acad. Sci. U.*

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

1 *S. A.* 96:6820–6823.

2 Duda TF, Palumbi SR. 2004. Gene expression and feeding ecology: Evolution of piscivory in the
3 venomous gastropod genus *Conus*. *Proc. R. Soc. London. Ser. B Biol. Sci.* 271:1165–
4 1174.

5 Duda TF, Remigio EA. 2008. Variation and evolution of toxin gene expression patterns of six
6 closely related venomous marine snails. *Mol. Ecol.* 17:3018–3032.

7 Eble JA. 2019. Structurally robust and functionally highly versatile—C-type lectin (-related)
8 proteins in snake venoms. *Toxins.* 11:136.

9 Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high
10 throughput. *Nucleic Acids Res.* 32:1792–1797.

11 Fox J, Weisberg S. 2019. An {R} Companion to Applied Regression. Third. Thousand Oaks
12 {CA}: Sage Available from: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

13 Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDAA, King GF, Nevalainen TJ,
14 Norman JA, Lewis RJ, Norton RS, et al. 2009. The toxicogenomic multiverse: Convergent
15 recruitment of proteins into animal venoms. *Annu. Rev. Genomics Hum. Genet.* 10:483–
16 511.

17 Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: Accelerated for clustering the next-generation
18 sequencing data. *Bioinformatics* 28:3150–3152.

19 Gendreau KL, Haney RA, Schwager EE, Wierschin T, Stanke M, Richards S, Garb JE. 2017.
20 House spider genome uncovers evolutionary shifts in the diversity and expression of black
21 widow venom proteins associated with extreme toxicity. *BMC Genomics* 18:178.

22 Gibbs HL, Chiucchi JE. 2011. Deconstructing a complex molecular phenotype: Population-level
23 variation in individual venom proteins in eastern massasauga rattlesnakes (*Sistrurus c.*
24 *catenatus*). *J. Mol. Evol.* 72:383–397.

25 Gibbs HL, Rossiter W. 2008. Rapid evolution by positive selection and gene gain and loss:
26 PLA2 venom genes in closely related *Sistrurus* rattlesnakes with divergent diets. *J. Mol.*
27 *Evol.* 66:151–166.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

1 Gibbs HL, Sanz L, Pérez A, Ochoa A, Hassinger ATB, Holding ML, Calvete JJ. 2020. The
2 molecular basis of venom resistance in a rattlesnake-squirrel predator-prey system. *Mol.*
3 *Ecol.* 29:2871–2888.

4 Giorgianni MW, Dowell NL, Griffin S, Kassner VA, Selegue JE, Carroll SB. 2020. The origin and
5 diversification of a novel protein family in venomous snakes. *Proc. Natl. Acad. Sci. U. S. A.*
6 117:10911–10920.

7 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
8 Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq
9 data without a reference genome. *Nat. Biotechnol.* 29:644–652.

10 Guignard MS, Crawley MJ, Kovalenko D, Nichols RA, Trimmer M, Leitch AR, Leitch IJ. 2019.
11 Interactions between plant genome size, nutrients and herbivory by rabbits, molluscs and
12 insects on a temperate grassland. *Proceedings. Biol. Sci.* 286:20182619.

13 Gutiérrez JM, Lomonte B. 2013. Phospholipases A2: Unveiling the secrets of a functionally
14 versatile group of snake venom toxins. *Toxicon* 62:27–39.

15 Haney RA, Clarke TH, Gadgil R, Fitzpatrick R, Hayashi CY, Ayoub NA, Garb JE. 2016. Effects
16 of gene duplication, positive selection, and shifts in gene expression on the evolution of the
17 venom gland transcriptome in widow spiders. *Genome Biol. Evol.* 8:228–242.

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

20 Ho LST, Ane C. 2014. A linear-time algorithm for Gaussian and non-Gaussian trait evolution
21 models. *Syst. Biol.* 63:397–408.

22 Hoekstra HE, Coyne JA. 2007. The locus of evolution: Evo devo and the genetics of adaptation.
23 *Evolution.* 61:995–1016.

24 Hofmann EP, Rautsaw RM, Strickland JL, Holding ML, Hogan MP, Mason AJ, Rokyta DR,
25 Parkinson CL. 2018. Comparative venom-gland transcriptomics and venom proteomics of
26 four Sidewinder Rattlesnake (*Crotalus cerastes*) lineages reveal little differential expression
27 despite individual variation. *Sci. Rep.* 8:15534.

28 Holding M, Margres M, Mason A, Parkinson C, Rokyta D, Holding ML, Margres MJ, Mason AJ,

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

1 Parkinson CL, Rokyta DR. 2018. Evaluating the performance of de novo assembly
2 methods for venom-gland transcriptomics. *Toxins.* 10:249.

3 Holding ML, Biardi JE, Gibbs HL. 2016. Coevolution of venom function and venom resistance in
4 a rattlesnake predator and its squirrel prey. *Proc. R. Soc. B Biol. Sci.* 283:20152841.

5 Holding ML, Drabeck DH, Jansa SA, Gibbs HL. 2016. Venom resistance as a model for
6 understanding the molecular basis of complex coevolutionary adaptations. *Integr. Comp.*
7 *Biol.* 56:1032–1043.

8 Holding ML, Strickland JL, Rautsaw RM, Hofmann EP, Mason AJ, Hogan MP, Nystrom GS,
9 Ellsworth SA, Colston TJ, Borja M, et al. 2021. Phylogenetically diverse diets favor more
10 complex venoms in North American pitvipers. *Proc. Natl. Acad. Sci. U. S. A.*
11 118:2015579118.

12 Hu Z, Sackton TB, Edwards S V, Liu JS. 2019. Bayesian detection of convergent rate changes
13 of conserved noncoding elements on phylogenetic trees. *Mol. Biol. Evol.* 36:1086–1100.

14 Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb
15 CO. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*
16 26:1463–1464.

17 Kondrashov FA. 2012. Gene duplication as a mechanism of genomic adaptation to a changing
18 environment. *Proc. R. Soc. B Biol. Sci.* 279:5048–5057.

19 Kondrashov FA, Kondrashov AS. 2006. Role of selection in fixation of gene duplications. *J.*
20 *Theor. Biol.* 239:141–151.

21 Krueger F. 2015. Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently
22 apply quality and adapter trimming to FastQ files.

23 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
24 9:357–359.

25 Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or
26 without a reference genome. *BMC Bioinformatics* 12:323.

27 Li M, Fry BG, Kini RM. 2005. Putting the brakes on snake venom evolution: The unique

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

1 molecular evolutionary patterns of *Aipysurus eydouxii* (Marbled Sea Snake) phospholipase
2 A 2 toxins. *Mol. Biol. Evol.* 22:934–941.

3 Li Q, Barghi N, Lu A, Fedosov AE, Bandyopadhyay PK, Lluisma AO, Concepcion GP, Yandell
4 M, Olivera BM, Safavi-Hemami H. 2017. Divergence of the venom exogene repertoire in
5 two sister species of *Turriconus*. *Genome Biol. Evol.* 9:2211–2225.

6 Lluisma AO, Milash BA, Moore B, Olivera BM, Bandyopadhyay PK. 2012. Novel venom
7 peptides from the cone snail *Conus pulicarius* discovered through next-generation
8 sequencing of its venom duct transcriptome. *Mar. Genomics* 5:43–51.

9 Lomonte B, Angulo Y, Sasa M, María Gutiérrez J. 2009. The phospholipase A 2 homologues of
10 snake venoms: Biological activities and their possible adaptive roles. *Protein Pept. Lett.*
11 16:860–876.

12 Margres MJ, Bigelow AT, Lemmon EM, Lemmon AR, Rokyta DR. 2017. Selection to increase
13 expression, not sequence diversity, precedes gene family origin and expansion in
14 rattlesnake venom. *Genetics* 206:1569–1580.

15 Margres MJ, Rautsaw RM, Strickland JL, Mason AJ, Schramer TD, Hofmann EP, Stiers E,
16 Ellsworth SA, Nystrom GS, Hogan MP, et al. 2021. The Tiger Rattlesnake genome reveals
17 a complex genotype underlying a simple venom phenotype. *Proc. Natl. Acad. Sci. U. S. A.*
18 118.

19 Margres MJ, Wray KP, Hassinger ATB, Ward MJ, McGivern JJ, Moriarty Lemmon E, Lemmon
20 AR, Rokyta DR. 2017. Quantity, not quality: Rapid adaptation in a polygenic trait
21 proceeded exclusively through expression differentiation. *Mol. Biol. Evol.* 34:3099–3110.

22 Margres MJ, Wray KP, Seavy M, McGivern JJ, Herrera ND, Rokyta DR. 2016. Expression
23 differentiation is constrained to low-expression proteins over ecological timescales.
24 *Genetics* 202:273–283.

25 Marigorta UM, Rodríguez JA, Gibson G, Navarro A. 2018. Replicability and prediction: Lessons
26 and challenges from GWAS. *Trends Genet.* 34:504–517.

27 Nachtigall PG, Rautsaw RM, Ellsworth SA, Mason AJ, Rokyta DR, Parkinson CL, Junqueira-de-
28 Azevedo ILM. 2021. ToxCodAn: A new toxin annotator and guide to venom gland

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

transcriptomics. *Brief. Bioinform.*

Nagy LG, Merényi Z, Hegedüs B, Bálint B. 2020. Novel phylogenetic methods are needed for understanding gene function in the era of mega-scale genome sequencing. *Nucleic Acids Res.* 48:2209–2219.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32:268–274.

Ohno S. 1970. Evolution by gene duplication. Springer Science & Business Media

Pease JB, Haak DC, Hahn MW, Moyle LC. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol.* 14:e1002379.

Pekár S, Bočánek O, Michálek O, Petráková L, Haddad CR, Šedo O, Zdráhal Z. 2018. Venom gland size and venom complexity—essential trophic adaptations of venomous predators: A case study using spiders. *Mol. Ecol.* 27:4257–4269.

Phuong MA, Mahardika GN, Alfaro ME. 2016. Dietary breadth is positively correlated with venom complexity in cone snails. *BMC Genomics* 17:1–15.

Raser JM, O'Shea EK. 2005. Noise in gene expression: Origins, consequences, and control. *Science.* 309:2010–2013.

Rausher MD, Delph LF. 2015. Commentary: When does understanding phenotypic evolution require identification of the underlying genes? *Evolution.* 69:1655–1664.

Rautsaw RM, Hofmann EP, Margres MJ, Holding ML, Strickland JL, Mason AJ, Rokyta DR, Parkinson CL. 2019. Intraspecific sequence and gene expression variation contribute little to venom diversity in sidewinder rattlesnakes (*Crotalus cerastes*). *Proc. R. Soc. B* 286:20190810.

Remigio EA, Duda TF. 2008. Evolution of ecological specialization and venom of a predatory marine gastropod. *Mol. Ecol.* 17:1156–1162.

Revell LJ. 2012. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3:217–223.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

1 Rockman M V. 2012. The QTN program and the alleles that matter for evolution: All that's gold
2 does not glitter. *Evolution.* 66:1–17.

3 Rokyta DR, Lemmon AR, Margres MJ, Aronow K. 2012. The venom-gland transcriptome of the
4 eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics* 13:312.

5 Rokyta DR, Margres MJ, Calvin K. 2015. Post-transcriptional mechanisms contribute little to
6 phenotypic variation in snake venoms. *G3* 5:2375–2382.

7 Rokyta DR, Wray KP, McGivern JJ, Margres MJ. 2015. The transcriptomic and proteomic basis
8 for the evolution of a novel venom phenotype within the Timber Rattlesnake (*Crotalus*
9 *horridus*). *Toxicon* 98:34–48.

10 Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, Gardner PP, Clarke JA, Baker
11 AJ, Clamp M, et al. 2019. Convergent regulatory evolution and loss of flight in
12 paleognathous birds. *Science.* 364:74–78.

13 Sanggaard KW, Bechsgaard JS, Fang X, Duan J, Dyrlund TF, Gupta V, Jiang X, Cheng L, Fan
14 D, Feng Y, et al. 2014. Spider genomes provide insight into composition and evolution of
15 venom and silk. *Nat. Commun.* 5:1–11.

16 Schield DR, Card DC, Hales NR, Perry BW, Pasquesi GM, Blackmon H, Adams RH, Corbin AB,
17 Smith CF, Ramesh B, et al. 2019. The origins and evolution of chromosomes, dosage
18 compensation, and mechanisms underlying venom regulation in snakes. *Genome Res.*
19 29:590–601.

20 Serrano SMT. 2013. The long road of research on snake venom serine proteinases. *Toxicon*
21 62:19–26.

22 Smith SD, Pennell MW, Dunn CW, Edwards S V. 2020. Phylogenetics is the new genetics (for
23 most of biodiversity). *Trends Ecol. Evol.* 35:415–425.

24 Stern DL, Orgogozo V. 2008. The loci of evolution: How predictable is genetic evolution?
25 *Evolution.* 62:2155–2177.

26 Strickland J, Mason A, Rokyta D, Parkinson C, Strickland JL, Mason AJ, Rokyta DR, Parkinson
27 CL. 2018. Phenotypic variation in Mojave Rattlesnake (*Crotalus scutulatus*) venom Is
28 driven by four toxin families. *Toxins.* 10:135.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

1 Sunagar K, Casewell NR, Varma S, Kolla R, Antunes A, Moran Y. 2014. Deadly innovations:
2 Unraveling the molecular evolution of animal venoms. In: Gopalakrishnakone P, Calvete
3 JJ, editors. Venom Genomics and Proteomics. Springer. p. 1–23.

4 Sunagar K, Moran Y. 2015. The rise and fall of an evolutionary innovation: Contrasting
5 strategies of venom evolution in ancient and young animals.Hoekstra HE, editor. *PLoS*
6 *Genet.* 11:1–20.

7 Sunagar K, Morgenstern D, Reitzel AM, Moran Y. 2016. Ecological venomics: How genomics,
8 transcriptomics and proteomics can shed new light on the ecology and evolution of venom.
9 *J. Proteomics* 135:62–72.

10 Tanksley SD. 1993. Mapping polygenes. *Annu. Rev. Genet.* 27:205–233.

11 Voyer AG, Garamszegi LZ. 2014. An introduction to phylogenetic path analysis. In: Modern
12 Phylogenetic Comparative Methods and their Application in Evolutionary Biology. Springer
13 Berlin Heidelberg. p. 201–229.

14 Whittington AC, Mason AJ, Rokyta DR. 2018. A single mutation unlocks cascading exaptations
15 in the origin of a potent pitviper neurotoxin. *Mol. Biol. Evol.* 35:887–898.

16 Wong ESWW, Belov K. 2012. Venom evolution through gene duplications. *Gene* 496:1–7.

17 Zancolli G, Calvete JJ, Cardwell MD, Greene HW, Hayes WK, Hegarty MJ, Herrmann HW,
18 Holycross AT, Lannutti DI, Mulley JF, et al. 2019. When one phenotype is not enough:
19 Divergent evolutionary trajectories govern venom variation in a widespread rattlesnake
20 species. *Proc. R. Soc. B Biol. Sci.* 286.

21 Zancolli G, Casewell NR. 2020. Venom systems as models for studying the origin and
22 regulation of evolutionary novelties. *Mol. Biol. Evol.*

23 Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: a fast and accurate Illumina Paired-
24 End reAd mergeR. *Bioinformatics* 30:614–620.

25

26

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

1 **Figure Legends**

2 **Figure 1.** Path analysis for models of venom evolution and diet breadth for an overall venom
3 model (a) and the CTL (b), PLA_2 (c), SVMP (d), and SVSP (e) toxin families. Path models test
4 for varying effects among gene content (GC), sequence diversity (SD), expression diversity
5 (ED), and diet breadth (DB) and are defined in Supplemental Figure S7. The barplots show
6 model weights for CICc comparisons. Numbers adjacent to the bars represent *p*-values for the
7 test of the null hypothesis that the model fits the data structure. Models with $p < 0.05$ are
8 statistically untenable. The best performing and averaged models are shown at the right with
9 path coefficients (partial regression coefficients standardized to the other independent variables)
10 indicated by numbers adjacent to arrows. Dashed lines in graphical models indicate negative
11 relationships. Averaged models were calculated based on model weighting of all top models
12 within 2 CICc.

13 **Figure 2.** Comparison of diet breadth and mean expression for CTLs (yellow), PLA_2s (red),
14 SVMPs (blue), and SVSPs (green). Mean expression is measured as center log-ratio
15 transformed TPM. Black dashed lines indicated lines of best fit inferred with phylogenetic linear
16 models. The red dashed line and corresponding red text in PLA_2s displays the line of best fit if
17 the outlying datapoint for *C. durissus* is excluded.

18 **Figure 3.** Graphical representation of how mean genetic distance (MGD) informs understanding
19 of patterns of gene family diversification. a) Three hypothetical lineages descending from a
20 common ancestor with differing patterns of gene diversification. Individual genes are shown as
21 colored circles on grey lines. b) Hypothetical gene family phylogeny derived from the three
22 lineages in a) and a representation of hypothetical MGD metrics based on the phylogeny.

23 **Figure 4.** Density distributions of standardized mean genetic distance (MGD) for the snake
24 venom metalloproteinase (SVMP; a) and snake venom serine protease (SVSP; b) gene
25 families. Correlations between expression weighted and unweighted standardized MGD are
26 shown as insets with *p*-values and R^2 values inferred by linear regression. Redlines shows the
27 fitted slopes and black lines show the one-to-one line.

28

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

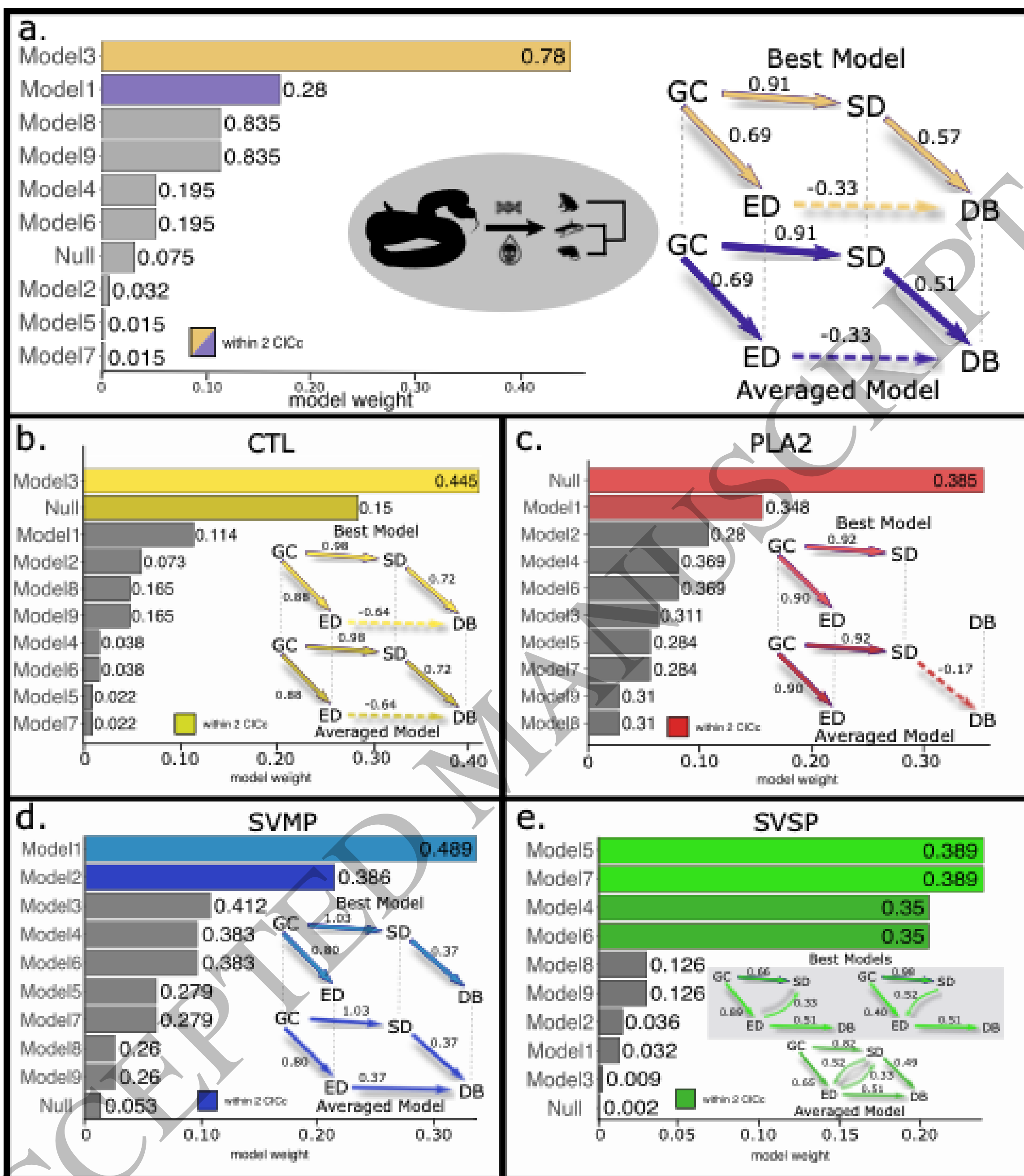
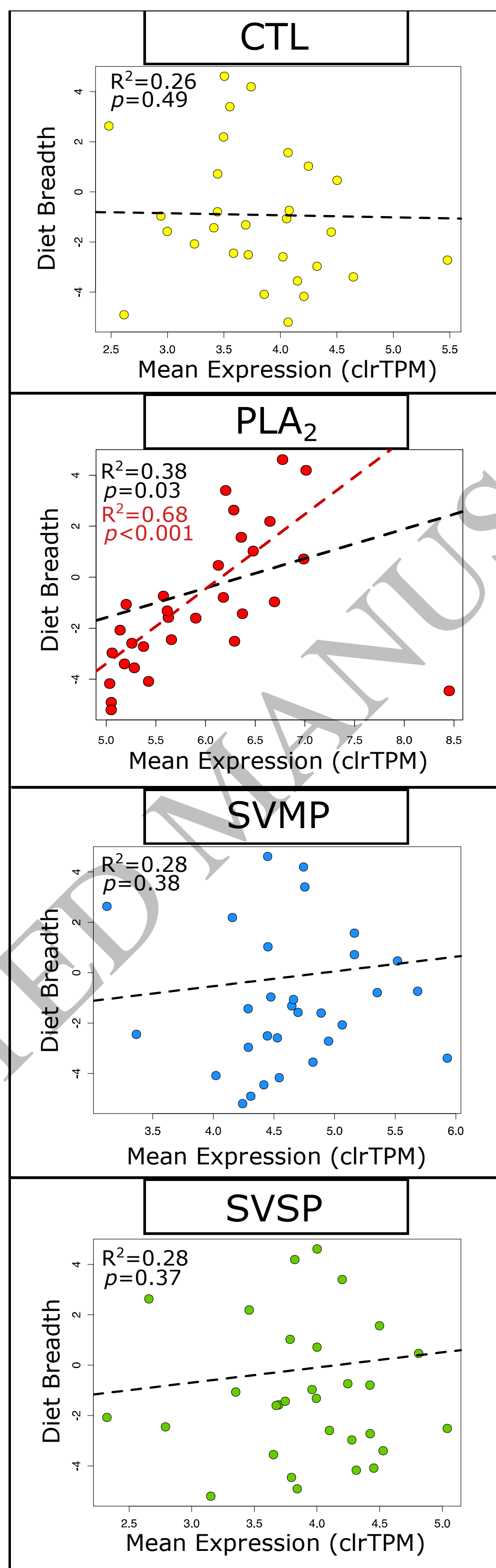


Figure 1

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

1

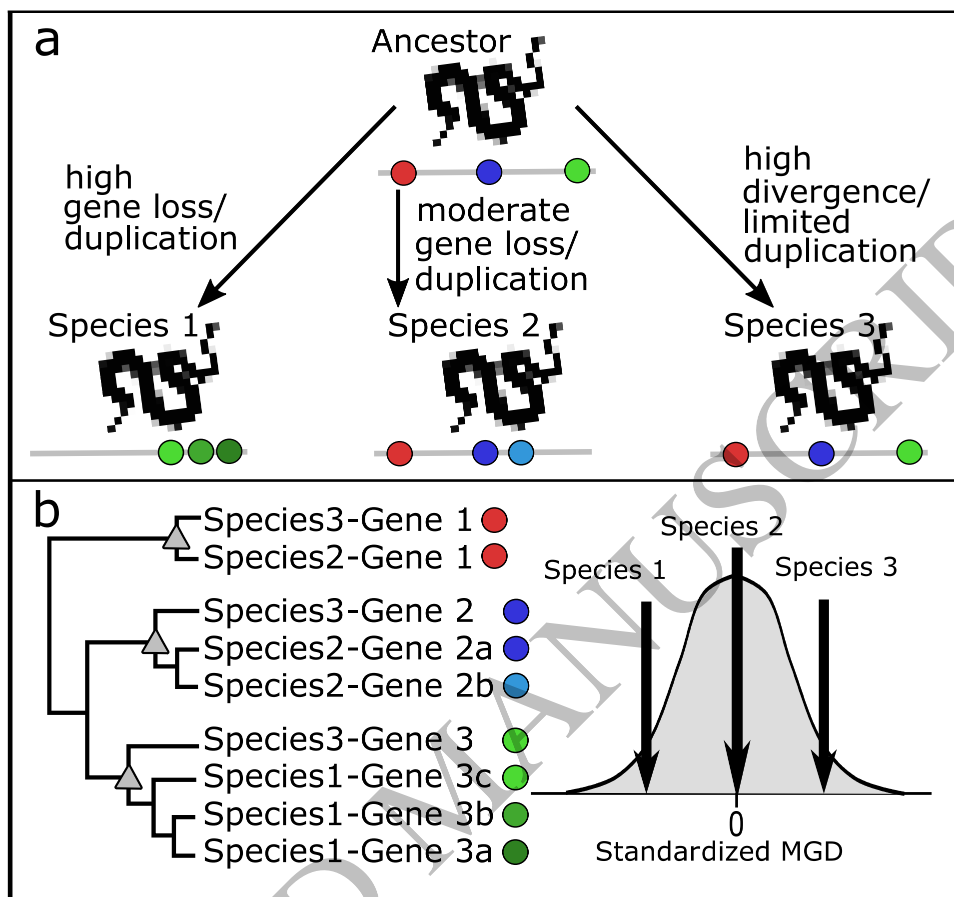


2 Figure 2

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022

1

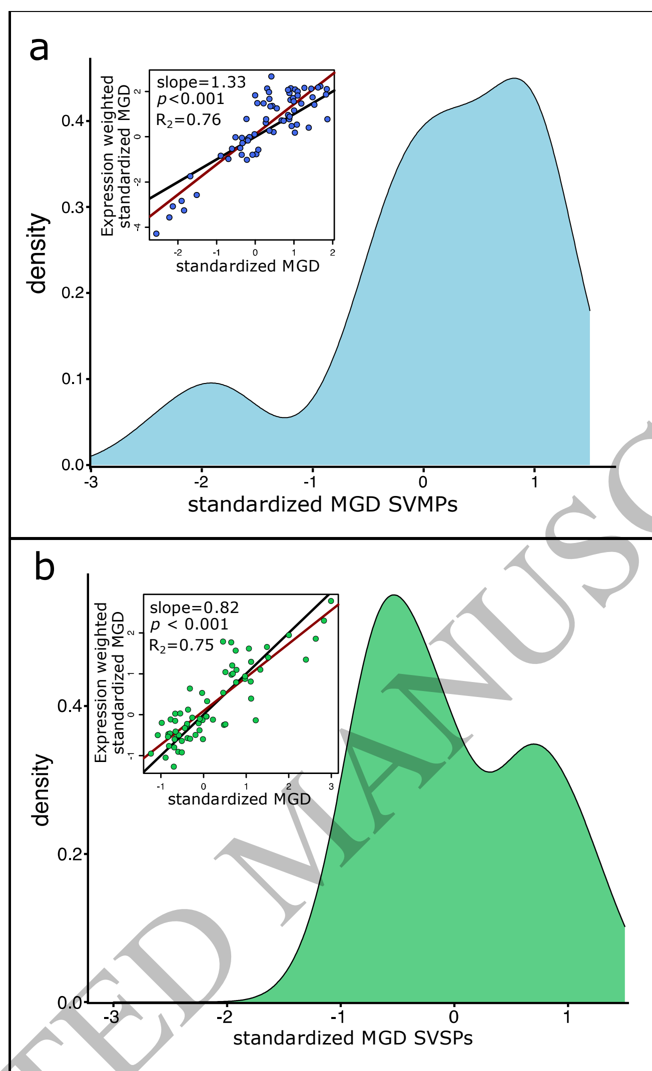


2

3 Figure 3

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022



1

2 Figure 4

3

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msac082/6567549 by Ohio State University Prior Health Sciences Library user on 13 April 2022