Main Manuscript for

Constraints on the evolution of toxin-resistant Na,K-ATPases have limited dependence on sequence divergence.

Shabnam Mohammadi[1,2,*], Lu Yang[3, §,*], Santiago Herrera-Álvarez[4,5,*], María del Pilar Rodríguez-Ordoñez[4,#], Karen Zhang[3], Jay F. Storz[1], Susanne Dobler[2], Andrew J. Crawford[4] & Peter Andolfatto[6]

[1]School of Biological Sciences, University of Nebraska, Lincoln, NE, USA
[2]Molecular Evolutionary Biology, Institute of Zoology, Universität Hamburg, Hamburg, Germany
[3]Department of Ecology and Evolution, Princeton University, Princeton, NJ, USA
[4]Department of Biological Sciences, Universidad de los Andes, Bogotá, 111711, Colombia
[5]Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA
[6]Department of Biological Sciences, Columbia University, New York, NY, USA

*Co-first authorship

[§] Current address: Wellcome Sanger Institute, Cambridge, United Kingdom
[#] Current address: Université Paris-Saclay Evry, Evry, France

Email: andrew@dna.ac, pa2543@columbia.edu

SM: 0000-0003-3450-6424, LY: 0000-0002-2694-1189, SHA: 0000-0002-0793-7811, MPRO: 0000-0002-0856-1297, KZ: 0000-0003-4406-9977, JFS: 0000-0001-5448-7924, SD: 0000-0002-0635-7719, AJC: 0000-0003-3153-6898, PA: 0000-0003-3393-4574

Classification
Biological Sciences; Evolution

Keywords
Epistasis, protein evolution, cardiotonic steroids, toxin resistance, adaptation

Author Contributions
PA and AJC conceived of and oversaw the project; SM, JFS, SD, AJC and PA designed experiments; KZ, LY, MPRO, SHA, SM collected data; SM, SHA and PA performed evolutionary and statistical analyses; SM, SHA, and PA wrote the paper; All authors edited the manuscript.

This PDF file includes:
Main Text
Figures 1 to 6

Abstract

A growing body of theoretical and experimental evidence suggests that intramolecular epistasis is a major determinant of rates and patterns of protein evolution and imposes a substantial constraint on the evolution of novel protein functions. Here, we examine the role of intramolecular epistasis in the case of the recurrent evolution of resistance to cardiotonic steroids (CTS) across diverse groups of tetrapods, which occurs via specific amino acid substitutions to the α-subunit family of Na,K-ATPases (ATP1A). First identifying a series of recurrent substitutions at two key sites of ATP1A that are predicted to confer CTS resistance in diverse tetrapods, we then performed protein engineering experiments to test the functional consequences of introducing these substitutions onto divergent species backgrounds. In line with previous results, we find that substitutions at these sites can have substantial background-dependent effects on CTS resistance. Globally, however, these substitutions also have pleiotropic effects that are consistent with additive rather than background-dependent effects. Moreover, the magnitude of a substitution's effect on activity does not depend on the overall extent of ATP1A sequence divergence between species. Our results suggest that epistatic constraints on the evolution of CTS-resistant forms of Na,K-ATPase likely depends on a small number of sites, with little dependence on overall levels of protein divergence. We propose that this dependence on a limited number sites may account for the observation of parallel CTS resistance substitutions observed among taxa with highly divergent Na,K-ATPases.

Significance Statement

Individual amino acid residues within a protein work in concert to produce a functionally coherent structure that must be maintained even as orthologous proteins in different species diverge over time. Given this dependence, we expect identical mutations to have more similar effects on protein function in more closely related species. We tested this hypothesis by performing protein-engineering experiments on ATP1A, an enzyme mediating target-site insensitivity to cardiotonic steroids (CTS) in diverse animals. These experiments reveal that that the phenotypic effects of substitutions can sometimes be background-dependent, but also that the magnitude of these phenotypic effects does not correlate with overall levels of ATP1A sequence divergence. Our results suggest that epistatic constraints are determined by states at a small number of sites, potentially explaining the frequent parallel CTS resistance substitutions among Na,K-ATPases of highly divergent taxa.

Main Text

Introduction

Instances of parallel and convergent evolution represent a useful paradigm to examine the factors that limit the rate of adaptation and the extent to which adaptive evolutionary paths are predictable (1, 2). The evolution of resistance to cardiotonic steroids (CTS) in animals represents one of the clearest examples of convergent molecular evolution. CTS are potent inhibitors of Na,K-ATPase (NKA, Fig. 1A), a protein that plays a critical role in maintaining membrane potential and is consequently vital for the maintenance of many physiological processes and signaling pathways in animals (3). CTS inhibit NKA function by binding to a highly conserved domain of the protein's α-subunit (ATP1A) and blocking the exchange of Na^+ and K^+ ions (3). NKA is thus often the target of parallel evolution of CTS resistance across widely divergent species, including insect herbivores that feed on toxic plants (4, 5) as well as predators that feed on toxic prey (6–10).

Patterns of parallelism and convergence in adaptive protein evolution are influenced by the mutational target size (i.e. the number of potentially adaptive mutations), the degree of pleiotropy (i.e. the effect of a given mutation on multiple phenotypes) and intramolecular epistasis (i.e. nonadditive interactions between mutant sites in the same protein) (11–19). If the phenotypic and fitness effects of mutations depend on the protein sequence background on which they arise (i.e. there is intramolecular epistasis), a given mutation is expected to have more similar phenotypic and fitness effects in orthologs from closely-related species. Therefore, the probability of parallel or convergent substitution is expected to decrease with increasing sequence divergence between orthologous proteins in different species. Consistent with this expectation, such a decline is observed in broad-scale phylogenetic comparisons of mitochondrial (20) and nuclear (21, 22) proteins. While these results suggest epistasis is an important global determinant of patterns of parallel and convergent protein evolution, studies linking these broad-scale observations with functional data at the level of individual proteins are lacking.

Functional investigations of CTS resistance-conferring substitutions at two key sites (111 and 122) in ATP1A orthologs of Drosophila (23, 24) and Neotropical grass frogs (25) revealed associated negative pleiotropic effects on protein function and showed that evolution at other sites in the protein mitigate these detrimental effects. In light of these pleiotropic and epistatic constraints, it is curious that parallel CTS-resistant substitutions are often observed among ATP1A orthologs of highly divergent species. Due to limited comparative functional data, the generality of pleiotropic

104 and epistatic constraints on ATP1A-mediated CTS resistance, and specifically the predicted
105 dependence on evolutionary distance, remain poorly understood.
106
107 To achieve a clearer picture of the phylogenetic distribution of CTS resistance substitutions, a more
108 complete and consistent sampling of ATP1A is needed in vertebrates. Broad phylogenetic
109 comparisons in vertebrates have primarily focused on the H1-H2 extracellular loop of ATP1A, which
110 represents only a subset of the CTS-binding domain. Further, most vertebrates possess three
111 paralogs of ATP1A that have different tissue-specific expression profiles and are associated with
112 distinct physiological roles (Fig. 1B) (3, 26). Previous studies of the ATP1A paralogs of vertebrate
113 taxa focused on ATP1A3 in reptiles (7, 8, 27, 28), ATP1A1 and/or ATP1A2 in birds and mammals
114 (10, 28), and either ATP1A1 or ATP1A3 in amphibians (6, 28). We therefore lack a comprehensive
115 and systematic survey of amino acid variation in the ATP1A protein family across vertebrates.
116
117 To bridge this gap, we first surveyed variation in near full-length coding sequences of the three
118 NKA α-subunit paralogs (ATP1A1, ATP1A2, ATP1A3) that are shared across major extant tetrapod
119 groups (mammals, birds, non-avian reptiles, and amphibians), and identified substitutions that
120 occur repeatedly among divergent lineages. Focusing on two key sites implicated in CTS resistance
121 across animals (111 and 122), we tested whether substitutions at these sites have increasingly
122 distinct phenotypic effects on more divergent genetic backgrounds. If the phenotypic effects of
123 these substitutions depend on states at a large number of sites throughout the protein, we expect
124 that identical substitutions should have increasingly distinct effects on more highly divergent
125 proteins. To test this, we engineered several common substitutions at sites 111 and 122 of ATP1A1
126 that differ between species to reveal potential 'cryptic' epistasis (16, 29). By quantifying the level of
127 CTS resistance conferred by these substitutions on different backgrounds, as well as their
128 pleiotropic effects on enzyme function, we evaluate the extent to which overall protein sequence
129 background has constrained the evolution of CTS-resistant forms of ATP1A1 across tetrapods.
130
131
132 **Results**
133
134 **Patterns of ATP1A sequence evolution across species and paralogs.**
135
136 To obtain a more comprehensive portrait of ATP1A amino acid variation among tetrapods, we
137 created multiple sequence alignments for near full-length ATP1A proteins for the three ATP1A
138 paralogs shared among vertebrates. In addition to publicly available data, we generated new RNA-
139 seq data for 27 non-avian reptiles (PRJNA754197) (Table S1-S2). We then *de novo* assembled

140 full-length transcripts of all ATP1A paralogs using these and RNA-seq data from 18 anuran species
141 (25) (PRJNA627222) to achieve better representation for these groups. In total, this dataset
142 comprises 429 species for ATP1A1, 197 species for ATP1A2 and 204 species for ATP1A3 (831
143 sequences total; Supplemental Dataset 1, Fig. S1).
144
145 Our survey reveals numerous substitutions at sites implicated in CTS resistance of NKA (Fig. 2;
146 Supplementary Dataset 2; for comparison to insects, see Supplemental file 1 of ref. (23)). As
147 anticipated from studies of full-length sequences in insects (4, 5, 23), most amino acid variation
148 among species and paralogs is concentrated in the H1-H2 extracellular loop (residues 111-122;
149 Fig 1A). Despite harboring just 28% of 43 sites previously implicated in CTS resistance (30), the
150 H1-H2 extracellular loop contains 81.4% of all substitutions identified among the three ATP1A
151 paralogs (Fig. S2).

152 Our survey reveals several clade- and paralog-specific patterns. Notably, ATP1A1 exhibits more
153 variation among species at sites implicated in CTS resistance (Fig. 2). Most of the variation in
154 ATP1A2 at these sites is restricted to squamate reptiles and ATP1A3 lacks substitutions at site 122
155 altogether, despite the well-known potential for substitutions at this site to confer CTS resistance
156 (25, 31). Looking across species and paralogs, the extent of parallelism at sites 111 and 122 is
157 remarkable (Figs. 2-3): for example, the substitutions Q111E, Q111T, Q111H, Q111L, and Q111V
158 all occur in parallel in multiple species of both insects and vertebrates. N122H and N122D also
159 frequently occur in parallel in both of these major clades. The frequent parallelism of CTS-sensitive
160 (i.e. Q111 and N122) to CTS-resistant states at these sites has been interpreted as evidence for
161 adaptive significance of these substitutions (4–7), but may also reflect mutational biases (32) and
162 the nature of physico-chemico constraints (21, 33).

163 In contrast, some parallelism is restricted to specific clades: for example, Q111R occurs in parallel
164 across tetrapods but has not been observed in insects. Similarly, the combination Q111R+N122D
165 has evolved three times independently in ATP1A1 of tetrapods but is not observed in insects.
166 Conversely, insects have evolved Q111V+N122H independently four times, but this combination
167 has never been observed in tetrapods. These patterns suggest that the fitness effects of some
168 CTS-resistant substitutions depend on genetic background (i.e. epistatic constraints), with the
169 result that CTS-resistance evolved via different mutational pathways in different lineages.

170 Beyond known CTS-resistant substitutions at sites 111 and 122, some taxa have evolved other
171 paths to CTS resistance. For example, the Pacman frog (genus *Ceratophyrs*) is known to prey on
172 CTS-containing toads (34) and its ATP1A1 harbors a known CTS-resistant substitution at site 121
173 (D121N, Supplementary Dataset 2). This substitution is rare among vertebrates but has been

174 previously reported in CTS-adapted milkweed bugs (4, 5). Similarly, the known CTS resistance
175 substitution C104Y is observed among many natricid snakes (Supplementary Dataset 2) and CTS-
176 adapted milkweed weevils (5). Chinchilla (*Chinchilla lanigera*) and yellow-throated sandgrouse
177 (*Pterocles gutturalis*) show distinct single-amino acid insertions in the H1–H2 extracellular loop, a
178 characteristic that has been previously associated with CTS resistance in pyrgomorphid
179 grasshoppers (30, 35). Further, *in lieu* of variation at site 122, ATP1A3 of tetrapods harbors
180 frequent parallel substitutions at site 120 (G120R, see also (7)). Interestingly, this site also shows
181 substantial parallel substitution in the ATP1A1 paralog of birds (where N120K occurs eight times
182 independently) but is mostly invariant in ATP1A1 of other tetrapods.

183

184 **Context-dependent CTS resistance of substitutions at sites 111 and 122**

185

186 The clade- and paralog-specific patterns of substitution among ATP1A paralogs outlined above
187 suggest that the evolution of CTS resistance may be highly dependent on sequence context.
188 However, the functional effects of the vast majority of these substitutions on the diverse genetic
189 backgrounds in which they occur remain largely unknown (25, 27, 31). Given the diversity and
190 broad phylogenetic distribution of parallel substitutions at sites 111 and 122, and the documented
191 effects of some of these substitutions on CTS resistance, we experimentally tested the extent to
192 which functional effects of substitutions at these sites are background-dependent.

193

194 We focused functional experiments on ATP1A1, because it is the most ubiquitously expressed
195 paralog and exhibits both the most sequence diversity and the broadest phylogenetic distribution
196 of parallel substitutions. Specifically, we considered ATP1A1 orthologs from nine representative
197 tetrapod species that possess different combinations of wild-type amino acids at 111 and 122 (Fig.
198 4A). Our taxon sampling includes two lizards, two snakes, two birds, two mammals and previously
199 published data for one amphibian (Fig. S5; Fig. S6; Table S3). The ancestral amino acid states of
200 sites 111 and 122 in tetrapods are Q and N, respectively. We found that the sum of the number of
201 derived states at positions 111 and 122 is a strong predictor of the level of CTS-resistance (Fig 4B,
202 IC_{50}, Spearman's r_S=0.85, p=0.001). Nonetheless, we also found greater than 10-fold variation in
203 CTS-resistance among enzymes that had identical paired states at 111 and 122 (e.g., compare
204 chinchilla (CHI) versus red-necked keelback snakes (KEE) or compare rat (RAT) versus the
205 resistant paralog of grass frogs (GRA_R)). These differences suggest that substitutions at other sites
206 also contribute to CTS resistance.

207

208 To test for epistatic effects of common CTS-resistant substitutions at sites 111 and 122, we used
209 site-directed mutagenesis to introduce 15 substitutions (nine at position 111 and six at position 122)

210 in the wildtype ATP1A1 backgrounds of 9 different species (Fig. S5). The specific substitutions
211 chosen were either phylogenetically broadly-distributed parallel substitutions and/or divergent
212 substitutions that distinguish closely related clades of species. We expressed each of these 24
213 ATP1A1 constructs with an appropriate species-specific ATP1B1 protein (Table S3). For each
214 recombinant NKA protein complex, we characterized its level of CTS resistance (IC_{50}) and
215 estimated enzyme activity as the rate of ATP hydrolysis in the absence of CTS (Table S4).
216
217 Among the 12 cases for which IC_{50} could be measured, substitutions had a 15-fold effect on
218 average (Fig. 4C, Table S4) and were equally likely to increase or decrease IC_{50}. To assess the
219 background-dependence of specific substitutions, we examined five cases in which a given
220 substitution (e.g., E111H), or the reverse substitution (e.g., H111E), could be evaluated on two or
221 more backgrounds. In the absence of intramolecular epistasis, the effect of a substitution on
222 different backgrounds should remain unchanged and the magnitude of the effect of the reverse
223 substitution should also be the same but with opposite sign. This analysis revealed substantial
224 background dependence for IC_{50} in two of the five informative cases (Fig. 4E; Table S5). In one
225 case, the N122D substitution results in a 200-fold larger increase in IC_{50} when added to the
226 chinchilla (CHI) background compared to the grass frog (GRA) background (p=1.2e-3 by ANOVA).
227 In the other case, the E111H substitution and the reverse substitution (H111E) produced effects in
228 the same direction (reducing CTS-resistance) when added to different backgrounds (false fer-de-
229 lance (FER) and red-necked keelback (KEE) snakes, respectively, p=1e-7 by ANOVA). Overall,
230 these results suggest that the effect of a given substitution on IC_{50} can be strongly dependent on
231 the background on which it occurs. The remaining three substitutions (H111T, Q111R and H122D)
232 showed no significant change in the magnitude of the effect on IC_{50} when introduced into different
233 species' backgrounds. These results suggest that, while some substitutions can have strong
234 background-dependent effects, strong intramolecular epistasis with respect to CTS resistance is
235 not universal.
236
237 **Pleiotropic effects on NKA activity exhibit no evidence for background-dependence.**
238
239 We next tested whether substitutions at sites 111 and 122 have pleiotropic effects on ATPase
240 activity. Because ion transport across the membrane is a primary function of NKA and its disruption
241 can have severe pathological effects (36), mutations that compromise this function are likely to be
242 under strong purifying selection. As suggested by previous work (23–25), CTS-resistant
243 substitutions at sites 111 and 122 can decrease enzyme activity. We evaluated the generality of
244 these effects across a broader phylogenetic scale by comparing enzyme activity of the 15 mutant
245 NKA proteins to their corresponding wild-type proteins.

246

247 Interestingly, the wild-type enzymes themselves exhibit substantial variation in activity, from 3-18
248 nmol/mg*min (*P* = 6e-7 by ANOVA, Fig 4D; Table S4). On average, substitutions at sites 111 and
249 122 on divergent orthologous protein backgrounds changed enzyme activity by 60% (Fig 4D; Fig
250 S4). In two cases, amino acid substitutions at position 122 (N122H and H122D) nearly inactivate
251 lizard NKAs and, in one case, a substitution at position 111 (Q111T) resulted in low expression of
252 the recombinant protein in the transfected cells (Fig S6; Fig. S7). A test of uniformity of pairwise t-
253 test p-values across substitutions suggests a significant enrichment of low p-values (Fig 4D inset;
254 p=2.5e-4, chi-squared test of uniformity). Thus, globally, this set of substitutions has significant
255 effects on NKA activity, but they were surprisingly not more likely to decrease than increase activity
256 (10 decrease: 5 increase, p>0.3, binomial test, Fig. 4D, Table S5).
257

258 We next asked to what extent pleiotropic effects of CTS resistant substitutions at positions 111 and
259 122 are dependent on genetic background. This question is motivated by recent studies in insects
260 which revealed that deleterious pleiotropic effects of some resistance-conferring substitutions at
261 sites 111 and 122 are background-dependent (23, 24). Likewise, recent work on ATP1A1 of toad-
262 eating grass frogs showed that effects of Q111R and N122D on NKA activity are background-
263 dependent (25). In contrast, among the five informative cases in which we compared the same
264 substitution (or the reverse substitution) on two or more backgrounds, there is no evidence for
265 background dependence (Fig 4E; Table S5). For example, N122D has similar effects on NKA
266 activity in grass frog and chinchilla despite the substantial divergence between the species' proteins
267 (8.4% protein sequence divergence; Fig. 4D). Similarly, the effects of Q111R in ostrich or the
268 reverse substitution R111Q in sandgrouse were not significantly different from the effect of Q111R
269 in grass frog (7.5% and 8% protein sequence divergence, respectively).
270

271 To further examine the evidence for background dependence, we tested whether changes to the
272 same amino acid state (regardless of starting state) at 111 and 122 produce different changes in
273 NKA activity (e.g., R111E on the rat background versus H111E on the false fer-de-lance
274 background). If epistasis is prevalent, involving a large number of sites, we expect that the
275 difference in effects of substitutions to a given amino acid state should increase with increasing
276 sequence divergence compared to ATP1A1 backgrounds in which that state is wild-type. However,
277 across the 11 possible comparisons, we found no relationship between the difference in the effect
278 of substitutions to the same state and the extent of amino acid divergence between the orthologous
279 proteins (Fig. 5). This pattern suggests that, while pleiotropic effects can be background dependent
280 (23, 25), these effects are not pervasive across species and do not correlate with overall sequence
281 divergence.

The overall rate of convergence across ATP1A proteins does not depend on sequence divergence.

If intramolecular epistasis is pervasive, rates of convergent substitution are predicted to decrease as a function of overall sequence divergence (20–22). In contrast to this expectation, our experiments suggest that, for ATP1A1, the extent of background sequence divergence is a poor predictor of the magnitude of effects of substitutions at sites 111 and 122 on CTS resistance and enzyme activity. Since our experiments were necessarily limited in scope, we carried out a broad phylogenetic analysis to evaluate how well our findings align with global estimates of rates of convergence for the ATP1A family beyond ATP1A1 and beyond sites implicated in CTS resistance.

Using a multisequence alignment of 831 ATP1A protein sequences, including the three ATP1A paralogs shared among tetrapods (i.e., amphibians, non-avian reptiles, birds, and mammals), we inferred a maximum likelihood phylogeny of the gene family (Fig. S1). We then used ancestral sequence reconstruction to infer the history of substitution events on all branches in the tree and counted the number of convergent amino acid substitutions per site along the entire protein (see Materials and Methods). Convergent substitutions are defined as substitutions on two branches at the same site resulting in the same amino acid state. Interestingly, we do not detect a correlation between the relative number of convergent substitutions with background ATP1A divergence across the tree (Fig. 6A). This result also holds true when considering only substitutions to the key CTS resistance sites 111 and 122 (Fig. S6).

To gain more insight into the factors that determine convergent evolution in ATP1A, we looked more closely at patterns of individual convergent substitutions at sites 111 and 122 by extracting each convergent substitution and visualizing its distribution along the sequence divergence axis (Fig. 6B). Under the expectation that rates of convergence should tend to decrease as a function of sequence divergence, the distribution of pairwise convergent events along the sequence divergence axis should be left-skewed, with a peak towards lower sequence divergence. In contrast to this expectation, the distribution is bimodal, with one peak at 0.33 and the other at 0.69 substitutions/site (Fig. 6B bottom panel). Parallel and convergent substitutions have occurred almost across the full range of protein divergence estimates. For example, if X is any starting state, the substitution X111R has occurred independently in 13 tetrapod lineages and X111L independently in 20 lineages. Both substitutions have a broad phylogenetic distribution, suggesting that their effects do not strongly depend on sequence states at a large number of sites throughout

318 the protein. Interestingly, however, the distributions for X111H and X111E substitutions are
319 relatively right-skewed, in line with epistasis for CTS resistance that we observed in experiments
320 for H111E/E111H (Fig 4E).
321
322
323
324 **Discussion**
325
326 Previous work has shown that the adaptive evolution of NKA-mediated CTS resistance in animals
327 is constrained by pleiotropy and epistasis (i.e. background-dependence) (23–25, 30, 37). Further,
328 based on broad phylogenetic analysis of proteomes, our *a priori* expectation is that epistasis should
329 represent a stronger constraint with increasing levels of divergence between species and ATP1A
330 paralogs (20). In light of these considerations, our broad survey of the ATP1A gene family in
331 tetrapods reveals two striking and seemingly contradictory patterns. The first is that some
332 substitutions underlying CTS resistance in tetrapods are broadly distributed phylogenetically and
333 even shared with insects (e.g., N122H is widespread among snakes and found in the monarch
334 butterfly and other insects; see Fig. 3 for more examples). Patterns like these suggest that epistatic
335 constraints have a limited role in the evolution of CTS resistance, as the same mutation can be
336 favored on highly divergent genetic backgrounds. On the other hand, there is also substantial
337 diversity in resistance-conferring states at sites 111 and 122, and some combinations of these
338 substitutions appear to be phylogenetically restricted. For example, the CTS-resistant combination
339 of Q111R+N122D has evolved multiple times in tetrapods but is absent in insects, whereas the
340 CTS-resistant combination Q111V+N122H evolved multiple times in insects but is absent in
341 tetrapods (Fig 3). Additionally, some substitutions also appear to be paralog-specific in tetrapods
342 (Fig 3). These phylogenetic signatures suggest at least some role for epistasis as a source of
343 contingency in the evolution of ATP1A-mediated CTS resistance in animals (i.e., that the fitness
344 effects of substitutions depend on the order in which they occur). How can these disparate patterns
345 be reconciled? To what extent do genetic background and contingency limit the evolution of CTS
346 resistance in animals?
347
348 In our functional analysis of diverse ATP1A1 proteins, we find that derived substitutions at sites
349 111 and 122 have largely predictable effects on CTS resistance, with salient exceptions that tend
350 to be in size rather than direction (Fig. 4). For example, Q111R contributes to CTS resistance on
351 many species' backgrounds, but not on that of sandgrouse (Fig. 4E). It is also notable that species
352 with identical paired states at 111 and 122 can vary in CTS resistance by more than an order of
353 magnitude (Fig 4B). Together, these patterns point to background determinants of CTS resistance

354 that may be additive rather than epistatic. Despite this, there are some broadly phylogenetically
355 distributed substitutions, such as N122D, that nonetheless do exhibit background-dependent
356 effects on CTS resistance (Fig. 4C and 4E).
357
358 With respect to pleiotropic effects on NKA activity, our functional analysis of substitutions at sites
359 111 and 122 on diverse ATP1A1 backgrounds suggest that interactions between these
360 substitutions and those backgrounds are largely additive. Specifically, we find that the effect of a
361 particular CTS-resistance substitution on NKA activity is no more severe if added to a protein
362 background that is ~49 or 86 amino acid substitutions away from the protein background on which
363 that substitution naturally occurs (Fig. 5). In light of previous results demonstrating background-
364 dependent effects of similar substitutions on NKA activity (23–25), our findings suggest that the
365 extent of epistasis does not have a monotonic dependence on the extent of ATP1A1 divergence.
366 Our findings further support increasing evidence that while epistasis is likely to be a pervasive
367 feature of protein evolution, many mutational effects on structural and functional properties of
368 proteins nonetheless seem to be additive (e.g., (38–40)).
369
370 We propose that our observations can be reconciled with previous results demonstrating epistatic
371 constraints if epistasis with respect to protein function is confined to a small number of sites in the
372 protein. If so, we might expect that the magnitude of epistasis may have little dependence on the
373 extent of protein-wide ATP1A1 divergence but would instead be better predicted by divergence at
374 a few key sites. In support of this view, Mohammadi et al. (25) showed that decreases in ATP1A1
375 enzyme activity due to substitutions Q111R and N122D can be rescued by 10 (or fewer) of the 19
376 amino acid differences distinguishing the backgrounds of CTS-resistant and sensitive ATP1A1
377 paralogs of grass frogs. Further, studies in *Drosophila melanogaster* (23, 24) show that severe
378 neural dysfunction associated with CTS resistance substitutions at sites 111 and 122 can largely
379 be rescued with one additional substitution (A119S) to the *Drosophila melanogaster* ATP1A
380 background.
381
382 It is worth noting that phenotypes like enzyme activity are not equivalent to organismal fitness and
383 that there may be a complex mapping between the two. The discussion above assumes that
384 changes in enzyme activity are most likely detrimental to organismal fitness, but this need not be
385 the case. Interestingly, we found that the activity of wildtype NKAs vary 6-fold among the species
386 surveyed (Fig. 4E), suggesting that most species are either robust to changes in NKA activity or
387 that changes have occurred in other genes (including other ATP1A paralogs) to compensate for
388 changes in NKA activity associated with ATP1A1. Thus, it may be that protein activity itself is either
389 not an important pleiotropic constraint on the evolution of CTS resistance of NKA or that constraint

390 depends not just on the protein background, but also on the broader genetic background of the
391 organism (e.g., other interacting proteins; see (23)).
392
393 Our work highlights the utility of comparative functional work in understanding the nature of epistatic
394 constraints on the evolution of novel protein functions. In this case-study of the evolution of CTS-
395 resistant Na,K-ATPases, we find that epistatic constraints are more likely depend on divergence at
396 a small number of key sites in the protein rather than overall levels of protein divergence. The
397 circumscribed nature of these constraints may account for the remarkable parallelism of CTS-
398 resistance substitutions observed among the Na,K-ATPases of highly divergent species.
399
400
401

402 **Materials and Methods**

403

404 **Sample collection and data sources.**

405 In order to carry out a comprehensive survey of vertebrate ATP1A paralogs (ATP1A1, ATP1A2 and
406 ATP1A3), we collated a total of 831 protein sequences for this study (Supplementary Dataset 1).
407 Mammals possess a fourth paralog (ATP1A4) that is expressed predominantly in testes (41) that
408 we don't consider here. In addition to publicly available data, we also generated RNA-seq data for
409 27 species of non-avian reptiles (Table S1; PRJNA754197) to achieve a better representation of
410 some previously underrepresented lineages. These included field-caught and museum-archived
411 specimens as well as animals purchased from commercial pet vendors. Purchased animals were
412 processed following the procedures specified in the IACUC Protocol No. 2057-16 (Princeton
413 University) and implemented by a research veterinarian at Princeton University. Wild-caught
414 animals were collected under Colombian umbrella permit *resolución No*. 1177 granted by the
415 *Autoridad Nacional de Licencias Ambientales* to the Universidad de los Andes and handled
416 according to protocols approved by the Institutional Committee on the Care and Use of Laboratory
417 Animals (abbreviated CICUAL in Spanish) of the Universidad de los Andes. In all cases, fresh
418 tissues (brain, stomach, and muscle) were taken and preserved in RNAlater (Invitrogen) and stored
419 at -80ºC until used.

420

421 **Reconstruction of ATP1A paralogs.**

422 RNA-seq libraries were prepared either using TruSeq RNA Library Prep Kit v2 (Illumina) and
423 sequenced on Illumina HiSeq2500 (Genomics Core Facility, Princeton, NJ, USA) or using NEBNext
424 Ultra RNA Library Preparation Lit (NEB) and sequenced on Illumina HiSeq4000 (Genewiz, South
425 Plainfield, NJ, USA) (Table S2). All raw RNA-seq data generated for this study have been deposited
426 in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under
427 bioproject PRJNA754197. Together with SRA datasets downloaded from public database, reads
428 were trimmed to Phred quality ≥ 20 and length ≥ 20 and then assembled *de novo* using Trinity
429 v2.2.0 (42). Sequences of ATP1A paralogs 1, 2 and 3 were pulled out with BLAST searches (blast-
430 v2.26), individually curated, and then aligned using ClustalW. Complete alignments of ATP1/2/3
431 can be found in Supplementary Dataset 1.

432

433 **Character state mapping and parameter estimation for the ATP1A1-3 paralogs**

434 Protein sequences from ATP1A1 (N=429), ATP1A2 (N=197) and ATP1A3 (N=205) including main
435 tetrapod groups (amphibians, non-avian reptiles, birds, and mammals) and lungfish+coelacanth as
436 outgroups were aligned using ClustalW with default parameters. The optimal parameters for
437 phylogenetic reconstruction were taken from the best-fit amino acid substitution model based on
438 Akaike Information Criterion (AIC) as implemented in ModelTest-NG v.0.1.5 (43), and was inferred

439 to be JTT+G4+F. An initial phylogeny was inferred using RAxML HPC v.8 (44) under the
440 JTT+GAMMA model with empirical amino acid frequencies. Branch lengths and node support
441 (aLRS) were further refined using PhyML v.3.1 (45) with empirical amino acid frequencies and
442 maximum likelihood estimates of rate heterogeneity parameters, I and Γ. Phylogeny visualization
443 and mapping of character states for each paralog was done using the R package ggtree (46).
444
445 **Ancestral sequence reconstruction and convergence calculations**
446 Ancestral sequence reconstruction (ASR) was performed in PAML using codeml (47) under the
447 JTT+G4+F substitution model. Statistical confidence in each position's reconstructed state for each
448 ancestor was determined from the posterior probability (PP), and only states with PP>0.8 were
449 considered. Ancestral sequences from all nodes in the ATP1A phylogeny were retrieved from the
450 codeml output, resulting in an alignment of 1,660 ATP1A proteins (831 extant species and 829
451 inferred ancestral sequences; Fig. S1). For each branch in the tree, we determined the occurrence
452 of substitutions by using the ancestral and derived amino acid states at each site using only states
453 with PP>0.8. All branch pairs were compared, except sister branches and ancestor-descendent
454 pairs (20, 21). When comparing substitutions on two distinct branches at the same site,
455 substitutions to the same amino acid state were counted as convergences, while substitutions away
456 from a common amino acid were counted as divergences. An alignment of 1,040 amino acids was
457 used to calculate the number of molecular convergences and divergences, excluding a putative 30
458 amino acid-long alternative spliced region (positions 834-864). Model-based estimates of sequence
459 divergence, number of convergences, number of divergences, and total number of substitutions
460 since the common ancestor were recorded for each pairwise comparison. We calculated the
461 proportion of observed convergent events per branch as (number of convergences +1) / (number
462 of divergences +1). The line describing the trend was calculated as a running average with window
463 size of 0.05 substitutions/site. 95% confidence intervals were calculated based on 100 bootstrap
464 replicates per window, resampling only variable sites.
465
466 For sites 111 and 122, molecular convergence was coded as "1" when the substitution along
467 $branch_i$ was to the same amino acid state as the substitution along $branch_j$, and "0" if substitutions
468 were to different states. Model-based estimates of sequence divergence, amino acid state, and
469 convergence event were recorded for each pairwise comparison (when convergence was "0",
470 amino acid state was set to "NA"). A logistic regression between molecular convergence (0 or 1)
471 and genetic distance was used to test for the correlation between variables (Fig 6B; Fig. S4)
472
473
474

475 **Construction of expression vectors**.
476 ATP1A1 and ATP1B1 wild-type sequences for the eight selected tetrapod species (Fig 4) were
477 synthesized by Invitrogen™ GeneArt. The β1-subunit genes were inserted into pFastBac Dual
478 expression vectors (Life Technologies) at the p10 promoter with XhoI and PaeI (FastDigest Thermo
479 Scientific™) and then control sequenced. The α1-subunit genes were inserted at the PH promoter
480 of vectors already containing the corresponding β1-subunit proteins using In-Fusion® HD Cloning
481 Kit (Takara Bio, USA Inc.) and control sequenced. All resulting vectors had the α1-subunit gene
482 under the control of the PH promoter and a β1-subunit gene under the p10 promoter. The resulting
483 eight vectors were then subjected to site-directed mutagenesis (QuickChange II XL Kit; Agilent
484 Technologies, La Jolla, CA, USA) to introduce the codons of interest. In total, 21 vectors were
485 produced (Table S3).
486
487 **Generation of recombinant viruses and transfection into Sf9 cells.**
488 *Escherichia coli* DH10bac cells harboring the baculovirus genome (bacmid) and a transposition
489 helper vector (Life Technologies) were transformed according to the manufacturer's protocol with
490 expression vectors containing the different gene constructs. Recombinant bacmids were selected
491 through PCR screening, grown, and isolated. Subsequently, Sf9 cells (4 x 10^5 cells*ml) in 2 ml of
492 Insect-Xpress medium (Lonza, Walkersville, MD, USA) were transfected with recombinant bacmids
493 using Cellfectin reagent (Life Technologies). After a three-day incubation period, recombinant
494 baculoviruses were isolated (P1) and used to infect fresh Sf9 cells (1.2 x 10^6 cells*ml) in 10 ml of
495 Insect-Xpress medium (Lonza, Walkersville, MD, USA) with 15 mg/ml gentamycin (Roth, Karlsruhe,
496 Germany) at a multiplicity of infection of 0.1. Five days after infection, the amplified viruses were
497 harvested (P2 stock).
498
499 **Preparation of Sf9 membranes.**
500 For production of recombinant NKA, Sf9 cells were infected with the P2 viral stock at a multiplicity
501 of infection of 10^3. The cells (1.6 x 10^6 cells*ml) were grown in 50 ml of Insect-Xpress medium
502 (Lonza, Walkersville, MD, USA) with 15 mg/ml gentamycin (Roth, Karlsruhe, Germany) at 27°C in
503 500 ml flasks (35). After 3 days, Sf9 cells were harvested by centrifugation at 20,000 x g for 10 min.
504 The cells were stored at -80 °C and then resuspended at 0 °C in 15 ml of homogenization buffer
505 (0.25 M sucrose, 2 mM EDTA, and 25 mM HEPES/Tris; pH 7.0). The resuspended cells were
506 sonicated at 60 W (Bandelin Electronic Company, Berlin, Germany) for three 45 s intervals at 0 °C.
507 The cell suspension was then subjected to centrifugation for 30 min at 10,000 x g (J2-21 centrifuge,
508 Beckmann-Coulter, Krefeld, Germany). The supernatant was collected and further centrifuged for
509 60 m at 100,000 x g at 4 °C (Ultra- Centrifuge L-80, Beckmann-Coulter) to pellet the cell
510 membranes. The pelleted membranes were washed once and resuspended in ROTIPURAN® p.a.,

511 ACS water (Roth) and stored at -20 °C. Protein concentrations were determined by Bradford assays
512 using bovine serum albumin as a standard. Three biological replicates were produced for each
513 NKA construct.
514
515 **Verification by SDS-PAGE/western blotting.**
516 For each biological replicate, 10 μg of protein were solubilized in 4x SDS-polyacrylamide gel
517 electrophoresis sample buffer and separated on SDS gels containing 10% acrylamide.
518 Subsequently, they were blotted on nitrocellulose membrane (HP42.1, Roth). To block non-specific
519 binding sites after blotting, the membrane was incubated with 5% dried milk in TBS-Tween 20 for
520 1 h. After blocking, the membranes were incubated overnight at 4 °C with the primary monoclonal
521 antibody α5 (Developmental Studies Hybridoma Bank, University of Iowa, Iowa City, IA, USA).
522 Since only membrane proteins were isolated from transfected cells, detection of the α subunit also
523 indicates the presence of the β subunit. The primary antibody was detected using a goat-anti-
524 mouse secondary antibody conjugated with horseradish peroxidase (Dianova, Hamburg,
525 Germany). The staining of the precipitated polypeptide-antibody complexes was performed by
526 addition of 60 mg 4-chloro-1 naphtol (Sigma-Aldrich, Taufkirchen, Germany) in 20 ml ice-cold
527 methanol to 100 ml phosphate buffered saline (PBS) containing 60 μl 30% H_2O_2. See Fig. S7.
528
529 **Ouabain inhibition assay.**
530 To determine the sensitivity of each NKA construct against cardiotonic steroids (CTS), we used the
531 water-soluble cardiac glycoside, ouabain (Acrōs Organics), as our representative CTS. 100 ug of
532 each protein was pipetted into each well in a nine-well row on a 96-well microplate (Fisherbrand)
533 containing stabilizing buffers (see buffer formulas in (48)). Each well in the nine-well row was
534 exposed to exponentially decreasing concentrations (10^{-3} M, 10^{-4} M, 10^{-5} M, 10^{-6} M, 10^{-7} M, 10^{-8} M,
535 dissolved in distilled H_2O) of ouabain, distilled water only (experimental control), and a combination
536 of an inhibition buffer lacking KCl and 10^{-2} M ouabain to measure background protein activity (48).
537 The proteins were incubated at 37°C and 200 rpms for 10 minutes on a microplate shaker
538 (Quantifoil Instruments, Jena, Germany). Next, ATP (Sigma Aldrich) was added to each well and
539 the proteins were incubated again at 37°C and 200 rpms for 20 minutes. The activity of NKA
540 following ouabain exposure was determined by quantification of inorganic phosphate (Pi) released
541 from enzymatically hydrolyzed ATP. Reaction Pi levels were measured according to the procedure
542 described in Taussky and Shorr (49) (see Petschenka et al. (48)). All assays were run in duplicate
543 and the average of the two technical replicates was used for subsequent statistical analyses.
544 Absorbance for each well was measured at 650 nm with a plate absorbance reader (BioRad Model
545 680 spectrophotometer and software package). See Table S4.
546

547 **ATP hydrolysis assay.**

548 To determine the functional efficiency of different NKA constructs, we calculated the amount of Pi
549 hydrolyzed from ATP per mg of protein per minute. The measurements were obtained from the
550 same assay as described above. In brief, absorbance from the experimental control reactions, in
551 which 100 μg of protein was incubated without any inhibiting factors (i.e., ouabain or buffer
552 excluding KCl), were measured and translated to mM Pi from a standard curve that was run in
553 parallel (1.2 mM Pi, 1 mM Pi, 0.8 mM Pi, 0.6 mM Pi, 0.4 mM Pi, 0.2 mM Pi, 0 mM Pi). See Table
554 S4.

555

556 **Statistical analyses of functional data.**

557 Background phosphate absorbance levels from reactions with inhibiting factors were used to
558 calibrate phosphate absorbance in wells measuring ouabain inhibition and in the control wells
559 measuring non-inhibited NKA activity (48). For ouabain sensitivity measurements, calibrated
560 absorbance values were converted to percentage non-inhibited NKA activity based on
561 measurements from the control wells (48). These data were plotted and log IC_{50} values were
562 obtained for each biological replicate from nonlinear fitting using a four-parameter logistic curve,
563 with the top asymptote set to 100 and the bottom asymptote set to zero. Curve fitting was performed
564 with the nlsLM function of the minipack.lm library in R. For comparisons of recombinant protein
565 activity, the calculated Pi concentrations of 100 μg of protein assayed in the absence of ouabain
566 were converted to nmol Pi/mg protein/min. IC_{50} values were log-transformed. We used pairwise *t*-
567 tests with Bonferroni corrections to identify significant differences between constructs with and
568 without engineered substitutions. We used a two-way ANOVA to test for background dependence
569 of substitutions (i.e., interaction between background and amino acid substitution) with respect to
570 ouabain resistance (log IC_{50}) and protein activity. Specifically, we tested whether the effects of a
571 substitution X->Y are equal on different backgrounds (null hypothesis: X->Y (background 1) = X-
572 >Y (background 2)). We further assumed that the effects of a substitution X->Y should exactly
573 match that of Y->X. All statistical analyses were implemented in R. Data were plotted using the
574 ggplot2 package in R.

575
576
577

578 **Acknowledgments**

579

580 We thank C. Natarajan, P. Kowalski, M. Winter, and V. Wagschal for assistance in the laboratory,
581 and D.A. Gómez-Sánchez for assistance in the field. We thank J. Oaks for providing tissue from
582 ring-necked snake. **Funding:** This study was funded by grants from the National Institutes of Health
583 to PA (R01-GM115523), JFS (R01-HL087216) and SM (F32–HL149172), the National Science
584 Foundation (OIA-1736249) to JFS, the Deutsche Forschungsgemeinschaft (Do 517/10-1) to SD,
585 and the Alexander von Humboldt Foundation to SM.

586

587 **References**
588

589 1. D. L. Stern, The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764
590 (2013).

591 2. J. F. Storz, Causes of molecular convergence and parallelism in protein evolution. *Nat. Rev.*
592 *Genet.* **17**, 239 (2016).

593 3. J. B. Lingrel, The physiological significance of the cardiotonic steroid/ouabain-binding site
594 of the Na, K-ATPase. *Annu. Rev. Physiol.* **72**, 395–412 (2010).

595 4. S. Dobler, S. Dalla, V. Wagschal, A. A. Agrawal, Community-wide convergent evolution in
596 insect adaptation to toxic cardenolides by substitutions in the Na, K-ATPase. *Proc. Natl.*
597 *Acad. Sci.* **109**, 13040–13045 (2012).

598 5. Y. Zhen, M. L. Aardema, E. M. Medina, M. Schumer, P. Andolfatto, Parallel molecular
599 evolution in an herbivore community. *Science* **337**, 1634–1637 (2012).

600 6. D. J. Moore, D. C. Halliday, D. M. Rowell, A. J. Robinson, J. S. Keogh, Positive Darwinian
601 selection results in resistance to cardioactive toxins in true toads (Anura: Bufonidae). *Biol.*
602 *Lett.* **5**, 513–516 (2009).

603 7. B. Ujvari, *et al.*, Widespread convergence in toxin resistance by predictable molecular
604 evolution. *Proc. Natl. Acad. Sci.* **112**, 11911–11916 (2015).

605 8. S. Mohammadi, *et al.*, Toxin-resistant isoforms of Na+/K+-ATPase in snakes do not closely
606 track dietary specialization on toads. *Proc. R. Soc. B Biol. Sci.* **283**, 20162111 (2016).

607 9. S. Mohammadi, L. Yang, M. Bulbert, H. M. Rowland, The evolutionary and behavioural
608 ecology of cardiotonic steroid resistance in predators (In prep).

609 10. S. Groen, N. Whiteman, Convergent evolution of cardiac-glycoside resistance in predators
610 and parasites of milkweed herbivores. *Curr. Biol.* **31**, R1465–R1466 (2021).

611 11. D. L. Stern, *Evolution, development, & the predictable genome* (Roberts and Co. Publishers,
612 2011).

613 12. P. C. Phillips, Epistasis—the essential role of gene interactions in the structure and
614 evolution of genetic systems. *Nat. Rev. Genet.* **9**, 855–867 (2008).

615 13. W. Fitch, Rate of change of concomitantly variable codons. *J. Mol. Evol.* **1**, 84–96 (1971).

616 14. B. Callahan, R. A. Neher, D. Bachtrog, P. Andolfatto, B. I. Shraiman, Correlated evolution of
617 nearby residues in Drosophilid proteins. *PLoS Genet.* **7**, e1001315 (2011).

618 15. T. N. Starr, J. W. Thornton, Epistasis in protein evolution. *Protein Sci.* **25**, 1204–1218
619 (2016).

620 16. J. F. Storz, Compensatory mutations and epistasis for protein function. *Curr. Opin. Struct.*
621 *Biol.* **50**, 18–25 (2018).

622 17. D. D. Pollock, G. Thiltgen, R. A. Goldstein, Amino acid coevolution induces an evolutionary
623 Stokes shift. *Proc. Natl. Acad. Sci.* **109**, E1352 (2012).

624 18. P. Shah, D. M. McCandlish, J. B. Plotkin, Contingency and entrenchment in protein
625 evolution under purifying selection. *Proc. Natl. Acad. Sci.* **112**, E3226 (2015).

626 19. V. O. Pokusaeva, *et al.*, An experimental assay of the interactions of amino acids from
627 orthologous sequences shaping a complex fitness landscape. *PLoS Genet.* **15**, e1008079
628 (2019).

629 20. R. A. Goldstein, S. T. Pollard, S. D. Shah, D. D. Pollock, Nonadaptive amino acid
630 convergence rates decrease over time. *Mol. Biol. Evol.* **32**, 1373–1381 (2015).

631 21. Z. Zou, J. Zhang, Are convergent and parallel amino acid substitutions in protein evolution
632 more prevalent than neutral expectations? *Mol. Biol. Evol.* **32**, 2085–2096 (2015).

633 22. Z. Zou, J. Zhang, Gene tree discordance does not explain away the temporal decline of
634 convergence in mammalian protein sequence evolution. *Mol. Biol. Evol.* **34**, 1682–1688
635 (2017).

636 23. A. M. Taverner, *et al.*, Adaptive substitutions underlying cardiac glycoside insensitivity in
637 insects exhibit epistasis in vivo. *eLife* **8**, e48224 (2019).

638 24. M. Karageorgi, *et al.*, Genome editing retraces the evolution of toxin resistance in the
639 monarch butterfly. *Nature* **574**, 409–412 (2019).

640 25. S. Mohammadi, *et al.*, Concerted evolution reveals co-adapted amino acid substitutions in
641 frogs that prey on toxic toads. *Curr. Biol.* **31**, 2530-2538.e10 (2021).

642 26. K. J. Sweadner, *et al.*, Genotype-structure-phenotype relationships diverge in paralogs
643 ATP1A1, ATP1A2, and ATP1A3. *Neurol. Genet.* **5**, e303–e303 (2019).

644 27. B. Ujvari, *et al.*, Isolation breeds naivety: island living robs Australian varanid lizards of
645 toad-toxin immunity via four-base-pair mutation. *Evol. Int. J. Org. Evol.* **67**, 289–294
646 (2013).

647 28. B. M. Marshall, *et al.*, Widespread vulnerability of Malagasy predators to the toxins of an
648 introduced toad. *Curr. Biol.* **28**, R654–R655 (2018).

649 29. M. Lunzer, G. B. Golding, A. M. Dean, Pervasive cryptic epistasis in molecular evolution.
650 *PLoS Genet* **6**, e1001162 (2010).

651 30. L. Yang, *et al.*, Predictability in the evolution of Orthopteran cardenolide insensitivity.
652 *Philos. Trans. R. Soc. B* **374**, 20180246 (2019).

653 31. E. M. Price, J. B. Lingrel, Structure-function relationships in the sodium-potassium ATPase.
654 alpha. subunit: site-directed mutagenesis of glutamine-111 to arginine and asparagine-122
655 to aspartic acid generates a ouabain-resistant enzyme. *Biochemistry* **27**, 8400–8408
656 (1988).

657 32. A. Stoltzfus, D. M. McCandlish, Mutational Biases Influence Parallel Adaptation. *Mol. Biol.*
658 *Evol.* **34**, 2163–2172 (2017).

659 33. J. Zhang, S. Kumar, Detection of convergent and parallel evolution at the amino acid
660 sequence level. *Mol. Biol. Evol.* **14**, 527–536 (1997).

661 34. L. F. Toledo, R. Ribeiro, C. F. Haddad, Anurans as prey: an exploratory analysis and size
662 relationships between predators and their prey. *J. Zool.* **271**, 170–177 (2007).

663 35. S. Dobler, *et al.*, New ways to acquire resistance: imperfect convergence in insect
664 adaptations to a potent plant toxin. *Proc. R. Soc. B* **286**, 20190883 (2019).

665 36. M. V. Clausen, F. Hilbers, H. Poulsen, The structure and function of the Na, K-ATPase
666 isoforms in health and disease. *Front. Physiol.* **8**, 371 (2017).

667 37. Y. Zhen, M. L. Aardema, E. M. Medina, M. Schumer, P. Andolfatto, Parallel molecular
668 evolution in an herbivore community. *science* **337**, 1634–1637 (2012).

669 38. J. A. Wells, Additivity of mutational effects in proteins. *Biochemistry* **29**, 8509–8517 (1990).

670 39. M. Lunzer, S. P. Miller, R. Felsheim, A. M. Dean, The biochemical architecture of an ancient
671 adaptive landscape. *Science* **310**, 499–501 (2005).

672 40. L. I. Gong, M. A. Suchard, J. D. Bloom, Stability-mediated epistasis constrains the evolution
673 of an influenza protein. *Elife* **2**, e00631 (2013).

674 41. A. Mobasheri, *et al.*, Na+, K+-ATPase isozyme diversity; comparative biochemistry and
675 physiological implications of novel functional interactions. *Biosci. Rep.* **20**, 51–91 (2000).

676 42. B. J. Haas, *et al.*, De novo transcript sequence reconstruction from RNA-seq using the
677 Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).

678 43. D. Darriba, *et al.*, ModelTest-NG: a new and scalable tool for the selection of DNA and
679 protein evolutionary models. *Mol. Biol. Evol.* **37**, 291–294 (2020).

680 44. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
681 phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

682 45. S. Guindon, *et al.*, New algorithms and methods to estimate maximum-likelihood
683 phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

684 46. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T. Lam, ggtree: an R package for visualization and
685 annotation of phylogenetic trees with their covariates and other associated data. *Methods*
686 *Ecol. Evol.* **8**, 28–36 (2017).

687 47. Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–
688 1591 (2007).

689 48. G. Petschenka, *et al.*, Stepwise evolution of resistance to toxic cardenolides via genetic
690 substitutions in the Na+/K+-ATPase of milkweed butterflies (Lepidoptera: Danaini).
691 *Evolution* **67**, 2753–2761 (2013).

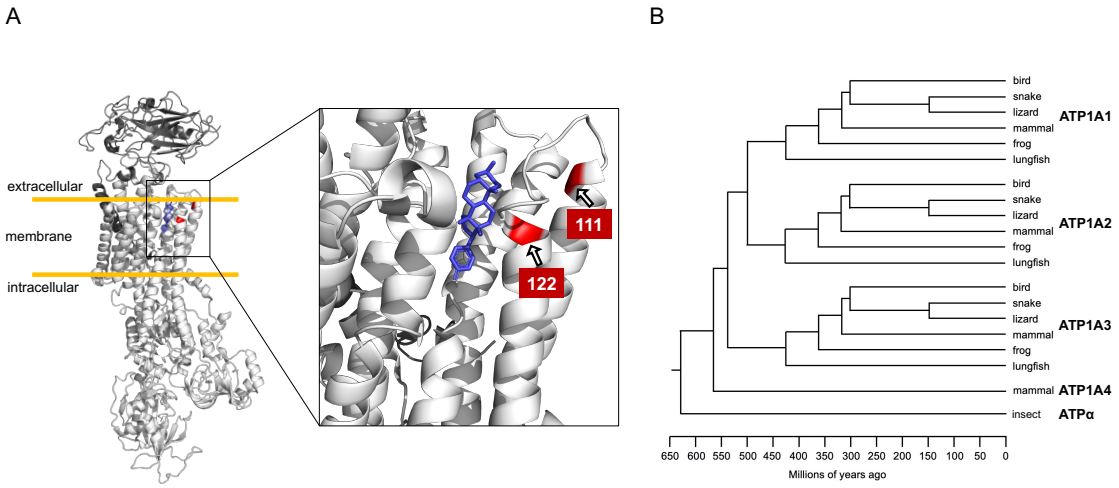
692 49. H. H. Taussky, E. Shorr, A microcolorimetric method for the determination of inorganic
693 phosphorus. *J. Biol. Chem.* **202**, 675–685 (1953).

694

695

696 **Figures and Tables**

697
698



699

700 **Figure 1. Na,K-ATPase structure and phylogenetic relationships of ATP1A paralogs among**
701 **vertebrates.** (A) Crystal structure of an Na^+,K^+-ATPase (NKA) with a bound the representative CTS
702 bufalin in blue (PDB 4RES). The zoomed-in panel shows the H1-H2 extracellular loop, highlighting
703 two amino acid positions (111 and 122 in red) that have been implicated repeatedly in CTS
704 resistance. We highlight key examples of convergence in amino acid substitutions at sites in the
705 H1-H2 extracellular loop associated with CTS resistance in Fig 3. (B) Phylogenetic relationships
706 among ATP1A paralogs of vertebrates and ATPα of insects.
707

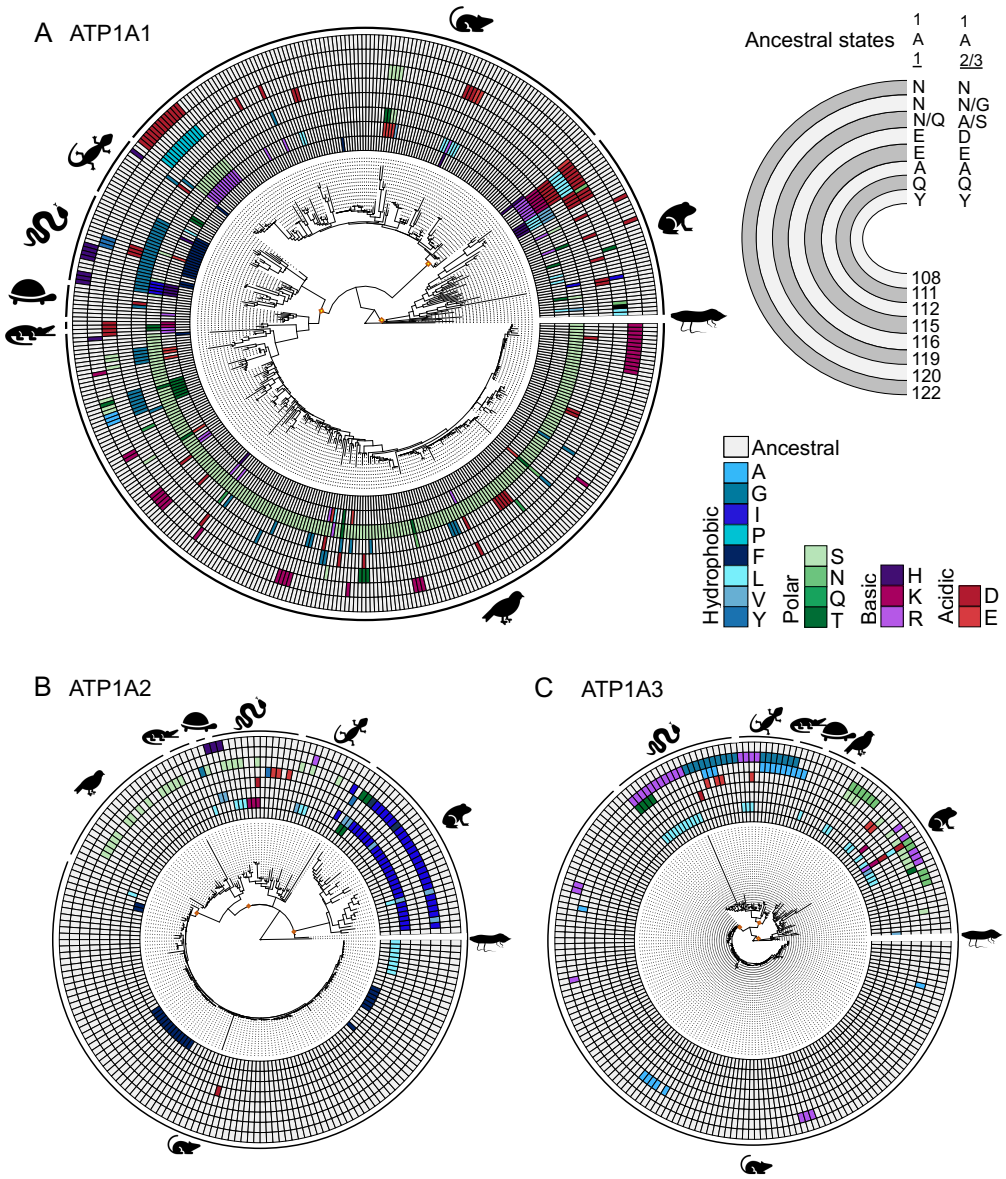
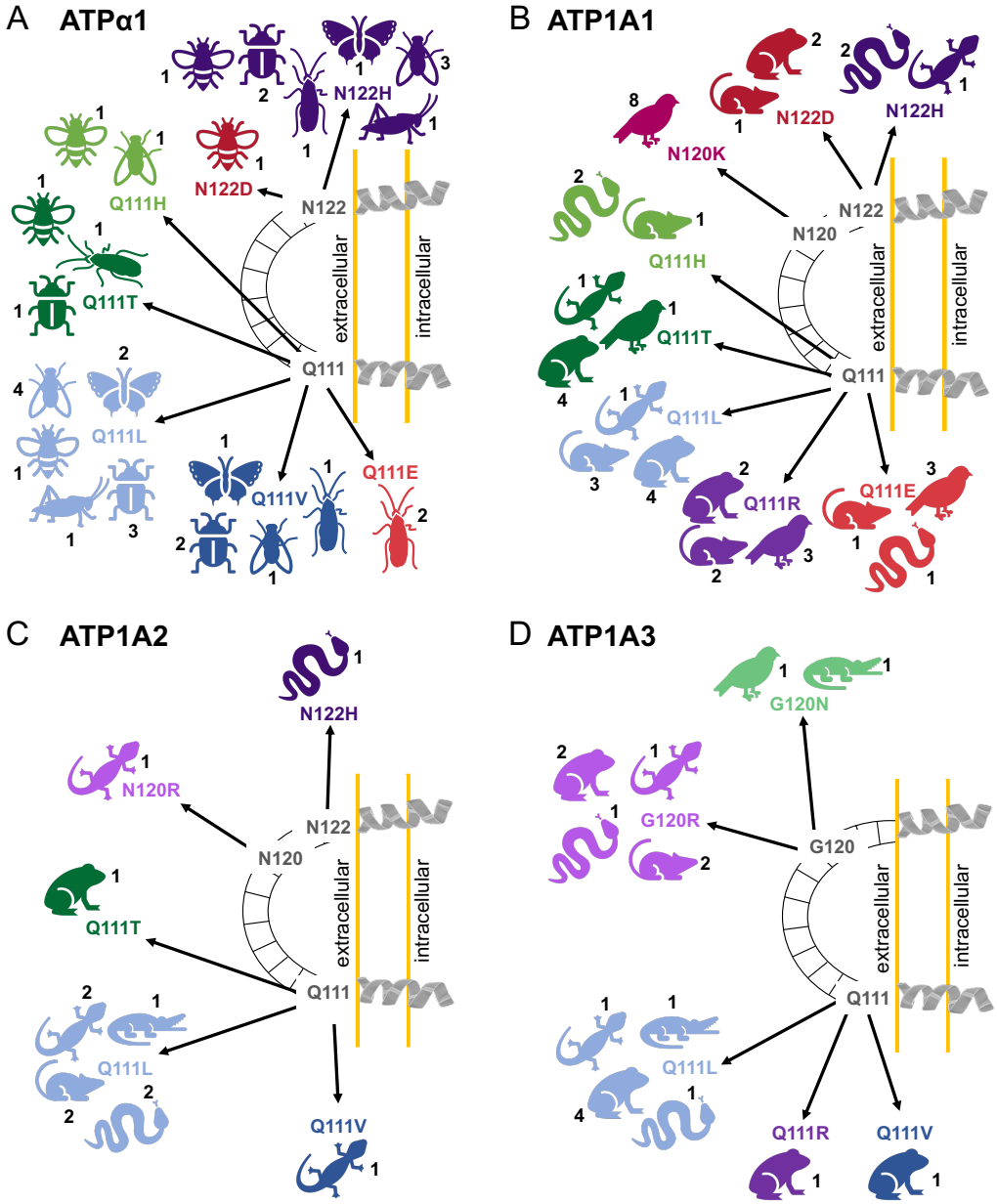


Figure 2. Patterns of molecular evolution in the α(M1–M2) extracellular loop of ATP1A paralogs shared among tetrapods. (A) Maximum likelihood phylogeny of tetrapod ATP1A1, (B) ATP1A2, and (C) ATP1A3. The character states for eight sites relevant to CTS resistance in and near the H1-H2 loop of the NKA protein are shown at the node tips. Yellow internal nodes indicate ancestral sequences reconstructed to infer derived amino acid states across clades to ease visualization; nodes reconstructed: MRCA of mammals, reptiles, and amphibians. *Top right*, each semi-circle indicates the site mapped in the main phylogeny with the inferred ancestral amino acid state for each of the three yellow nodes (posterior probability >0.8). In ATP1A1, site 119 was inferred as Q119 for amphibians and mammals, and N119 for reptiles (Table S6); in ATP1A2-3 site 119 was inferred as A119 for amphibians and reptiles, and S119 for mammals (Table S6). Site number corresponds to pig (*Sus scrofa*) reference sequence. Higher number and variation of substitutions in ATP1A1 stand out in comparison to the other paralogs.

723



724
725
726 **Figure 3. Parallel and divergent patterns of CTS-resistant substitutions across ATPα1 of**
727 **insects and the shared ATP1A paralogs of tetrapods.** Examples of convergence in ATPα1
728 across insects (A). Convergence in the **(B)** ATP1A1, **(C)** ATP1A2, and **(D)** ATP1A3 paralogs,
729 respectively, across tetrapods. Numbers indicate the number of independent substitutions in each
730 major clade depicted. For ATP1A3, resistance-conferring amino acid substitutions have been
731 identified at site 120, and not 122. A full list of amino acid substitutions can be found in
732 Supplementary Dataset 2 for tetrapods, and Taverner et al. (23) for insects.
733

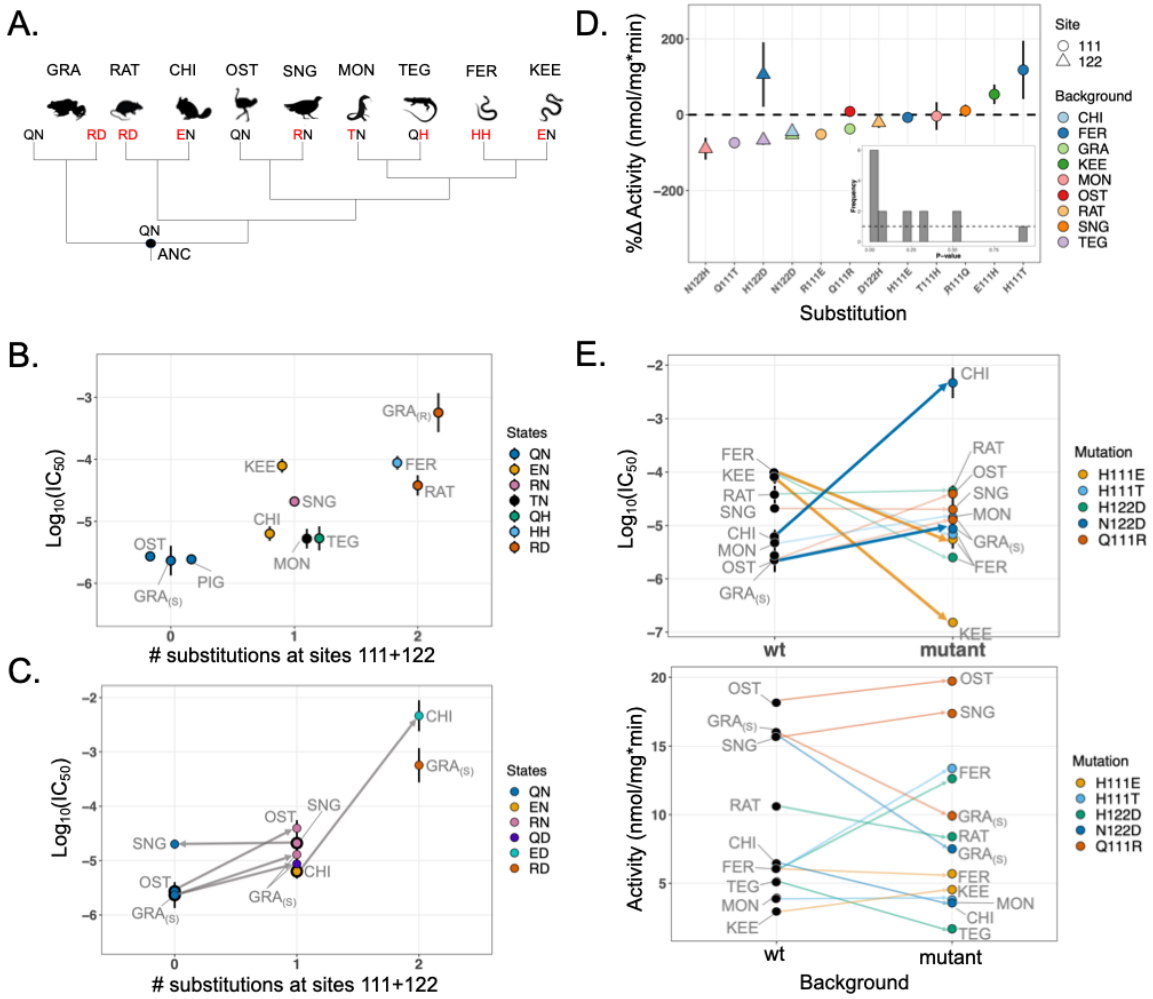


Figure 4. Functional properties of wild-type and engineered ATP1A1. **(A)** Cladogram relating the surveyed species. GRA: Grass Frog (*Leptodactylus*); RAT: Rat (*Rattus*); CHI: Chinchilla (*Chinchilla*); OST: Ostrich (*Struthio*); SNG: Sandgrouse (*Pterocles*); MON: Monitor lizard (*Varanus*); TEG: Tegu lizard (*Tupinambis*); FER: False fer-de-lance (*Xenodon*); KEE: Red-necked keelback snake (*Rhabdophis*). Two-letter codes underneath each avatar indicate native amino acid states at sites 111 and 122, respectively. Data for grass frog from Mohammadi et al. (2021). **(B)** Levels of CTS resistance (IC50) among wild-type enzymes. The x-axis distinguishes among ATP1A1 with 0, 1 or 2 derived states at sites 111 and 122. The subscripts S and R refer to the CTS-sensitive and CTS-resistant paralogs, respectively. **(C)** Effects of changing the number of substitutions at 111 or 122 on CTS resistance (IC50). Substitutions result in predictable changes to resistance except in the reversal R111Q in Sandgrouse (SNG). GRA_S represents Q111R+N122D on the sensitive paralog background. **(D)** Effects of single substitutions on Na^+,K^+-ATPase (NKA) activity. Each modified ATP1A1 is compared to the wild-type enzyme for that species. The inset shows the distribution of *t*-test p-values for all 15 substitutions, with the dotted line indicating the expectation**. (E)** Evidence for epistasis for CTS resistance (IC50, upper panel) and lack of such effects for enzyme activity (lower panel). Each line compares the same substitution (or the reverse substitution) tested on at least two backgrounds. Thicker lines correspond to substitutions with significant sequence-context dependent effects (Bonferroni-corrected ANOVA p-values < 0.05, Table S5).

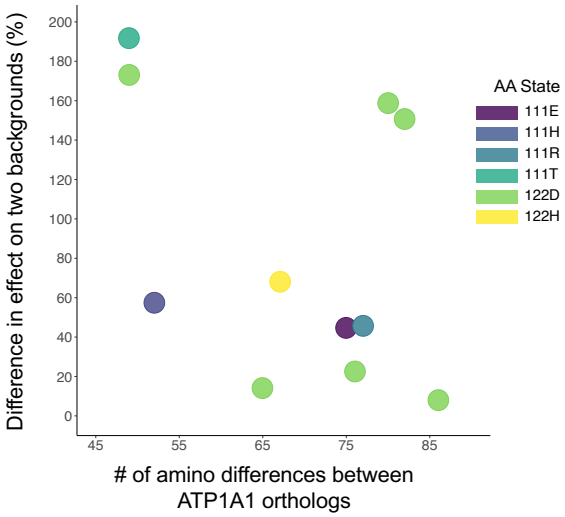


Figure 5. No relationship between the effect of substitution to a given amino state on activity and the extent of divergence between ATP1A1 orthologs. Each point represents a comparison between the effect (% change in activity relative to the wild-type enzyme) of a given amino acid state (e.g., 122D) on two different genetic backgrounds. For example, the effect of 122D between chinchilla and false fer-de-lance is measured as % change [chinchilla vs. chinchilla+N122D] minus the % change [false fer-de-lance vs. false fer-de-lance+H122D]. Comparisons were measured as the difference between the two effects. In total, 11 comparisons were possible. Note that comparisons are not independent. The x-axis represents the number of amino acid differences between two ATP1A1 proteins being compared. Assuming intramolecular epistasis for protein function is prevalent, a positive correlation is predicted. However, no such relationship is observed (Spearman's correlation, r_S = -0.42, p = 0.19).

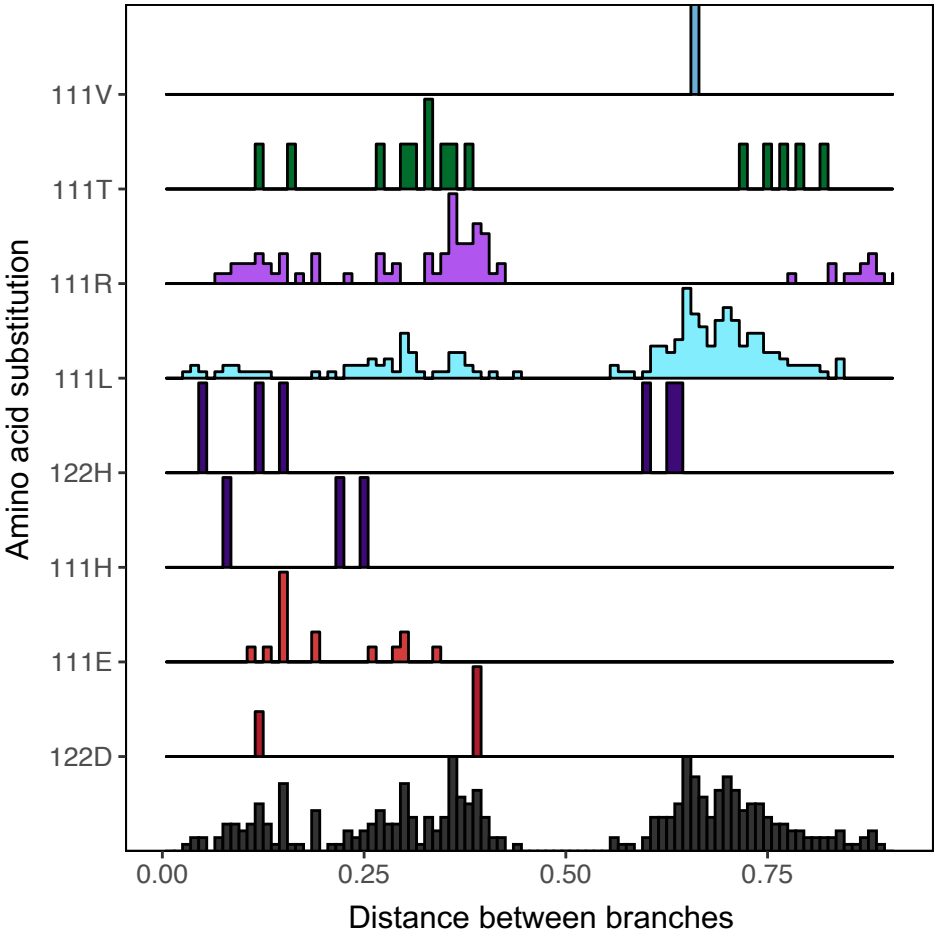


Figure 6. Rate of convergence across ATP1A sequences as a function of increasing sequence divergence. (A) Change in the rate of convergence (protein wide) over time for the ATP1A protein family. The proportion of convergent (C) over divergent (D) substitutions along the entire protein sequence was estimated for all pairs of branches in the ATP1A phylogeny, except for sister branches or ancestor-descendant pairs. Color scale shows the density of dots for both axes. The distance between branches corresponds to the expected number of amino acid substitutions per site between protein pairs being compared (under the JTT+G4+F model). The red line shows a running average with a window size of 0.05 substitutions/site. Dashed lines show the 95% confidence interval based on 100 bootstrap replicates per window. **(B)** For each derived amino acid state at sites 111 and 122, the histograms show the distribution of pairwise convergent events along the sequence divergence axis (expected number of substitutions per site). Substitutions are color coded as in Figure 2. The histogram at the bottom shows the combined distribution of pairwise convergent events for both sites.