Concerted evolution reveals co-adapted amino acid substitutions in frogs that prey on toxic toads

Authors: Shabnam Mohammadi[1]*, Lu Yang[2]*, Arbel Harpak[3]*, Santiago Herrera-Álvarez[4, ψ], María del Pilar Rodríguez-Ordoñez[4, #], Julie Peng, Karen Zhang[2], Jay F. Storz[1], Susanne Dobler[6], Andrew J. Crawford[4]** and Peter Andolfatto[3]**

Affiliations:

[1] School of Biological Sciences, University of Nebraska, Lincoln, NE, USA

[2] Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA

[3] Department of Biological Sciences, Columbia University, New York, NY, USA

[4] Department of Biological Sciences, Universidad de los Andes, Bogotá, 111711, Colombia

[5] Lewis-Sigler Institute, Princeton University, Princeton, NJ, USA

[6] Molecular Evolutionary Biology, Zoological Institute, Universität Hamburg, Hamburg, Germany

*Co-first authorship

**Correspondence to: andrew@dna.ac, pa2543@columbia.edu

[ψ] Current address: Dept. of Ecology and Evolution, University of Chicago, Chicago, IL, USA

[#] Current address: Université Paris-Saclay Evry, Evry, France

ABSTRACT (141 words):

Gene duplication is an important source of evolutionary innovation, but the adaptive division-of-labor between duplicates can be opposed by ongoing gene conversion between them. Here we document a tandem duplication of Na^+,K^+-ATPase subunit α1 (ATP1A1) shared by frogs in the genus *Leptodactylus*, a group of species that feeds on toxic toads. One ATP1A1 paralog evolved resistance to toad toxins while the other copy retained ancestral susceptibility. We show that the two *Leptodactylus* paralogs are distinguished by 12 amino acid substitutions that were maintained by strong selection that counteracted the homogenizing effect of gene conversion. Protein-engineering experiments show that two major-effect substitutions confer toxin resistance, whereas the 10 additional substitutions mitigate deleterious pleiotropic effects on enzyme function. Our results highlight how trans-specific, neofunctionalized gene duplicates can provide unique insights into interactions between adaptive substitutions and the genetic backgrounds on which they arise.

One Sentence Summary: Selection counteracts gene conversion to maintain an adaptive division-of-labor between tandemly duplicated genes.

49 **Main Text:**

50 The repeated evolution of toxin resistance in animals is one of the clearest examples of natural
51 selection at the molecular level and represents a useful paradigm to examine constraints on
52 the evolution of novel protein functions (*1*). Neotropical Grass Frogs of the genus *Leptodactylus*
53 (Leptodactylidae) are widely distributed throughout lowland South America and are known to
54 feed on chemically-defended toads – a predatory tendency that is rare among frogs (*2–6*). A
55 major component of the chemical defense secretions of toads is a class of cardiac glycosides
56 (CGs) called "bufadienolides" (*7*) that inhibit the α-subunit of Na^+,K^+-ATPases (ATP1A).
57 Na^+,K^+-ATPases are transmembrane proteins that are vital to numerous physiological processes
58 in animals including neural signal transduction, muscle contraction, and cell homeostasis (*8*, *9*).
59 CGs bind to the extracellular surface of ATP1A and block the flux of ions (*10*), making them
60 potent poisons to most animals. However, some vertebrates have independently evolved the
61 ability to prey on chemically-defended toads, partly via amino acid substitutions to the CG-
62 binding domain of ATP1A1 that confer molecular resistance to CGs (*11–14*).

63

64 Most vertebrates share several paralogous copies of ATP1A that have different tissue-specific
65 expression profiles (*15*). For example, ATP1A1 is the most ubiquitously expressed paralog and
66 ATP1A3 has enriched expression in nervous tissue and heart muscle (*16*, *17*; Fig. S1). Previous
67 studies on the molecular convergence of CG-resistance in reptiles have focused primarily on the
68 αM1–2 extracellular loop of ATP1A3 (*12–14*, *18*), whereas studies of birds, mammals, and
69 amphibians have focused on the same region of ATP1A1 (*11*, *18*). A survey of αM1–2 in toads
70 and frogs (*11*) revealed a possible duplication of ATP1A1 in the toad-eating frog, *Leptodactylus*
71 *latrans* (reported as *L. ocellatus*), where the resistant (R) copy includes substitutions known to

72 confer resistance to CGs while the sensitive (S) copy appears to have retained the ancestral
73 susceptibility to CGs. Duplication and neofunctionalization of ATP1A are a recurrent adaptive
74 mechanism in CG-resistant insects (*19–22*) but appear to be rare among CG-resistant vertebrates.
75 Further, the fate of duplicate genes and the probability that they will neofunctionalize is
76 predicted to depend on the strength of selection for duplicate diversification relative to the rate of
77 non-allelic gene conversion (NAGC)—nonreciprocal recombination between duplicates that
78 homogenizes sequence variation between them, thereby impeding divergence (*23–25*; Fig. 2A).
79 The ATP1A1 gene family in *Leptodactylus* therefore provide a unique opportunity to explore
80 this premise further in vertebrates.

81

82 We systematically surveyed the full-length coding sequences for all ATP1A paralogs in
83 *Leptodactylus* and other anurans using a combination of RNA-seq-based gene discovery (*19*)
84 (Tables S1 and S2), targeted sequencing of the full length ATP1A1 gene (Methods), and a *de*
85 *novo* genome assembly for *Leptodactylus fuscus* (Table S3). As predicted by previous findings
86 (*11*), we confirmed that ATP1A1 is duplicated in *L. latrans*. By generating a *de novo* genome
87 assembly for *L. fuscus* (Table S3), we find that the two duplicate copies are arranged in tandem
88 (Fig. S2), and are therefore potentially subject to non-allelic gene conversion. Our results also
89 indicate that duplication of ATP1A1 most likely occurred in the common ancestor of all
90 surveyed *Leptodactylus* species (Fig. 1A, Table S2). Two other ancient paralogs common to
91 vertebrates, ATP1A2 and ATP1A3, appear to be present as single-copy genes and lack any
92 known CG-resistant substitutions in *Leptodactylus* (Fig. S3).

93

94 The αM1–2 transmembrane domains of the ATP1A1 paralog in *L. latrans* are distinguished by
95 four amino acid substitutions (*11*) (Fig 1C, Fig. S4). Two of these substitutions, Q111R and
96 N122D, were first identified in rat ATP1A1 and have been shown to interact synergistically to
97 confer CG-resistance to the CG-sensitive sheep ATP1A1 protein *in vitro* (*26*, *27*). By comparing
98 ATP1A1 sequences in five distantly related *Leptodactylus* species (*28*), we find that all five
99 harbor both a putatively resistant paralog (R) that includes the Q111R and N122D substitutions
100 and a putatively sensitive ATP1A1 paralog (S) that lacks these substitutions. In addition to
101 Q111R and N122D, 10 other amino acid substitutions distinguish the R and S paralogs in most
102 of the five sampled species (Fig. 1C). We refer to these twelve substitutions as "R/S
103 distinguishing" substitutions henceforth.

104
105 To elucidate the history of gene duplication relative to speciation events, we estimated
106 phylogenies from alignments of ATP1A1 protein-coding nucleotide sequences. Phylogenies
107 estimated from synonymous and inferred protein sequences support strikingly different
108 topologies (Fig 1B and C). If we were to assume that R evolve independently within each
109 species, the phylogeny estimated from nucleotide sequences (primarily based on variation at
110 synonymous sites) would imply independent duplications in each of the *Leptodactylus* species,
111 followed by parallel substitutions at the same 12 R/S distinguishing amino acid positions (Fig.
112 1B). Instead, we propose that this genealogical pattern reflects a single ancestral duplication - as
113 estimated from inferred amino acid sequences (Fig. 1C, Table S4) - and a history of recurrent
114 NAGC between paralogs. Frequent NAGC produces a pattern of "concerted evolution" whereby
115 tandemly linked paralogs from the same species are more similar to one another than they are to
116 their orthologous counterparts in other species (Fig. 2A). We propose that the persistence of the
117 12 amino acid differences between the two paralogs is due to selection counteracting the

118 homogenizing effects of NAGC (*23*, *29*) (Fig. 2B), thereby maintaining an adaptive division-of-
119 labor between toxin resistance and cellular physiology.
120

121 The opposing forces of NAGC and selection are predicted to leave a characteristic genealogical
122 signature at neutral sites closely linked to the targets of selection (*29*) (Fig. 2B). We thus tested
123 the relationship between the genealogical signature and distance from nonsynonymous variants
124 putatively under selection. To do this, we classified all informative variant sites as supporting
125 either an ancient duplication of ATP1A1 that predated all speciation events versus an alternative
126 in which within-species paralogs are genealogical neighbors (defining support as a difference in
127 two log likelihoods for these alternative genealogical models). This analysis reveals that
128 synonymous (presumed to be neutral) variants congruent with an ancient duplication of R and S
129 have a median distance of 4 bp from nonsynonymous variants exhibiting the same pattern (Fig.
130 2C). In contrast, randomly sampled synonymous sites supporting the alternative genealogy have
131 a median distance of 88 bp from those nonsynonymous variants (bootstrap $p<10^{-5}$). This pattern
132 at synonymous sites is consistent with selection to maintain key amino acid substitutions
133 distinguishing neofunctionalized gene duplicates in the face of NAGC.
134

135 Similar qualitative patterns have been observed in the Amylase gene family in Drosophila (*30*),
136 the RHCE/RHD antigen genes of humans (*31*), and the red/green opsin genes of primates (*29*),
137 but the trans-specific sharing of R/S distinguishing substitutions here allows us to quantify the
138 intensity of these opposing evolutionary forces. We quantified the strength of selection required
139 at R/S distinguishing substitutions to counteract NAGC between R and S paralogs. We first
140 considered population genetics theory for the evolution of a single site in tandem duplicates (*32*)

141 (Supplementary Materials). This analytic model predicts that if, for example, gene conversion is
142 an order of magnitude more frequent than point mutation, then the maintenance of diverged,
143 distinct alleles is only likely under strong purifying selection (namely, a population size scaled
144 selection coefficient, *2Ns*, larger than one) (Fig. 3A). We next developed an inference method
145 based on simulations of ATP1A1 evolution to estimate the parameters that, jointly, best explain
146 divergence patterns throughout the gene, including both high and low levels of paralog
147 divergence observed near and far from the 12 R/S distinguishing substitutions, respectively
148 (Supplementary Materials). We estimate the rate of NAGC to be an order of magnitude higher
149 than the point mutation rate (posterior mode 9 with 80% credible interval 4-54 times the point
150 mutation rate), and *2Ns* substantially larger than one (posterior mode 9; 80% credible interval 5-
151 18; Fig. 3B). These estimates fall within the plausible range predicted by the theoretical single-
152 site model of Fig. 3A. Thus, our evolutionary inference supports the joint, reciprocal action of
153 frequent NAGC and strong purifying selection at R/S distinguishing substitutions to generate the
154 observed divergence patterns in ATP1A1.

155

156 The inference that selection maintains the co-occurrence of these 12 R/S distinguishing
157 substitutions on the R copy implies they are functionally important and collectively contribute to
158 organismal fitness. The functional importance for Q111R and N122D in the context of CG-
159 insensitivity has previously been demonstrated by *in vitro* enzyme inhibition assays (*9*).
160 Additionally, the potential importance of substitutions at sites 112 and 116 has been suggested
161 by molecular evolution analysis and structural studies, respectively (*11*, *33*). However, the
162 remaining eight R/S distinguishing substitutions are scattered among regions of the protein that
163 have not been implicated in CG-resistance. Since our analysis suggests that amino acid

164 divergence between R and S paralogs is maintained by selection, we performed protein-
165 engineering experiments to elucidate the functional significance of the 12 R/S distinguishing
166 substitutions. We synthesized and recombinantly expressed eight mutant Na^+,K^+-ATPase
167 proteins, each harboring different combinations of R-specific replacements on both S- and R-
168 type genetic backgrounds of the representative species, *L. latrans* (Fig. 4A, Table S6) and we
169 then quantified the level of CG-resistance of each genotype using enzyme-inhibition assays
170 (Table S7, Fig. S5, (*34*)). Individually, Q111R and N122D significantly increased CG-resistance
171 by 21-fold and 14-fold, respectively (ANOVA p=2.7e-13 and p=2.3e-6; Fig. 4B, Table S8).
172 When combined, Q111R and N122D produce a >100-fold increase in CG resistance relative to
173 the S-copy (Tukey's HSD test, adjusted p<4e-5, Fig. 4B, Tables S6-S7). In contrast, the
174 remaining 10 substitutions had no detectable effect on CG-resistance when added to the S
175 background (p=0.22, Fig. 4B).

176

177 Given the absence of detectable effects of R/S distinguishing substitutions other than Q111R and
178 N122D on CG-resistance, we asked whether these substitutions had effects on other aspects of
179 ATP1A1 function that relate to organismal fitness. Since ATP hydrolysis and ion co-transport
180 are strongly coupled functions of Na^+,K^+-ATPase (*35*), we used estimates of the rate of ATP
181 hydrolysis in the absence of ouabain as a proxy for overall protein activity. Based on this assay,
182 we found that CG-resistance substitutions Q111R and N122D significantly impair activity,
183 individually reducing ATPase activity by an average of 40% (*p*=0.024 and *p*=7.7e-4
184 respectively; Fig. 4B; Tables S8). We also detected a significant interaction between Q111R and
185 N122D that renders their joint effects somewhat less severe than predicted by the sum of their
186 individual effects (*i.e.*, a 30% reduction rather than the expected 78% reduction, p=0.022).

187 Additionally, adding the remaining 10 R-specific substitutions on the S background containing
188 Q111R and N122D restores normal ATPase activity close to wild-type levels. An analysis of
189 variance reveals a highly significant effect of these 10 R-specific substitutions ($p=1.2 \times 10^{-4}$, Fig.
190 4B, Table S8). These results indicate that these 10 R/S distinguishing substitutions play a vital
191 role in compensating for the negative pleiotropic effects of the resistance-conferring
192 substitutions, Q111R and N122D. We conclude that the evolution of the R protein from a CG-
193 sensitive ancestral state involved two epistatically interacting substitutions (Q111R and N122D)
194 in conjunction with compensatory effects of 10 additional substitutions that mitigate the trade-off
195 between toxin resistance and native enzyme activity.

196

197 Apart from the 12 R/S distinguishing substitutions shared by most *Leptodactylus* species, an
198 additional nine substitutions and a two-amino acid deletion distinguish the S and R paralogs in
199 the focal species, *L. latrans*. Our experiments revealed that these 10 species-specific
200 substitutions do not contribute to S vs. R differences in CG resistance or enzyme function (using
201 all 10 as one co-variate, ANOVA $p>0.5$).

202

203 The functional significance of protein divergence between species is a fundamental question in
204 biology and it is often difficult to assess in the absence of *a priori* functional information. The R-
205 and S-type ATP1A1 paralogs of *L. latrans* are distinguished by 22 amino acid substitutions
206 (including one deletion), only two of which are associated with protein function (in this case CG-
207 resistance). Our evolutionary analyses allowed us to focus experiments on just 12 of these 22
208 substitutions that we inferred were likely to be important for function. We found that an adaptive
209 division-of-labor between the R and S paralogs of ATP1A1 in *Leptodactylus* has been

210 maintained by strong selection that has counteracted the homogenizing effects of frequent
211 NAGC over the 35 million-year history of this genus. The possession of functionally distinct R-
212 and S-type ATP1A1 paralogs ensures that *Leptodactylus* frogs are resistant to the CG toxins
213 produced by their prey without compromising native ATPase function. Our work reveals the
214 mechanisms by which selection surmounts intragenic constraints on adaptation and demonstrates
215 how intramolecular epistasis and pleiotropy can shape adaptation.
216

217 **References and Notes:**

218 1. E. D. Brodie, Toxins and venoms. *Curr. Biol.* **19**, R931–R935 (2009).

219 2. K. Chen, A. L. CHEN, Notes on the poisonous secretions of twelve species of toads. *J.*
220 *Pharmacol. Exp. Ther.* **47**, 281–293 (1933).

221 3. W. R. Heyer, R. W. McDiarmid, D. L. Weigmann, Tadpoles, predation and pond habitats in
222 the tropics. *Biotropica*, 100–111 (1975).

223 4. M. R. Crossland, C. Azevedo-Ramos, Effects of Bufo (Anura: Bufonidae) toxins on
224 tadpoles from native and exotic Bufo habitats. *Herpetologica*, 192–199 (1999).

225 5. C. Azevedo-Ramos, W. E. Magnusson, Tropical tadpole vulnerability to predation:
226 association between laboratory results and prey distribution in an Amazonian savanna.
227 *Copeia*, 58–67 (1999).

228 6. . D. Guimaraes, R. M. Pinto, R. F. Juliano, Bufo granulosus (NCN). Predation. *Herpetol.*
229 *Rev.* **35**, 259 (2004).

230 7. L. Krenn, B. Kopp, Bufadienolides from animal and plant sources. *Phytochemistry*. **48**, 1–
231 29 (1998).

232 8. J.-D. Horisberger, Recent insights into the structure and mechanism of the sodium pump.
233 *Physiology*. **19**, 377–387 (2004).

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

236 10. M. Laursen, J. L. Gregersen, L. Yatime, P. Nissen, N. U. Fedosova, Structures and
237 characterization of digoxin-and bufalin-bound Na+, K+-ATPase compared with the
238 ouabain-bound complex. *Proc. Natl. Acad. Sci.* **112**, 1755–1760 (2015).

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

242 12. B. Ujvari, H. Mun, A. D. Conigrave, A. Bray, J. Osterkamp, P. Halling, T. Madsen,
243 Isolation breeds naivety: island living robs Australian varanid lizards of toad-toxin
244 immunity via four-base-pair mutation. *Evol. Int. J. Org. Evol.* **67**, 289–294 (2013).

245 13. B. Ujvari, N. R. Casewell, K. Sunagar, K. Arbuckle, W. Wüster, N. Lo, D. O'Meally, C.
246 Beckmann, G. F. King, E. Deplazes, Widespread convergence in toxin resistance by
247 predictable molecular evolution. *Proc. Natl. Acad. Sci.* **112**, 11911–11916 (2015).

248 14. S. Mohammadi, Z. Gompert, J. Gonzalez, H. Takeuchi, A. Mori, A. H. Savitzky, Toxin-
249 resistant isoforms of Na+/K+-ATPase in snakes do not closely track dietary specialization
250 on toads. *Proc. R. Soc. B Biol. Sci.* **283**, 20162111 (2016).

251 15. J. Orlowski, J. B. Lingrel, Tissue-specific and developmental regulation of rat Na, K-
252 ATPase catalytic alpha isoform and beta subunit mRNAs. *J. Biol. Chem.* **263**, 10436–10442
253 (1988).

254 16. L. Fagerberg, B. M. Hallström, P. Oksvold, C. Kampf, D. Djureinovic, J. Odeberg, M.
255 Habuka, S. Tahmasebpoor, A. Danielsson, K. Edlund, Analysis of the human tissue-specific
256 expression by genome-wide integration of transcriptomics and antibody-based proteomics.
257 *Mol. Cell. Proteomics*. **13**, 397–406 (2014).

258 17. S. Mohammadi, A. H. Savitzky, J. Lohr, S. Dobler, Toad toxin-resistant snake (Thamnophis
259 elegans) expresses high levels of mutant Na+/K+-ATPase mRNA in cardiac muscle. *Gene*.
260 **614**, 21–25 (2017).

261 18. B. M. Marshall, N. R. Casewell, M. Vences, F. Glaw, F. Andreone, A. Rakotoarison, G.
262 Zancolli, F. Woog, W. Wüster, Widespread vulnerability of Malagasy predators to the
263 toxins of an introduced toad. *Curr. Biol.* **28**, R654–R655 (2018).

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

266 20. G. Petschenka, V. Wagschal, M. von Tschirnhaus, A. Donath, S. Dobler, Convergently
267 evolved toxic secondary metabolites in plants drive the parallel molecular evolution of
268 insect resistance. *Am. Nat.* **190**, S29–S43 (2017).

269 21. J. N. Lohr, F. Meinzer, S. Dalla, R. Romey-Glüsing, S. Dobler, The function and
270 evolutionary significance of a triplicated Na, K-ATPase gene in a toxin-specialized insect.
271 *BMC Evol. Biol.* **17**, 1–10 (2017).

272 22. L. Yang, N. Ravikanthachari, R. Mariño-Pérez, R. Deshmukh, M. Wu, A. Rosenstein, K.
273 Kunte, H. Song, P. Andolfatto, Predictability in the evolution of Orthopteran cardenolide
274 insensitivity. *Philos. Trans. R. Soc. B*. **374**, 20180246 (2019).

275 23. J. B. Walsh, Sequence-dependent gene conversion: can duplicated genes diverge fast
276 enough to escape conversion? *Genetics*. **117**, 543–557 (1987).

277 24. J.-M. Chen, D. N. Cooper, N. Chuzhanova, C. Férec, G. P. Patrinos, Gene conversion:
278 mechanisms, evolution and human disease. *Nat. Rev. Genet.* **8**, 762–775 (2007).

279 25. H. Innan, F. Kondrashov, The evolution of gene duplications: classifying and distinguishing
280 between models. *Nat. Rev. Genet.* **11**, 97–108 (2010).

281 26. E. M. Price, J. B. Lingrel, Structure-function relationships in the sodium-potassium
282 ATPase. alpha. subunit: site-directed mutagenesis of glutamine-111 to arginine and
283 asparagine-122 to aspartic acid generates a ouabain-resistant enzyme. *Biochemistry*. **27**,
284 8400–8408 (1988).

285 27. E. Price, D. Rice, J. Lingrel, Structure-function studies of Na, K-ATPase. Site-directed
286 mutagenesis of the border residues from the H1-H2 extracellular domain of the alpha
287 subunit. *J. Biol. Chem.* **265**, 6638–6641 (1990).

288 28. R. O. de Sá, T. Grant, A. Camargo, W. R. Heyer, M. L. Ponssa, E. Stanley, Systematics of
289 the neotropical genus Leptodactylus Fitzinger, 1826 (Anura: Leptodactylidae): phylogeny,
290 the relevance of non-molecular evidence, and species accounts. *South Am. J. Herpetol.* **9**
291 (2014).

292 29. K. M. Teshima, H. Innan, Neofunctionalization of duplicated genes under the pressure of
293 gene conversion. *Genetics*. **178**, 1385–1398 (2008).

294 30. H. Shibata, T. Yamazaki, Molecular evolution of the duplicated Amy locus in the
295 Drosophila melanogaster species subgroup: concerted evolution only in the coding region
296 and an excess of nonsynonymous substitutions in speciation. *Genetics*. **141**, 223–236
297 (1995).

298 31. H. Innan, A two-locus gene conversion model with selection and its application to the
299 human RHCE and RHD genes. *Proc. Natl. Acad. Sci.* **100**, 8793–8798 (2003).

300 32. J. A. Fawcett, H. Innan, Neutral and non-neutral evolution of duplicated genes with gene
301 conversion. *Genes*. **2**, 191–209 (2011).

302 33. H. Ogawa, T. Shinoda, F. Cornelius, C. Toyoshima, Crystal structure of the sodium-
303 potassium pump (Na+, K+-ATPase) with bound potassium and ouabain. *Proc. Natl. Acad.*
304 *Sci.* **106**, 13742–13747 (2009).

305 34. S. Dalla, M. Baum, S. Dobler, Substitutions in the cardenolide binding site and interaction
306 of subunits affect kinetics besides cardenolide sensitivity of insect Na, K-ATPase. *Insect*
307 *Biochem. Mol. Biol.* **89**, 43–50 (2017).

308 35. G. G. Hammes, Unifying concept for the coupling between ion pumping and ATP
309 hydrolysis or synthesis. *Proc. Natl. Acad. Sci.* **79**, 6881–6884 (1982).

310 36. Y.-J. Feng, D. C. Blackburn, D. Liang, D. M. Hillis, D. B. Wake, D. C. Cannatella, P.
311 Zhang, Phylogenomics reveals rapid, simultaneous diversification of three major clades of
312 Gondwanan frogs at the Cretaceous–Paleogene boundary. *Proc. Natl. Acad. Sci.* **114**,
313 E5864–E5870 (2017).

314 37. A. Harpak, X. Lan, Z. Gao, J. K. Pritchard, Frequent nonallelic gene conversion on the
315 human lineage and its effect on the divergence of gene duplicates. *Proc. Natl. Acad. Sci.*
316 **114**, 12779–12784 (2017).

317 38. B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B.
318 Couger, D. Eccles, B. Li, M. Lieber, De novo transcript sequence reconstruction from
319 RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**,
320 1494 (2013).

321 39. D. R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de
322 Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).

323 40. M. H. Schulz, D. R. Zerbino, M. Vingron, E. Birney, Oases: robust de novo RNA-seq
324 assembly across the dynamic range of expression levels. *Bioinformatics*. **28**, 1086–1092
325 (2012).

326 41. G. W. Vurture, F. J. Sedlazeck, M. Nattestad, C. J. Underwood, H. Fang, J. Gurtowski, M.
327 C. Schatz, GenomeScope: fast reference-free genome profiling from short reads.
328 *Bioinformatics*. **33**, 2202–2204 (2017).

329 42. N. I. Weisenfeld, V. Kumar, P. Shah, D. M. Church, D. B. Jaffe, Direct determination of
330 diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).

331 43. M. Seppey, M. Manni, E. M. Zdobnov, in *Gene Prediction* (Springer, 2019), pp. 227–245.

332 44. S. M. Kiełbasa, R. Wan, K. Sato, P. Horton, M. C. Frith, Adaptive seeds tame genomic
333 sequence comparison. *Genome Res.* **21**, 487–493 (2011).

334 45. H. Li, seqtk Toolkit for processing sequences in FASTA/Q formats. *GitHub*. **767**, 69
335 (2012).

336 46. S. Koren, B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, A. M. Phillippy, Canu:
337 scalable and accurate long-read assembly via adaptive k-mer weighting and repeat
338 separation. *Genome Res.* **27**, 722–736 (2017).

339 47. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. **34**, 3094–
340 3100 (2018).

341 48. R. Vaser, I. Sović, N. Nagarajan, M. Šikić, Fast and accurate de novo genome assembly
342 from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).

343 49. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high
344 throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

345 50. M. Gouy, S. Guindon, O. Gascuel, SeaView version 4: a multiplatform graphical user
346 interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–
347 224 (2010).

348 51. B. Subramanian, S. Gao, M. J. Lercher, S. Hu, W.-H. Chen, Evolview v3: a webserver for
349 visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* **47**,
350 W270–W275 (2019).

351 52. M. Stanke, A. Tzvetkova, B. Morgenstern, AUGUSTUS at EGASP: using EST, protein and
352 genomic alignments for improved gene prediction in the human genome. *Genome Biol.* **7**,
353 1–8 (2006).

53. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, R. Lanfear, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

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

55. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

56. M. Kimura, On the probability of fixation of mutant genes in a population. *Genetics*. **47**, 713 (1962).

57. J. H. Gillespie, *Population genetics: a concise guide* (JHU Press, 2004).

58. Y.-B. Sun, Z.-J. Xiong, X.-Y. Xiang, S.-P. Liu, W.-W. Zhou, X.-L. Tu, L. Zhong, L. Wang, D.-D. Wu, B.-L. Zhang, Whole-genome sequence of the Tibetan frog Nanorana parkeri and the comparative evolution of tetrapod genomes. *Proc. Natl. Acad. Sci.* **112**, E1257–E1262 (2015).

59. A. J. Crawford, Relative rates of nucleotide substitution in frogs. *J. Mol. Evol.* **57**, 636–641 (2003).

60. S. P. Mansai, H. Innan, The power of the methods for detecting interlocus gene conversion. *Genetics*. **184**, 517–527 (2010).

61. D. Feng, L. Tierney, Computing and displaying isosurfaces in R. *J. Stat. Softw.* **28**, 1–24 (2008).

62. B. Venables WNRipley, Modern applied statistics with S. *Stat. Comput. N. Y. Springer* (2002).

63. V. A. Luckow, S. Lee, G. Barry, P. Olins, Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in Escherichia coli. *J. Virol.* **67**, 4566–4579 (1993).

64. G. Petschenka, S. Fandrich, N. Sander, V. Wagschal, M. Boppré, S. Dobler, Stepwise evolution of resistance to toxic cardenolides via genetic substitutions in the Na+/K+-ATPase of milkweed butterflies (Lepidoptera: Danaini). *Evolution*. **67**, 2753–2761 (2013).

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

Acknowledgments: We thank C. Natarajan, K. Rohlfing, V. Wagschal, and P. Kowalski for assistance in the laboratory. **Funding:** This study was funded by grants to PA from the National Institutes of Health (R01- GM115523) and to JFS from the National Institutes of Health (R01-HL087216) and the National Science Foundation (OIA-1736249), to SD from Deutsche Forschungsgemeinschaft (DFG grant DO527/10-1), and by a fellowship from to AH by The Simons Foundation's Society of Fellows (#633313). **Author contributions:** PA and AJC conceived of and oversaw the project; LY, MPRO, SHA, JP, and AJC collected samples and generated data; LY, AH, and KZ performed evolutionary and population genetics analyses; SM, JFS, SD, AJC and PA designed functional experiments; SM and PA performed experiments and statistical analyses; SM, JFS, LY, AH and PA wrote the paper; All authors edited the manuscript; **Competing interests:** None. **Data and materials availability:** BioProject PRJNA627222, SRA: SRR11583961-91, nucleotides: MT396181-92, genes: MT422192-MT422203, genome: TBD.

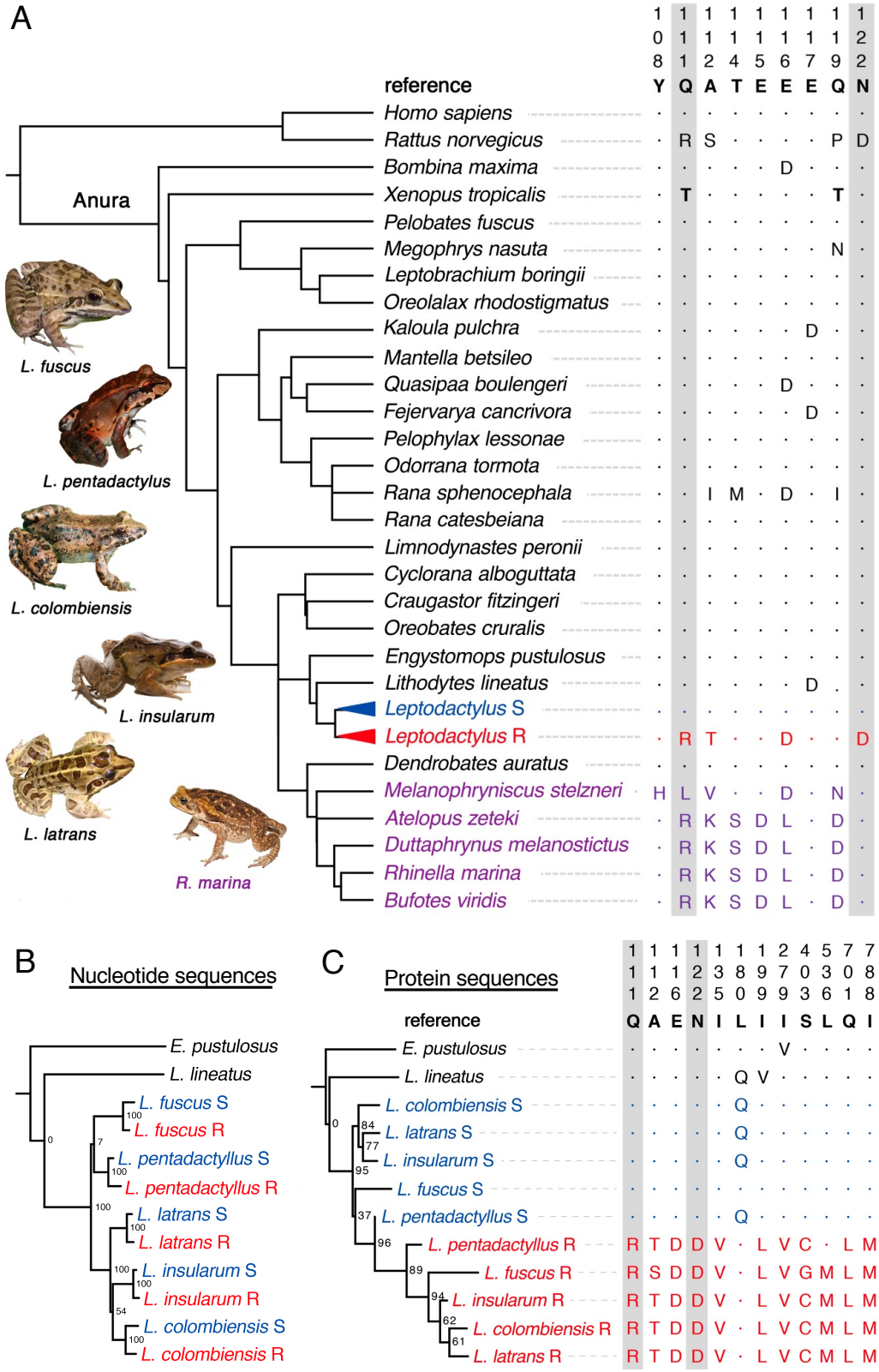
Supplementary Materials:

Materials and Methods

Figures S1-S6

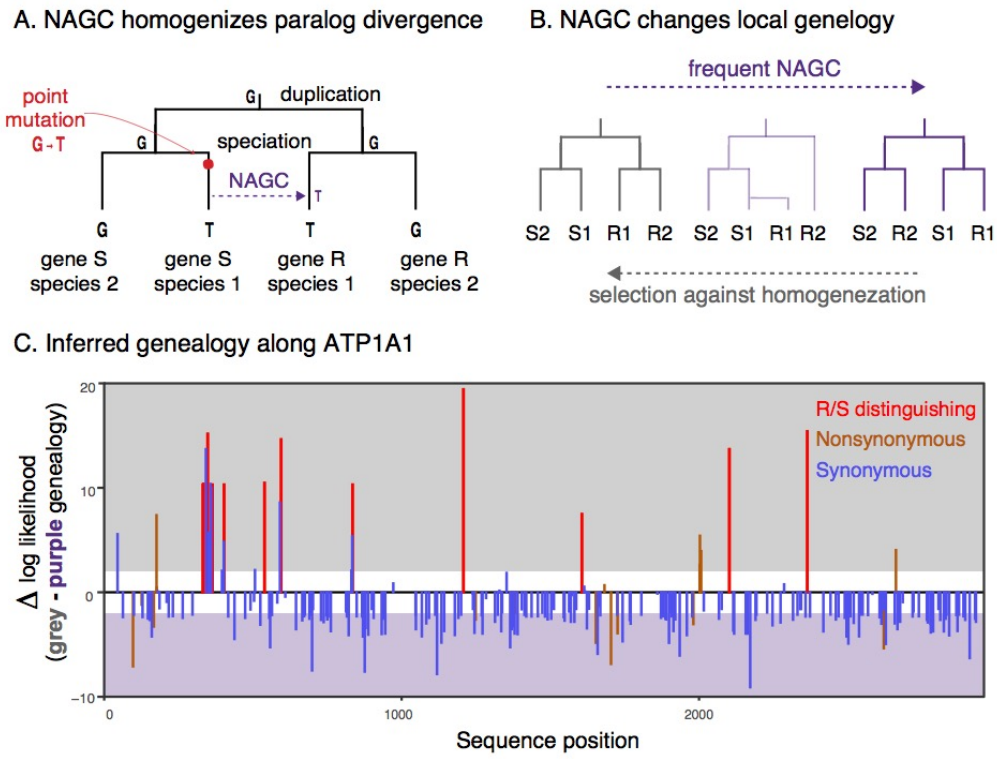
Tables S1-S8

References 38-65


A
reference
Y Q A T E E E Q N
Homo sapiens
Rattus norvegicus
Bombina maxima
Xenopus tropicalis
Pelobates fuscus
Megophrys nasuta
Leptobrachium boringii
Oreolalax rhodostigmatus
Kaloula pulchra
Mantella betsileo
Quasipaa boulengeri
Fejervarya cancrivora
Pelophylax lessonae
Odorrana tormota
Rana sphenocephala
Rana catesbeiana
Limnodynastes peronii
Cyclorana alboguttata
Craugastor fitzingeri
Oreobates cruralis
Engystomops pustulosus
Lithodytes lineatus
Leptodactylus S
Leptodactylus R
Dendrobates auratus
Melanophryniscus stelzneri
Atelopus zeteki
Duttaphrynus melanostictus
Rhinella marina
Bufotes viridis
Anura
L. fuscus
L. pentadactylus
L. colombiensis
L. insularum
L. latrans
R. marina
B
Nucleotide sequences
E. pustulosus
L. lineatus
L. fuscus S
L. fuscus R
L. pentadactyllus S
L. pentadactyllus R
L. latrans S
L. latrans R
L. insularum S
L. insularum R
L. colombiensis S
L. colombiensis R
C
Protein sequences
reference
Q A E N I L I I S L Q I
E. pustulosus
L. lineatus
L. colombiensis S
L. latrans S
L. insularum S
L. fuscus S
L. pentadactyllus S
L. pentadactyllus R
L. fuscus R
L. insularum R
L. colombiensis R
L. latrans R

404

405 **Fig. 1. Molecular evolution of ATP1A1 in anurans. (**A) A maximum likelihood phylogeny of
406 anuran species and mammalian outgroups derived from (*36*). Species names in purple
407 correspond to chemically defended toads, and blue and red colors correspond to the S and R
408 ATP1A1 paralogs in *Leptodactylus* species, respectively. Only variant sites with roles in the CG-
409 binding or sensitivity are shown (reviewed in (*22*)). The numbering of sites is based on sheep
410 ATP1A1 (*Ovis aries*, Genbank: NC019458.2) and appears at the top of the table (e.g. the first
411 position showed is 108). Dots indicate identity with the consensus sequence of anuran species
412 and letters represent amino acid substitutions relative to the reference. The images on the left are
413 the five surveyed *Leptodactylus* species and a representative prey toad (*Rhinella marina*). (B) A
414 maximum likelihood phylogeny estimated with nucleotide sequences with bootstrap support
415 values at nodes. (C) A maximum likelihood genealogy based on protein sequences with
416 bootstrap support values at nodes. To the right is the pattern of amino acid variation at 12
417 positions that most strongly distinguish the S and R paralogs.
418
419
420



421

422 **Fig. 2. (A)** Non-allelic gene conversion (NAGC) homogenizes sequence variation between
423 paralogous genes, and therefore changes the genealogical signal. (This illustration is adapted
424 from Harpak et al. (*37*)). **(B)** Frequent NAGC can result in a genealogy in which paralogous
425 genes in the same species appear to share a more recent common ancestor with one another than
426 with their orthologous counterparts in other species. In principle, the homogenizing effects of
427 NAGC can be counteracted by selection that favors the differentiation of paralogous genes. **(C)**
428 Site-by-site difference in the likelihood of two alternative tree topologies—generalizing the grey
429 and purple extremes of panel (B) to five *Leptodactylus* species. Shaded regions show a log
430 likelihood difference greater than 2 in support of the corresponding model. Only parsimony-
431 informative variants in the ATP1A1 coding sequence are shown. Red bars correspond to the 12
432 R/S distinguishing nonsynonymous substitutions (shown in red in Fig 1C; Table S4).
433

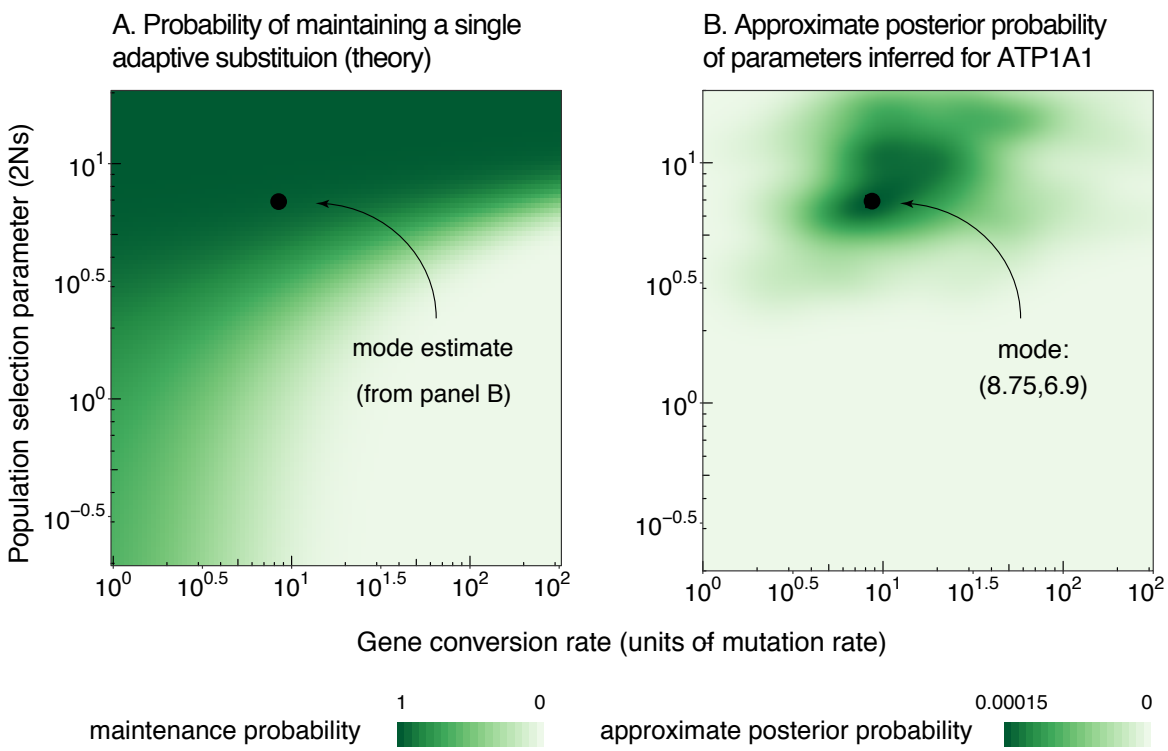


Fig. 3. (A) Theoretical probability of maintaining distinct alleles at a single site in the face of non-allelic gene conversion (NAGC). We used a theoretical model to compute the probability of maintaining alternative alleles at the same site in a pair of paralogous genes, given an NAGC rate and strength of selection against allele homogenization at the site. The black dot shows the approximate mode estimate from panel B, which falls in the range in which maintenance is likely according to this theoretical model. **(B) Estimates of evolutionary parameters.** Approximate posterior probabilities were inferred based on simulations of the evolution of ATP1A1 genes in *Leptodactylus*. The x-axis shows the NAGC rate across the gene, and the y-axis shows the population selection coefficient for the 12 substitutions that distinguish the R and S paralogs across species.

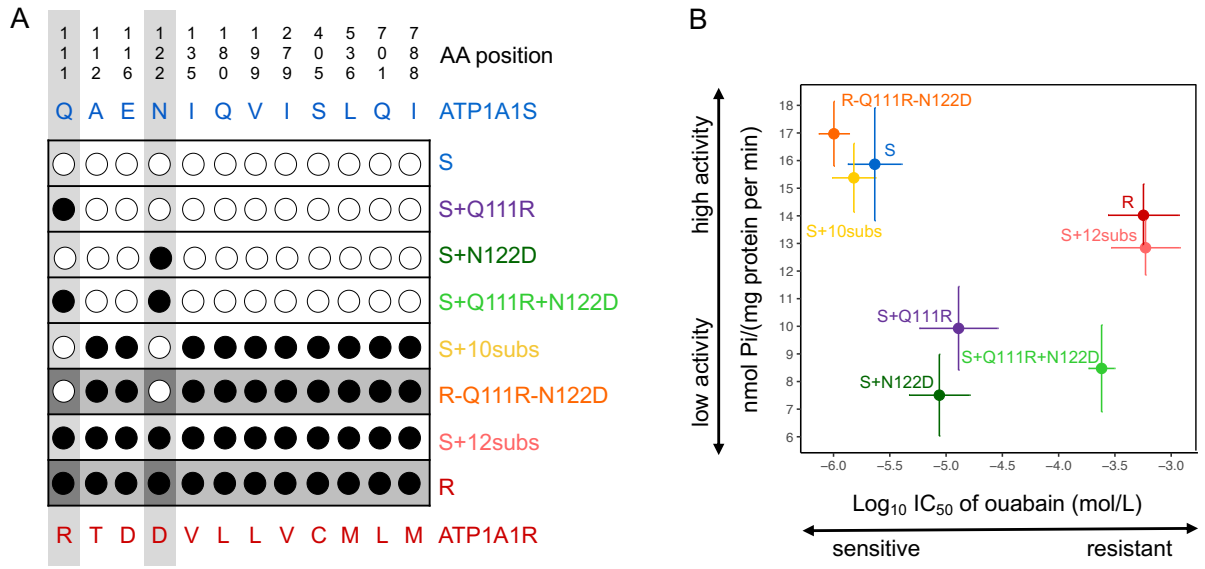


Fig. 4. Functional analysis of R-copy-specific substitutions. (A) ATP1A1 gene constructs designed to produce proteins with various combinations of the 12 variants distinguishing the S and R paralogs. Black circles indicate an amino acid matching the R copy whereas a white circle indicates a match with the S copy. A dark grey shaded background indicates that the construct was produced on the R genetic background and a white background indicates that the construct was produced on the S background. Light grey columns highlight two variants (Q111R and N122D) previously known to confer CG-resistance. (B) The joint functional properties of engineered Na^+,K^+-ATPases. Mean ± SEM Log IC_{50} (i.e., measurement of CG resistance) of the six biological replicates is plotted on the x-axis and mean ± SEM ATP hydrolysis rate (i.e., measurement of protein activity) for the same proteins is plotted on the y-axis. R and S copies differ by the 12 amino acid substitutions being studied and by 9 additional amino acid substitutions and one two amino acid insertion-deletion difference (see Table S6).