Title: Concerted evolution reveals co-adapted amino acid substitutions in Na^+K^+-ATPase of 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 paralog 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.

47 **Main Text:**

48

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

62

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

71 substitutions known to confer resistance to CGs while the sensitive (S) paralog appears to have
72 retained the ancestral susceptibility to CGs. Neofunctionalization of ATP1A paralogs has
73 contributed to the evolution of CG-resistance in numerous insect lineages (*19–22*) but appear to
74 be rare among CG-resistant vertebrates. Further, the fate of duplicated genes and the probability
75 that they will neofunctionalize is predicted to depend on the strength of selection for functional
76 differentiation relative to the rate of non-allelic gene conversion (NAGC), a form of
77 nonreciprocal genetic exchange that homogenizes sequence variation between duplicated genes,
78 thereby impeding divergence (*23–25*). The ATP1A1 duplication in *Leptodactylus* provides an
79 ideal opportunity to explore this process because the functional differentiation between R and S
80 paralogs has clear adaptive significance with regard to CG-resistance.

81

82 We surveyed the full-length coding sequences of all ATP1A paralogs in *Leptodactylus* and other
83 anurans using RNA-seq-based gene discovery (*19*; Tables S1). Our results confirm that ATP1A1
84 is duplicated in *L. latrans* (*11*) and indicate that the duplication of ATP1A1 most likely occurred
85 in the common ancestor of all surveyed *Leptodactylus* species (Fig. 1A, Table S2). Two other
86 ancient paralogs common to vertebrates, ATP1A2 and ATP1A3, appear to be present as single-
87 copy genes and lack any known CG-resistant substitutions in *Leptodactylus* (Fig. S2). The αM1–
88 2 transmembrane domains of the ATP1A1 paralogs in *L. latrans* are distinguished by four amino
89 acid substitutions (*11*; Fig 1C, Fig. S3). Two of these substitutions, Q111R and N122D, were
90 first identified in rat ATP1A1 and have been shown to interact synergistically to confer CG-
91 resistance to sheep ATP1A1 protein *in vitro* (*26*, *27*). Comparison of ATP1A1 sequences in five
92 distantly related *Leptodactylus* species (*28*) reveals that they each harbor a putatively resistant
93 paralog (R) that includes the Q111R and N122D substitutions and a putatively sensitive ATP1A1

94 paralog (S) that lacks these substitutions. In addition to Q111R and N122D, there are 10 other
95 amino acid substitutions distinguishing the R and S paralogs in most of the five sampled species
96 (Fig. 1C). Hereafter, we refer to these twelve substitutions as "R/S distinguishing".

97

98 To infer when duplication occurred relative to speciation events, we estimated phylogenies from
99 an alignment of ATP1A1 coding sequence. Phylogenies estimated from nucleotide and inferred
100 amino-acid sequences support strikingly different topologies (Fig 1B and C). Without further
101 information, the genealogy based on nucleotide sequences (primarily based on variation at
102 synonymous sites) would imply independent duplications in each of the *Leptodactylus* species,
103 followed by parallel substitutions at the same 12 R/S distinguishing amino acid positions (Fig.
104 1B). Instead, it seems much more likely that this genealogical pattern reflects a single ancestral
105 duplication — as indicated by genealogy based on amino acid sequences (Fig. 1C, Table S4) —
106 coupled with on-going NAGC between the R and S paralogs of each species. Frequent NAGC
107 produces a pattern of "concerted evolution" whereby tandemly linked paralogs from the same
108 species are more similar to one another than they are to their orthologous counterparts in other
109 species (*29;* Fig. 2A,B). By generating a *de novo* genome assembly of *L. fuscus* based on single-
110 molecule sequencing, we established that S and R copies are indeed arranged in tandem and in
111 the same orientation, and are therefore likely to be subject to NAGC (Table S3, Fig. S4). We
112 thus propose that the persistence of the 12 amino acid differences between the two paralogs is
113 due to selection counteracting the homogenizing effects of NAGC (*23*, *30*; Fig. 2B), thereby
114 maintaining an adaptive division-of-labor between the R and S copies.

115

116 The opposing forces of NAGC and selection are predicted to leave a characteristic genealogical
117 signature at neutral sites closely linked to the targets of selection (*30*; Fig. 2B). We therefore
118 tested the relationship between the genealogical signature and distance from nonsynonymous
119 variants putatively under selection. To this end, for all informative sites, we evaluated the level
120 of support for an ancient duplication of ATP1A1 in the common ancestor of all *Leptodactylus*
121 species (with no concerted evolution) relative to support for an alternative in which ATP1A1
122 paralogs within species are always more closely related to one another than they are to paralogs
123 in other species (as expected under concerted evolution). This analysis reveals that synonymous
124 (presumed to be neutral) variants congruent with an ancient duplication of R and S have a
125 median distance of 4 bp from nonsynonymous variants exhibiting the same pattern (Fig. 2C). In
126 contrast, randomly sampled synonymous sites supporting the alternative genealogy (i.e.
127 concerted evolution) have a median distance of 88 bp from those nonsynonymous variants
128 (bootstrap $p<10^{-5}$). This pattern at synonymous sites is consistent with a scenario in which
129 purifying selection maintains functionally important sequence differences between
130 neofunctionalized gene duplicates in the face of NAGC.

131

132 We next quantified the strength of purifying selection required to maintain the amino acid
133 differentiation between R and S duplicates in the face of NAGC. We first considered population
134 genetics theory for the evolution of a single site in tandem duplicates (*31*) (Supplementary
135 Materials). This analytic model predicts that if the rate of NAGC is an order of magnitude higher
136 than the rate of point mutation, then the maintenance of alternative amino acid states is only
137 likely under sufficiently strong purifying selection — namely, when the selection coefficient
138 scaled by population size, $2Ns$, is larger than one (Fig. 3A). We next developed an inference

139 method based on simulations of ATP1A1 evolution to estimate the combination of parameters
140 that best explain divergence patterns throughout the gene, including levels of paralog divergence
141 observed as a function of distance from the 12 R/S distinguishing substitutions (Supplementary
142 Materials). We estimate the rate of NAGC to be an order of magnitude higher than the point
143 mutation rate (posterior mode 9 with 80% credible interval 4-54 times the point mutation rate),
144 and 2*Ns* substantially larger than one (posterior mode 9; 80% credible interval 5-18; Fig. 3B).
145 These estimates fall within the plausible range predicted by the theoretical single-site model (Fig.
146 3A). These results indicate that the observed pattern of divergence between R and S paralogs
147 reflects a history of strong purifying selection that maintains fixed differences between them
148 despite high rates of NAGC.

149

150 The inference that selection maintains the co-occurrence of the 12 R/S distinguishing
151 substitutions implies they are functionally important and collectively contribute to organismal
152 fitness. The effects of Q111R and N122D on CG-insensitivity have previously been
153 demonstrated by *in vitro* enzyme inhibition assays (*9*). Additionally, while not related directly to
154 CG-resistance, the potential importance of substitutions at sites 112 and 116 has been suggested
155 by molecular evolution analysis and structural studies, respectively (*11*, *32*). However, the
156 remaining eight R/S distinguishing substitutions are located in structural domains that have not
157 been implicated in CG-resistance. Since our analysis suggests that amino acid divergence
158 between R and S paralogs is maintained by selection, we performed protein-engineering
159 experiments to elucidate the functional significance of the 12 R/S distinguishing substitutions.
160 We synthesized and recombinantly expressed eight mutant Na^+,K^+-ATPase proteins, each
161 harboring different combinations of R-specific replacements on both S- and R-type genetic

162 backgrounds of a representative species, *L. latrans* (Fig. 4A, Table S6), and we then quantified
163 the level of CG-resistance of each genotype using enzyme-inhibition assays (Table S7, Fig. S7,
164 (*33*)). Individually, Q111R and N122D significantly increased CG-resistance by 21-fold and 14-
165 fold, respectively (ANOVA p=2.7e-13 and p=2.3e-6; Fig. 4B, Table S8). When combined,
166 Q111R and N122D produce a greater than 100-fold increase in CG resistance relative to the S
167 paralog (Tukey's HSD test, adjusted p<4e-5, Fig. 4B, Tables S6-S8). In contrast, the remaining
168 10 substitutions had no detectable net effect on CG-resistance when jointly added to the S
169 background (p=0.22, Fig. 4B).
170
171 Given the absence of detectable effects of R/S distinguishing substitutions other than Q111R and
172 N122D on CG-resistance, we tested whether these substitutions had effects on other aspects of
173 ATP1A1 function. Since ATP hydrolysis and ion co-transport are strongly coupled functions of
174 Na^+,K^+-ATPase (*34*), we used estimates of the rate of ATP hydrolysis in the absence of ouabain
175 as a proxy for overall protein activity. Based on this assay, we found that CG-resistance
176 substitutions Q111R and N122D significantly impair activity, individually reducing ATPase
177 activity by an average of 40% (p=0.024 and p=7.7e-4 respectively; Fig. 4B; Table S8). We also
178 detected a significant interaction between Q111R and N122D that renders their joint effects
179 somewhat less severe than predicted by the sum of their individual effects (*i.e.*, a 30% reduction
180 rather than the expected 78% reduction, p=0.022). Critically, adding the remaining 10 R-specific
181 substitutions on the S background containing Q111R and N122D restores normal ATPase
182 activity close to wild-type levels. An analysis of variance reveals a highly significant effect of
183 these 10 R-specific substitutions ($p=1.2 \times 10^{-4}$, Fig. 4B, Table S8). Our results thus indicate that
184 these 10 R/S distinguishing substitutions play a vital role in compensating for the negative

185 pleiotropic effects of the resistance-conferring substitutions, Q111R and N122D. We conclude
186 that the evolution of the R protein from a CG-sensitive ancestral state involved two epistatically-
187 interacting substitutions (Q111R and N122D) in conjunction with compensatory effects of 10
188 additional substitutions that mitigate the trade-off between toxin resistance and native enzyme
189 activity.
190
191 The adaptive division-of-labor between the R and S paralogs of ATP1A1 in *Leptodactylus* has
192 been maintained by strong selection that has counteracted the homogenizing effects of frequent
193 NAGC over the 35 million-year history of this genus. Similar signatures of selection to maintain
194 sequence differentiation between neofunctionalized duplicates have been observed for the
195 RHCE/RHD antigen proteins of humans (*35*), "major facilitator family" transporter proteins in
196 *Drosophila* (*36*) and red/green opsins of primates (*30*). To our knowledge, only the case of
197 opsins has been linked directly to functional differentiation, notably two closely-linked amino
198 substitutions contributing to a red to green shift in absorbance maxima (*37*). Our study highlights
199 similar signatures of selection not only on the two amino acid substitutions directly linked to
200 adaptive differentiation for CG-resistance, but also at 10 more amino acid substitutions scattered
201 throughout the protein that facilitate this neofunctionalization. Thus, by identifying interactions
202 between adaptive substitutions and the genetic backgrounds that permit these changes, our
203 combination of evolutionary and functional analyses reveals how mechanisms of adaptation are
204 shaped by intramolecular epistasis and pleiotropy.
205
206
207

208 **References and Notes:**

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

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

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

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

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

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

221 7. L. Krenn, B. Kopp, Bufadienolides from animal and plant sources. *Phytochemistry*. **48**, 1–
222 29 (1998).

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

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

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

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

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

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

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

242 15. J. Orlowski, J. B. Lingrel, Tissue-specific and developmental regulation of rat Na, K-
243 ATPase catalytic alpha isoform and beta subunit mRNAs. *J. Biol. Chem.* **263**, 10436–10442
244 (1988).

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

249 17. S. Mohammadi, A. H. Savitzky, J. Lohr, S. Dobler, Toad toxin-resistant snake (Thamnophis
250 elegans) expresses high levels of mutant Na+/K+-ATPase mRNA in cardiac muscle. *Gene*.
251 **614**, 21–25 (2017).

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

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

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

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

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

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

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

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

272 26. E. M. Price, J. B. Lingrel, Structure-function relationships in the sodium-potassium
273 ATPase. alpha. subunit: site-directed mutagenesis of glutamine-111 to arginine and
274 asparagine-122 to aspartic acid generates a ouabain-resistant enzyme. *Biochemistry*. **27**,
275 8400–8408 (1988).

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

279 28. R. O. de Sá, T. Grant, A. Camargo, W. R. Heyer, M. L. Ponssa, E. Stanley, Systematics of
280 the neotropical genus *Leptodactylus* Fitzinger, 1826 (Anura: Leptodactylidae): phylogeny,
281 the relevance of non-molecular evidence, and species accounts. *South Am. J. Herpetol.* **9**
282 (2014).

283 29. K. M. Teshima, H. Innan, The effect of gene conversion on the divergence between
284 duplicated genes. *Genetics*. **166**, 1553–1560 (2004).

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

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

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

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

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

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

299 36. N. Osada, H. Innan, Duplication and gene conversion in the Drosophila melanogaster
300 genome. *PLoS Genet*. **4**, e1000305 (2008).

301 37. S. Yokoyama, F. B. Radlwimmer, The molecular genetics and evolution of red and green
302 color vision in vertebrates. *Genetics*. **158**, 1697–1710 (2001).

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

307 39. A. Harpak, X. Lan, Z. Gao, J. K. Pritchard, Frequent nonallelic gene conversion on the
308 human lineage and its effect on the divergence of gene duplicates. *Proc. Natl. Acad. Sci.*
309 **114**, 12779–12784 (2017).

310 40. B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B.
311 Couger, D. Eccles, B. Li, M. Lieber, De novo transcript sequence reconstruction from
312 RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**,
313 1494 (2013).

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

316 42. M. H. Schulz, D. R. Zerbino, M. Vingron, E. Birney, Oases: robust de novo RNA-seq
317 assembly across the dynamic range of expression levels. *Bioinformatics*. **28**, 1086–1092
318 (2012).

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

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

324 45. M. Seppey, M. Manni, E. M. Zdobnov, BUSCO: Assessing Genome Assembly and
325 Annotation Completeness in *Gene Prediction* (Springer, 2019), pp. 227–245.

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

328 47. H. Li, seqtk Toolkit for processing sequences in FASTA/Q formats. *GitHub*. **767**, 69
329 (2012).

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

333 49. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. **34**, 3094–
334 3100 (2018).

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

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

339 52. M. Gouy, S. Guindon, O. Gascuel, SeaView version 4: a multiplatform graphical user
340 interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–
341 224 (2010).

342 53. Z. He, H. Zhang, S. Gao, M. J. Lercher, W.-H. Chen, S. Hu, Evolview v2: an online
343 visualization and management tool for customized and annotated phylogenetic trees.
344 *Nucleic Acids Res.* **44**, W236–W241 (2016).

345 54. M. Stanke, A. Tzvetkova, B. Morgenstern, AUGUSTUS at EGASP: using EST, protein and
346 genomic alignments for improved gene prediction in the human genome. *Genome Biol.* **7**,
347 1–8 (2006).

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

351 56. Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**,
352 1586–1591 (2007).

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

355 58. M. Kimura, On the probability of fixation of mutant genes in a population. *Genetics*. **47**,
356 713 (1962).

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

358 60. Y.-B. Sun, Z.-J. Xiong, X.-Y. Xiang, S.-P. Liu, W.-W. Zhou, X.-L. Tu, L. Zhong, L. Wang,
359 D.-D. Wu, B.-L. Zhang, Whole-genome sequence of the Tibetan frog Nanorana parkeri and
360 the comparative evolution of tetrapod genomes. *Proc. Natl. Acad. Sci.* **112**, E1257–E1262
361 (2015).

362 61. A. J. Crawford, Relative rates of nucleotide substitution in frogs. *J. Mol. Evol.* **57**, 636–641
363 (2003).

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

366 63. D. Feng, L. Tierney, Computing and displaying isosurfaces in R. *J. Stat. Softw.* **28**, 1–24
367 (2008).

368 64. W. N. Venables, B. D. Ripley, Modern applied statistics with S. N. Y. Springer (2002).

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

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

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

377

Acknowledgments: We thank M. Przeworski for helpful comments on the manuscript. 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 a fellowship to AH from 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 sequence data; LY, AH, PA, SHA, 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.

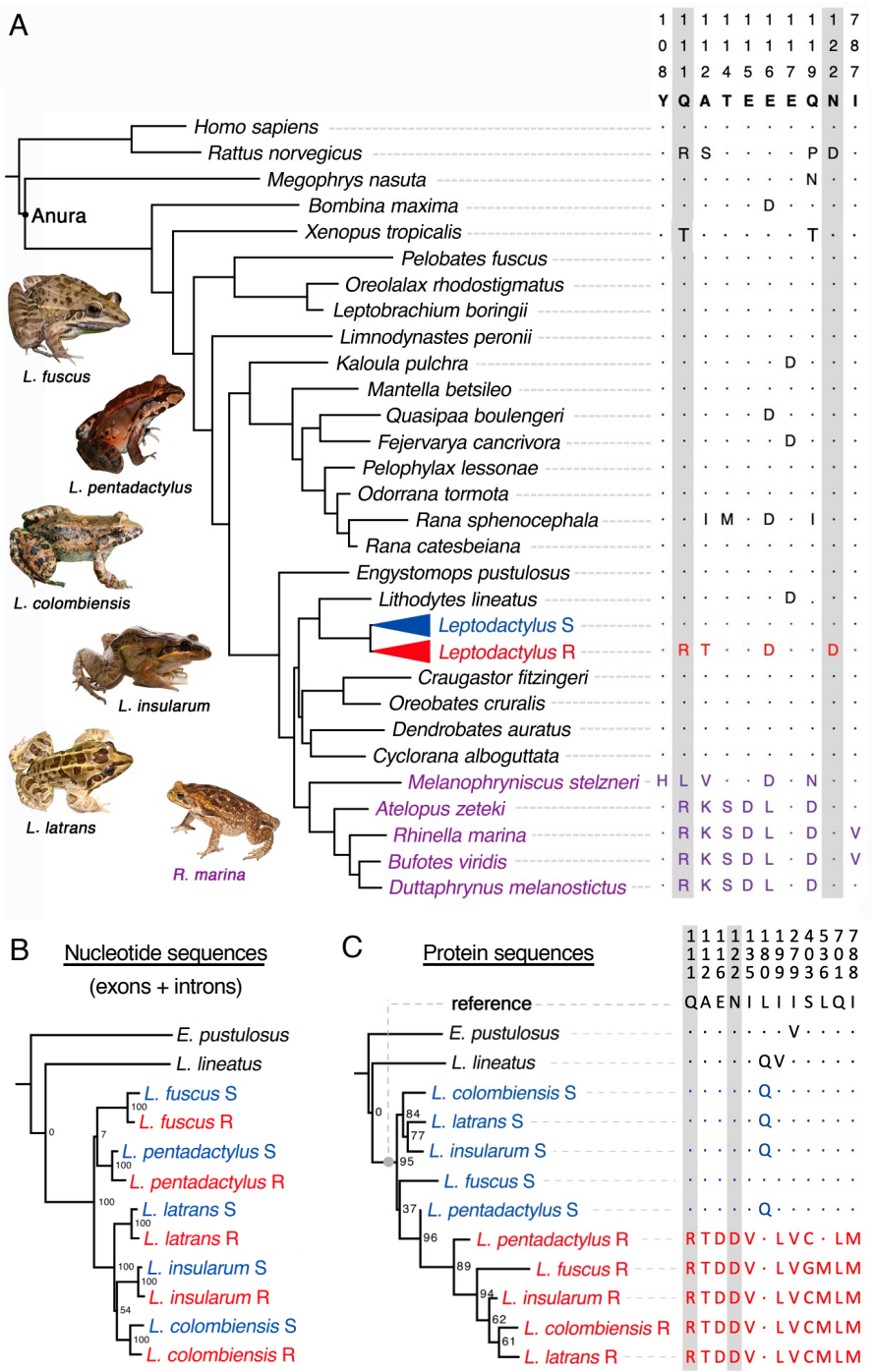
Supplementary Materials:

Methods

Figures S1-S7

Tables S1-S8

References 38-67



401

402 **Fig. 1. Molecular evolution of ATP1A1 in anurans. (**A) Maximum likelihood phylogeny of

403 anuran species and mammalian outgroups derived from (*38*). Species names in purple

404 correspond to chemically defended toads, and blue and red colors correspond to the S and R

405 ATP1A1 paralogs in *Leptodactylus* species, respectively. Only variable sites with documented

406 roles in CG-binding or sensitivity are shown (reviewed in (*22*)). The numbering of sites is based
407 on sheep ATP1A1 (*Ovis aries*, Genbank: NC019458.2) and appears at the top of the table (e.g.
408 the first position shown is 108). Dots indicate identity with the reference sequence and letters
409 represent amino acid substitutions relative to the reference. The images on the left depict the five
410 surveyed *Leptodactylus* species and a representative toad species (*Rhinella marina*) as potential
411 prey. Maximum likelihood phylogeny estimates based on nucleotide sequences (B) and amino
412 acid sequences (C) yield distinct topologies. Bootstrap support values are indicated at internal
413 nodes. To the right is the pattern of amino acid variation at 12 positions that distinguish the S and
414 R paralogs.

415

416

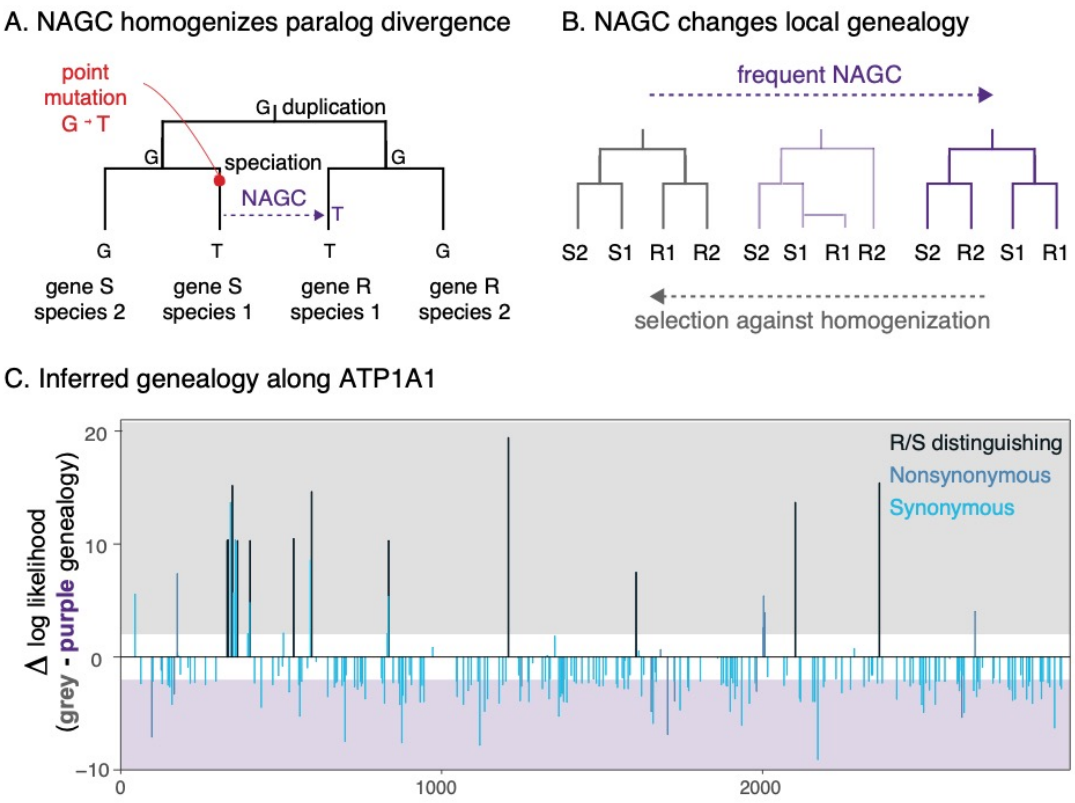


Fig. 2. (A) Non-allelic gene conversion (NAGC) homogenizes sequence variation between paralogous genes, and therefore changes the genealogical signal (adapted from (*39*)). **(B)** NAGC can result in a genealogy in which paralogous genes in the same species share a more recent common ancestor with one another than with their orthologous counterparts in other species ("concerted evolution"). The homogenizing effects of NAGC can be counteracted by selection that favors the differentiation of paralogous genes. **(C)** Site-wise difference in the $\log_{10}$-likelihood of two alternative tree topologies—generalizing the grey and purple extremes of panel B to five *Leptodactylus* species. Shaded regions show a log-likelihood difference greater than 2 in support of the corresponding model. Only parsimony-informative variants in the ATP1A1 coding sequence are shown. Black bars correspond to the 12 R/S distinguishing nonsynonymous substitutions (shown in red or blue in Fig 1C; Table S4).

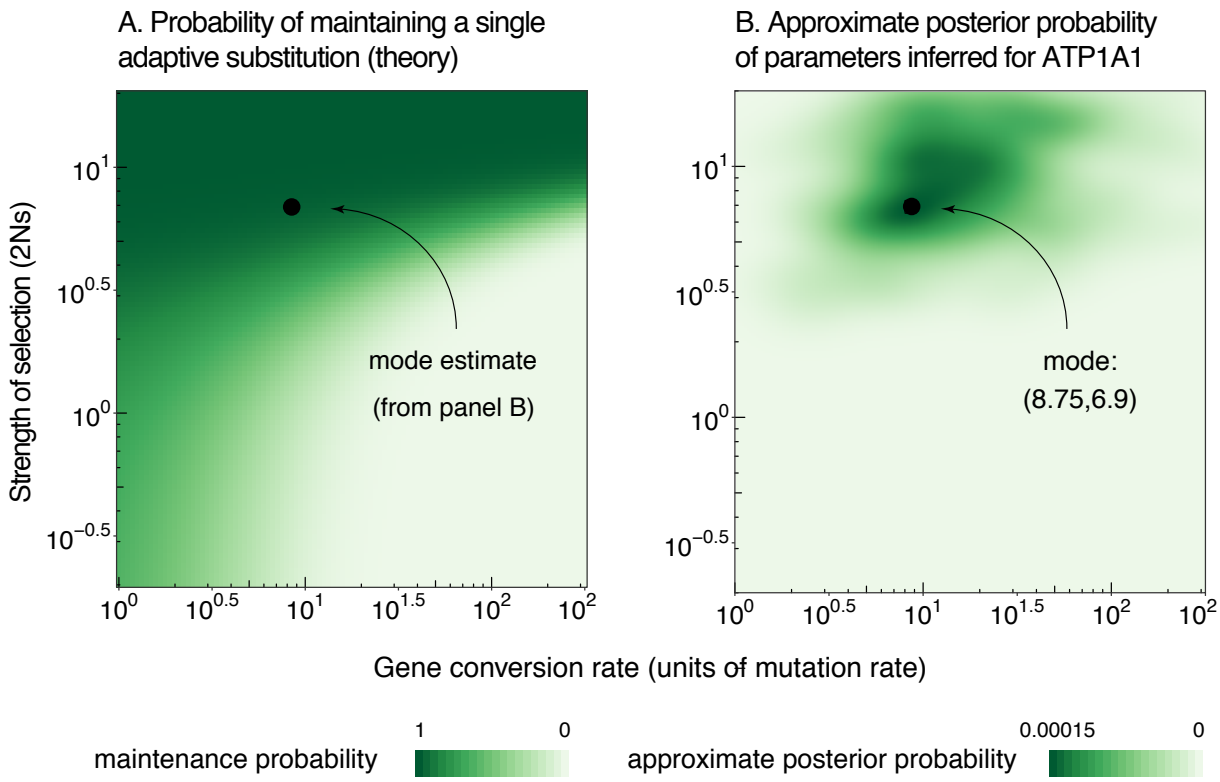


Fig. 3. (A) Theoretical probability of maintaining distinct alleles at a single site in the face of non-allelic gene conversion (NAGC). We used a theoretical model to compute the probability of maintaining alternative amino acid states at the same site in a pair of paralogous genes, given an NAGC rate 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.

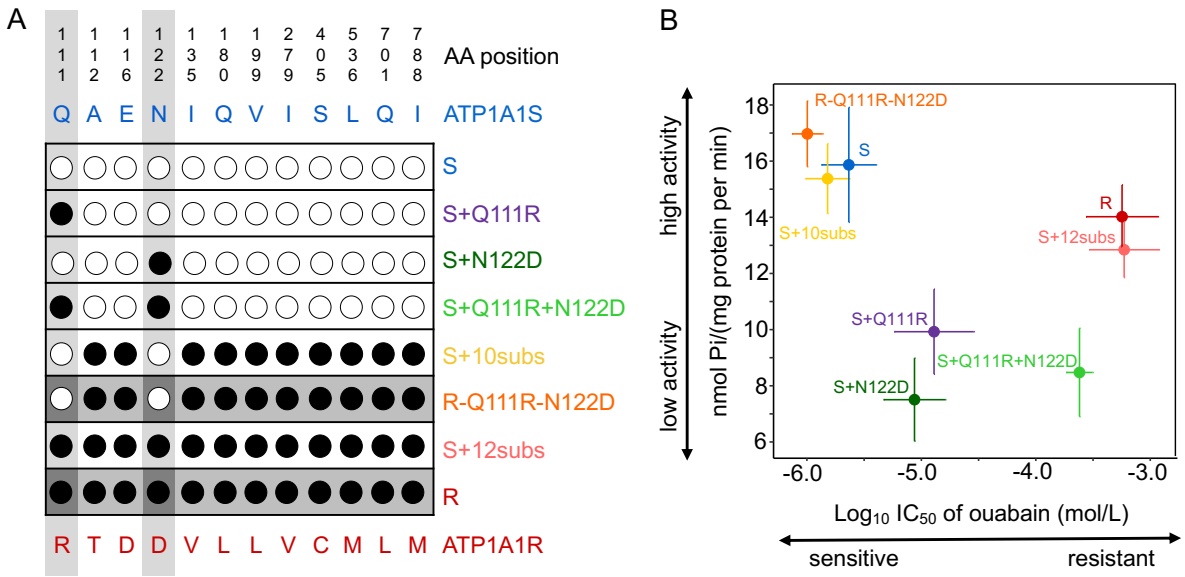


Fig. 4. Functional analysis of substitutions specific to the R-type ATP1A1 paralog. (A) ATP1A1 gene constructs with various combinations of the 12 substitutions that distinguish the S and R paralogs. Black circles indicate an amino acid matching the R paralog whereas a white circle indicates a match with the S paralog. Dark grey shading denotes the R background and white denotes the S background. Light grey columns highlight two substitutions (Q111R and N122D) that are known to confer CG-resistance. (B) Functional properties of engineered Na^+,K^+-ATPases. The mean ± SEM $\log_{10}IC_{50}$ (i.e., a measure of CG resistance) is plotted on the x-axis and the mean ± SEM ATP hydrolysis rate (i.e., a measure of protein activity) for the same proteins is plotted on the y-axis. Each estimate is based on six biological replicates.