Article

Current Biology

Concerted evolution reveals co-adapted amino acid substitutions in Na^+K^+-ATPase of frogs that prey on toxic toads

Highlights

- ATP1A1 has been duplicated and neofunctionalized in toad-eating *Leptodactylus* frogs
- Frequent non-allelic gene conversion (NAGC) homogenizes paralogs within species
- Selection counteracts NAGC to maintain 12 amino acid differences between paralogs
- Two substitutions confer toxin resistance and 10 mitigate their detrimental effects

Authors

Shabnam Mohammadi, Lu Yang, Arbel Harpak, ..., Susanne Dobler, Andrew J. Crawford, Peter Andolfatto

Correspondence

andrew@dna.ac (A.J.C.), pa2543@columbia.edu (P.A.)

In brief

In the frog genus *Leptodactylus*, a duplication of ATP1A1 has evolved toxin resistance. Using evolutionary and functional analyses, Mohammadi, Yang, Harpak et al. exploit a conflict between gene conversion and selection to identify amino acid substitutions underlying toxin resistance and maintaining the functional integrity of the resistant paralog.

Mohammadi et al., 2021, Current Biology *31*, 2530–2538
June 21, 2021
https://doi.org/10.1016/j.cub.2021.03.089

CellPress

CellPress

Current Biology

Article

Concerted evolution reveals co-adapted amino acid substitutions in Na^+K^+-ATPase of frogs that prey on toxic toads

Shabnam Mohammadi,[1,7,11] Lu Yang,[2,11,12] Arbel Harpak,[3,8,11,13] Santiago Herrera-Álvarez,[4,14] María del Pilar Rodríguez-Ordoñez,[4,15] Julie Peng,[5] Karen Zhang,[2] Jay F. Storz,[1] Susanne Dobler,[6] Andrew J. Crawford,[4,9,*] and Peter Andolfatto[3,10,16,*]

[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
[7]Twitter: @FrontlineEvo
[8]Twitter: @arbelharpak
[9]Twitter: @CrawfordAJ
[10]Twitter: @pandolfatto
[11]These authors contributed equally
[12]Present address: Wellcome Sanger Institute, Cambridge, UK
[13]Present address: Department of Population Health and Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA
[14]Present address: Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA
[15]Present address: Université Paris-Saclay Evry, Evry, France
[16]Lead contact
*Correspondence: andrew@dna.ac (A.J.C.), pa2543@columbia.edu (P.A.)
https://doi.org/10.1016/j.cub.2021.03.089

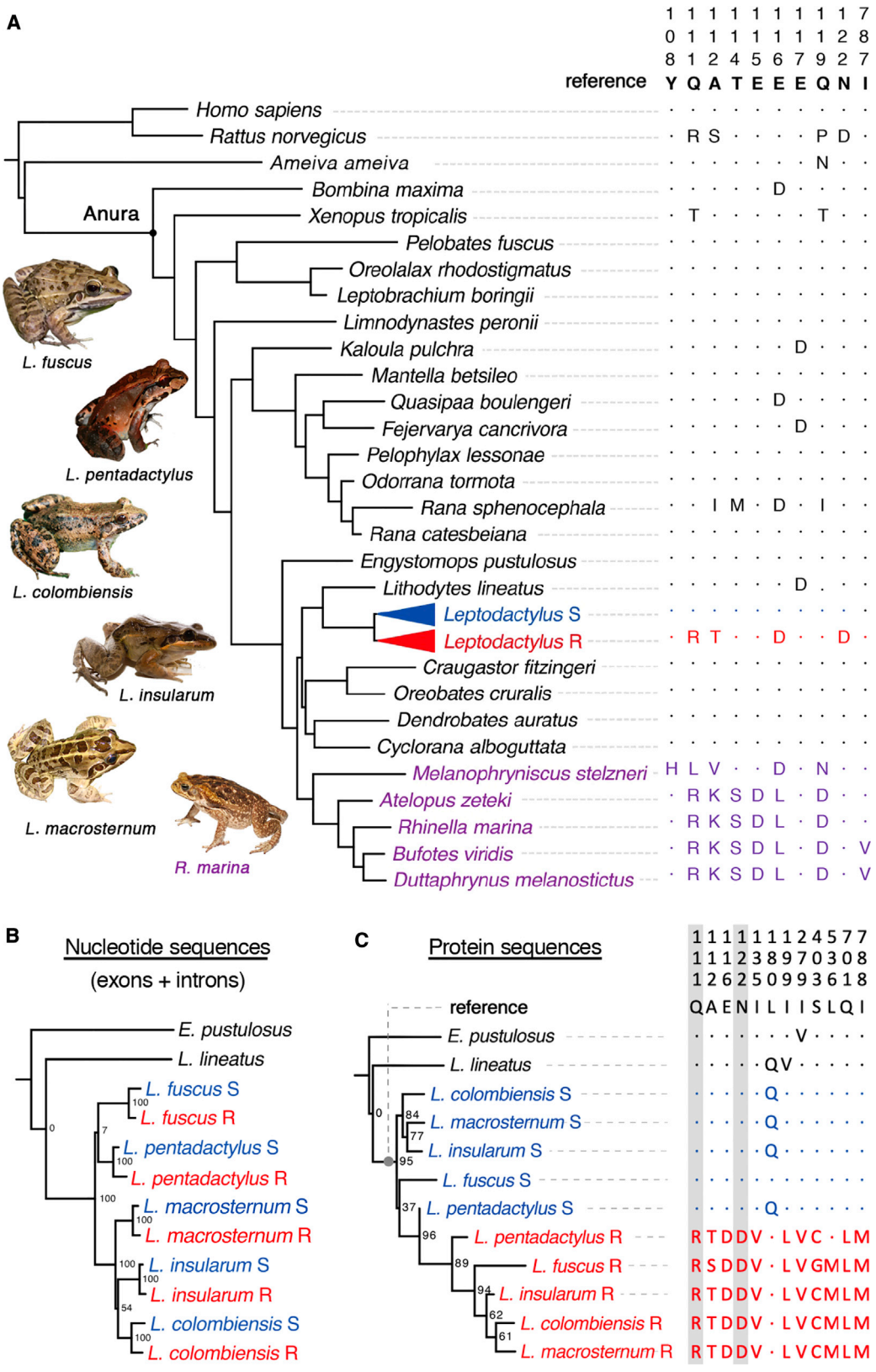
SUMMARY

Although gene duplication is an important source of evolutionary innovation, the functional divergence of duplicates can be opposed by ongoing gene conversion between them. Here, we report on the evolution of 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 although the other retained ancestral susceptibility. Within species, frequent non-allelic gene conversion homogenized most of the sequence between the two copies but was counteracted by strong selection on 12 amino acid substitutions that distinguish the two paralogs. Protein-engineering experiments show that two of these substitutions substantially increase toxin resistance, whereas the additional 10 mitigate their deleterious effects on ATPase activity. Our results reveal how examination of neo-functionalized gene duplicate evolution can help pinpoint key functional substitutions and interactions with the genetic backgrounds on which they arise.

INTRODUCTION

Along with other examples of parallel or convergent molecular evolution (e.g., color vision, pigmentation, and cold acclimatization),[1] the repeated emergence of toxin resistance in animals provides one of the clearest examples of natural selection at the genetic level and represents a useful paradigm to examine constraints on the evolution of novel protein functions.[2] Neotropical grass frogs of the genus *Leptodactylus* (Leptodactylidae) are widely distributed throughout lowland South America and are known to feed on chemically defended toads—a predatory tendency that is rare among frogs.[3–7] A major component of the chemical defense secretions of toads is a class of cardiotonic steroids (CTSs) called "bufadienolides"[8] that inhibit the α subunit of Na^+,K^+-ATPases (ATP1A). Na^+,K^+-ATPases are transmembrane proteins that are vital to numerous physiological processes in animals, including neural signal transduction, muscle contraction, and cell homeostasis.[9,10] CTSs bind to the extracellular surface of ATP1A and block the flux of ions,[11] making them potent poisons to most animals. However, some vertebrates have independently evolved the ability to prey on chemically defended toads, partly via amino acid substitutions to the CTS-binding domain of ATP1A1 that confer resistance to CTSs.[12–15]

Most vertebrates share several paralogous copies of ATP1A that have different tissue-specific expression profiles.[16] For example, ATP1A1 is the most ubiquitously expressed paralog and ATP1A3 has enriched expression in nervous tissue and heart muscle (Figure S1).[17,18] Previous studies on the molecular convergence of CTS resistance in reptiles have focused primarily



(legend on next page)

on the αM1–2 extracellular loop of ATP1A3,[13–15,19] whereas studies of birds, mammals, and amphibians have focused on the same region of ATP1A1.[12,19] A survey of ATP1A1 αM1–2 in toads and frogs[12] revealed a possible duplication of this gene in the toad-eating frog, *Leptodactylus latrans* (reported as *L. ocellatus*), where the resistant (R) paralog includes substitutions known to confer resistance to CTSs although the sensitive (S) paralog appears to have retained the ancestral susceptibility to CTSs. Neofunctionalization of ATP1A paralogs has contributed to the evolution of CTS resistance in numerous insect lineages[20–23] but appears to be rare among CTS-resistant vertebrates. Further, the fate of duplicated genes and the probability that they will neofunctionalize is predicted to depend on the strength of selection for functional differentiation relative to the rate of non-allelic gene conversion (NAGC), a form of nonreciprocal genetic exchange that homogenizes sequence variation between duplicated genes, thereby impeding divergence.[24–26] The ATP1A1 duplication in *Leptodactylus* provides an ideal opportunity to explore the results of the competition between evolutionary forces because the functional differentiation between R and S paralogs has clear adaptive significance with regard to CTS resistance.

RESULTS AND DISCUSSION

We surveyed the full-length coding sequences of all ATP1A paralogs in *Leptodactylus* and other anurans using RNA sequencing (RNA-seq)-based gene discovery (Table S1).[20] Our results confirm that ATP1A1 is duplicated in *Leptodactylus*, and the αM1–2 transmembrane domains of the ATP1A1 paralogs are distinguished by four amino acid substitutions (Figures 1C and 2).[12] Two of these substitutions, Q111R and N122D, were first identified in rat ATP1A1 and have been shown to interact synergistically to confer CTS resistance to sheep ATP1A1 protein *in vitro*.[27,28] Comparison of ATP1A1 sequences among five distantly related *Leptodactylus* species reveals that they each harbor a putatively resistant paralog (R) that includes the Q111R and N122D substitutions and a putatively sensitive ATP1A1 paralog (S) that lacks these substitutions. In addition to Q111R and N122D, there are 10 other amino acid substitutions (including two in the αM1–2 transmembrane domain) distinguishing the R and S paralogs in most of the five sampled species (Figure 1C). Hereafter, we refer to these twelve substitutions as "R/S-distinguishing substitutions." Because our sampling includes taxa from all four major species groups within *Leptodactylus*,[29] we infer that the duplication of ATP1A1 most likely occurred in the common ancestor of the genus (Figure 1A; Table S2). In contrast to the pattern for ATP1A1, two ancient paralogs common to vertebrates, ATP1A2 and ATP1A3, appear to be present as single-copy genes and lack any known CTS-resistant substitutions in *Leptodactylus* species (Figure S2).

To infer when the ATP1A1 duplication occurred relative to speciation events, we estimated phylogenies from a multiple alignment of gene sequences. Phylogenies estimated from nucleotide and inferred amino-acid sequences support dramatically different topologies (Figures 1B and 1C). Genealogies based on full gene sequences (Figure 1B) and intronic sites alone (Figure S4B) both suggest independent duplications in each of the *Leptodactylus* species, followed by parallel substitutions at the same 12 R/S-distinguishing amino acid positions (Figure 1B). Instead, the more parsimonious explanation is that of a single ancestral duplication—as indicated by the genealogy based on amino acid sequences (Figure 1C; Table 1)—coupled with ongoing NAGC between the R and S paralogs of each species. Frequent NAGC produces a pattern of "concerted evolution" whereby tandemly linked paralogs from the same species are more similar to one another than they are to their orthologous counterparts in other species (Figures 3A and 3B).[31] By generating a *de novo* genome assembly of *L. fuscus* based on linked-read sequencing technology (10X Genomics Chromium DNA sequencing), we established that S and R copies are indeed arranged in tandem and in the same orientation and are therefore likely to be subject to NAGC (Table S3; Figure S3). We thus propose that the unusual persistence of the 12 amino acid differences between the two paralogs is due to selection counteracting the homogenizing effects of NAGC (Figure 2B),[24,32] thereby maintaining an adaptive functional distinction between the R and S copies.

The opposing forces of NAGC and selection are predicted to leave a characteristic genealogical signature at neutral sites closely linked to the targets of selection (Figure 3B).[32] We tested the relationship between the genealogical signature and distance from nonsynonymous variants putatively under selection. To this end, for all informative sites, we evaluated the level of support for an ancient duplication of ATP1A1 in the common ancestor of all *Leptodactylus* species (with no concerted evolution) relative to support for an alternative in which ATP1A1 paralogs within species are always more closely related to one another than they are to paralogs in other species (as expected under concerted evolution). This analysis reveals that synonymous (presumed to be neutral) variants congruent with an ancient duplication of R and S have a median distance of 4 bp from nonsynonymous variants exhibiting the same pattern (Figure 3C). In contrast, equal numbers of randomly sampled synonymous sites supporting the alternative genealogy (i.e., concerted evolution) have a median distance of

Figure 1. Molecular evolution of ATP1A1 in anurans

(A) Maximum likelihood phylogeny of anuran species with mammalian and lizard outgroups derived from Feng et al.[30] Species names in purple correspond to chemically defended toads, and blue and red colors correspond to the S and R ATP1A1 paralogs in *Leptodactylus* species, respectively. Only variable sites with documented roles in CTS binding or sensitivity are shown (reviewed in Yang et al.[23]). The numbering of sites is based on sheep ATP1A1 (*Ovis aries*; GenBank: NC019458.2). Dots indicate identity with the reference sequence, and letters represent amino acid substitutions relative to the reference. The images on the left depict the five surveyed *Leptodactylus* species and a representative toad species (*Rhinella marina*) as potential prey.

(B and C) Maximum likelihood phylogeny estimates based on nucleotide sequences (B) and amino acid sequences (C) yield distinct topologies. Bootstrap support values are indicated at internal nodes. To the right is the pattern of amino acid variation at 12 positions that distinguish the S and R paralogs. The gray point indicates the inferred ancestral *Leptodactylus* lineage corresponding to the reference states. Amino acid positions 111–122 correspond to the αM1–2 transmembrane domain of ATP1A1. Two sites (111 and 122), previously implicated in CTS resistance, are shaded in gray.

See also Figures S1, S2, and S4 and Tables S1–S3 and S6.

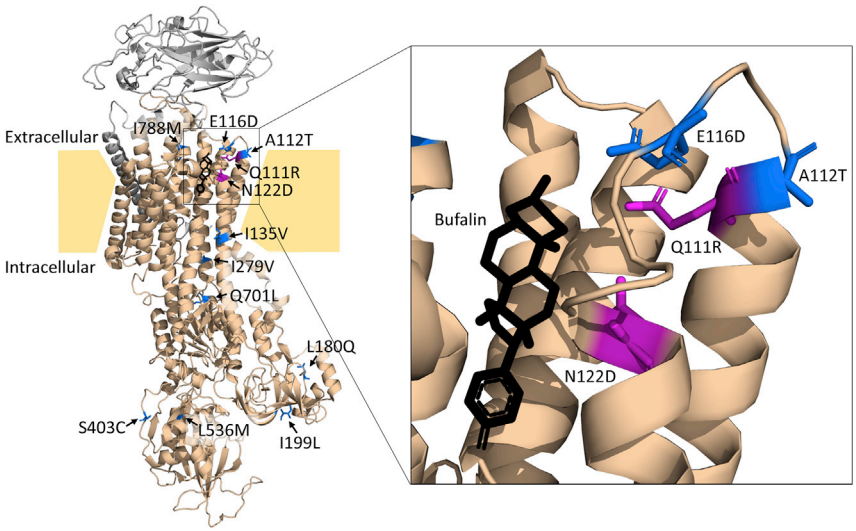


Figure 2. Positions of 12 R copy-specific amino acid substitutions on the crystal structure of pig Na^+K^+-ATPase (*Sus scrofa*; PDB: 4RES) bound to the cardiotonic steroid bufalin

Shown are the ATP1A1 (gold) and ATP1B1 (gray) subunits. The panel details the cardiotonic steroid-binding pocket of ATP1A1. Highlighted residues correspond to the 12 R/S-distinguishing amino acid substitutions in *Leptodactylus*. The two magenta residues correspond to key CTS resistance-conferring sites 111 and 122; blue residues correspond to 10 additional residues distinguishing the R and S proteins (Figure 1C). The span of the plasma membrane (in yellow) was estimated from Laursen et al.[11] See also Figure S2.

88 bp from those nonsynonymous variants (bootstrap $p < 10^{-5}$). This pattern at synonymous sites is consistent with a scenario in which purifying selection maintains functionally important sequence differences between neofunctionalized gene duplicates in the face of NAGC.

We next quantified the strength of purifying selection required to maintain the amino acid differentiation between R and S duplicates in the face of NAGC. We first considered population genetics theory for the evolution of a single site in tandem duplicates (STAR Methods).[34] This analytic model predicts that, if the rate of NAGC is an order of magnitude higher than the rate of point mutation, then the maintenance of alternative amino acid states is only likely under sufficiently strong purifying selection—namely, when the selection coefficient scaled by population size, $2Ns$, is larger than one (Figure 4A). We next developed an inference method based on simulations of ATP1A1 evolution to estimate the combination of parameters that best explains divergence patterns throughout the gene, including levels of paralog divergence observed as a function of distance from the 12 R/S-distinguishing substitutions (STAR Methods). We estimate the rate of NAGC to be an order of magnitude higher than the point mutation rate (posterior mode 9 with an 80% credible interval of 4- to 54-fold higher than the point mutation rate) and $2Ns$ substantially larger than one (posterior mode 9; 80% credible interval 5–18; Figure 4B). These estimates fall within the plausible range predicted by the theoretical single-site model (Figure 4A). These results indicate that the observed pattern of divergence between R and S paralogs reflects a history of strong purifying selection that maintains fixed differences between them despite high rates of NAGC.

The inference that selection maintains the co-occurrence of the 12 R/S-distinguishing substitutions implies they are functionally important and collectively contribute to organismal fitness. The effects of Q111R and N122D on CTS insensitivity have previously been demonstrated by *in vitro* enzyme inhibition assays.[10] Additionally, although not related directly to CTS resistance, the potential importance of substitutions at sites 112 and 116 has been suggested by molecular evolution analysis and structural studies, respectively.[12,35] However, the remaining eight R/S-distinguishing substitutions are located in structural domains that have not been implicated in CTS resistance. Because our analysis suggests that amino acid divergence between R and S paralogs is maintained by selection, we performed protein-engineering experiments to elucidate the functional significance of the 12 R/S-distinguishing substitutions. We synthesized and recombinantly expressed eight mutant Na^+,K^+-ATPase proteins, each harboring different combinations of R-specific replacements on both S- and R-type genetic backgrounds of a representative species, *L. macrosternum* (Figure 5A; Table S4). We then quantified the level of CTS resistance of each genotype using enzyme-inhibition assays (Table S4; Figure S6).[36] Individually, Q111R and N122D significantly increased CTS resistance by 21-fold and 14-fold, respectively (ANOVA $p = 2.7e{-}13$ and $p = 2.3e{-}6$; Figure 5B; Table S5). When combined, Q111R and N122D produce a greater than 100-fold increase in CTS resistance relative to

Table 1. Sitewise support for "non-concerted" and "concerted" topologies

Category	Informative sites	Non-concerted topology (NC)	Concerted topology (C)	Ratio (NC/C)	Fisher's exact test p value versus nonsynonymous
Nonsynonymous	32	15	9	1.67	–
Synonymous	207	12	112	0.11	8e−8
Intronic	421	14	337	0.04	3e−13

"Informative sites" refers to the number of sites analyzed, excluding those with singleton substitutions and sites containing gaps in the multi-alignment. The next two columns sum the number of sites for which there was >2 log-likelihood support for either the "non-concerted" topology or the "concerted" topology, respectively (Figure 3B). Synonymous and intronic sites were also significantly different (Fisher's exact test, $p = 0.02$).

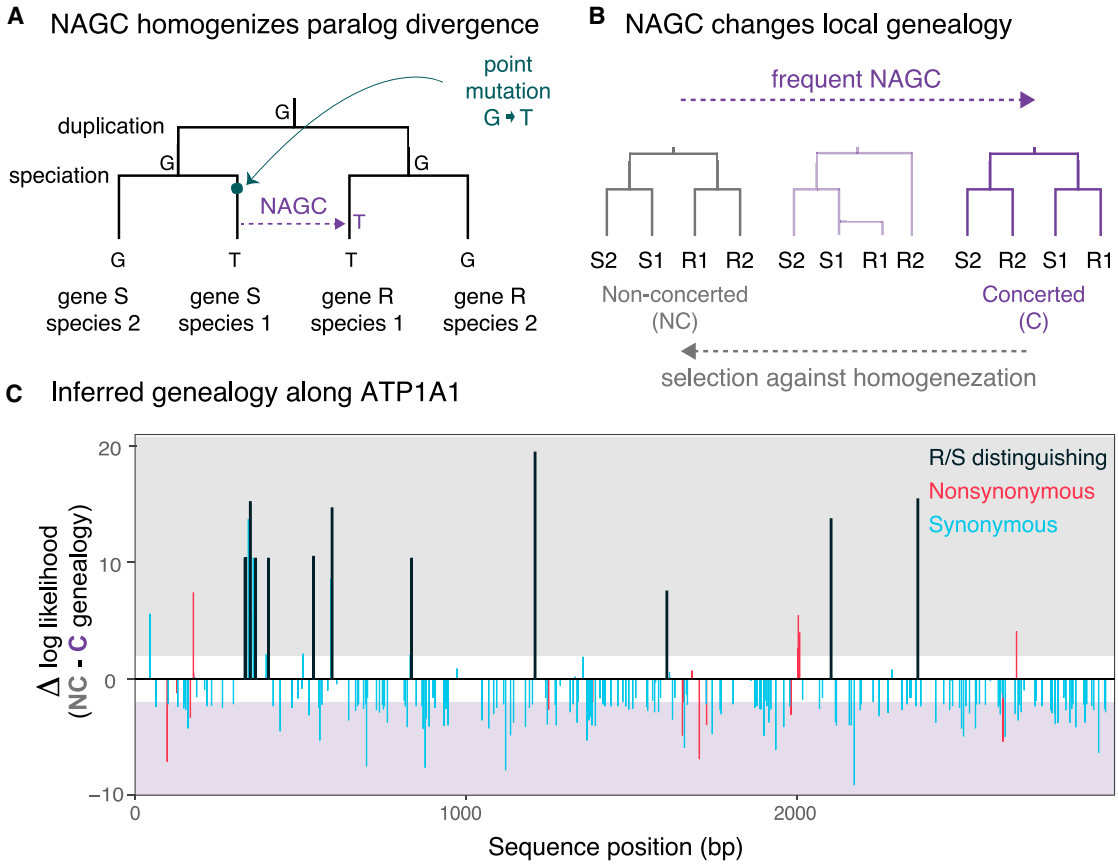


Figure 3. Non-allelic gene conversion (NAGC) and selection maintaining paralog specialization are opposing forces leading to the observed genealogical patterns

(A) NAGC homogenizes sequence variation between paralogous genes and therefore changes the genealogical signal (adapted from Harpak et al.[33]).

(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) Sitewise difference in the log-likelihood of two alternative tree topologies—generalizing the topological extremes of (B) to all five *Leptodactylus* species. Shaded regions indicate a log-likelihood difference greater than 2 in support of the corresponding model (gray, "NC"; purple, "C"). Only topology-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 Figure 1C).

See also Figure S3.

the S paralog (Tukey's HSD test; adjusted $p < 4e{-}5$; Figure 5B; Tables S5 and S6). In contrast, the remaining 10 substitutions had no detectable net effect on CTS resistance when jointly added to the S background ($p = 0.22$; Figure 5B).

Given the absence of detectable effects of R/S-distinguishing substitutions other than Q111R and N122D on CTS resistance, we tested whether these substitutions had effects on other aspects of ATP1A1 function. Because ATP hydrolysis and ion cotransport are strongly coupled functions of Na^+,K^+-ATPase,[37] we used estimates of the rate of ATP hydrolysis in the absence of ouabain as a proxy for overall protein activity. Based on this assay, we found that CTS resistance substitutions Q111R and N122D significantly impair activity, individually reducing ATPase activity by an average of 40% ($p = 0.024$ and $p = 7.7e{-}4$, respectively; Figure 5B; Table S5). We also detected a significant interaction between Q111R and N122D that renders their joint effects somewhat less severe than predicted by the sum of their individual effects (i.e., a 30% reduction rather than the expected 78% reduction; $p = 0.022$). Critically, adding the remaining 10 R-specific substitutions on the S background containing Q111R and N122D restores ATPase activity close to S levels—a significant effect even when controlling for the effects of Q111R and N122D (ANOVA $p = 1e{-}4$; Figure 5B; Table S5). Our results thus indicate that these 10 R/S-distinguishing substitutions play a vital role in compensating for the negative pleiotropic effects of the resistance-conferring substitutions, Q111R and N122D. We conclude that the evolution of the R protein from a CTS-sensitive ancestral state involved two epistatically interacting substitutions (Q111R and N122D) in conjunction with compensatory effects of 10 additional substitutions that mitigate the trade-off between toxin resistance and native enzyme activity.

Given that both paralogs maintain their ATPase function, it is interesting to speculate as to why the sensitive copy of

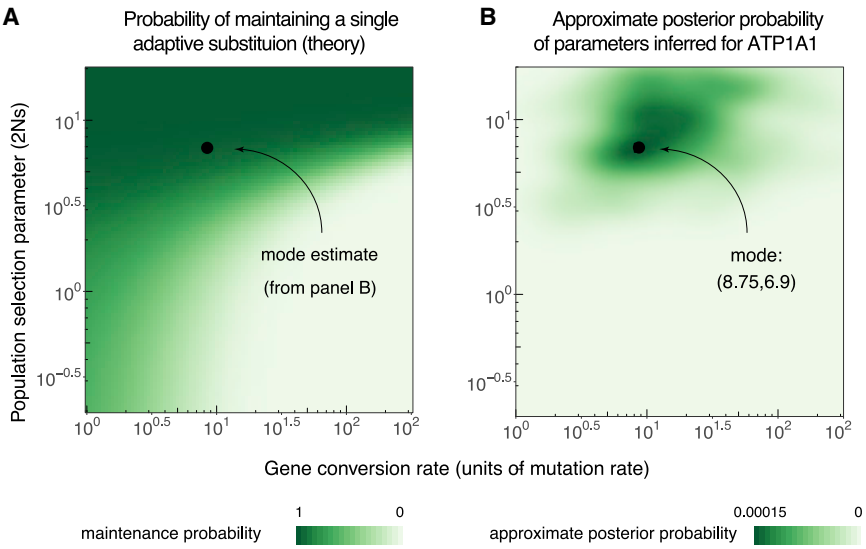


Figure 4. Modeling the competition between selection and NAGC and inference of evolutionary parameters

(A) Theoretical probability of maintaining distinct alleles at a single site in the face of 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 (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.

ATP1A1 is maintained at all in *Leptodactylus* species. This question is related to that of why the CTS-binding site itself is highly conserved across diverse animal taxa.[10] In addition to its ion-transport function, Na^+,K^+-ATPase also plays important and distinct roles in signaling pathways, linked to a variety of physiological processes, that are mediated by binding of endogenous CTSs.[10] Given that the R protein can no longer be regulated by CTSs, the S protein may be vital to maintaining these signaling pathways. Additionally, recent *in vivo* work has revealed that amino acid substitutions that may have a negligible effect on Na^+,K^+-ATPases at the level of ATPase activity can cascade to detrimental physiological effects at the whole-organism level.[38] We thus hypothesize that pleiotropy associated with the specialization of the R and S proteins extends beyond ATPase activity to physiological processes at the organismal level that cannot be straightforwardly probed with *in vitro* experiments.

The adaptive functional distinction between the R and S paralogs of ATP1A1 in *Leptodactylus* has been maintained by strong selection that has counteracted the homogenizing effects of frequent NAGC over the 35-Ma history of this genus. Similar signatures of selection to maintain sequence differentiation between neofunctionalized duplicates have been observed for the RHCE/RHD antigen proteins of humans,[39] "major facilitator family" transporter proteins in *Drosophila*,[40] and red/green opsins of primates.[32] To our knowledge, only in the case of opsins have differences between paralogs been linked directly to functional differentiation, notably two closely linked amino acid substitutions contributing to a red to green shift in absorbance maxima.[41] Our study highlights similar signatures of selection not only on the two amino acid substitutions directly linked to adaptive differentiation for CTS resistance but also at 10 more amino acid substitutions scattered throughout the protein that facilitate this neofunctionalization. Thus, by identifying interactions between adaptive substitutions and the genetic backgrounds that permit these changes, our combination of evolutionary and functional analyses reveals how mechanisms of adaptation are shaped by intramolecular epistasis and pleiotropy.

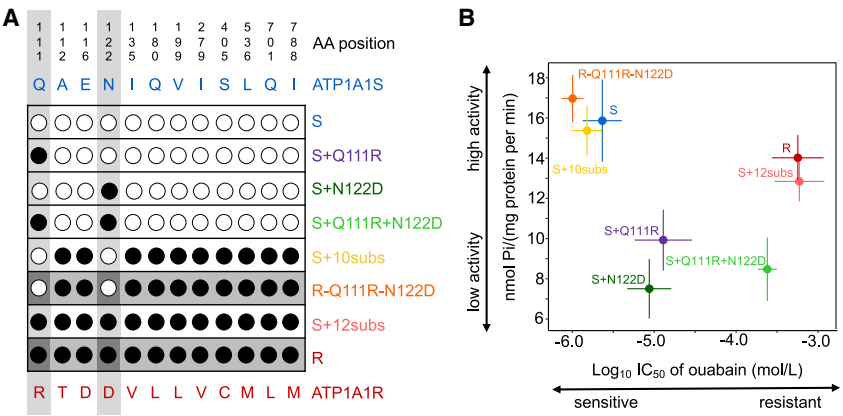


Figure 5. Functional analysis of substitutions specific to the R-type ATP1A1 paralog

(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 gray shading denotes the R background, and white denotes the S background. Light gray columns highlight two substitutions (Q111R and N122D) that are known to confer CTS resistance.

(B) Functional properties of engineered Na^+,K^+-ATPases. A measure of CTS resistance (i.e., mean $\log_{10}IC_{50}$ ± SEM) is plotted on the x axis, and a measure of protein activity (i.e., mean ATP hydrolysis rate ± SEM) for the same proteins is plotted on the y axis. Each estimate is based on six biological replicates.

See also Figures S5 and S6 and Tables S4–S6.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cultivation of *Escherichia coli* for production of expression vectors
  - Cultivation of Sf9 cells for expression of recombinant proteins
- METHOD DETAILS
  - Sample collection and data sources
  - RNA-seq based gene discovery of ATP1A paralogs
  - Targeted sequencing of protein-coding regions of ATP1A1 paralogs
  - *De novo* genome assembly of *Leptodactylus fuscus*
  - Targeted long-read sequencing of intronic sequences of ATP1A1
  - Estimation of genealogical relationships
  - Maximum likelihood analysis of site-wise support for alternative tree topologies
  - Theoretical single-site model for the probability of maintaining an adapted substitution
  - Simulations of ATP1A1 gene family evolution
  - Inference of evolutionary parameters using Approximate Bayesian Computation
  - Measuring similarity to observed divergence patterns
  - Analysis
  - Construction of expression vectors
  - Generation of recombinant viruses and transfection into Sf9 cells
  - Preparation of Sf9 cell membranes
  - Verification by SDS-PAGE and western blotting
  - Ouabain inhibition assay (measurement of CS resistance)
  - ATP hydrolysis assay (measurement of ATPase activity as a proxy for protein activity)
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analyses of biochemical assay results

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.03.089.

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. Thanks to M. Lyra for help in resolving issues of *Leptodactylus* taxonomy. This study was funded by grants to P.A. from the National Institutes of Health (R01-GM115523) and to J.F.S. from the National Institutes of Health (R01-HL087216) and the National Science Foundation (OIA-1736249), to S.D. from Deutsche Forschungsgemeinschaft (DFG) (grant DO527/10-1), and a fellowship to A.H. from The Simons Foundation's Society of Fellows (no. 633313).

AUTHOR CONTRIBUTIONS

P.A. and A.J.C. conceived of and oversaw the project; L.Y., M.d.P.R.-O., S.H.-Á., J.P., and A.J.C. collected samples and generated sequence data; L.Y., A.H., P.A., S.H.-Á., and K.Z. performed evolutionary and population genetics analyses; S.M., J.F.S., S.D., A.J.C., and P.A. designed functional experiments; S.M. and P.A. performed experiments and associated statistical analyses; S.M., J.F.S., L.Y., A.H., and P.A. wrote the paper; and all authors edited the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 5, 2021
Revised: March 12, 2021
Accepted: March 26, 2021
Published: April 21, 2021

REFERENCES

1. Carroll, S.B. (2006). Making of the Fittest: DNA and the Ultimate Forensic Record of Evolution (W. W. Norton & Company).
2. Brodie, E.D., 3rd. (2009). Toxins and venoms. Curr. Biol. *19*, R931–R935.
3. Chen, K.K., and Chen, A.L. (1933). Notes on the poisonous secretions of twelve species of toads. J. Pharmacol. Exp. Ther. *47*, 281–293.
4. Heyer, W.R., McDiarmid, R.W., and Weigmann, D.L. (1975). Tadpoles, predation and pond habitats in the tropics. Biotropica *7*, 100–111.
5. Crossland, M.R., and Azevedo-Ramos, C. (1999). Effects of *Bufo* (Anura: Bufonidae) toxins on tadpoles from native and exotic *Bufo* habitats. Herpetologica *55*, 192–199.
6. Azevedo-Ramos, C., and Magnusson, W.E. (1999). Tropical tadpole vulnerability to predation: association between laboratory results and prey distribution in an Amazonian savanna. Copeia *1999*, 58–67.
7. Guimaraes, L.D., Pinto, R.M., and Juliano, R.D.F. (2004). Bufo granulosus (NCN). Predation. Herpetol. Rev. *35*, 259.
8. Krenn, L., and Kopp, B. (1998). Bufadienolides from animal and plant sources. Phytochemistry *48*, 1–29.
9. Horisberger, J.-D. (2004). Recent insights into the structure and mechanism of the sodium pump. Physiology (Bethesda) *19*, 377–387.
10. Lingrel, J.B. (2010). The physiological significance of the cardiotonic steroid/ouabain-binding site of the Na,K-ATPase. Annu. Rev. Physiol. *72*, 395–412.
11. Laursen, M., Gregersen, J.L., Yatime, L., Nissen, P., and Fedosova, N.U. (2015). Structures and characterization of digoxin- and bufalin-bound Na+,K+-ATPase compared with the ouabain-bound complex. Proc. Natl. Acad. Sci. USA *112*, 1755–1760.
12. Moore, D.J., Halliday, D.C., Rowell, D.M., Robinson, A.J., and Keogh, J.S. (2009). Positive Darwinian selection results in resistance to cardioactive toxins in true toads (Anura: Bufonidae). Biol. Lett. *5*, 513–516.
13. Ujvari, B., Mun, H.C., Conigrave, A.D., Bray, A., Osterkamp, J., Halling, P., and Madsen, T. (2013). Isolation breeds naivety: island living robs Australian varanid lizards of toad-toxin immunity via four-base-pair mutation. Evolution *67*, 289–294.
14. Ujvari, B., Casewell, N.R., Sunagar, K., Arbuckle, K., Wüster, W., Lo, N., O'Meally, D., Beckmann, C., King, G.F., Deplazes, E., and Madsen, T. (2015). Widespread convergence in toxin resistance by predictable molecular evolution. Proc. Natl. Acad. Sci. USA *112*, 11911–11916.
15. Mohammadi, S., Gompert, Z., Gonzalez, J., Takeuchi, H., Mori, A., and Savitzky, A.H. (2016). Toxin-resistant isoforms of Na+/K+-ATPase in snakes do not closely track dietary specialization on toads. Proc. Biol. Sci. *283*, 20162111.
16. Orlowski, J., and Lingrel, J.B. (1988). Tissue-specific and developmental regulation of rat Na,K-ATPase catalytic alpha isoform and beta subunit mRNAs. J. Biol. Chem. *263*, 10436–10442.

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

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

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

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

21. Petschenka, G., Wagschal, V., von Tschirnhaus, M., Donath, A., and Dobler, S. (2017). Convergently evolved toxic secondary metabolites in plants drive the parallel molecular evolution of insect resistance. Am. Nat. *190* (S1), S29–S43.

22. Lohr, J.N., Meinzer, F., Dalla, S., Romey-Glüsing, R., and Dobler, S. (2017). The function and evolutionary significance of a triplicated Na,K-ATPase gene in a toxin-specialized insect. BMC Evol. Biol. *17*, 256.

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

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

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

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

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

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

29. de Sá, R.O., Grant, T., Camargo, A., Heyer, W.R., Ponssa, M.L., and Stanley, E. (2014). Systematics of the neotropical genus *Leptodactylus* Fitzinger, 1826 (Anura: Leptodactylidae): phylogeny, the relevance of non-molecular evidence, and species accounts. South Am. J. Herpetol. *9*, S1–S128.

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

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

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

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

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

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

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

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

38. Taverner, A.M., Yang, L., Barile, Z.J., Lin, B., Peng, J., Pinharanda, A.P., Rao, A.S., Roland, B.P., Talsma, A.D., Wei, D., et al. (2019). Adaptive substitutions underlying cardiac glycoside insensitivity in insects exhibit epistasis in vivo. eLife *8*, e48224.

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

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

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

42. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. *8*, 1494–1512.

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

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

45. Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics *27*, 764–770.

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

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

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

49. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

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

51. Li, H. (2012). seqtk Toolkit for processing sequences in FASTA/Q formats. https://github.com/lh3/seqtk.

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

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

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

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

56. Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. *33*, 1870–1874.

57. He, Z., Zhang, H., Gao, S., Lercher, M.J., Chen, W.-H., and Hu, S. (2016). Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. Nucleic Acids Res. *44* (W1), W236–W241.

58. Stanke, M., Tzvetkova, A., and Morgenstern, B. (2006). AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. Genome Biol. *7* (*Suppl 1*), 11.1–11.8.

59. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. *37*, 1530–1534.

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

61. Elzhov, T.V., Mullen, K.M., Spiess, A.-N., Bolker, B., Mullen, M.K., and Package, M. (2015). minpack. lm (CRAN Repository).

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

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

64. Kimura, M. (1962). On the probability of fixation of mutant genes in a population. Genetics *47*, 713–719.

65. Gillespie, J.H. (2004). Population Genetics: A Concise Guide (JHU).

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

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

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

69. Venables, W.N., and Ripley, B.D. (2002). Modern Applied Statistics with S (Springer).

70. Luckow, V.A., Lee, S.C., Barry, G.F., and Olins, P.O. (1993). 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.

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

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

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Chicken monoclonal antibody α5	Developmental Studies Hybridoma Bank, University of Iowa, Iowa City, IA, USA	RRID: AB_2166869
Goat-anti-mouse polyclonal secondary antibody conjugated with horseradish peroxidase	Dianova, Hamburg, Germany	Cat#115-035-003; RRID: AB_2617176
Bacterial and virus strains		
Escherichia coli MAX Efficiency DH10Bac Competent Cells	Thermo Fisher Scientific	Cat#10361012
Escherichia coli DH5α Competent Cells	Thermo Fisher Scientific	Cat#18265017
Escherichia coli XL 10-Gold Competent Cells	Agilent Technologies, La Jolla, CA, USA	Cat#200314
Biological samples		
Frog tissue samples, see Tables S1 and S2	This paper	See Tables S1 and S2
Chemicals, peptides, and recombinant proteins		
Cellfectin II reagent	(GIBCO) Thermo Fisher Scientific	Cat#10362100
Gentamycin	Roth, Karlsruhe, Germany	Cat#0233.1
Insect-Xpress medium	Lonza, Walkersville, MD, USA	Cat#BE12-730P10
RNA*later* Stabilization Solution	Thermo Fisher Scientific	Cat#AM7021
One*Taq* DNA Polymerase	NEB	Cat#M0480L
FastDigest XhoI	Thermo Fisher Scientific	Cat#FD0694
FastDigest NotI	Thermo Fisher Scientific	Cat#FD0593
FastDigest SpeI (also known as BcuI)	Thermo Fisher Scientific	Cat#FD1253
FastDigest KpnI	Thermo Fisher Scientific	Cat#FD0524
4-chloro-1 naphtol	(Merck) Sigma-Aldrich	Cat#C8890
Ouabain octahydrate 96%	Acros Organics	Cat#AC161730010
Adenosin-5-triphosphat Bis-(Tris)-salt hydrate (ATP)	(Merck) Sigma-Aldrich	CAS#102047-34-7
Adenosine 5′-Triphosphatase from porcine cerebral cortex	(Merck) Sigma-Aldrich	CAS 9000-83-3
Critical commercial assays		
Superscript III Reverse Transcriptase kit	Thermo Fisher Scientific	Cat#18080093
QuikChange II XL Site-Directed Mutagenesis Kit	Agilent Technologies, La Jolla, CA, USA	Cat#200521
Phusion Green High-Fidelity DNA Polymerase (2 U / μL)	Thermo Fisher Scientific	Cat#F534S
TRIzol Reagent	Thermo Fisher Scientific	Cat#15596026
TruSeq RNA Library Prep Kit v2	Ilumina	Cat#RS-122-2001
QIAquick PCR Purification Kit	QIAGEN	Cat#28104
TOPO TA Cloning Kit	Thermo Fisher Scientific	Cat#451641
Agencourt DNAdvance Kit	Beckman Coulter, France	Cat#A48705
LongAmp *Taq* PCR Kit	NEB	Cat#E5200S
Ligation Sequencing Kit	Oxford Nanopore Technology	SQK-LSK109

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw data for recombinant Na^+, K+-ATPase functional assays	This paper	Dryad DOI: https://doi.org/10.5061/dryad.qfttdz0f7
ATP1A1 alignment used to generate phylogenetic tree	This paper	Dryad DOI: https://doi.org/10.5061/dryad.qfttdz0f7
Sequences generated by this study are deposited at GenBank, see Table S2	This paper	GenBank, see Table S2
The de novo assembly of the genome of Leptodactylus fuscus	This paper	GitHub: https://github.com/AndolfattoLab/Leptodactylus-fuscus-genome
Experimental models: Cell lines		
Insect: Sf9 cells in Sf-900 II SFM	Thermo Fisher	Cat#11496015
Oligonucleotides		
All primers used in this study are listed in Table S6	This paper	N/A
Recombinant DNA		
Plasmid R-Q111R-N122D	This paper	Addgene Plasmid #167178
Plasmid S+12subs	This paper	Addgene Plasmid #167177
Plasmid S+10subs	This paper	Addgene Plasmid #167176
Plasmid S+Q111R+N122D	This paper	Addgene Plasmid #167175
Plasmid S+N122D	This paper	Addgene Plasmid #167174
Plasmid S+Q111R	This paper	Addgene Plasmid #167173
Plasmid S	This paper	Addgene Plasmid #167172
Plasmid R	This paper	Addgene Plasmid #167170
Software and algorithms		
Trinity v2.2.0	Haas et al.[42]	http://trinityrnaseq.sourceforge.net/
Velvet v1.2.10	Zerbino and Birney[43]	https://kbase.us/applist/apps/Velvet/run_velvet/release
Oases v0.2.8	Schulz et al.[44]	https://www.ebi.ac.uk/~zerbino/oases/
Long Ranger basic v2.2.2	10X Genomics	https://support.10xgenomics.com/genome-exome/software/downloads/latest
Jellyfish v2.2.7	Marçais and Kingsford[45]	https://github.com/gmarcais/Jellyfish/releases/tag/v2.2.7
GenomeScope	Vurture et al.[46]	http://qb.cshl.edu/genomescope/
Supernova v2.1	Weisenfeld et al.[47]	https://github.com/10XGenomics/supernova
BUSCOs v4.0.5	Seppey et al.[48]	https://busco.ezlab.org/
BLAST v2.2.26	Altschul et al.[49]	http://bioweb.pasteur.fr/packages/pack@blast@2.2.26
Albacore v2.3.4	Oxford Nanopore Technology	https://github.com/Albacore/albacore
LAST v980	Kiełbasa et al.[50]	http://last.cbrc.jp/
seqtk	Li[51]	https://github.com/lh3/seqtk
Canu v1.8	Koren et al.[52]	https://github.com/marbl/canu
minimap2	Li[53]	https://github.com/lh3/minimap2
racon v1.3.3	Vaser et al.[54]	https://github.com/isovic/racon
MUSCLE	Vaser et al.[54]	https://www.drive5.com/muscle/
SeaView	Crawford[55]	http://doua.prabi.fr/software/seaview
MEGA 7	Kumar et al.[56]	https://www.megasoftware.net/
EvolView	He et al.[57]	https://www.evolgenius.info:8443/evolview/
Augustus v3.2.2	Stanke et al.[58]	http://augustus.gobics.de/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
IQ-TREE 2 v.2.0.4	Minh et al.[59]	http://www.iqtree.org/
PAML 4.8	Yang[60]	http://abacus.gene.ucl.ac.uk/software/paml.html
R	The R Foundation	https://www.r-project.org/
minpack.lm package for R	Elzhov et al.[61]	https://cran.r-project.org/web/packages/minpack.lm/minpack.lm.pdf
PyMOL v2.4.0	Schrödinger, LLC	https://pymol.org/2/

RESOURCE AVAILABILITY

Lead contact

Further information and requests on methods can be directed to Dr. Peter Andolfatto pa2543@columbia.edu and Dr. Andrew J. Crawford andrew@dna.ac

Materials availability

Plasmids used in this study have been deposited to Addgene (see Key resources table for names and numbers). This study did not generate new unique reagents.

Data and code availability

Original data and alignments have been deposited to Dryad: https://doi.org/10.5061/dryad.qfttdz0f7. See also Key resources table.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cultivation of *Escherichia coli* for production of expression vectors

All *E. coli* strains used in this study (see Key resources table) for the production of expression vectors (see Method details) were grown and maintained in liquid media containing 5 g tryptone, 2.5 g yeast extract, 2.5 g NaCl, 0.5 mL 1M NaOH in 500 mL deionized H2O or agar plates containing the same media with the addition of 6 g agar. Bacteria grown in liquid media were incubated at 37°C and 225 rpm in a shaking incubator and those grown on plates were incubated at 37°C with no shaking.

Cultivation of Sf9 cells for expression of recombinant proteins

Sf9 cells used for the expression of recombinant proteins (see Method details) were maintained in T75 flasks (Sarstedt AG & Co., Nümbrecht, Germany) at 27°C in Insect-Xpress Medium (Lonza, Walkersville, MD, USA) with 15 mg/ml gentamycin. Cells were split every 3-4 days into new passages. Only cells between passage 5 and 30 were used for baculovirus infection and subsequent protein expression.

METHOD DETAILS

Sample collection and data sources

We sampled tissues from 16 anuran species. Five Leptodactylus species (*L. colombiensis*, *L. insularum*, *L. macrosternum*, *L. fuscus* and *L. pentadactylus*), two outgroup species (*Engystomops pustulosus* and *Lithodytes lineatus*) and one bufonid *Rhinella marina*, were collected from different geographic locations in Colombia (Table S1) and stored in RNAlater (Invitrogen) at −80°C until used. Field collections were made under permiso marco resolución No 1177 to the Universidad de los Andes from the Autoridad Nacional de Licencias Ambientales (ANLA), and animal use protocols were approved by the Institutional Committee on the Care and Use of Laboratory Animals (abbreviated CICUAL in Spanish) of the Universidad de los Andes. A tissue sample of the toad, *Atelopus zeteki*, was donated by the Smithsonian's National Zoo and came from a necropsied animal. The outgroup species, *Kaloula pulchra*, *Rana sphenocephala*, *Rana catesbeiana*, *Dendrobates auratus*, *Melanophryniscus stelzneri*, and *Duttaphrynus melanostictus* were obtained from the pet trade under IACUC Protocol No. 2057-16. Live animals were euthanized under the supervision of a research veterinarian at Princeton University. To capture all three paralogs of ATP1A, we collected tissue samples from brain, skeletal muscle, and stomach – each of which highly expresses at least one of the three paralogs.[16] To confirm identities of animals, we mined mitochondrial Cytochrome oxidase I (COI) sequences from RNA-seq *de novo* assemblies (described below) and performed BLAST[49] (blastn v2.26) searches against the GenBank nucleotide database. The species used in this study show 94%–100% identity to a corresponding record in NCBI, or 84%–90% identity with a sister species in the same genus where no mitochondrial DNA data were available.

RNA-seq based gene discovery of ATP1A paralogs

Full-length coding sequences of ATP1A1, ATP1A2 and ATP1A3 were reconstructed for several species using RNA-seq based gene discovery. Total RNA was extracted from multiple tissues of 16 anuran species (Table S2) using TRIzol Reagents (Ambion, Life

technologies) following the manufacturer's protocol. RNA-seq libraries were prepared with TruSeq RNA Library Prep Kit v2 (Illumina) and sequenced on Illumina HiSeq2500 (Genomics Core Facility, Princeton, NJ, USA) with either PE 75bp or SE 140bp (Table S2). Reads were trimmed and *de novo* assembled with Trinity v2.2.0.[42] ATP1A1 of *Xenopus laevis* (GenBank: NM_001090595) was initially used to BLAST against the assembled transcripts of *L. macrosternum* to recover ATP1A1S and ATP1A1R, which were later used as queries to reconstruct ATP1A1 genes from other species. ATP1A paralogs for the rest of the species used in this study were mined from publicly available data (Table S2) following the same pipeline.

Targeted sequencing of protein-coding regions of ATP1A1 paralogs

Total RNA was extracted from *L. fuscus*, *L. insularum*, and *L. colombiensis* as described above and reverse-transcribed to single-strand cDNA using SuperScript III Reverse Transcriptase (Invitrogen). ATP1A1 was amplified using Phusion High-Fidelity DNA polymerase (Invitrogen) using forward primer: 5′-ATAAGTATGAGCCCGCAGCC-3′ and reverse primer: 5′-CCAGGGCTGCGTCTGATT ATG-3′. PCR products were cleaned with QIAquick PCR Purification Kit (QIAGEN) and A-tailed with Taq Polymerase (NEB) before cloning into a pTOPO-TA vector (Invitrogen). The presence of the insert in the plasmid was confirmed by colony-PCR. Illumina-ready sequencing libraries of isolated plasmids were prepared with Tn5 transposase, charged with Illumina-ready indexed barcodes,[23] and sequenced on Illumina MiSeq (Genomics Core Facility, Princeton, NJ, USA). *De novo* assembly of the cloned PCR products was performed with Velvet v1.2.10[43] and Oases v0.2.8.[44] ATP1A1 paralogs were reconstructed by aligning with previously obtained ATP1A1 sequences of *L. macrosternum* and *L. pentadactylus*.

De novo genome assembly of *Leptodactylus fuscus*

High-molecular-weight genomic DNA was isolated from a single *Leptodactylus fuscus* individual (Table S1, JSM 205) and used to prepare a 10x Genomics Chromium library that was sequenced on Illumina HiSeq X sequencer (HudsonAlpha Institute of Biotechnology, Alabama, USA.). Barcodes were removed using the Long Ranger basic v2.2.2 (https://support.10xgenomics.com/genome-exome/software/downloads/latest). Trimmed reads were used for k-mer estimation in Jellyfish[45] (v2.2.7). The k-mer (k = 21) frequency distribution was processed in GenomeScope[46] to estimate the genome size, heterozygosity, and percentage of repeat content. The linked-reads were assembled using the Supernova v2.1.1 assembler[47] using default settings and the "–accept-extreme-coverage" flag. A summary of the assembly is provided in Table S3. The assembled genome is 2.42 Gb (16,530 scaffolds > = 10 kb, scaffold N50 = 363 kb, Table S3) and was outputted in the pseudohap2 format (*de novo* assembly; GitHub https://github.com/AndolfattoLab/Leptodactylus-fuscus-genome). The assembly size of contigs larger than 10 kb (1.26 Gb) is only ~1/2 of the estimated genome size (2.4 Gb). Effective depth coverage (48X) was in the middle of the recommended range (38-56X) which may have limited the success of the assembly. The completeness of the genome assembly was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCOs, v4.0.5[48]), and 72.6% of the BUSCO Tetrapoda gene annotations (version odb10) were identified (Table S3).

Targeted long-read sequencing of intronic sequences of ATP1A1

Intron annotations were determined using BLAST[49] (blastn v2.26) the protein-coding sequences of ATP1A1 S and ATP1A1 R against the *L. fuscus* genome assembly (Figure S3). For the other four *Leptodactylus* species (*L. pentadactylus*, *L. macrosternum*, *L. insularum*, and *L. colombiensis*) and two outgroup species (*Engystomops pustulosus* and *Lithodytes lineatus*), introns were obtained via targeted long-read sequencing using Oxford Nanopore MinION. Genomic DNA was extracted with Agencourt DNAdvance Kit (Beckman Coulter, France) and ATP1A1 was amplified using LongAmp *Taq* PCR kit (NEB) using customized species-specific barcoded primers (See Table S6). PCR products were gel confirmed and isolated using QIAquick PCR Purification kit (QIAGEN). Libraries were pooled and prepared for sequencing using Ligation Sequencing Kit SQK-LSK109 (Oxford Nanopore Technologies) following the manufacturer's protocol. 72,161 reads were generated within six hours, 89% passed the filter, and the real-time read length distribution matched that shown on the gel image of the amplicons. Base-calling from raw trace data was performed using Albacore v2.3.4 (Oxford Nanopore Technologies) and sequences were demultiplexed using LAST v980.[50] Reads that mapped to more than one barcode were discarded. Reads were assigned to each species based on barcodes using seqtk.[51] Only reads of the expected length ± 200 nt were used for downstream analyses. For *Leptodactylus* species with two ATP1A1 paralogs, reads were further split by perfectly matching the 111-122 region of the two copies, which exhibit 22%–25% difference in nucleotide sequences. Assembly was carried out using Canu v1.8[52] using -nanopore-raw with an estimated genome size of 5.3 kb. 1000 reads (1000x coverage) were randomly selected for better performance. Reconstructed sequences were identical when different sets of 1000 reads were used. Filtered reads were mapped back to the reconstructed reference with minimap2[53] and polished with racon v1.3.3.[54] Short-read sequencing data were generated using Tn5 transposase-based Illumina sequencing (as described above) to further correct and polish the sequences. Final sequences were aligned using MUSCLE[62] implemented in SeaView.[63] The boundaries between introns and exons were manually adjusted to start with GT and end with AG. Sequences are available at GenBank MT422192 - MT422203 (Table S2).

Estimation of genealogical relationships

A time-tree of anuran species in Figure 1A was derived from Feng et al.[30] Amino acid substitutions at sites that are implicated in cardenolide sensitivity[23] are shown. The nucleotide tree and protein tree (Figures 1B and 1C) of *Leptodactylus* and outgroup species were built with the exons and introns and protein sequences (Table S2), respectively. The best DNA and protein models were selected

using MEGA 7 based on AIC[56] (GTR+Γ+I for frog ATP1A1, K2P+Γ+I for *Leptodactylus* nucleotides and JTT+Γ+I for *Leptodactylus* protein). Phylogenies for ATP1A1 were reconstructed using a maximum likelihood method with 100 bootstraps and visualized in EvolView.[57] The alignment is available through a link provided in the Key resources table.

We estimated a species tree for three *Leptodactylus* species (*L. fuscus, L. pentadactylus*, *L. macrosternum*) and two outgroups (*Engystomops pustulosus* and *Lithodytes lineatus*) with high-confidence split time estimates specifically for use in the analyses described in sections "Theoretical single-site model for the probability of maintaining an adapted substitution" and "Simulations of ATP1A1 gene family evolution." Protein-coding genes were predicted from *de novo* transcriptome assemblies for each species using Augustus (v3.2.2)[58] and queried against the Tetrapoda ortholog database (odb10, https://www.orthodb.org) using BLAST (tblastn). A concatenated multi-alignment of cDNA sequences was created for 813 orthologous proteins longer than 100 amino acids that were shared among all five species. The best-fit nucleotide substitution model for each protein (i.e., each initial partition) was first determined using the "ModelFinder" function of IQ-TREE 2[59] (v.2.0.4) (command line: iqtree2 -s concat_813_mafft.fasta -p partition.txt -m MFP -nt AUTO -safe–prefix concat_813_partition_MFP). Proteins with the same inferred mutation model were subsequently concatenated into the same partition (using "-m TESTMERGE") prior to phylogenetic inference (command line: iqtree2 -s concat_813_mafft.fasta -p partition_MFP_best_scheme.nex -m TESTMERGE -nt AUTO–prefix concat_813_partition_MFP_merged).

Maximum likelihood analysis of site-wise support for alternative tree topologies

We used site-wise likelihoods to evaluate the relative level of statistical support for two alternative tree topologies relating to the origin of R/S ATP1A1 paralogs: Model 1 ("Non-Concerted") posits a single ancient origin of a R/S duplication with no concerted evolution: ((Lfus_S,(Lpen_S,(Lins_S,Llat_S,Lcol_S))),(Lfus_R,(Lpen_R,(Lins_R,Llat_R,Lcol_R)))). Model 2 ("Concerted") is the expected topology under concerted evolution: ((Lfus_S, Lfus_R), ((Lpen_S, Lpen_R), ((Lins_S, Lins_R), (Llat_S, Llat_R), (Lcol_S, Lcol_R)))). We note that the speciation events are assumed to follow the order inferred in the section "Estimation of genealogical relationships." For each nucleotide state (e.g., AAAATTTTTT, in the order of Lfus_S, Lfus_R, Lpen_S, LpenR, Llat_S, LlatR, LcolS, LcolR, LinsS, LinsR), likelihoods for the two topologies were calculated using *PAML 4.8 baseml*.[60] We consider $|\Delta\text{log-likelihood}| \geq 2$, as significant support for one topology over the other. 4-, 2-, 0-fold degenerate sites were classified using *MEGA* 7[56] and all variants at these sites were categorized as either synonymous or nonsynonymous. We used Fisher's Exact Test to test the hypothesis that the ratio of synonymous and nonsynonymous variants is independent of support for one of the topologies over the other (Table 1). The conclusions with respect to Nonsynonymous versus Synonymous/Intronic variants are not different if we assume the phylogenetic relationships to be ((Lfus, Lpen), (Lins, Llat, Lcol)) instead of (Lfus, Lpen, (Lins, Llat, Lcol)).

We further tested whether synonymous variants supporting alternative tree topologies (as outlined above) are equally distant from R/S distinguishing substitutions: We computed the distance of each variant from the nearest R/S distinguishing substitution, and compared the median distance of synonymous variants with $|\Delta\text{log-likelihood}| \geq 2$ support for the "Non-Concerted" genealogy to a random sample of synonymous variants supporting multiple origins.

Theoretical single-site model for the probability of maintaining an adapted substitution

Below, we describe the model and parameters used to compute the probability of maintaining a diverged substitution in two gene copies.

Model

We consider a single biallelic amino acid site in tandemly duplicated genes, evolving for t years. The two gene copies are initially fixed for the two distinct alleles. The site experiences mutation at rate 2μ (or 4μ for both copies) where μ is the per-nucleotide mutation rate, assuming for simplicity that all sites are biallelic, all mutations in the first two positions of the codon are nonsynonymous and all mutations at the third position are synonymous. The site also experiences non-allelic gene conversion at rate $4c$ (for both copies) and is under purifying selection with fitness cost $s > 0$, such that having two distinct alleles at the two copies confers a fitness of 1 and having the same allele confers to fitness $(1 - s)$.

De novo mutations (through point mutation or gene conversion) from the initial distinct-allele haplotype to a same-allele haplotype can occur in all haplotypes in the population. In a diploid population of size N, *de novo* same-allele haplotypes arise at rate

$$P(\textit{de novo same} - \textit{allele haplotype}) = 2N \cdot 4 \cdot (\mu + c).$$

The probability of fixation is bounded by the neutral case of $s = 0$, such that

$$P(\textit{same} - \textit{allele haplotype fixes}) < \frac{1}{2N}.$$

If

$$8N \cdot (\mu + c) \ll 1$$

and

$$\frac{1}{2N} \ll 1,$$

then the overall per-year rate of fixation for deleterious haplotypes, α, can be approximated by the product of these two,

$$\alpha = P(\textit{de novo same} - \textit{allele haplotype}) \cdot P(\textit{same} - \textit{allele haplotype fixes}) =$$

$$8N(c+\mu) \cdot \frac{e^{s}-1}{e^{2Ns}-1},$$

where we replaced $P(\textit{deleterious haplotype fixes})$ with Kimura's fixation probability for a deleterious allele.[64,65] Assuming a vanishingly small probability of back-mutations—namely, that no fixation of a same-allele haplotype is followed by another fixation reversing the haplotype back to the distinct alleles—the probability of maintaining the distinct-alleles haplotype for t years is:

$$P(\textit{maintenance of distinct alleles}) = (1-\alpha)^{t} = \left(1 - 8N(c+\mu)\frac{e^{s}-1}{e^{2Ns}-1}\right)^{t}. \quad \text{(Equation 1)}$$

Although we only use the general maintenance probability of Equation 1 in what follows, we note that if $s \ll 1$ then

$$e^{s} \approx 1+s,$$

and therefore

$$P(\textit{maintenance of distinct alleles}) \approx \left(1 - 4(c+\mu)\frac{2Ns}{e^{2Ns}-1}\right)^{t}, \quad \text{(Equation 2)}$$

giving a maintenance probability that is only dependent on the effective population size and the selection coefficient through the compound population parameter $2Ns$.

Parameters

To compute maintenance probabilities, we set the point mutation rate to its estimate by Sun et al.[66] (also supported by earlier work from Crawford[55]) of

$$\mu = 0.776 \cdot 10^{-9} \text{ mutations per bp per year.} \quad \text{(Equation 3)}$$

We wished to use the total branch length of the *Leptodactylus* phylogeny for t, the maintenance time, to reflect the observation of trans-specific maintenance. In considering the phylogenetic tree and split times here and in the evolutionary simulations of the section "Simulations of ATP1A1 gene family evolution" below, we only considered a subset of three *Leptodactylus* species— *L. fuscus, L. macrosternum* and *L. pentadactylus*—for which confident species split time estimates were available (see "Estimation of genealogical relationships" section; Figure S4): a split between *L. fuscus* and the common ancestor of the two other species 29,187,798 years ago, followed by a split between *L. macrosternum* and *L. pentadactylus* 27,426,120 years ago. Therefore, the total time on the species tree was set to

$$t = 2 \cdot 29,187,798 + 27,426,120 = 85,801,716 \textit{ years}. \quad \text{(Equation 4)}$$

The maintenance probabilities shown in Figure 3A were computed using Equation 1, plugging in the parameters in Equations 3 and 4 and across a grid of $Ns \in [-1, 1.5]$ and $c \in [0, 2.5]$ values.

Simulations of ATP1A1 gene family evolution

Overview

We developed evolutionary simulations with the goal of gauging the evolutionary parameters that could have produced the observed spatial divergence patterns along ATP1A1. Typically, and whenever possible, analytic likelihood or posterior probability functions are derived for such a task. Alternatively, backward-in-time simulations are used, because of their high computational efficiency. However, analytic or backward-in-time approaches were intractable for our purposes: both because we wished to account for the spatial divergence patterns and not consider sites independently—and because our model of ATP1A1 evolution in *Leptodactylus* includes complex interactions between point mutation, NAGC, and selection that violate typical assumptions of analytic / backward in time sequence evolution models. We therefore developed a forward-in-time simulation of R and S. The simulations take a set of parameters Θ as input (see section "Fitness model and other parameterization" below), start with two ancestral sequences and end with an output of contemporary R and S sequences in multiple *Leptodactylus* species, which we later compare to the observed data (see section "Inference of evolutionary parameters using Approximate Bayesian Computation").

Fitness model and other parameterization

At the heart of our simulation, we consider the possible fixation of new haplotypes in *Leptodactylus* lineages. These fixations follow random occurrence of *de novo* point mutations or NAGC in one of the haplotypes in the population; but the probability of fixation on the lineage will depend on the selection acting on the novel variant.

The ancestral haplotype with which the simulation begins is assumed to underlie the optimal function of R, S and interactions between them, and thus to be of optimal fitness. Therefore, the absolute fitness f of a haplotype X at any point of the simulation depends on its divergence from the ancestral haplotype with which the simulation begins, as follows:

$$f(X) = s_1X_1 + s_2X_2 + s_yY + s_zZ + s_{12}X_1X_2 + s_{1y}X_1Y + s_{2y}X_2Y,$$

where $X_1 \in \{0,1,2\}$ is the number of residue differences between X and the ancestral haplotype at position 111 of the amino acid sequences of both R and S; $X_2 \in \{0,1,2\}$ is the number of residue differences between X and the ancestral haplotype at position 122; $Y \in \{0,1,\ldots,20\}$ is the number of residue differences between X and the ancestral haplotype at the other 10 R/S distinguishing substitutions (referring to the substitutions strongly distinguishing R and S in the observed sequences); and Z is the number of total residue differences between X and the ancestral haplotype in the rest of the amino acid sequence. $\{s_1, s_2, s_y, s_z, s_{12}, s_{1y}, s_{2y}\}$ represent selection coefficients and are fixed parameters that are taken as input of the simulation.

Other parameters taken as input by our simulation (see pseudocode below) include:

N, the population size of each extant *Leptodactylus* lineage
μ, the per haplotype, per nucleotide per year mutation rate.
l, the mean NAGC tract length in base pairs. We model the tract length as Geometrically distributed.[41,67]
c, the NAGC per nucleotide per year rate. Note that this is the rate in which a site is included in a NAGC tract, not the rate at which NAGC events initiate at the site.
A rooted species tree, consisting of a bifurcating topology and branch lengths (split times) in years.

Simulation pseudocode

1. Initialize time t to the TMRCA of all species.
2. While t < today,

2.1. Advance t by t_w, the waiting time for the next mutational event, where

$$t_w \sim Exp((2N\ haplotypes)\cdot(extant\ species)\cdot(2\ paralogs\ per\ species)\cdot(ATP1A1\ sequence\ length)\cdot(rate\ per\ nucleotide\ c+\mu)).$$

2.2 If t > time for lineage split that had not yet occurred,
2.2.1 bifurcate lineage: copy R and S sequences of ancestral lineage into an identical copy and label each of the two sets as one of the lineages.
2.3 Draw $U_{event} \sim U(0,1)$. *If* $U_{event} < (\mu/\mu + c)$ then the de novo mutational event is a point mutation, else, it is a NAGC event.
2.4 Draw (uniformly) an extant species in which the event occurred.
2.5 Draw (uniformly) a paralog (R or S) in which the mutation occurred or served as the template for NAGC.
2.6 Draw (uniformly) a random nucleotide position where the mutational event occurred.
2.7 If the *de novo* event is a NAGC event,
2.7.1 Draw a tract length $L \sim Geo(l)$. Expand tract around initiation site, with a uniform fraction extending to the left and right of the site.
2.8 Translate the derived, *de novo* haplotype and the ancestral haplotype to amino acid sequences and calculate their fitness; calculate the resulting relative fitness of the derived haplotype.
2.9 Calculate p_{fix}, the fixation probability (see below) for a haplotype at frequency $(1/2N)$ conferring relative fitness as calculated in 2.8.
2.10 *Draw* $U_{fix} \sim U(0,1)$. *If* $U_{fix} < p_{fix}$,
2.10.1 Fix: Replace ancestral haplotype in the species with the *de novo* haplotype.

In step *2.9*, we consider a *de novo* haplotype arising in the population (namely, at frequency $1/2N$) with relative fitness $1+s$ to have probability

$$p_{fix} = \begin{cases} \dfrac{e^s - 1}{e^{2Ns} - 1} & if\ s < 0 (deleterious) \\ \dfrac{1}{2N} & if\ s = 0\ (neutral) \\ \dfrac{1 - e^{-s}}{1 - e^{-2Ns}} & if\ s > 0 (advantageous) \end{cases}$$

of fixing in the population, following Kimura.[64]

Inference of evolutionary parameters using Approximate Bayesian Computation

Overview

We used an Approximate Bayesian Computation (ABC) approach to estimate evolutionary parameters, including gene conversion rates and the strength of purifying selection acting at different sites in ATP1A1. In each iteration j, we sampled a set of parameters Θ_j from a predefined prior distribution. We approximated the posterior distribution of Θ_j by the empirical distribution given by a subset of this sample that generates divergence patterns that we inferred as closest to the true data. To infer the "distance" of simulated data from the observed data, we ran forward-in-time evolutionary simulations of ATP1A1 sequence evolution and quantified the similarity of the simulated divergence patterns to the observed divergence patterns. Simulations all begin with the same ancestral R and S genes in a common ancestor, and end with six evolved (simulated) contemporary sequences, corresponding to R and S in three

Leptodacylus species. From the divergence patterns between these six simulated sequences, we computed $d(\Theta_j)$, the distance between the simulated and the observed (real sequence data) ATP1A1 divergence patterns.

Parameter set and prior distribution

Our evolutionary simulations take as input a set of parameters as defined in the section "Simulations of ATP1A1 gene family evolution,"

$$\Theta = \{\mu, c, l, N, s_1, s_2, s_z, s_y, s_{12}, s_{1y}, s_{2y}\}.$$

The prior distributions of single parameters are mutually independent. Namely, the prior distribution on Θ was set as

$$\pi(\Theta) = \pi_c(c)\pi_{\tilde{s}}(\tilde{s})\pi_{s_z}(s_z),$$

where π_K is the marginal prior distribution of K, and $\tilde{s} := s_1 = s_2 = s_y$ such that all 12 sites distinguishing R and S in the observed data are under the same selective constraint, but it is free to differ from the selective constraint on other amino acids. The reason for setting $s_1 = s_2 = s_y$ is statistical: we have empirically found that our inference scheme has very little resolution on the strength of selection at individual sites (amino acid positions 111 and 122), and therefore focus on estimating the strength of selection against homogenization using this simplifying assumption. Similarly, there is very limited resolution given by our inference scheme on the selective interaction terms $s_{12}, s_{1y}$and s_{2y} when we allowed them to vary. We therefore set these fitness interaction terms to zero. The marginal priors on the gene conversion rate c and selection coefficients $\tilde{s}$, s_z were set as

$$\log_{10}\left(\frac{c}{\mu}\right) \sim U(0, 2.5),$$

$$\log_{10}(N\tilde{s}) \sim U(-1, 1)$$

and

$$\log_{10}(Ns_z) \sim U(-1, 1).$$

The other parameters were assumed fixed: we set the mutation rate to be $\mu = 0.776 \cdot 10^{-9}$ mutations per bp per year and the diploid population size (in each extant species at a given time in the simulation) to be N = 10 (2N = 20) as in the section "Theoretical single-site model for the probability of maintaining an adapted substitution." This small population size was chosen to allow for computational efficiency, because the simulation run time scaled linearly with N, and our inference became computationally infeasible with substantially larger population sizes. The mean tract length for gene conversion events was set to $l = 100bp$.

Measuring similarity to observed divergence patterns

Given y, a set of R and S nucleotide sequences in three species, we computed two summaries of the divergence at each nucleotide site i: $d_o(y_i)$, the sum of pairwise Hamming distances between R sequences in a pair of species (each $\in \{0, 1\}$ since only one site is considered) plus the sum of pairwise Hamming distances between S sequences; and $d_p(y_i)$, the sum—across the three species—of Hamming distances between paralogous R and S sequences. Let y^{obs} be the six observed sequences and y^{Θ_j} be the sequences output at the end of simulation run j. We measured the divergence between the simulated and observed data at site i as

$$d_i(\Theta_j) = d_i\left(y^{obs}, y^{\Theta_j}\right) = d_o\left(y_i^{obs}, y_i^{\Theta_j}\right) + d_p\left(y_i^{obs}, y_i^{\Theta_j}\right).$$

This per-site distance was computed for all positions I, namely nucleotide sites without missing data or insertions/deletions in any of the six observed sequences. Finally, the distance between simulation j and the observed data is given by

$$d(\Theta_j) = \sum_{\text{sites i}} w_i d_i\left(y^{obs}, y^{\Theta_j}\right),$$

where w_i are position-importance weights, giving extra weight for divergence patterns near R/S distinguishing sites—given that what we would like the parameters to recapitulate most are the spatial patterns around these sites. These weights were set as

$$w_i = 1 + \sum_{k=1}^{12} 10 \cdot e^{-|i - i_k|},$$

where $\{i_k\}$ is the set of $12 \cdot 3$ positions coding for one of the 12 R/S distinguishing substitution sites.

Analysis

We ran 23,323 simulations with Θ sampled from its prior distribution. We kept ~1% of these parameter sets—234 sets which produced simulations with the lowest $d(\cdot)$ values, and considered them as samples from the approximate posterior distribution. We then used the functions *kde3d* (for the approximate posterior distribution of c, s_z and $\tilde{s}$) and *kde2d* (for the marginal approximate posterior distribution of c and $\tilde{s}$) from the *R* packages *misc3d*[68] and *MASS*[69] to estimate the posterior with a spline fit using over

200 bins per dimension, in the range set by our prior distribution on each parameter, and with otherwise default settings of *kde3d and kde2d*. The approximate posterior mode was

$$(c = 18\mu, 2N\tilde{s} = 6, 2Ns_z = 1),$$

and the marginal posterior mode on the first two parameters was

$$(c = 9\mu, 2N\tilde{s} = 7).$$

The (single dimension) marginal credible interval mentioned in the main text are high posterior density credible intervals.

Construction of expression vectors

Na^+,K^+-ATPase is a multi-subunit protein that requires co-expression of the alpha (ATP1A) and beta subunits (ATP1B) in cell lines.[9] An RNA-seq analysis of *Leptodactylus* brain, stomach, and muscle tissues revealed that ATP1B1, one of four paralogous copies of ATP1B, is the most ubiquitously expressed. cDNA was reverse transcribed from *Leptodactylus macrosternum* stomach mRNA using the Superscript III Reverse Transcriptase kit (Invitrogen). The ATP1B1 gene was amplified from cDNA with the primers, 5′ATCCTCGAGATGGCCAGAGACAAAACCAAGGA 3′ and 5′ TGTGGTACCTCAGCTACTCTTAATCTCCAACTTTA 3′, which added a XhoI site at the 5′ end and a KpnI site at the 3′ end. ATP1B1 amplicons were inserted into pFastBac Dual expression vectors (Life Technologies) at the p10 promoter with XhoI and KpnI (FastDigest; Thermo Scientific), and then control sequenced. The vector insert sequence was an identical match to the *L. macrosternum* β1-subunit transcript generated in this study. ATP1A1S was amplified from cDNA with the primers 5′ TAATACTAGTATGGGATACGGGGCCGGACGTGAT 3′ and 5′ ACTGCGGCCGCTTAATAATAGGTT TCTTTCTCCA 3′ and ATP1A1R was amplified from a previously constructed vector containing a truncated copy of the gene with the overhang primers 5′ TAATACTAGTATGGGATACGGGGCCGGACGTGATGAGTATGAGCCCGCAGCCACTTCTGAACATGGCG GCAAGAAGAAAGGCAAAGGGAAGGATAAGGAT 3′ and 5′ ACTGCGGCCGCTTAATAATAGGTTTCTTTCTCCACCCAGCCGCCAGG GCTGCGTCTGATTATCAGTTTTCGGATTTCATCATATATGAAGATGAGCAGAGAGTAGGGGAAGGCACAGAACCACCATGTTGGTT TCAGTGGGTACATGCGGAGTGCCACATCCATGCCTGGG 3′. Both pairs of primers added a SpeI site at the 5′ end and a NotI site at the 3′ ends. All gene amplifications were performed using a high-fidelity proofreading polymerase (Phusion High-Fidelity DNA Polymerase; Thermo Fisher Scientific). ATP1A1S and ATP1A1R amplicons were inserted at the P_{PH} promoter of pFastBac Dual expression vectors already containing ATP1B1 with SpeI and NotI (FastDigest; Thermo Fisher Scientific), and then control sequenced. The ATP1A1S sequence was an identical match to the *L. macrosternum* sensitive α1-subunit transcripts and the ATP1A1R sequence was an identical match to *L. macrosternum* resistant α1-subunit transcripts generated from this study. Either *Escherichia coli* DH5α cells (Invitrogen) or *Escherichia coli* XL 10-Gold (Agilent Technologies, La Jolla, CA, USA) were transformed with the two resulting expression vectors (pFastBac Dual + ATP1B1 + ATP1A1S and pFastBac Dual + ATP1B1 + ATP1A1R). These completed vectors were then used to introduce the amino acid codons of interest by site-directed mutagenesis (QuikChange II XL Kit; Agilent Technologies, La Jolla, CA, USA) according to the manufacturer's protocol. One ATP1A1S gene construct was synthesized by Invitrogen GeneArt (S+12R). All resulting vectors had the α1-subunit gene under the control of the P_{PH} promoter and the β1-subunit gene under the p10 promoter (Table S4).

Generation of recombinant viruses and transfection into Sf9 cells

Escherichia coli DH10bac cells harboring the baculovirus genome (bacmid) and a transposition helper vector (Life Technologies) were transformed according to the manufacturer's protocol with expression vectors containing the different gene constructs. Recombinant bacmids were selected through PCR screening, grown, and isolated.[70] Subsequently, Sf9 cells (4×10^5 cells*ml) in 2 mL of Insect-Xpress medium (Lonza, Walkersville, MD, USA) were transfected with recombinant bacmids using Cellfectin reagent (Thermo Fisher). After a three-day incubation period, recombinant baculoviruses were isolated (P1) and used to infect fresh Sf9 cells (1.2×10^6 cells*ml) in 10 mL of Insect-Xpress medium (Lonza, Walkersville, MD, USA) with 15 mg/ml gentamycin (Roth, Karlsruhe, Germany) at a multiplicity of infection of 0.1. Five days after infection, the amplified viruses were harvested (P2 stock).

Preparation of Sf9 cell membranes

For production of recombinant Na^+,K^+-ATPase, Sf9 cells were infected with the P2 viral stock at a multiplicity of infection of 1000. The cells (1.6×10^6 cells per ml) were grown in 50 mL of Insect-Xpress medium (Lonza, Walkersville, MD, USA) with 15 mg/ml gentamycin (Roth, Karlsruhe, Germany) at 27°C in 500 mL flasks.[36] After 3 days, Sf9 cells were harvested by centrifugation at 20,000 x g for 10 min. The cells were stored at −80°C, and then resuspended at 0°C in 15 mL of homogenization buffer (0.25 M sucrose, 2 mM EDTA, and 25 mM HEPES/Tris; pH 7.0). The resuspended cells were sonicated at 60 W (Sonopuls 2070, Bandelin Electronic Company, Berlin, Germany) for three 45 s intervals at 0°C. The cell suspension was then subjected to centrifugation for 30 min at 10,000 x g (J2-21 centrifuge, Beckmann-Coulter, Krefeld, Germany). The supernatant was collected and further centrifuged for 60 min at 100,000 x g at 4°C (Ultra- Centrifuge L-80, Beckmann-Coulter) to pellet the cell membranes. The pelleted membranes were washed once and resuspended in ROTIPURAN p.a., ACS water (Roth) and stored at −20°C. Protein concentrations were determined by Bradford assays using bovine serum albumin as a standard. Six biological replicates were produced for each construct.

Verification by SDS-PAGE and western blotting

For each biological replicate, 50 ug of protein were solubilized in 4x SDS-polyacrylamide gel electrophoresis sample buffer and separated on SDS gels containing 10% acrylamide. Subsequently, they were blotted on nitrocellulose membrane (HP42.1, Roth). To block non-specific binding sites after blotting, the membrane was incubated with 5% dried milk in TBS-Tween 20 for 1 h. After blocking, the membranes were incubated overnight at 4°C with the primary monoclonal antibody α5 (Developmental Studies Hybridoma Bank, University of Iowa, Iowa City, IA, USA). Because only membrane proteins were isolated from transfected cells, detection of the α subunit also indicates the presence of the β subunit. The primary antibody was detected using a goat-anti-mouse secondary antibody conjugated with horseradish peroxidase (Dianova, Hamburg, Germany). The staining of the precipitated polypeptide-antibody complexes was performed by addition of 60 mg 4-chloro-1 naphtol (Sigma-Aldrich, Taufkirchen, Germany) in 20 mL ice-cold methanol to 100 mL phosphate buffered saline (PBS) containing 60 μl 30% H_2O_2. See Figure S5.

Ouabain inhibition assay (measurement of CS resistance)

To determine the sensitivity of each Na^+,K^+-ATPase construct against the water-soluble cardiotonic steroid, ouabain (Acrōs Organics), 100 ug of each protein was pipetted into each well in a nine-well row on a 96-well microplate (Fisherbrand) containing stabilizing buffers (see buffer formulas in Petschenka et al.[71]). Each well in the nine-well row was exposed to exponentially decreasing concentrations (10^{-3} M, 10^{-4} M, 10^{-5} M, 10^{-6} M, 10^{-7} M, 10^{-8} M, dissolved in distilled H_2O) of ouabain, distilled water only (experimental control), and a combination of an inhibition buffer lacking KCl and 10^{-2} M ouabain to measure background ATPase activity (see Petschenka et al.[71]). The proteins were incubated at 37°C and 200 rpms for 10 minutes on a microplate shaker (Quantifoil Instruments, Jena, Germany). Next, ATP (Sigma Aldrich) was added to each well and the proteins were incubated again at 37°C and 200 rpms for 20 minutes. The activity of Na^+,K^+-ATPases following ouabain exposure was determined by quantification of inorganic phosphate (Pi) released from enzymatically hydrolyzed ATP. Reaction Pi levels were measured according to the procedure described by Taussky and Shorr[72] (see Petschenka et al.[71]). All assays were run in duplicate and the average of the two technical replicates was used for subsequent statistical analyses. Absorbance for each well was measured at 650 nm with a plate absorbance reader (BioRad Model 680 spectrophotometer and software package).

ATP hydrolysis assay (measurement of ATPase activity as a proxy for protein activity)

To determine the functional efficiency of different Na^+,K^+-ATPase constructs, we calculated the amount of Pi hydrolyzed from ATP per mg of protein per minute. The measurements were obtained from the same assay as described above. In brief, absorbance from the experimental control reactions, in which 100 μg of protein was incubated without any inhibiting factors (i.e., ouabain or buffer excluding KCl), were measured and translated to mM Pi from a standard curve that was run in 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).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses of biochemical assay results

Background phosphate absorbance levels from reactions with inhibiting factors were used to calibrate phosphate absorbance in wells measuring ouabain inhibition and in the control wells.[71] For ouabain sensitivity measurements, calibrated absorbance values were converted to percentage non-inhibited Na^+,K^+-ATPases activity based on measurements from the control wells.[71] These data were plotted and log IC_{50} values were obtained for each biological replicate from nonlinear fitting using a four-parameter logistic curve, with the top asymptote set to 100 and the bottom asymptote set to zero (Figure S6). Curve fitting was performed with the nlsLM function of the minipack.lm library in R.[61] For comparisons of recombinant protein ATPase activity, the calculated Pi concentrations of 100 μg of protein assayed in the absence of ouabain were converted to nmol Pi/mg protein/min. We used ANOVA to test for effects of substitutions on ouabain resistance (log IC_{50}) and enzyme activity (Table S5; Levene's Test for Homogeneity of Variance for IC_{50}: $F_{7,40} = 0.68$ $p = 0.69$ and enzyme activity: $F_{7,40} = 0.31$ $p = 0.94$). We used linear regression to estimate effect sizes associated with substitutions and pairwise t tests to identify significant differences between substitution combinations (Table S5). All statistical analyses were implemented in R.