PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Yang L, Ravikanthachari N, Mariño-Pérez R, Deshmukh R, Wu M, Rosenstein A, Kunte K, Song H, Andolfatto P. 2019 Predictability in the evolution of Orthopteran cardenolide insensitivity. *Phil. Trans. R. Soc. B* **374**: 20180246.
http://dx.doi.org/10.1098/rstb.2018.0246

Accepted: 25 February 2019

One contribution of 16 to a theme issue 'Convergent evolution in the genomics era: new insights and directions'.

Subject Areas:
evolution, genomics, bioinformatics, computational biology, genetics

Keywords:
Orthoptera, cardenolide, parallel evolution, Na^+,K^+-ATPase, toxin insensitivity, pleiotropy

Authors for correspondence:
Lu Yang
e-mail: lyang24203@gmail.com
Peter Andolfatto
e-mail: pa2543@columbia.edu

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4472423.

THE ROYAL SOCIETY PUBLISHING

Predictability in the evolution of Orthopteran cardenolide insensitivity

Lu Yang[1], Nitin Ravikanthachari[2], Ricardo Mariño-Pérez[3], Riddhi Deshmukh[2], Mariana Wu[1], Adam Rosenstein[1], Krushnamegh Kunte[2], Hojun Song[3] and Peter Andolfatto[4]

[1]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA
[2]National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru, India
[3]Department of Entomology, Texas A&M University, College Station, TX 77843, USA
[4]Department of Biological Sciences, Columbia University, New York, NY 10027, USA

LY, 0000-0002-2694-1189; NR, 0000-0002-9474-7951; RM-P, 0000-0002-0566-1372; RD, 0000-0002-7634-2029; KK, 0000-0002-3860-6118; HS, 0000-0001-6115-0473; PA, 0000-0003-3393-4574

The repeated evolutionary specialization of distantly related insects to cardenolide-containing host plants provides a stunning example of parallel adaptation. Hundreds of herbivorous insect species have independently evolved insensitivity to cardenolides, which are potent inhibitors of the alpha-subunit of Na^+,K^+-ATPase (ATPα). Previous studies investigating ATPα-mediated cardenolide insensitivity in five insect orders have revealed remarkably high levels of parallelism in the evolution of this trait, including the frequent occurrence of parallel amino acid substitutions at two sites and recurrent episodes of duplication followed by neo-functionalization. Here we add data for a sixth insect order, Orthoptera, which includes an ancient group of highly aposematic cardenolide-sequestering grasshoppers in the family Pyrgomorphidae. We find that Orthopterans exhibit largely predictable patterns of evolution of insensitivity established by sampling other insect orders. Taken together the data lend further support to the proposal that negative pleiotropic constraints are a key determinant in the evolution of cardenolide insensitivity in insects. Furthermore, analysis of our expanded taxonomic survey implicates positive selection acting on site 111 of cardenolide-sequestering species with a single-copy of ATPα, and sites 115, 118 and 122 in lineages with neo-functionalized duplicate copies, all of which are sites of frequent parallel amino acid substitution.

This article is part of the theme issue 'Convergent evolution in the genomics era: new insights and directions'.

1. Introduction

Two enduring fundamental questions in modern evolutionary biology are *what factors limit the rate of adaptive evolution?* and *to what extent are adaptive evolutionary paths predictable?* [1] Theoretically, the predictability of adaptation depends on a number of factors that constrain the number of possible evolutionary paths. Among these is the number of potential targets for beneficial mutations [2–4]. However, the extent to which adaptation is constrained by mutation rate is unclear for most traits and some investigators have emphasized important roles for pleiotropy, the phenomenon by which one mutation affects multiple phenotypes, and epistasis, the effect of genetic background on the contribution of a mutation to a given phenotype, among other factors [1,5–7].

Evaluating the relative importance of these factors has been challenging. One approach has been to cobble together examples of adaptations from different traits in different species and contexts in an attempt to come to general conclusions (e.g. [8,9]). However, the heterogeneity inherent in such broad comparisons of different traits in different biological contexts may substantially limit the power to make inferences from such data [10]. An alternative is to

examine cases on a trait-by-trait basis in the context of adaptation to common selective pressure [10]. Instances of parallel evolution, the independent evolution of similar features in different lineages, can provide multiple portraits of the evolutionary process and offer insight into the factors that constrain adaptation and the extent to which adaptive evolutionary paths are predictable.

Examples of parallelisms from nature are abundant and occur at different scales from the resemblance of morphological traits to individual nucleotide substitutions that encode regulatory or protein changes [4,11,12]. Such examples will have greater power to make inferences about the factors determining the dynamics of adaptation for a given trait as the number of independent outcomes becomes larger, the more similar the selective pressure and the more is known about the genetic basis of the underlying trait. With these factors in mind, one fertile area for exploration is the repeated evolution of insensitivity of herbivorous insects to toxic secondary plant compounds. Plants are ubiquitously equipped with secondary chemical defences such as alkaloids, cyanogenic glucosides and terpenoids that contribute to defence against herbivory [13,14]. Despite these defences, herbivorous insects have in many cases repeatedly evolved mechanisms to render them insensitive to toxic compounds [14,15], and even sequester toxins for their own use [16].

A striking example is presented by herbivorous insects that have repeatedly evolved the ability to feed on and, in many cases, sequester cardenolides from Apocynaceae plants, which include milkweed [15]. Cardenolides represent a class of steroidal glycosides ('cardiac glycosides') that bind to and inhibit the alpha-subunit of Na^+,K^+-ATPase (ATPα). This protein is a ubiquitously distributed membrane-bound ion active-transporter present in animals with well-known roles in a variety of physiological processes including neural signal transduction, muscle contraction and osmoregulation [17]. Conservation of the cardenolide-binding domain of ATPα among distantly-related animals, including vertebrates and invertebrates, suggests important physiological roles for the regulation of Na^+,K^+-ATPase by endogenously produced cardenolides [17,18]. In fact, cardenolides have been used medicinally for hundreds of years as common treatments for conditions such as congestive heart failure and cardiac arrhythmias [18]. A growing number of studies have implicated the regulation of Na^+,K^+-ATPase by putatively endogenous cardenolides in signalling pathways linked to a variety of pathologies including hypertension and cancer [19,20].

Owing to its medical importance, the interaction between Na^+,K^+-ATPase and cardenolides has been well-studied. The binding of cardenolides arrests Na^+,K^+-ATPase in the phosphorylated state, where K^+ cannot be bound, Na^+ cannot be released to the extracellular side and ATP is not hydrolysed [21]. Mutagenesis experiments, enzyme-ligand co-crystal structures and evolutionary analyses have implicated 41 amino acid residues of ATPα, scattered throughout the protein, that either directly interact with cardenolides or affect their binding-affinity indirectly (references listed in the electronic supplementary material, table S1). These sites are largely concentrated near the site of cardenolide binding in ATPα, with some exceptions (electronic supplementary material, figure S1). As such, the evolution of cardenolide-insensitivity via the modification of ATPα (i.e. target-site insensitivity) is one of the rare traits for which we have a good *a priori* idea of the beneficial mutation target size for adaptation.

Broadly speaking, strategies employed by specialist herbivores to deal with toxin compounds include destroying and/or excreting the toxins [22,23], inactivating the toxins by chemical modifications [24], restricting the expression of the target protein to specific tissues [25,26], and/or the evolution of target-site insensitivity [24]. The evolution of ATPα insensitivity has so far been inferred in almost all Apocynaceae-specialist species surveyed from five insect orders, including Lepidoptera, Diptera, Coleoptera, Hymenoptera and Hemiptera [27–33].

Studies of the evolution of ATPα insensitivity in these five insect orders have revealed a remarkable degree of convergence of molecular mechanism at multiple levels. First, despite the identification of 41 residues in the protein that could potentially modulate sensitivity of ATPα to cardenolides (electronic supplementary material, table S1), there is a marked enrichment of substitutions in Apocynaceae-specialists observed at two sites in the protein, Q111 and N122, that flank the H1-H2 extracellular loop [30,33]; the substitutions Q111L, Q111T, Q111V and N122H occur in parallel in multiple lineages. Second, several rounds of duplication of ATPα have occurred in parallel in multiple species from three of the five surveyed insect orders, including Coleoptera, Hemiptera and Diptera [30,33]. In each case, at least one of the divergent copies harbours a number of cardenolide insensitivity-conferring substitutions [30,33,34]. Third, in most cases of ATPα duplication, the copies have been shown to exhibit parallel evolution of tissue-specific expression patterns, with putatively less cardenolide-sensitive copies predominating in the gut, and putatively more sensitive copies predominating in nervous tissue [30].

Taken together, these data suggest a key role for negative pleiotropy in the evolution of cardenolide-insensitivity in Apocynaceae-specialists: those species with a single copy are largely limited to evolution at just a few of the 41 possible sites, but those with duplicates can explore many other evolutionary paths using one of the differentially expressed duplicate copies [30]. Further support for the idea that negative pleiotropy plays a key role comes from site-directed mutagenesis studies demonstrating trade-offs in Na^+,K^+-ATPase function (e.g. efficiency of ATP hydrolysis) associated with some duplicate-specific substitutions observed in milkweed bugs (Hemiptera) [35].

The notion of predictability in evolution is often framed in terms of forecasting future events, for example, predicting the next steps in virus evolution [36]. However, it can also be used in the sense of looking at evolutionary patterns retrospectively and asking if a set of rules deduced from patterns of evolution in one group of organisms for a particular adaptation can reliably predict the genetic architecture of the same trait in another group. With this in mind, we survey a sixth insect order, the Orthoptera, which is phylogenetically positioned outside of the five insect orders that have previously been investigated (figure 1*a*). Within Orthoptera, we focus on the family Pyrgomorphidae, commonly known as gaudy grasshoppers. This group is a relatively small group of approximately 500 species that include some of the most colourful and showy grasshoppers in the world (figure 1*b–d*) [37]. Some members of this family that are known to feed on toxic plants, including Apocynaceae, possess aposematic coloration and are able to sequester plant secondary metabolites such as cardenolides and pyrrolizidine alkaloids [38–41]. In addition to the aposematic

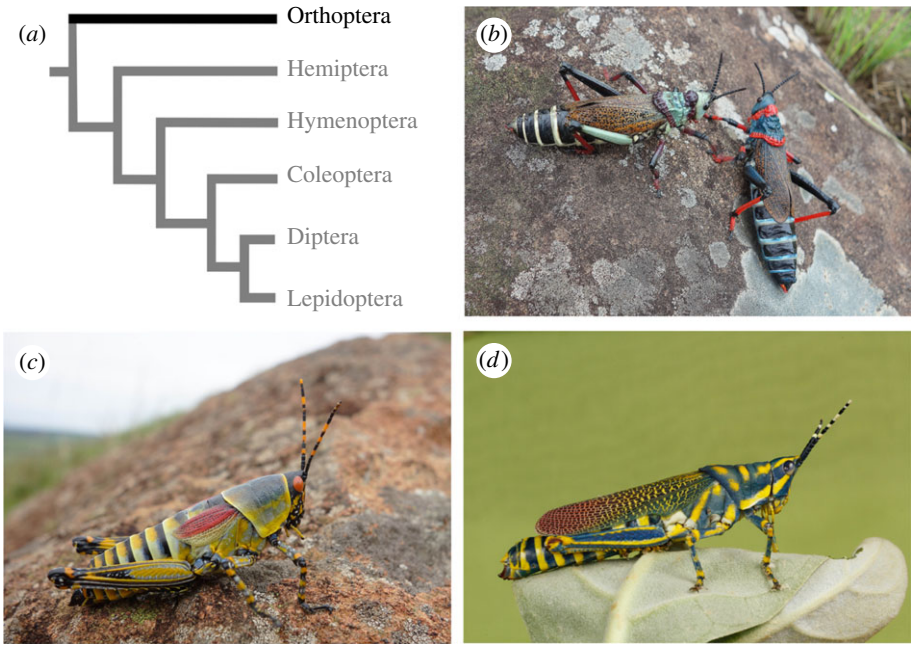

Figure 1. Phylogeny (*a*) and representative members of the Orthoptera: Pyrgomorphidae (*b*–*d*). Orthopteran species shown are (*b*) *Dictyophorus spumans* (South Africa); (*c*) *Zonocerus elegans* (South Africa); (*d*) *Poekilocerus pictus* (India).

coloration, some genera (such as *Phymateus*, *Poekilocerus*, *Zonocerus*) possess a unique mid-dorsal abdominal gland capable of squirting toxic chemical when disturbed, while others (*Aularches*, *Dictyophorus*, *Taphronota*) can produce foam as a result of haemolymph released through pores combined with air [41–44].

Studies of the common milkweed grasshopper *Poekilocerus bufonius* demonstrated that they were substantially less sensitive to the cardenolide ouabain injections (as measured by LD50) than species that do not feed on milkweed plants [45]. In addition, enzyme inhibition assays performed on extracts from *P. bufonius* suggest cardenolide insensitivity of Na^+,K^+-ATPase. In addition, observed heterogeneity among tissues in the degree of cardenolide-insensitivity was interpreted as possible evidence for distinct isoforms of the enzyme in this species [45]. While several species within Pyrgomorphidae show plant-mediated chemical defence, most of the species in the family do not possess aposematic coloration or feed on toxic plants [37], which presents an interesting opportunity to investigate the variation in cardenolide-insensitivity of ATPα.

To shed light on the evolution of cardenolide-insensitivity in Pyrgomorphidae, we generated RNA-Seq data for one representative grasshopper species from each of 10 genera in the family and reconstructed the ATPα by *de novo* transcriptomic assembly [30]. We find remarkably similar patterns of amino acid substitution to those observed in previous surveys of Apocynaceae-specialist herbivores, including a duplication event followed by neo-functionalization and tissue-specific differential expression in the genera *Phymateus* and *Poekilocerus*. This expanded dataset, now including data for 52 Apocynaceae-feeding species from six insect orders, further supports the view that adaptation in this system appears to be largely constrained by negative pleiotropic effects associated with otherwise adaptive substitutions. The dataset also affords increased power to detect positive selection acting on specific sites in the protein that are also sites of recurrent parallel amino acid substitution.

2. Material and methods

(a) Sequencing and *de novo* transcriptome assembly

For details on sample collection and preparation see the electronic supplementary material, table S2. Dissections were carried out in phosphate-buffered saline solution and stored either in TRIzol (Ambion, Life Technologies) or RNAlater (Ambion Inc.) at −80°C. For all insects, total RNA was extracted using TRIzol (Ambion, Life Technologies) following the manufacturer's protocol. RNA-seq libraries of *Aularches miliaris* and *Poekilocerus pictus* were prepared with TruSeq Stranded total RNA Library Prep Kit with Ribo-Zero Gold (Illumina) and sequenced on Illumina HiSeq2500 at AgriGenome (Cochin, Kerala, India). The libraries of *Taphronota calliparea* and *Dictyophorus griseus* were prepared with NEBNext Ultra RNA library Preparation Kit (NEB) and sequenced on Illumina HiSeq4000 (Genewiz, South Plainfield, NJ, USA). The libraries of *Chrotogonus hemipterus*, *Atractomorpha acutipennis*, *Zonocerus elegans*, *Phymateus leprosus*, *Ochrophlebia cafra* and *Sphenarium purpurascens* were prepared with TruSeq RNA Library Prep Kit v2 (Illumina) and sequenced either on Illumina HiSeq4000 (Genewiz, South Plainfield, NJ, USA) or HiSeq2500 (Genomics Core Facility, Princeton, NJ, USA). Reads were trimmed for adapters, quality and length (Phred quality ≥20 and ≥30 contiguous bases) using TQSfastq.py (http://code.google.com/p/ngopt/source/browse/trunk/SSPACE/tools/TQSfastq.py). All transcriptomes were *de novo* assembled with Trinity v. 2.2.0 [46]. ATPα of *Locusta migratoria* (GenBank: KF813097.1) was used to query the assembled transcripts using BLAST (blast-2.26). Reconstructions of ATPα for each species were used iteratively as query sequences to BLAST against each other using either tblastx or blastn to recover all ATPα copies.

We have also included previously unpublished full-length ATPα sequences for a number of other Apocynaceae-specialists including *Daphnis nerii* (Oleander hawk-moth, Lepidoptera), *Empyreuma pugione* (spotted Oleander caterpillar moth, Lepidoptera), *Euploea core* (the common crow, Lepidoptera), *Danaus chrysippus* (pain tiger, Lepidoptera), *Liriomyza asclepiadis* (milkweed leaf-miner fly, Diptera). The methods to reconstruct these sequences is identical to those used above following Zhen *et al.* [30]. Particular attention is paid to 41 sites implicated in

cardenolide-insensitivity (electronic supplementary material, table S1) established based on site-directed mutagenesis and protein–ligand co-crystal structure studies.

(b) Discovery and confirmation of duplicates

Given previous studies revealing duplications of ATPα associated with Apocynaceae-specialization [30], we evaluated evidence for duplication in our Orthopteran species data. Two ancient lineages of ATPα (ATPα1 and ATPα2) precede the diversification of multiple insect orders and form a distinct clade from the duplications of ATP1A found in vertebrates [30]. However, the expression level of ATPα2 is low in insects surveyed to date and its function remains largely unknown. In *Drosophila melanogaster*, expression of ATPα2 is limited to larval imaginal discs and adult male testes and accessory glands; RNAi and P-element knock-outs of the gene are homozygous viable but male-sterile (http://www.flybase.org/reports/FBgn0267363). Furthermore, we failed to recover an orthologue of ATPα2 from the *Locusta migratoria* genome assembly [47], and several other Orthopteran assemblies [48,49], suggesting that the ATPα1/α2 duplication may have arisen after the split of Orthopterans from the other insects we have surveyed (figure 1*a*). We thus decided to focus on ATPα1, which shows clear orthology across taxa and a strong correlation between evolution at known cardenolide-sensitivity sites in the protein and specialization on cardenolide-containing Apocynaceae plants [30].

The discovery of duplicated copies of ATPα1 in *Poekilocerus pictus* and *Phymateus leprosus* was verified by cloning and sequencing. Total RNA was extracted as described above, and reverse-transcribed to single-strand cDNA using SuperScript® III Reverse Transcriptase (Thermo Fisher Scientific). ATPα1 was polymerase chain reaction (PCR) amplified using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) using forward primer: 5′-ACATGG CGGCAAGAAGAAAG-3′ and reverse primer: 5′-AGTAGGG GAAGGCACAGAAC-3′. The PCR product was cleaned using QIAquick PCR Purification Kit (Qiagen), 3′A-tailed using Taq polymerase (NEB) and cloned into TOPO TA vector (Invitrogen) following the manufacturer's instructions. Ampicillin-resistant colonies were picked and screened by colony-PCR for the presence of inserts on a 1% agarose gel. Libraries of plasmids were constructed using Tn5 transposase [50] annealed with Tn5ME-A, 5′-TCGTCGGCAGCGTC AGATGTGTATAAGAGACAG-3′ and Tn5ME-B, 5′-GTCTCGTG GGCTCGGAGATGTGTATAAGAGACAG-3′ and indexed with customized Illumina i7, i5. Paired-end 150 nt reads were collected for the pooled library on an Illumina MiSeq Nano (Genomics Core Facility, Princeton University). *De novo* transcriptome assembly was performed with Velvet/Oases [51,52] using a random sub-sample of 10 000 reads for each indexed plasmid.

(c) Differential expression analysis

Head, muscle and foregut tissues of three male and three female *Poekilocerus pictus* were dissected and sequenced in two batches (see sample collection and sequencing). Adapter and quality trimmed reads were mapped back to our *de novo* assembly of the transcriptome as a reference (including ATPα1A and ATPα1B) using bwa mem [53] with default criteria, processed with SAMtools 0.1.18 [54] and mapped reads were counted with HTSeq 0.6.1 [55]. We used the inverted beta-binomial (ibb) test [56] to determine the significance of difference of expression level between tissues. The method uses a negative binomial distribution in a generalized linear model framework for paired-sample testing. We applied a standard Bonferroni correction to account for multiple tests. Paired-sample count data were normalized by either total number of mapped reads or the sum of reads mapping to ATPα1A and ATPα1B (electronic supplementary material, table S3).

(d) Re-analysis of Lygaeid ATPα1 evolution and expression

Using the recently completed *Oncopeltus fasciatus* genome [57], we detected a fourth copy of ATPα1 (ATPα1D) in these Lygaeid bugs that was missed by previous studies. Lygaeid ATPα1D is the least-derived copy at sites implicated in cardenolide-sensitivity and is expressed in *O. fasciatus* heads, yet has the lowest expression of the four copies (figure 3). We also confirmed that, like copies A-C, ATPα1D is also shared with *Lygaeus kalmii* (a sister-genus species) but we could only partially reconstruct it from our *de novo* transcriptome assembly owing to its low expression. Using RNA-seq data for *O. fasciatus*, we confirmed the finding of differential expression of duplicates documented by Zhen *et al.* [30] (electronic supplementary material, figure S3A). All four copies of ATPα1 in the Lygaeid *O. fasciatus* are more highly expressed in the head than the gut. The putatively most sensitive copy (ATPα1D), despite having the lowest expression level of the four copies, exhibits the greatest degree of upregulation in the head relative to other copies (electronic supplementary material, figure S3B).

(e) Evolutionary analyses

The ages of the duplicates were calculated from d*S* (the per site rates of substitution at synonymous sites) estimated using PAML4.8 codeml [58], with prior trees based on established cladistic relationships [59] and calibrated with divergence times obtained from www.timetree.org (*Locusta* and Pyrgomorphidae: 117.4 Mya; *Napomyza* and *Phytomyza*: 39 Mya). The divergence times of the large milkweed bug (*Oncopeltus fasciatus*), milkweed stem weevil (*Rhyssomatus lineaticollis*) and dogbane beetle (*Chrysochus auratus*) are taken from [30] where similar methods were used to date duplicates. In the Lygaeid bugs (*O. fasciatus* and *L. kalmii)*, a phylogenetic analysis strongly suggests a duplication order ((ABC),D), ((AB),C) and most recently (A,B).

To obtain distributions of lineage-specific evolutionary rates, ATPα1 lineages were grouped into four categories. ATPα1 of all non-specialist species were denoted as 'outgroup'. ATPα1 of Apocynaceae-specialists with a single copy of ATPα1 are denoted 'single'. For specialists with multiple copies of ATPα1, copies that are upregulated in the gut relative to the head are assumed to be relatively cardenolide-insensitive copies and marked as Dup^{I} [30]. Likewise, those upregulated in the head relative to the gut are assumed to be relatively sensitive copies and were grouped as Dup^{S}. We chose this criterion rather than the number of insensitivity-conferring substitutions because using the latter makes the designation of copies with intermediate numbers of substitutions ambiguous. In Lygaeid bugs (*O. fasciatus* and *L. kalmii)*, this implies that copies A and B are treated as relatively insensitive copies (Dup^{I}), whereas C and D are treated as relatively sensitive (Dup^{S}) (electronic supplementary material, figure S3). The d*N*/d*S* ratios (omega) for ATPα1 along a lineage were estimated within each insect order using PAML codeml under free ratio model. Parameters were set as follows: seqtype = 1, model = 1, NSsites = 0, clock = 0, CodonFreq = 2, fix_kappa = 0, kappa = 2.0, fix_omega = 0, omega = 0.02. Differences in the distribution of branch-specific estimates of omega between each group were tested with Dunn's test of multiple comparisons using rank sums as implement in R (dunn.test).

We also evaluated evidence for positive selection acting on individual sites of ATPα1 using PAML codeml. To do this, we defined and investigated several models (figure 7; electronic supplementary material, table S4). Model 1: positive selection on all Apocynaceae-specialist lineages; model 2: positive selection on all Apocynaceae-specialist lineages with single copies of ATPα1 (including ancestral lineages prior to duplication); model 3:

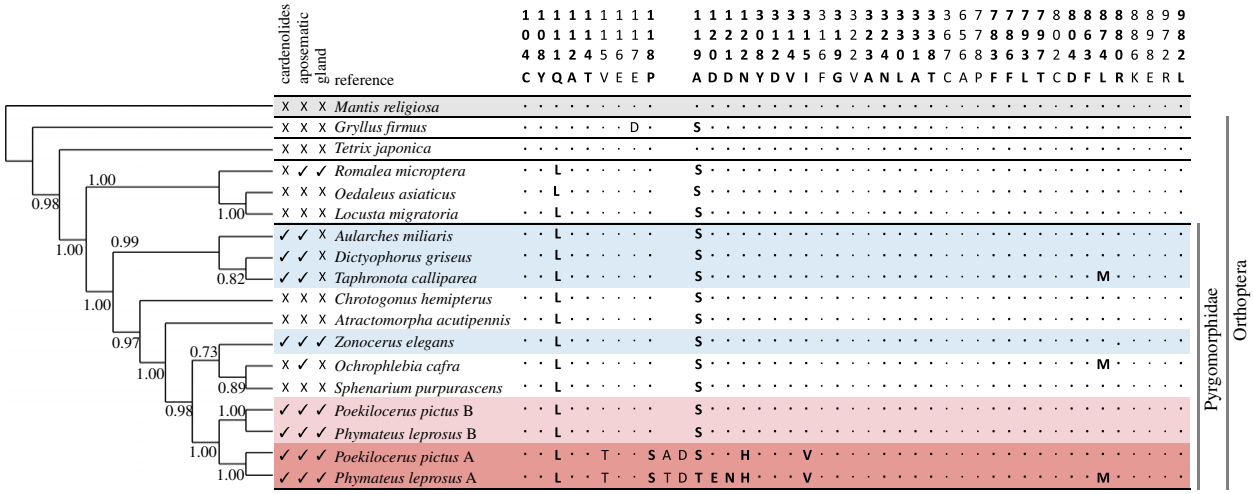

Figure 2. Amino acid substitutions at sites implicated in cardenolide-sensitivity for ATPα1 of Orthopterans in the family Pyrgomorphidae and outgroups. Each family is separated by black lines. The coloured rows correspond to putatively cardenolide-adapted species that either possess only one copy (light blue) or two copies (light/dark red) of ATPα1. The non-Orthopteran outgroup (*Mantis religiosa*, order Mantodea) is shaded in grey. The numbering of sites is based on sheep ATP1A1 (*Ovis aries*) (GenBank: NC019458.2). Bold columns correspond to the sites for which site-directed mutagenesis studies suggest a role in cardenolide-sensitivity, while the rest were identified by structural prediction. Dots indicate identity with the reference, which represents the consensus sequence among non-specialist Arthropods. Letters represent amino acid substitutions relative to the reference. The cladogram was constructed using a maximum-likelihood method implemented in SeaView based on protein-coding nucleotide sequences of ATPα1 with bootstrap values shown. 'Cardenolide' refers to species feeding on Apocynaceae, 'aposematic' refers to the presence of warning coloration patterns, 'gland' refers to the presence of a specialized abdominal defensive gland that secrete toxic chemicals.

positive selection on Dup^S and Dup^I lineages of Apocynaceae-specialists; model 4: positive selection on Dup^I lineages of Apocynaceae-specialists; model 5: positive selection on all outgroup lineages. Tests were carried out using the modified branch-site model A implemented in codeml [60–62] with parameters set as follows: seqtype = 1, model = 2, NSsites = 2, clock = 0, CodonFreq = 2, fix_kappa = 0, kappa = 2.0, fix_omega = 0, omega = 0.02. The ancestral sequences of duplicated ATPα1 were reconstructed with the function RateAncestor. Unrooted trees were used, and branch labels were added manually for each model. The method assigns each codon a Bayes Empirical Bayes (BEB) posterior probability that the codon belongs to a site class with omega greater than 1 (i.e. indicating positive selection). A BEB posterior probability of greater than 0.95 was considered evidence for positive selection.

3. Results

(a) Survey of ATPα1 of 15 Orthopteran genera

Using an RNA-seq-based gene discovery method, we reconstructed the complete coding sequences of the alpha subunit of Na^+,K^+-ATPase (ATPα1) of grasshoppers from species representing 10 genera in the family Pyrgomorphidae, as well as five outgroup species within, and one outside, the order Orthoptera (figure 2). Our broad survey of Apocynaceae-feeding Pyrgomorphidae revealed few amino acid substitutions among the 41 sites implicated in cardenolide-sensitivity (electronic supplementary material, table S1). The two most broadly distributed substitutions, Q111L and A119S, correlate only weakly with Apocynaceae-feeding, aposematism, and the presence of abdominal defensive glands in the group. The glaring exception is the lineage leading to the genera *Poekilocerus* and *Phymateus*, both containing multiple species, which appear to share a duplication of ATPα1. Both species surveyed retain an ancestral version of the protein (ATPα1B) while having a highly-derived copy (ATPα1A). The diverged ATPα1A copies of *Poekilocerus* and *Phymateus* share many amino acid substitutions relative to the ancestral copy, several at sites implicated in cardenolide-sensitivity. Phylogenetic analysis clearly indicates that the duplication of ATPα1 and functional divergence of ATPα1A predates the diversification of these clades into separate genera and species and we estimate the age of the duplication to be approximately 36 million years old.

(b) Patterns of amino acid substitution in ATPα1 of Orthopterans

Cross-referencing the pattern in Orthopterans with other Apocynaceae-specialists surveyed in other insect orders reveals a high level of parallel amino acid substitution (figure 3; electronic supplementary material, figure S2). Conspicuous among these are two parallel substitutions Q111L and A119S, which appear to predate the diversification of the Pyrgomorphidae. Cell transfection experiments have shown when substitutions at position 111 are introduced to a sensitive background of the *D. melanogaster* protein, the survival rate of HeLa cells increases three- to eight-fold [31]. Interestingly, A119S is observed in almost every Apocynaceae-specialist species that has been surveyed to date (electronic supplementary material, figure S2), including the *Drosophila subobscura* subgroup where resistant forms of ATPα1 have been documented segregating as polymorphisms within *D. subobscura* [63]. Exceptions include aphids (*Aphis nerii* and *Acyrthosiphon pisum*), the milkweed leaf beetle (*Labidomera clivicollis*) and several Hymenopteran species that harbour the similar substitution, A119N. Q111L and A119S are not associated with cardenolide-feeding or Apocynaceae-specialization in the Orthoptera and are possessed by a number of non-aposematic species not known

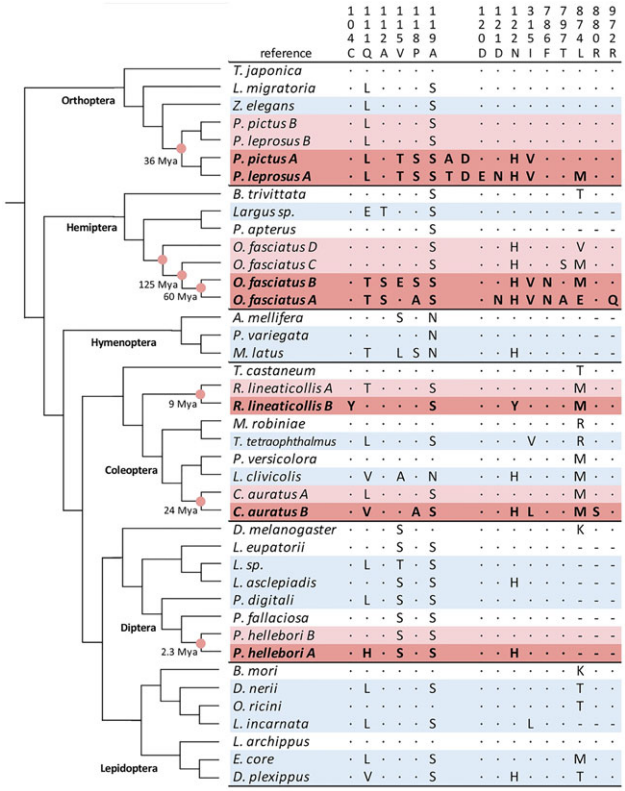

Figure 3. Patterns of amino acid substitution and duplication spanning six insect orders. Shown are sites implicated in cardenolide binding (electronic supplementary material, table S1) at which substitutions have occurred. The colour scheme is consistent with figure 2. Red circles indicate duplication events. The cladogram represents the relationships of insect orders and species phylogenies (see Methods).

to feed on cardenolide-containing plants (e.g. *L. migratoria*, *C. hemipterus*, *A. acutipennis*). The substitutions Q111L, A119S and A119N also occur sporadically among a number of other insects not known to feed on Apocynaceae (electronic supplementary material, figure S2).

Considering the cardenolide-insensitive ATPα1A copy-specific substitutions of *Poekilocerus* and *Phymateus*, N122H stands out as a substitution observed in parallel in many other Apocynaceae-specialists including at least one member of each of the six insect orders surveyed. The N122H substitution in isolation has been shown to increase *Drosophila* ATPα1 insensitivity to the cardenolide ouabain by 250-fold [64] and increase survival of HeLa cells challenged with ouabain [31]. N122H has also been reported to interact synergistically with substitutions at Q111, though Q111L was not tested [31]. The Orthopteran ATPα1A copy-specific substitutions V115T, P118S, D121N, I315V and L874M also occur in at least one other insect order. Of these, D121N is known to decrease cardenolide-sensitivity by approximately 100-fold [65].

(c) Unique substitutions associated with duplication of ATPα1 and differential expression of neo-functionalized copies

Using data from three insect orders, Zhen *et al.* [30] found a significant enrichment of unique substitutions at sites implicated in cardenolide-sensitivity in Apocynaceae-specialist lineages with duplicated copies of ATPα1 compared to those that retain a single copy. They also documented convergent patterns of differential gene expression of independently derived duplicates found in specialists. Specifically, they noted that duplicates inferred to be the most sensitive to cardenolides, based on the number of substitutions at sites implicated in cardenolide-insensitivity, consistently exhibited upregulation in the head relative to the gut. Zhen *et al.* [30] argued that this might be expected because the gut is the site of first-processing of cardenolides and sensitive forms of ATPα1 in nervous tissue are probably protected to some extent by the blood–brain barrier provided by the glial sheath surrounding neurons [30,66]. Zhen *et al.* [30] interpreted this pattern as being consistent with a key role for pleiotropy in the evolution of cardenolide-insensitivity. Specifically, they proposed that there might be trade-offs in enzyme performance associated with unique substitutions that are ameliorated by differential expression of neo-functionalized duplicate copies. Here, we re-evaluate this claim in the context of our expanded dataset, which now includes Orthoptera and previously published data for five other insect orders.

Despite the remarkable parallelism at the levels of gene duplication and amino acid substitutions, it is notable that 'unique' substitutions (i.e. those unique to one lineage) at sites implicated in cardenolide-sensitivity in Orthoptera are restricted to duplicated copies, as observed in other Apocynaceae-specialists with duplications. Notably, ATPα1A of *Poekilocerus* and *Phymateus* share a unique two amino acid insertion between residues 119 and 120 of the ancestral protein (figure 2). Insertions or deletions have not been observed so far among the other five insect orders surveyed to date (figure 3; electronic supplementary material, figure S2). Furthermore, *Phymateus leprosus* harbours an additional amino acid substitution (D120E) that also appears to be unique among Apocynaceae-specialists. Notably, no unique substitutions were observed among the other 12 Orthopteran species surveyed that appear to retain a single copy of ATPα1.

We also find that the functionally diverged duplicates of ATPα1 in the Orthopteran *Poekilocerus pictus* are differentially expressed in a similar manner to those found in Lygaeid bugs (see Methods) and other Apocynaceae-specialists [30]. Specifically, the putatively less cardenolide-sensitive copy (ATPα1A) is upregulated relative to the more sensitive copy (ATPα1B) in the gut compared to muscle and head (figure 4). The cardenolide-sensitive copy ATPα1B accounts for 18.7% of total ATPα1 expression in the brain but only 2.2% in the foregut, which is the primary location of cardenolide processing. This finding is consistent with previous studies of the related *Poekilocerus* species, *P. bufonius*, that noted Na^+,K^+-ATPase activity of the brain was more sensitive to cardenolide-inhibition than that of the gut [45]. Because the duplication of ATPα1 appears to predate the diversification of *Poekilocerus*, it is highly likely that the tissue-specific ouabain-sensitivity can be explained by the presence of the same duplication we have documented here.

Having established tissue-specific expression of duplicate copies in Lygaeids (see Methods) and Orthopterans, we then re-visited the pattern of unique versus parallel substitution with respect to duplication status and differential expression of ATPα1 in the full dataset now spanning six insect orders (figure 5). Consistent with Zhen *et al.* [30], we find a marked enrichment of unique substitutions at sites implicated in cardenolide-sensitivity in Apocynaceae-specialists with duplicated copies of ATPα1 (Fisher's exact

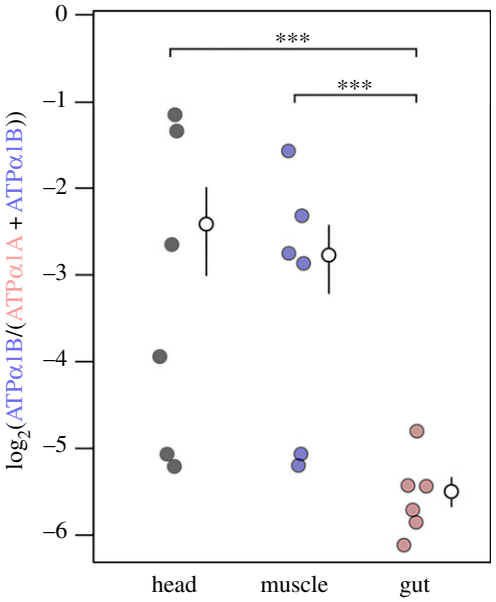

Figure 4. RNAseq-based estimates of ATPα1 expression by tissue in the duplication-bearing Orthopteran *Poekilocerus pictus*. As observed in other Apocynaceae-specialists [29], the relatively more cardenolide-sensitive ATPα1B is significantly less expressed in the gut than in head (corrected $p = 8 \times 10^{-4}$) and muscle (corrected $p = 9 \times 10^{-4}$) in *P. pictus* (electronic supplementary material, table S3). The mean proportion of ATPα1B of total ATPα1 is indicated with open circles with two standard errors as whiskers. *p*-values were estimated using the 'inverted beta-binomial' test for paired sample count data (see Methods) ***$p < 0.001$.

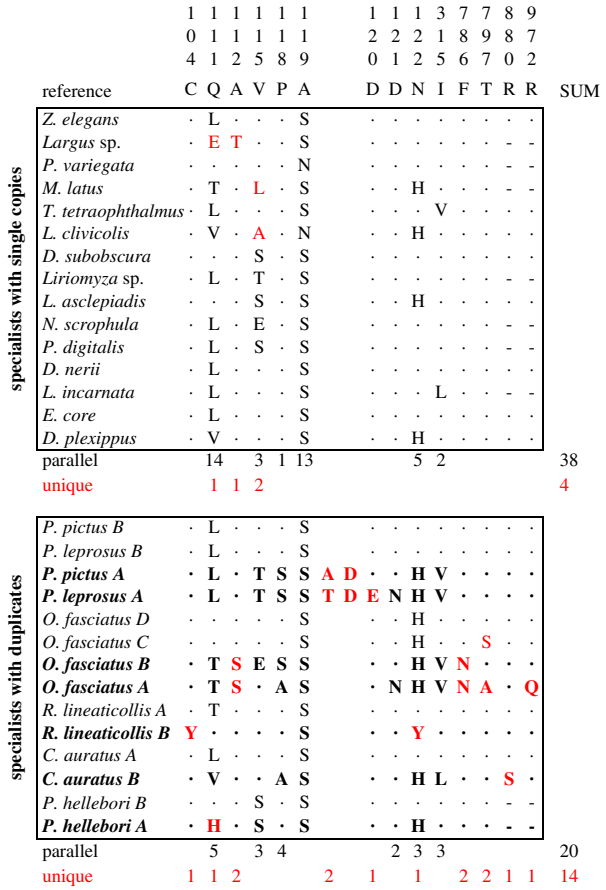

Figure 5. The pattern of parallel versus unique substitutions at sites implicated in cardenolide-sensitivity with respect to duplication status of ATPα1. Unique substitutions are indicated in red. Relatively less cardenolide-sensitive duplicate copies of ATPα1 are highlighted in bold. Site 874 was excluded from this analysis owing to uncertainty in the reconstruction of ancestral states.

test $p = 0.0022$). Examining the pattern in more detail, it is apparent that unique substitutions appear to be unequally distributed among copies. In each case of duplication, we can distinguish between less-sensitive and more-sensitive copies based on the number of derived amino acid-substitutions that have been implicated in cardenolide-sensitivity. We find a significant enrichment of unique substitutions in the dataset (13 out of 14) occur on putatively less-sensitive copies of ATPα1 that are substantially more expressed in the gut than more-sensitive copies (in the 4 out of 5 cases where expression patterns have been investigated).

(d) Relaxed constraints on ATPα1 duplicates and positive selection for insensitivity

It is clear from the above analyses that the evolution of cardenolide-insensitivity in some taxa is facilitated by duplication and differential expression of ATPα1, which is expected to relax constraints at sites known to confer insensitivity but are associated with negative pleiotropic effects. We further carried out a phylogenetic analysis to ask (i) whether this relaxation in constraint extends beyond sites directly implicated in cardenolide insensitivity, and (ii) whether there is evidence for positive selection associated with Apocynaceae-specialization at these and other sites in the protein. To examine patterns of constraint in more detail, we grouped both external and internal branches of the ATPα1 phylogeny into four categories: outgroup; Apocynaceae-specialist lineages with a single ATPα1 copy (single); and Apocynaceae-specialist lineages harbouring duplications that are inferred to be either relatively sensitive (Dup^S), or relatively insensitive (Dup^I) to cardenolides (see Methods). Examining the distributions of omega (dN/dS) estimates among these classes reveals that putatively less sensitive copies of ATPα1 (Dup^I) evolve approximately fivefold faster than their more sensitive counterparts (figure 6). We find that this pattern persists if we exclude the sites directly implicated in cardenolide-sensitivity (electronic supplementary material, figure S4), implying that relaxation of constraint on derived copies extends beyond this class of sites in the protein.

We next asked whether relaxed constraint in Apocynaceae-feeding lineages, or on insensitive duplicate ATPα1 lineages, is sufficient to account for the data or whether there is evidence for positive selection associated with insensitivity-conferring amino acid substitutions. We conducted a site-specific scan for positively selected substitutions using the improved branch site model implemented in PAML (see Methods). Site 111, a site directly implicated in cardenolide insensitivity and the target of frequent parallel amino acid substitution across insect orders, is identified as positively selected in this analysis under models assuming either positive selection in Apocynaceae-specialists, or only on Apocynaceae-specialist lineages bearing a single copy of ATPα1 (BEB posterior probability ≈ 1.0 under both models, figure 7, electronic supplementary material, table S4). Sites 115, 118 and 122, also directly implicated in cardenolide-sensitivity, are identified as positively selected under models of positive selection on the Dup^I lineages of ATPα1 (BEB posterior probabilities of 0.956, 0.998 and 0.993, respectively, figure 7). Interestingly, evidence for positive selection also emerges at several sites in the protein not previously implicated

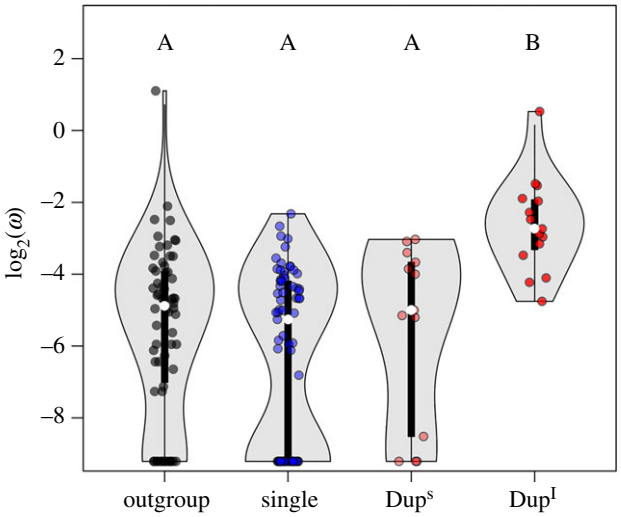

Figure 6. The distributions of omega (d*N*/d*S*) estimates for ATPα1 of non-specialists (outgroup), Apocynaceae-specialists with a single copy of ATPα1 (single), and specialists with duplicated ATPα1, where the relatively cardenolide sensitive and insensitive copies are noted as Dup^S and Dup^I, respectively. The median omega values are indicated with an open circle and bars represent 50% quantiles. There is a significant difference between the omega distributions for Dup^I and those of all three other groups (letters A and B indicate significantly different categories; Dup^I versus outgroup $p = 2 \times 10^{-5}$, Dup^I versus single $p = 3 \times 10^{-7}$, Dup^I versus Dup^S $p = 3 \times 10^{-3}$). *p*-values were estimated using Dunn's test of multiple comparisons using rank sums as implement in R (dunn.test) and adjusted using the Benjamini–Hochberg method.

in cardenolide-sensitivity. Seven sites show evidence for positive selection in lineages bearing duplicate copies of ATPα1 including a cluster of three sites (560, 563 and 566) that are located far from the cardenolide-binding domain (electronic supplementary material, table S4 and figure S1).

4. Discussion

Predictability in evolutionary biology not only refers to being able to forecast future evolutionary events but is also a statement about the ability to predict the genetic basis of adaptations outside a taxonomic group in which the rules governing the genetic basis of a particular adaptation were deduced. Previous surveys of ATPα1 in the context of insect adaptation to cardenolides established that, despite a reasonably large target size for evolving target-site insensitivity (i.e. 41 sites implicated in cardenolide-sensitivity, electronic supplementary material, table S1), a small proportion of these sites (and sites 111 and 122 in particular) are disproportionately used in the evolution of insensitivity to cardenolides [30]. The exception to this pattern is found in insects that have duplicated ATPα1 and differentially allocate a functionally-diverged copy to the gut, while retaining an ancestral copy that is more highly expressed in the head.

To test the generality of these patterns as predictors, we surveyed aposematic grasshoppers that sequester cardenolides belonging to Orthoptera, which is the first Polyneopteran order to be examined for cardenolide insensitivity (figure 1). Of the 10 species of Pyrgomorphidae included in the analysis, six species have been reported to feed on Apocynaceae and exhibit aposematic coloration and chemical defence (figure 2). The remaining four species

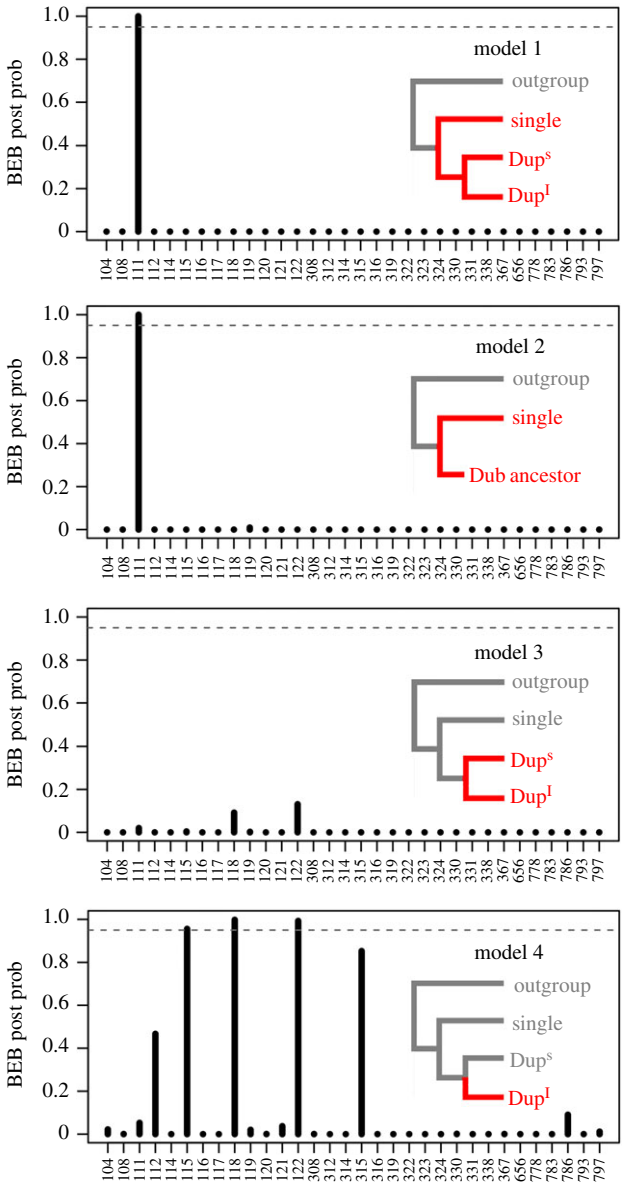

Figure 7. Positively selected sites of ATPα1 among 41 sites implicated in cardenolide-sensitivity (electronic supplementary material, table S1). Sites 802–982 were excluded from the analysis owing to missing data. Models 1–4 were tested to identify sites experiencing lineage-specific positive selection. The schematic diagram of each model is shown to the right in each panel. Foreground lineages where positive selection took place are coloured in red and corresponding background lineages are grey. BEB posterior probability greater than 0.95 (grey dashed line) is considered to be strong evidence for positive selection. See the electronic supplementary material, table S5 for the list of positively selected sites across the whole ATPα1 protein under each model.

feed on non-toxic plants and show cryptic coloration. While we initially expected all Apocynaceae-feeding species to show substitutions at sites implicated in cardenolide-insensitivity, we found that most Orthopteran species harbour only two substitutions (Q111L and A119S), that are not correlated with consumption and/or sequestration of cardenolide-containing plants in this group. The specific substitution Q111L appears to only weakly confer insensitivity to cardenolides based on previous functional studies. Furthermore, A119S does not interact with cardenolides directly or have an effect on cardenolide-affinity, although it does contribute to more rapid association/dissociation kinetics of the cardenolide ouabain [67]. Given this and their broader

phylogenetic distribution (figure 3; electronic supplementary material, figure S2), it is possible that these ancient and recurrent substitutions represent exaptations [68] that facilitate the evolution of more resistant forms of ATPα1 in some insect lineages [29].

Nevertheless, two of the six species (*Poekilocerus* and *Phymateus*) do share a duplication (ATPα1A) that exhibits multiple amino acid substitutions known to confer insensitivity to cardenolides. Why amino acid substitutions with large effects on cardenolide sensitivity or insensitive duplicates of ATPα1 are not found in the other four Apocynaceae-feeding species deserves an explanation. Of the six Apocynaceae-feeding species included in this study, *Phymateus*, *Poekilocerus* and *Zonocerus* have the mid-dorsal abdominal glands used for chemical defence [41] and they form a monophyletic group within the phylogeny of Pyrgomorphidae [37]. So far, cardenolides have been reported in the defence secretion from only *Phymateus* and *Poekilocerus* [41], and their common names, milkweed grasshoppers, suggest a strong association of these insects with the host plant. Although *Zonocerus* do feed on cardenolide-containing plants, their main defence chemical is pyrrolizidine alkaloid [69–71], rather than cardenolides. *Zonocerus variegatus* only excretes cardenolides when the toxins are included in its diet [72]. The other three Apocynaceae-feeding species (*Aularches*, *Dictyophorus* and *Taphronota*) do not have specialized abdominal glands, but produce toxic foams through various pores on the thorax and abdomen [41]. It is unknown whether they use cardenolides as their main chemical defence. The lack of cardenolide insensitivity of Na^+,K^+-ATPase in *Zonocerus*, *Aularches*, *Dictyophorus*, and *Taphronota* seems to suggest that they rely on other unknown mechanisms to cope with the toxicity of Apocynaceae. It is possible that while they are capable of consuming Apocynaceae, they might prefer a mixture of other toxic plants containing different kinds of secondary compounds to confer toxicity. Though the chemical ecology of these insects is under-studied, we can postulate that, at least in Pyrgomorphidae, only those species that are intimately associated with Apocynaceae have evolved the cardenolide insensitivity of Na^+,K^+-ATPase.

Considering the full dataset now including data from six insect orders, our study confirms previous findings [30] suggesting that, when ATPα1 has been duplicated, unique substitutions at known functionally important sites are significantly enriched specifically among insensitive copies (figure 5). This pattern of relaxation of constraint associated with one copy is also apparent at sites outside of those known to be functionally important sites (electronic supplementary material, figure S4). Both patterns are a strong indication of neo-functionalization rather than merely sub-functionalization of ATPα1. This may be expected given the age of the duplication event. Sub-functionalization is expected to occur rapidly after duplication while neo-functionalization takes place over longer timescales because it generally requires more mutations [73]. Most of the duplications of ATPα1 in Apocynaceae-specialist insects detected so far are indeed ancient and trans-specific (figure 3).

Despite the expected signature of strong purifying selection on ATPα1 (figure 7; electronic supplementary material, table S4), we have also detected the signature of positive selection on sites implicated in cardenolide-insensitivity and exhibiting frequent parallel substitution. Zou & Zhang [74] have pointed out that the observation of parallel substitution *per se* is not sufficient to infer adaptive significance as this pattern may be expected under a neutral model owing to among-site differences in physico-chemical constraints. Some evidence against this argument in the case of the evolution of ATPα1 target-site insensitivity is provided by the fact that Apocynaceae-specialist species are highly enriched for substitutions at sites implicated in cardenolide-sensitivity compared to non- specialist outgroups ([30]; this study). Our finding of positive selection at some of these sites establishes a more direct link between adaptive protein evolution and recurrent parallel substitutions at sites implicated in cardenolide-sensitivity.

Interestingly, we have also detected signatures of positive selection at sites not previously implicated in cardenolide-sensitivity (electronic supplementary material, table S4 and figure S1). Some of these sites (e.g. 301, 560, 563, 566 and 667) are located far from the cardenolide-binding domain of the enzyme and sites 560, 563 and 566 are known to have roles in binding ATP. While this might seem to preclude direct roles in cardenolide-insensitivity, the existence of sites exhibiting allosteric effects on sensitivity is not unprecedented (e.g. 367 and 656, electronic supplementary material, table S1, [75]). This being said, the detection of positive selection at these sites may reflect selection pressures that are either only indirectly related or even unrelated to the evolution of cardenolide-insensitivity. Future functional experiments could be aimed at understanding the effects of these substitutions on cardenolide-insensitivity and overall enzyme performance.

Duplications of ATPα1 feature prominently in the evolution of cardenolide-insensitivity in herbivorous insects. Given this feature, it will be interesting to compare patterns of recurrent parallel amino acid substitution in insects with vertebrates. Bufonid toads are among the few animals able to produce Na^+,K^+-ATPase-inhibiting compounds called 'bufadienolides' that closely resemble cardenolides and act in the same way [76]. As a result, predators of bufonid toads represented by a wide variety of vertebrates are under pressure to evolve insensitivity to these compounds [77,78]. In contrast to insects, vertebrates retain at least three copies of ATPα, (ATP1A1, ATP1A2 and ATP1A3), that are differentially expressed among tissues [79]. Previous studies have so far investigated the H1-H2 extracellular loop of ATP1A1 in bufonid toads and predatory frogs [77] and ATP1A3 of a number of predatory squamates [78,80]. It will be of considerable interest to further compare patterns of molecular evolution of cardenolide-sequestering insects to their bufonid predator analogues in the context of complete reconstructions of all three proteins.

Data accessibility. All raw RNA-seq sequence data generated for this study have been deposited in the National Center for Biotechnology Information Short Read Archive, www.ncbi.nlm.nih.gov/sra (BioProject PRJNA509040). Sequences of ATPα1 used in our analysis have been submitted to GenBank under Accession numbers MK294065-81, MK765670-72.

Authors' contributions. L.Y. and P.A. designed the study; N.R., R.D. and K.K. Collected and prepared RNA for Indian specimens; R.M.-P. and H.S. collected and provided tissue for African and South American Specimens; L.Y. performed the experimental work and analysed the data; M.W. contributed to electronic supplementary material, figure S3; A.R. contributed to electronic supplementary material, figure S2; L.Y. and P.A. wrote the manuscript; H.S. and R.M.-P. reviewed and edited.

Competing interests. We declare we have no competing interests.

Funding. This research was funded by NIH grant no. R01-GM115523 to P.A. Specimen collection and wet laboratory work in India was funded by a research grant from NCBS to K.K. Specimen collection in South Africa was funded by NSF grant no. DEB-1655097 to H.S.

Acknowledgements. We thank Patrick Reilly for help with *de novo* assemblies of transcriptomes and reconstruction of ATPα1. We thank Shurong Hou for constructive comments on protein structure inferences. We thank Piotr Naskrecki (Gorongosa Restoration Project, Mozambique) for help in sample collection.

References

1. Stern D. 2010 *Evolution, development, and the predictable genome*. Greenwood Village, CO: Roberts & Co.
2. Orr AH. 2005 The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* **6**, 119. (doi:10.1038/nrg1523)
3. Smith JM. 1970 Natural selection and the concept of a protein space. *Nature* **225**, 565–566. (doi:10.1038/225563a0)
4. Losos JB. 2011 Convergence, adaptation, and constraint. *Evolution* **65**, 1827–1840. (doi:10.1111/j.1558-5646.2011.01289.x)
5. Leroi A, Rose M, Lauder G. 1994 What does the comparative method reveal about adaptation? *Am. Nat.* **143**, 381–402. (doi:10.1086/285609)
6. Yukilevich R, Lachance J, Aoki F, True JR. 2008 Long-term adaptation of epistatic genetic networks. *Evolution* **62**, 2215–2235. (doi:10.1111/j.1558-5646.2008.00445.x)
7. Rosenblum E, Parent CE, Brandt EE. 2014 The molecular basis of phenotypic convergence. *Annu. Rev. Ecol. Evol. Syst.* **45**, 1–24. (doi:10.1146/annurev-ecolsys-120213-091851)
8. Hoekstra H, Coyne J. 2007 The locus of evolution: evo-devo and the genetics of adaptation. *Evolution* **61**, 995–1016. (doi:10.1111/j.1558-5646.2007.00105.x)
9. Stern D, Orgogozo V. 2008 The loci of evolution: how predictable is genetic evolution? *Evolution* **62**, 2155–2177. (doi:10.1111/j.1558-5646.2008.00450.x)
10. Streisfeld MA, Rausher MD. 2011 Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution. *Evolution* **65**, 629–642. (doi:10.1111/j.1558-5646.2010.01165.x)
11. Stern DL. 2013 The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764. (doi:10.1038/nrg3483)
12. Storz J. 2016 Causes of molecular convergence and parallelism in protein evolution. *Nat. Rev. Genet.* **17**, 239–250. (doi:10.1038/nrg.2016.11)
13. Mithöfer A, Boland W. 2012 Plant defense against herbivores: chemical aspects. *Annu. Rev. Plant Biol.* **63**, 431–450. (doi:10.1146/annurev-arplant-042110-103854)
14. Heckel DG. 2014 Insect detoxification and sequestration strategies. *Annual Plant Reviews*. **47**, 77–114. (doi:10.1002/9781118829783.ch3)
15. Dobler S, Petschenka G, Pankoke H. 2011 Coping with toxic plant compounds—the insect's perspective on iridoid glycosides and cardenolides. *Phytochemistry* **72**, 1593–1604. (doi:10.1016/j.phytochem.2011.04.015)
16. Erb M, Robert CA. 2016 Sequestration of plant secondary metabolites by insect herbivores: molecular mechanisms and ecological consequences. *Curr. Opin. Insect Sci.* **14**, 8–11. (doi:10.1016/j.cois.2015.11.005)
17. Lingrel JB. 2010 The physiological significance of the cardiotonic steroid/ouabain-binding site of the Na,K-ATPase. *Annu. Rev. Physiol.* **72**, 395–412. (doi:10.1146/annurev-physiol-021909-135725)
18. Schoner W. 2002 Endogenous cardiac glycosides, a new class of steroid hormones. *Eur. J. Biochem.* **269**, 2440–2448. (doi:10.1046/j.1432-1033.2002.02911.x)
19. Prassas I, Diamandis EP. 2008 Novel therapeutic applications of cardiac glycosides. *Nat. Rev. Drug Discov.* **7**, nrd2682. (doi:10.1038/nrd2682)
20. Blaustein MP. 2017 The pump, the exchanger and the holy spirit: origins of the endogenous ouabain-hypertension hypothesis and its 40 year evolution. *Am. J. Physiol. Cell Physiol.* **314**, ajpcell.00196. (doi:10.1152/ajpcell.00196.2017)
21. Lingrel J, Kuntzweiler T. 1994 Na^+,K^+-ATPase. *J. Biol. Chem.* **269**, 19 659–19 662.
22. Schuler MA. 2011 P450s in plant–insect interactions. *Biochim. Biophys. Acta Proteins Proteom.* **1814**, 36–45. (doi:10.1016/j.bbapap.2010.09.012)
23. Groen SC, LaPlante ER, Alexandre NM, Agrawal AA, Dobler S, Whiteman NK. 2017 Multidrug transporters and organic anion transporting polypeptides protect insects against the toxic effects of cardenolides. *Insect Biochem. Mol. Biol.* **81**, 51–61. (doi:10.1016/j.ibmb.2016.12.008)
24. Després L, David J-P, Gallet C. 2007 The evolutionary ecology of insect resistance to plant chemicals. *Trends Ecol. Evol.* **22**, 298–307. (doi:10.1016/j.tree.2007.02.010)
25. Jungreis AM, Vaughan GL. 1977 Insensitivity of lepidopteran tissues to ouabain: absence of ouabain binding and Na^+,K^+-ATPases in larval and adult midgut. *J. Insect Physiol.* **23**, 503–509. (doi:10.1016/0022-1910(77)90261-x)
26. Petschenka G, Pick C, Wagschal V, Dobler S. 2013 Functional evidence for physiological mechanisms to circumvent neurotoxicity of cardenolides in an adapted and a non-adapted hawk-moth species. *Proc. R. Soc. B* **280**, 20123089. (doi:10.1098/rspb.2012.3089)
27. Holzinger F, Frick C, Wink M. 1992 Molecular basis for the insensitivity of the Monarch (*Danaus plexippus*) to cardiac glycosides. *FEBS Lett.* **314**, 477–480. (doi:10.1016/0014-5793(92)81530-Y)
28. Labeyrie E, Dobler S. 2004 Molecular adaptation of *Chrysochus* leaf beetles to toxic compounds in their food plants. *Mol. Biol. Evol.* **21**, 218–221. (doi:10.1093/molbev/msg240)
29. Aardema ML, Zhen Y, Andolfatto P. 2012 The evolution of cardenolide-resistant forms of Na^+,K^+-ATPase in Danainae butterflies. *Mol. Ecol.* **21**, 340–349. (doi:10.1111/j.1365-294X.2011.05379.x)
30. Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P. 2012 Parallel molecular evolution in a herbivore community. *Science* **337**, 1634–1637. (doi:10.1126/science.1226630)
31. Dobler S, Dalla S, Wagschal V, Agrawal AA. 2012 Community-wide convergent evolution in insect adaptation to toxic cardenolides by substitutions in the Na,K-ATPase. *Proc. Natl Acad. Sci. USA* **109**, 13 040–13 045. (doi:10.1073/pnas.1202111109)
32. Dobler S, Petschenka G, Wagschal V, Flacht L. 2015 Convergent adaptive evolution: how insects master the challenge of cardiac glycoside-containing host plants. *Entomol. Exp. Appl.* **157**, 30–39. (doi:10.1111/eea.12340)
33. Petschenka G, Wagschal V, von Tschirnhaus M, Donath A, Dobler S. 2017 Convergently evolved toxic secondary metabolites in plants drive the parallel molecular evolution of insect resistance. *Am. Nat.* **190**, S29–S43. (doi:10.1086/691711)
34. Dalla S, Dobler S. 2016 Gene duplications circumvent trade-offs in enzyme function: insect adaptation to toxic host plants. *Evolution* **70**, 2767–2777. (doi:10.1111/evo.13077)
35. Dalla S, Baum M, 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. (doi:10.1016/j.ibmb.2017.08.005)
36. Morris DH, Gostic KM, Pompei S, Bedford T, Łuksza M, Neher RA, Grenfell BT, Lässig M, McCauley JW. 2018 Predictive modeling of influenza shows the promise of applied evolutionary biology. *Trends Microbiol.* **26**, 102–118. (doi:10.1016/j.tim.2017.09.004)
37. Mariño-Pérez R, Song H. 2017 Phylogeny of the grasshopper family Pyrgomorphidae (Caelifera, Orthoptera) based on morphology. *Syst. Entomol.* **43**, 90–108. (doi:10.1111/syen.12251)
38. Euw VJ, Fishelson L, Parsons J, Reichstein T, Rothschild M. 1967 Cardenolides (heart poisons) in a grasshopper feeding on milkweeds. *Nature* **214**, 35–39. (doi:10.1038/214035a0)
39. Pugalenthi P, Livingstone D. 1995 Cardenolides (heart poisons) in the painted grasshopper *Poecilocerus pictus* F. (Orthoptera: Pyrgomorphidae) feeding on the milkweed *Calotropis gigantea*

L. (Asclepiadaceae). *J. N. Y. Entomol. Soc.* **103**, 191–196.

40. Seibt U, Kasang G, Wickler W. 2000 Suggested pharmacophagy of the African bushhopper *Phymateus leprosus* (Fabricius) (Pyrgomorphidae, Orthoptera). *Z. Naturforsch. C J. Biosci.* **55**, 442–448. (doi:10.1515/znc-2000-5-621)
41. Whitman DW. 1990 Grasshopper chemical communication. In *Biology of grasshoppers* (eds RF Chapman, A Joern), pp. 357–391. New York, NY: John Wiley & Sons.
42. Rowell FC. 1967 Experiments on aggregations of *Phymateus purpurascens* (Orthoptera, Acrididae, Pyrgomorphinae). *J. Zool.* **152**, 179–193. (doi:10.1111/j.1469-7998.1967.tb01884.x)
43. Chapman R, Page W, McCaffery A. 1986 Bionomics of the variegated grasshopper (*Zonocerus variegatus*) in West and Central Africa. *Annu. Rev. Entomol.* **31**, 479–505. (doi:10.1146/annurev.en.31.010186.002403)
44. Rentz DC, Lewis RC, Su YN, Upton MS. 2003 *A guide to Australian grasshoppers and locusts*. Borneo, Malaysia: Natural History Publications.
45. Al-Robai AA. 1993 Different ouabain sensitivities of Na^+/K^+-ATPase from *Poekilocerus bufonius* tissues and a possible physiological cost. *Comp. Biochem. Physiol. B* **106**, 805–812. (doi:10.1016/0305-0491(93)90034-3)
46. Haas BJ *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. (doi:10.1038/nprot.2013.084)
47. Wang X *et al.* 2014 The locust genome provides insight into swarm formation and long-distance flight. *Nat. Commun.* **5**, 2957. (doi:10.1038/ncomms3957)
48. Blankers T, Oh K, Genes SK. 2018 The genetics of a behavioral speciation phenotype in an island system. *Genes* **9**, 346. (doi:10.3390/genes9070346)
49. Zhao L, Zhang X, Qiu Z, Huang Y. 2018 *De novo* assembly and characterization of the *Xenocatantops brachycerus* transcriptome. *Int. J. Mol. Sci.* **19**, 520. (doi:10.3390/ijms19020520)
50. Picelli S, Björklund ÅK, Reinius B, Sagasser S, Winberg G, Sandberg R. 2014 Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040. (doi:10.1101/gr.177881.114)
51. Zerbino DR, Birney E. 2008 Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829. (doi:10.1101/gr.074492.107)
52. Schulz MH, Zerbino DR, Vingron M, Birney E. 2012 Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086–1092. (doi:10.1093/bioinformatics/bts094)
53. Li H. 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint. See https://arxiv.org/abs/1303.3997.
54. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079. (doi:10.1093/bioinformatics/btp352)
55. Anders S, Pyl P, Huber W. 2015 HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169. (doi:10.1093/bioinformatics/btu638)
56. Pham TV, Jimenez CR. 2012 An accurate paired sample test for count data. *Bioinformatics* **28**, i596–i602. (doi:10.1093/bioinformatics/bts394)
57. Panfilio K *et al.* 2018 Molecular evolutionary trends and feeding ecology diversification in the Hemiptera, anchored by the milkweed bug genome. *Biorxiv*, 201731. (doi:10.1101/201731)
58. Yang Z. 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556. (doi:10.1093/bioinformatics/13.5.555)
59. Scheffer SJ, Winkler IS, Wiegmann BM. 2007 Phylogenetic relationships within the leaf-mining flies (Diptera: Agromyzidae) inferred from sequence data from multiple genes. *Mol. Phylogenet. Evol.* **42**, 756–775. (doi:10.1016/j.ympev.2006.12.018)
60. Yang Z, Nielsen R. 2002 Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**, 908–917. (doi:10.1093/oxfordjournals.molbev.a004148)
61. Yang Z, Wong W, Nielsen R. 2005 Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**, 1107–1118. (doi:10.1093/molbev/msi097)
62. Zhang J, Nielsen R, Yang Z. 2005 Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **22**, 2472–2479. (doi:10.1093/molbev/msi237)
63. Pegueroles C, Ferrés-Coy A, Martí-Solano M, Aquadro CF, Pascual M, Mestres F. 2016 Inversions and adaptation to the plant toxin ouabain shape DNA sequence variation within and between chromosomal inversions of *Drosophila subobscura*. *Sci. Rep.* **6**, 23754. (doi:10.1038/srep23754)
64. Dalla S, Swarts HG, Koenderink JB, Dobler S. 2013 Amino acid substitutions of Na, K-ATPase conferring decreased sensitivity to cardenolides in insects compared to mammals. *Insect Biochem. Mol. Biol.* **43**, 1109–1115. (doi:10.1016/j.ibmb.2013.09.006)
65. Price EM, Rice DA, Lingrel JB. 1989 Site-directed mutagenesis of a conserved, extracellular aspartic acid residue affects the ouabain sensitivity of sheep Na,K-ATPase. *J. Biol. Chem.* **264**, 21 902–21 906.
66. Freeman MR, Doherty J. 2006 Glial cell biology in *Drosophila* and vertebrates. *Trends Neurosci.* **29**, 82–90. (doi:10.1016/j.tins.2005.12.002)
67. Crambert G, Schaer D, Roy S, Geering K. 2004 New molecular determinants controlling the accessibility of ouabain to its binding site in human Na,K-ATPase alpha isoforms. *Mol. Pharmacol.* **65**, 335–341. (doi:10.1124/mol.65.2.335)
68. Gould S, Vrba E. 1982 Exaptation—a missing term in the science of form. *Paleobiology* **8**, 4–15. (doi:10.1017/S0094837300004310)
69. Ananthakrishnan TN. 1994 Profiles of insect diversity. *Curr. Sci.* **66**, 271–281.
70. Bernays E, Chapman R, Leather E, McCaffery A, Modder W. 1977 The relationship of *Zonocerus variegatus* (L.) (Acridoidea: Pyrgomorphidae) with cassava (*Manihot esculenta*). *Bull. Entomol. Res.* **67**, 391–404. (doi:10.1017/S0007485300011202)
71. Bernays E, Edgar JA, Rothschild M. 1977 Pyrrolizidine alkaloids sequestered and stored by the aposematic grasshopper, *Zonocerus variegatus*. *J. Zool.* **182**, 85–87. (doi:10.1111/j.1469-7998.1977.tb04142.x)
72. Rafaeli-Bestein A, Mordue W. 1978 The transport of the cardiac glycoside ouabain by the Malpighian tubules of *Zonocerus variegatus*. *Physiol. Entomol.* **3**, 59–63. (doi:10.1111/j.1365-3032.1978.tb00133.x)
73. He X, Zhang J. 2005 Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**, 1157–1164. (doi:10.1534/genetics.104.037051)
74. Zou Z, Zhang J. 2015 Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations? *Mol. Biol. Evol.* **32**, 2085–2096. (doi:10.1093/molbev/msv091)
75. Kirley TL, Peng M. 1991 Identification of cysteine residues in lamb kidney (Na,K)-ATPase essential for ouabain binding. *J. Biol. Chem.* **266**, 19 953–19 957.
76. Krenn L, Kopp B. 1998 Bufadienolides from animal and plant sources. *Phytochemistry* **48**, 1–29. (doi:10.1016/S0031-9422(97)00426-3)
77. Moore DJ, Halliday DC, Rowell DM, Robinson AJ, Keogh J. 2009 Positive Darwinian selection results in resistance to cardioactive toxins in true toads (Anura: Bufonidae). *Biol. Lett.* **5**, 513–516. (doi:10.1098/rsbl.2009.0281)
78. Ujvari B *et al.* 2015 Widespread convergence in toxin resistance by predictable molecular evolution. *Proc. Natl Acad. Sci. USA* **112**, 11 911–11 916. (doi:10.1073/pnas.1511706112)
79. Herrera V, Emanuel J, Ruiz-Opazo N, Levenson R, Nadal-Ginard B. 1987 Three differentially expressed Na,K-ATPase alpha subunit isoforms: structural and functional implications. *J. Cell Biol.* **105**, 1855–1865. (doi:10.1083/jcb.105.4.1855)
80. Ujvari B, Mun HC, Conigrave AD, Bray A, Osterkamp J, Halling P, 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. (doi:10.1111/j.1558-5646.2012.01751.x)