Extended Materials and Methods:

Acoustic data

We collected a total of 16,657 advertisement calls from 351 individual recordings of poison frogs from 172 species ($\bar{X}$ = 2.04 ± 1.33 SD recordings per species; range 1–7). Of the species sampled, 127 are described, 19 are undescribed, and 26 are cryptic species from within 6–10 species complexes. These cryptic species are acoustically and genetically distinct. All genera and major lineages of poison frogs are represented by at least one recording per species (Dataset S1, Table S1). All recordings were obtained from field collections, museum archives, and detailed published acoustic descriptions. Our Dataset S1 and Table S1 also include information about locality of collection, geographic coordinates, branch tip number assigned to each taxon in the chronogram (Phy ID; Fig. S1), calling behavior (perching site), recording distance from the vocalizing male, hour of collection, recording temperature, voucher identification number, snout-vent length (SVL), oscillograms, spectrograms, power spectra, collectors, published references, and recording equipment.

Some recordings lacked data on temperature (11 taxa) or SVL of the recorded individual (64 taxa) because field notes did not provide recording temperatures or calling animals were not collected (Dataset S1). We used proxies for each variable. For temperature, the geographic coordinates of the recording collection site were used to obtain the corresponding annual mean temperature from the WorldClim database [1]. We considered this approximation adequate for poison frogs by the following reasons. First, most dendrobatid species call during the day and they are vocally active though the year [2]. Second, all recordings with missing temperature data are from Neotropical lowlands (e.g., Chocó and Amazon Basin). These regions have relatively stable (e.g., ± 2 °C) daily and annual mean temperatures [3]. Likewise, among all call variables only temporal features (e.g., pulse rate and duration) are likely to correlate positively with body temperature [4]. Finally, the best estimate of body temperature is from within the animal (e.g., cloacal readings) and all other environmental measurements (e.g., air or substrate temperatures) including our proxy temperature should be considered moderate approximations [5]. Most dendrobatids will probably call at some optimal temperature (microclimate) and this value is likely correlated to the broad-scale climate [2]. For body

size, the mean SVL of male conspecifics was used as proxy. As much as possible, we used measurements from syntopic males collected with the recorded individual.

All call recordings were digitized using Adobe Audition v 4.0 [6] with a sample size of 16 bits at a rate of 22 or 44 kHz into high resolution digital files (.wav or .aiff). Some recordings were filtered from background noise to enable good contrast of each call. For this goal, we used a bandpass filtering approach between 1–5 kHz. Calls temporal and spectral features were measured using RavenPro v 1.4 [7]. Oscillograms were used to measure temporal features and spectrograms and power spectra were used to measure the spectral properties [8, 9]. Spectrograms and power spectra were estimated after a Fast Fourier Transform (FFT) using a Blackman window type, a length of 900 samples of overlap among subsequent FFTs, and 3-dB filter bandwidth of 87.5 Hz.

Acoustic variables

Poison frog vocalizations have relatively simple structure with pulse-like sound units, produced with variable repetition and with little frequency modulation (Dataset S1, Table S1). Consequently, we used a physiological definition of call homology as the unit of sound produced by single cycle of trunk muscles contraction resulting in an expiratory event [10, 11]. Likewise, we found that only note-pulses presented uniform temporal, spectral, and taxon-specific features [12, 13]. To ensure that only advertisement calls were included in our analyses, we explored the entire extent of our recordings and used only frog vocalizations that were given with high redundancy, similar intensity, and without agonistic interactions [14, 15].

We measured 29 acoustic parameters of the advertisement call (Fig. S2 and Table S1). These describe two main sets of call features that include temporal and spectral properties [4]. We analyzed the temporal domain parameters and found that poison frog calls range from single note to multinote calls [16]. For this reason we divided temporal properties into gross- and fine-scale/temporal acoustic variables. Call data of individuals of each taxon were averaged and mean values were used for subsequent comparative analyses.

Gross temporal variables included features quantifiable in multinote calls. These variables are (*i*) multinote call duration or the time in seconds from beginning to end of a

single multinote call; (*ii*) multinote call interval or the time in seconds between multinote calls measured from the start of the preceding multinote call to the beginning of the current multinote call; (*iii*) multinote call rise time or the time in seconds from multinote call onset to the pulse-note that was at the level of 75% of maximal amplitude; (*iv*) number of pulse-notes in a multinote call; (*v*) multinote call rate or the total number of multinote calls minus 1, divided by time in seconds between the start of the first multinote call to the beginning of the last multinote call; and (*vi*) pulse-note rate on a multinote call or the total number of notes minus 1, divided by the time in seconds between the start of the first pulse-note to the beginning of the last pulse-note in a continuous series of multinote calls.

Fine-scale temporal variables included call features present in both single and multinote calls. These are (*vii*) number of pulse-notes on a single unit of repetition (UR) as defined in Fig. S2; (*viii*) UR duration or the time in seconds from beginning to end of one UR; (*ix*) UR interval or the time in seconds measured from the start of the preceding UR to the beginning of the current UR; (*x*) UR rate or the total number of URs minus 1, divided by the time in seconds between the start of the first UR to the beginning of the last UR; (*xi*) number of pulse-notes at the beginning of the call at ≤ 75% of the call's maximal amplitude, in the case of single pulse-note call this value is always 1; (*xii*) initial pulse-note duration or the time in seconds from beginning to end of a initial pulse-note, in the case of single pulse-note calls this value is the same as the middle pulse-note estimate; (*xiii*) initial pulse-note interval or the time in seconds measured from the start of the preceding initial pulse-note to the beginning of the current pulse-note, in the case of single pulse-note calls this value is the same as the middle pulse-note estimate; (*xiv*) initial pulse-note rate or the total number of initial pulse-notes minus 1, divided by the time between the start of the first initial pulse-note to the beginning of last the initial pulse-notes; (*xv*) number of pulse-notes in the middle of the call or the number of notes with >75% of the maximum amplitude of the call; in the case of single pulse-note calls this is always 1; (*xvi*) middle pulse-note duration or the time in seconds between the beginning to end of a middle note; (*xvii*) middle pulse-note interval or the time in seconds between notes measured from the start of the preceding middle pulse-note to the beginning of the current middle pulse-note; (*xviii*) middle pulse-note rate or the total

number of middle pulse-notes minus 1, divided by the time in seconds between the start of the first middle pulse-note to the beginning of the last middle pulse-note; (*xix*) middle pulse-note rise time or the time in seconds from onset to the maximum amplitude of each middle pulse-note; (*xx*) middle pulse-note shape, which is a unitless index of the shape of the amplitude envelope derived by dividing middle pulse-note rise time by its duration; and (*xxi*) middle pulse-note duty, which is a unitless index of the proportion of time spent producing an acoustic signal, which is determined by dividing the pulse-note duration by the interval between two consecutive middle note-pulses.

Spectral and call frequency breadth variables are defined as follows: (*xxii*) first quartile frequency of the initial pulse-note, or initial Q_1, in Hz; in the case of single pulse-note calls this value is the same as the middle pulse-note estimate; (*xxiii*) initial pulse-note peak frequency in Hz; in the case of single pulse-note calls this value is the same as the middle pulse-note estimates; (*xxiv*) initial pulse-note third quartile frequency or initial Q_3, in Hz; in the case of single pulse-note calls this value is the same as the middle pulse-note estimate; (*xxv*) initial pulse-note frequency modulation is the Q_3 minus Q_1 frequency of the initial note-pulses, and is defined as the IQR (Inter-Quartile Range) bandwidth; in the case of single pulse-note calls this value is the same as the middle pulse-note estimate; (*xxvi*) middle pulse-note first quartile frequency or middle Q_1, in Hz; (*xxvii*) middle pulse-note peak frequency in Hz; (*xxviii*) middle pulse-note third quartile frequency or Q_3, in Hz; and (*xxix*) middle pulse-note frequency modulation is Q_3 minus Q_1 frequency of the middle note-pulses.

Alkaloid sequestration and conspicuousness variables

We compiled skin alkaloid information of 97 taxa from published accounts (Table S1 – S2). Species were characterized by their ability to sequester alkaloids as state 1 (able) or 0 (unable). All species scored as able to sequester alkaloids are taxa with lipophilic skin alkaloids in natural populations or experimentally demonstrated to sequester these alkaloids (e.g., after artificial feeding). However, we scored *Colostethus panamansis* as state 0 despite being defended with tetrodotoxin (TTX) [17]. TTX is a potent hydrophilic neurotoxin probably derived from bacterial symbionts [18], and its mechanism of acquisition is hypothesized to be different from the dietary arthropod

origin of the lipophilic alkaloids in the aposematic dendrobatids [19].

We used a binary assessment of poison frog conspicuousness derived from six color contrast thresholds against a leaf litter background. This methodology is based on the chromatic contrast of frog segments of color against their natural background (i.e., leaf litter) as seen by a mammal trichromat (i.e., humans). Three features describe human perception of color including uniqueness (hue), intensity or distinctiveness from gray (chroma), and luminosity (brightness). We define chromatic contrast as the human qualitative sensation of coloration (i.e., hue and chroma) differing from that of leaf litter. For a detailed discussion of the limitations of our conspicuousness assessment see the next section in the *Supporting Text*.

Our methodology for qualitative assessment of conspicuousness used a threshold framework. Discrete variables were defined is as follows: (*i*) color descriptions of live male specimens were derived from species descriptions, color photographs, and field notes by as many independent observers as possible ($\bar{X} = 2.8 \pm 1.17$ SD descriptions per taxa; range 1–6 observations; see Table S2); (*ii*) the frog's body was divided into 11 non-overlapping sections (patches) of skin covering dorsal, ventral, and lateral views; dorsal and lateral stripes; and flash markings in arms and legs (Fig. S3); (*iii*) a positive perception of chromatic contrast against natural background [20, 21] by human vision was scored as conspicuous or 1 for any body section with coloration that is different from leaf litter (i.e., gray, brown, and black) or 0, if otherwise; (*iv*) the total score of each taxa was determined by the sum of all binary values or total contrast score (TCS or ΣS_i) and ranged from 0 (all sections 0 or 0*11 = 0; no contrast) to 11 (all sections 1 or 1*11 = 11; maximum contrast); (*v*) given that our scoring might be different for non-human viewers [22-24] and changing incident light [25], we determined six increasing color contrast thresholds as measures of conspicuousness raging from liberal (any taxon with TCS3 or $\Sigma S_i \geq 3$ is conspicuous) to strict (only taxa with TCS8 or $\Sigma S_i \geq 8$ are conspicuous); we found too few (i.e., <10) species were classifiable under $\Sigma S_i \geq 1$, $\Sigma S_i \geq 2$, $\Sigma S_i \geq 9$, and $\Sigma S_i \geq 10$ so these thresholds were excluded; (*vi*) as a result, binary estimates of coloration contrast were determined per each taxon as proxy of its conspicuousness under six categorical variables (i.e., TCS3-TCS8). For example, one taxon with total score $\Sigma S_i = 7$

is considered cryptic under the threshold $\Sigma S_i \geq 8$ (TCS8); but it is conspicuous under the thresholds $\Sigma S_i \geq 3$, $\Sigma S_i \geq 4$, $\Sigma S_i \geq 5$, $\Sigma S_i \geq 6$, and $\Sigma S_i \geq 7$ (i.e., TCS3–TCS7).

To compare the aposematism score among conspicuousness variables, we used a joint criterion for binary classifiers [26] using alkaloid sequestration ability as an unbiased predictor of aposematism. For this purpose, we used indices of accuracy, precision, and sensitivity. The accuracy index is defined as the proportion of true results (true positives or conspicuous/able to sequester alkaloids and true negatives or cryptic/unable) among all species with scored ability or inability to sequester alkaloids. The precision index indicates the proportion of the true positives against all positive results regardless of these being correct (i.e., both true positives and false positives). Finally, the sensitivity index indicates the proportion of actual positives that were correctly identified (i.e., the percentage of conspicuous/able taxa who are correctly identified as such). The best-performing binary conspicuousness variables are those that have the combined highest value in all accuracy, precision, and sensitivity indices (i.e., all scores close to 1.00).

Limitations and rationale of quantifying conspicuousness of poison frogs from a human perspective

Several authors [22, 24, 27] have strongly criticized the assessment of coloration in animals based on human perception (i.e., a non-UV sensitive trichromat). We agree that these criticisms are valid if the only intended receivers have drastically different visual sensitivities than humans (e.g., tetrachromats as birds). However, skin coloration in poison frogs in the context of conspicuousness seems to be intended to multiple receivers [28-30]. Thus, we think that humans are valid assessors of poison frog conspicuousness for several reasons.

First, the importance of tetrachromat viewers (e.g., birds) as primary receivers of aposematic colors of poison frogs needs further confirmation. Visual signals in poison frogs seem to be intended for conspecifics [31, 32] as well as a mixture natural predators including and not limited to snakes, crabs, and birds [33]. Some authors have argued that poison frogs direct their visual warning signal mostly to birds based on a single species,

Dendrobates pumilio [28], but natural history evidence supports that avian predators account less than 3% of all documented predation events across the ~300 species in the dendrobatid lineage [33]. At this level of Dendrobatidae, the most important predators are snakes with 71% (25 of 35 reported events), followed by spiders 17% (6/35), and only 3% by birds (1/35).

Although these observations, in general, reflect the number of individual frogs found in the stomachs of predators, a key question is which of those predators are frequently attacking (i.e., sampling) but not necessarily eating poison frogs in nature. A three-year field study followed anuran predation by 91 species of Neotropical understory birds in central Panama [34]. These authors demonstrated neither a single event of predation (individuals found in birds' stomachs) nor reported observations of attack upon any cryptic or aposematic species of dendrobatid. Yet, some evidence suggests that birds account for a significant proportion of attacks (i.e., sampling) or intended predation upon clay models painted as poison frogs [35-40]. However, stationary clay models do not behave like living frogs (i.e., do not vocalize, move, taste, smell, and interact with the environment) and more likely behave like immobile brightly colored fruits or seeds. Only recently have clay model experiments incorporated motion in the assessment of poison frog predation [41]. This study showed a significant increase in avian and mammal predation in comparison with previous studies based only on stationary clay models. Not surprisingly, the introduction of motion to predation experiments shows the importance of movement for prey selection by visual guided predators. Other field studies showed that aposematic dendrobatids (e.g., *Dendrobates auratus* and *D. pumilio*) exhibit complex escape behavior, capacity to assess predation risk, and adjust their behavior to minimize predation attack [42, 43]. Thus, more sophisticated predation experiments that mimic living animals will provide information that is more relevant to predation dynamics.

Second, several studies have supported the conspicuousness of poison frogs for both non-human and human viewers based on chromatic conspicuousness [29, 30, 35-38, 42, 44-50]. These studies are supported by the fact that the actual predators of poison frogs are a mixture of viewers that includes dichromats (e.g., crabs), trichromats (e.g., snakes) and tetrachromats (e.g., birds). Likewise, these predators have visual sensitivities that completely or partially overlap the range of 400 – 700 nm of human vision [51, 52].

Thus, human assessment of conspicuousness is a rough, but likely a good, estimator of chromatic conspicuousness of poison frogs to at least a subset of predators.

Third, UV reflectance by the skin of some poison frogs seems to be absent [47] and receivers with UV sensitivity (e.g., tetrachromats such as birds) might not perceive new information at the UV range [53]. Likewise, even if some poison frogs do reflect UV, experimental evidence has shown UV-visually sensitive predators such as birds do not perceive UV reflectance as aposematic [54]. Thus, human inability to see UV might not necessarily bias the assessment of conspicuousness of poison frogs based only on this light sensitivity range.

Fourth, poison frogs use visual cues and skin coloration to communicate with their conspecifics [32]. Evidence suggests that at least one species, *Dendrobates pumilio*, has trichromatic-like vision [30] and at least other three species may have color vision sensitivities [55]. Interestingly, the maxima spectral sensitivity (λ_{max}) of the three putative color photopigments of *Dendrobates pumilio* [30] and those of humans [51] are very similar: (1) putative short-wavelength-sensitivity receptor or SWS is $\lambda_{max:D.pumilio}$ = 466 nm versus $\lambda_{max:H.sapiens}$ = 420 nm; (2) putative middle-wavelength-sensitivity receptor or MWS is $\lambda_{max:D.pumilio}$ = 489 nm versus $\lambda_{max:H.sapiens}$ = 535 nm; and (3) putative long-wavelength-sensitivity receptor or LWS is $\lambda_{max:D.pumilio}$ = 561 nm versus $\lambda_{max:H.sapiens}$ = 562 nm. Thus, humans appear to visually perceive wavelengths very close to what poison frogs perceive in their conspecifics.

Finally, the number of taxa included in our dataset is extremely large (172 species). Our dataset includes some highly threatened or nearly extinct species (e.g., *Allobates olfersioides, Ameerega erythromos,* and *Hyloxalus delatorreae*). Obtaining a consistent spectral data of all these taxa will be very difficult. Despite our arguments for the validity of our procedures, we emphasize that best assessment of conspicuousness will be had from direct techniques including conspecific recognition tests, total reflectance flux at different light conditions, and models of predator perception [22, 28]. We hope that future field experiments provide a better assessment of conspicuousness across the poison frog family (299 species) from ecologically relevant and diverse receivers.

Metabolic rate variables

Table S1 provides the metabolic rate parameters of 54 species used in the present analyses that are derived from a previous study [33]. This data is corresponds to the same species that have at the same time corresponding acoustic data. The physiological variables were (*i*) resting metabolic rate (RMR, oxygen consumption while resting or VO_2rest mL*hour^{-1}), (*ii*) active metabolic rate after non-sustainable exercise (AMR, oxygen consumption after forced activity or VO_2active mL*hour^{-1}), and (*iii*) mean body mass to the nearest 0.01g of all the individuals tested. From these raw data, mass-specific metabolic rates (AMR and RMR) were estimated by dividing the metabolic rates by the body mass of each individual. The average of all conspecific rates was used as the species mass-specific metabolic rate.

Phylogenetic analyses

The molecular data for the poison frogs included the mitochondrial rRNA genes (12S and 16S) and the tRNA-Val (a total: ~2400 bp of mtDNA). The molecular phylogeny of poison frogs was inferred from the same 172 taxa for which acoustic data were collected and 6 outgroups (Fig. 1 and S1). Sequences were already available for most taxa, based on our previous work [33], and GenBank accession numbers are given in Table S1. We sequenced 17 taxa that lacked the corresponding molecular information. The list of primers and conditions used were described on previous study [33]. Sequence validation was performed using NCBI-BLAST to rule out contamination.

Sequence alignment was performed under an iterative approach (i.e., simultaneous alignment and tree estimation) using SATé [56]. This alignment was used to estimate the phylogeny of the poison frogs using a gene-partitioned model for 12S, tRNA-val, and 16S. Tree estimation and nodal support were calculated as follows. The molecular model for each gene was determined using jModelTest v 0.1.1 [57], and it was found to be GTR+Γ+I for all genes. Then, a maximum likelihood (ML) phylogeny was estimated using RAxML v 7.0.4 [58] and Garli v 2.0 [59], 400 nonparametric bootstrap searches provided support for the nodes. Bayesian approaches were performed using MrBayes v 3.4 [60] with default settings for all priors. The Markov Chain Monte Carlo (MCMC) setup for the Bayesian analysis included six independent runs, each one with

two chains of 75 million generations with a sampling rate of 1,500 generations. The convergence of the runs was determined using Tracer v 1.4 [61] and 300,000 trees were discarded as burnin. We inferred a 50% majority rule consensus tree and the posterior probabilities of each node from the remaining 300,000 trees. ML and Bayesian methodologies gave similar tree topologies and RAxML phylogeny was used as a starting tree topology for the chronogram estimation. Although, our ML phylogeny is only based on mitochondrial genes, this tree topology is similar to other studies that included more species, more markers, and additional nuclear genomic data [33, 62, 63]. The only major topological difference is the placement of Hyloxinae and Dendrobatinae as sister lineages. However, this node has weak support even with ~10 kb of nuclear and mitochondrial data [33]. Chronogram tree file is deposited in the TreeBASE database under the accession number XXXX. Few taxonomic changes are necessary based on our phylogenetic results. Specifically, we propose two new combinations for the species *Ameerega erythromos* and *Colostethus jacobuspetersi* that we found to be well-nested within *Hyloxalus* clade (i.e., bootstrap support > 95% and posterior probability ~ 1.0). For these reasons, we propose these new combination *Hyloxalus erythromos* and *H. jacobuspetersi*. We identified a list of unambiguous reconstructions on our aligned 12S-16S matrix that placed *H. erythromos* within *Hyloxalus* (positions indicated in parenthesis): A => C (1863, 2180); A => T (416, 1326); C => A (496); C => T (957, 1318, 2566, 2923); G => A (1669, 2580, 2946); and T => C (462, 682, 1449, 1760, 1862, 2014, 2512, 2701). Similarly, we identified a list of unambiguous reconstructions that placed *H. jacobuspetersi* within *Hyloxalus*: A => C (426, 451, 1863, 2180, 2192); A => G (460, 526, 2685, 3003); A => T (416, 1326, 1769, 1877, 1892, 2345, 2757, 2764); C => A (427, 496, 1107, 2627); C => T (164, 463, 957, 1249, 1318, 1320, 1645, 1772, 2062, 2064, 2257, 2432, 2566, 2636, 2725, 2803, 2923); G => A (179, 1669, 2132, 2220, 2580, 2629, 2946); T => A (1759, 1761, 2256, 2700, 2802, 2830); and T => C (221, 462, 682, 1349, 1449, 1760, 1862, 2014, 2056, 2264, 2273, 2512, 2701, 2761, 2804, 3005).

A chronogram of the poison frogs was estimated using a Bayesian approach and this was used for all comparative analyses. The age of the nodes and their confidence intervals were determined using Bayesian MCMC analysis of molecular sequences [64] and summarized in the following steps. A user-specified starting ultrametric tree was

estimated using the best RAxML topology under a penalized likelihood rate smoothing approach [65] with the Agastorophrynia (Bufonidae + Dendrobatidae) crown node set to 59.4 million of years ago (MYA) using the chronogram function in TreeEdit v. 1.0 [66]; the age of this node is the weighted average of the mean values from published Neobatrachia chronograms [67, 68]. The chronogram of poison frogs was estimated using a relaxed clock model approach implemented in BEAST v. 1.5.3 [64] with five user-defined sets of priors, namely, a GTR+Γ+I molecular model for each gene; the user-specified starting ultrametric tree; a normally distributed age of the root ($\bar{X}$ = 59.4 ± 5.336) in MYA; five node-age constraints; and *U*(0, 100) hyperprior for ucld.mean. The node-age constraints were in MYA, normally distributed, and correspond to the weighted averages of the crown node of Dendrobatidae ($\bar{X}$ = 38.073 ± 4.190 MYA) from [63, 67]; the crown node of Dendrobatinae + Hyloxinae + Colostethinae ($\bar{X}$ = 31.642 ± 5.625 MYA) from [63, 67, 68]; the crown node of Dendrobatinae ($\bar{X}$ = 24.117 ± 3.068 MYA) from [63, 67]; the crown node of Colostethinae ($\bar{X}$ = 27.090 ± 3.197 MYA) from [63]; and the crown node of *Ameerega* ($\bar{X}$ = 8.722 ± 1.936 MYA) from [63]. The suggested modifications for default MCMC operators were determined after two runs of two million generations with a sampling rate every 1,000 generations. The chronogram estimation included the suggested MCMC operator calibrations and four runs of 25 million generations sampled every 1,000 generations. The convergence of the runs and the optimal burnin was determined using Tracer v 1.4 [61]. The tree files were combined using LogCombiner [64], and ~60,000 initial trees were discarded as burnin. The maximum clade credibility summary tree was determined with the retained trees using TreeAnnotator [64]. The poison frog chronogram was determined from a summary tree using FigTree v 1.2.3 [69]. We assessed the robustness of the calibrations on the chronogram by dropping one node-age constraint at a time and comparing the node-age estimates with those of the chronogram with all node-age constraints. We did not find major age differences with those of the chronogram estimated with all the nodal constraints. Consequently, this last tree was used for all subsequent analyses.

Comparative Analyses

We used univariate and multivariate comparative analyses to determine trends of aposematic and acoustic signals evolution in poison frogs using the 172 taxa with call data. Our analyses aimed to identify the relationship between binary components aposematism (i.e., conspicuousness and alkaloid presence) and predictors based on call variables. Temperature and body size variables were also included to account for the strong influences by both variables on temporal and spectral call properties demonstrated for ectotherms [4]. The comparative methods are divided in the following sets of analyses: (*i*) aposematism diversification analyses, (*ii*) multivariate data exploration and variable reduction of call variables; (*iii*) phylogenetic signal and model of trait evolution; (*iv*) bivariate phylogenetic correlations and exploratory factor analyses; and (*v*) phylogenetic logistic regressions.

Diversification analyses were performed for the binary alkaloid sequestration and color binary variables (Figs. 2 and S4). For these variables, we estimated the likelihood of the ancestral states, rates of speciation, extinction, and transition between alternative character states. These parameters were estimated using the binary-state speciation and extinction (BiSSE) models [70] using the diversitree R-package v. 0.9 [71]. The BiSSE approach [70] estimates the likelihood for a model with up to six diversification parameters under different character states (i.e., 0-absent, 1-present): Speciation rates are indicated by λ_0 and λ_1, extinction rates are indicated by μ_0 and μ_1, and transition rates between each character state are indicated by q_{01} and q_{10} (i.e., from state 1 to 0 and 0 to 1 respectively). In BiSSE analyses, fixing parameters allows testing of alternative diversification models [71]. Optimal models are determined based on a combination of best fit to the data and minimization of the number of predictor variables. In our analyses, we tested alternative models for speciation (H_0: $\lambda_0 = \lambda_1$ vs. H_a: $\lambda_0 \neq \lambda_1$), extinction (H_0: $\mu_0 = \mu_1$ vs. H_a: $\mu_0 \neq \mu_1$), and transition (H_0: $q_{01} = q_{10}$ vs. H_a: $q_{01} \neq q_{10}$) rates between character states. Because pairs of compared models are nested, the difference between them can be approximated by χ^2 and its significance estimated with α-level set to 0.05 for a two-tailed distribution [71].

We also performed MCMC analyses following the procedure described in the diversitree-tutorial (www.zoology.ubc.ca/prog/diversitree/doc/diversitree-tutorial.pdf).

We used an exponential prior 1/(2*r*) where *r* is the character-independent diversification rate, an optimized step size window, and initial values obtained from starting BiSSE parameters. We ran the chains for 20,000 steps to estimate the marginal distributions for all diversification, extinction, and transition parameters. This procedure provided an alternative way to test and visualize (Fig. 2) the parameter space associated with the diversification models.

We explored the statistical power of our BiSSE analyses using simulations. The authors of BiSSE cautioned the use of their method with phylogenies with less than 300 tips [70] because the power of BiSSE analyses diminishes with low sample sizes [72]. This low power might reduce the probability of correctly rejecting a false null hypothesis (i.e., no differences in the diversification rates). Consequently, smaller sample sizes in BiSSE analyses should tend to decrease the ability to find differences between parameters. To explore the power of our BiSSE analyses, we used a modified approach described previously [72, 73]. Our power analyses can be summarized as follows. First, sets of 200 trees each were simulated with 50, 100, 150, 200, 300, and 400 terminals using the diversitree R-package v. 0.9 [71]; for these, we used the diversification parameters from our empirical (observed) data (Fig. S4). We compared the estimated parameters versus those from our observed data and calculated their differences (i.e., estimated – observed). Then, we simulated sets of 200 trees each with the corresponding number of terminals (i.e., 50, 100, 150, 200, 300, and 400 terminals) but constrained the speciation, extinction, and transition rates to be equal between alternative character states. Finally, we determined the percentage of simulated trees that rejected the constrained model and plot them as a function of the number of terminals (Fig. S4). We considered that number of terminals within a tree was adequate (i.e., provides enough power to the analyses) if the estimated parameters were enough to detect the differences in diversification parameters in >40% of the simulations. We found that our sample size was adequate for the speciation and transition rate tests, but more limited for the extinction rate analyses (Fig. S4).

Multivariate data exploration identified species-specific call features that were homologous and non-redundant across the poison frogs. We chose the variables for the analyses that met these criteria: (*i*) all taxa included in the phylogeny must have an

estimate; (*ii*) variables must have high interspecific and low inter-populational variation among dendrobatids [11, 74, 75] or in related hyloid frogs [8, 9]; (*iii*) variables must have been demonstrated to be relevant for conspecific recognition or phonotaxis during behavioral experiments [4]; and (*iv*) variables must be non-redundant and temporally closer to the shortest acoustic unit in poison frogs (i.e., pulse-note). As a consequence, we excluded all gross-scale temporal variables and those based on units of repetition (URs). The variables used for subsequent comparative analyses (Table S1) were reduced to 18 variables: body size (SVL), temperature, and non-redundant acoustic variables (Table S3).

The variables retained were further explored for their suitability for standard multivariate analyses by checking their distributions in SPSS v 16.0.2 [76]. Skewness and kurtosis were used to determine if the variables needed to be transformed [77]. Linearity and multivariate outliers were also explored using pairwise scatterplots and calculation of Mahalanobis distances [77]. Three taxa were excluded as multivariate outliers or lacking appropriate data (e.g., unknown SVL size). These taxa were *Ameerega bilinguis* (outlier), *Allobates* sp. Neblina (SVL not available), and *Hyloxalus jacobuspetersi* (outlier). Therefore, our starting dataset of 172 species was reduced to 169.

We used the reduced dataset to perform a standard principal component analysis (PCA). The PCAs were estimated with varimax orthogonal rotation by maximizing the variance explained by components using SPSS v 16.0.2 [76]. Three components were retained with eigenvalues > 1.0 and these explained 84.21% of the variance (Table S3). We described these components as follows: PC1→ "morphology," defined by SVL and call spectral features; PC2→ "behavior/physiology," defined by all fine-scale temporal call features; and PC3→ "environment," defined by temperature of recording, frequency modulation, and pulse-note shape. No variables were found to be crossloading (i.e., loadings > 0.4 or <–0.4 in two or more components). The three components retained were used as summary variables in the phylogenetic comparative analyses.

Phylogenetic signal for components and individual variables was assessed using Pagel's λ [78] and Blomberg *et al*.'s K-statistic [79]. For Pagel's λ, phylogenetic signal was estimated using 'fitDiscrete' and 'fitContinuous' functions of the geiger R-package v 1.3 [80]. Pagel's λ describes a phylogeny transformation parameter that gradually

excludes underlying phylogenetic structure under certain values of λ [78, 81]. The significance of λ was determined against a null hypothesis of no phylogenetic signal (i.e., H_0: λ = 0) by contrasting ML scores using a likelihood-ratio test (LRT) [80]. The K-statistic was determined from each continuous variables using 'phylosignal' function in the picante R-package v 1.3 [82]. The significance of each K-statistic was determined by comparing it to a randomized distribution of 10,000 replicates [79]. We tested four models of character evolution with our continuous variables (i.e., no signal or white noise, Brownian motion or BM, Ornstein-Uhlenbeck or OU, and Pagel's λ). We fit of each model to the data using the 'fitContinuous' function of the geiger R-package v 1.3 [80]. The best model was determined by contrasting ML scores using LTR for nested models or comparing Akaike's Information Criteria corrected for small samples (AICc) scores for non-nested models with a criterion of >3 units for "strong" support [80]. The estimates of phylogenetic signal and fit of evolutionary model are in Table S3.

Bivariate phylogenetic correlations between continuous variables were determined using phylogenetic independent contrasts (PIC) and generalized least squares-PGLS [83, 84]. For the PIC method, we estimated the pairwise correlations after calculating PICs for each continuous variable under its best model of character evolution [85]. This approach can be summarized as follows. The branch lengths of the poison frog chronogram were transformed in accordance to each trait best model of character evolution from Table S3. The best model-PICs were estimated using the transformed tree with the 'pic' function of the ape R-package v. 2.7 [86]. The presence of multivariate λ-PIC outliers was determined by calculating Mahalanobis distances [77]; no outliers were found. For the PGLS method, pairwise correlations were determined using 'pgls' function and 'ML' method of the caper R-package v. 1.0 [87]. This approach addresses phylogenetic signal by taking into account phylogenetic non-independence of the data in the estimation of correlation coefficients. The correlation matrices obtained are presented in Table S3.

Phylogenetic exploratory factor analyses were done using phylogenetic principal component analyses (PPCA) with varimax orthogonal rotation [33, 85]. The estimation of PPCA can be summarized as follows. The phylogenetic variance-covariance (VCV) matrix was estimated using the PICs and their standard deviations from the phylogenetic

correlation matrix (Table S3). The PPCAs were estimated using the VCV matrix as input in SPSS v 16.0.2 [76]. Phylogenetic principal components were retained if their eigenvalues were >1.0. Three components were found for both the best model- and the λ-PICs (Table S3). These components explained 77.88% (best model-PICs) and 78.45% (λ-PICs) of the variance. Each component was defined by the same variables as in the standard PCA. No variables were found to be crossloading between the retained components.

Phylogenetic logistic regressions (PLRs) were evaluated to determine if call variables predicted the two components of aposematism (Table S4). We tested two dependent binary variables: alkaloid sequestration (able/unable) and the set of conspicuousness variables (six binary color variables; cryptic/conspicuous). Our predictors were two sets of continuous call variables. First, we used the summary variables: PC1→ morphology, PC2→ behavior/physiology, and PC3→ environment. Second, we performed tests using individual call variables (e.g., note-pulse rate) with body size (SVL) and temperature as covariates. To account for phylogeny, we converted the poison frog chronogram to its phylogenetic VCV matrix using the 'vcv.phylo' function of the ape R-package v. 2.7 [86]. The phylogenetic VCV can be described as a square matrix where the elements in its main diagonal are root-to-tip branch lengths, and off-diagonal elements are branch lengths from the root to the last common ancestor of each pair of tips. The logistic regressions were performed with the PLogReg routine [88] using MATLAB v 7.9 [89]. This routine simultaneously tests for phylogenetic signal while conducting the regressions. We used the following options to run PLogReg: (*i*) three continuous predictors with no interactions, (*ii*) all continuous variables were standardized, (*iii*) estimation of phylogenetic and standard logistic regressions with the Firth corrections, (*iv*) bootstrap confidence intervals and statistical significance of the regression slope and intercept estimated after 1,000 simulations, and (*v*) α-level set to 0.05 for a two-tailed distribution. Convergence of model parameters was achieved in all cases (Table S4). Finally, outlier cases on the PLRs were also explored using these join criteria for their identification in logistic regressions [90]: standardized residuals with absolute values >3.0 and Cook's distance >1.0. No outliers were found.

To interpret significant regression coefficients, we applied the percentage increase in odds and the "divide by 4 rule" [91]. The increase in the odds percentage is defined as the likelihood increase of a positive outcome (i.e., able to sequester alkaloids or conspicuousness) by the change of one standard-unit in a continuous predictor while assuming all other predictors are fixed (Table S4). The "divide by 4 rule" states that the steepest point of the logistic curve is at its center [91]. The slope of the curve (i.e., the derivative of the logistic function) is maximized at the center point and it can be approximated by $\beta/4$ where β is the regression coefficient. The interpretation of $\beta/4$ is the upper bound of the predictive difference derived by a change of one standard-unit in a continuous predictor while assuming all other predictors are fixed. For example, the regression coefficient of the summary variable of body size-spectral properties (i.e., PC1→ morphology) on alkaloid sequestration using the phylogenetic logistic regression is defined as follows:

$$\Pr(\text{Alkaloid sequestration}) = \text{logit}^{-1}(-0.339 - 0.806*\text{PC1} - 1.583*\text{PC2} + 0.152*\text{PC3})$$

where the β coefficient of PC1 is – 0.806. By applying the "divide by 4 rule", $-0.806/4 \sim -0.202$, which corresponds to no more than a 20.2% negative difference or 1.2-fold decrease in the probability of having the ability to sequester alkaloids as a result of one standard-unit increase in PC1. This is correct if PC2 and PC3 are held constant. Further interpretations of these results are magnitude and relationship (i.e., positive or negative) of the loading of individual variables on PC1. For example, SVL loaded negatively (i.e., the larger the PC1 score, the smaller the SVL) while call spectral variables loaded positively (e.g., the larger the PC1 score, the higher the dominant frequency). Therefore, the significant result of the analysis of PC1 suggests that species with larger SVLs and lower call frequencies have higher and significant probabilities of sequestering alkaloids.

To corroborate our PLR results, we performed phylogenetic regressions using only continuous variables. These analyses confirmed that conspicuousness is predicted by spectral and temporal call variables (Table S4). Specifically, the continuous conspicuousness variable (i.e., total contrast score TCS or ΣS_i) is also predicted by the PCs (i.e., PC1→morphology and PC2→behavior/physiology) and individual temporal

variables (e.g., note-pulse rate). For the subset of taxa with metabolic data (52 species; Table S3), we performed pairwise correlation analyses between metabolic rates, body mass, and call variables. We estimated these phylogenetic regressions using the caper R-package v. 1.0 [87]. The significance of each PGLS correlation coefficient was determined at α-level set to 0.05 for a two-tailed test.

We also tested the pairwise correlations between the discrete dependent variables: conspicuousness variables, alkaloid sequestration, and perching behavior while vocalizing (Table S3). We used PLRs with Firth correction [88] and Pagel's 1994 test for correlation of two binary characters [92]. For the PLR procedures, we used the same methodology as described above for PCs on aposematism components. For the Pagel's 1994 test, we used two different implementations: The BayesTraits approach using BayesTraits v. 1.0 [93] and the Midford and Maddison approach using Mesquite v. 2.75 [94].

Supporting Information

Supporting Dataset S1. Poison frog male acoustic signal description including ecological and natural history data. Description for each species includes locality of collection (source), voucher, phylogeny identifier (Phy ID) number, recording equipment, temperature, body size (snout vent length, SVL), oscillogram, spectrogram, and power spectrum.

Supporting Text: Detailed description of all methods, taxonomic changes, and rationale for quantifying conspicuousness of poison frogs.

Supporting Figure S1. Chronogram, nodal support, and nodal age uncertainty of the Dendrobatidae (poison frog) Tree of Life. The chronogram shows nodal support under each different estimation method and node age uncertainty (blue bars). The species name and the phylogenetic identifier (Phy ID) number within square brackets are used in Tables S1– S3 and Dataset S1. Nodal support is given by non-parametric bootstrap proportions (ML-RAxML and ML-Garli) and Bayesian posterior probabilities (PP). The node-age calibration points are shown as open stars. These include (node A) crown node of Dendrobatidae ($\bar{X}$ = 38.1 ± 4.2 millions of years-MYA); (node B) the crown node of Dendrobatinae + Hyloxinae + Colostethinae ($\bar{X}$ = 31.6 ± 5.6 MYA); (node C) the crown node of Dendrobatinae ($\bar{X}$ = 24.1 ± 3.1 MYA); (node D) the crown node of Colostethinae ($\bar{X}$ = 27.1 ± 3.2 MYA); and (E) the crown node of *Ameerega* ($\bar{X}$ = 8.7 ± 1.9 MYA).

Supporting Figure S2. Definitions of call variables for (A) multinote calls and units of repetition (URs); and (B) single-note pulses (smallest acoustic units).

Supporting Figure S3. Binary characterization of conspicuousness of three exemplar poison frog species using color contrast thresholds. We characterized eleven frog segments: dorsal background (a), dorsal stripe (b), dorsolateral stripe (c), lateral background (d), ventrolateral stripe (e), oblique lateral stripe (f), arm dorsal (g), flash mark (h), thigh dorsal (i), throat (j), and abdomen (k). Frog diagrams were modified from s previous characterization [95]. Total contrast score (TCS or ΣS_i) was determined after adding the conspicuousness or cryptic states from all frog segments (Table S1 – S2). Poison frogs might have multiple natural predators (e.g., snakes, crabs, and birds) with diverse visual sensitivities under different light conditions. For this reason, each species was considered conspicuous (a red dot) or cryptic (brown dot) based on its ΣS_i in

relationship (i.e, more or equal than, ≥) to an increasing series an integer threshold values from 3, 4, …, 8. In the case of *Epipedobates tricolor*, its ΣS_i was 10 and this species is considered conspicuous under all color contrast thresholds (i.e., value of 1-conspicuous under binary variables $\Sigma S_i \geq 3$, $\Sigma S_i \geq 4$, …, $\Sigma S_i \geq 8$). For *Ameerega bilinguis*, its ΣS_i was 7 and this species is considered conspicuous under the thresholds 3 – 7 (i.e., value of 1-conspicuous under binary variables $\Sigma S_i \geq 3$, $\Sigma S_i \geq 4$, …, $\Sigma S_i \geq 7$ and 0-cryptic under the variable $\Sigma S_i \geq 8$). For *Hyloxalus awa*, its ΣS_i was only 3 and this species is considered conspicuous only under the threshold 3 (i.e., value of 1-conspicuous under binary variable $\Sigma S_i \geq 3$ and 0-cryptic under variables $\Sigma S_i \geq 4$, $\Sigma S_i \geq 5$, …, $\Sigma S_i \geq 8$.

Supporting Figure S4. Diversity analyses of the aposematic syndrome in poison frogs. (A) Distribution of conspicuousness and alkaloid sequestration. To visualize the relationship between call and aposematism, we depicted each species based on its conspicuousness, alkaloid sequestration ability, and plotted them in call space (PC1: morphology, versus PC2: behavior/physiology). Note the location of *Allobates zaparo* (visual Batesian mimic) and its aposeme model *Ameerega parvula*. (B) Probability values for presence of phylogenetic signal (Pagel's lambda test) and correlations between each conspicuousness variable with alkaloid sequestration ability. Significant correlations ($P < 0.05$) indicate the aposmatic phenotype. No phylogenetic signal was detected in the variables TCS7 and TCS8 variables. All other conspicuousness variables are significantly correlated with alkaloid sequestration. (C) Diversification analysis results under each characterization of conspicuousness: λ = speciation, μ = extinction, and q = transition between character states. Alternative states are indicated by subscripts: 0 (cryptic coloration or unable or lack of alkaloid sequestration) and 1 (conspicuousness or alkaloid sequestration). The results of power analyses are indicated by bar charts. These bars indicate mean and ± one standard deviation of the differences between the values of the parameters estimated (200 simulated data) and observed (observed data). Line plots indicate the percentage of simulated trees that rejected the constrained model as a function of the number of terminals under different constraints (i.e., $\lambda_0 \neq \lambda_1$, $\mu_0 \neq \mu_1$, and $q_{10} \neq q_{01}$). (D) Distribution of taxa based on their aposematic phenotypic value. The variable TCS5 or $\Sigma S_i \geq 5$ is the best qualitative (binary) measurement for classifying taxa

as aposematic (i.e., conspicuous when they also are able to sequester alkaloids) based on joint criterion for accuracy, precision, and sensitivity indices when they are closer to 1.00 (see *Supporting Text* for details).

Supporting Table S1. Temporal, spectral, conspicuousness (total contrast score TCS or ΣS_i and corresponding binary values $\Sigma S_i \geq 3$, $\Sigma S_i \geq 4$, …, $\Sigma S_i \geq 8$), alkaloid sequestration ability, and scaling variables (metabolic rates, body mass, and SVL). The call variables include the phylogeny identifier number (Phy ID from Fig. S1), locality, call behavior, habit, temperature, size, number of recordings, multinote call features, units of repetition (UR), initial pulse-note, and middle pulse-note parameters.

Supporting Table S2. Binary coloration classification and composite conspicuousness (i.e., total contrast score TCS or ΣS_i measured as the sum of all binary S_i points) of the species of poison frogs included in the analysis. Description of coloration is based on specimen descriptions from the literature, as well as from scoring of photographs of live animals. Nomenclature of the regions in the male frog's body is provided in the Fig. S3.

Supporting Table S3. Phylogenetic signal and model character evolution of the poison frog call parameters, and the three principal components summarizing the call parameters. Lambda (λ) estimates for color variables (TCS3, TCS4, ..., TCS8) and ability to sequester alkaloids in poison frogs.

Supporting Table S4. Standard and phylogenetic logistic regression analyses between discrete binary variables (conspicuousness variables TCS3, TCS4, ..., TCS8 and ability to sequester alkaloids) and principal components.

Supporting Information References:

[1] Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. 2005 Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25, 1965-1978. (doi:10.1002/joc.1276).
[2] Wells, K.D. 2007 *The Ecology and Behavior of Amphibians*. Chicago (IL), University Of Chicago Press; 1400 p.
[3] Salati, E. & Vose, P.B. 1984 Amazon Basin - A system in equilibrium. *Science* 225, 129-138. (doi:10.1126/science.225.4658.129).
[4] Gerhardt, H.C. & Huber, F. 2002 *Acoustic Communication in Insects and Anurans*. Chicago, IL, University of Chicago Press; 531 p.
[5] Gayou, D.C. 1984 Effects of temperature on the mating call of *Hyla versicolor*. *Copeia* 1984, 733-738.
[6] Adobe-Systems-Incorporated. 2011 Adobe Audition. (ed. Adobe-Systems-Incorporated), 4.0 ed. San Jose, CA.
[7] Charif, R.A., Waack, A.M. & Strickman, L.M. 2010 Raven Pro 1.4 (Ithaca, NY, Cornell Lab of Ornithology.
[8] Cocroft, R.B. & Ryan, M.J. 1995 Patterns of advertisement call evolution in toads and chorus frogs. *Animal Behaviour* 49, 283-303.
[9] Lemmon, E.M. 2009 Diversification of conspecific signals in sympatry: geographic overlap drives multidimensional reproductive character displacement in frogs. *Evolution* 63, 1155-1170. (doi:10.1111/j.1558-5646.2009.00650.x).
[10] McLister, D., Stevens, E.D. & Bogart, J.P. 1995 Comparative contractile dynamics of calling and locomotor muscles in three hylid frogs. *J Exp Biol* 198, 1527-1538.
[11] Erdtmann, L. & Amezquita, A. 2009 Differential evolution of advertisement call traits in dart-poison frogs (Anura: Dendrobatidae). *Ethology* 115, 801-811. (doi:10.1111/j.1439-0310.2009.01673.x).
[12] Gerhardt, H.C. 1998 Acoustic signals of animals: Recording, field measurements, analysis and description. In *Animal Acoustic Communication: Sound Analysis and Research Methods* (eds. S.L. Hopp, M.J. Owren & C.S. Evans), pp. 1-25. Berlin, Heidelberg, New York, Springer-Verlag.
[13] Littlejohn, M.J. 2001 Patterns of differentiation in temporal properties of acoustic signals of anurans. In *Anuran Communication* (ed. M.J. Ryan), pp. 102-120. Washington, Smithsonian Institution Press
[14] Pröhl, H. 2003 Variation in male calling behaviour and relation to male mating success in the strawberry poison frog (*Dendrobates pumilio*). *Ethology* 109, 273-290.
[15] Zimmermann, E. 1990 Behavioral signals and reproduction modes in the neotropical frog family Dendrobatidae. In *Fortschritte der Zoologie: Biology and Physiology of Amphibians* (ed. W. Hanke), pp. 61-73. Stuttgart, New York, Gustav Fischer Verlag.
[16] Lötters, S., Reichle, S. & Jungfer, K.H. 2003 Advertisement calls of Neotropical poison frogs (Amphibia : Dendrobatidae) of the genera *Colostethus*, *Dendrobates* and *Epipedobates*, with notes on dendrobatid call classification. *Journal of Natural History* 37, 1899-1911.

[17] Daly, J.W., Gusovsky, F., Myers, C.W., Yotsu-Yamashita, M. & Yasumoto, T. 1994 First occurrence of tetrodotoxin in a dendrobatid frog (*Colostethus inguinalis*), with further reports for the bufonid genus *Atelopus*. *Toxicon* 32, 279-285.
[18] Miyazawa, K. & Noguchi, T. 2001 Distribution and origin of tetrodotoxin. *Journal of Toxicology-Toxin Reviews* 20, 11-33.
[19] Saporito, R.A., Spande, T.F., Garraffo, H.M. & Donnelly, M.A. 2009 Arthropod alkaloids in poison frogs: A review of the 'dietary hypothesis'. *Heterocycles* 79, 277-297.
[20] Cott, H.B. 1940 *Adaptative Coloration in Animals*. London, Methuen & Co LTD; 508 p.
[21] Sherratt, T.N. & Beatty, C.D. 2003 Th evolution of warning signals as reliable indicators of prey defense. *Am Nat* 162, 377-389.
[22] Endler, J.A. 1990 On the measurement and classification of color in studies of animal color patterns. *Biological Journal of the Linnean Society. Linnean Society of London* 41, 315-352.
[23] Endler, J.A. & Mappes, J. 2004 Predator mixes and the conspicuousness of aposematic signals. *Am Nat* 163, 532-547.
[24] Bennett, A.T.D., Cuthill, I.C. & Norris, K.J. 1994 Sexual selection and the mismeasure of color. *Am Nat* 144, 848-860. (doi:10.1086/285711).
[25] Endler, J.A. 1978 A predators view of animal color patterns. *Evolutionary Biology* 11, 319-364.
[26] Schütze, H., Manning, C. & Raghavan, P. 2008 *Introduction to Information Retrieval* Cambridge, UK Cambridge University Press; 482 p.
[27] Endler, J.A. & Mielke, P.W. 2005 Comparing entire colour patterns as birds see them. *Biological Journal of the Linnean Society. Linnean Society of London* 86, 405-431. (doi:10.1111/j.1095-8312.2005.00540.x).
[28] Maan, M.E. & Cummings, M.E. 2012 Poison frog colors are honest signals of toxicity, particularly for bird predators. *Am Nat* 179, E1-E14. (doi:10.1086/663197).
[29] Crothers, L.R. & Cummings, M.E. 2013 Warning signal brightness variation: Sexual selection may work under the radar of natural selection in populations of a polytypic poison frog. *Am Nat* 181, E116-E124.
[30] Siddiqi, A., Cronin, T.W., Loew, E.R., Vorobyev, M. & Summers, K. 2004 Interspecific and intraspecific views of color signals in the strawberry poison frog *Dendrobates pumilio*. *J Exp Biol* 207, 2471-2485.
[31] Summers, K., Symula, R., Clough, M. & Cronin, T. 1999 Visual mate choice in poison frogs. *Proc Biol Sci* 266, 2141-2145.
[32] Hödl, W. & Amezquita, A. 2001 Visual signaling in anuran amphibians. In *Anuran Communication* (ed. M.J. Ryan), pp. 121-141. Washington, Smithsonian Institution Press.
[33] Santos, J.C. & Cannatella, D.C. 2011 Phenotypic integration emerges from aposematism and scale in poison frogs. *Proc Natl Acad Sci USA* 108, 6175-6180.
[34] Poulin, B., Lefebvre, G., Ibáñez, R., Jaramillo, C., Hernández, C. & Rand, A.S. 2001 Avian predation upon lizards and frogs in a neotropical forest understorey. *Journal of Tropical Ecology* 17, 21-40.
[35] Noonan, B.P. & Comeault, A.A. 2009 The role of predator selection on polymorphic aposematic poison frogs. *Biol Lett* 5, 51-54. (doi:10.1098/rsbl.2008.0586).

[36] Saporito, R.A., Zuercher, R., Roberts, M., Gerow, K.G. & Donnelly, M.A. 2007 Experimental evidence for aposematism in the dendrobatid poison frog *Oophaga pumilio*. *Copeia* 2007, 1006-1011.
[37] Hegna, R.H., Saporito, R.A. & Donnelly, M.A. 2012 Not all colors are equal: predation and color polytypism in the aposematic poison frog *Oophaga pumilio*. *Evolutionary Ecology*, DOI 10.1007/s10682-10012-19605-z.
[38] Hegna, R.H., Saporito, R.A., Gerow, K.G. & Donnelly, M.A. 2011 Contrasting colors of an aposematic poison frog do not affect predation. *Annales Zoologici Fennici* 48, 29-38.
[39] Chouteau, M. & Angers, B. 2011 The role of predators in maintaining the geographic organization of aposematic signals. *Am Nat* 178, 810-817. (doi:10.1086/662667).
[40] Richards-Zawacki, C.L., Yeager, J. & Bart, H.P.S. 2013 No evidence for differential survival or predation between sympatric color morphs of an aposematic poison frog. *Evolutionary Ecology* 27, 783-795. (doi:10.1007/s10682-013-9636-0).
[41] Paluh, D.J., Hantak, M.M. & Saporito, R.A. 2013 A test of aposematism in the dendrobatid poison frog *Oophaga pumilio*: The importance of movement in clay model experiments. *Journal of Herpetology* DOI: 10.1670/13-027.
[42] Cooper, W.E., Caldwell, J.P. & Vitt, L.J. 2009 Conspicuousness and vestigial escape behaviour by two dendrobatid frogs, Dendrobates auratus and Oophaga pumilio. *Behaviour* 146, 325-349. (doi:10.1163/156853909x410946).
[43] Cooper, W.E., Caldwell, J.P. & Vitt, L.J. 2009 Risk assessment and withdrawal behavior by two species of aposematic poison frogs, *Dendrobates auratus* and *Oophaga pumilio*, on forest trails. *Ethology* 115, 311-320. (doi:10.1111/j.1439-0310.2009.01615.x).
[44] Daly, J.W. & Myers, C.W. 1967 Toxicity of Panamanian poison frogs (Dendrobates): some biological and chemical aspects. *Science* 156, 970-973.
[45] Summers, K. & Clough, M. 2001 The evolution of coloration and toxicity in the poison frog family (Dendrobatidae). *Proc Natl Acad Sci USA* 98, 6227-6232.
[46] Wang, I.J. & Shaffer, H.B. 2008 Rapid color evolution in an aposematic species: a phylogenetic analysis of color variation in the strikingly polymorphic strawberry poison-dart frog. *Evolution* 62, 2742-2759. (doi:10.1111/j.1558-5646.2008.00507.x).
[47] Summers, K., Cronin, T.W. & Kennedy, T. 2003 Variation in spectral reflectance among populations of *Dendrobates pumilio*, the strawberry poison frog, in the Bocas de Toro Archipelago, Panama. *Journal of Biogeography* 30, 35-53.
[48] Comeault, A.A. & Noonan, B.P. 2011 Spatial variation in the fitness of divergent aposematic phenotypes of the poison frog, *Dendrobates tinctorius*. *J Evol Biol* 24, 1374-1379. (doi:10.1111/j.1420-9101.2011.02258.x).
[49] Stuart, Y.E., Dappen, N. & Losin, N. 2012 Inferring predator behavior from attack rates on prey-replicas that differ in conspicuousness. *PloS One* 7, e48497. (doi:10.1371/journal.pone.0048497).
[50] Richards-Zawacki, C.L., Wang, I.J. & Summers, K. 2012 Mate choice and the genetic basis for colour variation in a polymorphic dart frog: inferences from a wild pedigree. *Mol Ecol* 21, 3879-3892. (doi:10.1111/j.1365-294X.2012.05644.x).

[51] Surridge, A.K., Osorio, D. & Mundy, N.I. 2003 Evolution and selection of trichromatic vision in primates. *Trends Ecol Evol* 18, 198-205. (doi:10.1016/s0169-5347(03)00012-0).
[52] Goldsmith, T.H. 1990 Optimization, constraint, and history in the evolution of eyes. *Q Rev Biol* 65, 281-322. (doi:10.1086/416840).
[53] Vorobyev, M., Marshall, J., Osorio, D., de Ibarra, N.H. & Menzel, R. 2001 Colourful objects through animal eyes. *Color Research and Application* 26, S214-S217. (doi:10.1002/1520-6378(2001)26:1+<::aid-col45>3.0.co;2-a).
[54] Lyytinen, A., Alatalo, R., V, Lindstrom, L. & Mappes, J. 2001 Can ultraviolet cues function as aposematic signals? *Behavioral Ecology* 12, 65-70.
[55] Hellman, J.P. & Jaeger, R.G. 1974 Phototactic responses to spectrally dominant stimuli and use of colour vision by adult anuran amphibians - a comparative survey. *Animal Behav* 22, 757-795.
[56] Liu, K., Raghavan, S., Nelesen, S., Linder, C.R. & Warnow, T. 2009 Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. *Science* 324, 1561-1564. (doi:10.1126/science.1171243).
[57] Posada, D. 2008 jModelTest: Phylogenetic Model Averaging. *Mol Bio Evol* 25, 1253-1256.
[58] Stamatakis, A. 2006 RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.
[59] Zwickl, D. 2006 Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. http://www.bio.utexas.edu/faculty/antisense/garli/Garli.html [Ph.D. dissertation]. Austin (TX), The University of Texas at Austin.
[60] Huelsenbeck, J.P. & Ronquist, F.P. 2001 MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754-755.
[61] Rambaut, A. & Drummond, A.J. 2007 Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer (
[62] Grant, T., Frost, D.R., Caldwell, J.P., Gagliardo, R., Haddad, C.F.B., Kok, P.J.R., Means, B.D., Noonan, B.P., Schargel, W. & Wheeler, W.C. 2006 Phylogenetic systematics of dart-poison frogs and their relatives (Anura: Athesphatanura: Dendrobatidae). *Bulletin of the American Museum of Natural History* 299, 1-262.
[63] Santos, J.C., Coloma, L.A., Summers, K., Caldwell, J.P., Ree, R. & Cannatella, D.C. 2009 Amazonian amphibian diversity is primarily derived from late Miocene Andean lineages. *PLoS Biol* 7, e56 doi:10.1371/journal.pbio.1000056
[64] Drummond, A.J. & Rambaut, A. 2007 BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7, 214.
[65] Sanderson, M.J. 2002 Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach *Mol Bio Evol* 19, 101-109.
[66] Rambaut, A. & Charleston, M.A. 2002 TreeEdit: Phylogenetic Tree Editor. http://tree.bio.ed.ac.uk/software/treeedit/. (1.0 alpha 10 ed. Oxford, UK.
[67] Heinicke, M.P., Duellman, W.E., Trueb, L., Means, D.B., MacCulloch, R.D. & Hedges, S.B. 2009 A new frog family (Anura: Terrarana) from South America and an expanded direct-developing clade revealed by molecular phylogeny. *Zootaxa*, 1-35.

[68] Roelants, K., Gower, D.J., Wilkinson, M., Loader, S.P., Biju, S.D., Guillaume, K., Moriau, L. & Bossuyt, F. 2007 Global patterns of diversification in the history of modern amphibians. *Proc Nat Acad Sci USA* 104, 887-892.
[69] Rambaut, A. 2009 FigTree: Tree Figure Drawing Tool. http://tree.bio.ed.ac.uk/software/figtree/. (1.2.3 ed. Edinburgh, UK.
[70] Maddison, W.P., Midford, P.E. & Otto, S.P. 2007 Estimating a binary character's effect on speciation and extinction. *Syst Biol* 56, 701-710.
[71] FitzJohn, R.G., Maddison, W.P. & Otto, S.P. 2009 Estimating Trait-Dependent Speciation and Extinction Rates from Incompletely Resolved Phylogenies. *Syst Biol* 58, 595-611. (doi:10.1093/sysbio/syp067).
[72] Davis, M.P., Midford, P.E. & Maddison, W. 2013 Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. *BMC Evol Biol* 13. (doi:38
10.1186/1471-2148-13-38).
[73] Maliska, M.E., Pennell, M.W. & Swalla, B.J. 2013 Developmental mode influences diversification in ascidians. *Biol Lett* 9, 1-4.
[74] Amezquita, A., Lima, A.P., Jehle, R., Castellanos, L., Ramos, O., Crawford, A.J., Gasser, H. & Hodl, W. 2009 Calls, colours, shape, and genes: a multi-trait approach to the study of geographic variation in the Amazonian frog Allobates femoralis. *Biological Journal of the Linnean Society. Linnean Society of London* 98, 826-838.
[75] Amezquita, A., Flechas, S.V., Lima, A.P., Gasser, H. & Hödl, W. 2011 Acoustic interference and recognition space within a complex assemblage of dendrobatid frogs. *Proc Natl Acad Sci USA* 108, 17058-17063. (doi:10.1073/pnas.1104773108).
[76] SPSS. 2008 Statistical Package for the Social Sciences (SPSS) for Windows, Rel. 16.0.2 (Chicago.
[77] Tabachnick, B.G. & Fidell, L.S. 2007 *Using multivariate statistics*. 5th ed. Boston (MA), Pearson Education.
[78] Pagel, M. 1993 Seeking the evolutionary regression coefficient: an analysis of what comparative methods measure. *J Theor Biol* 164, 191–205.
[79] Blomberg, S., Garland, T. & Ives, A. 2003 Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57, 717-745.
[80] Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E. & Challenger, W. 2008 GEIGER: investigating evolutionary radiations. *Bioinformatics* 24, 129-131. (doi:10.1093/bioinformatics/btm538).
[81] Pagel, M. 1999 Inferring the historical patterns of biological evolution. *Nature* 401, 877-884.
[82] Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P. & Webb, C.O. 2010 Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26, 1463-1464.
[83] Felsenstein, J. 1985 Phylogenies and the comparative method. *Am Nat* 125, 1-15.
[84] Garland, T. & Ives, A.R. 2000 Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *Am Nat* 155, 346-364.
[85] Santos, J.C. 2012 Fast molecular evolution associated with high active metabolic rates in poison frogs. *Mol Bio Evol* 29, 2001-2018.

[86] Paradis, E., Claude, J. & Strimmer, K. 2004 APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20, 289-290
[87] Orme, D. 2011 The caper package: comparative analysis of phylogenetics and evolution in R. http://cran.r-project.org/web/packages/caper/vignettes/caper.pdf. (
[88] Ives, A.R. & Garland, T. 2010 Phylogenetic logistic regression for binary dependent variables. *Syst Biol* 59, 9-26. (doi:10.1093/sysbio/syp074).
[89] MathWorks. 2009 MATLAB. (version 7.9 ed. Natick, MA, The MathWorks Inc.
[90] Hosmer, D.W., Lemeshow, S. & Sturdivant, R.X. 2013 *Applied Logistic Regression*. Hoboken, NJ, John Wiley & Sons, Inc.; 528 p.
[91] Gelman, A. & Hill, J. 2006 *Data Analysis Using Regression and Multilevel/Hierarchical Models*. 1st ed. New York, NY, Cambridge University Press; 648 p.
[92] Pagel, M. 1994 Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete data. *Proc Biol Sci* 255, 37-45.
[93] Pagel, M. & Meade, A. 2007 BayesTraits. (ver 1.0 ed. Reading, UK, University of Reading.
[94] Maddison, W.P. & Maddison, D.R. 2009 Mesquite: a modular system for evolutionary analysis. http://mesquiteproject.org. (ver 2.75 ed.
[95] Silverstone, P.A. 1976 A revision of the poison-arrow frogs of the genus *Phyllobates* Bibron in Sagra (Family Dendrobatidae). *Bulletin of the Natural History Museum of Los Angeles County Science* 27, 1-53.