



# Phylogenetic influence on mating call preferences in female túngara frogs, *Physalaemus pustulosus*

MICHAEL J. RYAN\*† & A. STANLEY RAND†

\*Department of Zoology, University of Texas

†Smithsonian Tropical Research Institute

(Received 27 April 1998; initial acceptance 22 June 1998;  
final acceptance 16 November 1998; MS. number: A7938R)

We evaluated how various phylogenetic models for estimating ancestral characters can influence studies of behavioural evolution. Previously we used a single model of evolution to estimate the values of call characters at ancestral nodes for the *Physalaemus pustulosus* species group and some close relatives (Ryan & Rand 1995, *Science*, **269**, 390–392). We then synthesized these ancestral calls and measured the females' responses to such calls in phonotaxis experiments. We repeated the above procedure to determine the sensitivity of these results and conclusions to various models used to estimate the ancestral call characters. We asked whether: (1) different models gave different call estimates for the same nodes; (2) different call estimates at the same node were perceived as different by females; and (3) differences in female responses influenced previous conclusions. We used seven different models that varied in at least one of the following parameters: tree topology (bifurcating versus pectinate in-group trees), algorithms (local squared-change versus squared-change parsimony), tempo (gradual or punctuated evolution), and outgroups (two or three outgroup taxa used). Although different models often gave different call estimates for the same node, these different estimates often were not perceived as different by the females. These data reinforce our previous conclusions that: (1) the range of female preferences exceeds the known variation of the conspecific call; (2) females do not discriminate between the conspecific call and the call of their most recent ancestor; and (3) female responses may be context dependent, given that females differ in their responses to the same signal variation in discrimination and recognition experiments.

© 1999 The Association for the Study of Animal Behaviour

Most studies of behavioural evolution address the current selection forces acting on behaviour (Brown 1982). There is general agreement, however, that the past history of selection and the constraints under which behaviours evolve also have a strong influence on current phenotypes (Lorenz 1941; Gould & Vrba 1982; Brooks & McLennan 1991; Harvey & Pagel 1991; Ryan 1996). Relative to studies of current function, however, studies of past history require a different set of methods.

Recent re-emphasis of historical aspects of behaviour has resulted in the use of explicit phylogenetic approaches to uncover historical patterns of behavioural evolution (e.g. Felsenstein 1985; Huey & Bennett 1987; Brooks & McLennan 1991; Martins & Garland 1991b; de Queiroz & Wimberger 1993; Brooks et al. 1995; Martins 1996a, b). Typically, behavioural characters are mapped onto a phylogenetic tree, which is a graphical

representation of a hypothesis of phylogenetic relationships. The resulting data might then be used to describe the general patterns of behavioural evolution (e.g. Prum 1990; Lanyon 1992; Cocroft & Ryan 1995), to test hypotheses of coevolution (e.g. plant–insect: Mitter et al. 1991; parasite–host: McLennan & Brooks 1991), or to describe the sequence by which pairs of characters evolve (Höglund 1989; Basolo 1990, 1995; Ryan 1990; Proctor 1991, 1992; McClintock & Uetz 1996). The behavioural characters being mapped may or may not be used to generate the phylogenetic hypothesis, depending on the study.

The interpretation of a phylogenetic analysis of behaviour is crucially dependent on the assumed hypothesis of phylogenetic relationships (Reeve & Sherman 1993; Ryan 1996). It would be an exaggeration, however, to state that if the phylogenetic hypothesis is 'wrong' then so is the behavioural interpretation. A phylogenetic hypothesis is a series of hypotheses about the relationships of all taxa to one another. It is more appropriate to view any smaller set of relationships within the proposed phylogeny as

Correspondence: M. J. Ryan, Department of Zoology, University of Texas, Austin TX 78712, U.S.A. (email: [mryan@mail.utexas.edu](mailto:mryan@mail.utexas.edu)). A. S. Rand is at the Smithsonian Tropical Research Institute, APO 2072, Balboa, Panama.

being more or less supported by the available data rather than viewing the entire set of relationships as right or wrong. The sum of such support for specific relationships indicates the support for the overall hypothesis. There are various phylogenetic algorithms that can be used to estimate such support (e.g. retention index, consistency index; reviewed in [Wimberger & de Queiroz 1996](#)).

For some purposes the support for specific networks of relationships within the overall phylogeny might be more relevant than the support for the overall phylogenetic hypothesis. For example, [Höglund \(1989\)](#) tested the hypothesis that the presence of a lek mating system influenced the evolution of sexual dimorphism in some birds. In this example, the interpretations depend on the phylogenetic hypothesis only to the extent that the hypothesis can correctly predict when transitions to lek mating systems occurred, but the interpretations are not at all dependent on the phylogenetic relationships among taxa that share lekking or nonlekking mating systems through a common ancestor. Therefore, an estimate of the overall support for a phylogenetic hypothesis could give a false impression of the degree to which that phylogeny is consistent with a specific interpretation of behavioural evolution.

Another use of phylogenetics in animal behaviour is to estimate past behavioural phenotypes. Regardless of the proposed phylogeny, the method used when characters are estimated at ancestral nodes can also bias the interpretations. There are several methods that can be used to reconstruct value of nodal characters, including: squared-changed parsimony, local squared-change parsimony, and linear parsimony (e.g. [Swofford & Maddison 1987](#); [Maddison 1991](#); [Martins & Garland 1991b](#); [Maddison & Maddison 1992](#); [Martins & Hansen 1997](#); [Schluter et al. 1997](#)). Different reconstruction algorithms, however, can give different estimates of nodal character values. The particular tempo of evolution (e.g. punctuated or gradual change) can similarly influence these values. Another potential bias, not often acknowledged, is the outgroups that are used in the analysis. It is possible that the inclusion or exclusion of a particular outgroup taxon might not influence the tree topology but might substantially alter estimates of character values at ancestral nodes.

Phylogeneticists are well aware of how different assumptions underlying character reconstruction models might bias ancestral character estimates and subsequent interpretations (e.g. [Maddison & Maddison 1992](#); [Losos & Miles 1994](#); [Martins 1996b](#)). The robustness of character estimates to such assumptions has been investigated in simulation studies by [Losos \(1994\)](#) and [Martins \(1996b\)](#), for example. Our approach in estimating the effects of these various biases on specific interpretations of behavioural evolution is to repeat the behavioural analyses under conditions in which several factors are varied, such as: tree topology, outgroups, algorithms for estimating nodal values, and the specific tempo of evolution. Such a sensitivity analysis should allow us to evaluate to some extent the robustness of our interpretations. We have taken this approach to extend a previous study of

historical effects on acoustic mate recognition in túngara frogs, *Physalaemus pustulosus* ([Ryan & Rand 1995](#)). In that study we used an analysis of the advertisement calls of extant species, combined with a molecular phylogeny of the *P. pustulosus* species group and some close relatives to estimate and then synthesize advertisement calls at ancestral nodes. These calls were then used in phonotaxis experiments to ascertain the degree to which they elicited phonotactic responses from female túngara frogs.

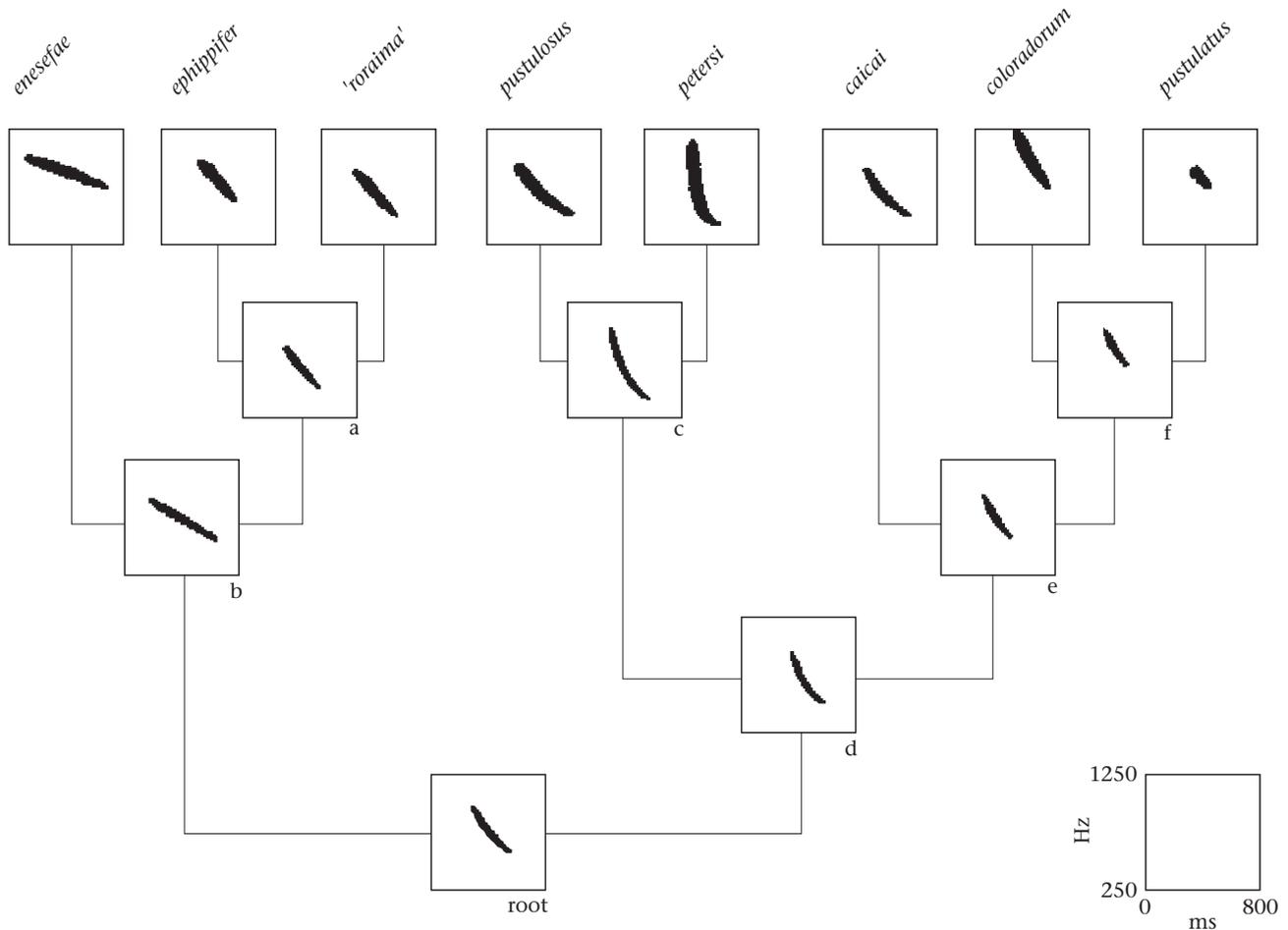
The main purpose of this study was to determine the robustness of our conclusions about call-preference evolution in relation to assumptions used to estimate the ancestral calls. To do this we addressed three questions. (1) Do different models for reconstruction of mating call characters yield different estimates of ancestral calls for the same nodes? (2) If calls differ at the same node, are they perceived as different by females in phonotaxis tests? (3) If different calls at the same node are perceived as different by the females, do we need to alter the conclusions of our previous study of historical effects on female preferences.

### Female Preference for Ancestral Calls in Túngara Frogs

As in many other organisms, female túngara frogs show a strong preference for conspecific mating signals when these are presented in concert with a heterospecific mating signal. Nevertheless, the strength of the preferences for the conspecific call varies, as estimated by the number of times females are attracted to (move towards) the heterospecific call. When presented with a heterospecific call paired with white noise, however, female preference is even more variable; some calls do not elicit positive phonotaxis while others elicit a stronger and statistically significant response ([Ryan & Rand 1993a, b, c](#)). This result is not peculiar to túngara frogs, as similar results were found in phonotaxis experiments with another close relative, *P. coloradorum* ([Ryan & Rand 1993a](#)).

The salience of some heterospecific calls in eliciting phonotaxis led us to consider how females would respond to calls of ancestral species. Obviously, these calls are not accessible for experimentation. Thus we estimated what such calls might sound like ([Ryan & Rand 1995](#)). To do this we used the most parsimonious hypothesis for the phylogenetic relationships within the *P. pustulosus* species group and three close relatives ([Fig. 1](#); [Cannatella et al. 1998](#)) based on morphological characters, allozymes and mitochondrial DNA sequences, as well as calls. Inclusion or exclusion of the call data did not change the most parsimonious hypothesis. Independent analyses of the morphological and molecular data tended to result in trees congruent with one another and with the tree combining all data sets. Independent analysis of the call data, however, yielded a phylogenetic tree quite unlike all the rest ([Cannatella et al. 1998](#)).

We used the local squared-change parsimony algorithm and a model of gradual evolution (Brownian motion) in which branch lengths were derived from the most parsimonious estimate of changes in DNA base pairs (see [Figure 1](#) in [Ryan & Rand 1995](#)); this assumes that



**Figure 1.** The tree illustrating the most parsimonious hypothesis for the relationships among members of the *Physalaemus pustulosus* species group and the three species we used as outgroups. *Physalaemus 'roraima'* is an undescribed species. Sonograms illustrate the synthetic advertisement calls for each taxa; the calls estimated for the ancestral nodes were derived from a local squared-change parsimony model assuming a gradual model of evolution. This phylogenetic tree and the estimated ancestral calls is LSG in Table 1 and Figs 3 and 4.

changes in DNA are proportional to the total amount of evolutionary change accumulated on each branch (Olmand 1997). We then estimated the value of eight quantitative call characters at each of the ancestral nodes in the phylogeny (Fig. 2), and using these call variables we synthesized our estimates of the ancestral calls at each node (Fig. 1). There is, of course, no way to determine whether there were ancestors that ever produced calls similar to our estimates (our reference to 'ancestral calls', below, does not suggest such confidence but is merely a shorthand for 'our estimates of calls at ancestral nodes'). This approach, however, does yield a parsimonious estimate of the pathways by which these multivariate characters might have evolved.

We conducted phonotaxis experiments with female túngara frogs similar to those using extant heterospecific calls. Females show statistically significant discrimination between the conspecific call and all other calls, with one exception; they did not discriminate between the conspecific call and the call of their immediate ancestor. Thus, the range of female call preferences exceeds the variation of calls displayed by conspecific males.

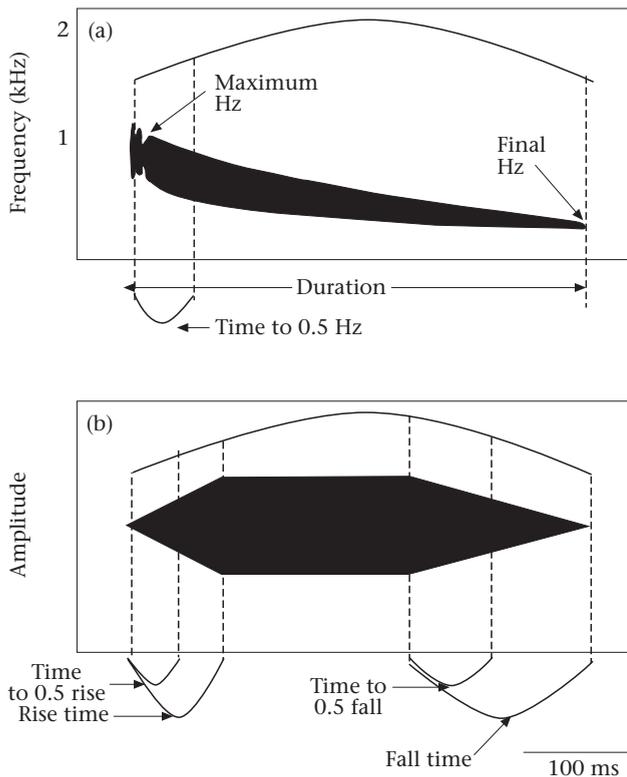
Furthermore, the strength of discrimination is better explained by phylogenetic distance than by overall call similarity, while the reverse is true of the strength of recognition. This suggests that the females' response to calls in the two contexts might rely on different weighting of call parameters.

## METHODS

### Call Characters

Conspecific and heterospecific advertisement calls were recorded during other studies, and details are presented there (Ryan & Rand 1993a, b, c). We used a variety of analog and digital techniques to measure 12 call variables that then were reduced to the following eight variables (Fig. 2, Table 1).

(1) Maximum frequency. In most calls this is the frequency of the onset of the frequency-modulated whine. In some individuals, the whine exhibits a rapid and slight increase in frequency, thus the initial frequency might be slightly lower than the maximum



**Figure 2.** Illustration of call measures made. (a) Sonagram, representing how frequency changes with time; (b) oscillogram, representing how amplitude changes with time. The lines demarcate portions of the call that correspond to various measures.

frequency. In call synthesis we began all calls at the maximum frequency.

(2) Final frequency. This the whine's frequency at the end of the call and is always the lowest frequency.

(3) Duration. The duration of the entire call.

(4) Rise time. The time from the call onset to its maximum amplitude.

(5) Fall time. The time from the call's maximum amplitude to the end of the call.

(6) Whine shape. The proportion of the call's duration from the onset to its mid-frequency.

(7) Rise shape. The proportion of the call's duration from the onset of the call to one-half the maximum amplitude during the rise.

(8) Fall shape. The proportion of the call's duration from the maximum amplitude to one-half the maximum amplitude during the fall.

### Signal Synthesis and Female Phonotaxis

We synthesized stimuli using the call parameters described above and listed in Table 1 at a sampling rate of 20 kHz on an Amiga computer using a program made available by J. Schwartz. Calls were broadcast from the computer through an amplifier and ADS L200C speakers. Calls were presented antiphonally at a rate of 1 call/2 s. Speakers were placed directly opposite one another 3 m apart in a small room in which the walls were covered

with foam to reduce acoustic reverberation. Stimuli were broadcast such that the peak intensity of the whines were 82 dB SPL (re 20  $\mu$ Pa) at the site of the female. We placed a female equidistant between the speakers under a funnel. After 2 min, the female was released and we noted a response if she approached to within 10 cm of a speaker within 15 min. Females were usually tested in more than one test, but previous analysis had shown no order effect.

We tested females in two kinds of phonotaxis experiments, discrimination ('discriminate: to make or constitute a difference in or between', *The Compact Oxford English Dictionary*, 2nd edn, 1993) and recognition ('recognize: to treat as valid', *The Compact Oxford English Dictionary*, 2nd edn, 1993). During discrimination tests we presented females with a choice between a conspecific and a heterospecific signal; 'heterospecific' includes ancestral calls. During recognition tests, we presented females with a heterospecific call versus noise to determine whether female túngara frogs mistakenly recognize a heterospecific call as indicating an appropriate mate (show phonotaxis towards that call). Our use of these terms here is consistent with Ryan & Rand (1993c) and Ryan & Rand (1995), although in the former we used the term 'preference' instead of 'discrimination'. These terms are also consistent with some uses in psychophysics in which identification is analogous to recognition: '*Discrimination* requires a subject to tell apart stimuli presented in pairs (by indicating whether they are the same or different). *Identification* requires the subject to categorize individuals using labels...' [italics in original] (Harnad 1987, page 3). We realize that this dichotomy between discrimination and recognition is not absolute, and that the use of these terms might differ among fields.

In the discrimination experiments, we tested the null hypothesis of no female preference with a two-tailed exact binomial probability. In the recognition experiments, we determined empirically the null hypothesis for the female's response to a heterospecific call versus noise. We used the number of times a female approached a silent speaker as an estimate of the number of random approaches to a speaker without reference to the stimulus being broadcast. Eighteen of 20 females showed no response and two came into contact with the silent speaker. Thus, if females ignore the heterospecific call as a communication signal, the null expectation is that they would still approach the call by chance in the ratio of 2:18. We compared this expectation to the actual responses with a Fisher's exact test. We recorded 'no response' in the recognition experiments if the female either approached the speaker broadcasting noise or did not exhibit phonotaxis after 15 min. In the latter case these data were included only if the female subsequently responded to the conspecific call, suggesting that her previous lack of phonotaxis was due to lack of a meaningful stimulus rather than lack of motivation.

Females were collected at choruses within hours prior to testing in Gamboa, Panama. Females usually are only found at a chorus when they are ready to mate, and most females were in amplexus (i.e. clasped by a male) when

**Table 1.** Call parameters used in synthesis of phonotaxis stimuli under different assumptions of evolutionary history, and female responses to those stimuli in phonotaxis experiments

Call stimulus/ species	Call								Female response	
	Frequency (Hz)		Time (ms)			Shape			HVC	HVN
	Max	Final	Duration	Rise	Fall	Whine	Fall	Rise		
<i>Physalaemus</i>										
<i>pustulosus</i>	884	484	369.7	24.0	342.8	0.33	0.49	0.33		
<i>petersi</i>	1220	384	246.1	13.7	230.3	0.11	0.79	0.84	0	4
<i>coloradorum</i>	1180	628	209.3	53.4	161.7	0.39	0.71	0.44	1	11
<i>pustulatus</i>	964	676	206.0	99.5	104.3	0.43	0.49	0.95	0	2
<i>caicai</i>	888	444	394.5	105.1	293.7	0.29	0.68	0.66	3	10
<i>ephippifer</i>	944	576	266.4	83.5	177.4	0.53	0.66	0.47	1	7
<i>roraima</i>	876	460	339.1	94.6	251.6	0.47	0.72	0.60	1	16
<i>enesefae</i>	976	692	745.7	301.5	445.7	0.52	0.54	0.55	0	0
Assumptions										
Node a										
TLSG	910	518	302.0	89.0	213.0	0.51	0.69	0.54	1	3
Node b										
TLSG	949	622	568.0	216.0	353.0	0.51	0.58	0.55	1	1
LSG	937	589	483.0	176.0	308.0	0.51	0.60	0.54	0	0
Node c										
TLSG	974	466	333.0	32.0	300.0	0.29	0.57	0.53	14	15
LSG	1050	434	309.0	19.0	287.0	0.25	0.61	0.53	7	17
Node d										
TLSG	1120	439	274.0	44.0	230.0	0.21	0.75	0.73	4	13
LSG	1015	507	297.0	59.0	238.0	0.30	0.63	0.69	5	9
LSP	1008	479	314.0	55.0	259.0	0.28	0.64	0.67	3	11
LSG*	1015	507	297.0	59.0	238.0	0.30	0.63	0.69	6	13
LSP*	1008	479	314.0	55.0	259.0	0.28	0.64	0.67	8	19
SG	998	526	339.0	85.0	254.0	0.36	0.62	0.62	10	16
SP	995	512	372.0	92.0	280.0	0.36	0.62	0.60	11	19
Node e										
TLSG	961	572	320.0	94.0	230.0	0.32	0.67	0.69	3	13
LSG	988	564	287.0	90.0	200.0	0.34	0.65	0.72	0	10
LSP	962	527	320.0	94.0	229.0	0.32	0.67	0.69	0	14
LSG*	988	564	287.0	90.0	200.0	0.34	0.65	0.72	4	9
LSP*	962	527	320	94	229	0.32	0.67	0.69	6	16
SG	992	547	311	88	225	0.35	0.64	0.68	5	16
SP	974	521	339	93	249	0.34	0.65	0.67	3	15
Node f										
TLSG	1072	652	208	76	133	0.41	0.62	0.78	2	5
LSG	1062	654	208	78	130	0.41	0.61	0.79	1	4
Root										
TLSG	961	545	451	125	326	0.43	0.57	0.54	2	8
LSG	995	528	345	89	257	0.37	0.63	0.62	1	12
LSP	977	554	448	140	308	0.44	0.60	0.57	6	12
LSG*	990	510	298	66	233	0.34	0.64	0.65	0	12
LSP*	955	506	308	73	235	0.40	0.66	0.59	3	12
SG	995	528	345	89	257	0.37	0.63	0.62	3	19
SP	977	554	448	140	308	0.44	0.60	0.57	8	18

Definitions of categories and call parameters: whine shape: the proportion of call duration when the frequency sweep reaches midfrequency; fall shape: the proportion of the call duration when the call reaches half the amplitude from the peak amplitude to the end of the call; rise shape: the proportion of the call duration when the call reaches half the amplitude from the initial to the peak amplitude. Female responses: HVC: the number of female responses to the heterospecific call when presented in concert with a conspecific call (maximum  $N=20$ ); HVN: the number of female responses to the heterospecific call when presented in concert with white noise (maximum  $N=20$ ).

Assumptions: Local (L), squared-change (S) parsimony; gradual (G), punctuated (P) evolution; pectinate tree (T).  
\*Without *P. enesefae*.

collected. After testing, we toe-clipped females and released them within 12 h of capture, allowing them the opportunity to nest. Recaptured females were not tested again in the same experiment.

We tested female discrimination in 35 separate experiments in which the conspecific call was presented in concert with a heterospecific (including ancestral) call; this included a control stimulus, in which two conspecific

calls were presented simultaneously. The sample size was 20 in each experiment, thus there were 700 choices in total. Similarly, we tested recognition in 35 separate experiments in which the heterospecific call (and the conspecific as a control) was paired with a white-noise stimulus; the sample sizes were the same as in the discrimination experiments for a total of 700 choice tests in the recognition experiments. Thus we report here the results of 70 separate experiments involving a total of 1400 female choices. In the analyses of female responses, we excluded the two experiments in which túngara frogs were tested with only conspecific calls (discrimination tests) or the conspecific call versus noise (recognition tests).

### Character Reconstruction

Estimates of ancestral calls in the previous study (Ryan & Rand 1995) were derived from a phylogeny with all three outgroup species, one tree topology (the most parsimonious one; Cannatella et al. 1998), one algorithm for estimating ancestral values (local squared-change parsimony; Maddison 1991), and one model of evolution (gradual). We conducted further analyses to estimate calls at ancestral nodes in which the following parameters were varied.

#### Tree topology

We explored the effects of using two different tree topologies, one having two clades within the species group, and another having a pectinate structure within the species group (Fig. 3). Both the initial hypothesis of the relationships of the *P. pustulosus* species group (Cannatella & Duellman 1984), and the hypothesis currently considered to be the most parsimonious (Cannatella et al. 1998) suggest two monophyletic groups within the species group. A previous preliminary analysis suggested that a pectinate tree within the species group was the most parsimonious phylogenetic hypothesis (Fig. 3; see also Ryan & Rand 1993b). Here we determine whether these differences in tree topology would influence our interpretations regarding the degree of responsiveness to ancestral calls.

#### Algorithm to estimate ancestral values

In addition to local squared-changed parsimony, we also used squared-change parsimony to estimate the quantitative call characters at ancestral nodes (Maddison 1991). Martins (1996a) suggested the latter was a more appropriate analysis for this purpose. We used a program provided by Martins & Garland (1991a, b) to estimate the nodal values of the call characters. The fact that we did not use other algorithms such as linear parsimony does not suggest we considered them inappropriate (see Discussion).

#### Tempo of evolution

Although we feel a gradual tempo of evolution is most appropriate, one could argue that a punctuated tempo might better apply to changes in species recognition

systems. Gradual evolution is modelled by scaling branch lengths to the hypothesized number of DNA base changes, and punctuated evolution is modelled by having all branch lengths equal (branch lengths are presented in Figure 1 in Ryan & Rand 1995). We contrasted the two tempos to determine how robust our previous conclusions are to varying this parameter.

### Outgroups

One of the outgroups, *P. enesefae*, has a call that differs substantially from most other calls in this study. To remove undue bias due to this single datum, we repeated the analysis without this call (as suggested by D. McLennan, personal communication).

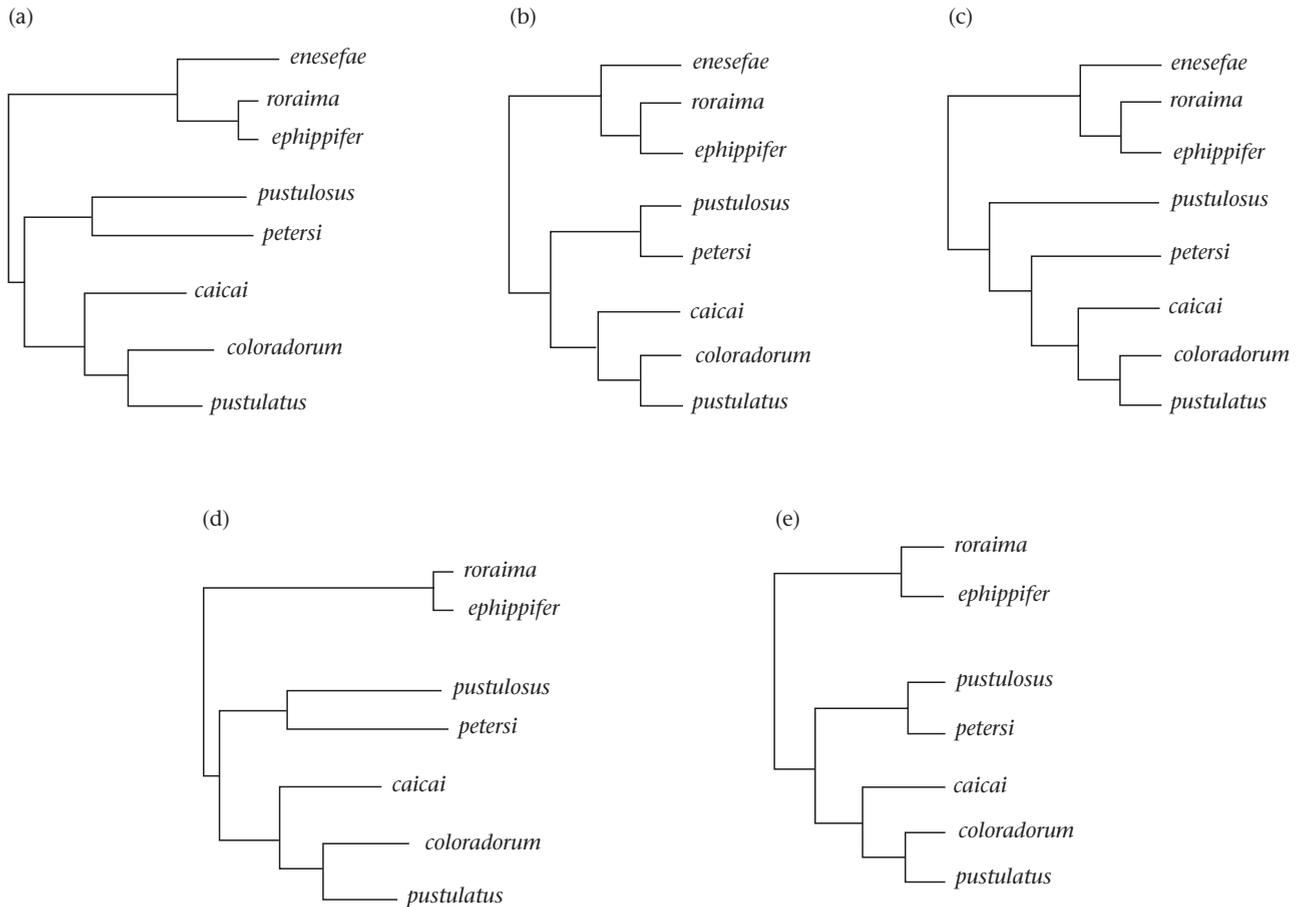
We used seven different models (i.e. unique combinations of algorithm, mode, tree topology and outgroups) to estimate the eight quantitative call characters at the seven ancestral nodes (in the model in which *P. enesefae* was omitted, there were only six ancestral nodes; Fig. 3). This gave a total of 384 call variables estimated for the ancestral nodes. We used an arbitrary criterion of 10% to determine what call estimates were sufficiently different to warrant testing female responses to that call. Thus if the estimates of calls at any ancestral node differed by more than 10% in any of the eight call parameters, these calls were tested in female phonotaxis experiments. The results of these experiments allowed us to determine whether different call estimates that were derived from making different evolutionary assumptions also resulted in call differences that were meaningful to females. Table 1 shows the call parameters for all of the calls at each of the nodes that were considered to be different enough from each other to warrant further testing in female phonotaxis experiments (see also Fig. 4).

### Statistical Analysis

All statistical analyses were conducted with SYSTAT (Wilkinson 1991). To determine if females responded differently to different calls estimated for the same ancestral nodes we compared all the phonotaxis results at each node with a *G* test.

We determined overall call similarity by first computing a principal component analysis (PCA) on the standardized (i.e. mean=0, standard deviation =1) call variables. Using the first three axes of variation from the PCA, we then computed the Euclidean distances among calls.

The phylogenetic distances between *P. pustulosus* and other taxa and nodes was based on the most parsimonious estimate of changes in DNA base sequence. We used a multiple regression analysis to determine the degree to which call similarity and phylogenetic distance predicted the strength of female responses in discrimination and recognition tests. To determine the contribution of these variables, we first stepped one variable and then the other out of the regression model. We also used a Pearson's product moment correlation to calculate the relationship between various pairs of variables.



**Figure 3.** The various evolutionary models used to estimate calls at ancestral nodes. Variables include tree topology, presence or absence of the outgroup species *P. enesefae*, branch length (trees with branch lengths equal represents an assumption of punctuated evolution), branch lengths that differ indicate a gradual model of evolution), and the algorithm used to calculate the characters at ancestral nodes. (a) LSG, SG; (b) LSP, SP; (c) TLSG; (d) LSG\*; (e) LSP\* (refer to Table 1 for abbreviations for different models).

## RESULTS

### Ancestral Call Estimates

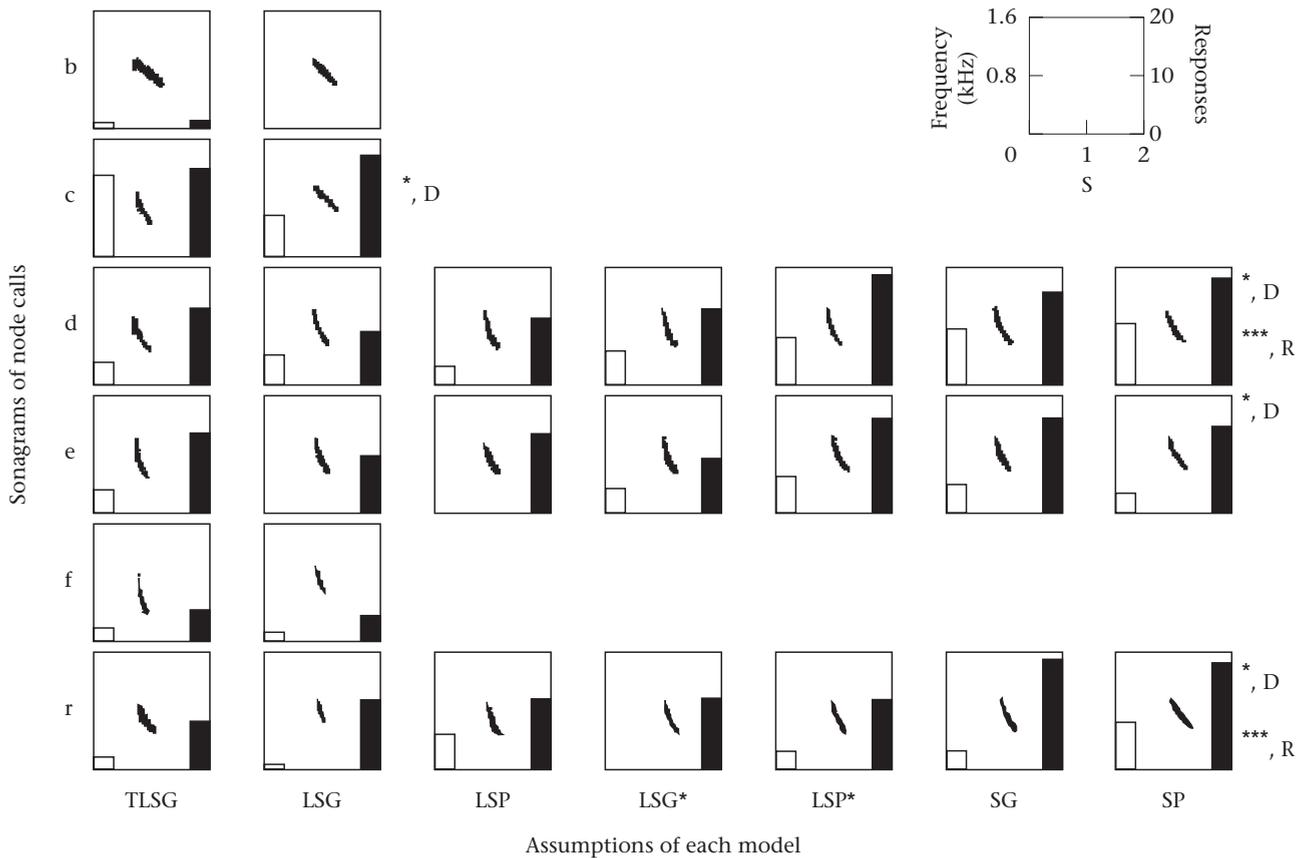
We estimated the eight call variables under study using seven different models of evolution for all seven ancestral nodes, except for the model in which *P. enesefae* was excluded, which contained only six nodes (Fig. 3). This resulted in 48 estimates of ancestral calls, which included our earlier estimates of one call for each of the seven ancestral nodes (Ryan & Rand 1995; local squared-change parsimony; 'LSG' in Table 1 and Fig. 4). We then compared the 41 estimates from this study to these seven estimates from our previous study. Of these 41 estimates, 21 (51%) differed from the previous estimate by at least 10% in one of the eight call parameters being estimated. We used these 21 ancestral calls in tests of female discrimination and recognition (Table 1; Fig. 4).

Ancestral nodes differed in the degree to which their call estimates were influenced by the different models we used (Table 1). The call estimates at the deepest nodes were the least robust to varying the models; for the root, node d, and node e, each of the seven models gave call estimates in which at least one call parameter differed by 10%. For the other four nodes (a, b, c, f), all seven models

gave the same estimate (within 10% for all call parameters) or only one model gave an estimate that differed by our criterion.

### Variation in Call Estimates and Meaningful Differences to Females

The fact that different models of character estimation yielded different call estimates does not show, by itself, that these call differences are meaningful to females. Thus we tested female responses in discrimination and recognition tests for calls that differed by our criterion from previous call estimates at the same node (Table 1, Fig. 4). There were multiple call estimates that differed at all of the nodes except for node a. For the other six nodes, we determined female phonotactic responses to all of the call estimates at a single node in both the discrimination tests and the recognition tests (Fig. 4). In six of the 12 sets of experiments there were significant differences in female responses to multiple call estimates at the same node. In the discrimination experiments, females responded differently to calls at node c ( $G=5.0$ ,  $P=0.025$ ), node d ( $G=12.6$ ,  $P=0.049$ ), node e ( $G=17.6$ ,  $P=0.007$ ), and the root ( $G=18.7$ ,  $P=0.004$ ). In the recognition experiments,



**Figure 4.** Sonograms of the calls of the various nodes (rows; see Fig. 1) estimated from different models of evolution (columns; abbreviations are defined in Table 1) and the number of females responding to each call when it was presented in comparison with the túngara frog call (□); and a white-noise stimulus (■) (all  $N=20$ ). Asterisks indicate significant difference ( $G$  test) in the female responses to various calls at the same node during discrimination (D) and recognition (R) tests: \* $P<0.05$ , \*\* $P<0.01$ , \*\*\* $P<0.001$ .

females responded differently to calls at node d ( $G=24.8$ ,  $P<0.001$ ), and the root ( $G=23.1$ ,  $P=0.021$ ; Fig. 4). All but one of these differences were at the nodes where the call estimates differed for all seven models of estimation. These differences in female response were more likely to be exhibited in the discrimination tests (four) than in the recognition tests (two).

Most of the conclusions reached in our previous study dealt with patterns of female responses among calls. The fact that females showed significantly different responses to call estimates at a single node may or may not influence our interpretations about patterns of female response. This is discussed below. One conclusion of our previous study, however, was based upon the female response to the ancestral call at a single node: the immediate ancestor of *P. pustulosus* and *P. petersi*, node c. The results of our previous study showed that females did not significantly discriminate between the conspecific call and the call estimated for the immediate ancestor (13 to conspecific versus seven to ancestral call; two-tailed exact binomial probability  $P=0.263$ ). The seven different models of evolution gave two different call estimates for this node. The pectinate tree gave one estimate, and the other six models, which varied in some assumptions but shared the same tree topology, gave another. The results of the  $G$  test, when not adjusted for experimentwide

error, suggest that the females responded to these two estimates differently in the discrimination test. Using the estimate from the pectinate tree, 14 females were attracted to the ancestral call, while six were attracted to the conspecific call ( $P=0.115$ ). Although the strength of female response differed between the two experiments, the results of both experiments failed to reject the hypothesis of statistically significant discrimination. We do not argue strongly for a true lack of discrimination when  $P=0.115$ . However, the trend in preference was actually in the opposite direction predicted: more females were attracted to the call estimate for the immediate ancestor than to the conspecific call. Thus, either females did not discriminate, or they showed a weak preference for the ancestral call over the conspecific call.

### History and Call Similarity as Predictors of Female Responsiveness

We determined the degree to which two variables predicted female performance in the discrimination and recognition tests; each variable was estimated as the total number of female responses to the call. We determined the role of the phylogenetic distance between the túngara frog and each node or taxon from which the competing call was derived, as well as the overall call similarity

between the túngara frog call and competing calls as predictors of the strength of discrimination and recognition. Phylogenetic distance and call similarity were not significantly correlated with one another (Pearson's product moment correlation:  $r=0.206$ ,  $N=35$ ,  $P=0.234$ ).

A multiple regression model in which female response in the discrimination experiments was the dependent variable and phylogenetic distance and call similarity were the independent variables was statistically significant (adjusted  $r^2=0.452$ ,  $F=15.04$ ,  $P<0.001$ ). When call similarity was stepped out of the model, phylogenetic distance still explained a significant portion of the variation in female discrimination ( $r^2=0.393$ ,  $F=21.39$ ,  $P<0.001$ ). When phylogenetic distance was removed from the model, call similarity significantly predicted female discrimination but much less so than phylogenetic distance alone ( $r^2=0.181$ ,  $F=7.28$ ,  $P=0.011$ ).

Phylogenetic distance and call similarity also explained a significant portion of the variation in the strength of female recognition in a multiple regression model (adjusted  $r^2=0.544$ ,  $F=21.26$ ,  $P<0.001$ ). Each variable alone predicted a significant and similar amount of the variation in the strength of female recognition (phylogenetic distance:  $r^2=0.380$ ,  $F=20.26$ ,  $P<0.001$ ; call similarity:  $r^2=0.307$ ,  $F=14.61$ ,  $P=0.001$ ).

There was a significant correlation between the females' responses to the same heterospecific call in the discrimination and recognition experiments ( $r^2=0.386$ ,  $N=36$ ,  $P<0.001$ ), although 61% of the variation in female responses was not explained by this correlation.

## DISCUSSION

The purpose of this study was to determine the degree to which our inferences about the influence of history on acoustic mate recognition are sensitive to variation in estimates of ancestral calls derived from different evolutionary models. We approached this problem in three steps by asking the following. (1) How often do different models result in significantly different call estimates at the same node? (2) Are these different call estimates at the same node meaningful to females, and if so, (3) do these meaningful differences alter our previous interpretations? The short answers to these three questions are: (1) usually; (2) sometimes; (3) no. We elaborate below.

It is clear that estimates of call variables at ancestral nodes are quite sensitive to the different models of call evolution. The seven models generated calls that differed at the same nodes more than half of the time. The number of estimates that differed by our arbitrary criterion of at least 10% for any of the eight call variables, differed drastically among the nodes, from all estimates being essentially the same to all estimates being different. The sensitivity of the nodes to different models was not randomly distributed across the phylogeny. In general, the number of descendants per node was indicative of the variability in estimates of the call for that node. For nodes a, b, c and f, the seven models gave only one (node a) or two (nodes b, c, f) estimates, and these nodes had either two (nodes a, c and f) or four (node d) descendants. For nodes d, e, and the root, the seven models gave seven

different call estimates, and these nodes had eight, four and 15 descendants, respectively. These variations in estimates by different models suggest a general caution: the more distant in history the target of the character estimates, the more circumspect we should be.

None of our models used linear parsimony. There might have been an advantage in doing so. Some suggest that this is a superior method for estimating quantitative ancestral characters (Maddison & Maddison 1992). Also, squared-change and local squared-change parsimony might be more likely to give estimates similar to one another than either might to linear parsimony. As new methods for estimating ancestral characters are developed, the túngara frog mate recognition system can serve as one means of evaluating such algorithms with a behavioural assay.

The second question we addressed is whether the significantly (by our 10% criterion) different call estimates for the same node lead to meaningful differences to females. It is crucial that this difference between statistical and biologically meaningful variation be addressed. There are many examples, including our own studies of túngara frog call preferences (e.g. Rand et al. 1992; Wilczynski et al. 1995), in which substantial stimulus variation has no effect on an animal's behavioural response. This phenomenon is central to the concept of the sign stimuli in which Tinbergen (1963) and other ethologists clearly demonstrated that only a portion of a signal might be involved in releasing a behavioural response in the receiver. Furthermore, animal responses to signal variation are based on just-meaningful rather than just-noticeable differences (Nelson & Marler 1990); the former being differences in signals that release different behavioural responses (including no response) while the latter refers to differences that can be detected by the sensory system. In this study we found that less than half of the statistically different call estimates translated into meaningful differences. This is a conservative estimate because it represents the number of sets of experiments (i.e. type of phonotaxis experiment-by-node, see Fig. 4) in which the strength of the female response in at least one of the experiments differed from the others.

Finally, we asked whether the meaningful differences in female responses to statistically different estimates of calls at the same nodes would cause us to alter interpretations of our previous study. There were three such major interpretations: (1) the range of female preferences exceeds the known variation of the conspecific call; (2) females do not discriminate between the conspecific call and the call of their most recent ancestor; and (3) the stimulus features that influence a female's response may be context dependent, differing if females are being asked to discriminate or recognize stimuli.

We found that the results of this expanded study of female preferences for ancestral calls were generally consistent with all of the major conclusions from our earlier study. In this study, as well as in our earlier study, we found that there are a number of heterospecific (including ancestral) signals that elicit substantial, and in many cases, statistically significant responses. Furthermore, we found that the range of preferences usually exceeds the

known conspecific call variation. These responses can be viewed as sensory biases which could promote the evolution of male signals without invoking the more complicated notions of linkage disequilibrium and genetic correlations that are necessary assumptions for preference–trait evolution under runaway sexual selection and good genes models (e.g. West-Eberhard 1979; Ryan 1990, 1997; Kirkpatrick & Ryan 1991; Endler 1992; Guilford & Dawkins 1993; Christy 1995; Shaw 1995). Thus, these results further confirm a general conclusion of many other empirical studies that suggest an incongruence between extant male traits and female preferences can generate sexual selection (e.g. Basolo 1990, 1995; Christy & Salmon 1991; Proctor 1991, 1992; Fleishman 1992; Searcy 1992; Weary et al. 1993; McClintock & Uetz 1996).

This study also confirms the result from our previous study that females do not discriminate between the conspecific call and the call at node c. In fact, in this study the majority of the females responded to the ancestral call over the conspecific call. Thus, this experiment is almost the exception that proves the rule. This result is interesting in the context of studies that have shown an asymmetry in mate recognition between extant species or populations. In fruit flies (Kaneshiro 1980, 1983), swordtails (Ryan & Wagner 1987; McLennan & Ryan 1997), and salamanders (Arnold et al. 1996) there is an asymmetry in the degree to which one species or population is attracted to another. Such an asymmetry can occur if there are differences in the rates in which preferences and traits evolve, or if, as Kaneshiro suggested, a species pair consists of an extant ‘ancestral’ species and its descendant or daughter species. This study of túngara frogs shows that the evolution of calls and preferences have not closely tracked one another. Ideally, to gain further insights into the dynamics of signal–receiver evolution in this part of the clade we would like to examine the preferences of *P. petersi* for their own conspecific call versus both the túngara frog call and the node c call.

The third conclusion from the previous study is that females might be using different decision rules to govern their responses to heterospecific calls depending on whether these calls are contrasted with a conspecific call or a heterospecific call. In both studies there was a significant correlation between the females’ responses to the same heterospecific call in the discrimination and the recognition experiments, although these correlations left unexplained 67 and 61% of the variation in the previous and current studies, respectively. In the previous study, phylogenetic distance but not call similarity significantly predicted the strength of response in the discrimination experiments, and only call similarity explained female response in the recognition tests. The results in this study were similar but not identical. In the discrimination tests, phylogenetic distance was a much better predictor of female responses, but both phylogenetic distance and call similarity explained a similar amount of variation in female recognition. What we refer to as discrimination and recognition tests have both been used independently in other studies to test the strength of female preferences for male signals. To our knowledge, the possibility has

not been raised that these experiments might in fact be testing different perceptual processes. If this is the case, it would have obvious implications for the proximate bases of mate choice. But there would also be serious implications for ultimate explanations as well, such as whether different preference functions for discrimination versus recognition might be predicted by the costs of females making type I versus type II errors in different situations. For example, if an ovulating female is presented with a heterospecific call and noise, would her thresholds for making a type II error (mistakenly accepting a heterospecific) be lowered if ovulation placed a premium on quickly obtaining a mate?

In general, we feel that this study, combined with our previous study (Ryan & Rand 1995), indicates that using explicit phylogenetic approaches to study the history of animal communication systems provides a number of insights into how such systems evolve. These phylogenetic methods, however, are quite sensitive to the assumptions employed; different models can give different estimates of ancestral characters. To evaluate the impact of different models, it is important not only to determine whether different assumptions result in statistically different estimates of the same character, but whether these differences are biologically meaningful. Only if the latter is determined, can one evaluate how such meaningful differences might alter interpretations of behavioural evolution. As this study should indicate, this is a tractable but an onerous task.

### Acknowledgments

This study was supported by grants from the National Science Foundation (IBN-9316185) and the Smithsonian Institution’s Scholarly Studies Program. We thank E. Martins and D. McLennan and an anonymous referee for comments on the manuscript, and the many assistants who conducted the phonotaxis experiments.

### References

- Arnold, S. J., Verrell, P. A. & Tilley, S. G. 1996. The evolution of asymmetry in sexual isolation: a model and a test case. *Evolution*, **50**, 1024–1033.
- Basolo, A. L. 1990. Female preference predates the evolution of the sword in swordtail fish. *Science*, **250**, 808–810.
- Basolo, A. L. 1995. A further examination of a pre-existing bias favouring a sword in the genus *Xiphophorus*. *Animal Behaviour*, **50**, 365–375.
- Brooks, D. R. & McLennan, D. A. 1991. *Phylogeny, Ecology and Behaviour*. Chicago: University of Chicago Press.
- Brooks, D. R., McLennan, D. A., Carpenter, J. M., Weller, S. G. & Coddington, J. A. 1995. Systematics, ecology and behaviour. *Bioscience*, **45**, 687–695.
- Brown, J. L. 1982. The adaptationist program. *Science*, **217**, 884–885.
- Cannatella, D. C. & Duellman, W. E. 1984. Leptodactylid frogs of the *Physalaemus pustulosus* species group. *Copeia*, **1984**, 902–921.
- Cannatella, D. C., Hillis, D. M., Chippendale, P. T., Weigt, L., Rand, A. S. & Ryan, M. J. 1998. Phylogeny of frogs of the *Physalaemus pustulosus* species group, with an examination of data incongruence. *Systematic Biology*, **47**, 311–355.

- Christy, J. H. 1995. Mimicry, mate choice and the sensory trap hypothesis. *American Naturalist*, **146**, 171–181.
- Christy, J. H. & Salmon, M. 1991. Comparative studies of reproductive behaviour in mantis shrimp and fiddler crabs. *American Zoologist*, **31**, 329–337.
- Cocroft, R. C. & Ryan, M. J. 1995. Patterns of mating call evolution in toads and chorus frogs. *Animal Behaviour*, **49**, 283–303.
- Endler, J. A. 1992. Signals, signal conditions and the direction of evolution. *American Naturalist*, **131**, S125–S153.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *American Naturalist*, **125**, 1–15.
- Fleishman, L. J. 1992. The influence of the sensory system and the environment on motion patterns in the visual displays of anoline lizards and other vertebrates. *American Naturalist (Supplement)*, **139**, 36–61.
- Gould, S. J. & Vrba, E. S. 1982. Exaptation: a missing term in the science of form. *Paleobiology*, **8**, 4–15.
- Guilford, T. & Dawkins, M. S. 1993. Receiver psychology and the design of animal signals. *Trends in Neuroscience*, **16**, 430–436.
- Harnad, S. 1987. *Categorical Perception: The Groundwork of Cognition*. Cambridge: Cambridge University Press.
- Harvey, P. H. & Pagel, M. D. 1991. *The Comparative Method in Evolutionary Biology*. Oxford: Oxford University Press.
- Höglund, J. 1989. Size and plumage dimorphism in lekking birds: a comparative analysis. *American Naturalist*, **134**, 72–87.
- Huey, R. B. & Bennett, A. F. 1987. Phylogenetic studies of coadaptation: preferred temperatures versus optimal performance temperatures of lizards. *Evolution*, **41**, 1098–1115.
- Kaneshiro, K. Y. 1980. Sexual isolation, speciation and the direction of evolution. *Evolution*, **34**, 437–444.
- Kaneshiro, K. Y. 1983. Sexual selection and direction of evolution in the biosystematics of Hawaiian *Drosophila*. *Annual Review of Entomology*, **28**, 161–178.
- Kirkpatrick, M. & Ryan, M. J. 1991. The paradox of the lek and the evolution of mating preferences. *Nature*, **350**, 33–38.
- Lanyon, S. M. 1992. Interspecific brood parasitism in blackbirds: a phylogenetic perspective. *Science*, **255**, 77–79.
- Lorenz, K. 1941. Comparative studies of the motor patterns of Anatinae (translated by R. Martin 1971). *Studies in Animal and Human Behaviour*, **2**, 14–18.
- Losos, J. B. 1994. An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. *Systematic Biology*, **43**, 117–123.
- Losos, J. B. & Miles, D. B. 1994. Adaptation, constraint, and the comparative method: phylogenetic issues and methods. In: *Ecological Morphology* (Ed. by P. C. Wainwright & S. M. Reilly), pp. 60–98. Chicago: University of Chicago Press.
- McClintock, W. J. & Uetz, G. W. 1996. Female choice and preexisting bias: visual cues during courtship in two *Schizocosa* wolf spiders (Araneae: Lycosidae). *Animal Behaviour*, **52**, 167–181.
- McLennan, D. A. & Brooks, D. R. 1991. Parasites and sexual selection: a macroevolutionary perspective. *Quarterly Review of Biology*, **66**, 255–286.
- McLennan, D. A. & Ryan, M. J. 1997. Responses to conspecific and heterospecific olfactory cues in the swordtail *Xiphophorus cortezi* (Pisces: Poeciliidae). *Animal Behaviour*, **54**, 1077–1088.
- Maddison, W. P. 1991. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. *Systematic Zoology*, **40**, 304–314.
- Maddison, W. P. & Maddison, D. R. 1992. *MacClade. Version 3*. Sunderland, Massachusetts: Sinauer.
- Martins, E. P. (Ed.) 1996a. *Phylogenies and the Comparative Method in Animal Behaviour*. Oxford: Oxford University Press.
- Martins, E. P. 1996b. Phylogenies, spatial autoregression, and the comparative method: a computer simulation test. *Evolution*, **50**, 1750–1765.
- Martins, E. P. & Garland, T. 1991a. *CMAP—Comparative Method Analysis Program*. Available at <http://evolution.uoregon.edu/~emartins/programs/cmap.html>.
- Martins, E. P. & Garland, T. 1991b. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. *Evolution*, **45**, 534–557.
- Martins, E. P. & Hansen, T. F. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist*, **149**, 646–667.
- Mitter, C., Farrell, B. & Futuyma, D. J. 1991. Phylogenetic studies of insect–plant interactions: insights into the genesis of diversity. *Trends in Ecology and Evolution*, **6**, 290–293.
- Nelson, D. A. & Marler, P. 1990. The perception of bird song and an ecological concept of signal space. In: *Comparative Perception* (Ed. by E. C. Stebbins & M. A. Berkeley), pp. 443–478. New York: J. Wiley.
- Olmstead, K. E. 1997. Correlated rates of molecular and morphological evolution. *Evolution*, **51**, 1381–1393.
- Proctor, H. C. 1991. Courtship in the water mite *Neurmania papillator*: males capitalize on female adaptations for predation. *Animal Behaviour*, **42**, 589–598.
- Proctor, H. C. 1992. Sensory exploitation and the evolution of male mating behaviour: a cladistic test using water mites (Acari: *Parasitengona*). *Animal Behaviour*, **44**, 745–752.
- Prum, R. O. 1990. Phylogenetic analysis of the evolution of display behaviour in neotropical manikins (Aves: Pipridae). *Ethology*, **84**, 202–231.
- de Queiroz, A. & Wimberger, P. H. 1993. The usefulness of behaviour for phylogeny estimation: levels of homoplasy in behavioural and morphological characters. *Evolution*, **47**, 46–60.
- Rand, A. S., Ryan, M. J. & Wilczynski, W. 1992. Signal redundancy and receiver permissiveness in acoustic mate recognition by the túngara frog, *Physalaemus pustulosus*. *American Zoologist*, **32**, 81–90.
- Reeve, H. K. & Sherman, P. W. 1993. Adaptation and the goals of evolutionary research. *Quarterly Review of Biology*, **68**, 1–32.
- Ryan, M. J. 1990. Sensory systems, sexual selection, and sensory exploitation. *Oxford Surveys in Evolutionary Biology*, **7**, 157–195.
- Ryan, M. J. 1996. Phylogenetics and behaviour: some cautions and expectations. In: *Phylogenies and the Comparative Method in Animal Behaviour* (Ed. by E. Martins), pp. 1–21. Oxford: Oxford University Press.
- Ryan, M. J. 1997. Sexual selection and mate choice. In: *Behavioural Ecology, An Evolutionary Approach* (Ed. by J. R. Krebs & N. B. Davies), pp. 179–202. Oxford: Blackwell.
- Ryan, M. J. & Rand, A. S. 1993a. Phylogenetic patterns of behavioural mate recognition systems in the *Physalaemus pustulosus* species group (Anura: Leptodactylidae): the role of ancestral and derived characters and sensory exploitation. *Linnean Society Symposium Series*, **14**, 251–267.
- Ryan, M. J. & Rand, A. S. 1993b. Sexual selection and signal evolution: the ghost of biases past. *Philosophical Transactions of the Royal Society, Series B*, **340**, 187–195.
- Ryan, M. J. & Rand, A. S. 1993c. Species recognition and sexual selection as a unitary problem in animal communication. *Evolution*, **47**, 647–657.
- Ryan, M. J. & Rand, A. S. 1995. Female responses to ancestral advertisement calls in the túngara frog. *Science*, **269**, 390–392.
- Ryan, M. J. & Wagner, W. E., Jr 1987. Asymmetries in mating preferences between species: female swordtails prefer heterospecific mates. *Science*, **236**, 595–597.
- Searcy, W. A. 1992. Song repertoire and mate choice in birds. *American Zoologist*, **32**, 71–80.

- Schluter, D., Price, T., Moores, A. Ø. & Ludwig, D. 1997. Likelihood of ancestral states in adaptive radiation. *Evolution*, **51**, 1699–1711.
- Shaw, K. 1995. Phylogenetic tests of the sensory exploitation model of sexual selection. *Trends in Ecology and Evolution*, **10**, 117–120.
- Swofford, D. L. & Maddison, W. P. 1987. Reconstructing ancestral character states under Wagner parsimony. *Mathematical Bioscience*, **87**, 199–229.
- Tinbergen, N. 1963. On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, **20**, 410–433.
- Weary, D. M., Guilford, C. & Weisman, R. G. 1993. A product of discriminative learning may lead to female preferences for elaborate males. *Evolution*, **47**, 333–336.
- West-Eberhard, M. J. 1979. Sexual selection, social competition, and evolution. *Proceedings of the American Philosophical Society*, **123**, 222–234.
- Wilczynski, W., Rand, A. S. & Ryan, M. J. 1995. The processing of spectral cues by the call analysis system of the túngara frog, *Physalaemus pustulosus*. *Animal Behaviour*, **49**, 911–929.
- Wilkinson, L. 1991. *SYSTAT: The System for Statistics*. Evanston, Illinois: SYSTAT.
- Wimberger, P. H. & de Queiroz, A. 1996. Comparing behavioural and morphological characters as indicators of phylogeny. In: *Phylogenies and the Comparative Method in Animal Behavior* (Ed. by E. Martins), pp. 206–233. Oxford: Oxford University Press.