# Phylogeny and biogeography of midwife toads (*Alytes*, Discoglossidae): a rebuttal

J.W. Arntzen[1] & M. García-París[2]

[1]*School of Biological Sciences, University of Wales, Bangor LL57 2UW, United Kingdom;* [2]*Museum of Vertebrate Zoology, University of California, 3101 Valley Life Sciences Building, Berkeley CA 94720, USA*

## Abstract

Three competing phylogenetic hypotheses for the genus *Alytes* (midwife toads) are evaluated. Based on quantitative coding of protein characters the most parsimonious solution shows a sister taxon relationship for *Alytes dickhilleni* and *A. muletensis*. The alternatives in which *A. obstetricans* has its sister group in either *A. dickhilleni* or *A. muletensis* lack support. Using calibrations derived from protein evolutionary rates, the vicariant events giving rise to *A. obstetricans* and the lineage leading to the *A. muletensis* and *A. dickhilleni* clade and the subsequent splitting between *A. muletensis* and *A. dickhilleni* cannot be placed much earlier than the Miocene-Pliocene boundary. Biogeographical scenarios invoking an earlier time of divergence should be rejected.

## Resumen

Se evaluan tres hipótesis filogenéticas alternativas sobre la evolución del género *Alytes* (sapos parteros). Utilizando una codificación cuantitativa para los datos de proteinas, la solucion mas favorecida de acuerdo con el criterio de máxima parsimonia, incluye a *Alytes dickhilleni* y *A. muletensis* como taxa hermanos. Las otras alternativas, en las que *A. obstetricans* aparece como grupo hermano de *A. dickhilleni* o, alternativamente, de *A. muletensis*, carecen del suficiente apoyo. De acuerdo con las diferentes propuestas de un reloj molecular para evolución de proteinas, los eventos vicariantes que dieron lugar a la separación de los linajes correspondientes a *A. obstetricans* y luego a los de *A. muletensis* y *A. dickhilleni*, no pueden localizarse mucho antes del límite Mioceno-Plioceno. Por lo tanto otros escenarios paleogeográficos que requieran una divergencia anterior son descartables.

## Introduction

Our taxonomic revision of midwife toads, genus *Alytes*, resulted in the recognition of four species and a number of subspecies (Arntzen & García-París, 1995). *Alytes muletensis* (Sanchíz & Adrover, 1979) is found on the island of Mallorca. *Alytes cisternasii* Boscá, 1879 and *A. obstetricans* (Laurenti, 1768) have contiguous distributions in continental western Europe with marginal range overlap in central parts of the Iberian Peninsula. Hybridization between the taxa is exceedingly rare (Rosa, 1995). *Alytes dickhilleni* Arntzen & García-París, 1995 is restricted to the southeastern corner of the Iberian Peninsula with no recorded overlap with *A. cisternasii* and *A. obstetricans*. Each species is reproductively isolated from the others. The Iberian subspecies of *Alytes obstetricans* on the other hand are neither geographically nor genetically isolated (Arntzen & García-París, 1995). We gathered morphological and protein genetic data for 12 populations representing all European taxa. We observe that the procedure of molecular (or palaeontological, or palaeogeographical) dating of evolutionary events is independent from the choice for phenetic or phylogenetic methods of tree reconstruction. Secondly, the study of descent has to be distinguished from the study of spatial variation. We focused on phenetic methods of classification when dealing with variation at the intraspecific level and on phylogenetic methods of tree construction when dealing with interspecific differentiation. However, both approaches could not be completely disentangled because we dealt with variation and descent across species, subspecies, and populations. Moreover, the specific status of *A. dickhilleni* was only elucidated as the result of our studies.

The instructive critique to our work by Altaba (1997) deals exclusively with matters of phylogeny and biogeography. No disagreement appears to exist about the recognition of four European species in the genus *Alytes*, nor is doubt expressed regarding the correct allocation of populations to species and subspecies. Consensus also exists about the phylogenetic position of *A. cisternasii* outside the *dickhilleni – muletensis – obstetricans* clade. However, Altaba challenges the phylogenetic hypothesis that considers *A. dickhilleni* to be the sister species of *A. muletensis*, and he outlines a biogeographical scenario different from ours. We will show that the alternative phylogenetic hypothesis lacks strong support and that the data coding scheme on which it is based does not behave well for the data at hand. The alternative biogeographical hypothesis is incompatible with documented rates of molecular evolution and, with no calibration it is only one out of an undetermined number of equally straightforward scenarios. Finally, we will outline the research currently undertaken for further testing scenarios on the evolutionary history of midwife toads.

Three phylogenetic hypotheses are under discussion. We follow a terminology in which 'solution 1' refers to *A. dickhilleni – A. muletensis*, 'solution 2' to *A. muletensis – A. obstetricans* and 'solution 3' to *A. dickhilleni – A. obstetricans* as sister taxa. Phylogenetic analysis was done by hand, or by using Jelly (Ellis, 1987) and PAUP 3.1.1 (Swofford, 1993).

**Phylogeny**

Fig. 3 of Arntzen & García-París (1995) in which *A. dickhilleni* and *A. muletensis* are sister species is taken by Altaba (1997) to represent a distance-Wagner tree and he suggests that using genetic distance measures in the tree construction implies uniform evolutionary rates along the lineages. This misrepresents our work, in particular when phylogenetic conclusions are perceived to be based on the application of phenetic methods. The figure in fact represents a character-Wagner tree. It was constructed employing the locus as a

character and it did not involve the assumption of equal rates of evolutionary change along lineages.

The essential difference between the analyses by Altaba and ourselves is not in the phenetic versus phylogenetic treatment of the data but resides in the procedure used for coding protein data prior to phylogenetic analysis. Altaba codes allelic character states as present or absent. This procedure is known as 'qualitative coding' (Buth, 1984). Qualitative coding has the virtue of rigidity but it does not cope well with small and varying sample sizes. For example, three apomorphic character states supporting solution 3 (*GP.M-1*[a], *GP.M-5*[a], and *GP.S-3*[b]) are derived from alleles observed in single specimens of *A. obstetricans*. We do not question the phylogenetic signal provided by these character states; rather we point out that many additional apomorphies may have been missed due to sampling error. As shown in Fig. 1, the magnitude of this effect is dependent on allele frequency and sample size. Average sample size was 26.3 for *A. obstetricans*, 16 for *A. cisternasii*, 10.7 for *A. dickhilleni* and 3.7 for *A. muletensis*. Hence, the probability of not observing an allele with a frequency of 0.1 was 46% in *A. muletensis*, 11% in *A. dickhilleni* and negligible in *A. obstetricans* and *A. cisternasii*. The sampling of *A. muletensis* alleles was even less effective than the figure indicates. Some toads may have been siblings because they were the offspring from a small number of adults in a captive breeding programme. Moreover, loss of genetic variation in populations will result from the combined processes of population bottlenecking and genetic drift. The loss of particular alleles will be most pronounced in small and isolated populations, such as found in *A. muletensis* and *A. dickhilleni*. Thus, ideally a larger sample of *A. muletensis* should be analysed (and more populations), in particular when choosing for qualitative coding of the protein data. However, *Alytes muletensis* is a rare and endangered species and sampling was done without any risk to affecting natural populations.

In coding the protein genetic data for phylogenetic analysis we employed a scheme in which Rogers' genetic distance was calculated between populations for individual loci. The software used
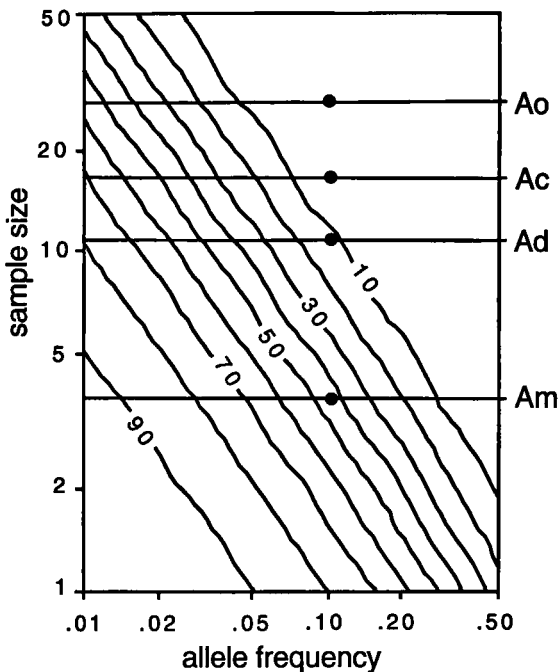
*Fig. 1.* Percent probability (*p*) of not observing an allele at a particular locus as a function of its frequency (*f*) and sample-size (*N*). Indicated are average sample size for *Alytes obstetricans* (Ao), *A. cisternasii* (Ac), *A. dickhilleni* (Ad) and *A. muletensis* (Am). Dots refer to an example discussed in the text. The dependent variable follows a binomial distribution: $p = 100 \times (1\text{-}f)^{2N}$ (Swofford & Berlocher, 1987). Note that the axes are scaled logarithmically.

to calculate single locus genetic distances was BIOSYS-1 (Swofford & Selander, 1981). Coding techniques that do include gene frequency information are known as 'quantitative' (Buth, 1984). The reason for using quantitative coding was not that we feel that frequency information itself is particularly helpful in uncovering phylogenetic relationships (Crother, 1990). Instead, the scheme was devised to give less weight to rare alleles such as those that would have been missed had the sampling effort been less. With quantitative coding rare alleles are not given undue influence in the phylogenetic analysis. The ingroup taxa Altaba considers sister species are the better sampled ones. We argue that this is an artefact, resulting from the equal weight given to synapomorphies representing rare and common alleles alike in a situation where sample sizes are small and varying. Our coding scheme in contrast deals

elegantly with sample-size limitations.

As Altaba (1997) points out, less than half of the characters we uncovered provide phylogenetic information at the species level. Re-analysing our data using Altaba's method (i.e., with qualitative coding), we prefer not to discard the data for locus *Icd-1* since the *d*-allele constitutes a synapomorphic character state (be it in favour of a solution preferred by neither Altaba nor ourselves) and we dismiss the interpretation in which information provided by the *Mpi-1* locus is taken to support solution 3 and not solution 1. (The reader is reminded that the allelic profile observed at this *Mpi*-locus was 'ab' for *A. dickhilleni*, 'b' for *A. muletensis*, 'acd' for *A. obstetricans* and 'e' for the outgroup *A. cisternasii*.) This multistate character we coded in PAUP as a 'polymorphism' rather than as an 'uncertainty', a decision that does not affect the outcome, neither in terms of relative parsimony, nor in the bootstrapping result. Thus, the number of characters compatible with solution 1, 2 and 3 respectively is not 6, 3 and 8 out of 17 (Altaba, 1997: table II), but 6–7, 4 and 7–8 out of 18, with minimum lengths for the corresponding trees of 31, 33 and 30 steps. Solution 3 displays maximum parsimony. However, with a bootstrap replication score of 56% the result is unconvincing. Our position, that we share with Berry & Gascuel (1996), is that if the (recoded) data contain little phylogenetic signal the bootstrap method will detect the flaw and lead to an irresolution, which is better than a false resolution. Hence, to avoid error, Altaba should not favour tree 3, he should favour no tree at all. His additional observation that tree 3 would coincide with current nomenclature constitutes a phenetic argument with which we have no affinity.

Following Murphy et al. (1983) in their strict cladistic approach, loci are considered phylogenetically uninformative if no alleles are shared between the outgroup and the ingroup (because in such a situation the plesiomorphic condition cannot be identified; cf. Watrous & Wheeler, 1981; Murphy, 1993). This is the case for five loci out of eighteen : *Nadhdh-2*, *GP.M-1*, *GP.M-2*, *GP.S-3*, and *Mpi-1*. Excluding these loci from the analysis, the number of characters compatible with

solution 1, 2 and 3 is 6, 3 and 4, respectively, with minimum lengths for the corresponding trees of 20, 23 and 22 steps. So, under the most stringent character selection regime that retains 13 loci and with qualitative coding, solution 3 is not the one showing maximum parsimony. It might be noted that five out of six characters supporting the most parsimonious solution (solution 1) were identified in our evolutionary classification by their synapomorphic character states (Arntzen & García-París, 1995, fig. 5: *Acph-2$^d$*, *Ak$^a$*, *α-Gly$^b$*, *Mpi-2$^a$*, *6-Pgd$^b$*; the allele *Me-2$^b$* was inadvertently shown as an autapomorphy for *A. muletensis* instead of as an synapomorphy for *A. dickhilleni* and *A. muletensis*.

In our phylogenetic reconstruction (Arntzen & García-París, 1995: fig. 3) we opted for the quantitative coding of data from 50 protein loci. Altaba (1997) states that the internodal distances separating *A. obstetricans*, *A. muletensis*, and *A. dickhilleni* are 'quite short' and hence provide 'little support' for solution 1. However, the issue at stake is not about relative branch lengths but about absolute branch lengths, or, in other words, whether or not there is sufficient evidence for recognizing *A. dickhilleni – A. muletensis* as a monophyletic group. Our work satisfies this criterion. Solution 1 is more parsimonious than the alternatives and with a bootstrap replication score of over 80% the phylogenetic hypothesis is strongly supported by the data (cf. Hillis & Bull, 1993).

We repeated the phylogenetic analysis with populations pooled into species and with data for 18 loci (see above) coded quantitatively. Tree 1 has a length of 27.5. This compares favourably with the lengths of 29.5 and 29.9 for tree 2 and 3, respectively. Lengths are expressed in Rogers' genetic distance per locus, subsequently summed over all loci. Optimization of solution 1 with the hyperboloid approximation procedure (Rogers, 1984; Ellis, 1987) results in a marginal improvement (gain 0.3) but involves the allocation to the internal branch of small amounts of plesiomorphic character state change, at the loci *GP.S-3* (length 0.1) and *Icd-1* (length 0.3). As before, the bootstrap analysis was run for 1000 permutations in PAUP and again solution 1 came out on top, with a bootstrap replication score exceeding 80%. Deleting the five loci for which the evolutionary polarity of character states cannot unambiguously be assigned, lengths are 17.8, 20.0 and 20.3 for the trees corresponding to solution 1, 2 and 3, respectively and solution 1 has a bootstrap score of > 80%. Once more we conclude that the monophyly of *A. dickhilleni – A. muletensis* is strongly supported by the data. Selection of tree 1 is not an artefact resulting from the inclusion of phylogenetically non-informative characters or from constraints put over the phylogeny by the inclusion of lower taxonomic units.

## Biogeography

Altaba describes our vicariant biogeographical model as 'an unlikely series of events', but surely the best reason to reject the model – or to reject any biogeographical reconstruction – would be lack of support for the phylogeny on which it is based. Leaving this matter as it stands, some of the objections to our scenario cannot be substantiated while Altaba's alternative has some problems that he fails to address.

Altaba proposes that large inland saline lakes in association with tectonic activity are account for speciation in allopatry of *A. cisternasii* and a proto-*A. obstetricans* at ca. 16 mY (Altaba, 1997). However, the available palaeogeographical data do not support a disruption of gene flow between two incipient *Alytes* lineages in western or northern Iberia (Altaba, 1997: fig. 2.2). The contemporary expansion of the northern Betic Sea Strait on the other hand constitutes an unambiguous vicariant event. Rosa (1995) agrees with our phylogenetic hypothesis, but, like Altaba (1997), positions the vicariant event between *A. cisternasii* and a proto-*A. obstetricans* on the Iberian Peninsula, to which he invokes a 'sea-arm' in the west of the Peninsula at 16–18 mY (Rosa, 1995: fig. 5.3), rather than 'inland saline lakes' in the northeast of the Peninsula (Salvador, 1974; Altaba, 1997). A problem with our biogeographical scenario – spotted by neither Rosa (1995) or Altaba (1997) – is how to explain the regression of *A. cisternasii* to the southwestern corner of the Ibe-

rian Peninsula and its adaptation to a fossorial mode of life. To identify the causes underlying such a process is almost beyond approach. Indeed, similar questions can be asked for almost any organism. The urodele *Chioglossa lusitanica* Bocage, 1864 is restricted to the northwestern corner of the Iberian Peninsula and shows an extreme adaptation to swift running mountain brooks (Arntzen, 1981). What could be the driving force for *C. lusitanica* to become so strongly specialized ?

Altaba wonders how an insular species could become successful on the continent. There is no compelling intrinsic reason why island-to-mainland colonizations should be less successful than the reverse combination, although they may be less numerous considering the generally small size of the source populations, and the presence of ecological competitors (such as perhaps *A. cisternasii* relative to *A. obstetricans*) may be more likely on continents than on islands (MacArthur & Wilson, 1967). Examples of successful island-to-mainland colonization include the frogs *Eleutherodactylus coqui* Thomas, 1965 and *Osteopilus septentrionalis* (Boulenger, 1882) into Florida from respectively Puerto Rico and Cuba (Conant & Collins, 1991) and the lizard *P. pityusensis* Boscá, 1883 into continental Spain from the Balearic Islands (Carretero et al., 1991; 1995). Considering that *Alytes cisternasii* was restricted to the southwestern corner of the Iberian Peninsula and with ample time available for dispersal, we see no problem with the scenario in which *A. obstetricans* expands its range from a small source.

Three speciation events leading to four extant species of midwife toads are palaeogeographically dated by Altaba as occurring around 16 mY, 14 mY and 8 mY. As can be shown from the data in table IV of Arntzen & García-París (1995), this would involve highly unequal rates of molecular evolution along lineages, differing by as much as a factor 12. This contradicts our knowledge on relatively constant rates of molecular evolution among closely related lineages (Britten, 1986; Martin & Palumbi, 1993). Also the dating is at odds with calibrations from the 'molecular clock' (Maxson & Maxson, 1979; Thorpe, 1982;

Beerli et al., 1996) according to which the two more recent speciation events in *Alytes* are much younger than allowed for by Altaba's scenario. The vicariant events giving rise to *A. obstetricans* and the lineage leading to the *A. muletensis* and *A. dickhilleni* clade and the subsequent splitting between *A. muletensis* and *A. dickhilleni* cannot be placed much earlier than the Miocene-Pliocene boundary. Biogeographical scenarios invoking an earlier time of divergence should be rejected.

One may wish to discard the concept of a 'molecular clock' altogether. Then, however, there is no time frame associated to the phylogenetic hypothesis and no temporal bounds to the biogeographical reconstruction other than the age of the genus *Alytes*. By consequence, several equally plausible explanations can be postulated. Altaba's biogeographical scenario is only one of these. Altaba's model has strong appeal, but, unfortunately, it does not acknowledge some of the constraints presented by the data.

### Recommendations for further research

We pointed out that the presence of *A. obstetricans* in northern Africa poses a difficult question (Arntzen & García-París, 1995). Previous genetic analyses indicated the possibility of recent colonization of the Rif Mts. by *A. o. maurus* Pasteur & Bons, 1962 (Arntzen & Szymura, 1984). If new data would show a sister-group relationship between *A. dickhilleni* and *A. (o.) maurus*, the only compelling vicariant event would be the opening of the Strait of Gibraltar at ca. 5.5 mY. This would not allow us to discriminate between the competing biogeographical models. Additionally, the capability of anurans to cross sea straits may be underestimated. Beerli et al. (1996) cite the observation of a living adult water frog on a piece of wood floating in the Mediterranean 5 km off coast. Dispersal across the Strait of Gibraltar was also invoked by Busack (1986) to explain relatively low levels of genetic differentiation between African and European populations in some of the more mobile and ubiquitous amphibian and reptile species.

We agree in principle that palaeontological data could test for hypotheses on phylogeny and biogeography, but the chance that well-preserved material of informative age and locality will become available is remote. A more constructive suggestion, therefore, is to put the phylogenetic hypotheses to the test, preferably by the use of new and independent information, such as osteological and mitochondrial DNA sequence data, both of which studies are in progress.

# References

Altaba, C.R., 1997. Phylogeny and biogeography of midwife toads (Alytes, Discoglossidae): a reappraisal. Contr. Zool., 66(4): 257–262.

Arntzen, J.W., 1981. Ecological observations on Chioglossa lusitanica (Caudata, Salamandridae). Amphibia-Reptilia, 1: 187–203.

Arntzen, J.W. & M. García-París, 1995. Morphological and allozyme studies of midwife toads (genus Alytes) including the description of two new taxa from Spain. Contr. Zool., 65: 5–34.

Beerli, P., H. Hotz & T. Uzzell, 1996. Geologically dated sea barriers calibrate a protein clock for Aegean water frogs. Evolution, 50: 1676–1687.

Berry, V. & O. Gascuel, 1996. On the interpretation of bootstrap trees: appropriate threshold of clade selection and induced gain. Mol. Biol. Evol., 13: 999–1011.

Britten, R.J., 1986. Rates of DNA sequence evolution differ between taxonomic groups. Science, 231: 1393–1398.

Busack, S.D., 1986. Biogeographical analysis of the herpetofauna separated by the formation of the Strait of Gibraltar. Nat. geogr. Res., 2: 17–36.

Buth, D.G., 1984. The application of electrophoretic data in systematics studies. Annual Rev. Ecol. Syst., 15: 501–522.

Carretero, M.A., O. Arribas, G.A. Llorente, A. Montori, X. Fontanet, C. Llorente, X. Santos & J. Rivera, 1991. Una poblacion de Podarcis pityuensis en Barcelona. Bol. Asoc. Herpetol. Esp., 2: 18–19.

Carretero, M.A., G. A. Llorente, X. Santos & A. Montori, 1995. Características reproductoras de una población introducida de Podarcis pityusensis. Rev. esp. Herpetol., 9: 93–102.

Conant, R. & J.T. Collins, 1991. A field guide to reptiles and amphibians. Eastern and central North America (3rd ed.): 1–450 (Houghton Mifflin Co., Boston).

Crother, B.I., 1990. Is "some better than none" or do allele frequencies contain phylogenetically useful information ? Cladistics, 6: 277–281.

Ellis, W.N., 1987. Jelly: a program for the Macintosh computer for the generation of Wagner character state networks using allele frequency characters (available from the author, c/o Zoological Museum, University of Amsterdam, Entomology Department, Plantage Middenlaan 64, 1018 DH Amsterdam, The Netherlands).

Hillis, D.M. & J.J. Bull, 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol., 42: 182–192.

MacArthur, R.H. & E.O. Wilson, 1967. The theory of island biogeography: i-xi, 1–203 (Princeton University Press, Princeton).

Martin, A.P. & S.R. Palumbi, 1993. Body size, metabolic rate, generation time, and the molecular clock. Proc. natn. Acad. Sci. USA, 90: 4087–4091.

Maxson, L.R. & R.D. Maxson, 1979. Comparative albumin and biochemical evolution in plethodontid salamanders. Evolution, 33: 1057–1062.

Murphy, R.W., 1993. The phylogenetic analysis of allozyme data: invalidity of coding alleles by presence / absence and recommended procedures. Biochem. Syst. Ecol., 21: 25–38.

Murphy, R.W., W.E. Cooper & W.S. Richardson, 1983. Phylogenetic relationships of the North American five-lined skinks, genus Eumeces (Sauria: Scincidae). Herpetologica, 39: 200–211.

Rogers, J.S., 1984. Deriving phylogenetic trees from allele frequencies. Syst. Zool., 33: 52–63.

Rosa, H.D.U.C., 1995. Estrutura e diferenciação genética de populações de anuros da fauna portuguesa: i-vii, 1–214 (Ph.D. thesis, University of Lisbon, Lisbon).

Salvador, A., 1974. Guía de los anfibios y reptiles españoles: 1–237 (ICONA, Madrid).

Swofford, D.L., 1993. PAUP: phylogenetic analysis using parsimony. Computer program distributed by the Illinois Natural History Survey. Champaign, Illinois, USA.

Swofford, D.L. & S.H. Berlocher, 1987. Inferring evolutionary trees from gene frequency data under the principle of maximum parsimony. Syst. Zool., 36: 293–325.

Swofford, D.L. & R.B. Selander, 1981. BIOSYS-1: A FORTRAN program for the comprehensive analysis of electrophoretic data in population genetics and systematics. J. Hered., 72: 281–283.

Thorpe, J.S., 1982. The molecular clock hypothesis: biochemical evolution, genetic differentiation and systematics. Annual Rev. Ecol. Syst., 13: 139–168.

Watrous, L.E. & Q.D. Wheeler, 1981. The out-group comparison method of character analysis. Syst. Zool., 30: 1–12.