Distinguishing the distributions of two cryptic frogs (Anura: Discoglossidae) using molecular data and environmental modeling

Raimundo Real, A. Márcia Barbosa, Íñigo Martínez-Solano, and Mario García-París

Abstract: Currently, the identification of two cryptic Iberian amphibians, *Discoglossus galganoi* Capula, Nascetti, Lanza, Bullini and Crespo, 1985 and *Discoglossus jeanneae* Busack, 1986, relies on molecular characterization. To provide a means to discern the distributions of these species, we used 385-base-pair sequences of the cytochrome *b* gene to identify 54 Spanish populations of *Discoglossus*. These data and a series of environmental variables were used to build up a logistic regression model capable of probabilistically designating a specimen of *Discoglossus* found in any Universal Transverse Mercator (UTM) grid cell of 10 km × 10 km to one of the two species. Western longitudes, wide river basins, and semipermeable (mainly siliceous) and sandstone substrates favored the presence of *D. galganoi*, while eastern longitudes, mountainous areas, severe floodings, and impermeable (mainly clay) or basic (limestone and gypsum) substrates favored *D. jeanneae*. Fifteen percent of the UTM cells were predicted to be shared by both species, whereas 51% were clearly in favor of *D. galganoi* and 34% were in favor of *D. jeanneae*, considering odds of 4:1. These results suggest that these two species have parapatric distributions and allow for preliminary identification of potential secondary contact areas. The method applied here can be generalized and used for other geographic problems posed by cryptic species.

Résumé : L'identification de deux espèces cryptiques d'amphibiens ibériques (*Discoglossus galganoi* Capula, Nascetti, Lanza, Bullini et Crespo, 1985 et *Discoglossus jeanneae* Busack, 1986) requiert actuellement une caractérisation moléculaire. Afin de fournir un moyen de déterminer la répartition de ces espèces, nous avons identifié 54 populations espagnoles de *Discoglossus* à partir de séquences de 385 paires de bases du gène cytochrome *b*; nous avons combiné ces données à des variables environnementales pour élaborer un modèle de régression logistique capable de prédire avec une probabilité si un spécimen trouvé dans une cellule de 10 km × 10 km d'une grille de Mercator transverse universelle (UTM) appartient à l'une ou l'autre des deux espèces. Des longitudes occidentales, des vallées fluviales élargies et des substrats semi-perméables (surtout siliceux) et gréseux favorisent *D. galganoi*, alors que des longitudes orientales, des régions montagneuses, des inondations importantes et des substrats imperméables (surtout glaiseux) ou basiques (calcaires et gypseux) favorisent *D. jeanneae*. Quinze pour cent des cellules UTM sont désignées comme possédant potentiellement les deux espèces, alors que 51 % d'entre elles sont nettement attribuées à *D. galganoi* et 34 % à *D. jeanneae*, avec une probabilité de 4/1. Ces résultats indiquent que les deux espèces ont des répartitions parapatriques et ils permettent l'identification préliminaire de zones potentielles de contact secondaire. La méthode utilisée ici peut être généralisée et utilisée pour résoudre d'autres problèmes géographiques posés par des espèces cryptiques.

[Traduit par la Rédaction]

Introduction

Taxonomic changes that result in the splitting of a species into new, morphologically cryptic species are frequent among amphibians in temperate regions of Europe, North America, and Asia (Arntzen and García-París 1995; Jockusch and Wake 2002). Problems associated with species identification are relevant at any operational level of decision making involving such species. These problems affect the delimitation of geographic ranges, the evaluation of conservation statuses, and even the inference of evolutionary patterns and processes. They also entail uncertainties in the assignation of the previous species' distribution data to the new species.

The genus *Discoglossus* (Amphibia: Anura: Discoglossidae) comprises at least five species distributed over western Europe, northwestern Africa, the Middle East, and some Mediterranean islands (Capula et al. 1985; Busack 1986; García-París and Jockusch 1999). The Iberian Peninsula,

Received 22 June 2004. Accepted 23 March 2005. Published on the NRC Research Press Web site at http://cjz.nrc.ca on 27 May 2005.

R. Real[1] and A.M. Barbosa. Departamento de Biología Animal, Facultad de Ciencias, Universidad de Málaga, E-29071 Málaga, Spain.
I. Martínez-Solano and M. García-París. Museo Nacional de Ciencias Naturales (MNCN), Consejo Superior de Investigaciones Científicas (CSIC), c/ José Gutiérrez Abascal 2, E-28006 Madrid, Spain.

[1] Corresponding author (e-mail: rrgimenez@uma.es).

doi: 10.1139/Z05-040
© 2005 NRC Canada

with three species, shows the highest diversity (Pleguezuelos et al. 2002). However, similarity in external morphology among Iberian *Discoglossus* species has precluded correct species identification for decades and has resulted in misinterpretations regarding the extent of their respective distribution areas. In fact, before 1985 all the *Discoglossus* populations in the Iberian Peninsula were assigned to *D. pictus* Otth, 1837, which is also found on the Mediterranean islands of Sicily, Malta, and Gozo, as well as in North Africa. However, subsequent analyses revealed unexpected genetic diversity among *Discoglossus* populations and led to a revision of the taxonomy of the genus in the Iberian Peninsula. First, Capula et al. (1985) analyzed morphological and genetic differentiation among a number of western Mediterranean populations and described a new species from the Iberian Peninsula, *D. galganoi* Capula, Nascetti, Lanza, Bullini and Crespo, 1985, which occurred in several localities of the western part of the territory. Many of the populations classified as *D. pictus* before then belonged, according to them, to this new species. One year later, Busack (1986) analyzed allozymes and morphological data and described a new species from the south of the Guadalquivir River (southern Spain): *D. jeanneae* Busack, 1986. The two species pose one of the most extreme cases of morphological crypsis among the European herpetofauna. They are separated by more than a 7% cytochrome *b* sequence divergence (García-París and Jockusch 1999; Martínez-Solano 2004) and are thus estimated to be evolutionarily isolated from each other by more than 5 million years (Fromhage et al. 2004; Martínez-Solano 2004).

These taxonomic changes profoundly affected the interpretation of the chorological information available, as reflected in the different distribution maps shown in successive Iberian amphibian distribution atlases and field guides (Salvador 1974, 1985; Andrada 1985; Barbadillo 1987; Pleguezuelos 1997; Barbadillo et al. 1999; Ferrand de Almeida et al. 2001; Salvador and García-París 2001; Pleguezuelos et al. 2002). Immediately after the studies of Capula et al. (1985) and Busack (1986), no clear distinction was made between the three *Discoglossus* species' distributions, which were represented together on one distribution map (Barbadillo 1987). Later, the distribution of *D. pictus*, limited to the northeastern extreme of Spain, was separated from those of the two Iberian species (Pleguezuelos 1997), which were still represented together on the same map. García-París and Jockusch (1999), using data on mitochondrial DNA variation, considered that *D. galganoi* would occupy the western half of the Iberian Peninsula, whereas *D. jeanneae* would extend from the south of the Guadalquivir River throughout all eastern Iberia up to the Pyrenees. This was indicated by Barbadillo et al. (1999), who still represented the distributions of both endemic species on a single map. However, the assignation of populations from large geographic areas was still problematic, as contact zones were expected to exist in Sierra Morena, the Sistema Central range, Montes de Toledo, and the Northern Meseta. Salvador and García-París (2001) had already represented the distributions of *D. galganoi* and *D. jeanneae* on separate maps, but there were still a number of records on 20 km × 20 km quadrats that were not specifically assigned to either species. Finally, in Pleguezuelos et al. (2002), the distributions of both endemics on 10 km × 10 km Universal Transverse Mercator (UTM) squares were shown completely separated, but the assignation of many of the contact-zone records to one of the two species was made on the basis of fragmentary molecular data complemented with subjective appreciations (Martínez-Solano 2002; Martínez-Solano and García-París 2002). From a conservation perspective, this uncertainty is relevant because it impedes correct estimation of the extent of occurrence of each species. Accordingly, these species should be catalogued as "Data Deficient" in the regional Red Books of, at least, the Spanish Autonomous Regions of Navarra, País Vasco, Castilla y León, Madrid, Castilla-La Mancha, and Andalucía.

The combining of ecological modeling and molecular data may shed light on doubtful distribution records, especially for controversial species, which is important in conservation biology and environmental management. The use of multivariate statistics to model habitat–wildlife relationships, and thus determine potential distribution areas of species from a sample of distribution data, has increased considerably over the last two decades (Hilborn and Mangel 1997; Burnham and Anderson 2002), especially after geographic information systems (GIS) became widely available. Several methods can be used to determine potential distributions, some of which yield a probability of occurrence of the species in each geographic unit (Guisan and Zimmermann 2000; Seoane and Bustamante 2001).

In this work we used molecular analyses to obtain a sample of unequivocal presences of *D. galganoi* and *D. jeanneae*. With these data and a set of environmental variables, we built up a model capable of predicting the probability that any *Discoglossus* record in peninsular Spain corresponds to one or the other species, or eventually both.

Methods

Study area

Peninsular Spain (i.e., Spain excluding islands and the northern African towns) covers an area of approximately 493 000 km^2, nearly 85% of the Iberian Peninsula, in southwestern Europe. It has a complex orography and most of the main mountain ranges and river basins are oriented longitudinally. The climate is heterogeneous, with a latitudinal gradient of southward-increasing temperatures and a longitudinal and latitudinal gradient of northward- and westward-increasing precipitation (Capel 1981; Font 2000). Granite and shale substrates are common in western Spain, whereas gypsum and limestone are predominant in the eastern part of the country (IGME 1979).

Spain is considered to enclose the whole distribution area of *D. jeanneae* as well as a significant part of that of *D. galganoi* (Martínez-Solano 2002; Martínez-Solano and García-París 2002), which is also present in continental Portugal (Ferrand de Almeida et al. 2001). The contact zone between the two species and the conflicting distribution data are, therefore, likely within Spain.

Sample collection and identification

We collected, between 1999 and 2001, one to three *Discoglossus* specimens in each of 54 UTM 10 km × 10 km grid cells all over the distribution ranges of *D. galganoi* and

© 2005 NRC Canada

D. jeanneae in Spain (Fig. 1; Appendix A) and isolated genomic DNA from muscle from toe tips of adults or tail tips of larvae using a standard proteinase K – phenol extraction protocol (Sambrook et al. 1989). We amplified 385 base pairs corresponding to part of the cytochrome *b* gene via polymerase chain reaction (PCR) using the primers cytb2 and MVZ15 (Kocher et al. 1989; Moritz et al. 1992). This marker has been previously used to unambiguously distinguish between *D. galganoi* and *D. jeanneae* (García-París and Jockusch 1999). Corrected genetic distances (Kimura 2-parameter) between the two species are 7.56%–8.89%, whereas maximum intraspecific genetic distance is 1.42% for *D. galganoi* and 0% for *D. jeanneae* (Martínez-Solano 2004). PCR reactions consisted of 40 cycles of denaturing at 94 °C (60 s), annealing at 50 °C (60 s), and extension at 72 °C (60 s). PCR reactions were run in a total volume of 25 µL containing 0.2 µL of Taq polymerase (5 U/mL; Biotools, Madrid, Spain), 0.5 µL of BSA (10 mg/mL), 0.5 µL of each primer (10 µmol/L), 1 µL of dNTPs (10 mmol/L), and 1 µL of $MgCl_2$ (25 mmol/L) plus 2.5 µL of a reaction buffer (Biotools). After PCR product purification by ethanol – sodium acetate precipitation, samples were cycle-sequenced with the ABI Prism® dGTP BigDye™ Terminator v 3.0 Ready Reaction Cycle Sequencing Kit in 10-µL reactions, following the manufacturer's instructions (Applied Biosystems, Foster City, California), with 3.25 pmol of primer, 3 µL of Terminator Ready Reaction Mix, and 5% dimethyl sulfoxide. The cycling profile for the sequencing reaction consisted of 25 cycles of 10 s at 96 °C, 5 s at 50 °C, and 4 min at 60 °C. Forward and reverse sequences were obtained independently. Cycle-sequencing products were purified using MultiScreen filter plates (Millipore, Billerica, Massachusetts) and analyzed on an ABI Prism 3700 DNA Analyzer (Applied Biosystems). For species identification, the sequences obtained were aligned using the Sequence Navigator™ software, version 1.0.1 (Applied Biosystems), and compared with the sequences of *D. galganoi* and *D. jeanneae* published by García-París and Jockusch (1999) (available in GenBank, http://www.ncbi.nlm.nih.gov/Genbank/, under accession Nos. AF128895–AF128916). Some of the sequences used for the molecular characterization of populations have already been used in previous studies (García-París and Jockusch 1999; Martínez-Solano 2004) (Appendix A).

Predictor variables

For the modeling of the distribution data obtained by molecular identification, we recorded 29 variables related to topography, climate, lithology, and human activity (Table 1). We digitized the variables (except for *Alti*, which was already available in a digital version) using CartaLinx software (version 1.2; Clark Labs, Worcester, Massachusetts) and processed them using Idrisi32 GIS software (Clark Labs). Isoline variables (*HJan* through *Long*) were interpolated from a triangulated irregular network performing parabolic bridge and tunnel edge removal. Secondary variables, defined in Table 1 by an algebraic operation in parentheses, were calculated from primary variables using the Idrisi Image Calculator. Distance variables (*DHi*, *U100*, and *U500*) were calculated from the digitized highways and major towns using the Idrisi DISTANCE module. *Perm* was obtained from a map of synthesis of groundwater aquifers, a categorical map with four classes ranging from "areas with practically no aquifers" to "areas with predominant aquifers in porously permeable formations", which were assigned values from 1 to 4 (IGME 1979). We determined *Perm* by calculating the mean of the values assigned to the pixels within each 10 km × 10 km UTM square. Categorical variables *Aqmin* and *Aqmax* are the minimum and maximum aquifer categories in each UTM square, respectively. The resolution scale adopted for all variables was 1 pixel ≈ 1 km^2. We extracted the mean values of the quantitative variables for each 10 km × 10 km UTM grid cell of peninsular Spain ($n = 5167$) using a digital UTM grid map provided by the Área de Defensa Contra Incendios Forestales (Dirección General de Conservación de la Naturaleza, Ministerio de Medio Ambiente, Spain).

Statistical analyses

All statistical analyses were carried out using the SPSS® statistical package (SPSS Inc. 1999). To achieve maximum predictive power while avoiding the problems associated with multicollinearity among variables, we performed a principal components analysis of the 27 quantitative variables, thereby obtaining a set of orthogonal axes that absorbed most of the variance (see Table 2). Following the Kaiser–Guttman criterion (Legendre and Legendre 1998), we interpreted only the principal axes whose eigenvalues were higher than 1. The ecological interpretation of the axes was based on the contributions of the variables to each principal component. A particular contribution was considered greater than expected under the hypothesis of an equal contribution to all principal axes if the score of the variable in the axis was higher than an equilibrium circle with radius = (number of interpretable axes / number of variables)$^{1/2}$ (see Legendre and Legendre 1998).

We then carried out a forward stepwise logistic regression of the distribution data of our 54 genetically identified populations on the interpretable axes together with the two qualitative variables. Logistic regression is a widely used technique for modeling species' distributions (see, for example, Brito et al. 1999; Teixeira et al. 2001; Barbosa et al. 2003) that yields a probabilistic model relating a binary (1–0) target variable to a set of discrete or continuous predictor variables. The logistic regression model has the form

$$P = \frac{e^y}{1 + e^y}$$

where P is the probability of an output of 1 for the binary dependent variable, e is the basis of the Napierian logarithm, and y is a linear regression equation of the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

where β_0 is a constant and $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients of the n independent variables $x_1, x_2, \ldots, x_n$ (Hosmer and Lemeshow 1989). Logistic regression is generally used with presence/absence data, attributing ones to presences and zeros to absences to calculate the probability that a particular species is present in each territorial unit. In our case, the aim was not to use probabilities of presence to determine the po-

Table 1. Variables used to discriminate between the distributions of *Discoglossus galganoi* and *D. jeanneae* in Spain.

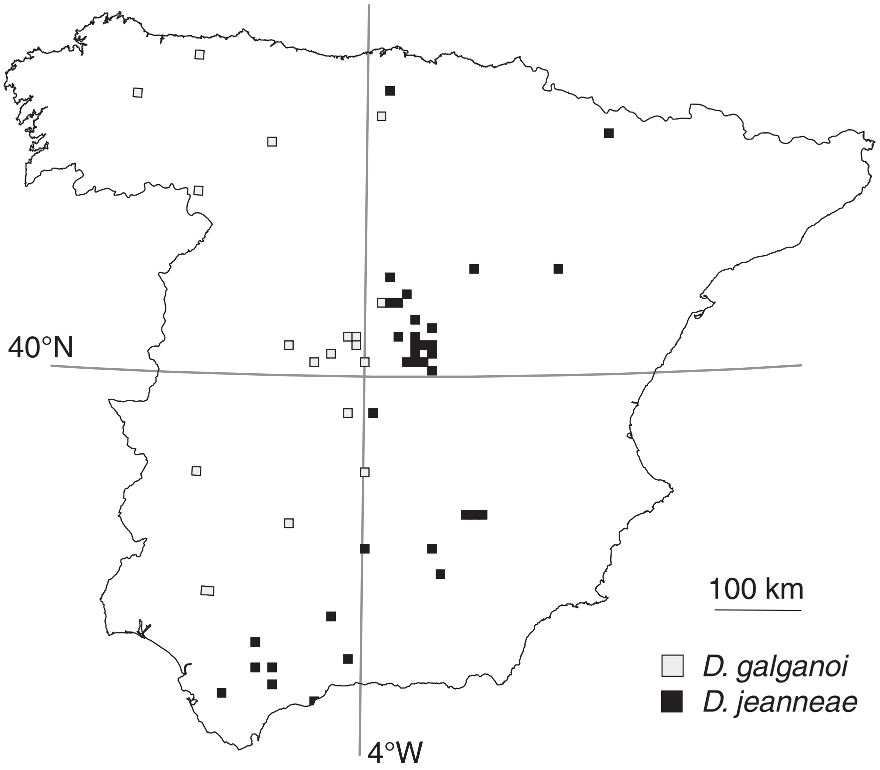
Code	Variable	Reference
Alti	Altitude (m)	US Geological Survey (1996)
ARan	Altitude range (m) (= maximum *Alti* – minimum *Alti*)	
Slop	Slope (°) (calculated from *Alti*)	
HJan	Mean relative air humidity in January at 0700 (%)	Font (1983, 2000)
HJul	Mean relative air humidity in July at 0700 (%)	Font (1983, 2000)
HRan	Annual relative air humidity range (%) (= $\lvert HJan–HJul \rvert$)	
PET	Mean annual potential evapotranspiration (mm)	Font (1983, 2000)
AET	Mean annual actual evapotranspiration (mm) (= min[*PET*, *Prec*])	
Inso	Mean annual insolation ($h \cdot year^{-1}$)	Font (1983, 2000)
SRad	Mean annual solar radiation ($kWh \cdot m^{-2} \cdot day^{-1}$)	Font (1983, 2000)
TJan	Mean temperature in January (°C)	Font (1983, 2000)
TJul	Mean temperature in July (°C)	Font (1983, 2000)
Temp	Mean annual temperature (°C)	Font (1983, 2000)
TRan	Annual temperature range (°C) (= *TJul* – *TJan*)	
DFro	Mean annual number of frost days (minimum temperature ≤ 0 °C)	Font (1983, 2000)
DPre	Mean annual number of days with precipitation ≥ 0.1 mm	Font (1983, 2000)
Prec	Mean annual precipitation (mm)	Font (1983, 2000)
MP24	Maximum precipitation in 24 h (mm)	Font (1983, 2000)
RMP	Relative maximum precipitation (= *MP24/Prec*)	
PIrr	Pluviometric irregularity	Montero de Burgos and González-Rebollar (1974)
ROff	Mean annual runoff (mm)	IGME (1979)
Lati	Latitude (°N)	IGN (1999), data on the population of urban centers taken from http://www.ine.es
Long	Longitude (°E)	IGN (1999), data on the population of urban centers taken from http://www.ine.es
DHi	Distance to the nearest highway (km)	IGN (1999), data on the population of urban centers taken from http://www.ine.es
U100	Distance to the nearest town with more than 100 000 inhabitants (km)	IGN (1999), data on the population of urban centers taken from http://www.ine.es
U500	Distance to the nearest town with more than 500 000 inhabitants (km)	IGN (1999), data on the population of urban centers taken from http://www.ine.es
Perm	Soil permeability	IGME (1979)
Aqmin	Minimum aquifers category	IGME (1979)
Aqmax	Maximum aquifers category	IGME (1979)

tential distributions of *D. galganoi* and *D. jeanneae*, but rather to distinguish between the records of these two species in areas of confirmed presence where identification is dubious. So, we attributed zeros to *D. galganoi* records and ones to *D. jeanneae* records to calculate the probability that a *Discoglossus* species in a given UTM square is *D. jeanneae* instead of *D. galganoi* (the complementary value would be the probability that it is *D. galganoi* instead of *D. jeanneae*). We therefore assessed whether the odds of finding one species instead of the other are high, or whether the odds are similar for both species. Since this procedure does not consider the possibility that neither of the two species is present at a particular site, we extrapolated the model only to the 10 km × 10 km grid cells for which there are *D. galganoi* or *D. jeanneae* records in Spain ($n = 1244$; Martínez-Solano 2002; Martínez-Solano and García-París 2002).

Extrapolation of logistic regression models should be carried out carefully, as the probability values yielded depend not only on the predictor variables included in the model but also on the random probability that is given by the proportion of both states of the target variable in the training sample (Hosmer and Lemeshow 1989). In this way, the probabilities are biased towards the state that is more frequent within the sample but not necessarily as frequent outside it. Consequently, if the source sample has a different number of cases of each state of the dependent variable, which is the most common situation, the probability values yielded by the logistic function cannot be applied universally. Some authors circumvent this problem by using subsets of data containing approximately equal proportions of occurrences of each state (e.g., Brito et al. 1999), but this implies the discarding of relevant information and the subjective choice of one specific subset among all possible subsets of data. Other authors equilibrate the impact of ones and zeros through a weighting procedure (Teixeira et al. 2001), but this alters the original data. We chose to use all data available and adjusted the probabilities yielded by the model by assigning a value of 0.5 to the relative proportion of both species in the training sample, so that 0.5 corresponds to an equal environmental probability that the *Discoglossus* record corresponds to *D. galganoi* or *D. jeanneae* (Rojas et al. 2001). This is equivalent to considering the probability value (P) as the degree of membership of the fuzzy set of samples whose environmental conditions are characteristic of *D. jean-*

© 2005 NRC Canada

Fig. 1. Results of molecular identification of *Discoglossus* specimens collected on 54 Universal Transverse Mercator grid cells (10 km × 10 km) of peninsular Spain.



neae, and the complementary value 1 – *P* as the degree of membership of the complementary fuzzy set of samples whose environmental conditions are characteristic of *D. galganoi*.

We grouped the UTM grid cells into three categories depending on their probability value: cells for which the probability of *D. galganoi* is at least four times greater than that of *D. jeanneae* ($P \leq 0.2$); cells for which the probability of *D. jeanneae* is at least four times greater than that of *D. galganoi* ($P \geq 0.8$); and cells with intermediate odds ($0.2 < P < 0.8$). As a continuity constraint, we deleted from the final map the isolated predictions (more than 50 km away from any other square predicted as favorable to the same species) that invaded the other species' range and were not supported by actual molecular observations (Hausser 1995). In these isolated favorable areas, the presence of the species is improbable owing to the geographic distance to the edge of its range.

Finally, we checked the performance of our model on a set of *Discoglossus* samples obtained and identified a posteriori ($n = 9$).

Results

Discoglossus samples taken from the western half of Spain were mostly *D. galganoi*, while those from the south and east of the country were mostly *D. jeanneae* (Fig. 1). The two species' distributions seem rather segregated, although in contact. The species were not found together in the same square, but in Madrid province (central Spain) they were found in two contiguous 10 km × 10 km grid cells.

The principal components analysis of the quantitative predictor variables yielded six interpretable axes that absorbed more than 82% of the variance (Table 2). The radius of the equilibrate contribution circle was $(6/27)^{1/2} = 0.47$. The contributions to each axis that were greater than expected, and their related factors, are also presented in Table 2.

The regression equation obtained was

$$y = -10.807 + 1.598\text{Axis1} + 1.837\text{Axis2} + 6.554\text{Axis3} - 5.001\text{Axis4} - 2.346\text{Axis5} - 2.919\text{Axis6}$$

$$+\begin{cases} 3.276 \text{ (if } Aqmax = 1) \\ -2.129 \text{ (if } Aqmax = 2) \\ -2.136 \text{ (if } Aqmax = 3) \\ 0 \text{ (if } Aqmax = 4) \end{cases} + \begin{cases} 15.617 \text{ (if } Aqmin = 1) \\ 6.728 \text{ (if } Aqmin = 2) \\ 13.655 \text{ (if } Aqmin = 3) \\ 0 \text{ (if } Aqmin = 4) \end{cases}$$

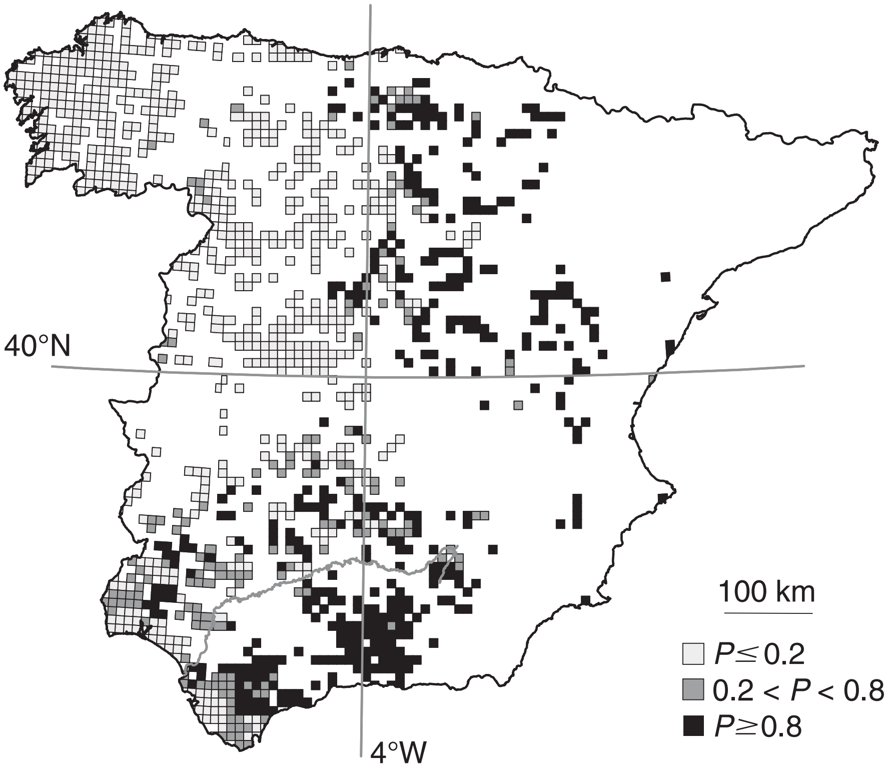
The probability values yielded by the logistic regression model, divided into three categories, are shown in Fig. 2. The continuity constraint entailed the elimination of 39 isolated predictions (≈3% of the total number of grid cells in the analysis).

Table 2. Results of principal components analysis of 27 quantitative predictor variables, with the environmental factors involved and associated variables with greater than expected contributions (under the hypothesis of an equal contribution to all principal axes) to each of six interpretable axes (only the axes with eigenvalues >1 are shown).

Axis	Eigenvalue	% of variance absorbed	% of variance accumulated	Negative	Positive
1	9.7	35.94	35.94	Water availability (*HJul*, *DPre*, *Prec*, *ROff*, *AET*) Latitude (*Lati*)	Energy (*PET*, *Inso*, *SRad*, *TJan*, *TJul*, *Temp*) Climatic instability (*PIrr*, *HRan*, *TRan*) Disturbance (*RMP*)
2	4.83	17.89	53.83	Temperature (*TJan*, *Temp*)	Altitude (*Alti*) Distance to urban centers (*U100*) Days of frost (*DFro*) Climatic instability (*HRan*, *TRan*)
3	3	11.13	64.95	Width of river basins (*HJan*)*	Mountainous areas (*ARan*, *Slop*) Longitude (*Long*) Floodings (*MP24*)
4	2.38	8.82	73.77	Longitude (*Long*)	
5	1.21	4.47	78.24		Distance to major urban centers (*U500*)
6	1.09	4.04	82.29		Soil permeability (*Perm*)

*See discussion for a detailed interpretation.

Fig. 2. Probability classes obtained for the 10 km × 10 km Universal Transverse Mercator grid cells for which there are reported presences of *Discoglossus galganoi* or *D. jeanneae* in Spain. *P* is the probability that the species present is *D. jeanneae* instead of *D. galganoi*, so that black squares represent odds greater than 4:1 favorable to *D. jeanneae*, and light grey squares represent odds greater than 4:1 favorable to *D. galganoi*. The irregular grey line is the Guadalquivir River.



In Table 3 we compare the results of the a posteriori sample identification with the outcomes predicted by our model. The model predicted the presence of both species at three localities. Among the remaining six localities, the molecular identification of populations confirmed four of the model's predictions, whereas two records from Madrid province (central Spain) were incorrectly classified. However, both of these records belonged to UTM grid cells contiguous to at least two others where the other species was predicted to occur.

Discussion

The results obtained characterize the environments occupied either by *D. galganoi* or by *D. jeanneae* and the envi-

© 2005 NRC Canada

Table 3. *Discoglossus* populations identified a posteriori compared with the predictions of the model.

Locality and province	UTM grid cell	Observed species	Predicted species	GenBank accession No.
Alfoz de Lloredo (Cantabria)	VP00	*D. galganoi*	*D. galganoi*	AY920563
Hoyo de Manzanares (Madrid)	VK29	*D. galganoi*	*D. jeanneae*	AY920564
El Pardo (Madrid)	VK38	*D. jeanneae*	*D. galganoi*	AY920565
Cerceda (Madrid)	VL10	*D. galganoi*	*D. galganoi*	AY920566
Puerto de Canencia (Madrid)	VL32	*D. jeanneae*	Both	AY920567
Puebla de la Sierra (Madrid)	VL63	*D. jeanneae*	*D. jeanneae*	AY920568
El Cardoso (Guadalajara)	VL64	*D. jeanneae*	*D. jeanneae*	AY920569
Rosal de la Frontera (Huelva)	PC60	*D. galganoi*	Both	AY920570
Constantina (Sevilla)	TG79	*D. jeanneae*	Both	AY920571

Note: UTM, Universal Transverse Mercator.

ronments shared by both species. Probability values close to 0.5 indicate areas for which the model cannot distinguish between the two species because the degrees of environmental favorability for the two species are similar. Since our model was applied only to UTM squares where *Discoglossus* presences are confirmed, these values likely indicate environments that are favorable, rather than unfavorable, to both species.

Our model predicts a quite sharp segregation between these two species, with opposite predictions for contiguous squares. Moreover, only 181 squares (15%) are predicted to be shared by both species (Fig. 2, light grey squares, where the odds of finding one species instead of the other are less than 4:1), whereas 641 (51%) are clearly in favor of *D. galganoi* and 422 (34%) are in favor of *D. jeanneae*. This is an indication of parapatric distributions and cannot be considered an artifact of the inference method, since logistic regression generally predicts more records in the intermediate probability range than in the extremes. For example, Romero and Real (1996), using the same method for identifying areas favorable to *Bufo bufo* (L., 1758) or *Bufo calamita* Laurenti, 1768 in the south of Spain, found that more than 70% of the localities were potentially shared ($0.2 < P < 0.8$).

According to the coefficients for the principal axes in the regression equation and their associated factors (Table 2), *D. galganoi* tends to occur in regions with higher water availability, temperature, relative air humidity in January, distance to major urban centers, and soil permeability than *D. jeanneae*, which seems to be favored by higher environmental energy, climatic instability, disturbance, altitude, distance to urban centers, number of frost days, mountainous areas, longitude (which increases towards the east), and floodings. The relative air humidity in January at 0700 is linked to the size and form of river basins and the flow of the rivers. In winter, nocturnal thermal inversions produce morning mists in the low areas around Spanish rivers (Capel 1981). In large basins with broad valleys and a high volume of water, these mists affect much larger areas than in shorter and narrower basins, so that the mean air humidity is greater. January air humidity has also been related to the distribution of other Iberian species associated with rivers: Vargas et al. (1998) found that it characterizes a biotic boundary for freshwater fishes in the Iberian Peninsula, separating Atlantic from Mediterranean basins, and Barbosa et al. (2001) found a positive influence on the proportion of otter (*Lutra lutra* (L., 1758)) presences in the Spanish provinces. The effect of latitude can be partially due to the fact that most of the samples from the north of Spain were collected in the western region and therefore corresponded mostly to *D. galganoi*, whereas most of the southern samples were collected in the *D. jeanneae* range (see Fig. 1). In fact, the model predicts that it is in the south that most of the overlap between the two distributions occurs, mainly because *D. jeanneae* penetrates into areas traditionally considered within the range of *D. galganoi*.

García-París and Jockusch (1999) suggested the type of geological substrate as a relevant feature for delimiting these two species' distributions and predicted the existence of zones of secondary contact in areas characterized by a transition in the type of geological substrate. Our results are in accordance with this general pattern but, in addition, suggest that it might be more complex. Regression coefficients, which can be compared because the six axes have similar ranges of values, show that Axes 3 and 4 and *Aqmin* are the most important factors in distinguishing these two species' distributions. According to this, *D. galganoi* shows particular preference for western longitudes, wide river basins, and semipermeable (mainly siliceous) and sandstone substrates, whereas *D. jeanneae* prefers eastern longitudes, mountainous areas, regions with severe floodings, and impermeable (mainly clay) or basic (limestone and gypsum) substrates.

The role of the Guadalquivir River in the separation of these species should also be reevaluated. This river has been considered a natural boundary between the distributions of *D. galganoi* and *D. jeanneae* since the description of the latter species (Busack 1986). It has also been suggested that the Guadalquivir river basin has acted as a geographic barrier for other taxa. For example, Doadrio (1988) considered the formation of the Guadalquivir fluvial system to be an important speciation event for Iberian freshwater fishes, and García-París et al. (1998) invoked this barrier to explain the differentiation between the southern subspecies *Salamandra salamandra longirostris* Joger and Steinfartz, 1994 and other subspecies. However, according to our results, there are several *Discoglossus* records to the south of the Guadalquivir River where the odds are very favorable for *D. galganoi*, and many records favorable for *D. jeanneae* to the north of this river (Fig. 2). In fact, one of the samples identified a posteriori as *D. jeanneae* was from the northern side of this river (Constantina, in Sevilla province; see Table 3). Whatever the event that yielded two vicariant species on opposite sides of the Guadalquivir, our results suggest that the river itself might not be a barrier for Iberian *Discoglossus* species, but the line where the distributions of *D. galganoi* and *D. jeanneae* meet.

The method employed here provides a way to make better use of chorological information on splitting species and subspecies for which a great deal of undifferentiated information is available, and it allows ascription of ambiguous distribution data to particular species and subspecies with greater confidence, especially for cryptic taxa. The increasing number of studies on the phylogeny and phylogeography of a large variety of taxa is producing large amounts of molecular data readily available to be integrated with GIS for management purposes, although this is yet a largely underexplored area (Alexandrino et al. 2002). The geographical mapping of genetic variability enables identification of areas characterized by maximum genetic diversity (Moritz 2002). In this context, the results of our model can be used to identify potential areas of sympatry between *D. galganoi* and *D. jeanneae* in different areas of the Iberian Peninsula in order to carry out more detailed studies of the evolutionary processes operating at these contact zones. According to our model, some of these areas would include the southern slopes of the Cantabrian Mountains in northern Spain, the Sierra de Guadarrama in Central Spain, and the western Sierra Morena in southern Spain. Moreover, the identification of different haplotype lineages within *D. galganoi* (Martínez-Solano 2004) suggests that the model could also be extended to predict areas of maximum haplotype diversity.

The integration of genetic information and environmental modeling across taxa with different evolutionary histories may provide land managers in the near future with powerful tools with which to design networks of protected areas that allow adequate preservation of evolutionary processes and their outcomes (Moritz 2002). But, in practice, this integration is not possible when the species involved are cryptic taxa. In these cases, recording of geographic occurrences is not possible without the aid of molecular tools, and the need to identify each population makes broad-scale analyses impracticable. Environmental modeling can circumvent large-scale DNA sampling and sequencing to delineate the ranges of cryptic species. We believe that the method presented here overcomes many of the problems imposed by crypsis and can be generalized to many other cases in which taxonomic ascription based on external characters fails. The accuracy of the model ultimately depends upon the total number of populations genetically identified but, as shown here, even with a relatively low number of populations, the predictions are mostly correct. The use of this method allows for the inclusion of old published records or even the use of already extirpated populations for which molecular data are impossible to obtain.

Acknowledgements

E. Albert, J. Barbadillo, R. Esteban, C. Grande, M. Lapeña, B. Lorente, M. París, E. Recuero, and I. Sánchez assisted in collecting samples. J.W. Arntzen, A. Guisan, and R. Jehle made useful comments that helped us improve previous versions of this manuscript. The Consejerías de Medio Ambiente of Andalucía, Extremadura, Castilla-La Mancha, Castilla y León, Asturias, Galicia, Aragón, and Madrid (Spain) provided the permits to collect the samples used in this work, and Patrimonio Nacional the permits to study the amphibians from El Pardo. A.M.B. is supported by a doctoral grant from Fundação para a Ciência e a Tecnologia (Portugal), and I.M.S. is supported by a Comunidad de Madrid (MNCN, CSIC) predoctoral fellowship. This work has also been partially funded by projects REN2000-1541/GLO, HP2002/0098, and REN2003-02471/GLO (Ministerio de Ciencia y Tecnología, Spain).

References

Alexandrino, J., Teixeira, J., Arntzen, J.W., and Ferrand, N. 2002. Historical biogeography and conservation of the golden-striped salamander (*Chioglossa lusitanica*) in northwestern Iberia: integrating ecological, phenotypic and phylogeographic data. *In* International Symposium on Phylogeography in Southern European Refugia: Evolutionary Perspectives on the Origins and Conservation of European Biodiversity, 11–15 March 2002, Vairão, Portugal. p. 34.

Andrada, J. 1985. Guía de campo de los Anfibios y Reptiles de la Península Ibérica. Ed. Omega, Barcelona.

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

Barbadillo, L.J. 1987. La guía de Incafo de los Anfibios y Reptiles de la Península Ibérica, Islas Baleares y Canarias. Incafo, Madrid.

Barbadillo, L.J., Lacomba, J.I., Pérez-Mellado, V., Sancho, V., and López-Jurado, L.F. 1999. Anfibios y Reptiles de la Península Ibérica, Baleares y Canarias. GeoPlaneta, Barcelona.

Barbosa, A.M., Real, R., Márquez, A.L., and Rendón, M.A. 2001. Spatial, environmental and human influences on the distribution of otter (*Lutra lutra*) in the Spanish provinces. Divers. Distrib. **7**: 137–144.

Barbosa, A.M., Real, R., Olivero, J., and Vargas, J.M. 2003. Otter (*Lutra lutra*) distribution modeling at two resolution scales suited to conservation planning in the Iberian Peninsula. Biol. Conserv. **114**: 377–387.

Brito, J.C., Crespo, E.G., and Paulo, O.S. 1999. Modelling wildlife distributions: logistic multiple regression vs. overlap analysis. Ecography, **22**: 251–260.

Burnham, K.P., and Anderson, D.R. 2002. Model selection and multimodel inference. A practical information–theoretic approach. 2nd ed. Springer-Verlag New York, Inc., New York.

Busack, S.D. 1986. Biochemical and morphological differentiation in Spanish and Moroccan populations of *Discoglossus* and the description of a new species from southern Spain (Amphibia, Anura, Discoglossidae). Ann. Carnegie Mus. **55**: 41–61.

Capel, J.J. 1981. Los climas de España. Oikos-Tau, Barcelona.

Capula, M., Nascetti, G., Lanza, B., Bullini, L., and Crespo, E.G. 1985. Morphological and genetic differentiation between the Iberian and the other west Mediterranean *Discoglossus* species (Amphibia Salientia Discoglossidae). Monit. Zool. Ital. (N.S.), **19**: 69–90.

Doadrio, I. 1988. Delimitation of areas in the Iberia Peninsula on the basis of freshwater fishes. Zool. Beitr. **39**: 113–128.

Ferrand de Almeida, N., Ferrand de Almeida, P., Gonçalves, H., Sequeira, F., Teixeira, J., and Ferrand de Almeida, F. 2001. Guias FAPAS: Anfíbios e Répteis de Portugal. FAPAS, Oporto.

Font, I. 1983. Atlas climático de España. Instituto Nacional de Meteorología, Madrid.

Font, I., 2000. Climatología de España y Portugal. Ediciones Universidad de Salamanca, Salamanca, Spain.

Fromhage, L., Vences, M., and Veith, M. 2004. Testing alternative vicariance scenarios in Western Mediterranean discoglossid frogs. Mol. Phylogenet. Evol. **31**: 308–322.

García-París, M., and Jockusch, E.L. 1999. A mitochondrial DNA perspective on the evolution of Iberian *Discoglossus* (Amphibia: Anura). J. Zool. (Lond.), **248**: 209–218.

García-París, M., Alcobendas, M., and Alberch, P. 1998. Influence of the Guadalquivir River basin on the evolution of *Salamandra salamandra* (Caudata: Salamandridae) from southern Spain. Copeia, 1998(1): 173–176.

Guisan, A., and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. Ecol. Model. **135**: 147–186.

Hausser, J. 1995. Mammifères de la Suisse. Birkhäuser Verlag, Basel.

Hilborn, R., and Mangel, M. 1997. The ecological detective. Confronting models with data. Princeton University Press, Princeton, N.J.

Hosmer, D.W., and Lemeshow, S. 1989. Applied logistic regression. John Wiley and Sons, New York.

IGME 1979. Mapa hidrogeológico Nacional. Explicación de los mapas de lluvia útil, de reconocimiento hidrogeológico y de síntesis de los sistemas acuíferos. Vol. 81. 2nd ed. Instituto Geológico y Minero de España, Ministerio de Industria y Energía, Madrid.

IGN. 1999. Mapa de carreteras. Península Ibérica, Baleares y Canarias. Instituto Geográfico Nacional / Ministerio de Fomento, Madrid.

Jockusch, E.L., and Wake, D.B. 2002. Falling apart and merging: diversification of slender salamanders (Plethodontidae: *Batrachoseps*) in the American West. Biol. J. Linn. Soc. **76**: 361–391.

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X., and Wilson, A.C. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. U.S.A. **86**: 6196–6200.

Legendre, P., and Legendre, L. 1998. Numerical ecology. 2nd ed. Elsevier, Amsterdam.

Martínez-Solano, I. 2002. *Discoglossus galganoi* Capula, Nascetti, Lanza, Bullini y Crespo, 1985. Sapillo pintojo ibérico. *In* Atlas y Libro Rojo de los Anfibios y Reptiles de España. *Edited by* J.M. Pleguezuelos, R. Márquez, and M. Lizana. Dirección General de Conservación de la Naturaleza – Asociación Herpetológica Española, Madrid. pp. 85–87.

Martínez-Solano, I. 2004. Phylogeography of Iberian *Discoglossus* (Anura: Discoglossidae). J. Zool. Syst. Evol. Res. **42**: 298–305.

Martínez-Solano, I., and García-París, M. 2002. *Discoglossus jeanneae* Busack, 1986. Sapillo pintojo meridional. *In* Atlas y Libro Rojo de los Anfibios y Reptiles de España. *Edited by* J.M. Pleguezuelos, R. Márquez, and M. Lizana. Dirección General de Conservación de la Naturaleza – Asociación Herpetológica Española, Madrid. pp. 88–90.

Montero de Burgos, J.L., and González-Rebollar, J.L. 1974. Diagramas bioclimáticos. ICONA, Madrid.

Moritz, C. 2002. Strategies to protect biological diversity and the evolutionary processes that sustain it. Syst. Biol. **51**: 238–254.

Moritz, C., Schneider, C.J., and Wake, D.B. 1992. Evolutionary relationships within the *Ensatina eschscholtzii* complex confirm the ring species interpretation. Syst. Biol. **41**: 273–291.

Pleguezuelos, J.M. (*Editor*). 1997. Distribución y biogeografía de los Anfibios y Reptiles en España y Portugal. Universidad de Granada – Asociación Herpetológica Española, Granada. pp. 134–139.

Pleguezuelos, J.M., Márquez, R., and Lizana, M. (*Editors*). 2002. Atlas y Libro Rojo de los Anfibios y Reptiles de España. Dirección General de Conservación de la Naturaleza – Asociación Herpetológica Española, Madrid.

Rojas, A.B., Cotilla, I., Real, R., and Palomo, L.J. 2001. Determinación de las áreas probables de distribución de los mamíferos terrestres en la provincia de Málaga a partir de las presencias conocidas. Galemys, **13**(N.E.): 217–229.

Romero, J., and Real, R. 1996. Macroenvironmental factors as ultimate determinants of the distribution of common toad and natterjack toad in the south of Spain. Ecography, **19**: 305–312.

Salvador, A. 1974. Guía de los Anfibios y Reptiles Españoles. ICONA, Madrid.

Salvador, A. 1985. Guía de campo de los Anfibios y Reptiles de la Península Ibérica, Islas Baleares y Canarias. Santiago García, León.

Salvador, A., and García-París, M. 2001. Anfibios españoles. Identificación, historia natural y distribución. Canseco Editores, Talavera de la Reina, Spain.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular cloning. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Seoane, J., and Bustamante, J. 2001. Modelos predictivos de la distribución de especies: una revisión de sus limitaciones. Ecologia, **15**: 9–21.

SPSS Inc. 1999. SPSS®. Version 9.0 [computer program]. SPSS Inc., Chicago.

Teixeira, J., Ferrand, N., and Arntzen, J.W. 2001. Biogeography of the golden-striped salamander, *Chioglossa lusitanica*: a field survey and spatial modelling approach. Ecography, **24**: 618–623.

US Geological Survey. 1996. GTOPO30. Land Processes Distributed Active Archive Center [online]. Available from http://edcdaac.usgs.gov/gtopo30/gtopo30.asp.

Vargas, J.M., Real, R., and Guerrero, J.C. 1998. Biogeographical regions of the Iberian Peninsula based on freshwater fish and amphibian distributions. Ecography, **21**: 371–382.

Appendix A

Table A1. *Discoglossus* populations genetically identified to obtain unequivocal distribution records of *D. galganoi* and *D. jeanneae* with which to build up a logistic regression model.

Locality and province	*N*	UTM grid cell	Species	GenBank accession No(s).
Tineo (Asturias)	3	QJ00	*D. galganoi*	AY442042–AY442044
Monbeltrán (Ávila)	1	UK25	*D. galganoi*	AY442045
Mérida (Badajoz)	2	QD21	*D. galganoi*	AY442046–AY442047
Quintanilla-Escalada (Burgos)	1	VN33	*D. galganoi*	AY442048
Poblete (Ciudad Real)	2	VJ10	*D. galganoi*	AY442049–AY442050

Table A1 (*concluded*).

Locality and province	*N*	UTM grid cell	Species	GenBank accession No(s).
Alcaracejos (Córdoba)	1	UH35	*D. galganoi*	AF128900
Reliegos (León)	2	UN00	*D. galganoi*	AY442051–AY442052
Gomeán (Lugo)	1	PH35	*D. galganoi*	AY442053
Cenicientos (Madrid)	1	UK75	*D. galganoi*	AY442054
Navalagamella (Madrid)	1	VK07	*D. galganoi*	AY920550
Navas del Rey (Madrid)	1	UK97	*D. galganoi*	AY920551
Soto del Real (Madrid)	1	VL31	*D. galganoi*	AY442055
Villamantilla (Madrid)	1	VK06	*D. galganoi*	AY442056
El Ronquillo (Sevilla)	1	QB47	*D. galganoi*	AY442057
Casarrubios a Chozas (Toledo)	1	VK14	*D. galganoi*	AY442058
El Real de San Vicente (Toledo)	2	UK54	*D. galganoi*	AY442059–AY442060
Ventas con Peña Aguilera (Toledo)	1	UJ98	*D. galganoi*	AY442061
Codesal (Zamora)	2	QG14	*D. galganoi*	AY442062–AY442063
Bienservida (Albacete)	1	WH36	*D. jeanneae*	AF128905
Riópar (Albacete)	1	WH46	*D. jeanneae*	AF128906
Entrambosríos (Burgos)	1	VN46	*D. jeanneae*	AY442064
Algodonales (Cádiz)	1	TF88	*D. jeanneae*	AY442065
San José del Valle (Cádiz)	1	TF55	*D. jeanneae*	AY442066
Cabra (Córdoba)	1	UG74	*D. jeanneae*	AF128903
Belinchón (Cuenca)	2	VK93	*D. jeanneae*	AY442067–AY442068
Zafarraya (Granada)	1	UF89	*D. jeanneae*	AY920552
Azuqueca (Guadalajara)	1	VK79	*D. jeanneae*	AF128909
Driebes (Guadalajara)	1	VK95	*D. jeanneae*	AY442069
Mondéjar (Guadalajara)	1	VK96	*D. jeanneae*	AY442070
Uceda (Guadalajara)	1	VL62	*D. jeanneae*	AY442071
Yebes (Guadalajara)	1	VK98	*D. jeanneae*	AY442072
Jaca (Huesca)	1	YN01	*D. jeanneae*	AF128911
Cazorla (Jaén)	1	WG09	*D. jeanneae*	AY920553
Iznatoraf (Jaén)	1	VH92	*D. jeanneae*	AY920554
Marmolejo (Jaén)	2	VH12	*D. jeanneae*	AY442073–AY442074
Aoslos (Madrid)	3	VL44	*D. jeanneae*	AY442075–AY442077
Belmonte de Tajo (Madrid)	1	VK74	*D. jeanneae*	AY920555
Carabaña (Madrid)	1	VK75	*D. jeanneae*	AY920556
Chinchón (Madrid)	1	VK64	*D. jeanneae*	AY442078
Estremera (Madrid)	1	VK95	*D. jeanneae*	AY442079
Fuentidueña de Tajo (Madrid)	1	VK84	*D. jeanneae*	AY442080
Nuevo Baztán (Madrid)	1	VK77	*D. jeanneae*	AY920557
Orusco (Madrid)	1	VK86	*D. jeanneae*	AY442081
Redueña (Madrid)	1	VL51	*D. jeanneae*	AY442082
Rivas de Jarama (Madrid)	1	VK57	*D. jeanneae*	AF128908
Tielmes (Madrid)	1	VK75	*D. jeanneae*	AY442083
Valdilecha (Madrid)	1	VK76	*D. jeanneae*	AY920558
Villarejo de Salvanés (Madrid)	1	VK74	*D. jeanneae*	AY920559
Alcaucín (Málaga)	1	UF08	*D. jeanneae*	AF128904
Fuengirola (Málaga)	1	UF54	*D. jeanneae*	AY920560
Ronda (Málaga)	1	UF07	*D. jeanneae*	AY920561
Morón de la Frontera (Sevilla)	1	TG81	*D. jeanneae*	AY920562
Ambrona (Soria)	1	WL45	*D. jeanneae*	AF128910
Los Yébenes (Toledo)	1	VJ28	*D. jeanneae*	AF128907
Romanos (Zaragoza)	1	XL45	*D. jeanneae*	AY442084

Note: *N*, number of individuals collected; UTM, Universal Transverse Mercator.