



Bayesian estimation of post-Messinian divergence times in Balearic Island lizards

R.P. Brown^{a,*}, B. Terrasa^b, V. Pérez-Mellado^c, J.A. Castro^b, P.A. Hoskisson^d, A. Picornell^b, M.M. Ramon^b

^a School of Biological and Earth Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK

^b Laboratori de Genètica, Departament de Biologia, Universitat de les Illes Balears, Palma de Mallorca, Spain

^c Departamento de Biología Animal, Universidad de Salamanca, Salamanca, Spain

^d Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Royal College, Glasgow G1 1XW, UK

ARTICLE INFO

Article history:

Received 15 January 2008

Revised 19 March 2008

Accepted 6 April 2008

Available online 14 April 2008

Keywords:

Bayesian MCMC

Messinian salinity crisis

Mediterranean

mtDNA

Phylogeny

Phylogeography

Podarcis

Rate-drift

ABSTRACT

Phylogenetic relationships and timings of major cladogenesis events are investigated in the Balearic Island lizards *Podarcis lilfordi* and *P. pityusensis* using 2675 bp of mitochondrial and nuclear DNA sequences. Partitioned Bayesian and Maximum Parsimony analyses provided a well-resolved phylogeny with high node-support values. Bayesian MCMC estimation of node dates was investigated by comparing means of posterior distributions from different subsets of the sequence against the most robust analysis which used multiple partitions and allowed for rate heterogeneity among branches under a rate-drift model. Evolutionary rates were systematically underestimated and thus divergence times overestimated when sequences containing lower numbers of variable sites were used (based on ingroup node constraints). The following analyses allowed the best recovery of node times under the constant-rate (i.e., perfect clock) model: (i) all cytochrome *b* sequence (partitioned by codon position), (ii) cytochrome *b* (codon position 3 alone), (iii) NADH dehydrogenase (subunits 1 and 2; partitioned by codon position), (iv) cytochrome *b* and NADH dehydrogenase sequence together (six gene-codon partitions), (v) all unpartitioned sequence, (vi) a full multipartition analysis (nine partitions). Of these, only (iv) and (vi) performed well under the rate-drift model. These findings have significant implications for dating of recent divergence times in other taxa. The earliest *P. lilfordi* cladogenesis event (divergence of Menorcan populations), occurred before the end of the Pliocene, some 2.6 Ma. Subsequent events led to a West Mallocan lineage (2.0 Ma ago), followed 1.2 Ma ago by divergence of populations from the southern part of the Cabrera archipelago from a widely-distributed group from north Cabrera, northern and southern Mallocan islets. Divergence within *P. pityusensis* is more recent with the main Ibiza and Formentera clades sharing a common ancestor at about 1.0 Ma ago. Climatic and sea level changes are likely to have initiated cladogenesis, with lineages making secondary contact during periodic landbridge formation. This oscillating cross-archipelago pattern in which ancient divergence is followed by repeated contact resembles that seen between East-West refugia populations from mainland Europe.

Crown copyright © 2008 Published by Elsevier Inc. All rights reserved.

1. Introduction

One of the more significant contributions made by molecular studies has been to provide dates for cladogenesis events, particularly among closely-related taxa that cannot be distinguished in the fossil record. Obtaining accurate estimates can be complicated by variation in rates of molecular evolution within and between both lineages and loci. Simple estimates were formerly obtained for molecules with quite well-characterized rates that varied little over given (usually recent) periods of time. More refined methods allow for rate variation between and within lineages in a phylogeny, removing the constant-rate assumption and therefore promising improved estimates (Thorne et al., 1998; Sanderson, 2002).

The introduction of Bayesian Markov Chain Monte Carlo (MCMC) methods that incorporate rate heterogeneity between branches/nodes have added a further level of refinement (Thorne et al., 1998; Huelsenbeck et al., 2000). One of the most frequently-applied Bayesian approaches is based on divergence-time priors that are obtained both recursively and input as upper and lower bounds on node times within the known topology (Kishino et al., 2001). Although other models have been considered (see Aris-Brosou and Yang, 2002) rate heterogeneity is usually modeled as drift in the logarithm of the rate over time (geometric Brownian motion: hereafter “rate-drift”) (Thorne et al., 1998). The methods have advantages over alternative maximum likelihood methods, primarily by allowing for uncertainty in calibration dates. A number of studies on real and simulated data suggest that they can provide stable posterior distributions of node times on trees that cover long periods of evolutionary time (see Thorne and Kishino, 2002; Yang and Yoder, 2003; Ho et al., 2005a). Leaving aside potential

* Corresponding author. Fax: +44 151 2073224.

E-mail address: r.p.brown@ljmu.ac.uk (R.P. Brown).

problems due to coalescence preceding species/population node dates, one area that has not yet been examined is whether they may also be useful for dating nodes on intraspecific/intrageneric phylogenies that cover much shorter time scales. Some proponents have expressed concerns about their application to intraspecific studies (Kishino et al., 2001), while failure of these methods could explain the sharp transition in evolutionary rates across the 2 Ma boundary detected by Ho et al., 2005a,b (see also Ho et al., 2007 and Emerson, 2007). Reliable recovery of node times in shallow phylogenies is often of considerable importance in revealing the recent historical biogeography of a species/genus because it can allow assessment of the temporal coincidence with physical events (e.g., Renner and Zhang, 2004; Brown et al., 2006). Furthermore, these studies often lend themselves to the Bayesian approach because knowledge of events such as island appearance and flooding of landbridges (and the associated uncertainty surrounding these data) can be readily incorporated as node time priors.

Podarcis lizards from the Balearic Islands (Spain) represent a specific case for analyzing post-Miocene evolution in the Mediterranean basin. The two species *P. lilfordi* (Mallorca, Menorca, Cabrera) and *P. pityusensis* (Ibiza, Formentera) appear to have originated at the end of the Messinian salinity crisis (MSC) when their respective island groups were separated (Terrasa et al., 2004). It is well-known that closure of the Rif and Betic marine connections between the Atlantic and Mediterranean (Krijgsman et al., 1999) led to desiccation of the Mediterranean basin during the MSC (5.96–5.33 Ma ago), with sea levels falling by up to 1500 m below present sea levels (Clauzon et al., 1996). The Balearic Islands represent the exposed peaks of a submarine promontory that is an elongation of the Spanish Betic system and were connected during this period. Fragmentation of the two main island groups would have occurred after the reopening of the Gibraltar strait 5.33 Ma ago, which led to a rapid refilling of the Mediterranean basin (in possibly less than 20 ka Pierre et al., 2006). This event has been associated with speciation in several other Mediterranean groups (see Veith et al., 2003). Subsequent sea level fluctuations have been insufficient to reconnect the groups (which are separated by depths of >640 m), or to connect Ibiza to mainland Spain (water depths > 800 m). However, all other islands and islets are separated by shallower channels, the deepest being the Mallorca–Menorca channel at ~70 m, while many other channels are <10 m. This has meant repeated connections during Quaternary falls in sea level (Emig and Geistdoerfer, 2004), with some islands/islets even having been connected since the last Würm glaciation, 18000 years ago.

Fragmentation-mediated cladogenesis events through flooding of landbridges could have occurred during any of the post-Messinian periods of heightened sea levels. If cladogenesis events are found to have occurred prior to these periods, and the resultant lineages are not shared across geographical areas, then this indicates limited or no introgression when land connections were subsequently reestablished. Such a pattern would reflect findings in continental Europe, where repeated contact has occurred during Quaternary interglacials between populations of many species after they expanded out of south-western (Iberia peninsula) and south-eastern (the Balkans) glacial refugia (Hewitt, 2000). While repeated historical contact appears well-established in these mainland species, complete isolation following a single fragmentation event is often assumed for islands.

This paper aims to establish phylogenetic relationships (i.e., tree topology) within Balearic *Podarcis* and to estimate node dates using a Bayesian MCMC approach. The impact of different sets of sequence data with very different numbers of variable positions on the means of the posterior distributions is examined. This allows assessment of the amount and type of sequence data required to obtain stable time estimates in phylogeography studies that typi-

cally span short time scales. It also provides new insights into a Mediterranean island clade that has undergone significant diversification since the start of the Pliocene, similar to many other intrageneric and intraspecific lineages from this region.

2. Materials and methods

2.1. Samples and sequencing

Podarcis pityusensis and *P. lilfordi* were sampled from all main parts of their natural distributions during the period 2001–05. Preliminary analyses of short mtDNA fragments (Terrasa et al., 2004; and unpublished data) from across the species ranges allowed selection of specimens to represent all major clades within these species (Fig. 1). Inclusion of more specimens (e.g., representing all subspecies) would have led to a large number of polytomies near the tips and potentially perturbed the dating analyses. *P. sicula* (the Italian wall lizard) was also obtained from an introduced population in Menorca and used as an outgroup to the *P. pityusensis/lilfordi* clade. Specimens were captured by noose and tail-tips removed, before being released at the site of capture (under licenses from the Balearic Islands Government). Tissue samples were kept in 100% ethanol for subsequent DNA extractions.

DNA was extracted using standard phenol–chloroform extractions. Primers were used to amplify six mtDNA and one nuclear gene fragments for each specimen (Table 1). The genes sequenced here were chosen to encompass a range of evolutionary rates, although the difficulty of finding nuclear genes that showed significant sequence diversity (due to the short evolutionary timescales involved) meant that they were mainly of mitochondrial origin. PCR products were sequenced using an ABI3130 sequencer.

Sequences were initially aligned using Clustal V, and then adjustments were made by eye. Positions within genes with known secondary structures (rRNA, and tRNA) were assigned to stems or loops, and stem positions matched with corresponding positions with which they were paired. Secondary structures for 12S rRNA were based on those for *Plestiodon egregius* (EMBL AB016606; Van de Peer et al., 2000), although some refinements were made by examination of base compatibility and covariation between sites (see Brown 2005). The tRNA structures were based on standard tRNAs (Kumazawa and Nishida, 1993).

2.2. Phylogenetic inference

The sequence data were assigned to the following twelve partitions: control region (CR); codon positions 1 (cytb/1), 2 (cytb/2), 3 (cytb/3) for cytochrome *b*; codon positions 1 (ND/1), 2 (ND/2) and 3 (ND/3) for the ND1 and ND2 genes together; stem (tRNA/s) and loop (tRNA/l) bases in the three transfer RNAs together, tRNA-Gln, tRNA-Ile, tRNA-Met; stem (12S/s) and loop (12S/l) regions of the 12S rRNA; *C-mos* (CMOS). Evidence suggests that appropriate partitioning, particularly when based on functional aspects of the sequence (e.g., codon position), is preferable to no partitioning (Yang and Yoder, 2003; Brandley et al., 2005). Log-likelihoods for 24 different DNA substitution models were calculated for a single topology (Neighbor Joining tree [NJ] from Jukes Cantor [JC] distances, using all partitions). This was run on each data partition and the “best” model of DNA substitution selected, using hierarchical likelihood-ratio tests (as implemented in MrModeltest 2.2: Nylander, 2004).

Bayesian phylogenetic analyses were performed using MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001). The above 12 data partitions were used with their respective models (parameters were unlinked). Doublet models for non-independent evolution of paired sites were applied to the stem RNA partitions. Two MCMC samplers were run in parallel (four chains each, temp parameter

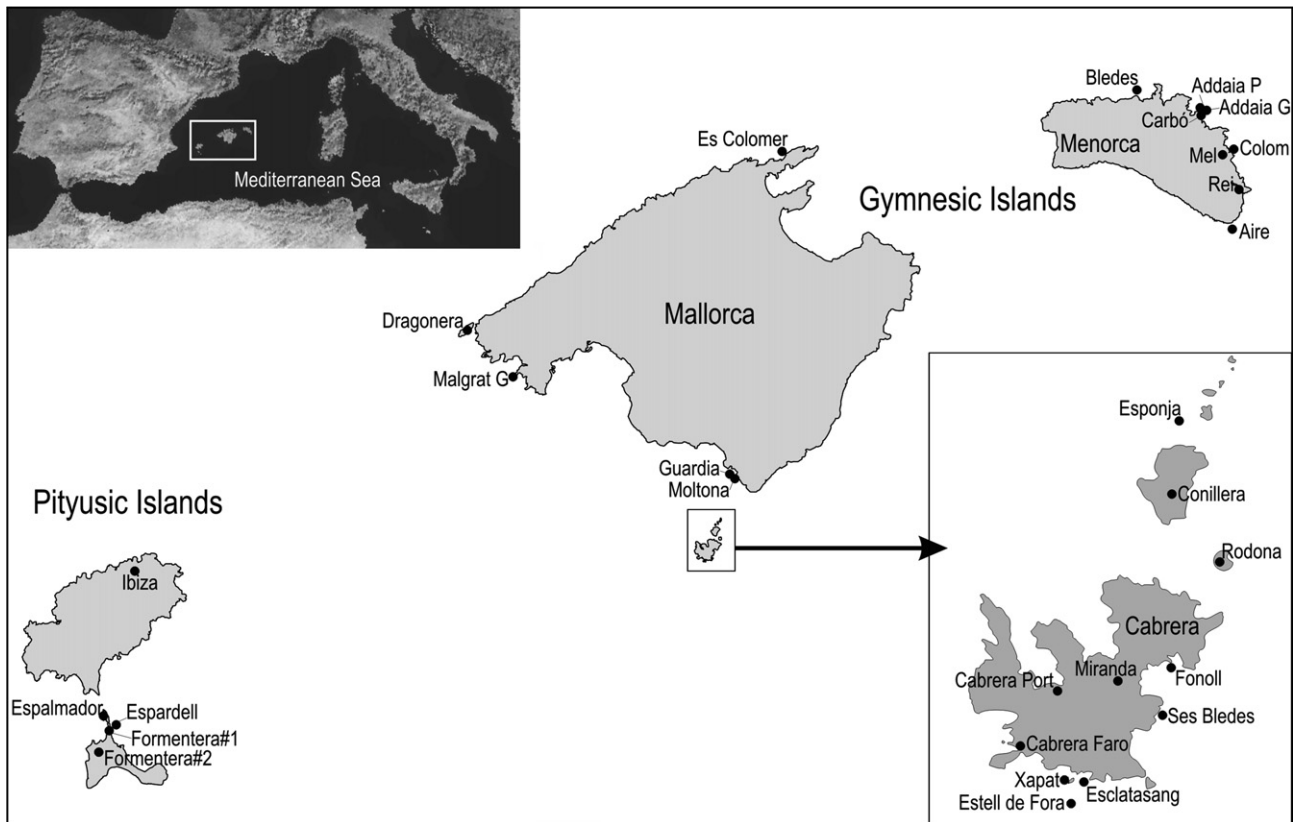


Fig. 1. Map of the Balearic Islands with the 29 locations corresponding to the *P. pityusensis* (Pityusic Islands) and *P. lilfordi* (Gymnesic Islands) individuals used in the present study (one individual per location).

Table 1
Genes and primers

Genes	Primers and reference	Sequence length (bp) analysed
Cytochrome <i>b</i>	L14841 and H15149 (Kocher et al., 1989)	306
Cytochrome <i>b</i>	L14724 and H15175 (Palumbi, 1996)	
Cytochrome <i>b</i>	L15369 and H15915 (Fu, 2000)	506
ND1, ND2, tRNA _{Lie} , tRNA _{Gln} , tRNA _{Met}	L4178 and H4980 (Macey et al., 1998)	667
Control region	L15022 (Brehm et al., 2003) and 5'-GTCTTGTGACTGTAATTAACCGATA-3'	481
12s rRNA	L1091 and H1478 (Kocher et al., 1989)	373
<i>C-mos</i>	G73 and G74 (Saint et al., 1998)	342

Light and heavy strand primer are generally referred to as L and H, respectively. All genes are mitochondrial except *C-mos*.

set at 0.5), starting from a random tree, up to 2.6×10^6 generations (samples recorded every 100 generations). The first 15,000 samples of each run were discarded to ensure that stationarity had been reached (assessed from the standard deviation of the split frequencies from the two runs). The runs were compared and combined to provide an estimate of the posterior distribution based on 22,000 samples. Maximum Parsimony (MP) analyses (unweighted; gaps excluded) were also performed using PAUP* version 4.0b10 (Swofford, 2002), on 1000 bootstrap samples from the sequence data (heuristic search, 10 random addition replicates), to provide an alternative approach (see Simmons et al., 2004; Lewis et al., 2005).

2.3. Estimates of divergence times

A Bayesian molecular dating approach was applied to the above topology (Thorne et al., 1998; Thorne and Kishino, 2002), using a

practical approach described in detail elsewhere (Renner and Zhang, 2004; Rutschmann, 2005). Maximum likelihood (ML) estimates of model parameters were obtained for each partition using BASEML from the PAML version 3.14 suite of programs (Yang, 1997) for individual NJ trees based on JC distances. ML estimates of tree branch lengths and their corresponding variance-covariance matrices were obtained under each model for each individual partition, for the Bayesian topology using the program ESTBRANCHES_DNA from the MULTIDIVTIME package (Thorne and Kishino, 2002). MULTIDIVTIME was used to run a Markov Chain Monte Carlo (MCMC) chain and provide posterior distributions of times and substitution rates, based on: (1) multipartition analyses based on all partitions; (2) multipartition analyses that included only subsets of all partitions; (3) single data partitions; (4) all data entered as a single-partition (i.e., concatenated). The two partitions with the least phylogenetic information were excluded (tRNA/s helices and CMOS). Other single partitions with low numbers of variable sites were included (e.g., *cytb*/2) for completeness (e.g., to allow analysis of the entire cytochrome *b* sequence) and because they allowed assessment of performance under poor conditions.

The ingroup node priors were specified as hard upper and lower bounds (under the model used here, proposed states falling outside these constraints will have a prior probability of zero and therefore be rejected). Priors for unspecified nodes are obtained recursively from a Dirichlet distribution based on the time between a particular node and a terminal node as a proportion of the time from the ingroup root to the present (Kishino et al., 2001). The well-known physical events in the Mediterranean during the MSC as well as evidence from other studies (e.g., Martínez-Solano et al., 2004) allow narrow constraints to be applied to the ingroup node (time = T_0) for *P. lilfordi*/*P. pityusensis*. The upper limit for T_0 was 5.33 Ma corresponding to independent dating of the end of the

MSC, when refilling of the Mediterranean basin began (see Introduction). Exactly when the two island groups would have become separated is not known, but the deep intermediate channel together with evidence of a very rapid refilling of the basin (Pierre et al., 2006) suggests that this would have occurred within a few thousand years and so the lower constraint was set at 5.32 Ma.

Individual rate priors were input into single-partition analyses. These were based on rates estimated from the divergence between the in-group node and the tip sequences. This was used as a prior in a preliminary MULTIDIVTIME analysis, with the mean of the posterior distribution of rates at this node then being input into the final analysis (despite this, the standard deviation was set equal to the mean to ensure the prior had little influence on the posterior distribution). Only one rate prior can be entered for multipartition analyses in MULTIDIVTIME and so the mean of the rates for all partitions (at the within-group node) was used. Analyses that assumed rate constancy (i.e., clock-like evolution) across the tree and analyses that allowed for rate-drift were computed. In the latter, the intention was to use a diffuse rate autocorrelation prior, ν , and so 0.3752 was used for both the mean and standard deviation (this way the product $\nu T_0 = 2$; see MULTIDIVTIME manual, and Kishino et al., 2001). Previous work has supported the use of equal rate autocorrelation between genes (Thorne and Kishino, 2002).

The MCMC chain was sampled every 100 cycles, until 10,000 samples had been obtained. In each analysis, between 3×10^5 and 1×10^6 cycles were discarded as burn-in.

3. Results

3.1. Phylogeny of Balearic Island *Podarcis*

Sequence accession numbers are given in the Appendix. The partitioned Bayesian analysis provided a well-supported topology, and all major nodes showed clade credibility values ≥ 0.99 , with values >0.95 for 19 of the 21 nodes (Fig. 2). The MP analysis provided an almost identical topology between island groups, with only negligible differences at two recent internal nodes. High levels of bootstrap support were detected for the six major clades shown in Fig. 2 (91–100%), although support for more recent clades was lower, ranging from 61% to 97%.

The most basal Balearic Island node represents separation of the two Balearic *Podarcis* species (Fig. 2). In *P. lilfordi* subsequent cladogenesis events led to a Menorcan lineage and then a west Mallorcan lineage (represented by individuals from the islets of Malgrats and Dragonera). The next cladogenesis event led to the separation of populations from (N Mallorca, S Mallorca, and N Cabrera) from the main island of Cabrera and its associated islets. The latter clade is further subdivided between northern and southern parts of the Cabrera archipelago. Fewer *P. pityusensis* were sampled because of the lower diversity within this group. The present analyses support preliminary work that suggested two main clades, with the ancestral node representing divergence of an Ibiza haplotype from Formentera haplotypes.

3.2. Node dates

The general validity of the constraints applied to the ingroup node can be examined by consideration of the resultant rates of sequence divergence relative to those in other lizard studies. The mean sequence divergence between *P. lilfordi* and *P. pityusensis* for cytochrome *b* (one of the most widely used markers in phylogeography studies) was 0.1093 under a Tamura 3-parameter model (approximating the model used in the Bayesian estimation of dating) with the corresponding *p*-distance being 0.08546. If the

node coincides with the loss of landbridges between the two main island groups at the end of the MSC, then it equates to a between-lineage rate of 2.05% divergence per million years (1.6% based on raw divergence). This is close to cytochrome *b* divergence rates in other island lizards (i.e., 2.3%/Ma (unadjusted) calculated from Brown and Pestano, 1998; 1.43%/Ma (adjusted) from Malhotra and Thorpe, 2000; 1.5–1.9%/Ma and 2.0%/Ma (unadjusted) for two lizard species discussed in Pestano et al., 2003). Additional support is obtained by examining corresponding rates of NADH subunit evolution in the same manner. The mean between-lineage rate for this sequence is found to be 1.12%/Ma (unadjusted) and 1.28%/Ma (adjusted). These values are very close to calibrated rates for other reptiles and amphibia (1.2–1.4%/Ma (unadjusted) in Macey et al., 1998; 1.28%/Ma (unadjusted) in Weisrock et al., 2001).

A likelihood-ratio test was used to compare the difference (Δ) between likelihoods with and without a molecular clock assumption for the topology identified by the Bayesian analysis. The deviation was not quite significant at the 5% significant level: $2\Delta = 40.3$ [compared with $\chi^2_{[df=28]}$ distribution], $P = 0.062$. This rather equivocal result justified the use of both rate-drift and constant-rate analyses of divergence times.

Estimated dates and rates at internal nodes were analyzed to assess the impact of the different partitions on the analyses. In the analyses allowing for rate-drift, the mean of the posterior distribution for rate at the root was lower in all but one of the single-partition analyses than the corresponding value for the same partition in the full multipartition analysis, and this amounted to a several-fold difference for the most slowly-evolving partitions (Table 2). The exception was the partition with the highest rate, namely codon position 3 for the cytochrome *b* gene. Tendency to change rate (ν) did not differ greatly between multi- and single-partition analyses, although it did appear higher in genes with higher rates (at the ingroup root node). Levels of correlation between partitions (multipartition analysis) are mixed (Table 3), and there is no strong tendency for correlations between partitions from the same gene.

Deviations between node times estimated under the distinct-partitions schemes and those estimated under the multiple partition rate-drift analysis are summarized in Fig. 3. The assumption that the latter provided the most robust analysis for comparison was justified by the fact that it contained the greatest amount of appropriately-partitioned sequence data, as well as accommodating possible changes in rate between branches. The results indicate that low levels of sequence variability lead to considerable overestimation of node dates. This bias clearly diminishes as the number of variable sites increases and can also depend on whether rate homogeneity is assumed. Under a strict molecular clock (i.e., under $\nu = 0$), estimated dates became very close to those under the full multipartition analysis when 54 or more variable sites are used, while 172 or more sites were required to achieve a similar effect in the rate-drift analyses. The unpartitioned (concatenated sequence) analyses showed exactly the same pattern, i.e., the strict-clock method provided better estimates than the rate-drift analysis. This underlines the importance of sequence partitioning, as both types of analysis gave quite similar estimates when the data were fully partitioned.

4. Discussion

4.1. Bayesian estimation of evolutionary rates and node times

Using Balearic Island lizard sequence data, we present evidence that substitution rates are considerably underestimated and thus time since divergence overestimated by the Bayesian MCMC method, particularly when short, slowly-evolving sequence data are

Table 3
Rank correlation coefficients for rates between partitions (multipartition analysis)

Partition	Partition									
	CR	Cytb/1	Cytb/2	Cytb/3	ND/1	ND/2	ND/3	tRNA/l	12S/s	
Cytb/1	0.526									
Cytb/2	0.824*	0.293								
Cytb/3	0.136	0.755	-0.117							
ND/1	0.142	0.722	0.071	0.549						
ND/2	0.652	0.326	0.813	0.020	0.242					
ND/3	0.334	0.806	0.193	0.629	0.853*	0.375				
tRNA/l	0.547	0.974***	0.273	0.743*	0.719	0.361	0.800*			
12S/s	0.215	0.818	0.094	0.598	0.865*	0.200	0.786	0.801*		
12S/l	0.681	0.330	0.818*	-0.006	0.170	0.921**	0.402	0.330	0.190	

* $P < 0.05$.

** $P < 0.01$.

*** $P < 0.001$.

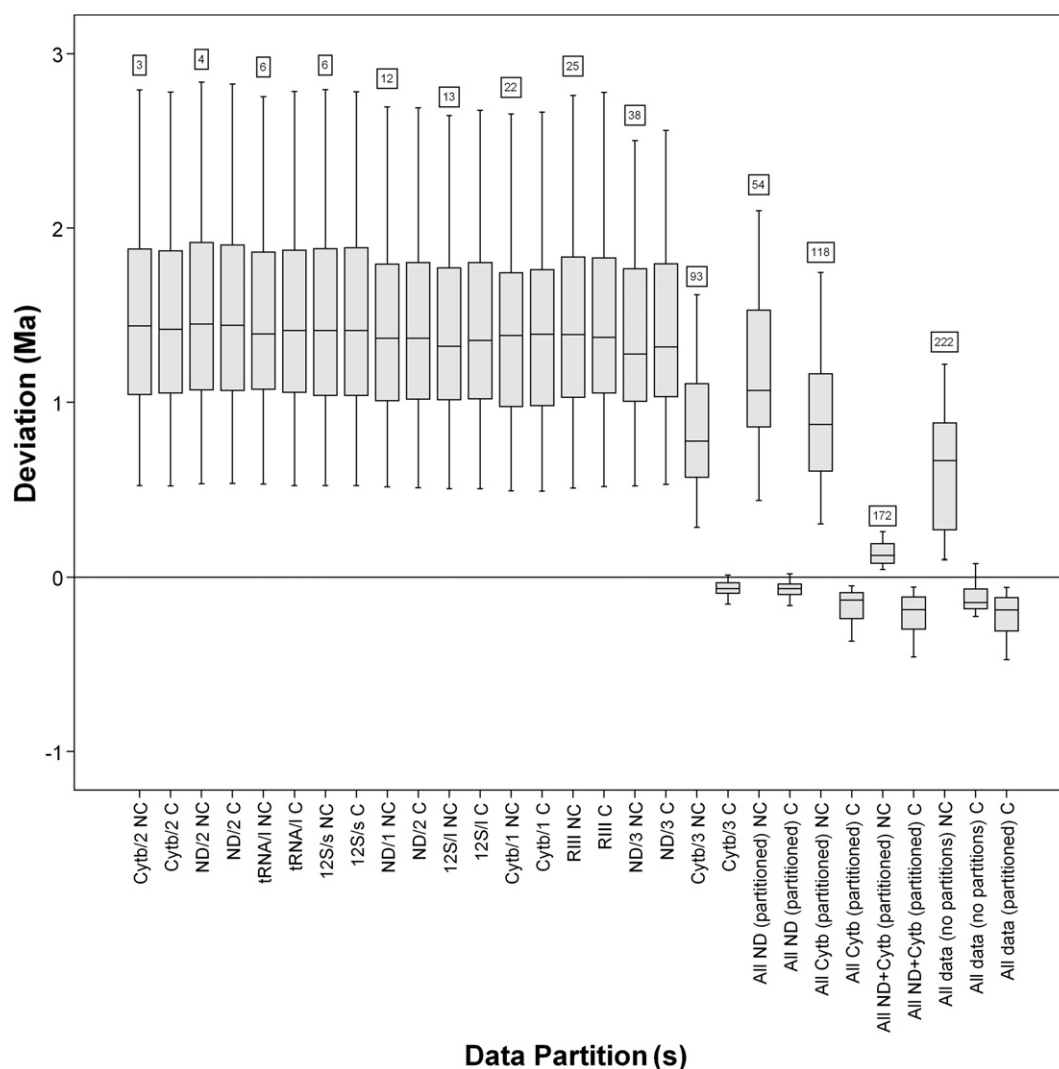


Fig. 3. Box and whisker plot displaying summaries (median, interquartile range and range) for deviations between means of the posterior distributions of node ages from specified Bayesian MCMC analyses and those from the full multipartition rate-drift analysis. NC after the descriptors of the partition(s) used indicates a rate-drift analysis, C indicates a constant-rate analyses. Numbers above NC whiskers indicate the number of ingroup variable sites in the partition(s).

used. This supports concerns about applying current Bayesian MCMC methods to estimate substitution rates and node times in shallow phylogenies (Kishino et al., 2001; Emerson, 2007).

The practical impact of this can be illustrated by considering a typical phylogeography study on ~30 ingroup taxa that share a common ancestor (mtDNA) at 5–6 Ma ago, based on a reasonable amount of rapidly-evolving mtDNA sequence. We show that over

1250 bp of partitioned NADH and cytochrome *b* data are needed to obtain reasonable node time estimates in the rate-drift analysis. In contrast, the strict-clock analysis did allow quite accurate dating with less sequence. It performed significantly better for all datasets containing ostensibly reasonable numbers of variable sites (i.e., 3rd codon from cytochrome *b*, partitioned NADH sequence, partitioned cytochrome *b* data), as well as for the non-partitioned analysis

based on all sequence data. Thorne and Kishino (2002) showed that the constant-rates assumption worked well with simulated multi-gene data, possibly because rates between simulated genes were not correlated. Rate correlations between mitochondrial partitions appeared higher here, but this is probably of little importance relative to the impact of low sequence divergence.

The considerably improved performance of strict-clock analyses at intermediate levels of sequence variability does not seem to be attributable to evolution being relatively clock-like. This is because the rate-drift analysis should also accommodate clock-like evolution. In particular, the prior distribution of the autocorrelation rate parameter ν (modeled with a gamma distribution) was set equal to its standard deviation and so should be exponentially distributed (under these conditions the shape parameter, $\alpha = 1$, and scale parameter, $\theta = \text{mean}[\nu]$) and thus provide high probability density close to zero (i.e., as modeled under a strict clock). While alternative reasons for this remain to be investigated, one tentative explanation is that the rate-drift analyses suffer as a result of an increased number of priors, some of which have a major influence on the analysis when the sequence data are less informative.

We also considered whether biased date estimation for data sets containing fewer variable sites was due to poorer performance at more recent internal nodes. Daughter lineages from older nodes are likely to be well differentiated, while those from younger nodes could have almost identical sequences and so impede the ability of the method to estimate recent node times. To test for a temporal bias, rank correlations were computed between node ages (estimated by the multipartition analysis), and node-age deviations between multipartition and single-partition estimates. They were found to be large and positive for each of the most conservative single partitions (results not shown) which supports the opposite trend, i.e., times for older nodes are more greatly overestimated than times for younger nodes. Another way of exploring this effect is to test the correlation between number of variable sites in the sequence and the estimated dates for a particular node, using data from each of the 10 single partitions. This analysis revealed strong negative associations both for the oldest estimated node ($r = -0.913$, $P < 0.001$), as well as the most recent internal node ($r = -0.931$, $P < 0.001$). This also suggested that the overestimation of node ages (when sequence divergence is low) occurs at both old and young nodes on the tree.

Finally, it should be stressed that some components of the observed biases may be related to specific aspects of the data set. In particular, and due to a lack of additional external evidence, we do not investigate how the position of the node constraints on the tree might affect the observed bias. Here, upper and lower bounds could be applied to the prior for the oldest ingroup node. Future studies could investigate whether similar biases are obtained for phylogenies in which node priors are specified on very recent nodes.

4.2. Timing of divergence in the Mediterranean basin

The two Balearic *Podarcis* species are clearly quite old and appear to have been evolving within separate island groups since the refilling of the Mediterranean basin at the end of the MSC. Subsequent cladogenesis events within each of these taxa are much more recent, and assuming that the multigene analysis is reliable (based on convergence of date estimates with increasing levels of phylogenetic information content), we shall consider their timings in relation to geoclimatic events.

The oldest within-species node corresponds to the separation of Menorcan from Mallorca/Cabrera *P. lilfordi* in the late Pliocene, some 2.6 Ma ago. This period coincides with the beginning of the Quaternary and was notable because of very significant global cooling and growth of the major ice sheets in the Northern Hemi-

sphere (Webb and Bartlein, 1992) and its concomitant effect on the world's ecosystems (Stanley and Ruddiman, 1995). Many phylogenetic studies have revealed cladogenesis events in diverse species from around the world that coincide with this period (e.g., Paulo et al., 2001; Veith et al., 2003; Guo et al., 2005). Significant changes in nanofossil ratios have been observed for the Mediterranean at about this time (Siesser and de Kaenel, 1999). The alternative hypotheses to be considered here are whether this node represents a transmarine dispersal or colonization of islands via temporary landbridges followed by vicariance as sea levels rose. Sea levels were generally above current levels since before the mid-late Pliocene and do not appear to have dropped sufficiently to expose the 70 m channel between Mallorca and Menorca (Emig and Geistdoerfer, 2004). This strongly suggests transmarine dispersal and fits recent findings that a significant proportion of reptile colonization events in the Mediterranean occurred when no landbridges were present and can therefore be attributed to rafting (e.g., Carranza et al., 2006). Posterior island connections during eustatic Pleistocene sea level changes (see Emig and Geistdoerfer, 2004) imply episodes of secondary contact between these clades, but with little or no introgression.

The subsequent node in *P. lilfordi* represents cladogenesis of NW Mallorca populations from the remaining Mallorca and Cabrera populations some ~ 2 Ma. The node estimates encompass the Pliocene/Pleistocene boundary, characterized by widespread marine bivalve extinctions (Stanley and Campbell, 1981). It is possible that sea levels fell below present levels at this time (Emig and Geistdoerfer, 2004), which could suggest dispersal over a temporary landbridge between Mallorca and Cabrera (with subsequent recolonization of Eastern Mallorca from Cabrera) but this hypothesis remains highly speculative. Difficulties with interpretation of this node are partly because of the extinction of *P. lilfordi* on the main island of Mallorca. This means that the current clade distribution does not provide a detailed picture of the original phylogeography across the island, which limits assessment of its cause.

Specimens from Cabrera and adjacent islets form a clade that diverged from the remaining populations in North-eastern and Southern Mallorca, together with islets from the north of the Cabrera group during the Pleistocene. The pattern is slightly discordant with the relatively deep channel (50 m) that separates all Cabreran islands and Mallorca: a monophyletic Cabrera group might have been predicted, as seen in Menorca. Our analyses suggest that this cladogenesis event may be dated at ~ 1.2 Ma ago. The only "deep" node within *P. pityusensis*, the node separating the Ibiza specimen from the rest, also corresponds to this period. These events occurred during the long period between the Donau (2.35 Ma ago) and Guntz glaciations (0.65 Ma ago). Very high sea levels (around 100 m above current levels) appear to have been present twice during the Pleistocene (1.8–0.7 Ma and ~ 0.6 Ma ago) (Emig and Geistdoerfer, 2004) and meant that much of the Cabrera archipelago (and Ibizan islands) would be submerged, although some subaerial peaks would be present. This again points to transmarine colonization events leading to the observed cladogenesis within the two species.

Two further cladogenesis events within the aforementioned *P. lilfordi* clades seem to correspond to periods around 0.7–0.8 Ma ago broadly coinciding with the Guntz (0.65 Ma) glaciation. Around this period the range of eustatic sea level fluctuations between glacials and interglacials may have reached 200 m. No further clues allow additional inferences concerning these clades.

In summary, the post-Messinian evolution of Balearic *Podarcis* appears to have occurred during quite early Pliocene/Pleistocene cladogenesis events. Several of these events occurred when sea levels were higher than present suggesting transmarine colonization. These patterns have persisted despite subsequent periods when islands would have been joined, indicating that there was lit-

the introgression of mtDNA during secondary contact. This reflects findings on continental populations from Europe that appear to have expanded out of refugia several times during consecutive interglacial periods (Hewitt, 2004).

Acknowledgments

This work was funded by the Grant REN2003-08432 from the Spanish Ministry of Education and Science. We wish to thank the Servei de Protecció d'Espècies, Conselleria de Medi Ambient, Govern de les Illes Balears for permits (ref.: 30/2004) to study *Podacis* in the Balearic Islands. We also thank Pep Amengual, José Angel Fernández-Estevéz, Teresa García Díez, Paloma Gimenez, Joan Mayol and the Cabrera National Park wardens for help with various aspects of the work. We are grateful to Ziheng Yang, Alan Larson and an anonymous referee for helpful comments on an earlier draft of the manuscript. RPB was funded by a HEFCE sabbatical during the writing of this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2008.04.013.

References

- Aris-Brosou, S., Yang, Z., 2002. The effects of models of rate evolution on estimation of divergence dates with a special reference to the metazoan 18S rRNA phylogeny. *Systematic Biology* 51, 703–714.
- Brandley, M.C., Schmitz, A., Reeder, T.W., 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. *Systematic Biology* 54, 373–390.
- Brehm, A., Harris, D., Alves, C., Jesus, J., Thomarat, F., Vincente, L., 2003. Structure and evolution of the mitochondrial DNA complete control region in the lizard *Lacerta dugesii* (Lacertidae, Sauria). *Journal of Molecular Evolution* 56, 46–53.
- Brown, R.P., Pestano, J., 1998. Phylogeography of Canary Island skinks inferred from mtDNA sequences. *Molecular Ecology* 7, 1183–1191.
- Brown, R.P., 2005. Large subunit mitochondrial rRNA secondary structures and site-specific rate variation in two lizard lineages. *Journal of Molecular Evolution* 60, 45–56.
- Brown, R.P., Hoskisson, P.A., Welton, J.H., Báez, M., 2006. Geological history and within-island diversity: a debris avalanche and the Tenerife lizard *Gallotia galloti*. *Molecular Ecology* 15, 3631–3640.
- Carranza, S., Harris, D.J., Arnold, E.N., Batista, V., Gonzalez, J.P., 2006. Phylogeography of the lacertid lizard, *Psammotromus algirus*, in Iberia and across the Strait of Gibraltar. *Journal of Biogeography* 33, 1279–1288.
- Clauzon, G., Suc, J.P., Gautier, F., Berger, A., Loure, M.F., 1996. Alternate interpretation of the Messinian salinity crisis; controversy resolved? *Geology* 24, 363–366.
- Emerson, E., 2007. Alarm bells for the molecular clock? No support for Ho et al.'s model of time-dependent molecular rate estimates. *Systematic Biology* 56, 337–345.
- Emig, C.C., Geistdoerfer, P., 2004. The Mediterranean deep-sea fauna: historical evolution, bathymetric variations and geographical changes. *Carnets de Géologie/Notebooks on Geology*, Maintenon, Article 2004/01 (CG2004_A01_CCE-PG).
- Fu, J., 2000. Toward the phylogeny of family Lacertidae: why 4708 base pairs of mtDNA sequences cannot draw the picture. *Biological Journal of the Linnean Society* 71, 203–217.
- Guo, X., He, S., Zhang, Y., 2005. Phylogeny and biogeography of Chinese sisorid catfishes re-examined using mitochondrial cytochrome *b* and 16S rRNA gene sequences. *Molecular Phylogenetics & Evolution* 35, 344–362.
- Hewitt, G.M., 2000. The genetic legacy of the quaternary ice ages. *Nature* 405, 907–913.
- Hewitt, G.M., 2004. Genetic consequences of climatic oscillations in the quaternary. *Philosophical Transactions of the Royal Society of London B* 359, 183–195.
- Ho, S.Y.W., Phillips, M.J., Drummond, A.J., Cooper, A., 2005a. Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. *Molecular Biology and Evolution* 22, 1355–1363.
- Ho, S.Y.W., Phillips, M.J., Cooper, A., Drummond, A.J., et al., 2005b. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution* 22, 1561–1568.
- Ho, S.Y.W., Shapiro, B., Phillips, M.J., Cooper, A., Drummond, A.J., 2007. Evidence for time dependency of molecular rate estimates. *Systematic Biology* 56, 515–522.
- Huelsenbeck, J.P., Larget, B., Swofford, D.L., 2000. A compound poisson process for relaxing the molecular clock. *Genetics* 154, 1879–1892.
- Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755.
- Kishino, H., Thorne, J.L., Bruno, W.J., 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology and Evolution* 18, 352–361.
- Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X., Wilson, A.C., 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing of conserved regions. *Proceedings of the National Academy of Sciences of the United States of America* 86, 6196–6200.
- Krijgsman, W., Hilgen, F., Raffi, I., Sierro, F., Wilson, D., 1999. Chronology, causes and progression of the Messinian salinity crisis. *Nature* 400, 652–654.
- Kumazawa, Y., Nishida, M., 1993. Sequence evolution of mitochondrial tRNA genes and deep branch animal phylogenetics. *Journal of Molecular Evolution* 37, 380–398.
- Lewis, P.O., Holder, M.T., Holsinger, K.E., 2005. Polytomies and Bayesian phylogenetic inference. *Systematic Biology* 54, 241–253.
- Macey, J.R., Schulte II, J.A., Ananjeva, N.B., Larson, A., Rastegar-Pouyani, N., Shammakov, S.M., Papenfuss, T.J., 1998. Phylogenetic relationships among agamid lizards of the *Laudakia caucasia* species group: testing hypotheses of biogeographic fragmentation and an area cladogram for the Iranian plateau. *Molecular Phylogenetics & Evolution* 10, 118–131.
- Malhotra, A., Thorpe, R.S., 2000. The dynamics of natural selection and vicariance in the Dominican anole: patterns of within-island molecular and morphological divergence. *Evolution* 54, 245–258.
- Martínez-Solano, I., Gonçalves, H.A., Arntzen, J.W., García-París, M., 2004. Phylogenetic relationships and biogeography of midwife toads (Discoglossidae: *Alytes*). *Journal of Biogeography* 31, 603–618.
- Nylander, J.A.A., 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.
- Palumbi, S.R., 1996. Nucleic acids II: The polymerase chain reaction. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), *Molecular Systematics*, second ed. Sinauer Associates, Sunderland, pp. 205–247.
- Paulo, O.S., Dias, C., Bruford, M.W., Jordan, W.C., Nichols, R.A., 2001. The persistence of Pliocene populations through the Pleistocene climatic cycles: evidence from the phylogeography of an Iberian lizard. *Proceedings of the Royal Society of London B* 268, 1625–1630.
- Pestano, J., Brown, R.P., Suárez, N.M., Fajardo, S., 2003. Phylogeography of pipistrelle-like bats in the Canary Islands based on mtDNA sequences. *Molecular Phylogenetics & Evolution* 26, 56–63.
- Pierre, C., Caruso, A., Blanc-Valleron, M.-M., Rouchy, J.-M., Orszag-Sperber, F., 2006. Reconstruction of the paleoenvironmental changes around the Miocene-Pliocene boundary along a west-east transect across the Mediterranean. *Sedimentary Geology* 188–189, 319–340.
- Renner, S.S., Zhang, L.-B., 2004. Biogeography of the *Pistia* clade (Araceae) based on cp and mtDNA sequences and Bayesian divergence time inference. *Systematic Biology* 53, 422–432.
- Rutschmann, F., 2005. Bayesian Molecular Dating Using PAML/Multidivtime. A Step-By-Step Manual. University of Zurich, Switzerland. Available from: <http://www.plant.ch>.
- Saint, K.M., Austin, C.C., Donnellan, S.C., Hutchinson, M.N., 1998. *C-mos*, a nuclear marker useful for squamate phylogenetic analysis. *Molecular Phylogenetics & Evolution* 10, 259–263.
- Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* 19, 101–109.
- Siesser, W., De Kaenel, E., 1999. Neogene calcareous nanofossils: western Mediterranean biostratigraphy and paleoclimatology. In: Zahn, R., Comas, M.C., Klaus, A. (Eds.), *Proc. ODP Sci. Results* 161. College Station, Texas, pp. 23–237.
- Simmons, M.P., Pickett, K.M., Miya, M., 2004. How meaningful are Bayesian support values? *Molecular Biology and Evolution* 21, 188–199.
- Stanley, S.M., Campbell, L.D., 1981. Neogene mass extinction of Western Atlantic molluscs. *Nature* 293, 457–459.
- Stanley, S.M., Ruddiman, W.F., 1995. Neogene ice age in the north Atlantic region: climatic changes, biotic effects, and forcing factors. In: *Effects of Past Global Changes on Life*. National Academy Press, Washington, DC, pp. 118–133 (Board on Earth Sciences and Resources, Commission on Geosciences, and Resources, National Research Council).
- Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony 4.0 (* and other methods). Sinauer Associates, Sunderland, Massachusetts.
- Terrasa, B., Picornell, A., Castro, J.A., Ramon, M.M., 2004. Genetic variation within endemic *Podarcis* lizards from the Balearic Islands inferred from partial cytochrome *b* sequences. *Amphibia-Reptilia* 25, 407–414.
- Thorne, J.L., Kishino, H., Painter, I.S., 1998. Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution* 15, 1647–1657.
- Thorne, J.L., Kishino, H., 2002. Divergence time estimation and rate evolution with multilocus data sets. *Systematic Biology* 51, 689–702.
- Van de Peer, Y., De Rijk, P., Wuyts, J., Winkelmann, T., de Wachter, R., 2000. The European small subunit ribosomal RNA database. *Nucleic Acids Research* 28, 175–176.
- Veith, M., Kosuch, J., Vences, M., 2003. Climatic oscillations triggered post-Messinian speciation of Western Palearctic brown frogs (Amphibia, Anura, Ranidae). *Molecular Phylogenetics and Evolution* 26, 310–327.
- Webb III, T., Bartlein, P.J., 1992. Global changes during the last 3 million years: climatic controls and biotic responses. *Annual Review of Ecology and Systematics* 23, 141–173.
- Weisrock, D.W., Macey, J.R., Ugurtas, I.H., Larson, A., Papenfuss, T.J., 2001. Molecular phylogenetics and historical biogeography among Salamandrids of the “True” salamander clade: rapid branching and numerous highly divergent lineages in

- Mertensiella luschani* associated with the rise of Anatolia. *Molecular Phylogenetics and Evolution* 18, 434–448.
- Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* 15, 555–556.
- Yang, Z., Yoder, A.D., 2003. Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. *Systematic Biology* 52, 705–716.