MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 542–556 doi: 10.1111/j.1365-294X.2009.04451.x

Nuclear gene phylogeography reveals the historical legacy of an ancient inland sea on lineages of the western pond turtle, *Emys marmorata* in California

PHILLIP Q. SPINKS, ROBERT C. THOMSON and H. BRADLEY SHAFFER

Department of Evolution and Ecology, University of California, Davis, CA 95616, USA, and Center for Population Biology, University of California, Davis, CA 95616, USA

Abstract

The historical biogeography of California's taxa has been the focus of extensive research effort. The western pond turtle (*Emys marmorata*) is an example of a wide-ranging taxon that spans several well-known California diversity hotspots. Using a dataset comprised of one mitochondrial and five nuclear loci, we elucidate the major biogeographic patterns of the western pond turtle across the California landscape. By employing a combination of phylogenetic and network-based approaches, we recovered a relatively ancient (*c.* 2–8 Ma) north/south split among populations of *E. marmorata* and find an area of intergradation centred in the Central Coast Ranges of California. In addition, discordant mitochondrial/nuclear genetic patterns suggest subsequent gene flow from northern populations and from San Joaquin Valley populations into the Central Coast Ranges after the Pliocene-Pleistocene marine embayment of the Great Central Valley subsided. Our results emphasize the utility of nuclear DNA phylogeography for recovering the impact of relatively ancient biogeographic events, and suggest that the Central Coast Ranges of California have played a major role in the geographic structuring of the western pond turtle, and possibly other co-distributed taxa.

Keywords: *Actinemys*, *Clemmys*, *Emys marmorata*, Intron, nuclear gene phylogeography

Received 14 August 2009; revision received 23 October 2009; accepted 28 October 2009

Introduction

Understanding how geographic barriers have contributed to species' distributions and population structure is a focal point of phylogeographic analyses (e.g. Avise *et al.* 1987). Along the west coast of North America, central/southern California is recognized as a phylogeographic 'hot spot' with numerous taxa showing genetic discontinuities around this topographically complex region. Interactions of orogeny, plate tectonics, climatic fluctuations, and historic marine embayments have all contributed to the high endemism and complex genetic structure of many organisms distributed across this landscape (see Calsbeek *et al.* 2003; Lapointe & Rissler 2005; Feldman & Spicer 2006; Chatzimanolis & Caterino 2007; and references therein). The inherent geographic complexity of this region poses a challenge for understanding phylogeographic patterns, especially when the original, historical causes of vicariance are no longer present. Mountain ranges, for example, are obvious, persistent genetic barriers for some taxa. More ephemeral barriers—such as Miocene-Pleistocene marine embayments of the Great Central Valley—are often not readily apparent as barriers to gene flow, but were nevertheless historical landscape features that are increasingly recognized as having had a profound affect on the genetic structure of a wide diversity of taxa distributed across the San Joaquin Valley (Wake 1997; Rodríguez-Robles *et al.* 1999, 2001; Matocq 2002; Shaffer *et al.* 2004; Spinks & Shaffer 2005; Feldman & Spicer 2006; Chatzimanolis & Caterino 2007; Parham & Pappenfuss 2008).

Our understanding of the biogeographic history of California has dramatically increased in the last decade, based in large part on the comparative analyses of

Correspondence: Phillip Q. Spinks, Fax: 530 752 1449;
E-mail: pqspinks@ucdavis.edu

mitochondrial gene trees from codistributed taxa (e.g. Calsbeek *et al.* 2003; Feldman & Spicer 2006; Rissler *et al.* 2006; Chatzimanolis & Caterino 2007). However, phylogeographic analyses are increasingly turning from single-marker analyses toward more sophisticated multilocus approaches incorporating nuclear DNA (nuDNA) sequences in conjunction with coalescent and non-tree like methods to infer complex demographic histories (Hare 2001; Huson & Bryant 2006; Chatzimanolis & Caterino 2007; Leavitt *et al.* 2007; Peters *et al.* 2007; Smit *et al.* 2007; Good *et al.* 2008).

Analyses of multiple nuclear loci have the potential to provide additional insights into historical processes and evolution of taxa, but incorporating nuDNA sequences in phylogeographic analyses is not as straightforward as for mitochondrial DNA (mtDNA). Challenges include intragenic recombination, marker availability, and low variability of nuclear sequences (Hare 2001; Zhang & Hewitt 2003). Most of these issues can probably be addressed, but low variability of nuclear sequences might be a particularly difficult challenge for some taxa. For example, recombination rates have to be relatively high to bias phylogeographic analyses (Schierup & Hein 2000), and network analyses can be used to help elucidate intraspecific polytomies (Posada & Crandall 2001; Huson & Bryant 2006). Low sequence divergence levels probably pose the greatest methodological challenge for nuDNA phylogeographic studies. Empirical results from turtles, our study taxon, have revealed relatively low levels of intraspecific nuclear genetic variation (Caccone *et al.* 2004; Spinks & Shaffer 2005, 2007, 2009; Velo-Antón *et al.* 2007; Fritz *et al.* 2008; McGaugh *et al.* 2008), limiting the impact of nuclear DNA data on phylogeographic analyses. More generally, the prospects for nuclear DNA phylogeography of organisms with relatively low nuclear sequence variation seem poor, even as an ever-increasing body of work demonstrates that single-marker analyses are often plagued by introgressive hybridization and/or incomplete lineage sorting (e.g. Bernatchez *et al.* 1995; Melo-Ferreira *et al.* 2005; Robertson *et al.* 2006; Linnen & Farrell 2007; Peters *et al.* 2007; Good *et al.* 2008; Spinks & Shaffer 2009).

Here we assessed the feasibility of nuclear DNA phylogeography for a relatively widespread species of freshwater turtle (*Emys marmorata*, the western pond turtle) that we examined previously using mtDNA analyses (Spinks & Shaffer 2005). This previous work demonstrated that two nuclear loci had relatively low intraspecific nuclear genetic variation in *E. marmorata*, making this an interesting test case for phylogeographic analysis in the face of limited per-locus nuclear variation. We increased our nuclear locus sampling and employed several methods to determine major patterns of nuclear genetic variation across the landscape, including the existence and pattern of lineage diversification, the degree of admixture between these lineages, and the biogeography of *Emys marmorata* over the last several million years. Finally, we explicitly test several competing hypotheses for mitochondrial, nuclear, and morphological variation within the species.

Emys marmorata *biology*

Emys marmorata is distributed along the Pacific coast of North America, west of the Sierra Nevada/Cascade Range from Washington south to Baja California Norte, Mexico (Storer 1930; Seeliger 1945). There is also a disjunct, possibly introduced population in western Nevada (Holland 1991; Lovich & Meyer 2002; Spinks & Shaffer 2005). Two subspecies are generally recognized; the northwestern pond turtle (*Emys marmorata marmorata*), and the southwestern pond turtle (*E. m. pallida*) (Seeliger 1945). *Emys marmorata marmorata* is morphologically diagnosed by the presence of triangular inguinal plates and dull neck markings, and ranges from the San Francisco Bay region and Sacramento Valley north to Washington, including the Nevada isolate. *Emys m. pallida* usually lacks or has relatively small inguinal plates, has lighter neck markings, and ranges from the San Francisco Bay region south to Baja California Norte including the disjunct Mojave River population in inland southern California. The San Joaquin Valley has been described as a region of intergradation (Fig. 1) (Seeliger 1945). A previous phylogeographic analysis of *E. marmorata* revealed no significant variation from two nuclear introns sampled across the range of the species. However, analyses of mtDNA revealed four deeply divergent clades that were not geographically congruent with the two subspecies, including: (i) a Northern clade comprised of populations from the Coast Ranges from San Luis Obispo County, northwards to Oregon, Nevada and Washington; (ii) a San Joaquin Valley clade which was composed of populations east of the Coast Ranges and west of the Sierra Nevada from the San Francisco Bay Delta south to the Tehachapi Mountains; (iii) a geographically restricted Santa Barbara clade which included populations from the Pacific Coast of San Luis Obispo County, and Transverse Ranges of Santa Barbara and Ventura Counties; and (iv) a Southern clade that included some individuals from Santa Barbara County, plus all remaining populations from south of the Transverse Ranges to Baja California Norte including the Mojave Desert (Fig. 1). Phylogeographic breaks between the Southern, Santa Barbara, and Northern + San Joaquin Valley clades were fairly well demarcated by the Transverse and Coast Ranges, but there are no obvious contemporary barriers between the San

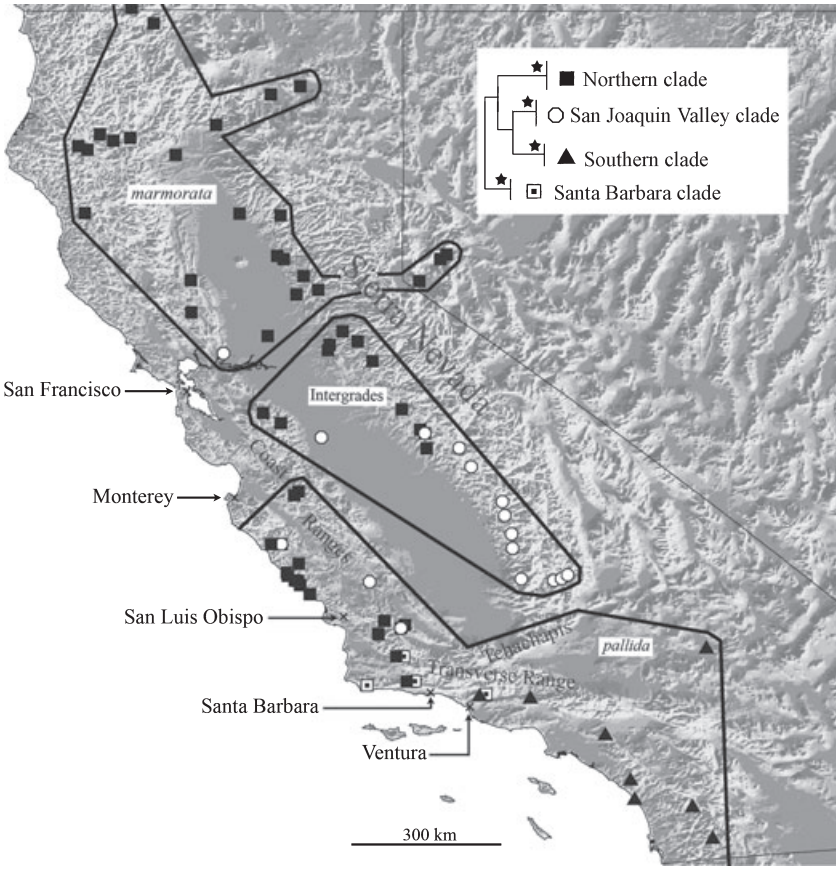

Fig. 1 Map showing collection localities, and mitochondrial haplotypes for California samples. All samples from Oregon and Washington had Northern clade mitotypes while all samples from Baja California Norte, Mexico had Southern clade mitotypes (see Supporting Information for geographic coordinates). The dark lines encompass the region of intergradation defined by Seeliger (1945), as well as the range of *E. m. marmorata* and *E. m. pallida*. All remaining samples occur within the range of *E. m. marmorata*. Stars on the phylogenetic tree indicate nodes with Bayesian posterior probabilities ≥0.99/ML bootstrap support values of ≥80.

Joaquin Valley and Northern clades. However, Pliocene/Pleistocene marine embayments of the Great Central Valley of California (Peabody & Savage 1958; Yanev 1980; Dupré 1990; Dupré *et al.* 1991) have been hypothesized to be the cause of this phylogeographic break (Spinks & Shaffer 2005).

Materials and methods

Taxon and data sampling

Available GenBank sequences included a fragment of the mitochondrial nicotinamide adenine dehydrogenase subunit 4 gene and flanking $tRNA^{His}$ (hereafter referred to as ND4), intron 1 of the fingerprint protein 35 (R35), and a fragment of intron XI of the glyceraldehyde-3-phosphate dehydrogenase, GAPDH gene. We downloaded and included 68 ND4 sequences and 40 R35 sequences from GenBank, but did not include the GAPDH sequences since they were mostly invariant across populations of *Emys marmorata* (Spinks & Shaffer 2005). Based on previous mtDNA analysis, most genetic variation appears to be among populations from central California to Baja California Norte. Thus, we intensified our taxon/data sampling from central California and Baja California Norte (Supporting Information). We also generated ND4 sequence data from an additional six individuals from Nevada since the provenance of this population remains controversial (Holland 1991; Spinks & Shaffer 2005). Relationships among the turtles in the subfamily Emydinae are somewhat obscure (Spinks & Shaffer 2009; Spinks *et al.* 2009), so we included samples of the emydine genera *Clemmys*, *Glyptemys* and *Terrapene* as outgroups. (Supporting Information). The R35 sequences from GenBank were ~ 500 bp in length, but we resequenced these individuals for a final read length of ~ 770 bp. DNA extraction methods and PCR conditions for ND4 and R35 follow Spinks & Shaffer (2005). We generated sequences from four additional nuclear loci from 90 *E. marmorata* and most outgroup taxa

including *RELN* (Spinks & Shaffer 2007), *HNF-1α*, (Primmer *et al.* 2002), *TGFB2* (Primmer *et al.* 2002), and *c-myc* (Harshman *et al.* 2003). All PCR products were sequenced in both directions on ABI 3730 automated sequencers at the UC Davis Division of Biological Sciences Sequencing Facility (http://dnaseq.ucdavis.edu/).

Phylogenetic analyses

Given previously demonstrated nuclear-mitochondrial gene tree discordances in *Emys* (Spinks & Shaffer 2009), all analyses were performed separately for mitochondrial and nuclear loci. Putative nuclear mitochondrial pseudogenes (numts) have been encountered in the ND4 region of *E. marmorata* (Spinks & Shaffer 2005). Thus, the ND4 coding regions were translated using MacClade 4.06 (Maddison & Maddison 2003) to check for numts. Maximum likelihood (ML), and Bayesian analyses were carried out in PAUP* 4.0b10 (Swofford 2002) using subtree pruning and regrafting (SPR) branch swapping and ten random stepwise heuristic searches with each search limited to 1 million rearrangements. Models of molecular evolution for parameter estimation were selected using the DT-Modsel software (Minin *et al.* 2003) with parameter values estimated in PAUP* 4.0b10. Support for each phylogeny was assessed with 100 bootstrap pseudoreplicates (Felsenstein 1985), with each pseudoreplicate limited to 1 h of computation time.

Single and partitioned model Bayesian analyses were performed using MrBayes V3.1.1 (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003). Models of molecular evolution for individual partitions or combinations of partitions of the mtDNA sequences were estimated using the DT-Modsel software. The optimal partitioning strategy for the mtDNA sequences was determined using comparisons of Bayes factors from preliminary Bayesian analyses (Kass & Raftery 1995; Nylander *et al.* 2004). All preliminary analyses were performed with two replicates and four chains for ten million generations, and the chains were sampled every 1000 generations. Final analyses of the preferred partitioning strategy were performed as above but with 50 million generations sampled every 5000 generations. Stationarity was determined as the point when the potential scale reduction factor (PSRF) reached 1, and the −log likelihood (−ln*L*) scores plotted against generation time stabilized (Huelsenbeck & Ronquist 2001). The first 25% of samples were discarded as burnin. To accommodate among-partition rate variation (APRV) in the partitioned model analyses, we invoked the *prset ratepr = variable* option in MrBayes V3.1.1 (Ronquist & Huelsenbeck 2003).

Population assignment of individuals

Seeliger's (1945) hypothesis includes two subspecies as well as an area of intergradation, and we used the Structure v2.2.3 software (Pritchard *et al.* 2000) and allelic data from our five nuclear locus data set generated via the Phase2.1.1 software (Stephens *et al.* 2001; Stephens & Donnelly 2003) to test this hypothesis. We used the Bayesian approach implemented in the Phase2.1.1 software to reconstruct probable pairs of haplotypes for each sequence. We used the default settings except that we accepted haplotype reconstructions with Bayesian posterior probabilities (BPP) of ≥95% only; characters with ≤94% BPP were coded as ambiguous data. The Structure software implements a Bayesian clustering method to estimate the number of populations (*K*) represented within the data, and then assigns individuals probabilistically to the recovered set of populations so as to maximize Hardy–Weinberg equilibrium within populations. We set *K* to 2 in order to test the hypothesis that there are two populations (subspecies) contributing to the gene pool of our sample. Putatively admixed individuals (intergrades) can be estimated by assuming that admixed individuals inherit some fraction of their genome (*q*) from each parental population (Pritchard *et al.* 2000). Simulations have shown that a threshold *q* value of 0.10 is most efficient for correctly classifying purebred and hybrid individuals (Vähä & Primmer 2005), thus we considered individuals with *q* values between 0.10 and 0.90 to be admixed. We employed the correlated allele frequency model and the admixture ancestry model because they appear to perform better than the alternatives in the face of low sequence variation and subtle population structure (Falush *et al.* 2003). We used the default settings for the remaining parameters. We ran five independent analyses, each for 1 million generations with a burnin of 100 000 MCMC generations. In addition, we used the DnaSP V5.1 software (Librado & Rozas 2009) to estimate genetic diversity and fixation indices within and among the northern and southern groups.

Network analyses

Molecular phylogenies are reconstructed under the assumption that the data arose through a branching process, but at the intraspecific level population processes such as recombination, for example produce reticulate relationships. Phylogenetic networks account for reticulations by displaying conflicting signal as loops or reticulations (Huson 1998). Thus, we used SplitsTree version 4.10 (Huson 1998; Huson & Bryant 2006) with uncorrected 'p' genetic distances and the NeighborNet algorithm (Bryant & Moulton 2004) to

generate a phylogenetic network of our concatenated data set. In addition, we performed a bootstrap analysis with 1000 replicates, and tested for recombination within each locus using the pairwise homoplasy index (PHI) statistic (Bruen *et al.* 2006) implemented in SplitsTree4.10. Although admixture is a real component of the evolutionary history of a lineage, it can also obscure relationships and reduce statistical confidence for widespread, non-admixed lineages. We therefore performed two analyses, one on the entire nuclear sequence data set and one with putatively admixed individuals (identified via the Structure V2.2.3 analysis) removed.

Phylogenetic analyses of allelic data

Although they are potentially phylogenetically informative, most current phylogenetic reconstruction methods do not utilize allelic data (Joly & Bruneau 2006). We used the POFAD software (Phylogeny of Organisms From Allelic Data, Joly & Bruneau 2006) to generate a multilocus phylogeny. The POFAD algorithm combines genetic distance matrices generated from allelic data from individual loci into a single genetic distance matrix. We used PAUP* to generate distance matrices (uncorrected 'p' genetic distances) for all five loci, and then used POFAD to generate a standardized combined-locus distance matrix from the five loci which was then imported into PAUP* and used to reconstruct a neighbour-joining (NJ) phylogram.

To assess statistical support for the POFAD tree, we performed a nonparametric bootstrap procedure. First, we produced 100 nonparametric bootstrap pseudoreplicate datasets for each locus using the Seqboot module of the PHYLIP 3.66 package (Felsenstein 2005) where each pseudoreplicate data set was of the same length as the original locus. Pseudoreplicate data sets from each locus were then drawn without replacement and concatenated to create 100 five-locus data sets. Next, POFAD was used to generate a NJ tree from each pseudoreplicate data set. Finally, the resulting 100 NJ trees were imported into PAUP* and used to generate a majority-rule consensus tree. The preceding bootstrap procedure was largely automated in Perl (code available from http://www.eve.ucdavis.edu/rcthomson). As in the network analyses, we performed this analysis on the entire nuDNA sequence data set, and with the putatively admixed individuals removed prior to analysis.

Genetic landscape shapes

In order to visualize genetic variation across the landscape, we used the Alleles in Space (AIS) software (Miller 2005) which employs an inverse distance weighted interpolation method to generate three-dimensional surface plots of genetic/geographic variation or 'genetic landscape shapes' (Miller 2005). The genetic landscape shapes (GLS) are analogous to geographic topography maps such that regions with high genetic distances between individuals from different sampling localities are equivalent to high elevations, and regions of low genetic distances are equivalent to low elevations (Vignieri 2007). For input settings, we used 'Midpoints of pairwise distances of all observations', 'Residual genetic distances' and a distance weight value of 0.25. Sampling localities were input as latitude/longitude, but samples from Oregon and Washington were excluded from these analyses since they showed virtually no variation in previous analyses, and this large region was relatively sparsely sampled (Spinks & Shaffer 2005). The AIS software can handle nucleotide sequence data from one locus only, but can handle data from multiple loci in the form of haplotypes. Thus, for the mtDNA analysis we used nucleotide sequences, but for the nuDNA analysis we used haplotypes from the Phase2.1.1 output. To visually display the AIS results in a geographical context, we imported the GLS *X*, *Y*, *Z* output coordinates (where *X* and *Y* values are latitude and longitude, respectively and Z values are interpolated genetic distances represented as heights) into ArcView GIS 3.2 (ESRI), and plotted them on a map of western North America.

Hypothesis testing

Using our mtDNA data, we explicitly tested the nu DNA and morphology-based hypotheses using Bayesian tests of monophyly and Shimodaira–Hasegawa (SH) tests. Under the Bayesian framework, the probability that a tree is correct is the posterior probability of a tree conditioned on the model and data (Huelsenbeck & Rannala 2004), therefore the Bayesian framework offers a straightforward way to evaluate monophyly (Linnen & Farrell 2007). First, we used MacClade v4.06 to construct constraint trees corresponding to the phylogenetic network, POFAD phylogeny, and the morphology-based subspecies split. The constraint trees were constructed such that the hypothesized major groupings or clades (morphological subspecies, phylogenetic network or POFAD clades) were monophyletic, but with no constraints on substructure within groups/clades. We excluded all outgroups and all samples that did not occur in the nuclear sequence data set. For the subspecies hypothesis only, we excluded sequences collected from specimens inhabiting the region of intergrade proposed by Seeliger (1945) (see Supporting Information). These constraint trees were then imported into PAUP* and used to filter the corresponding post-burnin posterior distribution of trees from the mtDNA analyses. If

less than 1.7% of trees ($\alpha = 0.05$, after Bonferroni correction $\alpha = 0.05/3$) were retained then the null hypotheses (no conflict between trees) were rejected (Miller *et al.* 2002; Buschbom & Barker 2006; Linnen & Farrell 2007). Likelihood scores of constrained and unconstrained searches were assessed using SH tests carried out in PAUP* using the RELL approximation (1000 replicates), and the same constraint trees as above.

Dating phylogenetic splits

To approximate the timing of splits within *E. marmorata*, we used a slightly expanded dataset and the identical dating approach used in Spinks & Shaffer (2009). We used the penalized likelihood (PL) method (Sanderson 2002) implemented in the program r8s v1.70 (Sanderson 2003), fossil calibrations and estimated divergence times reported by Near *et al.* (2005), and the cytochrome *b* (cyt*b*) dataset described in Spinks & Shaffer (2009). We retained two individuals of each of the four mitochondrial clades (except Santa Barbara, where we retained one) described in Spinks & Shaffer (2009). Based on these seven sequences, we placed both a date and confidence intervals on the divergence of crown group *marmorata*.

Results

Mitochondrial DNA: phylogeny and GLS

Sequencing chromatograms from the ND4 sequences generated here showed no indication of heterozygous positions, protein-coding regions successfully translated to amino acid sequences, and we observed the expected bias against guanine nucleotides (e.g. Kocher *et al.* 1989) (A = 0.35%, C = 0.27%, G = 0.12%, and T = 0.26%). Thus, we consider our sequences to represent authentic mtDNA. Our mitochondrial data set was composed of up to 725 bp from 152 individuals including 147 *Emys marmorata* and five outgroups (for details, see Supporting Information). This data set was almost complete with 0.8% missing data. Analyses of this expanded ND4 data set recovered a phylogeny that was very similar to our previous analysis (Spinks & Shaffer 2005; their Fig. 3). Thus, we do not show the mtDNA tree, but indicate the mitochondrial clade membership of each sample in Fig. 1 (see Supporting Information for the mtDNA phylogeny). Our additional mtDNA sampling further refined the geographic distribution of both the Northern and San Joaquin Valley clades such that Northern clade haplotypes were recovered from additional sites in the Sierra Nevada foothills (Amador County), northeastern California (Modoc County) and coastal southern California (Santa Barbara County), while San Joaquin Valley clade haplotypes were recovered from additional sites in coastal (Monterey, San Luis Obispo Counties) and southern (Santa Barbara County) California (Fig. 1). The new Nevada samples were identical to those recovered previously (Spinks & Shaffer 2005, Supporting Information).

The two-dimensional GLS plot of the mtDNA data revealed that the region containing the greatest genetic distances (highest elevation) between sampling localities was centred on the San Joaquin Valley, Central Coast Ranges and Transverse Ranges of central California, corresponding to the intersection/overlap region of all mitochondrial clades (Fig. 2a). In addition, the relatively high mitochondrial sequence variability of coastal southern California compared to northern California was reaffirmed by the higher genetic diversity of this region. Conversely, the relatively low mitochondrial sequence variability in northern California was reflected in the low genetic diversity among populations in the northern portion of the state (Fig. 2a).

Nuclear loci

Our nuclear data was composed of up to 3376 bp per specimen including 423 bp of *c-myc*, 582 bp of *HNF-1α*, 777 bp of *R35*, 1014 bp of *RELN*, and 580 bp of *TGFB2*. This data set was almost complete, containing $\sim$ 1.4% missing data. All sequences were deposited into GenBank (for accession #s, see Supporting Information). Alignments were made in MacClade v4.06 with final modifications made by hand, and deposited into TreeBase (http://www.treebase.org, accession # S2498). Most fragments were trimmed in order to minimize missing data. The number of inferred haplotypes/locus varied from 5 (TGFB2) to 28 (RELN) (Table 1), and sequence divergence (uncorrected p) among *E. marmorata* samples varied from 0.51% to 0.87% (R35 and HNF-1α, respectively, Table 2). Phylogenetic analyses of individual loci recovered little phylogenetic structure (not shown).

The nuclear GLS plot (Fig. 2b) was qualitatively similar to the mtDNA GLS plot in several important respects. In both cases, parts of the San Joaquin Valley and coastal Southern California were recovered as regions of highest genetic variation between sampling localities, while California north of San Francisco was a region of lowest variation among sampling localities. However, the nuDNA GLS plot indicated that the variation fell off quickly as one progresses from the southern San Joaquin Valley west to the central coast, with the region from Santa Barbara to Monterey one of the lowest regions of genetic distances among sampling localities. In addition, the nuDNA data show the extreme western edge of the Mojave Desert and virtually all of southern California as a region containing high interpo-

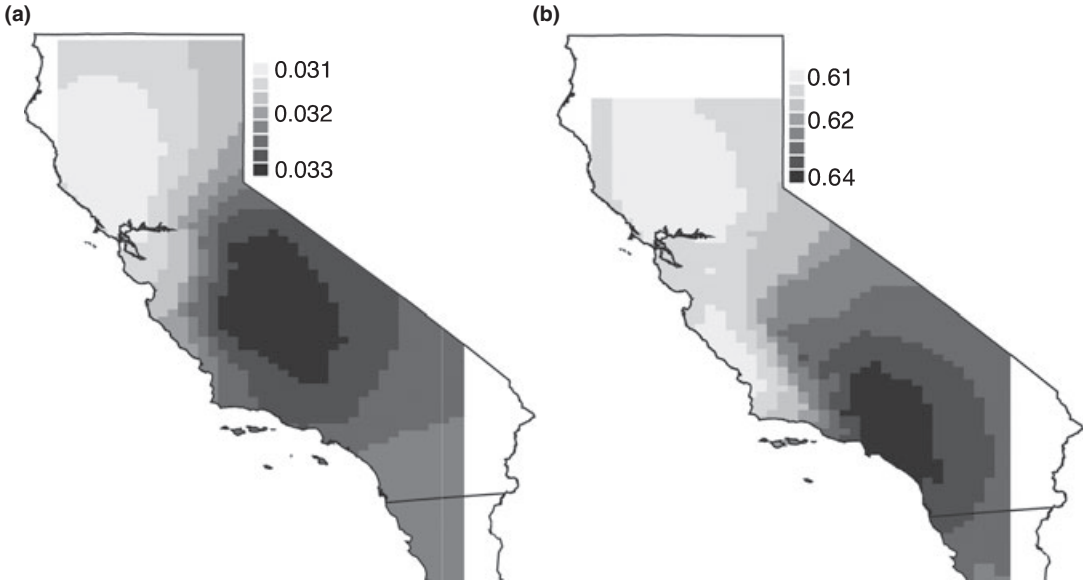

Fig. 2 Genetic heat maps generated from (a) 725 bp mtDNA, and (b) nuclear allelic data (five loci) generated via the AIS software. Genetic topography varies from low (light grey) to high (black). Interpolated genetic distances were arbitrarily assigned to eight bins. Note: scales are not comparable because genetic distances generated from mitochondrial sequences vs. allelic data are not equivalent.

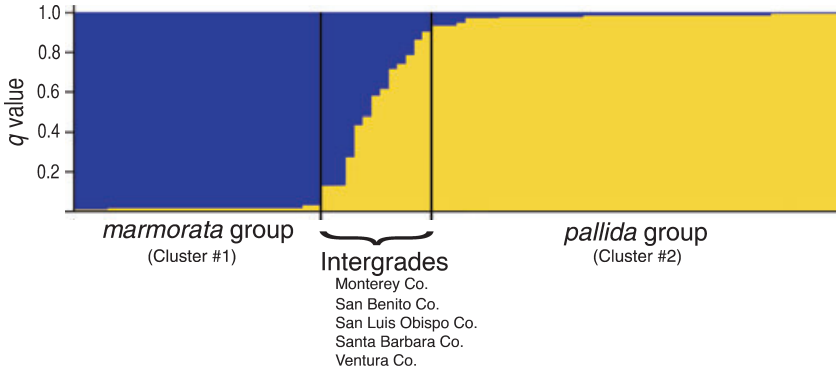

Fig. 3 Bar plot of population assignment of individuals generated using Structure v2.2.3. Individuals with q values > 0.90 were assigned to either the *marmorata* (blue colour) or *pallida* (gold) group. Individuals with q values between 0.10 and 0.90 were putatively admixed, and all putatively admixed individuals were collected from Monterey, San Benito, San Luis Obispo, Santa Barbara, and Ventura County.

Table 1 Geographic distribution of nuclear alleles with special emphasis on Central Coast Range (CCR) and Baja California Norte (BCN) populations. Alleles in bold are unique to that area. For example, *cmyc* alleles 5 and 8 were recovered form Central Coast Range populations only while all Baja California Norte samples were homozygous for *cmyc* allele 3

Geographic distribution of nuclear alleles					
mtDNA clade	*cmyc*	*HNFL*	*R35*	*RELN*	*TGFB2*
Northern	1,3,**10**	1,3,4,13,14,**15**	**1**,2,**4**,12,13	1,**3**,9,12,15,23,24	1,2,3
CCR	3,4,**5**,6,**8**	1,2,3,**5**,6,9,**11**,16	2,5,9,**11**,12	**2**,4,**5**,6,**7**,8,**13**,**14**,15,16,**17**,**18**,**19**,**20**,21,22,**26**,**27**,**28**	1,2,**5**
San Joaquin Valley	1,**2**,3,4,6,**9**	1,2,3,4,6,9,**13**,14	2,**3**,5,**6**,**7**,13	1,4,9,**10**,12,15,16,23,24	1,2,3
Santa Barbara	1,3,4,**7**	1,3,**7**,8,9,**10**,14	2,8,9,**10**	1,4,6,8,16,21,24	1
Southern	3,4,**11**	8,9,**12**	2,8,9	6,8,**11**,21	1
BCN	3	8,9	8	8,21,22,**25**	**4**

Table 2 Table of uncorrected 'p' sequence divergence levels/locus and basic descriptive statistics including estimates of allelic and haplotype diversity, and population F_{ST} values

	cmyc	*HNFL*	*R35*	*RELN*	*TGFB2*
Uncorrected 'P' sequence divergence	0.62%	0.87%	0.51%	0.56%	0.55%
Nucleotide diversity, Pi					
Northern group	0.0014	0.002	0.0007	0.0019	0.0001
Southern group	0.0015	0.004	0.0014	0.0021	0.0016
Total	0.002	0.004	0.0013	0.0029	0.0014
Haplotype diversity, Hd					
Northern group	0.526	0.735	0.47	0.799	0.503
Southern group	0.571	0.758	0.491	0.85	0.352
Total	0.667	0.82	0.497	0.91	0.424
F_{ST}	0.452	0.349	0.204	0.491	0.076

lated genetic distances among sampling localities while the mtDNA shows more modest among-locality variation in these areas (Fig. 2).

Population assignment of individuals

Of the 90 individuals in the nuDNA data set, the 29 individuals from the San Francisco Bay area and San Joaquin Valley (except for Stanislaus Co. samples) north to Washington were assigned to the cluster #1 while the 49 samples from the vicinity of Monterey Bay south to Baja California Norte were assigned to cluster #2 (Fig. 3). Cluster #1 and #2 were roughly equivalent to *E. m. marmorata* and *E. m. pallida*, respectively. The remaining 12 samples were identified as putatively admixed, and all of the admixed samples were collected from central/southern coastal California (San Benito, Monterey, San Luis Obispo, Santa Barbara or Ventura County, see Supporting Information). Nucleotide and haplotype diversity were higher in the southern group at 5/5 and 4/5 loci, respectively, while estimates of gene flow varied from a low of 0.076 (*TGFB2*) to a high of 0.491 (*RELN*, Table 2).

Phylogenetic network

The PHI test detected no significant level of recombination within any locus ($P = 1.0$ in all cases), thus these data were appropriate for phylogenetic network analysis. We analyzed each locus individually, but recovered little structure (not shown). However, analysis of the concatenated nuclear sequences recovered two major groups with a clear but weakly supported split (bootstrap support value = 42) between northern and central/southern California populations, with some individuals from central California intermediate between the two major groups (Fig. 4a). Importantly, most of these intermediate individuals were identified as admixed in the Bayesian clustering analysis (Supporting Information). When these admixed individuals were removed, bootstrap support values for the north/south split increased substantially to 82 (Fig. 4b) such that populations from the San Joaquin Valley and east-flowing (that is, interior) drainages of the Coast Ranges north to Washington formed an exclusive '*marmorata* group' with respect to populations from the Coast Ranges south to Baja California Norte in the '*pallida* group' (Fig. 4). Essentially, the *marmorata* group was mostly equivalent to the mitochondrial Northern clade + San Joaquin Valley clade samples while the *pallida* group was mostly equivalent to the mitochondrial Southern clade + Santa Barbara clade samples (compare Figs 1, 3, 4).

Phylogeny of organisms from allelic data

The POFAD nuDNA phylogeny was mostly unsupported; the single exception was the strongly supported monophyly of the Baja California Norte turtles (Fig. 5). Removing the putatively admixed individuals had little impact on these results, and we show the phylogeny generated from the entire nuclear sequence data set. Although not strongly supported, four mostly geographically coherent major clades were recovered in this analysis (clades a-d, Fig. 5) that are largely congruent with the *marmorata* and *pallida* groups recovered from the network analysis. Clades a and d contained all of the *marmorata* group individuals, while clades b and c contained all of the *pallida* group individuals; the sole exceptions were five central/southern California samples that fell within clades a and d (Fig. 5).

Discussion

Historical biogeography of Emys marmorata *in California*

Our phylogenetic results indicate three evolutionary features in the *E. marmorata* tree that may be associated with the historical geology of California. The first is the north-south split that separates turtles from roughly Santa Barbara south, and those from San Francisco north. Second is the considerable admixture of northern and southern groups in the central Coast Range from about Monterey to Santa Barbara/Ventura. Third is the further clarification of the distribution of regional clades in the San Joaquin and Santa Barbara regions. The first two are features of both the mitochondrial (Fig. 1) and nuclear data sets (Figs 3, 4 and 5), while the third is solely a mitochondrial feature (Fig. 1). Based on our cyt*b* analysis, we estimated a mean diversification age

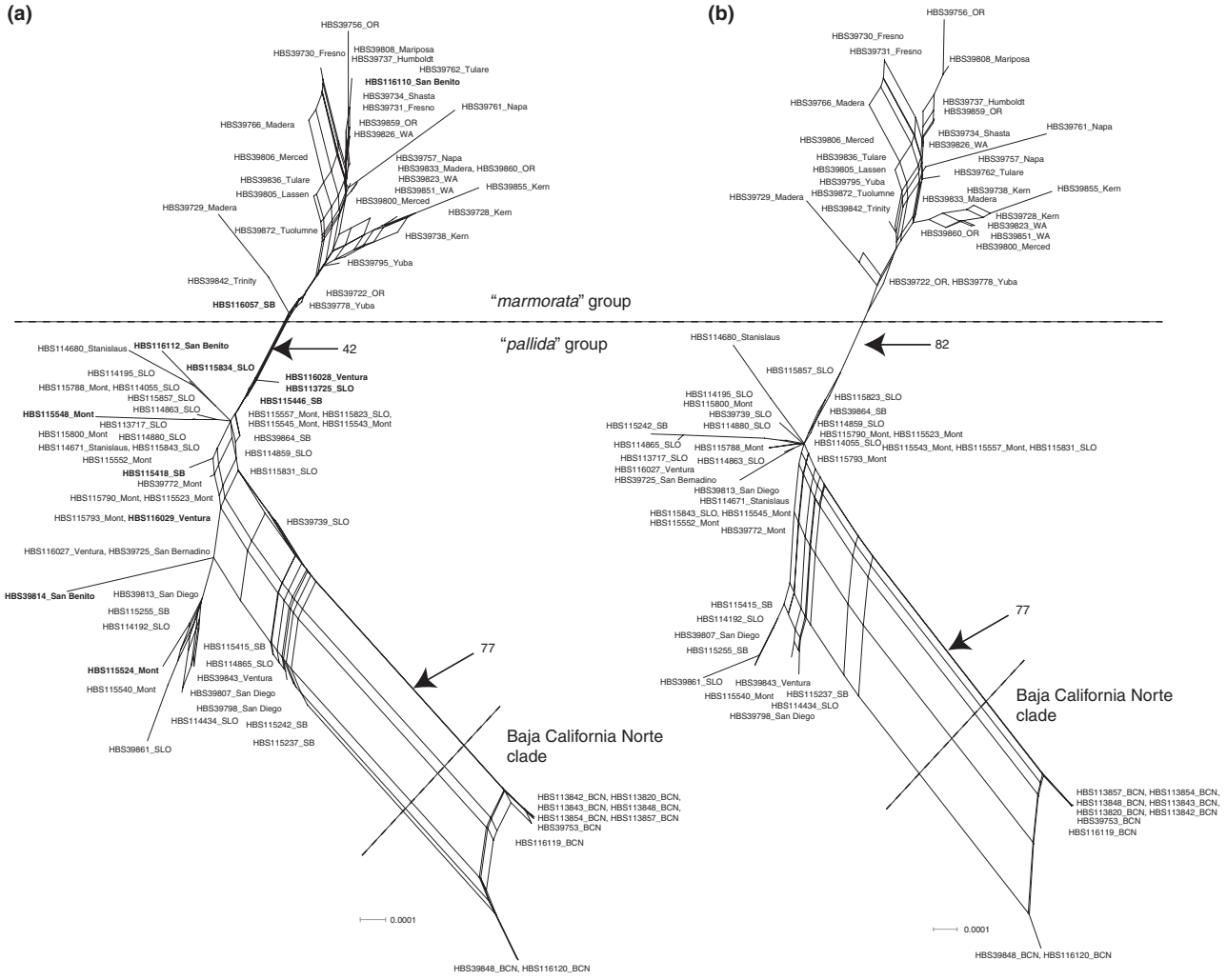

Fig. 4 Phylogenetic networks of the concatenated nuclear loci using uncorrected 'p' genetic distances, and the NeighborNet algorithm. Abbreviations: BCN, Baja California Norte; Mont, Monterey; SLO, San Luis Obispo; SB, Santa Barbara. Analyses were performed on (a) the entire nuclear sequence data set including putatively admixed individuals (in bold), and (b) with putatively admixed individuals removed. Bands of parallel edges represent incompatible splits. For example, some nucleotide characters in the putatively admixed individuals conflict with the *marmorata/pallida* split that is represented in panel (a) as multiple parallel edges. Thus, removing the conflicting data results in fewer parallel edges. With the exception of one sample (HBS116057) from Santa Barbara County, California the *marmorata* group contained populations from the San Joaquin Valley north to Washington while the *pallida* group contained populations from the Central Coast Ranges south to Baja California Norte, Mexico. Bootstrap support values for the split between the *marmorata* and *pallida* groups (indicated with arrows) varied from 42 for the full data set to 82 with admixed individuals removed. Bootstrap support values for the split between Baja California Norte populations from all others was 77 in either analysis.

of living *E. marmorata*-group animals at about 4 Myr (SD = 1.13, min = 2.33, max = 8.8 Myr). Therefore, our working hypothesis is that the divergence of the major *marmorata* lineages initiated as early as *c.* 8–9 Ma in the late Miocene, but more likely *c.* 4 Ma in the mid-late Pliocene.

The paleogeography of California is extremely complex, particularly for the Pliocene-Pleistocene of the Central Coast Ranges and San Joaquin Valley. There is general agreement that from the Miocene (about 24 Ma) through the Pleistocene, both the San Joaquin Valley and the Central Coast Ranges experienced several marine embayments resulting in a large inland sea that covered much of the San Joaquin Valley and isolated most of the Central Coast Ranges from the remaining California mainland (Durham 1974; Hall 2002; Bowersox 2005). Of particular note was the large marine embayment that occupied much of the current San Joaquin Valley and connected to the Pacific at what is now Monterey Bay (Hall 2002). This connection was present continuously

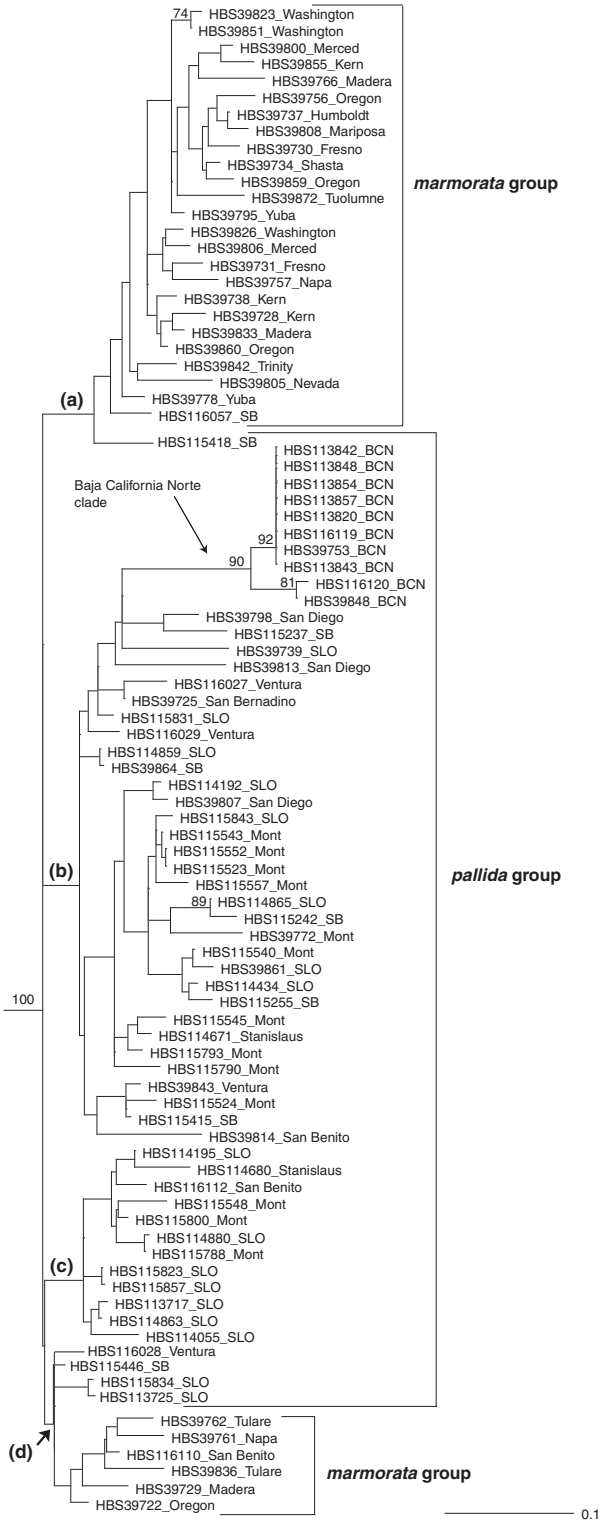

Fig. 5 Neighbour joining phylogram based on a combined distance matrix generated using the POFAD algorithm. Except for five individuals, clades a and d are equivalent to the *marmorata* group while clades b and c are equivalent to the *pallida* group. Bootstrap support values (above branches) were generated using a novel nonparametric approach (see text).

from *c.* 8 to 2.5 Ma, effectively cutting off the central California coastal mountains from freshwater or terrestrial contact with the landscape north of Monterey (Fig. 6). Also, throughout this period, much of current Santa Barbara County existed as an isolated peninsula that was never inundated, but was isolated from most of the California mainland (Dupré *et al.* 1991; Hall 2002). Finally, the extensive Miocene drying period that characterized California may have resulted in range contractions of freshwater organisms including turtles, with subsequent expansion in the Pleistocene and Holocene (Jacobs *et al.* 2004).

Based on Bayesian tests of monophyly and SH tests, the mtDNA phylogeny was not congruent with the topology recovered from the phylogenetic network or POFAD analysis, nor was it congruent with the subspecies split ($P = 0.00$ in all cases). However, from a landscape perspective, both the nuclear (Fig. 6a) and mitochondrial (Fig. 6b) datasets show a very strong north-south division in *E. marmorata* with considerable admixture in the Central Coast Ranges. The nuclear phylogenetic network indicates distinct northern and southern groups with considerable admixture in the Central Coast region as well (Fig. 4). Although the mtDNA demonstrates four distinct clades, their among-clade relationships are not well resolved (Fig. S1). If, however, the mitochondrial San Joaquin Valley and Northern clades are combined, their range is virtually identical with the nuclear northern group; the same is true of the combined mitochondrial Southern + Santa Barbara clades and the nuclear southern group (Figs 4 and 6). Thus, while the mtDNA phylogeny and nuDNA/morphology-based hypotheses are topologically incongruent, the distribution of pure nuDNA northern group individuals (which we refer to as '*marmorata*') includes all terrestrial habitat north and east of the reconstructed inland sea that dominated the San Joaquin Valley from *c.* 8 to 2 Ma and emptied into current-day Monterey Bay. The sole exception to this pattern is a small area of southern-group ('*pallida*') nuclear leakage in the current region of Monterey Bay. The same is true for pure nuDNA (*pallida*) individuals south of the transverse ranges from Santa Barbara south (Fig. 6). However, the central coast region from Monterey to Santa Barbara, which was isolated from northern California until the Pleistocene (Hall 2002), is an area of north-south admixture for *marmorata* and *pallida* nuDNA groups, and all four mitochondrial clades also overlap in this area.

Taken together, our combined mtDNA and nuclear results provide a novel perspective on the biogeographic evolution of *E. marmorata* that is consistent with the known geology of Central and Southern California over the last *c.* 2–8 Myr of *Emys* history. First, our nuDNA

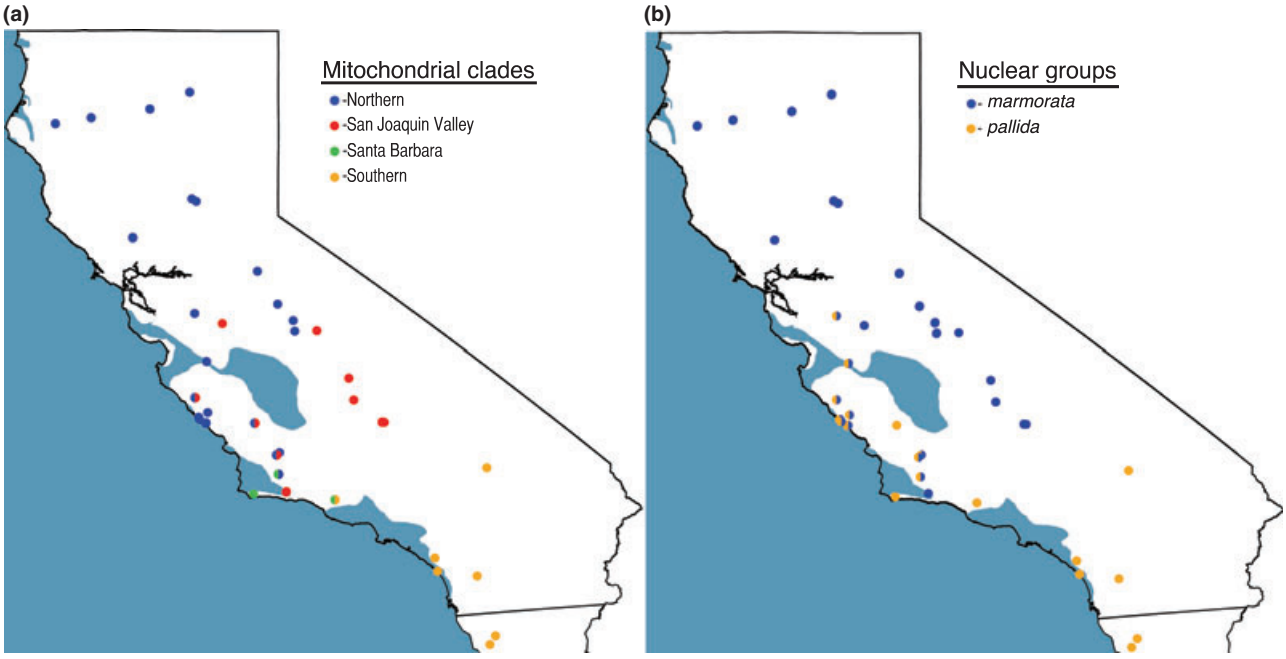

Fig. 6 Maps showing distribution of mitochondrial clades (a) and nuclear groups (b) across California. All samples from Oregon and Washington have mitochondrial northern clade mitotypes and were assigned to the *marmorata* nuclear group. Conversely, all samples from Baja California Norte have mitochondrial southern clade mitotypes and were assigned to the *pallida* nuclear group. Blue indicates sea level *c.* 8–5 Ma (after Hall 2002).

results reveal a relatively ancient north/south genetic break in the mid Central Coast Ranges with limited gene flow between these groups. The *marmorata* and *pallida* groups were probably historically allopatric along the west coast of North America, with a strong biogeographic separation formed by the San Joaquin Valley inland sea. Currently, most individuals sampled from the Central Coast Ranges have northern or San Joaquin Valley mitotypes, but possess nuclear haplotypes that represent pure southern or admixtures of southern and northern group haplotypes. Given that there has been a continuous connection between the Central Coast Ranges and coastal southern California during the entire *c.* 2–8 Myr crown-clade *E. marmorata* group history, and the presence of nuclear southern group genotypes throughout the region, it appears most likely that the Central Coast Range was historically a part of the Southern *pallida* group. When the marine connection between the San Joaquin Valley and the Pacific Ocean at Monterey Bay was interrupted in the late Pliocene (Dupré *et al.* 1991; Shaffer *et al.* 2004) southerly movement and admixture of northern turtles both from the San Joaquin Valley and from coastal sites would have led to the mitochondrial and nuclear admixture patterns seen today, with complete replacement of southern mtDNA but some admixture at the nuclear level (Fig. 6). Whether this southerly expansion of northern genes is continuing or has stopped at the Transverse Ranges, a major biogeographic boundary for many taxa (Tan & Wake1995; Rodríguez-Robles *et al.* 1999, 2001; Maldonado *et al.* 2001; Calsbeek *et al.* 2003; Spinks & Shaffer 2005; Feldman & Spicer 2006; Kuchta & Tan 2006; Rissler *et al.* 2006; Parham & Pappenfuss 2008), is currently unknown.

Taxonomy and management of the western pond turtle

Understanding and describing diversity is at the core of phylogenetic and phylogeographic analyses. However, describing diversity unambiguously can be extremely difficult, especially in the early stages of speciation when lineages are not expected to be reciprocally monophyletic (de Queiroz 1998; Hudson & Coyne 2002; Shaffer & Thomson 2007). Even so, the genetic insights presented here should form a solid foundation for both management and taxonomic decisions for this declining Species of Special Concern in California (Jennings & Hayes 1994; http://arssc.ucdavis.edu).

Our nuclear results, in conjunction with variation in morphology, illuminated a relatively ancient split between northern and southern populations of *E. marmorata*. However, the four mitochondrial clades represent cohesive evolutionarily significant units (Moritz 1994) that are not at odds with the nuclear results. Rather, they suggest that the southern San Joaquin Valley and Santa Barbara clades are younger splits

than the fundamental *marmorata/pallida* division. From a management perspective, we feel that the safest and most prudent strategy is to consider the Northern, San Joaquin Valley, Southern and Santa Barbara clades as four separate entities. Taxonomic decisions are less clear, although we are making progress toward the resolution of species boundaries within the *marmorata* complex. We recovered a deep north/south split within the western pond turtle, suggesting that an alternative to the current subspecies descriptions could be to elevate the *marmorata* and *pallida* groups (Fig. 4) to full species status as *Emys pallida* and *Emys marmorata* (Seeliger 1945). However, conflicts between the morphologically and genetically defined regions of intergradation render this decision problematic. For example, based on morphological studies, Seeliger (1945:158) reported the range of *E. pallida* as 'Mojave River and the coastal region of southern California northward from the Mexican border, west of the San Joaquin Valley, to the vicinity of Monterey Bay; intergrades in the San Joaquin Valley and in the San Francisco Bay region with *Emys marmorata marmorata*'. However, based on our combined data, the Coast Ranges from the San Francisco Bay area south to the Santa Ynez Mountains is a region of genetic admixture among mtDNA clades as well as nuclear groups (Figs 3, 4 and 5) such that there is significant uncertainty regarding the genetic or morphologic assignment of turtles inhabiting the San Joaquin Valley and Central Coast Ranges. Perhaps the additional nuclear DNA genotyping underway by our group will help clarify the taxonomic status of the Central Coast Range populations, and we await these results before suggesting final taxonomic resolution.

Acknowledgements

We thank Dan Holland for tissue samples, Laura Patterson, Sean Cross and LaDel Bonham (Modoc National Wildlife Refuge), Richard Cain (Oroville Wildlife Area), T Cox, Michelle Workman and Kent Lambert (East Bay Municipal Utilities District), and Mike Connor (Sierra Foothills Research and Extension Center) for field assistance, and Mark Miller for advice on AIS. This work was supported by grants from the NSF (DEB 0516475, 0507916, 0817042), an NSF Doctoral Dissertation Improvement grant (to RCT; DEB-0710380), the California Department of Fish and Game, and the UC Davis Agricultural Experiment Station.

References

Avise JC, Arnold J, Ball RM *et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489–522.

Bernatchez L, Glémet H, Wilson CC *et al.* (1995) Introgression and fixation of arctic char (*Salvelinus alpinus*) mitochondrial genome in an allopatric population of brook trout (*Salvelinus fontinalis*). *Canadian Journal of Fisheries and Aquatic Sciences*, **52**, 179–185.

Bowersox JR (2005) Reassessment of extinction patterns of Pliocene molluscs from California and environmental forcing of extinction in the San Joaquin Basin. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **221**, 55–82.

Bruen TC, Philippe H, Bryant R (2006) A simple and robust statistical test for detecting the presence of recombination. *Genetics*, **172**, 2665–2681.

Bryant D, Moulton V (2004) NeighborNet: an agglomerative algorithm for the construction of planar phylogenetic networks. *Molecular Biology Evolution*, **21**, 255–265.

Buschbom J, Barker D (2006) Evolutionary history of vegetative reproduction in *Porpidia* s.l. (lichen-forming Ascomycota). *Systematic Biology*, **55**, 471–484.

Caccone A, Gentile G, Burns CE *et al.* (2004) Extreme difference in rate of mitochondrial and nuclear DNA evolution in a large ectotherm, Galápagos tortoises. *Molecular Phylogenetics and Evolution*, **31**, 794–798.

Calsbeek R, Thompson JN, Richardson JE (2003) Patterns of molecular evolution and diversification in a biodiversity hotspot: the California Floristic Province. *Molecular Ecology*, **12**, 1021–1029.

Chatzimanolis S, Caterino MS (2007) Toward a better understanding of the ''Transverse Range break'': lineage diversification in Southern California. *Evolution*, **61** 9, 2127–2141.

Dupré WR (1990) Quaternary geology of the Monterey Bay region, California. In: *Geology and Tectonics of the Central California Coastal Region, San Francisco to Monterey* (eds Garrison RE, Greene HG, Hicks KR *et al.*), pp. 185–192. Pacific Section of the American Association of Petroleum Geologists, Bakersfield, California.

Dupré WR, Morrison RB, Clifton HE *et al.* (1991) Quaternary geology of the Pacific margin. In: *Quaternary Nonglacial Geology; Conterminous US* (ed. Morrison RB), pp. 141–214. Geological Society of America, Boulder, Colorado.

Durham DL (1974) *Geology of the Southern Salinas Valley Area, California*. U. S. Geological Survey professional, paper 819 (111 ps). Washington, D.C., USA.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Feldman CR, Spicer GS (2006) Comparative phylogeography of woodland reptiles in California: repeated patterns of cladogenesis and population expansion. *Molecular Ecology*, **15**, 2201–2222.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.

Felsenstein J (2005) *PHYLIP (Phylogeny Inference Package) Version 3.6*. Distributed by the author. Department of Genome Sciences. University of Washington, Seattle, WA.

Fritz U, Ayaz D, Buschbom J *et al.* (2008) Go east: phylogeographies of *Mauremys caspica* and *M. rivulata* —discordance of morphology, mitochondrial and nuclear

genomic markers and rare hybridization. *Journal of Evolution Biology*, **21**, 527–540.

Good JM, Hird S, Reid N *et al.* (2008) Ancient hybridization and mitochondrial capture between two species of chipmunks. *Molecular Ecology*, **17**, 1313–1327.

Hall CA (2002) *Nearshore Marine Paleoclimatic Regions, Increasing Zoogeographic Provinciality, Mollsucan Extinctions, and Paleoshorelines, California: Late Oligocene (27 Ma) to Late Pliocene (2.5 Ma)*. Geological Society of America, Special Paper 357. Boulder, Colorado, USA.

Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology & Evolution*, **12**, 700–706.

Harshman J, Huddleston CJ, Bollback JP *et al.* (2003) True and false gharials: a nuclear gene phylogeny of Crocodylia. *Systematic Biology*, **52**, 386–402.

Holland DC (1991) *A Synopsis of the Ecology and Status of the Western Pond Turtle Clemmys Marmorata in 1991*. Report to National Ecological Research Center, United States Fish and Wildlife Service, San Simeon, California.

Hudson RR, Coyne JA (2002) Mathematical consequences of the genealogical species concept. *Evolution*, **56**, 1557–1565.

Huelsenbeck JP, Rannala RB (2004) Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology*, **53**, 904–913.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*, **17**, 754–755.

Huson DH (1998) SplitsTree: a program for analyzing and visualizing evolutionary data. *Bioinformatics*, **14**, 68–73.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Jacobs DK, Haney TA, Louie KD (2004) Genes, diversity, and geologic process on the Pacific coast. *Annual Review of Earth and Planetary Science*, **32**, 601–652.

Jennings MR, Hayes MP (1994) *Amphibian and Reptile Species of Special Concern in California*. pp. 1–255. California Department of Fish and Game, Rancho Cordova, California, USA.

Joly S, Bruneau A (2006) Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: an Example from Rosa in North America. *Systematic Biology*, **55**, 623–636.

Kass RE, Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.

Kocher TD, Thomas WK, Meyer A *et al.* (1989) Dynamics of mitochondrial evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences USA*, **86**, 6196–6200.

Kuchta SR, Tan AM (2006) Lineage diversification on an evolving landscape: phylogeography of the California newt, *Taricha torosa* (Caudata: Salamandridae). *Biological Journal of the Linnean Society*, **89**, 213–239.

Lapointe FJ, Rissler LJ (2005) Notes and comments—congruence, consensus, and the comparative phylogeography of codistributed species in California. *American Naturalist*, **166**, 290–299.

Leavitt DH, Bezy RL, Crandall KA *et al.* (2007) Multi-locus DNA sequence data reveal a history of deep cryptic vicariance and habitat-driven convergence in the desert night lizard *Xantusia vigilis* species complex (Squamata: Xantusiidae). *Molecular Ecology*, **16**, 4455–4481.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Linnen CR, Farrell BD (2007) Mitonuclear discordance is caused by rampant mitochondrial introgression in *Neodiprion* (Hymenoptera:Diprionidae) sawflies. *Evolution*, **61**, 1417–1438.

Lovich J, Meyer K (2002) The western pond turtle (*Clemmys marmorata*) in the Mojave River, California, USA: highly adapted survivor or tenuous relict? *Journal of the Zoological Society of London*, **256**, 537–545.

Maddison WP, Maddison DR (2003) *MacClade*. Sinauer Associates, Sunderland, MA.

Maldonado JE, Vila C, Wayne RK (2001) Tripartite genetic subdivisions in the ornate shrew (*Sorex ornatus*). *Molecular Ecology*, **10**, 127–147.

Matocq M (2002) Phylogeographical structure and regional history of the dusky-footed woodrat, *Neotoma fuscipes*. *Molecular Ecology*, **11**, 29–242.

McGaugh SE, Eckerman CM, Janzen FJ (2008) Molecular phylogeography of *Apalone spinifera* (Reptilia, Trionychidae). *Zoologica Scripta*, **37**, 289–304.

Melo-Ferreira J, Boursot P, Suchentrunk F *et al.* (2005) Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in Northern Iberia. *Molecular Ecology*, **14**, 2459–2464.

Miller MP (2005) Alleles in space (AIS): computer software for the joint analysis of interindividual spatial and genetic information. *Journal of Heredity*, **96**, 722–724.

Miller RE, Buckley TR, Manos PS (2002) An examination of the monophyly of morning glory taxa using Bayesian phylogenetic inference. *Systematic Biology*, **51**, 740–753.

Minin V, Abdo Z, Joyce P *et al.* (2003) Performance-based selection of likelihood models for phylogeny estimation. *Systematic Biology*, **52**, 674–683.

Moritz C (1994) Defining evolutionarily significant units for conservation. *Trends in Ecology and Evolution*, **9**, 373–375.

Near TJ, Meylan PA, Shaffer HB (2005) Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. *American Naturalist*, **165**, 137–146.

Nylander JAA, Ronquist F, Huelsenbeck JP *et al.* (2004) Bayesian phylogenetic analysis of combined data. *Systematic Biology*, **53**, 47–67.

Parham JFP, Pappenfuss TJ (2008) High genetic diversity among fossorial lizard populations (*Anniella pulchra*) in a rapidly developing landscape (Central California). *Conservation Genetics*, **10**, 169–176.

Peabody FE, Savage JM (1958) Evolution of a coast range corridor in California and its effect on the origin and dispersal of living amphibians and reptiles. In: *Zoogeography* (ed. Hubbs CL), pp. 159–186. American Association for the Advancement of Science, Washington, D.C.

Peters JL, Zhuravlev Y, Fefelov I *et al.* (2007) Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas* spp.). *Evolution*, **61**, 1992–2006.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution*, **16**, 37–45.

Primmer CR, Borge T, Lindell J *et al.* (2002) Single-nucleotide polymorphism characterization in species with limited

available sequence information: high nucleotide diversity revealed in the avian genome. *Molecular Ecology*, **11**, 603–612.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

de Queiroz K (1998) The general lineage concept of species. In: *Endless Forms: Species and Speciation* (eds Howard DJ, Berlocher SH), pp. 57–75. Oxford University Press, New York.

Rissler LJ, Hijmans RJ, Graham CH *et al.* (2006) Phylogeographic lineages and species comparisons in conservation analyses: a case study of California herpetofauna. *American Naturalist*, **167**, 655–666.

Robertson AV, Ramsden C, Niedzwiecki J *et al.* (2006) An unexpected recent ancestor of unisexual *Ambystoma*. *Molecular Ecology*, **15**, 3339–3351.

Rodríguez-Robles JA, Denardo DF, Staub RE (1999) Phylogeography of the California mountain kingsnake, *Lampropeltis zonata* (Colubridae). *Molecular Ecology*, **8**, 1923–1934.

Rodríguez-Robles JA, Stewart GR, Papenfuss TJ (2001) Mitochondrial DNA-based phylogeography of North American rubber boas, *Charina bottae* (Serpentes: Boidae). *Molecular Phylogenetics and Evolution*, **18**, 227–237.

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution*, **19**, 101–109.

Sanderson MJ (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*, **19**, 301–302.

Schierup MH, Hein J (2000) Consequences of recombination on traditional phylogenetic analysis. *Genetics*, **156**, 879–891.

Seeliger LM (1945) Variation in the Pacific mud turtle. *Copeia*, **1945**, 150–159.

Shaffer HB, Thomson RC (2007) Using single nucleotide polymorphisms (SNPs) to delimit species in recent radiations. *Systematic Biology*, **56**, 896–906.

Shaffer HB, Pauly GB, Oliver JC *et al.* (2004) The molecular phylogenetics of endangerment: cryptic variation and historical phylogeography of the California tiger salamander, *Ambystoma californiense*. *Molecular Ecology*, **13**, 3033–3049.

Smit HA, Robinson TJ, Van Vuuren BJ (2007) Coalescence methods reveal the impact of vicariance on the spatial genetic structure of *Elephantulus edwardii* (Afrotheria, Macroscelidea). *Molecular Ecology*, **16**, 2680–2692.

Spinks PQ, Shaffer HB (2005) Range-wide molecular analysis of the western pond turtle (*Emys marmorata*): cryptic variation, isolation by distance, and their conservation implications. *Molecular Ecology*, **14**, 2047–2064.

Spinks PQ, Shaffer HB (2007) Conservation phylogenetics of the Asian box turtles (Geoemydidae, *Cuora*): mitochondrial introgression, numts, and inferences from multiple loci. *Conservation Genetics*, **8**, 641–657.

Spinks PQ, Shaffer HB (2009) Conflicting mitochondrial and nuclear phylogenies for The Widely Disjunct *Emys* (Testudines: Emydidae) Species Complex, And What They Tell Us About Biogeography And Hybridization. *Systematic Biology*, **58**, 1–20.

Spinks PQ, Thomson RC, Lovely GA *et al.* (2009) Assessing what is needed to resolve a molecular phylogeny: simulations and empirical data from emydid turtles. *BMC Evolutionary Biology*, **9**, 56.

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction. *American Journal of Human Genetics*, **73**, 1162–1169.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Storer TI (1930) Notes on the range and life history of the Pacific fresh-water turtle, *Clemmys marmorata*. *University of California Publications in Zoology*, **32**, 429–441.

Swofford DL (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4*. Sinauer Associates, Sunderland, Massachusetts.

Tan A, Wake DB (1995) MtDNA phylogeography of the California newt, *Taricha torosa* (Caudata, Salamandridae). *Molecular Phylogenetics and Evolution*, **4**, 383–394.

Vähä JP, Primmer CR (2005) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, **15**, 63–72.

Velo-Antón G, García-París M, Rivera AC (2007) Patterns of nuclear and mitochondrial DNA variation in Iberian populations of *Emys orbicularis* (Emydidae): conservation implications. *Conservation Genetics*, **9**, 1263–1274.

Vignieri SN (2007) Cryptic behaviours, inverse genetic landscapes, and spatial avoidance of inbreeding in the Pacific jumping mouse. *Molecular Ecology*, **16**, 853–866.

Wake DB (1997) Incipient species formation in salamanders of the *Ensatina* complex. *Proceedings of the National Academy of Sciences USA*, **94**, 7761–7767.

Yanev KP (1980) Biogeography and distribution of three parapatric salamander species in coastal and borderland California. In: *The California Islands: Proceedings of a Multidisciplinary Symposium* (ed. Power DM), pp. 531–550. Santa Barbara Museum of Natural History, Santa Barbara, California.

Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, **12**, 563–584.

Phil Spinks works at the intersection of population genetics, phylogeography, and conservation genetics, with a strong emphasis on freshwater turtles. Bob Thomson has broad interests in phylogenetics, phylogenomics/informatics, and conservation biology of amphibians and reptiles. Brad Shaffer and his lab group have maintained a long-standing interest in the evolution, ecology, and conservation biology of reptiles (particularly turtles) and amphibians (particularly ambystomatid salamanders).

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Formatting matches that of Spinks & Shaffer 2005.

Fig. S2 One of 63 equally likely trees recovered from maximum likelihood (ML) analysis of up to 725 base pairs of mitochondrial ND4 data from 147 *Emys marmorata* plus five outgroup taxa.

Fig. S3 Models of molecular evolution for parameter estimation, partitioning strategy, and Bayes factor comparisons for the 152-taxon ND4 data set.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.