Journal of Biogeography (J. Biogeogr.) (2016) 43, 639–652

ORIGINAL ARTICLE

Ancient river systems and phylogeographical structure in the spring salamander, *Gyrinophilus porphyriticus*

Shawn R. Kuchta[1]*, Michael Haughey[1], Addison H. Wynn[2], Jeremy F. Jacobs[2] and Richard Highton[3]

[1]*Department of Biological Sciences, Ohio Center for Ecology and Evolutionary Studies, Ohio University, Athens, OH 45701, USA,* [2]*Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA,* [3]*Department of Biology, University of Maryland, College Park, MD 20742, USA*

*Correspondence: Shawn R. Kuchta, Department of Biological Sciences, Ohio Center for Ecology and Evolutionary Studies, Oho University, 107 Irvine, Athens, OH 45701, USA.
E-mail: kuchta@ohio.edu

ABSTRACT

Aim The river drainages of the Appalachian Mountains have experienced a dynamic history as glacial cycles, stream capture and other geological processes have led to the fragmentation and fusion of formerly isolated palaeodrainages. Some ancient rivers have gone extinct, including portions of the great Teays River. Here we investigate the contribution of contemporary and historical drainages to patterns of phylogeographical structure in the spring salamander complex, *Gyrinophilus porphyriticus*.

Location Eastern North America, USA.

Methods Sampling spanned the range of the *G. porphyriticus* complex, and included representative samples of the cave species of *Gyrinophilus* as well. Molecular sequence data included the mitochondrial DNA locus cytochrome *b* and the nuclear locus recombination-activating gene 1 (RAG-1). Time-calibrated phylogenies were inferred, and Bayes-LAGRANGE was used to reconstruct ancestral distributions. Contemporary and historical river influences on patterns of genetic diversity were tested using distance-based redundancy analysis (db-RDA).

Results The *G. porphyriticus* complex originated prior to the Pleistocene glacial cycles, and historical river systems explained more genetic variation than did contemporary drainages or geographical distance. Patterns of genetic variation suggest that extinct or remodelled palaeodrainages, including the Teays River, played an important role in structuring contemporary patterns of genetic variation.

Main conclusions The hydrogeological history of eastern North American drainage basins has been instrumental in structuring patterns of regional biodiversity in freshwater species. Here we show that hydrological remodelling has also left its genetic signature in the semi-aquatic spring salamander complex, *G. porphyriticus.* Historical drainages accounted for the largest fraction of phylogeographical structure, more so than did contemporary drainages or geographical distance, with spatial and temporal patterns of variation associated with the extinct Teays River.

Keywords
Appalachian River, cytochrome *b*, distance-based redundancy analysis, *Gyrinophilus*, Plethodontidae, phylogeography, RAG-1, stream capture, Teays River

INTRODUCTION

The geological and climatic history of the Appalachian Mountains of eastern North America included the waxing and waning of glaciers during the Pleistocene, secondary bouts of uplift that rejuvenated topographical relief, and the hydrological remodelling of ancient river systems (Thornbury, 1965; Prince *et al.*, 2010; Gallen *et al.*, 2013). Such changes impact populations by promoting or constraining dispersal, vicariance and extinction (Berendzen *et al.*, 2003;

Journal of Biogeography

Jones *et al.*, 2006; Soltis *et al.*, 2006), and can lead to the origin of distinct evolutionary lineages (Highton, 1995; Kuchta *et al.*, 2009a; Kozak & Wiens, 2010). Even within species and species complexes, phylogeographical structure often reflects the joint influence of contemporary and historical connections (Kozak *et al.*, 2006; Kuchta *et al.*, 2009b).

The Appalachian Mountains of eastern North America are one of the most biologically diverse temperate regions in the world, and are a biodiversity hotspot for salamanders (Highton, 1995; Duellman & Sweet, 1999). In part, this is due to the dynamic history of the region, which includes the extensive hydrological remodelling of river systems (Hocutt *et al.*, 1986; Mayden, 1988) (Fig. 1). For example, the former Teays River was a major drainage in eastern North America that flowed westward from the central Appalachians, but was rerouted during the Pleistocene when it was dammed by advancing glaciers 0.78–1.3 Ma (Melhorn & Kempton, 1991; Andrews, 2004) (Fig. 1). This impoundment created a massive lake, Lake Tight, which existed for several thousand years before breaching drainage divides and carving new drainage channels, thus initiating the formation of the present-day Ohio River (Hocutt *et al.*, 1978; Melhorn & Kempton, 1991).

Another mechanism by which formerly isolated river systems can fragment and fuse is stream capture. This occurs when the headwaters of a river system breach a drainage divide and assimilate the headwaters of another stream. Examples are evident all along the eastern continental divide (ECD) of eastern North America, as the higher gradient and faster flowing streams of Atlantic river systems have continually eroded westward into the Appalachian Mountains (Thornbury, 1965; Hocutt *et al.*, 1978; Prince *et al.*, 2010).

For organisms associated with headwater streams, hydrological rearrangements can profoundly influence population structure. Populations once connected become isolated when rivers break apart, and formerly isolated populations come into contact when new connections are formed. The consequences of river system alterations have been especially well studied in fishes (Hocutt *et al.*, 1986; Mayden, 1988; Near *et al.*, 2001; Berendzen *et al.*, 2003), but increasing evidence indicates that phylogeographical structure in aquatic and semi-aquatic plethodontid salamanders has been impacted as well (Jones *et al.*, 2006; Kozak *et al.*, 2006).

The aim of this study was to investigate the relative contribution of historical and contemporary river systems to phylogeographical diversity in the spring salamander complex, *Gyrinophilus porphyriticus* Green, 1827. To this end, we examined patterns of genetic diversity using DNA sequence data from mitochondrial DNA (cytochrome *b*; hereafter Cyt-*b*) and the nuclear DNA marker recombination-activating gene 1 (RAG-1). Specific hypotheses include: (1) diversification in *Gyrinophilus* is pre-Pleistocene in origin, (2) the Teays River was a long-standing barrier to dispersal that separated groups of populations, and (3) stream capture functioned to translocate populations across the ECD from west to east.

Figure 1 Historical and present-day river systems of eastern North America. (a) Major river systems prior to the Pleistocene glaciations. (b) Contemporary river systems. The dashed grey line demarcates the extent of the Nebraskan glaciation.

MATERIALS AND METHODS

Distribution and natural history

Salamanders in the genus *Gyrinophilus* are widely distributed throughout the Appalachian Mountains. Most of our

sampling is of the widespread spring salamander complex, *Gyrinophilus porphyriticus,* which ranges from northern Alabama to southern Québec (Fig. 2, Appendix S1 in Supporting Information). Following Brandon (1966), four subspecies have historically been recognized (*porphyriticus, duryi, danielsi, dunni*). Individuals of *G. porphyriticus* are found in and near headwater streams, springs, seepages, and caves. Larvae are tied to streams. Adults, by contrast, are semi-aquatic, but when terrestrial tend to remain within the riparian corridor (Greene *et al.*, 2008; Lowe *et al.*, 2008). In contrast with *G. porphyriticus*, the cave taxa *G. palleucus palleucus, G. p. necturoides* and *G. gulolineatus* have relatively restricted distributions and are endemic to cave systems (Fig. 2). Previous research has suggested that cave taxa are phylogenetically nested within *G. porphyriticus* (Niemiller *et al.*, 2008, 2009).

Population sampling and laboratory techniques

Blood and tissue samples were collected from throughout the range *G. porphyriticus*, as well as from representative samples of *G. p. palleucus, G. p. necturoides* and *G. gulolineatus* (Fig. 2, Appendix S1). Genomic DNA was extracted using Qiagen DNeasy Blood and Tissue Kits (Qiagen, Valencia, CA, USA). We amplified and sequenced the Cyt-*b* locus using the primers OUCYTBF (5′ AAACCAATGGCCCACACCCTACGC 3′) and OUCYTBR (5′ TCTGCCGTCC CCGTTATAGGAATAAT 3′). Polymerase chain reactions (PCR) were carried out in 25 µl reaction volumes using 20 ng of template DNA, 0.4 µM of each primer, 0.2 mM of each dNTP, 1.25 U of Green Taq (Genscript, Piscataway, NJ, USA) and 1X Green GoTaq buffer (Promega, Madison, WI, USA). PCR conditions were as follows: initial denaturation at 94 °C for 3 min; 30 cycles at 94 °C for 30 s, 55.5 °C for 30 s, and 72.0 °C for 40 s; and a final elongation at 72.0 °C for 7 min. PCR products were visualized on 1% agarose gels, and purified using QIAquick PCR purification kits (Qiagen, Valencia, CA, USA). We obtained Cyt-*b* data from 122 individuals from 72 populations (see Appendix S1). All sequences were trimmed to 902 bp so that they were of equal length, with limited missing data.

The RAG-1 was amplified and sequenced using the primers OURAG1F (5′ TCCCTTCACTTGCCCAAGCGCCA 3′) and OURAG1R (5′ TGCAGAGAAAGCCCTCCTTCCAGGCT 3′), which were modified from Niemiller *et al.* (2008). PCR conditions were as follows: initial denaturation at 94 °C for 4 min; 30 cycles at 94 °C for 30 s, 64.2 °C for 30 s, and 72.0 °C for 1 min; and a final elongation at 72.0 °C for 7 min. Sequence data were obtained from 107 individuals from 74 populations (see Appendix S1). We trimmed all sequences to 1161 bp so they were of equal length, with no missing data.

DNA was sequenced on an Applied Biosystems (Foster City, CA, USA) 3130xl Genetic Analyzer and a 3730 DNA Analyzer. Most samples were sequenced in the forward and reverse directions, including all individuals heterozygous for RAG-1. Sequences were aligned using MUSCLE in GENEIOUS 6.1 (Biomatters, Ltd., San Francisco, CA, USA). The phase of heterozygotes was estimated using PHASE 2.1.1 (Stephens *et al.*, 2001). We ran PHASE for 1000 iterations, with a thinning interval of two steps and a burn-in of 100 iterations. For heterozygotes not resolved by PHASE, PCR products were cloned using a TOPO-TA Cloning Kit (Invitrogen, Carlsbad, CA, USA). Four separate colonies were sequenced from each clone, and all heterozygotes were resolved. For RAG-1, we tested for intragenic recombination using the difference in sum-of-squares (DSS) test in TOPALI (Milne *et al.*, 2009), including a 10 bp increment, a window size of 100 and 500 parametric bootstraps.

We compared patterns of genetic differentiation among clades using polymorphism (P), the number of segregating sites (S), haplotype diversity (h), sequence diversity (κ) and nucleotide diversity (π) (Nei, 1987). Calculations were carried out in DNAsp 5.10.1 (Librado & Rozas, 2009).

Phylogenetic analyses and divergence time estimation

We inferred time-calibrated phylogenies in BEAST 2.1.3 (Bouckaert *et al.*, 2014) using Cyt-*b* and a concatenated data set including Cyt-*b* and RAG-1. The concatenated data set was trimmed to include only individuals that had both markers (n = 85; Appendix S1). As there are no spelerpine fossils from the pre-Pleistocene (Holman, 2006), we adopted two different approaches to time calibration: (1) we used the rate of substitution for Cyt-*b* as estimated by Mueller (2006), and (2) we used several outgroups and fossil data to constrain the age of extant plethodontids.

To estimate a phylogeny from the concatenated data using the substitution prior, we used a normal distribution for the clock rate with a rate of 0.62 substitutions per nucleotide site per 100 Myr (SD = 0.16) (Mueller, 2006). No outgroups were included, as recommended by Drummond & Bouckaert (2015). An optimal partitioning strategy and model of sequence evolution was estimated using PARTITIONFINDER 1.0.1 (Lanfear *et al.*, 2012). For nucleotide positions 1, 2, and 3, respectively, the best-fit model for Cyt-*b* was HKY+I+Γ, GTR+I and GTR+Γ, and for RAG-1, it was GTR+I+Γ, HKY and HKY+I. We employed a strict clock model and a constant population size coalescent tree prior. Analyses were run for 50 million generations, with samples saved every 5000 generations. Effective sample sizes (ESS) and the stationarity of likelihood values were examined in Tracer 1.6 (Rambaut *et al.*, 2014). No ESS values were < 200, and most were several thousand. Results were summarized using a maximum clade credibility (MCC) tree in TreeAnnotator 2.0.2 (Heled & Bouckaert, 2013). The first 25% of trees was discarded as burn-in, which was well beyond stationarity.

To estimate a time-calibrated phylogeny from the concatenated data using node constraints, we included 24 outgroup plethodontid species from 11 genera (Bonett *et al.*, 2014), as

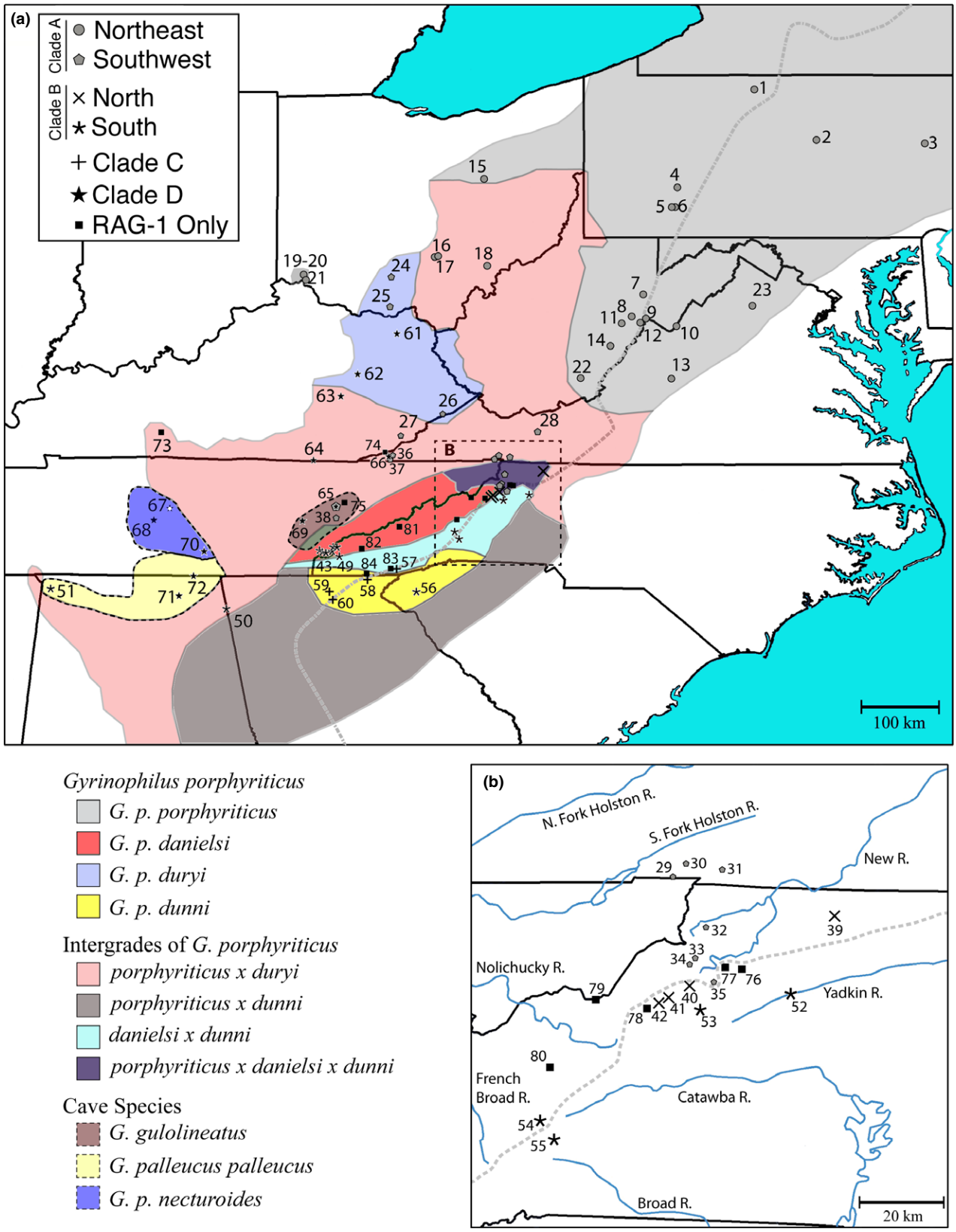

Figure 2 (a) Map of the range of *Gyrinophilus porphyriticus*, with turquoise representing Lake Erie (to the north) and the Atlantic Ocean (to the east). Sample localities are numbered, and match Fig. 3 and Appendix S1. Symbols identify the mtDNA clade of individuals in that population (Fig. 3). Subspecies and intergradation zones, as proposed by Brandon (1966), are differentially shaded. The dotted grey line identifies the Eastern Continental Divide. The range of *G. gulolineatus* overlaps *G. porphyriticus*. (b) Sampling at the junction of Virginia, North Carolina and Tennessee.

listed in Table S1. *Amphiuma means* was used to root the tree, as the Amphiumidae is consistently recovered as closely related to plethodontids (Zhang & Wake, 2009; Pyron & Wiens, 2011; Bonett *et al.*, 2014). Because different tree priors apply to ingroup and outgroup taxa, we adopted a two-step approach to inferring a time-calibrated phylogeny. First, we inferred a phylogeny including all the outgroup taxa and a sample of individuals (n = 13) from throughout *Gyrinophilus* (see Appendix S1). Following Bonett *et al.* (2014), we applied a normally distributed calibration prior for the crown group of plethodontids, with a mean age of 73 Myr and a standard deviation of 6 Myr. This provides a 95% prior distribution that ranges from 61.2–84.8 Myr, which encompasses the range of dates commonly inferred for the age of plethodontids (Mueller, 2006; Wiens *et al.*, 2006; Zhang & Wake, 2009; Bonett *et al.*, 2014). The best-fit model of sequence evolution was chosen using the Akaike information criterion corrected for small sample size (AICc) in jModeltest 2.1.5 (Darriba *et al.*, 2012): HKY+I+Γ for Cyt-*b*, and GTR+I+Γ for RAG-1. Topologies and branch lengths were linked, and we used an uncorrelated lognormal molecular clock and Yule tree prior. This analysis estimated that the most recent common ancestor (MRCA) of *Gyrinophilus* existed 5.2 Ma [95% highest posterior density (HPD): 3.9–6.7 Ma) (Table 1). Next, we used this estimate as a prior in a second analysis that included only *Gyrinophilus*, including a normal distribution and a standard deviation of 0.76. The models of evolution were the same as used in the analysis using the substitution rate prior. The BEAST analysis, Markov chain Monte Carlo (MCMC) sampling, and summarizing of results were as described above.

Because our geographical sampling was most complete for Cyt-*b* (n = 122), we also inferred a time-calibrated gene tree for this locus alone. The partitioning scheme and model of evolution were selected using PARTITIONFINDER (K80+I+Γ, GTR+I+Γ and GTR+Γ for positions 1, 2 and 3 respectively). Tree and clock models were linked, and no outgroups were used. The substitution rate prior was applied (Mueller, 2006). Markov chain Monte Carlo sampling and the summarizing of results were as described above.

For comparative purposes, maximum likelihood (ML) analyses were also conducted on the concatenated data and on Cyt-*b*. These are described in Appendix S2.

Gene trees for RAG-1 were poorly resolved due to a lack of variation (not shown). To test for a correspondence between mtDNA clades and patterns of variation in RAG-1, we used an analysis of molecular variation (AMOVA) as implemented in Arlequin 3.5 (Excoffier *et al.*, 2005). To infer relationships among RAG-1 alleles, we generated a haplotype network using statistical parsimony in TCS 1.21, including a 95% connection limit (Clement *et al.*, 2000).

Ancestral area reconstruction

The distributions of ancestral lineages of *Gyrinophilus* were reconstructed using the dispersal-extinction-cladogenesis (DEC) model of Bayes-LAGRANGE (Ree & Smith, 2008; Smith, 2009), as implemented in RASP 3.2 (Yu *et al.*, 2015). The analysis takes into account topology, branch lengths and phylogenetic uncertainty when making inferences. The MCC trees and posterior distributions obtained from our analyses of the concatenated and Cyt-*b* data were used, but trimmed to include only single individuals from monophyletic populations using the *drop.tip* function in the 'APE' package (Paradis *et al.*, 2004) in R (R Development Core Team, 2014). Analyses included 1000 trees randomly sampled from our Bayesian posterior distributions. The watersheds used were (historical drainages in parentheses): (A) upper Ohio (Teays); (B) lower Ohio (Old Ohio); (C) upper Tennessee (Appalachian); (D) lower Tennessee (drained into Old Mississippi); (E) Mid-Atlantic, between the West Branch Susquehanna and James Rivers; (F) South Atlantic, between the Yadkin and Savannah Rivers; (G) Northern mid-Atlantic (Erigan system), including the Allegheny and Monongahela Rivers. Ancestral ranges included up to three areas, and no constraints on dispersal were imposed.

Table 1 Estimated ages of clades within *Gyrinophilus,* dated in BEAST 2. Concatenated data included Cyt-*b* and RAG-1. Clade ages were estimated using rates of substitution for Cyt-*b* (Substitution), or by constraining the age of the family Plethodontidae (Node). Ages are reported as median values (95% HPD).

	Estimated age of clade (Ma ± 95% HPD)		
Clade	Cyt-*b* Substitution	Concatenated Substitution	Node
Gyrinophilus	8.4 (4.1–14.1)	9.3 (5.8–13.4)	4.9 (3.5–6.6)
Clade A (pop. 37 included)	2.9 (1.2–4.8)	3.2 (1.9–4.8)	3.2 (1.9–4.5)
Clade A (pop. 37 excluded)	1.7 (0.8–3.0)	2.1 (1.2–3.2)	1.8 (1.0–2.7)
North-east	1.2 (0.4–2.1)	1.4 (0.7–2.2)	0.8 (0.4–1.3)
South-west	1.3 (0.6–2.3)	1.6 (0.9–2.5)	0.9 (0.5–1.4)
Clade B	4.1 (1.8–7.0)	4.3 (2.6–6.4)	2.4 (1.4–3.5)
North	0.3 (< 0.1–0.5)	0.3 (< 0.1–0.0.7)	0.2 (< 0.1–0.4)
South	2.4 (1.1–4.0)	2.6 (1.5–3.8)	1.5 (0.9–2.2)
Clade C	5.2 (2.3–9.0)	5.3 (2.9–7.6)	2.9 (1.6–4.2)
Clade D	1.3 (0.4–2.3)	1.6 (0.8–2.7)	0.9 (0.4–1.5)

HPD, highest posterior density; RAG-1, recombination-activating gene 1.

Riverine influences on phylogenetic patterns of genetic diversity

Distance-based Redundancy Analysis (db-RDA) was used to test for a relationship between patristic distance, drainage association and geographical location (Legendre & Fortin, 2010). db-RDA is a multivariate analogue to regression that quantifies the proportion of variation in an explanatory data matrix that is explained by predictor variables. All db-RDA

calculations were carried out using the 'vegan' package in R. As R^2 in RDA is inflated, we used an adjusted R^2 calculation (Bocard *et al.*, 2011). Samples were assigned to contemporary river systems based on collection site (Fig. 2, Appendix S1). Historical drainage assignments were determined using historical drainage maps (Hocutt *et al.*, 1986; Mayden, 1988). Differentiation among haplotypes was estimated using patristic distances, which were calculated from our Bayesian Cyt-*b* phylogeny using the 'APE' package in R 3.1.2. Binary drainage matrices were created by assigning a value of zero to individuals inhabiting the same drainage, and one to individuals inhabiting different drainages. Geographical distances were calculated using the program Geographic Distance Matrix Generator 1.2.3 (http://biodiversityinformatics.amnh.org/open_source/gdmg).

RESULTS

Patterns of sequence diversity

Fifty haplotypes of Cyt-*b* from 122 individuals from 72 sampling localities were recovered for Cyt-*b*. In general, populations in the south harboured the highest levels of diversity (Table 2). For example, sequence diversity (κ) was 11.8 in Clade A (north), but in Clades B and C (south) it was 17.6 and 34.2 respectively. On the other hand, Clade D, which is located in the south and potentially includes haplotypes from cave species (see below), had relatively low levels of diversity (e.g. κ = 8.3).

For RAG-1, we recovered 48 unique alleles from 107 individuals from 74 sampling localities (Table 3). No evidence of allelic recombination was detected. RAG-1 displayed lower diversity than Cyt-*b*, with only 3% of 1161 sites polymorphic. We analysed patterns of diversity by assigning individuals to mitochondrial clades and subclades. An AMOVA attributed 19.3% of genetic variation to variation among mtDNA clades, 32% of variation to variation among populations within clades, and 48.3% of variation to variation within populations (all $P < 0.0001$; Table 4). Patterns of variation did not exhibit a clear latitudinal trend. For example, sequence diversity (κ) was 2.1 in Clade A and 2.3 in Clade B. The lowest diversity was documented in Clade C (κ = 1.3), while the highest diversity was found in Clade D (4.4). The high level of variation in Clade D contrasts sharply with mtDNA variation, where Clade D had the lowest levels of diversity (Tables 2 & 3).

Phylogenetic relationships

The topology and support levels were similar among the trees inferred using different methodologies. The time-calibrated phylogeny inferred using Cyt-*b* and substitution rate priors is presented in Fig. 3, while the ML analyses are summarized in Fig. S1, and the Bayesian inferences from the concatenated data (calibrated using either a substitution rate prior or a node constraint prior) are in Figs. S2–S3. All Bayesian analyses recovered four clades, which we labelled Clades A–D (Fig. 3). Relationships among the four clades are poorly supported (Fig 3, Figs. S2–S3). Clade A [posterior probability (pp) = 1] is distributed throughout the northern portion of the range of *G. porphyriticus*. It includes a 'north-east' subclade (pp = 0.96) in the north-eastern portion of the range, including geographically disjunct populations in Hamilton Co, Ohio (populations 19–21, Fig. 2). A 'south-west' subclade ranges from the headwaters of the New River into eastern Kentucky, northern Tennessee, and southern Ohio (pp = 0.96). Haplotypes from populations 38 and 65, which are south of the rest of Clade A, were sister to the remaining populations; however, one haplotype from population 65 was also recovered in Clade D, consistent with

Table 2 Diversity indices for Cyt-*b* (see Appendix S1).

Clade	No. sequences	No. pops	Sequence length	No. haplotypes	S*	*h*†	π‡	κ§
G. porphyriticus	122	72	776	50	155	0.96	0.031	23.9
Clade A¶	80	39	902	24	59	0.93	0.013	11.8
North-east	46	23	902	15	22	0.89	0.004	3.2
South-west	33	16	902	8	22	0.77	0.007	5.9
Clade B	24	18	892	17	61	0.97	0.020	17.6
North	5	4	902	2	1	0.60	0.001	0.6
South	19	14	892	15	45	0.98	0.013	11.6
Clade C	4	4	824	3	58	0.83	0.042	34.2
Clade D	13	12	854	7	40	0.85	0.010	8.3

*Number of segregating sites in a sample.
†Haplotype diversity (the probability that two randomly sampled sequences are different).
‡Nucleotide diversity (the probability that two randomly sampled homologous nucleotides are different, or, equivalently, the average number of nucleotide differences per site between two randomly chosen sequences).
§Sequence diversity (the average number of nucleotide differences between pairs of sequences).
¶Specimen 52157 (pop. 31) is excluded due to missing data. In addition, a unique haplotype from population 37 is sister to rest of Clade A (Fig. 3), and was not included in either subclade.

Table 3 Diversity indices for 1161 base pairs of RAG-1. The number of individuals in each grouping is equal to the number of alleles ÷ 2.

Clade	No. pops	No. alleles	No. unique alleles	S	*h*	π	κ
*G. porphyriticus**	74	214	48	38	0.92	0.0023	2.69
Clade A	33	110	21	19	0.86	0.0018	2.11
North-east	20	64	6	5	0.66	0.0009	1.08
South-west	13	44	17	15	0.93	0.0024	2.82
Clade B	17	46	15	14	0.89	0.0020	2.30
North	4	14	5	6	0.86	0.0022	2.60
South	13	32	12	11	0.86	0.0015	1.72
Clade C	6	14	5	7	0.51	0.0012	1.34
Clade D	10	22	11	16	0.91	0.0038	4.43

RAG, recombination-activating gene 1.

*Seven populations (76–82) were analysed as part of *G. porphyriticus*, but were not assigned to a clade because we lacked mtDNA for those populations and they are close to secondary contacts between mtDNA clades. In addition, population 65 was not assigned to a clade, because it is a hybrid population between Clades A and D.

hybridization between these two clades (see below). Clade A also includes a divergent haplotype from population 37 that is sister to the two subclades, while another individual from population 37 possessed a haplotype that was deeply nested within the south-west subclade. The divergent haplotype may belong to a narrowly distributed clade that we failed to otherwise sample (perhaps hybridizing with members of the south-west subclade), or it may be an example of retained ancestral polymorphism.

Populations outside of Clade A occupy the southern portion of the range of *Gyrinophilus*, and have relatively restricted distributions. Clade B (pp = 1) is found along the southern edge of the range. It is includes a northern subclade (pp = 1) located near the headwaters of the New River, and a southern subclade (pp = 1) located along the southern border of the range of *G. porphyriticus*. Clade C (pp = 0.98) occupies a small geographical area in north-eastern Georgia and western North Carolina, while Clade D constitutes a geographically cohesive group in eastern Kentucky, eastern Tennessee and northern Alabama (Fig. 2). This last clade includes *G. porphyriticus*, the cave-dwelling taxa *G. gulolineatus* and *G. palleucus,* and hybrids between *G. porphyriticus* and cave taxa. Individuals from population 65 were hybrids between *G. porphyriticus* × *G. gulolineatus* (A.H. Wynn & J.F. Jacobs, unpublished data), and included haplotypes belonging to both Clade D and Clade A. Finally, a single divergent haplotype from population 72 was not assigned to a clade because its placement is poorly resolved (*c.f.,* Fig. 3, Figs. S2–S5), but the specimen was identified as *G. p. palleucus*. By contrast, an individual from populations 71 was also identified as *G. p. palleucus*, but its haplotype was deeply nested within Clade D.

Table 4 Results of the hierarchical analysis of molecular variance using RAG-1, with populations sorted into mtDNA clades.

Source of variation	*df*	Sum of squares	Percentage of variation	*P*-value
Among clades	6	52.62	19.3	< 0.0001
Among populations within clades	59	111.43	32.07	< 0.0001
Within populations	126	82.92	48.26	< 0.0001
Total	191	246.96		

RAG-1, recombination-activating gene 1.

In general, the ML analyses were less resolved than the Bayesian analyses, with lower support values (Fig. S1). A phylogeny of Cyt-*b* inferred using PAUP* recovered Clades A–C with moderate to high bootstrap support (69–95%), but Clade D was not recovered. A phylogeny inferred using the concatenated data in RAxML was similar, except that Clade C, which was composed of deeply divergent haplotypes in all other analyses, was also unresolved. However, RAxML is designed for rapid ML inference from large data sets, and necessarily compromises thoroughness for speed (Stamatakis, 2014).

Finally, a haplotype network of RAG-1 alleles revealed limited structure (Fig. 4). Four alleles were abundant (1, 3, 4, 5). Allele 3 was largely associated with Clade A, and allele 4 was entirely restricted to the north-east subclade of Clade A. On the other hand, allele 1 was common in Clades A, B, and D, and allele 5 was recovered in all clades (Fig. 4).

DIVERGENCE TIME ESTIMATES

We obtained divergence time estimates using two methods: a substitution rate prior and a node age constraint (Table 1). All analyses indicated that the *G. porphyriticus* complex originated prior to the Pleistocene. The oldest estimate for the complex was the rate-based calibration of the concatenated data at 9.3 Ma (95% HPD = 5.8–13.4 Ma), while the youngest estimate was the node-based calibration of the concatenated data at 4.9 Ma (95% HPD = 3.5–6.6 Ma). Mean estimates for the age of Clade A ranged from 2.9–3.2 Ma (95% HPD: 1.2–4.8 Ma), while the MRCA of the north-east and south-west subclades is estimated to have existed 1.7–2.1 Ma (95% HPD: 0.8–3.2 Ma). Clades B and C were estimated to have originated 2.4–5.3 Ma (95% HPD: 1.4–9.0 Ma). Within Clade B, the southern subclade is relatively old, with mean estimates of 1.5–2.6 Ma (95% HPD = 0.9–4.0 Ma), while the southern subclade is relatively young, with mean estimates ≤ 0.3 Myr. Finally, the age estimates for Clade D ranged from 0.9–1.6 Myr (95% HPD: 0.4–2.7 Ma).

Ancestral area inference

All analyses produced similar results (Fig. 5, Figs. S4–S5, Table S2). The distribution of the ancestor of the *G. porphyriticus* complex was not clearly determined, but the most probable inferences included the upper Tennessee (former Appalachian) River system (Table S2). The ancestor of Clade

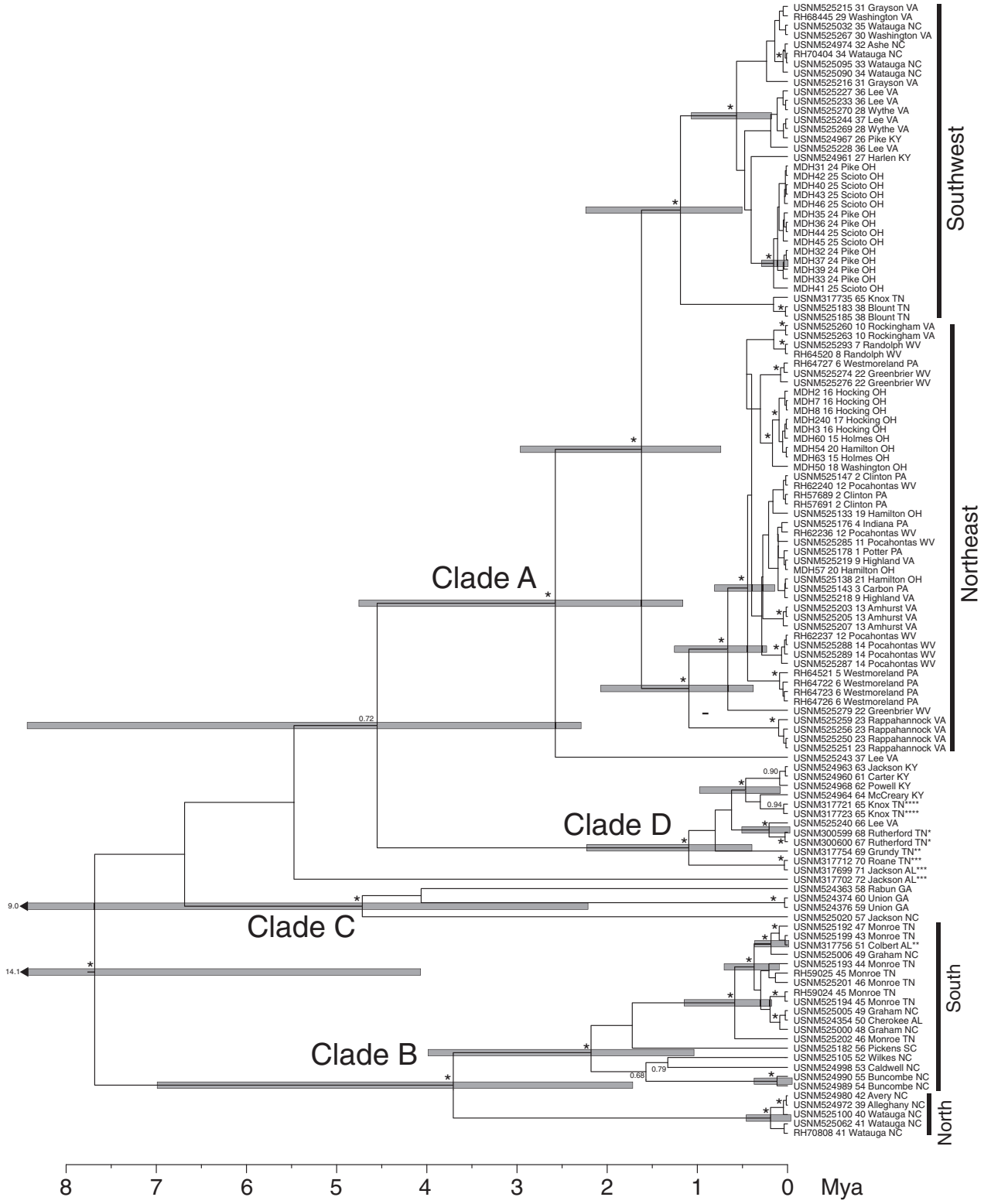

Figure 3 Bayesian maximum clade credibility (MCC) tree inferred using Cyt-*b*. Taxon labels include specimen identification number, the population number from Fig. 2, and county plus state information. All samples are *G. porphyriticus*, except as follows: **G. palleucus palleucus*, ***G. palleucus necturoides*, ****G. gulolineatus*, *****G. porphyriticus* × *G. gulolineatus*. Numbers adjacent to each node are Bayesian posterior probabilities. Negative branch lengths are not an error, but occur because node heights are averages from the posterior distribution. Negative branches here are limited to poorly supported nodes.

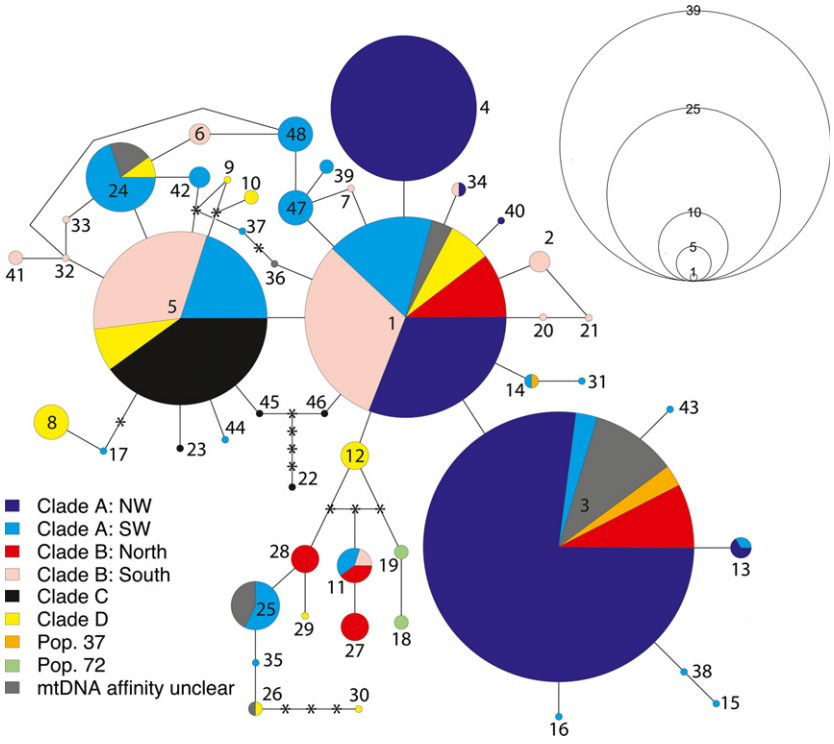

Figure 4 Haplotype network of RAG-1 alleles, inferred using statistical parsimony in TCS v 1.21 (Clement *et al.*, 2000). Circles represent haplotypes, with the radius proportional to sample size (as illustrated by the concentric circles in the upper right). Colours designate the mtDNA clades to which individuals belong. Lines connecting haplotypes represent single mutations; asterisks represent inferred mutations. The numbers are haplotype numbers, as identified in Appendix S1. RAG-1, recombination-activating gene 1.

A was inferred to have been broadly distributed, including the Teays, Appalachian and mid-Atlantic drainages. The ancestor of the north-east subclade of Clade A was inferred to have occupied the mid-Atlantic drainages, with the second most probable distribution also including the upper Ohio. The south-west subclade, by contrast, was inferred to have occupied either the upper Tennessee or the upper Tennessee + upper Ohio. The ancestor of Clade B was inferred to have occupied the upper Tennessee, or the upper Tennessee + south Atlantic drainages. All models supported the northern subclade of Clade B as having occupied the upper Ohio + south Atlantic drainages. This is an odd distribution at first blush, but stems from sampled populations currently occupying a restricted distribution near the headwaters of the New River (upper Ohio), with populations in Atlantic streams just over the ECD from the New River. The ancestor of the southern subclade of Clade B was estimated to have occupied the upper Tennessee + south Atlantic. All models supported the Tennessee as the ancestral region of Clade C. Finally, the ancestor of Clade D was inferred to be widespread in all models, with only Atlantic regions excluded.

Historical versus contemporary river systems

Distance-based redundancy analysis was used to partition variation among contemporary drainages, historical drainages, and geography (Table 5, Fig. 6). For Cyt-*b*, the full model explained 52.6% of the variation ($P < 0.0001$). Geography accounted for 30.7% of the variation ($P < 0.001$); however, only 4.0% of the variation was explained by geography after factoring out the contributions of drainages ($P < 0.001$). Similarly, contemporary drainages accounted for 35.3% of the variation ($P < 0.001$), but only 2.1% of the variation was explained by contemporary drainages alone ($P = 0.025$). Historical drainages accounted for the most variation: 47% of the total variation was accounted for by historical drainages, including 12.7% of the variation by historical drainages alone ($P < 0.001$).

For RAG-1, the full model explained 21% of the variation ($P < 0.001$) (Table 5, Fig. 6). Much of this was accounted for by historical drainages, which explained 11.2% of the variation after conditioning on contemporary drainages and geography ($P < 0.001$). By contrast, contemporary drainages explained 1.6% of the variation after conditioning on historical drainages and geography ($P = 0.027$). There was no relationship between geography and patristic distance after conditioning on drainages, and little variation (< 3%) was explained jointly by any set of factors.

DISCUSSION

In this study, we sampled throughout the range of the *G. porphyriticus* complex and documented large amounts of phylogeographical structure. Patterns of variation are likely a consequence of the old age of the complex, an extensive geographical range relative to dispersal ability, and hydrological remodelling of drainage systems since the Pleistocene. We found that Cyt-*b* was substantially more variable than our nuclear marker, RAG-1, which is consistent with other studies (Ballard & Whitlock, 2004). Nonetheless, the differences in variability and obvious genetic structure are striking (Tables 2 & 3). This could simply be a consequence of a

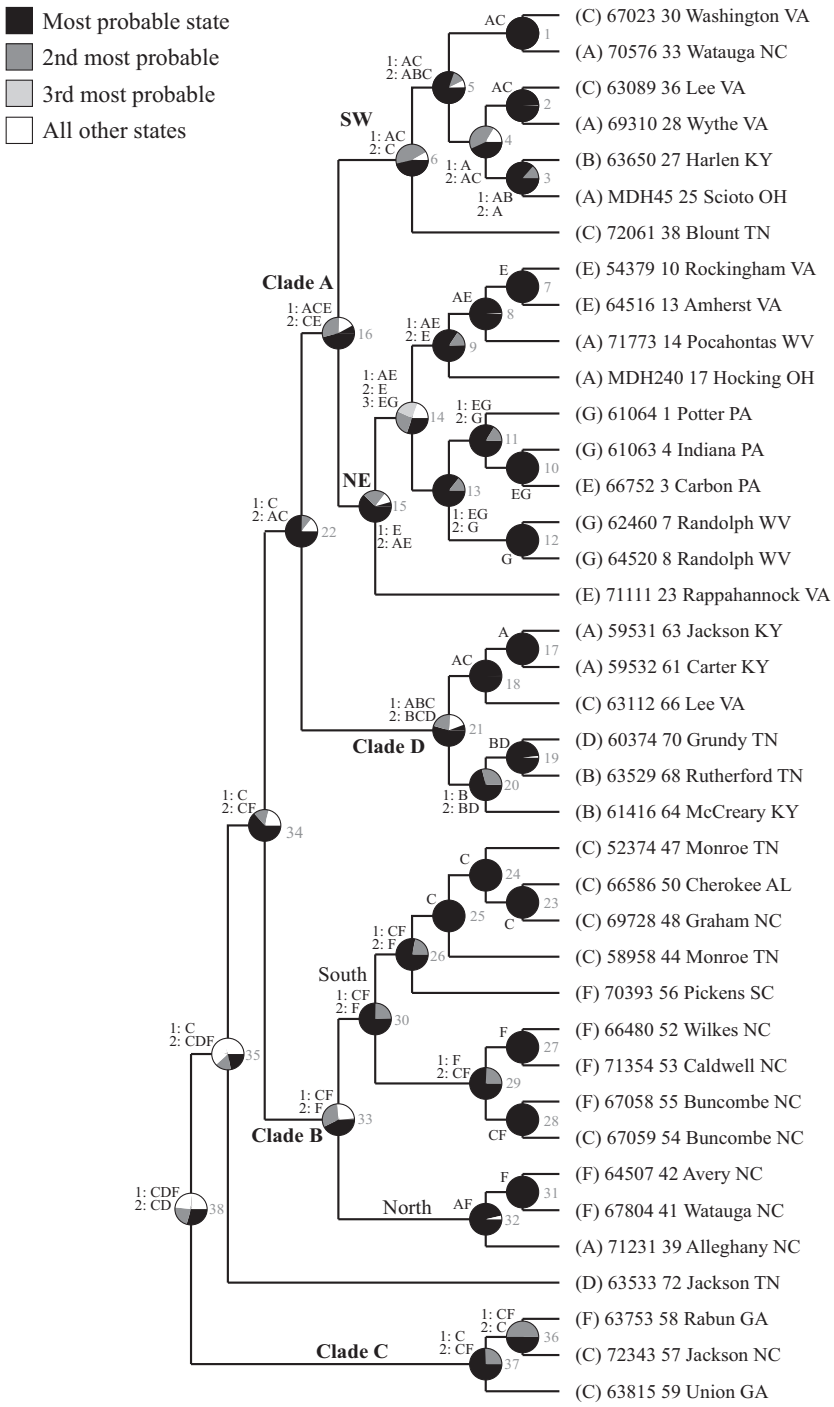

Figure 5 Bayes-LAGRANGE ancestral range analysis. The analysis was run on 1000 dated topologies randomly sampled from the posterior distribution of phylogenies produced in BEAST 2. The topology here is the maximum clade credibility (MCC) tree from a concatenated data set including Cyt-*b* and RAG-1, and calibrated using a node-based age prior. However, to improve readability branch lengths are not proportional to time (see Table 1, Fig. S3 for ages and support values). Pie charts present the most likely ancestral geographical ranges at each node, with values less than 20% not shown. Numbers next to nodes correspond with Table S2 in Appendix S3. A = upper Ohio River, of the former Teays River system; B = lower Ohio River, not of the former Teays; C = upper Tennessee River, of the former Appalachian River system; D = lower Tennessee, not in the former Appalachian River system; E = Rivers draining into the Mid-Atlantic, between the West Branch Susquehana and James Rivers; F = Rivers draining into the south Atlantic, between Yadkin and Savannah Rivers; G = Rivers that were potentially members of the former Laurentian system, including the Allegheny and Monongahela Rivers. Taxon labels include specimen identification number, the population number from Fig. 2, and county plus state information. RAG-1, recombination-activating gene 1.

lower mutation rate and higher effective population size in RAG-1. Alternatively, male-biased dispersal could lead to less geographical variation in nuclear markers if the nuclear genome is experiencing higher levels of gene flow (Ballard & Whitlock, 2004).

It is likely that the strong signal in mtDNA dominated our phylogeographical inferences. MtDNA has a lower effective population size than nuclear markers because it is haploid and maternally inherited. As a consequence, genetic drift is relatively strong, increasing the probability that mtDNA variation will coalesce sooner, and increasing the likelihood that mtDNA gene tree will be concordant with the species tree (Moore, 1995; Maddison, 1997). Nonetheless, lineage sorting is a stochastic process, and mtDNA will not always reflect the species tree. In addition, mtDNA readily introgresses across contact zones (Jackman & Wake, 1994; Kuchta & Tan, 2006; Kuchta, 2007), increasing the opportunity for introgressive hybridization to be a confounding factor (see also Niemiller *et al.*, 2008, 2009). In our study, individuals in population 65 were hybrids between *G. porphyriticus* and

Table 5 Proportion of genetic variance explained by contemporary drainage associations (contemporary), historical drainage associations (historical) and geographical distance. The 'Combined' factor includes all main effects as well as variation shared jointly by the main effects (Fig. 4).

Locus	Predictors	Variance	Adjusted R^2	*P*-value
Cyt-*b*	Contemporary drainages	1.76	3.6%	0.024
	Historical drainages	6.94	12.7%	< 0.001
	Geographical distance	2.21	4.0%	0.002
	Combined	28.175	52.6%	< 0.001
	Total variance	49.49		
RAG-1	Contemporary drainages	0.00006	1.6%	0.03
	Historical drainages	0.00026	11.2%	< 0.001
	Geographical distance	0.00002	0.4%	ns
	Combined	0.00052	21.0%	< 0.001
	Total variance	0.00208		

G. gulolineatus, population 37 contained two highly divergent haplotypes, and population 72 possessed a haplotype that was not closely related to surrounding populations. Apart from differences in variability, however, it is not clear that Cyt-*b* and RAG-1 were strongly discordant in our data, and an AMOVA showed that patterns of variation in RAG-1 are correlated with mtDNA clades.

Historical drainage connections and phylogeographical divergence

We inferred divergence times using two alternative approaches: a substitution rate prior (Mueller, 2006) and a node constraint (Bonett *et al.*, 2014). The rate prior resulted in older age estimates and wider confidence intervals (Table 1). For instance, for the concatenated data, the MRCA of the *G. porphyriticus* complex was inferred to be 9.3 Myr old (95% HPD = 5.8–13.4), whereas under the node constraint the complex was inferred to be 4.9 Myr old (95% HPD = 3.5–6.6). For comparison, Bonett *et al.* (2014), using a species tree approach, estimated the age of the *G. porphyriticus* complex at ~ 5–6 Myr old (based on their Fig. 2). These estimates all agree that the species complex originated prior to the onset of Pleistocene glaciations, when many rivers in eastern North America tracked different courses than they do today.

We found that phylogeographical divergence in the *G. porphyriticus* complex was associated with both contemporary and historical drainages. However, variance partitioning showed that historical drainages explained substantially more variation than did contemporary drainages, both for mtDNA (12.7% vs. 2.1%) and RAG-1 (11.2% vs. 1.6%) (Tables 4–5; Fig. 6). Geographical distance accounted for only 4.0% of the variation in Cyt-*b*, and its influence was not detected using RAG-1.

These results implicate historical drainage systems in the fashioning of phylogeographical diversity within the *G. porphyriticus* complex. A leading hypothesis is that the Teays River functioned as a barrier to dispersal, resulting in divergent phylogeographical groups to the north-east and south-west. This is because the Teays River was massive and would have been a barrier for headwater species. Today, the headwaters of the New River correspond with the headwaters of the former Teays (Fig. 2b), and are located in a relatively narrow strip between the ECD and the headwaters of the Tennessee River system. It is here that one would expect phylogeographical units to meet. Our findings are consistent with the hypothesis that the Teays functioned as a biogeographical barrier. The northern-most clade we recovered was Clade A (Figs. 2 & 3). Populations in the north-east subclade of Clade A are found north and east of the former Teays River, with the exception of populations in Hamilton Co, Ohio, which are disjunct from the rest of the range (Figs. 2 & 3). These disjunct populations were likely recently founded, as they are deeply nested within the north-east subclade. Ancestral range reconstruction using Bayes-LAGRANGE indicated that the ancestral range of the north-east subclade occupied either the mid-Atlantic drainages, or the mid-Atlantic + upper Ohio. We prefer the latter because *Gyrinophilus* currently has a limited distribution east of the ECD, and the occupation of these drainages is consistent with stream capture (see below). The south-west subclade of Clade A, on the other hand, is found exclusively west of the former Teays (Fig. 2). Bayes-LAGRANGE reconstructs the ancestor as occupying the upper Tennessee drainage, and perhaps also the upper Ohio. The south-west and north-east subclades likely contact one another north of the headwaters of the New River, between populations 22 and 28 (Fig. 2). The 95% HPD of our divergence time estimates for the split between the two subclades straddle the damming of the

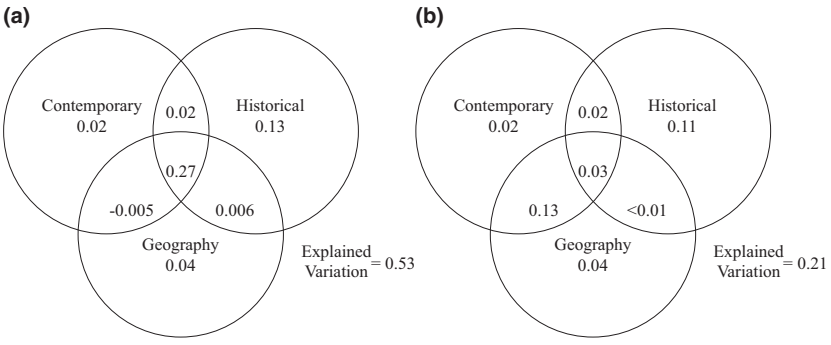

Figure 6 Venn diagrams showing the partitioning of genetic variation between contemporary drainages, historical drainages and geographical distance. (a) Cyt-*b*. (b) RAG-1. RAG-1, recombination-activating gene 1.

Teays, which occurred roughly 0.78–1.3 Ma (Andrews, 2004). This is the expected result if re-routing of the Teays played a role in the formation of phylogeographical diversity. Whether the damming of the Teays played a role in the origin of the subclades is unclear, however, and dispersal (as opposed to vicariance) is also consistent with this pattern. In either case, our findings are consistent with the hypothesis that the Teays River impacted population structure in *G. porphyriticus*, and is congruent with other studies that have identified the Teays as an important isolating barrier in aquatic taxa (Hocutt *et al.*, 1978, 1986; Berendzen *et al.*, 2003; Kozak *et al.*, 2006).

In addition to hydrological remodelling, we also predicted that historical stream capture played a role in the phylogeographical history of the *G. porphyriticus* complex. Stream capture is an important process across the ECD, where Atlantic-draining river systems have long eroded westward into the Appalachian Mountains (Hocutt & Wiley, 1986; Prince *et al.*, 2010). Stream capture was found to play a fundamental role in the phylogeographical history of the Northern two-lined salamander complex, *Eurycea bislineata*, as well as in the distributions of several fish species (Hocutt *et al.*, 1978; Kozak *et al.*, 2006). In *G. porphyriticus*, mitochondrial haplotypes from populations inhabiting mid-Atlantic drainages (populations 10, 12–13) are phylogenetically nested among haplotypes from populations that drain into the Ohio River (Figs. 2 & 3). To the south, Clade C, which is composed of divergent haplotypes (Fig. 3), is distributed along the border of the Tennessee, Savannah, and Chattahoochee/Apalachicola rivers. Stream capture may have played a role in this distribution, although long-distance dispersal is a viable alternative explanation. For example, Campbell *et al.* (2010) have shown that newly metamorphosed individuals of the salamander genus *Desmognathus* engage in overland movement, including movement between headwater streams.

Not all populations east of the ECD fit the stream capture hypothesis. For example, haplotypes from population 23 are sister to the remainder of the north-east subclade (Fig. 2), and we estimated this split occurred during the Pleistocene. Whether this is the consequence of an older stream capture event or some form of vicariance awaits further investigation. It is also conceivable the haplotypes descended from an ancestor that inhabited the Erigan River system, which was a part of the Laurentian River system and drained out the present-day St. Lawrence Valley (Fig. 1a); however, this is speculative and we did not find convincing evidence that the *G. porphyriticus* complex inhabited the Erigan River system. Finally, populations 52–53 and 55 in the upper Catawba, Broad, and Pee Dee River systems (mid-Atlantic drainages) formed a clade. Population 54 in the upper reaches of the Tennessee River system near the ECD is the only other member of this clade. Ancestral range reconstruction in BayesLAGRANGE predicted that this clade diversified on the eastern side of the ECD, then expanded westward across the ECD (Fig. 5). This pattern contradicts our stream capture hypothesis, although an east–west stream capture event during the Pleistocene is possible, as is long-distance dispersal across the ECD.

SUMMARY

River systems in eastern North America have experienced a dynamic history as glacial impoundments, stream capture, and other geological events have led to the fragmentation and fusion of formerly isolated palaeodrainages. Zoogeographical studies are increasingly finding that drainage history has had a profound influence on regional biodiversity, especially in obligate aquatic organisms. We studied the phylogeography of the semi-aquatic spring salamander complex, *Gyrinophilus porphyriticus* and recovered a complex genetic structure that is reflective of the influence of historical drainage connections. Historical drainage systems explained more variation in phylogeographical structure than did contemporary drainages or geographical distance, and it is likely that the prehistorical Teays River played a major role in promoting isolation and divergence.

ACKNOWLEDGEMENTS

Financial support was provided to S.R.K. by Ohio University, including the Program to Aid in Career Exploration (PACE) and the Honors Tutorial College (HTC). We thank Ashley Brown and Elizabeth Makley for their assistance in the DNA lab, and Paul Converse, Maggie Hantak, and Tom Radomski for their comments on drafts of the manuscript. We thank Ronald A. Brandon, John R. MacGregor, David M. Sever, R. Wayne Van Devender, and the many other people who provided specimens and helped with fieldwork.

REFERENCES

Andrews, W.M., Jr (2004). Geologic controls on Plio-Pleistocene drainage evolution of the Kentucky River in central Kentucky. PhD dissertation, University of Kentucky, Lexington.

Ballard, J.W.O. & Whitlock, M.C. (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Berendzen, P.B., Simons, A.M. & Wood, R.M. (2003) Phylogeography of the northern hogsucker, *Hypentelium nigricans* (Teleostei: Cypriniformes): genetic evidence for the existence of the ancient Teays River. *Journal of Biogeography*, **30**, 1139–1152.

Bocard, D., Gillet, F. & Legendre, P. (2011) *Numerical ecology with R*. Springer, New York.

Bonett, R.M., Steffen, M.A., Lambert, S.M., Wiens, J.J. & Chippindale, P.T. (2014) Evolution of paedomorphosis in plethodontid salamanders: ecological correlates and re-evolution of metamorphosis. *Evolution*, **68**, 466–482.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A. & Drummond,

A.J. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Brandon, R.A. (1966) Systematics of the salamander genus *Gyrinophilus*. *Illinois Biology Monographs*, **35**, 1–85.

Campbell, G.E.H., Nichols, J.D., Lowe, W.H. & Fagan, W.F. (2010) Use of multiple dispersal pathways facilitates amphibian persistence in stream networks. *Proceedings of the National Academy of Sciences USA*, **107**, 6936–6940.

Clement, M., Posada, D. & Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**(8), 772.

Drummond, A.J. & Bouckaert, R.R. (2015) *Bayesian evolutionary analysis with BEAST 2*. Cambridge University Press.

Duellman, W.E. & Sweet, S.S. (1999) Distribution patterns of amphibians in the Nearctic region of North America. *Patterns of distribution of amphibians* (ed. by W.E. Duellman), pp. 31–109. Johns Hopkins University Press, Baltimore.

Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Gallen, S.F., Wegmann, K.W. & Bohnenstiehl, D.R. (2013) Miocene rejuvenation of topographic relief in the southern Appalachians. *GSA Today*, **23**(2), 4–10.

Greene, B.T., Lowe, W.H. & Likens, G.E. (2008) Forest succession and prey availability influence the strength and scale of terrestrial-aquatic linkages in a headwater salamander system. *Freshwater Biology*, **53**, 2234–2243.

Heled, J. & Bouckaert, R.R. (2013) Looking for trees in the forest: summary tree from posterior samples. *BMC Evolutionary Biology*, **13**, 221.

Highton, R. (1995) Speciation in eastern North American salamanders of the genus *Plethodon*. *Annual Review of Ecology and Systematics*, **26**, 576–600.

Hocutt, C.H. & Wiley, E.O. (1986) *The zoogeography of North American freshwater fishes*. John Wiley and Sons, New York.

Hocutt, C.H., Denoncourt, R.F. & Stauffer, J.R., Jr (1978) Fishes of the Greenbrier river, West Virginia with drainage history of the Central Appalachians. *Journal of Biogeography*, **5**, 59–80.

Hocutt, C.H., Jenkins, R.E. & Stauffer, J.R., Jr (1986) Zoogeography of the fishes of the Central Appalachians and Central Atlantic Coastal Plain. *The zoogeography of North American freshwater fishes* (ed. by C.H. Hocutt and E.O. Wiley), pp. 161–211. John Wiley and Sons, New York.

Holman, J.A. (2006) *Fossil salamanders of North America*. Indiana University Press, Bloomington, Indiana.

Jackman, T. & Wake, D.B. (1994) Evolutionary and historical analysis of protein variation in the blotched forms of salamanders of the *Ensatina* complex (Amphibia, Plethodontidae). *Evolution*, **48**, 876–897.

Jones, M.T., Voss, S.R., Ptacek, M.B., Weisrock, D.W. & Tonkyn, D.W. (2006) River drainages and phylogeography: an evolutionary significant lineage of shovel-nosed salamander (*Desmognathus marmoratus*) in the southern Appalachians. *Molecular Phylogenetics and Evolution*, **38**, 280–287.

Kozak, K.H. & Wiens, J.J. (2010) Accelerated rates of climatic niche evolution underlie rapid species diversification. *Ecology Letters*, **13**, 1378–1389.

Kozak, K.H., Blaine, R.A. & Larson, A. (2006) Gene lineages and eastern North American palaeodrainage basins: phlyogeography and speciation in salamanders of the *Eurycea bislineata* species complex. *Molecular Ecology*, **15**, 191–207.

Kuchta, S. (2007) Contact zones and species limits: hybridization between lineages of the California Newt, *Taricha torosa*, in the southern Sierra Nevada. *Herpetologica*, **63**, 332–350.

Kuchta, S. & Tan, A.-M. (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.

Kuchta, S.R., Parks, D.S., Mueller, R.L. & Wake, D.B. (2009a) Closing the ring: historical biogeography of the salamander ring species *Ensatina eschscholtzii*. *Journal of Biogeography*, **36**, 982–995.

Kuchta, S.R., Parks, D.S. & Wake, D.B. (2009b) Pronounced phylogeographic structure on a small spatial scale: geomorphological evolution and lineage history in the salamander ring species *Ensatina eschscholtzii* in central coastal California. *Molecular Phylogenetics and Evolution*, **50**, 240–255.

Lanfear, R., Calcott, B., Ho, S.Y. & Guindon, S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Legendre, P. & Fortin, M.-J. (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**, 831–844.

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

Lowe, W.H., McPeek, M.A., Likens, G.E. & Cosentino, B.J. (2008) Linking movement behaviour to dispersal and divergence in plethodontid salamanders. *Molecular Ecology*, **17**, 4459–4469.

Maddison, W.P. (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.

Mayden, R.L. (1988) Vicariance biogeography, parsimony, and evolution in North America freshwater fishes. *Systematic Zoology*, **37**, 329–355.

Melhorn, W.N. & Kempton, J.P., eds. (1991) Geology and hydrogeology of the Teays-Mahomet bedrock valley system. *Geological Society of America Special Paper*, **258**, 128.

Milne, I., Lindner, D., Bayer, M., Husmeier, D., McGuire, G., Marshall, D.F. & Wright, F. (2009) TOPALi v2: a rich

graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics*, **25**, 126–127.

Moore, W.S. (1995) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, **49**, 718–726.

Mueller, R.L. (2006) Evolutionary rates, divergence dates, and the performance of mitochondrial genes in Bayesian phylogenetic analysis. *Systematic Biology*, **55**, 289–300.

Near, T.J., Page, L.M. & Mayden, R.L. (2001) Intraspecific phylogeography of *Percina evides* (Percidae: Etheostomatinae): an additional test of the Central Highlands pre-Pleistocene vicariance hypothesis. *Molecular Ecology*, **10**, 2235–2240.

Nei, M. (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.

Niemiller, M.L., Fitzpatrick, B.M. & Miller, B.T. (2008) Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: *Gyrinophilus*) inferred from gene genealogies. *Molecular Ecology*, **17**, 2258–2275.

Niemiller, M.L., Miller, B.T. & Fitzpatrick, B.M. (2009) Systematics and evolutionary history of subterranean *Gyrinophilus* salamanders. *Proceedings of the 15th International Congress of Speleology*, **1**, 242–248.

Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Prince, P.S., Spotila, J.A. & Henika, W.S. (2010) New physical evidence of the role of stream capture in active retreat of the Blue Ridge escarpment, southern Appalachians. *Geomorphology*, **123**, 305–319.

Pyron, R.A. & Wiens, J.J. (2011) A large-scale phylogeny of Amphibia including over 2,800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution*, **61**, 543–583.

R Development Core Team (2014) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. Available at: http://www.R-project.org/.

Rambaut, A., Suchard, M.A., Xie, D. & Drummond, A.J. (2014) Tracer v1.6. Available at: http://beast.bio.ed.ac.uk/Tracer.

Ree, R.H. & Smith, S.A. (2008) Maximum-likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, **57**, 4–14.

Smith, S.A. (2009) Taking into account phylogenetic and divergence-time uncertainty in a parametric biogeographical analysis of the Northern Hemisphere plant clade Caprifolieae. *Journal of Biogeography*, **36**, 2324–2337.

Soltis, D.E., Morris, A.B., McLachlan, J.S., Manos, P.S. & Soltis, P.S. (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261–4293.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

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

Thornbury, W.D. (1965) *Regional geomorphology of the United States*. John Wiley and Sons, New York.

Wiens, J.J., Engstrom, T.N. & Chippindale, P.T. (2006) Rapid diversification, incomplete isolation, and the "speciation clock" in North American salamanders (genus *Plethodon*): testing the hybrid swarm hypothesis of rapid radiation. *Evolution*, **60**, 2585–2603.

Yu, Y., Harris, A.J., Blair, C. & He, X.J. (2015) RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Molecular Phylogenetics and Evolution*, **87**, 46–49.

Zhang, P. & Wake, D.B. (2009) Higher-level salamander relationships and divergence dates inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution*, **53**, 492–508.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1: Detailed sampling information.
Appendix S2: Additional phylogenies inferred using RAxML, PAUP* and BEAST 2.
Appendix S3: Ancestral area reconstructions using BayesLagrange in RASP.

DATA ACCESSIBILITY

All DNA sequence data were deposited in GenBank; see Appendix S1 for accession numbers, including detailed locality information.

BIOSKETCH

Shawn R. Kuchta is an assistant professor at Ohio University. He is interested in the processes that generate and maintain patterns of diversity within and among closely related species, with an emphasis on the evolution of geographical variation, species formation, natural selection in the wild, the evolution of prey to predators, and everything about the biology of salamanders.

Author contributions: S.R.K. and M.D.H. conceived the study; A.H.W., J.F.J. and R.H. collected and curated blood samples from across Appalachia over the course of years; M.D.H. collected the DNA sequence data; S.R.K. led the writing and analyses; A.H.W., J.F.J. and R.H. provided expertise.

Editor: Robert Bryson Jr.