Molecular Phylogenetics and Evolution 106 (2017) 241–253

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenomics and species delimitation in the knob-scaled lizards of the genus *Xenosaurus* (Squamata: Xenosauridae) using ddRADseq data reveal a substantial underestimation of diversity

Adrián Nieto-Montes de Oca [a,*], Anthony J. Barley [b], Rubi N. Meza-Lázaro [a], Uri O. García-Vázquez [c], Joan G. Zamora-Abrego [d], Robert C. Thomson [b], Adam D. Leaché [e]

[a] Laboratorio de Herpetología, Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad Universitaria, México, D.F., C.P. 04510, Mexico
[b] Department of Biology, University of Hawai'i at Mānoa, Honolulu, HI 96822, USA
[c] Facultad de Estudios Superiores Zaragoza, Universidad Nacional Autónoma de México, Batalla 5 de Mayo s/n, Ejército de Oriente, México, D.F., C.P. 09230, Mexico
[d] Departamento de Ciencias Forestales, Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Medellín, Antioquía 050034, Colombia
[e] Department of Biology & Burke Museum of Natural History and Culture, University of Washington, Seattle, WA 98195, USA

ARTICLE INFO

Article history:
Received 28 January 2016
Revised 3 August 2016
Accepted 1 September 2016
Available online 5 October 2016

Keywords:
ddRADseq
Middle America
Phylogenomics
Species delimitation
Xenosaurus

ABSTRACT

Middle American knob-scaled lizards of the genus *Xenosaurus* are a unique radiation of viviparous species that are generally characterized by a flattened body shape and a crevice-dwelling ecology. Only eight species of *Xenosaurus*, one of them with five subspecies (*X. grandis*), have been formally described. However, species limits within *Xenosaurus* have never been examined using molecular data, and no complete phylogeny of the genus has been published. Here, we used ddRADseq data from all of the described and potentially undescribed taxa of *Xenosaurus* to investigate species limits, and to obtain a phylogenetic hypothesis for the genus. We analyzed the data using a variety of phylogenetic models, and were able to reconstruct a well-resolved and generally well-supported phylogeny for this group. We found *Xenosaurus* to be composed of four major, allopatric clades concordant with geography. The first and second clades that branch off the tree are distributed on the Atlantic slopes of the Sierra Madre Oriental and are composed of *X. mendozai*, *X. platyceps*, and *X. newmanorum*, and *X. tzacualtipantecus* and an undescribed species from Puebla, respectively. The third clade is distributed from the Atlantic slopes of the Mexican Transvolcanic Belt in west-central Veracruz south to the Pacific slopes of the Sierra Madre del Sur in Guerrero and Oaxaca, and is composed of *X. g. grandis*, *X. rectocollaris*, *X. phalaroanthereon*, *X. g. agrenon*, *X. penai*, and four undescribed species from Oaxaca. The last clade is composed of the four taxa that are geographically closest to the Isthmus of Tehuantepec (*X. g. arboreus*, *X. g. rackhami*, *X. g. sanmartinensis*, and an undescribed species from Oaxaca). We also utilized a variety of molecular species delimitation approaches, including analyses with GMYC, PTP, BPP, and BFD*, which suggested that species diversity in *Xenosaurus* is at least 30% higher than currently estimated.

1. Introduction

1.1. Taxonomy of *Xenosaurus*

Middle American knob-scaled lizards of the genus *Xenosaurus* (Fig. 1) occur from southwestern Tamaulipas and eastern Guerrero on the Atlantic and Pacific versants of Mexico, respectively, south and east to Alta Verapaz, Guatemala (Ballinger et al., 2000; King and Thompson, 1968; Nieto-Montes de Oca et al., 2001, 2013). They can be found between 300 m and 2360 m of elevation in a wide variety of habitats ranging from xerophytic tropical scrub to cloud forest to tropical rain forest (Ballinger et al., 2000; King and Thompson, 1968; Smith and Iverson, 1993; Zamora-Abrego et al., 2007), where they occupy moderately diverse places, including crevices and holes in limestone, spaces under volcanic boulders, crevices in volcanic rocks, karst limestone, limestone terrain, and hollow logs in dry areas where trees are sparse (King and Thompson, 1968).

In their monograph of the genus, King and Thompson (1968) recognized only three species of *Xenosaurus* (Fig. 2): *X. newmanorum*, from southeastern San Luis Potosí; *X. platyceps*, from southwestern Tamaulipas; and *X. grandis*, with five subspecies occurring collectively from central Veracruz and southern Oaxaca

* Corresponding author.
E-mail address: anietomontesdeoca@me.com (A. Nieto-Montes de Oca).

http://dx.doi.org/10.1016/j.ympev.2016.09.001
1055-7903/

Fig. 1. Representatives of *Xenosaurus*. From left to right and from top to bottom (author of photograph in parentheses): *X. platyceps* (E. García), *X. mendozai* (L. Canseco), *X.* sp Vista Hermosa (E. Centenero), *X. g. grandis* (U. Garcia), *X. rectocollaris* (L. Canseco), *X.* sp Pápalos (E. Centenero), *X.* sp Tejocote (U. García), *X. phalaroantereon* (L. Canseco), and *X. arboreus* (L. Canseco).

on the Atlantic and Pacific versants of Mexico, respectively, south and east to Alta Verapaz in Guatemala: *X. grandis agrenon*, from southern Oaxaca; *X. g. arboreus*, from extreme southeastern Oaxaca; *X. g. grandis*, distributed from west-central Veracruz south-southeast to north-northwestern Oaxaca; *X. g. rackhami*, distributed from eastern Oaxaca east to Alta Verapaz in Guatemala; and *X. g. sanmartinensis*, from southern Veracruz. In addition, King and Thompson (1968) reported intergrades between *X. g. grandis* and *X. g. rackhami* from north-central Oaxaca (campamento Vista Hermosa) and east-central Oaxaca (San Lucas Camotlán).

Since the work of King and Thompson (1968), five species of *Xenosaurus* have been described: *X. rectocollaris*, from southeastern Puebla and adjacent Oaxaca (Smith and Iverson, 1993), *X. penai*, from eastern Guerrero (Pérez-Ramos et al., 2000), *X. phalaroanthereon*, from central and south-central Oaxaca (Nieto-Montes de Oca et al., 2001), *X. tzacualtipantecus*, from east-central Hidalgo and adjacent Veracruz (Woolrich-Piña and Smith, 2012), and *X. mendozai*, from northeastern Querétaro and adjacent north-central Hidalgo (Nieto-Montes de Oca et al., 2013).

Also, Canseco-Márquez (2005) and Zamora-Abrego (2009), in an unpublished M. S. thesis and an unpublished Ph. D. dissertation, respectively, performed phylogenetic analyses of *Xenosaurus* based on external morphology and mtDNA sequences that provided evidence suggesting the distinctness of *X. g. agrenon*, *X. g. arboreus*, *X. g. grandis*, and *X. g. rackhami*, and Zamora-Abrego (2009) also found several significantly supported, exclusive clades concordant with geography within *X. g. grandis*, which suggested the possible existence of cryptic species within this taxon. Other workers have treated some of the subspecies of *X. grandis* as distinct species (e.g., Bhullar, 2011; Nieto-Montes de Oca et al., 2013; Woolrich-Piña and Smith, 2012), but have not provided evidence to support them. To date, no molecular study of the status of *X. grandis* subspecies has been published.

1.2. Additional diversity within Xenosaurus

Canseco-Márquez (2005), and especially Zamora-Abrego (2009), provided evidence suggesting that some previously known and recently discovered populations of *Xenosaurus* may represent distinct, undescribed species. These populations included those from Campamento Vista Hermosa and San Lucas Camotlán, previously regarded as intergrades between *X. g. grandis* and *X. g. rackhami* by King and Thompson (1968); the population from Tejocote, previously assigned to *X. g. agrenon* also by King and Thompson (1968), and populations from Hidalgo (La Mojonera region), Puebla (Huehuetla region), Oaxaca (Monteverde and Pápalos regions), and Querétaro, most of them discovered in the last two decades during field work by herpetologists at the Museo de Zoología Alfonso L. Herrera, Universidad Nacional Autónoma de México. The populations from Hidalgo and Querétaro were later described as *X. tzacualtipantecus* (Woolrich-Piña and Smith, 2012) and *X. mendozai* (Nieto-Montes de Oca et al., 2013), respectively. However, the status of the remaining putative undescribed species (hereafter referred to as *X.* sp Camotlán, *X.* sp Huehuetla, *X.* sp Monteverde, *X.* sp Pápalos, *X.* sp Tejocote, and *X.* sp Vista Hermosa) remains uncertain, and no study of their systematics, molecular or otherwise, has been published to determine their validity.

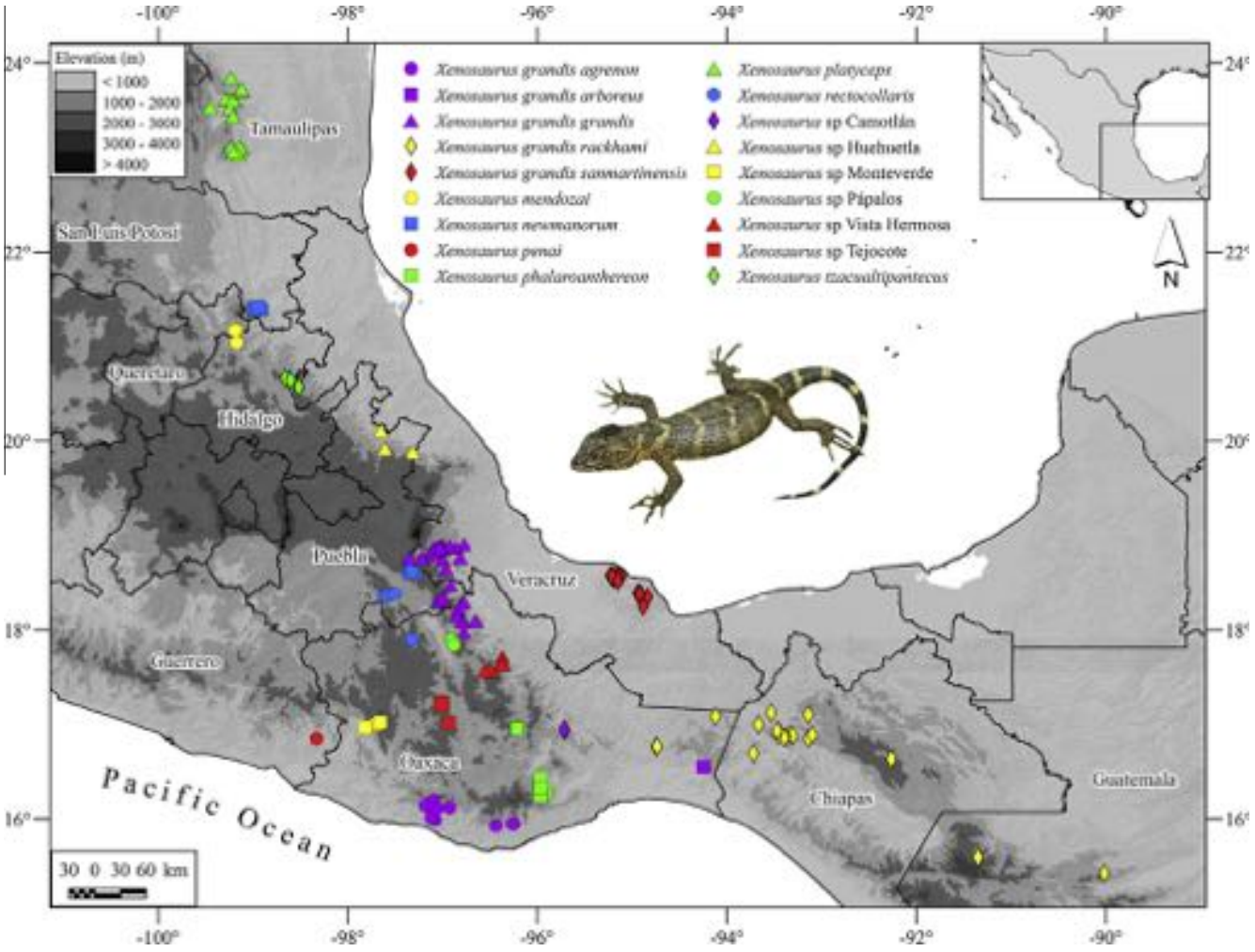

Fig. 2. Geographic distribution of the described species and subspecies and putative undescribed species of *Xenosaurus*.

Similarly, the only published phylogenetic study of *Xenosaurus* was provided by Bhullar (2011). This study included several related fossil taxa and was based on osteological characters. However, although it recovered a fully resolved and well-supported tree, it only included six of the extant species of *Xenosaurus*.

Herein, we use a genome-scale dataset to elucidate the phylogenetic relationships of this unique evolutionary radiation of crevice-dwelling lizards and to address species delimitation problems in the genus (i.e., the status of the subspecies of *X. grandis* and the putative undescribed species, and the possible existence of cryptic species within some taxa). By employing a diversity of available methodologies for phylogenetic inference and molecular species delimitation, we are able to assess consistency across analytical methods, as well as identify several instances of incongruence, which require further study.

2. Materials and methods

2.1. Taxon sampling

We sampled 2–3 representatives of each described species, subspecies, and each putative undescribed species in the genus in our study with two exceptions: to investigate the possible existence of cryptic species within *X. g. grandis*, we included 12 individuals from throughout its geographic distribution, whereas for *X.* sp Tejocote there was only one sample available to us. The list of vouchers used in this study is provided in Table 1. Representatives of *Xenosaurus* have been included in a number of molecular and morphological, higher-level phylogenetic studies of squamates. Most of these studies (e.g., Townsend et al., 2004; Wiens et al., 2010, 2012; Reeder et al., 2015) have placed Helodermatidae as the sister to Xenosauridae (Anniellidae + Anguidae), typically with strong support. However, in another recent study (Pyron et al., 2013), Xenosauridae was strongly supported as the sister taxon to the Anguidae + Helodermatidae clade. We included one representative each of the genera *Abronia* and *Heloderma* in our study as outgroups. Unfortunately, too few loci were shared between the ingroup and the outgroups to reliably root the tree. We rooted our trees on the basis of an ultrametric tree obtained using BEAST (see below).

2.2. ddRADseq data collection

We collected ddRADseq data following the protocol described by Peterson et al. (2012). High-molecular weight DNA was extracted from tissue samples using standard protocols, examined for quality on agarose gels, and quantified with a Qubit version 2.0 fluorometer (Thermo Fisher Scientific). We double-digested 500–1000 ng of genomic DNA for each sample with 20 units each of a rare cutter SbfI (restriction site 5′-CCTGCAGG-3′) and a common cutter MspI (restriction site 5′-CCGG-3′) in a single reaction with the manufacturer recommended buffer (New England Biolabs) for

Table 1
Vouchers used in this study. Voucher IDs are catalogue numbers in the Museo de Zoología *Alfonso L. Herrera* (MZFC-) or field numbers of specimens catalogued at the herpetological collection of the University of Texas at Arlington (JAC-) or to be catalogued in the MZFC.

ID	Taxon	Locality	Latitude	Longitude
ANMO-961	*X. grandis agrenon 1*	Mexico: Oaxaca: San Gabriel Mixtepec	16.1127	−97.0661
UOGV-1784	*X. g. agrenon 2*	Mexico: Oaxaca: San Juan Lachao-Temazcaltepec road	16.1408	−97.1778
MZFC-12476	*X. g. agrenon 3*	Mexico: Oaxaca: Cafetal Alemania	15.9494	−96.2608
UOGV-931	*X. g. arboreus 1*	Mexico: Oaxaca: Mountains NW of Santo Domingo Zanatepec	16.5479	−94.2464
UOGV-932	*X. g. arboreus 2*	Mexico: Oaxaca: Mountains NW of Santo Domingo Zanatepec	16.5479	−94.2464
UOGV-2709	*X. g. grandis 1*	Mexico: Veracruz: Zongolica	18.6514	−96.9660
UOGV-579	*X. g. grandis 2*	Mexico: Veracruz: Cuautlapan	18.8720	−97.0302
UOGV-572	*X. g. grandis 3*	Mexico: Puebla: Zoquitlán	18.3297	−97.0099
UOGV-577	*X. g. grandis 4*	Mexico: Puebla: Zoquitlán	18.3297	−97.0099
MZFC-22176	*X. g. grandis 5*	Mexico: Puebla: Tepequezquiapan, Eloxochitlan-Tlacotepec de Diaz road	18.4749	−96.9180
ANMO-947	*X. g. grandis 6*	Mexico: Oaxaca: Puente de Fierro	18.1538	−96.8528
MZFC-9529	*X. g. grandis 7*	Mexico: Oaxaca: Joya Maria	18.2155	−96.8359
MZFC-9577	*X. g. grandis 8*	Mexico: Oaxaca: San Martín Caballero	18.1088	−96.6336
MZFC-9574	*X. g. grandis 9*	Mexico: Oaxaca: 2 km W of San Martín Caballero	18.1102	−96.6515
MZFC-9575	*X. g. grandis 10*	Mexico: Oaxaca: 2 km W of San Martín Caballero	18.1102	−96.6515
LCM-1272	*X. g. grandis 11*	Mexico: Oaxaca: Chiquihuitlán	17.9956	−96.7588
MZFC-9533	*X. g. grandis 12*	Mexico: Oaxaca: Santa Rosa	18.2505	−96.8059
JAC-16882	*X. g. rackhami* 1	Guatemala: Huehuetenango: Santa Cruz Barilla, Finca Chiblac	15.8207	−91.3372
UOGV-997	*X. g. rackhami* 2	Mexico: Chiapas, Coapilla	17.1033	−93.1361
MZFC-9564	*X. g. sanmartinensis* 1	Mexico: Veracruz: Bastonal, Santa Martha	18.3883	−94.9400
UOGV-2683	*X. g. sanmartinensis* 2	Mexico: Veracruz: Trail Along Road NE San Andres Tuxtla	18.5219	−95.1491
MZFC-7489	*X. mendozai 1*	Mexico: Querétaro: Santa Ines-Tilaco road	21.1781	−99.1678
UOGV-1577	*X. mendozai 2*	Mexico: Hidalgo: Jacala, El Pinalito	21.0419	−99.1703
MZFC-8454	*X. newmanorum 1*	Mexico: San Luis Potosí: 5 km N of Xilitla	21.3969	−98.9675
MZFC-8452	*X. newmanorum 2*	Mexico: San Luis Potosí: 5 km N of Xilitla	21.3969	−98.9675
MZFC-7099	*X. penai 1*	Mexico: Guerrero: 1 km NW of Pico del Aguila, Yucuichinio	16.8500	−98.3300
MZFC-8479	*X. penai 2*	Mexico: Guerrero: 1 km NW of Pico del Aguila, Yucuichinio	16.8500	−98.3300
ANMO-2741	*X. phalaroanthereon 1*	Mexico: Oaxaca: Santa María Albarradas, 30 km NE of Mitla	16.9587	−96.2081
MZFC-15609	*X. phalaroanthereon 2*	Mexico: Oaxaca: Santa Maria Albarradas, 32.7 km NE of Mitla	16.9600	−96.2073
WSB-689	*X. phalaroanthereon 3*	Mexico: Oaxaca: San Juan Acaltepec	16.4333	−95.9667
UOGV-519	*X. platyceps 1*	Mexico: Tamaulipas, Gómez Farías, Alta Cima	23.0554	−99.2026
MZFC-8519	*X. platyceps 2*	Mexico: Tamaulipas, Gómez Farías, Ejido Azteca	23.0294	−99.1465
MZFC-9559	*X. platyceps 3*	Mexico: Tamaulipas: 18.9 km SW of Ciudad Victoria	23.7083	−99.1317
JADE-29	*X. rectocollaris* 1	Mexico: Puebla: Chapulco	18.6167	−97.3250
UOGV-1021	*X. rectocollaris* 2	Mexico: Puebla: Zapotitlán, Cerro El Pajarito	18.3801	−97.5075
TJO-12	*X. rectocollaris* 3	Mexico: Oaxaca: Tepelmeme	17.8959	−97.3279
UOGV-413	*X.* sp Camotlán 1	Mexico: Oaxaca: San Lucas Camotlán	16.9450	−95.7111
UOGV-416	*X.* sp Camotlán 2	Mexico: Oaxaca: San Lucas Camotlán	16.9450	−95.7111
IDF-20	*X.* sp Huehuetla 1	Mexico: Puebla: Huehuetla, Chilocoyo del Carmen	20.1108	−97.6529
MZFC-9579	*X.* sp Huehuetla 2	Mexico: Puebla: Huehuetla, Chilocoyo del Carmen	20.1108	−97.6529
ANMO-959	*X.* sp Huehuetla 3	Mexico: Puebla: Xochitlán, Apulco River	19.9225	−97.6046
ANMO-812	*X.* sp Monteverde 1	Mexico: Oaxaca: Altamira Monteverde, La Unión	16.9948	−97.7011
ANMO-814	*X.* sp Monteverde 2	Mexico: Oaxaca: Altamira Monteverde, La Unión	16.9948	−97.7011
LCM-1126	*X.* sp Pápalos 1	Mexico: Oaxaca: San Isidro Buenos Aires	17.9394	−96.8657
UOGV-437	*X.* sp Pápalos 2	Mexico: Oaxaca: 1.5 km SW of San Lorenzo Pápalo	17.8859	−96.8879
EPR-27	*X.* sp Vista Hermosa 1	Mexico: Oaxaca: La Esperanza	17.6333	−96.3667
JAC-10284	*X.* sp Vista Hermosa 2	Mexico: Oaxaca: La Esperanza	17.6333	−96.3667
MZFC-12870	*X.* sp Tejocote 1	Mexico: Oaxaca: 4 km SW of El Tejocote	17.2100	−97.0200
MZFC-9517	*X. tzacualtipantecus* 1	Mexico: Hidalgo: Zacualtipán, La Mojonera	20.6399	−98.5990
UOGV-581	*X. tzacualtipantecus* 2	Mexico: Hidalgo: Zacualtipán, La Mojonera	20.6399	−98.5990

2 h at 37 °C. Fragments were purified with Agencourt AMPure beads before ligation of barcoded Illumina adaptors onto the fragments. The oligonucleotide sequences used for barcoding and adding Illumina indexes during library preparation are provided in Peterson et al. (2012). Equimolar amounts of each sample were pooled, with each pool containing up to eight unique barcoded samples, in a 96-well plate format. The barcodes differed by at least two base pairs to reduce the chance of errors caused by inaccurate barcode assignment. The pooled libraries were size-selected (between 415 and 515 bp after accounting for adapter length) on a Pippin Prep size fractionator (SageScience) according to manufacturer instructions. We sequenced the final pool of libraries on a single Illumina HiSeq 2000 under a 50 bp single-end read protocol.

2.3. ddRADseq bioinformatics

We processed raw Illumina reads with the software pipeline pyRAD (Eaton, 2014) version 3.0.63. We demultiplexed the samples using their unique barcode and adapter sequences. After the removal of the 6-bp restriction site overhang and the 5-bp barcode, each locus was reduced from 50 to 39 bp. Sites with accuracy of the base call under 99% (Phred quality score = 20) were changed into "N" characters, and reads with $\geqslant$10.2% N's (i.e., $\geqslant$4) were discarded. Also, four samples with <50,000 reads passing the quality filter were excluded from further analysis. Within the pyRAD pipeline, reads from each sample are clustered using the program VSEARCH version 1.9.3 (https://github.com/torognes/vsearch) and aligned with MUSCLE version 3.8.31 (Edgar, 2004). The first clustering step establishes homology among reads within samples. We determined the optimal value for the clustering parameter using the clustering threshold series approach described by Ilut et al. (2014). This method seeks to assemble reads into loci such that false homozygosity (splitting reads from a single locus into two) and false heterozygosity (due to clustering of paralogs) is minimized (i.e., the optimum clustering threshold). We generated a clustering threshold series (*sensu* Ilut et al., 2014) using similarity

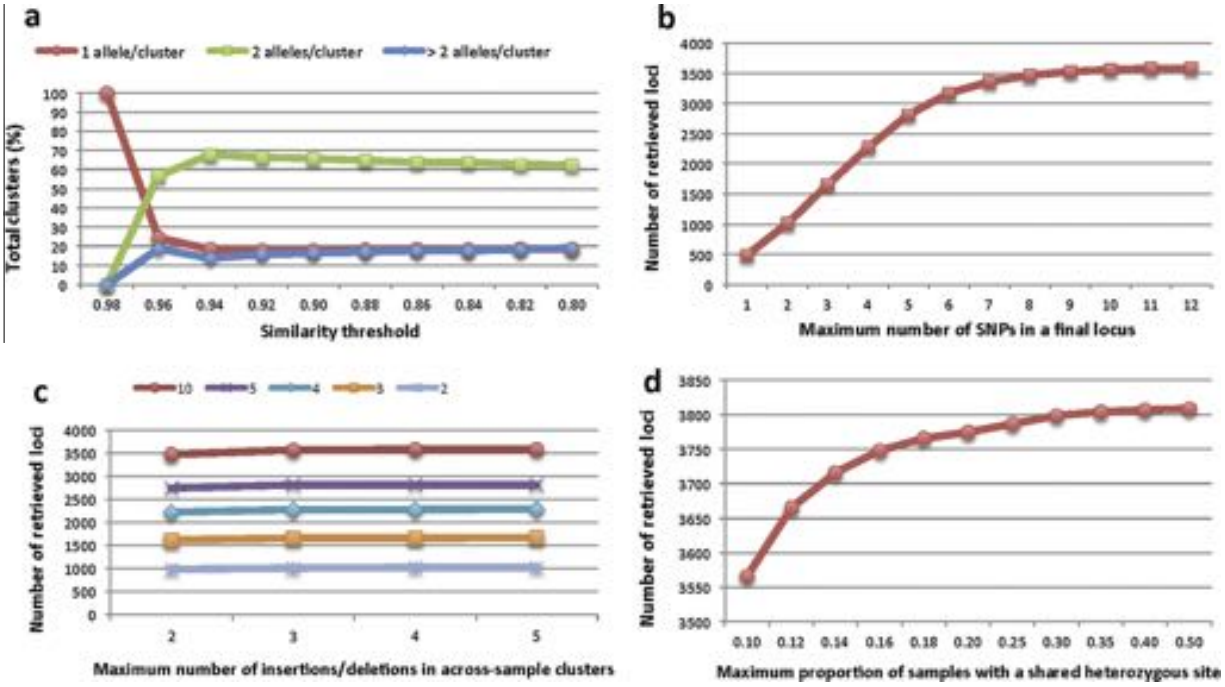

Fig. 3. Variation in the proportion of clusters with 1, 2, and >2 alleles retrieved with different similarity thresholds (a); variation in the number of retrieved loci with different maximum numbers of SNPs in a final locus (b); variation in the number of retrieved loci with different maximum numbers of insertions/deletions in across-sample clusters (c); variation in the number of retrieved loci with different maximum proportions of samples with a shared heterozygous site (d). The different line colors in (a) and (c) represent different numbers of alleles/cluster and different numbers of insertions/deletions in across sample-clusters, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

thresholds ranging from 0.80 to 0.98. The optimal clustering threshold was 0.92 (Fig. 3a), which is used for both within and among sample clustering.

After the clustering of reads within samples, we estimated the error rate and heterozygosity from the base counts in each site across all clusters, and these values were used to generate consensus sequences for each cluster. Consensus sequences were then clustered across samples and aligned as described above. We examined the sensitivity of the final dataset to changing the maximum number of SNPs allowed in a locus. The number of retained loci increased linearly with higher numbers of SNPs allowed, until it began to plateau at a maximum of 10 SNPs (Fig. 3b). We used this value for the assembly of the final dataset under the rationale that above this value, the small number of additional loci that were retained potentially represented paralogs. In addition, for any maximum number of SNPs allowed per locus between 1 and 10, the number of retained loci increased slightly when the maximum number of insertions/deletions allowed in across-sample clusters increased from 2 to 3, but not appreciably further with higher numbers (e.g., 4 and 5; Fig. 3c). Thus, we allowed a maximum of 3 insertions/deletions in across-sample clusters for the assembly of the final dataset. We also examined the sensitivity of the final dataset to changing the maximum proportion of samples allowed to share a heterozygous site (between 0.1–0.5). The number of loci retained increased roughly linearly (Fig. 3d) until it first plateaued at a value of 0.2 (corresponding to ~9 samples), which we again chose for the final value under the rationale that loci exhibiting higher shared heterozygosity potentially represented paralogs. We also discarded loci that had >4 undetermined or heterozygous sites (default pyRAD settings), or >2 haplotypes (to filter paralogs), and used a minimum depth of coverage of 10 for genotype calls. We set the minimum number of ingroup samples with data for a given locus to be retained in the final dataset to about 50% of the samples (24). The total number of retained loci in this dataset was 4077 (161,476 characters; 33% of missing data). Unless otherwise stated, all of the phylogenetic analyses were performed on this dataset.

2.4. Phylogeny reconstruction

As in the majority of RADseq phylogenetic studies, we did not attempt to estimate gene trees because short read lengths cause the individual loci to be minimally phylogenetically informative. However, we used several approaches to investigate the phylogenetic history of *Xenosaurus*.

First, we used maximum likelihood and Bayesian methods to infer a phylogenetic tree from the concatenated ddRAD loci in the final dataset. We performed the maximum likelihood analysis using RAxML-HPC version 8.0 (Stamatakis, 2014) with the GTR + GAMMA model of nucleotide substitution and the –f a option, which searches for the best-scoring tree and performs a rapid bootstrap analysis to estimate node support. The analyses were run on the CIPRES Science Gateway, version 3.3 (Miller et al., 2010). We also conducted a concatenated analysis using MrBayes version 3.2.6 (Ronquist et al., 2012). The analysis consisted of two runs for 1×10^7 generations with the GTR + GAMMA substitution model and four Markov chains (heating parameter = 0.05) sampled every 1000 generations. To evaluate convergence on a stationary distribution, we used two convergence diagnostics: the values for the average standard deviation of split frequencies and the Potential Scale Reduction Factor of the branch lengths. As the two runs converge onto a stationary distribution, these values should approach 0 and 1.0, respectively (Ronquist et al., 2011). In addition, we evaluated convergence with Tracer version 1.6.1 (Rambaut et al., 2014). We discarded 25% of the samples as burn-in, and used the remaining samples to compute a majority consensus tree. Finally, we also used BEAST version 1.8.2 (Drummond et al., 2012) to perform a concatenated phylogenetic analysis using a GTR + GAMMA substitution model, an uncorrelated lognormal relaxed clock, and

a birth-death process tree prior. The analysis was run for 20 million generations, sampling every 2000 generations.

We also performed a phylogenetic analysis of the data using PhyloBayes-MPI version 1.5 (Lartillot et al., 2009, 2013). Compared to other phylogenetic MCMC samplers, the main distinguishing feature of PhyloBayes is the use of Dirichlet process mixtures to modeling among-site variation in nucleotide propensities. Treating site-specific profiles and rates as random variables and estimating them from the data has been shown to result in better model fit and greater phylogenetic accuracy (Lartillot et al., 2013). However, due to the complexity of the PhyloBayes model, computational constraints prohibited us from analyzing the entire ~160,000 bp dataset under the model. Additionally, including constant sites in datasets analyzed under infinite mixture models frequently causes mixing problems (Lartillot et al., 2013; a result we also noted during preliminary analyses of the dataset). Thus, we tried running several different datasets that were trimmed until the analysis became computationally feasible. We conducted the analysis on a dataset composed of 47,332 aligned positions, which represent a subset of the total loci with the least missing data, and eliminated the constant sites from the alignment. After removing the constant sites, 3608 sites were included in the analysis. Removing invariant sites causes acquisition bias that can inflate branch length estimates, and in extreme cases produce an inaccurate phylogeny (Leaché et al., 2015a). However, the PhyloBayes tree did not show conspicuously longer branches than the trees from the above methods, and the topology of all of these trees was the same (see below). We ran the analysis under the CAT-GTR model for 30,000 cycles, sampling the chain every 10 cycles. Convergence was assessed with the tracecomp and bpcomp programs with burn-in = 10% and Tracer version 1.6.1 (Rambaut et al., 2014).

In addition, we analyzed the final dataset using the SVDquartets species tree model (Chifman and Kubatko, 2014) as implemented in PAUP* version 4a146 (Swofford, 2002). This method infers relationships among quartets of taxa under a coalescent model and then estimates the species tree using a quartet assembly method. Because of this, the method is only able to estimate a specie tree topology lacking branch lengths. For this analysis, we extracted the SNP with the least missing data from each locus of the final dataset. We evaluated all the possible quartets from the data matrix, and used the QFM quartet amalgamation and multispecies coalescent options in PAUP* to infer a species tree from the quartets. We used non-parametric bootstrapping with 100 replicates to assess the variability in the estimated tree.

2.5. Species delimitation

We used several methods to address the following species delimitation problems in *Xenosaurus*: (1) the distinctness of all of the subspecies of *X. grandis*; (2) the distinctness of all of the putative undescribed species; and (3) the potential existence of multiple species within *X. g. grandis* (see above). First, we used two species discovery methods that infer putative species limits on a given phylogenetic tree. The GMYC method (Fujisawa and Barraclough, 2013; Pons et al., 2006) models speciation branching events via a pure birth process and within-species branching events as neutral coalescent processes, and infers the transition point between inter-and intra-species branching rates on a time-calibrated ultrametric tree. We ran the analyses on the GMYC web server (http://species.h-its.org/gmyc/) under the single threshold and multiple threshold models using the ultrametric tree from the concatenated BEAST analysis (see above).

The PTP method (Zhang et al., 2013) is similar to the GMYC method in that it seeks to identify significant changes in the rate of branching in a phylogenetic tree. However, rather than using time to estimate branching rates as in the GMYC model, PTP directly uses the number of substitutions. We implemented the PTP method on the majority consensus tree from the above concatenated Bayesian analysis of the final dataset in the bPTP web server (http://species.h-its.org/ptp/) for 500,000 generations, with thinning = 100 and burn-in = 10%. The bPTP web server runs both the original maximum likelihood version of PTP and an updated version which adds Bayesian support to delimited species on the input tree (Bayesian implementation of the PTP model or bPTP). We report the results of both the maximum likelihood and Bayesian implementations of the model.

Second, we also used BPP version 3.1 (Yang, 2015; Yang and Rannala, 2010, 2014) to jointly perform species delimitation and species tree inference under the multispecies coalescent model. Computational constraints of the method again prohibited us from analyzing the full dataset; however, to ensure consistency of the species delimitation analyses across runs, and examine the sensitivity of the method to how individuals are assigned to species, we set up several different analyses. The BPP program uses a rjMCMC algorithm (Rannala and Yang, 2013) to estimate a posterior distribution of species delimitations and species trees, while integrating over uncertainty in gene trees. The assignment of individuals to populations or putative species is fixed, and the model is able to merge different specified groups into a single species, but not to split pre-defined populations into multiple species. In the first analysis, we grouped individuals into 12 species corresponding to every described species and subspecies, and 6 species corresponding to every putative undescribed species, for a total of 18 species. We also set up several analyses splitting out groups of populations as suggested by the GMYC and PTP analyses, and previous taxonomic treatments (with the numbers of species being 20, 21, and 23; see Table 2). Because these analyses are computationally intensive and included all sampled individuals from the study, we performed each analysis using relatively few loci. Since setting the minimum number of ingroup samples with data for a given locus to be retained in the final data to 47 and 48 returned 129 and 91 loci, respectively, we decided to perform each analysis with 129 loci. Finally, we ran two smaller analyses, one that only included samples from the *X. g. rackhami*, *X. g. sanmartinensis*, and *X.* sp Camotlán populations, and another with only the *X. grandis grandis* samples. Because these analyses were performed on smaller datasets, we arbitrarily chose to run each analysis utilizing the 300 loci with the least amount of missing data for their respective clades.

In all of the BPP analyses, all population size parameters (θ) were assigned the gamma prior G(1, 10), and the divergence time at the root of the species tree ($\tau 0$) was assigned the gamma prior G (2, 2000). These priors correspond to large ancestral population sizes and shallow divergences, and are considered to favor conservative models containing fewer species (Leaché and Fujita, 2010; Yang, 2015). We used the topology of the majority-rule consensus trees from the above concatenated ML and Bayesian analyses of the final dataset as the starting tree. The analysis was run for a total of 2,500,000 iterations (sampling interval of 5) with a burn-in of 1000.

We also performed Bayes Factor species delimitation (BFD*; Grummer et al., 2014; Leaché et al., 2014) using the SNAPP model (Bryant et al., 2012). SNAPP models the probability of allele frequency change across ancestor/descendent nodes directly from biallelic character data, thus bypassing the necessity of having to explicitly estimate gene trees for each locus, and estimates a posterior distribution for the species tree, species divergence times, and effective population sizes. Comparisons among candidate species delimitation models can be performed using Bayes factors (Kass and Raftery, 1995), which require the estimation of marginal likelihoods for each competing model.

Table 2
Summary of species delimitation results. PTP ML = PTP analysis, maximum likelihood results; bPTP = PTP analysis, Bayesian results. PTP values are posterior delimitation probabilities. BPP analyses of the complete dataset: BPP 1: Populations of all of the described taxa and putative undescribed species each assigned to a single species (18 species). BPP 2: Same, except populations of *Xenosaurus g. grandis* split into four species as per the clades in Fig. 4, and populations of *X. rectocollaris* split into three species. BPP 3: Same, except only populations of *X. g. grandis* split. BPP 4: Same, except only populations of *X. rectocollaris* split. BPP 5: Only populations of *X. g. grandis*, split into four species as above. BPP 6: Only populations of *X. g. rackhami*, *X. g. sanmartinensis*, and *X.* sp. Camotlán. BPP values are posterior probabilities.

Taxon/method	GMYC single threshold	GMYC multi threshold	PTP ML	bPTP	BPP 1	BPP 2	BPP 3	BPP 4	BPP 5	BPP 6
	Xenosaurus grandis subspecies									
X. grandis agrenon	√	√	0.53	0.53	1.00	1.00	1.00	1.00		
X. grandis arboreus	√	√	0.92	0.92	1.00	1.00	1.00	1.00		
X. grandis grandis			0.72	0.72	1.00			1.00		
X. g. grandis 1	√	√				0.62	0.70		1.00	
X. g. grandis 2						0.62	0.61		1.00	
X. g. grandis 3						1.00	0.73		1.00	
X. g. grandis 4						1.00	0.80		1.00	
X. g. grandis 1 + 2						0.38	0.12			
X. g. grandis 2 + 3							0.07			
X. g. grandis 2 + 3 + 4	√	√					0.02			
X. g. grandis 1 + 2 + 3 + 4							0.18			
X. grandis rackhami					0.94	0.94	0.77	0.63		1.00
X. grandis rackhami 1				0.35						
X. grandis rackhami 2				0.35						
X. grandis sanmartinensis				0.58	0.94	0.94	0.77	0.63		1.00
X. g. rackhami + *X. g. sanmartinensis*	√	√	0.31		0.06	0.06	0.23	0.37		
	Other described species									
X. mendozai	√	√	0.77	0.77	1.00	1.00	1.00	1.00		
X. newmanorum	√	√	0.80	0.80	1.00	1.00	1.00	1.00		
X. penai	√	√	0.50		0.99	0.97	1.00	1.00		
X. penai 1				0.50						
X. penai 2				0.50						
X. phalaroanthereon	√	√	0.61	0.61	1.00	1.00	1.00	1.00		
X. platyceps	√	√	0.55	0.55	1.00	1.00	1.00	1.00		
X. rectocollaris		√			1.00		1.00			
X. rectocollaris 1	√		0.91	0.91		0.99		0.99		
X. rectocollaris 2	√		0.91	0.91		0.99		0.99		
X. rectocollaris 3	√		1.00	1.00		1.00		1.00		
X. tzacualtipantecus	√	√	0.85	0.85	1.00	1.00	1.00	1.00		
	Putative undescribed species									
X. sp Huehuetla	√	√	0.75	0.75	1.00	1.00	1.00	1.00		
X. sp Monteverde	√	√	0.51	0.51	0.99	0.97	1.00	1.00		
X. sp Pápalos	√	√	0.79	0.79	1.00	1.00	1.00	1.00		
X. sp Camotlán	√	√	0.62	0.62	1.00	1.00	1.00	1.00		1.00
X. sp Tejocote	√	√	1.00	1.00	0.99	1.00	1.00	1.00		
X. sp Vista Hermosa	√	√	0.83	0.83	1.00	1.00	1.00	1.00		

We performed species delimitation using SNAPP for the two scaled-down analyses discussed above. For each clade, we also included the sister taxon (or a representative of the sister taxon) as an outgroup in order to test the hypothesis that all populations represent a single species. The competing species delimitation models and out groups used in these analyses are provided in Table 3.

We used the software package phrynomics version 2.0 (Leaché et al., 2015a) to extract the unlinked SNPs with the least missing data across taxa and transform the SNP data matrix from pyRAD into SNAPP format. We used SNAPP (Bryant et al., 2012) in BEAST2 version 2.3.1 (Bouckaert et al., 2014). We used mutation rates U = V = 1.0, sampled the coalescence rate, and used the log likelihood correction. We used the average sequence divergence among samples known to belong to a single species as an estimation of the mean for the expected divergence (theta) prior, and a hyperprior (gamma distribution) to sample the lambda parameter of the Yule prior on the species tree. We used the maximum observed sequence divergence between any pair of taxa (including the outgroup) as an estimation of the mean of the gamma distribution for

Table 3
Competing species delimitation models and their rank in the BFD* analyses. ML = marginal likelihood. BF = Bayes factor. The outgroup for the populations of *X. g. grandis* was *X.* sp Vista Hermosa, and the outgroup for the populations of *X. g. rackhami*, *X. g. sanmartinensis*, and *X.* sp Camotlán was *X. arboreus*. The population of *Xenosaurus* from San Lucas Camotlán was regarded as an intergrade between *X. g. grandis* and *X. g. rackhami* by King and Thompson (1968).

Species delimitation model	Species	ML	Sites	Rank	BF
	X. grandis grandis				
A: Current taxonomy	1	−4533.76	401	4	
B: Split into *X. g. grandis* 1 + 2 + 3 and *X. g. grandis* 4	2	−4133.81	401	3	−799.90
C: Split into *X. g. grandis* 1 + 2, *X. g. grandis* 3, and *X. g. grandis* 4	3	−4006.23	401	2	−1055.06
D: Split into *X. g. grandis* 1, *X. g. grandis* 2, *X. g. grandis* 3, and *X. g. grandis* 4	4	−3882.79	401	1	−1301.94
	X. g. rackhami–*X. g. sanmartinensis*				
A: Current taxonomy: *X. g. rackhami*, *X. g. sanmartinensis*, and *X. g. grandis* x *X. g. rackhami*	1	−2302.62	430	4	
B: Split into (*X. g. sanmartinensis* + *X. g. grandis* x *X. g. rackhami*) and *X. g. rackhami*	2	−2137.60	430	3	−330.04
C: Split into (*X. g. rackhami* + *X. g. sanmartinensis*) and *X.* sp Camotlán	2	−1998.70	430	2	−607.84
D: Split into *X. g. rackhami*, *X. g. sanmartinensis*, and *X.* sp Camotlán	3	−1833.20	430	1	−938.84

lambda. The MCMC chain was run for 1,000,000 generations, sampling every 1000th. We conducted path sampling with 48 steps (300,000–500,000 MCMC steps, 10,000 pre-burnin steps, alpha = 0.3) to estimate the marginal likelihood of each species delimitation model. Convergence was evaluated with Tracer version 1.6.1 (Rambaut et al., 2014). We ranked the alternative species delimitation models by their marginal likelihood and calculated Bayes factors to compare the models. The strength of support from Bayes Factors comparisons of competing models was evaluated using the framework of Kass and Raftery (1995).

3. Results

3.1. Phylogeny of Xenosaurus

The total number of raw and variable loci obtained was 54,420 and 33,095, respectively, and the average coverage of loci was 16.477.

The ML and Bayesian consensus trees (from BEAST, MrBayes, and PhyloBayes) were identical in topology (Fig. 4). Nearly all nodes in the trees from the RAxML, BEAST, and MrBayes analyses were strongly supported, whereas the nodes in the PhyloBayes tree generally had lower support values. The phylogenetic analyses found *Xenosaurus* to be composed of four major, allopatric clades concordant with geography. We name each of these clades after the first species described in the clade (Figs. 4 and 5). The first major clade that branches off the tree (the *newmanorum* clade) was composed of *X. mendozai* as sister taxon to *X. platyceps* and *X. newmanorum*, and is distributed from southwest Tamaulipas to northern Querétaro and Hidalgo on the Atlantic slopes of the Sierra Madre Oriental. The second major clade (the *tzacualtipantecus* clade) was composed of *X. tzacualtipantecus* and *X.* sp Huehuetla, which collectively occur from central-east Hidalgo and adjacent Veracruz to the Mexican Transvolcanic Belt in Puebla, also on the Atlantic slopes of the Sierra Madre Oriental, and was the sister group to the other two major clades.

The third major clade was the *grandis* clade, which consisted of nine taxa that occur collectively from the Atlantic slopes of the Mexican Transvolcanic Belt in west-central Veracruz south to the Pacific slopes of the Sierra Madre del Sur in Guerrero and Oaxaca

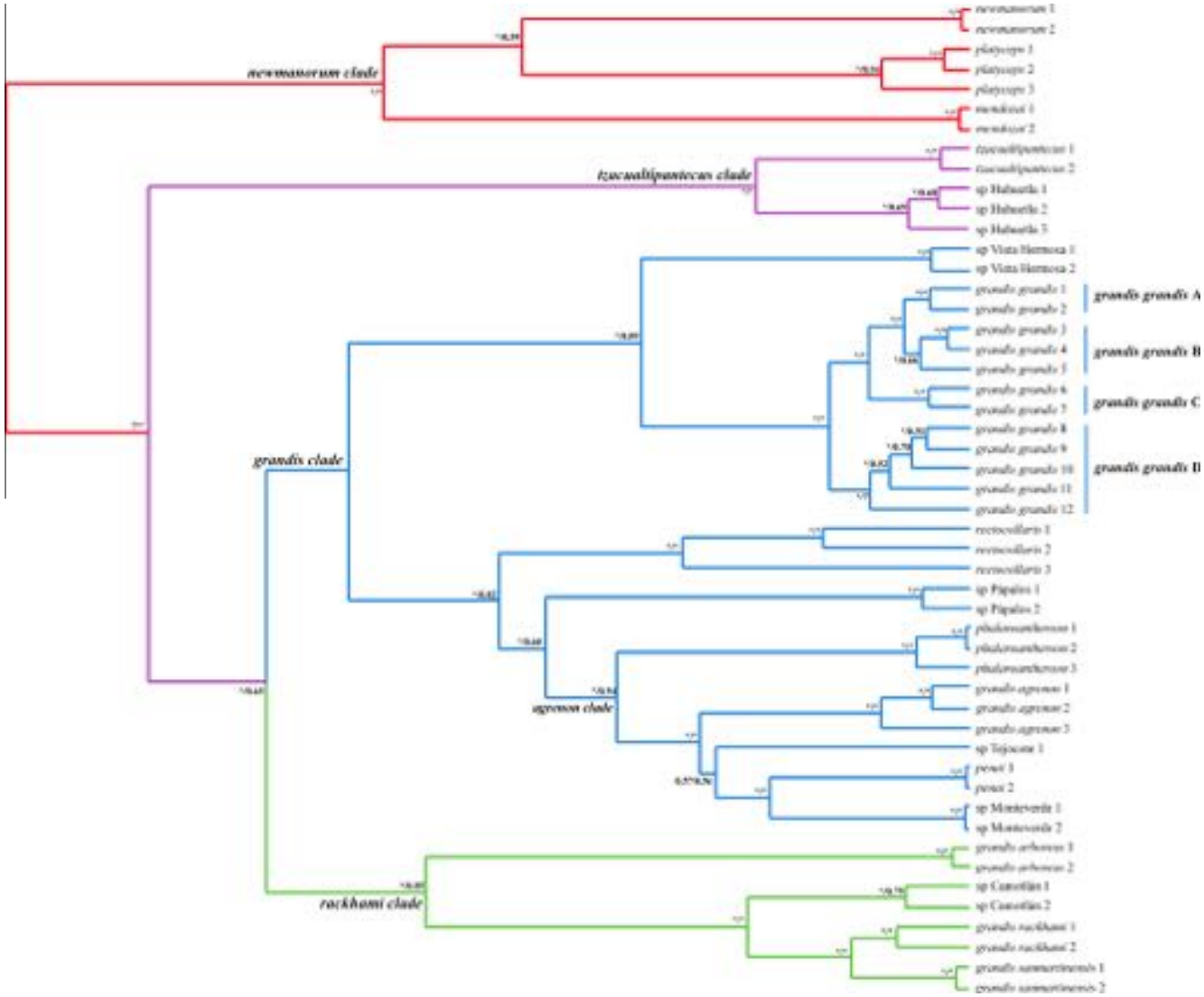

Fig. 4. Majority-rule consensus tree from the concatenated BEAST analysis. The clades with red, purple, blue, and green branches are the *newmanorum*, *tzacualtipantecus*, *grandis*, and *rackhami* major clades, respectively. *Xenosaurus g. grandis* clades A, B, C, and D are potential cryptic species. Nodal support values are posterior probabilities (BEAST/PhyloBayes trees; * = posterior probability ⩾0.95). All of the nodes with posterior probabily ⩾0.95 in the BEAST tree had bootstrap values ⩾75% in the RAxML tree and posterior probability ⩾0.95 in the MrBayes tree. Support for the *X.* sp Tejocote + (*X. penai* + *X.* sp Monteverde) clade in the latter trees was 41% and 0.52, respectively.

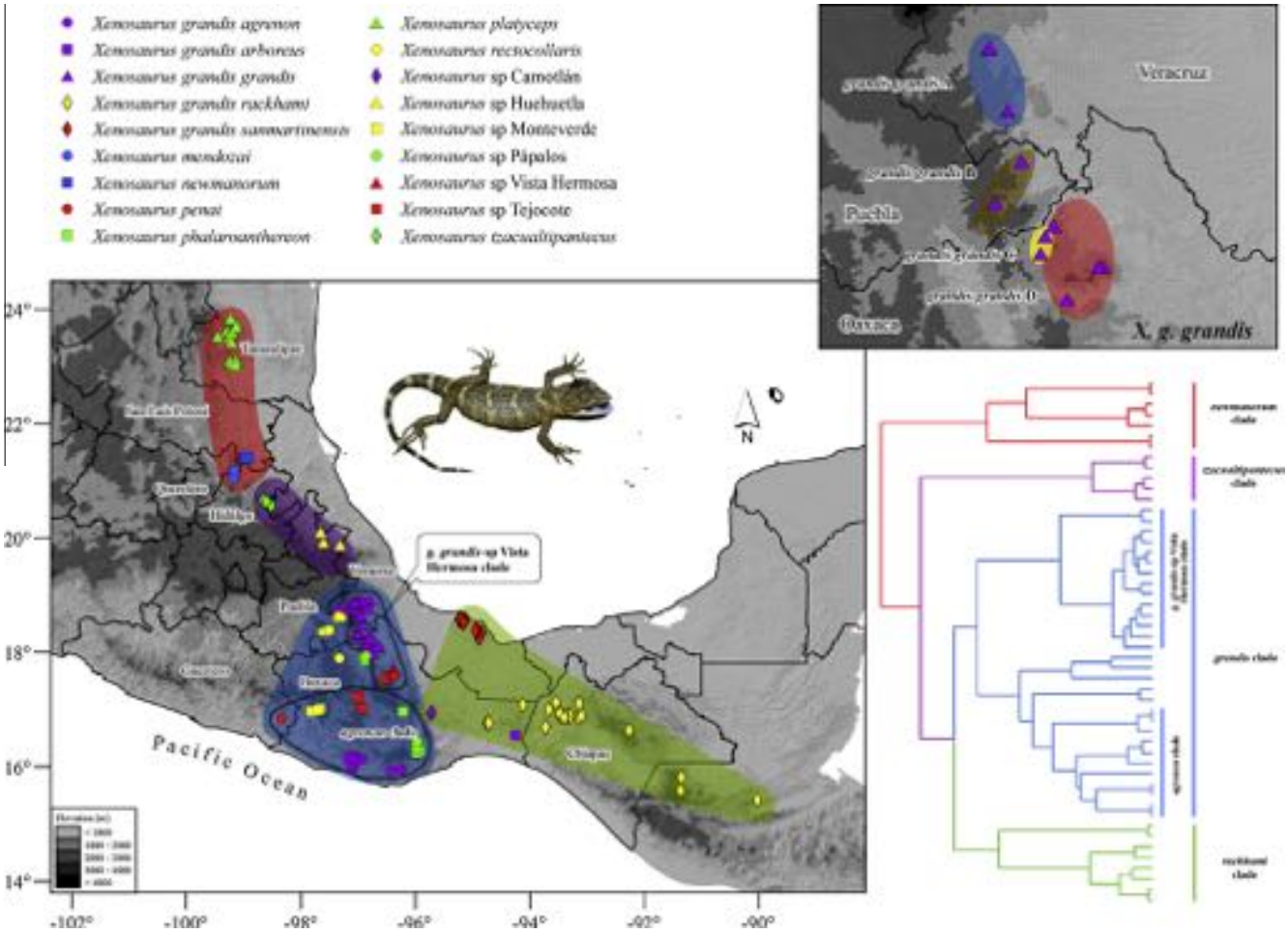

Fig. 5. Geographic distribution of the major clades of *Xenosaurus*. Inset: geographic distribution of the putative cryptic species within *X. g. grandis*.

(Figs. 4 and 5). Within this clade, *X. g. grandis* and *X.* sp Vista Hermosa, from the Atlantic slopes of the mountain ridges extending from the Mexican Transvolcanic Belt in west-central Veracruz south-southeast to north-central Oaxaca (the Sierra de Zongolica, Sierra Mazateca, and Sierra de Juárez), comprised the sister group of the remaining seven taxa, which occur west and south of these mountain ridges. Of these, five comprise a clade on the Pacific versant of eastern Guerrero and western and central Oaxaca (the *agrenon* clade). Within this clade, *X.* sp Tejocote, from the Sierra de Cuatro Venados in central Oaxaca, was the sister taxon to *X. penai* and *X.* sp Monteverde, from the Sierra Madre del Sur in eastern Guerrero and western Oaxaca, respectively, and the closest relative to these three taxa within the *agrenon* clade was *X. agrenon*, followed by *X. phalaroanthereon*, both from the Pacific slopes of the Sierra Madre del Sur in south-central Oaxaca and the slopes of the Sierra Mixe (*X. phalaroanthereon*) in central Oaxaca. The closest relative to the *agrenon* clade was *X.* sp Pápalos, followed by *X. rectocollaris*, both from the Tehuacán-Cuicatlán-Quiotepec Depression. This basin, although lying west of the Sierra de Zongolica and Sierra Mazateca in southeastern Puebla and north-northwestern Oaxaca (i.e., to the side opposite to their Atlantic versant), drains to the Gulf of Mexico in the Atlantic versant.

The last major clade was the *rackhami* clade (Figs. 4 and 5). In this clade, *X.* sp Camotlán, known from a single locality west of the Isthmus of Tehuantepec in Oaxaca, was the sister taxon to *X. g. sanmartinensis*, endemic to Los Tuxtlas range at the northern end of the Isthmus, and *X. g. rackhami*, which ranges from eastern Oaxaca east to Guatemala east of the Isthmus. *Xenosaurus g. arboreus*, from the Sierra Madre in extreme southeastern Oaxaca east of the Isthmus, was the sister taxon to the other three taxa.

The topology of the SVDquartets species tree (Fig. 6) was generally similar to the topologies from the other analyses, except that *X. g. rackhami* was paraphyletic with respect to *X. g. sanmartinensis* and *X. platyceps* was the sister taxon to *X. mendozai*, rather than to *X. newmanorum*. However, only the former relationship was strongly supported. In addition, the *grandis* clade was not strongly supported.

3.2. Species delimitation

The results of the species delimitation analyses with the GMYC, PTP, and BPP methods are summarized in Table 2. The GMYC method (single and multiple threshold) and PTP method (ML and Bayesian results) identified all of the described species of *Xenosaurus* as distinct lineages. Additionally, most of the subspecies of *X. grandis* and all of the putative undescribed species were identified as distinct in each of the analyses. However, some analyses identified additional lineages within some of the above taxa. The confidence intervals for the number of ML entities identified in the GMYC single and multi threshold analyses were broad (4–24 and 6–29, respectively). Also, the posterior delimitation probabilities of the identified taxa in the PTP analyses varied widely across taxa.

Similarly, the BPP analyses identified all of the described species of *Xenosaurus*, all of the subspecies of *X. grandis*, and all of the putative undescribed species as distinct lineages. Also, the BPP analyses

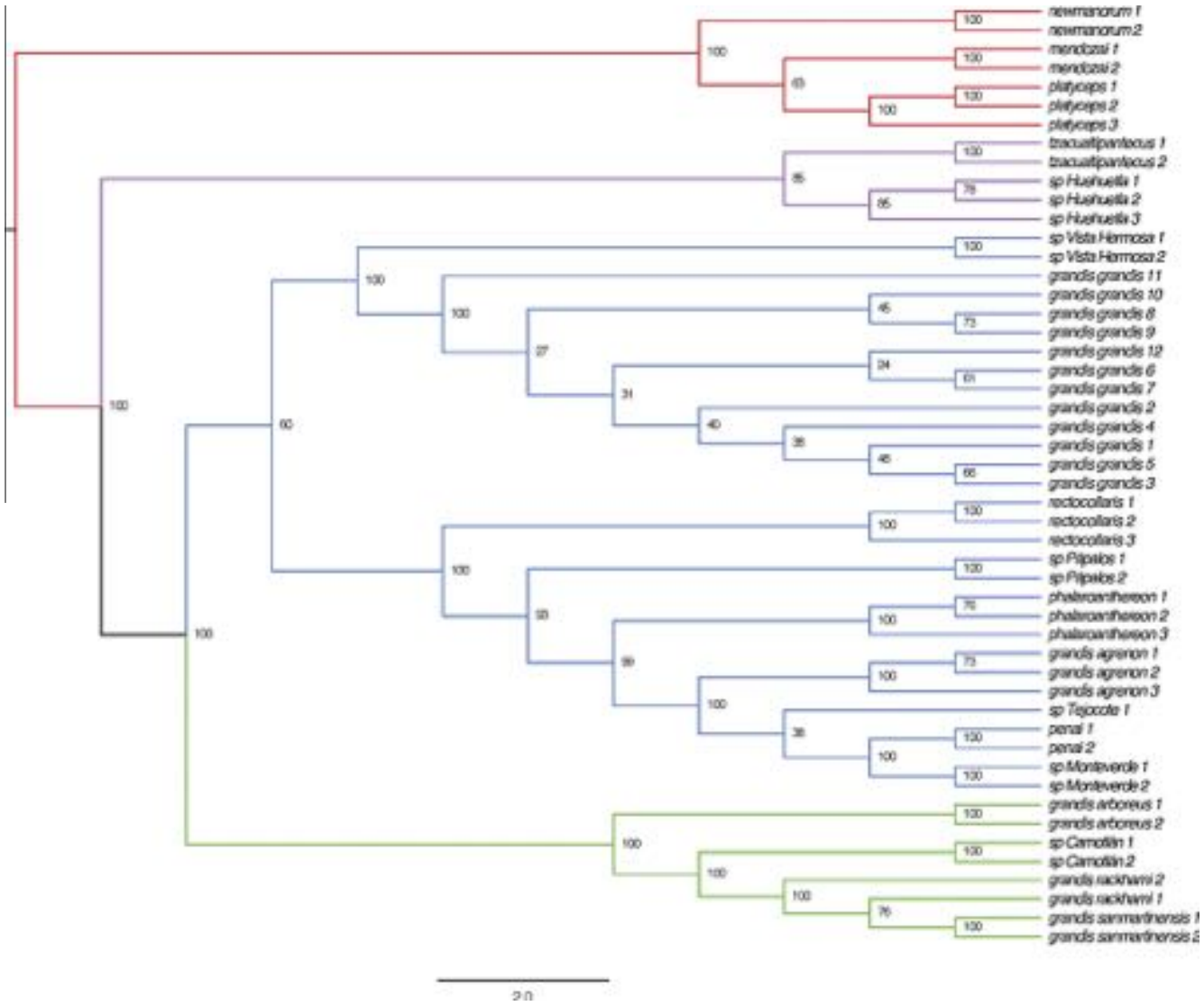

Fig. 6. Majority-rule consensus tree from the SVDquartets analysis. The clades with red, purple, blue, and green branches are the *newmanorum*, *tzacualtipantecus*, *grandis*, and *rackhami* major clades, respectively. Nodal support values are bootstrap proportions.

generally split species into the most liberal groupings in each of the analyses. Results from the different analyses were largely consistent, although we did note several interesting differences between several runs.

The results of the BFD* analyses are provided in Table 3. In each clade, the top-ranked model was the one with the highest number of species, and the BFD in support of this model was decisive.

4. Discussion

4.1. Phylogeny of Xenosaurus

Knob-scaled lizards of the genus *Xenosaurus* are a unique radiation of species that are highly distinct (both genetically and morphologically) from their closest living relatives (the anguids and helodermatids). A flattened body shape and a crevice-dwelling ecology generally characterize species in the group. In part because of their habitat specialization, all of the known species and subspecies are allopatric, and several have small geographic distributions. In this study, we provide the first molecular phylogeny for *Xenosaurus* in order to elucidate the evolutionary history of this enigmatic clade and provide insights into the unique biogeographic patterns they exhibit. To do so, we generated a genome-scale dataset for all described species and subspecies in the genus. By analyzing the data using a variety of phylogenetic models, we were able to reconstruct a well-resolved phylogeny for this group. We also utilized a variety of molecular species delimitation approaches to investigate multiple cases of putative undescribed species diversity in *Xenosaurus*.

Although the use of genomic datasets for phylogenetics has provided increased power for resolving the tree of life, the field currently faces substantial computational and analytical challenges (Philippe et al., 2005; Jeffroy et al., 2006; Kumar et al., 2012). Because genomic datasets are becoming relatively easy to collect for nearly any group of organism and tools for analyzing these data are rapidly being developed, researchers are frequently confronted with a broad array of analytical choices. Some models trade complexity for computational efficiency. However, in many cases, choosing an appropriate model is not a straightforward exercise in model selection. In some circumstances, different approaches to analyzing phylogenetic datasets can have profound impacts on phylogenetic inference (e.g., Leaché et al., 2015b; Pisani et al., 2015; Shaffer et al., 2013; Whelan et al., 2015). Because of this,

we chose to employ a broad suite of phylogenetic models to analyzing our data to assess congruence across analytical approaches.

Some *in silico* studies have been applied to divergences dating back to 55–60 myr in mammals, *Drosophila*, and fungi (Rubin et al., 2012; Cariou et al., 2013). In addition, empirical ddRADseq data were able to resolve the deepest divergence in the phylogeny of phrynosomatid lizards, which began diversifying approximately 55 myr (Leaché et al., 2015b). Herein, phylogenetic inference with ddRADseq data was feasible at a relatively deep evolutionary timescale. However, there seems to be a low probability of obtaining large numbers of shared loci among distantly related species using ddRADseq, because the number of homologous loci obtained decreases in relation to time since divergence (Leaché et al., 2015b). For instance, in this study only 54 shared loci were obtained between the ingroup and a representative of *Abronia*. The split between Anguidae + Helodermatidae and Xenosauridae seems to have taken place between 99 Ma and 118 Ma (Wiens et al., 2006).

The phylogenetic hypotheses inferred for *Xenosaurus* were mostly consistent across all of the methods used in this study, and were thus able to adequately resolve most of the nodes in the *Xenosaurus* phylogeny. In addition, nearly all of the clades recovered in all of the trees, except for the PhyloBayes tree, were strongly supported. This is likely because the model in PhyloBayes is more complex than the model in the concatenated analysis, and because the analyzed dataset was much smaller due to computational limits, which decreased our power to infer parameters of the model (including the phylogeny). The only exception was the Xenosaurus sp Tejocote (X. penai + X. sp Monteverde) clade. The low support of this clade and the comparatively short branch subtending it suggest a case of rapid speciation within the *agrenon* clade (see below).

Furthermore, we found that the inferred phylogeny of *Xenosaurus* is congruent with the geography (see below). Several sister taxa are geographically closer to each other than they are to any other taxon (e.g., *X. tzacualtipantecus* and *X.* sp Huehuetla, the daughter clades of the *grandis* clade, *X. penai* and *X.* sp Monteverde, *X. arboreus* and the *X.* sp Camotlán + [*X. g. rackhami* + *X. g. sanmartinensis*] clade).

4.2. Biogeography of Xenosaurus

The two northernmost major clades (i.e., the *newmanorum* and *tzacualtipantecus* clades) and the *grandis* major clade of *Xenosaurus* are restricted to different morphotectonic provinces (the Sierra Madre Oriental and Sierra Madre del Sur, respectively), whereas the *rackhami* major clade is composed of those taxa with the closest distributions to the Isthmus of Tehuantepec in the neighboring Gulf Coastal Plain foreland, Sierra Madre del Sur, and Sierra de Chiapas provinces.

Bhullar (2011) identified the extinct taxon *Exostinos serratus*, from the Orellan sediments (early Oligocene; 33,900,000–33,300,000 years BP) as the sister taxon to *Xenosaurus*, and suggested that this genus may be very ancient. However, no taxa that fall within the crown clade *Xenosaurus* have been identified. The closest branches to the root of *Xenosaurus* are the two northernmost major clades (i.e., the *newmanorum* and *tzacualtipantecus* clades), distributed north of the Mexican Transvolcanic Belt, whereas the following clades up the tree are the *grandis* and *rackhami* clades, distributed south and southeast from the Belt. This is consistent with a Nearctic origin of *Xenosaurus*, as stem xenosaurs have been found in Colorado, Wyoming, and farther north still (Bhullar, 2011). The *newmanorum* and *tzacualtipantecus* clades are separated by the canyons of the Pánuco River.

The *grandis* major clade of *Xenosaurus* is restricted to the highlands of Veracruz, Puebla, and Oaxaca west of the Isthmus of Tehuantepec. The daughter clades contained within the *grandis* clade have nearly contiguous distributions (Fig. 5). However, the *X. g. grandis* + *X.* sp Vista Hermosa clade is restricted to the humid pine-oak and cloud forests of the Atlantic versant of the Sierras de Zongolica, Sierra Mazateca, and Sierra de Juárez, whereas the earliest branches of its sister clade (e.g., *X. rectocollaris* and *X.* sp Pápalos) are found in the arid pine-oak forests and tropical scrubs in the much drier Tehuacán-Cuicatlán-Quiotepec Depression west of those sierras. This suggests that these different hydric regimes may have contributed to this speciation event. In contrast, the distributions of *X. g. grandis* and *X.* sp Vista Hermosa in the mesic forests of the Sierras de Zongolica and Sierra Mazateca, and Sierra de Juárez, respectively, are separated by the low course of the Papaloapan River in northern Oaxaca, which suggests that, in this clade, speciation may have resulted from the fragmentation of the mesic forests. All of the species in the *agrenon* clade inhabit mostly cloud or pine-oak forests on the highlands of the Sierra Madre del Sur morphotectonic province, and except for *X.* sp Tejocote, which occurs in the Sierra de Cuatro Venados, all of them are distributed on the Sierra Madre del Sur mountain range. This range is roughly oriented from west to east, and it is crossed by several rivers that drain the central valleys to the Pacific Ocean (e.g., the Putla, Verde, Copalita, and Zimatán Rivers), which may have promoted speciation along the Sierra. Also, the most recently derived species in the clade are the ones on the western end of its distribution. This suggests a sequence of events of speciation in the clade accompanied by dispersal to the west, except for *X.* sp Tejocote (see above).

Whereas the *grandis* major clade of *Xenosaurus* is restricted to highlands west of the Isthmus of Tehuantepec, taxa in the *rackhami* clade inhabit relatively cool and humid montane forests at moderately high elevations (>800 m) west, north, and east of the Isthmus (Fig. 5). Thus, the degree at which the Isthmus was responsible for the split of the *grandis* and *rackhami* major clades is uncertain. Nonetheless, *X. grandis sanmartinensis* and *X.* sp Camotlán appear to be isolated from *X. arboreus* and *X. grandis rackhami* and from each other by the lowlands of the Isthmus of Tehuantepec. Cloud forest and, in general, mesic forests on either side of the Isthmus of Tehuantepec enjoyed a much more extensive distribution in the past during colder and wetter times in the Pleistocene (Campbell, 1984; Toledo, 1982). In such periods, cloud forest might have descended low enough to permit cloud-forest taxa to attain more extensive distributions, which subsequently were fragmented along with the cloud forest itself in warmer, drier periods. This fragmentation is potentially responsible for at least some of the speciation events in this clade, as it seems to have been the case in other cloud forest groups such as the genera *Abronia* or *Plectrohyla*. A discussion on the herpetogeography of the region can be found in Campbell (1984).

4.3. Species delimitation

All of the described species of *Xenosaurus* were consistently supported as distinct lineages in all of the analyses. Additionally, in all of the analyses *X. grandis* was found to be composed of at least four allopatric, mutually monophyletic, and divergent lineages, none of which are sister taxa (e.g., *X. g. agrenon*, *X. g. arboreus*, *X. g. grandis*, and the *X. g. rackhami* + *X. g. sanmartinensis* clade). These lineages are also distinct morphologically (King and Thompson, 1968). Thus, there seems to be little question that they each represent a distinct evolutionary species. *Xenosaurus g. rackhami* and *X. g. sanmartinensis* were supported as distinct taxa in some analyses, but not others. However, they were recovered as mutually monophyletic in most of the phylogenetic analyses, and strongly supported by the BPP and BFD* analyses of the *rackhami* clade. Also, *X. g. rackhami* and *X. g. sanmartinensis* are widely allo-

patric (Fig. 5), which makes gene flow between them unlikely. The main tectonic events that formed the Sierra de Los Tuxtlas took place during the late Miocene. However, the latest eruptions of the San Martín Pajapan and Santa Marta volcanoes took place between 2.4 and 1.0 million years ago, while the San Martín volcano erupted for the last time in 1793, destroying the vegetation on its flanks (Martin-Del Pozzo, 1997). We hypothesize that the lack of support for the distinctness of these two taxa in some analyses may reflect the fact that they are recently derived, and argue that they are nonetheless on evolutionary independent trajectories and therefore both should be elevated to species level.

King and Thompson (1968) listed several scalation, morphometric, and color pattern characters that are shared by the subspecies of *X. grandis* and intergrades between them (i.e., a distinctly developed canthus temporalis, a longitudinal row of 3–5 enlarged hexagonal supraoculars that are wider than long, a large V-shaped nape blotch that is attenuate posteriorly, among others) and absent in *X. newmanorum* and *X. platyceps*, the only other species of the genus known at the time. Because of those shared characters, it could be expected that the forms involved in the polytypic *X. grandis* formed a monophyletic group. In our phylogenetic hypothesis, all of these forms were indeed placed within the group composed of the *grandis* + *rackhami* clades, a group that excludes *X. newmanorum*, *X. platyceps*, and their closest relative. However, the subspecies of *X. grandis* did not form a clade. This is likely because the remaining forms in the *grandis* + *rackhami* clades (*X. penai*, *X. phalaroanthereon*, *X. rectocollaris*, *X.* sp Monteverde, and *X.* sp Pápalos) were not known at the time the subspecies of *X. grandis* were described (the 1940–1960s), and thus some or all of the characters shared by the forms in the polytypic *X. grandis* are either simplesiomorphies or homoplasies.

All of the putative undescribed species of *Xenosaurus* also were consistently corroborated as distinct lineages in all of the analyses. This is additionally supported by their small allopatric distributions and preliminary morphological data (Nieto-Montes de Oca, personal observation). Thus, results of our species delimitation analyses suggest that species diversity in *Xenosaurus* is significantly higher than currently estimated, including the 12 currently described species and subspecies and the six putative undescribed species.

Additionally, many of our analyses suggest that there may be additional cryptic species diversity within *X. g. grandis* and *X. rectocollaris*. Because many *Xenosaurus* species are habitat specialists with restricted ranges, gene flow among populations may be low, thereby causing molecular species delimitation methods to oversplit taxa. Therefore, we recommend some of these putative taxa be investigated further using additional taxonomic and genetic sampling that could provide robust estimates of the demographic history of these populations, and identify any morphological divergence that has occurred.

Our study has important implications for the conservation of the species of *Xenosaurus*. By their nature, these species may be particularly susceptible to the impacts of habitat loss and degradation. Although their restricted distributions make habitat conservation planning straightforward in some ways, because most individual populations are highly genetically distinct, each has a unique, important value to the conservation of the species as a whole.

5. Conclusions

By analyzing the data using a variety of phylogenetic models, we were able to reconstruct a well-resolved phylogeny for this group. This phylogeny was generally strongly supported and congruent with the geography. The five subspecies of the traditionally recognized, polytypic species *Xenosaurus grandis* actually represent each a distinct evolutionary species, and the genus *Xenosaurus* appears to be composed of at least 18 species.

Acknowledgments

Many people helped to collect *Xenosaurus* specimens for this study. We thank all of them, and especially W. Schmidt Ballardo, the late F. Mendoza Quijano, and L. Canseco Márquez. We thank J. A. Campbell for the donation of several tissue samples. We thank M. McElroy, R. Harris, and J. Grummer for assistance with molecular lab work. Permits were provided by the Secretaría de Medio Ambiente y Recursos Naturales, Dirección General de Fauna Silvestre to A. Nieto-Montes de Oca (FAUT-0093). Financial support for field work was provided by grants from DGAPA, UNAM (PAPIIT no. IN224009) and CONACYT (no. 154093) to A. Nieto-Montes de Oca. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303. AJB and RCT acknowledge financial support from NSF grants (DEB- 1354506 and DBI-1356796) and a Beckman Postdoctoral Fellowship, and the use of high performance computing resources provided by the University of Hawai'i.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2016.09.001.

References

Ballinger, R.E., Smith, G.R., Lemos-Espinal, J.A., 2000. Xenosaurus. Catal. Am. Amphibians Reptiles 712, 1–3.
Bhullar, B.A.S., 2011. The power and utility of morphological characters in systematics: a fully resolved phylogeny of *Xenosaurus* and its fossil relatives (Squamata: Anguimorpha). Bull. Mus. Compar. Zool. 160, 65–181.
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 Comput. Biol. 10, e1003537. http://dx.doi.org/10.1371/journal.pcbi.1003537.
Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29, 1917–1932.
Campbell, J.A., 1984. A new species of *Abronia* (Sauria: Anguidae) with comments of the herpetogeography of the highlands of southern Mexico. Herpetologica 40, 373–381.
Canseco-Márquez, L., 2005. Filogenia de las lagartijas del género Xenosaurus Peters (Sauria: Xenosauridae) basada en morfología externa Master Thesis. Universidad Nacional Autónoma de México, p. 79p.
Cariou, M., Duret, L., Charlat, S., 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecol. Evol. 3, 846–852.
Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30, 3317–3324.
Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.
Eaton, D.A.R., 2014. PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. Bioinformatics 30, 1844–1849.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res., 1792–1797
Fujisawa, T., Barraclough, T.G., 2013. Delimiting species using single-locus data and the generalized mixed Yule coalescent approach: a revised method and evaluation on simulated data sets. Syst. Biol. 62, 707–724.
Grummer, J.A., Bryson Jr., R.W., Reeder, T.W., 2014. Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). Syst. Biol. 63, 119–133.
Ilut, D.C., Nydam, M.L., Hare, M.P., 2014. Defining loci in restriction-based reduced representation genomic data from nonmodel species: sources of bias and diagnostics for optimal clustering. BioMed Res. Int. 2014, 9p. http://dx.doi.org/10.1155/2014/675158.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.
King, W., Thompson, F.J., 1968. A review of the American lizards of the genus *Xenosaurus* Peters. Bull. Florida State Mus. 12, 93–123.
Kumar, S., Filipski, A.J., Battistuzzi, F.U., Kosakovsky, S.L., Tamura, K., 2012. Statistics and truth in phylogenomics. Mol. Biol. Evol. 29, 457–472.

Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286–2288.

Lartillot, N., Rodrigue, N., Stubbs, D., Richer, J., 2013. A Bayesian Software for Phylogenetic Reconstruction Using Mixture Models. MPI version, p. 22p. Available at <http://megasun.bch.umontreal.ca/People/lartillot/www/pb_mpiManual1.5.pdf>.

Leaché, A.D., Fujita, M.K., 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). Proc. R. Soc. B 277, 3071–3077.

Leaché, A.D., Fujita, M.K., Minini, V.N., Bouckaert, R.R., 2014. Species delimitation using genome-wide SNP data. Syst. Biol. 63, 534–542.

Leaché, A.D., Banbury, B.L., Felsenstein, J., Nieto-Montes de Oca, A., Stamatakis, A., 2015a. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. Syst. Biol. 64, 1032–1047.

Leaché, A.D., Chavez, A.S., Jones, L.N., Grummer, J.A., Gottscho, A.D., Linkem, C.W., 2015b. Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. Genome Biol. Evol. 7, 706–719.

Martin-Del Pozzo, A.L., 1997. Geología. In: González-Soriano, E., Dirzo, R., Vogt, R.C. (Eds.), Historia Natural De Los Tuxtlas. Universidad Nacional Autónoma de México, México, D.F., pp. 25–31.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE). 14 Nov. 2010, New Orleans, LA, pp. 1–8.

Nieto-Montes de Oca, A., Campbell, J.A., Flores-Villela, O.A., 2001. A new species of *Xenosaurus* (Squamata: Xenosauridae) from the Sierra Madre del Sur of Oaxaca, Mexico. Herpetologica 57, 32–47.

Nieto-Montes de Oca, A., García-Vázquez, U.O., Zúñiga-Vega, J.J., Schmidt-Ballardo, W., 2013. A new species of *Xenosaurus* (Squamata: Xenosauridae) from the Sierra Gorda Biosphere Reserve of Querétaro, Mexico. Revista Mexicana de Biodiversidad 84, 485–498.

Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H., 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22, 225–231.

Pérez-Ramos, E., Saldaña de la Riva, L., Campbell, J.A., 2000. A new allopatric species of *Xenosaurus* (Squamata: Xenosauridae) from Guerrero, Mexico. Herpetologica 56, 500–506.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7, 1–11.

Philippe, H., Delsuc, F., Brinkmann, H., Lartillot, N., 2005. Phylogenomics. Ann. Rev. Ecol. Evol. System. 36, 541–562.

Pisani, D., Pett, W., Dohrmann, M., Feuda, R., Rota-Stabelli, O., Philippe, H., Lartillot, N., Wörheide, G., 2015. Genomic data do not support comb jellies as the sister group to all other animals. PNAS 112, 15402–15407.

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D., Vogler, A.P., 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst. Biol. 55, 595–609.

Pyron, A., Burbrink, F.T., Wiens, J.J., 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. BMC Evol. Biol. 2013, 13–93.

Rambaut, A., Suchard, M.A., Xie, W., Drummond, A.J., 2014. Tracer, version 1.6, available at <http://beast.bio.ed.ac.uk/Tracer>.

Rannala, B., Yang, Z., 2013. Improved reversible jump algorithms for Bayesian species delimitation. Genetics 194, 245–253.

Reeder, T.W., Townsend, T.M., Mulcahy, D.G., Noonan, B.P., Wood Jr., P.L., Sites Jr., J. W., Wiens, J.J., 2015. Integrated analyses resolve conflicts over Squamate reptile phylogeny and reveal unexpected placements for fossil taxa. PLoS ONE, 1–22. http://dx.doi.org/10.1371/journal.pone.0118199.

Ronquist, F., Huelsenbeck, J., Teslenko, M., 2011. MrBayes version 3.2 Manual: Tutorials and model summaries. 172 p. Available at <http://mrbayes.sourceforge.net/manual.php>. Last accessed on January 21, 2016.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.

Rubin, B.E.R., Ree, R.H., Moreau, C.S., 2012. Inferring phylogenies from RAD sequence data. PLoS ONE 7, e33394.

Shaffer, H.B. et al., 2013. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. Genome Biol. 14, R28.

Smith, H.M., Iverson, J.B., 1993. A new species of knob-scale lizard (Reptilia: Xenosauridae) from Mexico. Bull. Maryland Herpetol. Soc. 29, 51–66.

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

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4. Springer, Sunderland, MA.

Toledo, V.M., 1982. Pleistocene changes of vegetation in tropical Mexico. In: Prance, G.T. (Ed.), Biological Diversification in the Tropics. Columbia University Press, New York, pp. 93–111.

Townsend, T.M., Larson, A.L.E., Macey, J.R., 2004. Molecular phylogenetics of Squamata: the position of snakes, amphisbaenians, and dibamids, and the root of the squamate tree. Syst. Biol. 53, 735–757.

Wiens, J.J., Brandley, M.C., Reeder, T.W., 2006. Why does a trait evolve multiple times within a clade? Repeated evolution of snakelike form in squamate reptiles. Evolution 60, 123–141.

Wiens, J.J., Kuczynski, C.A., Townsend, T., Reeder, T.W., Mulcahy, D.G., Sites, J.W., 2010. Combining phylogenomics and fossils in higher-level squamate reptile phylogeny: molecular data change the placement of fossil taxa. Syst. Biol. 59, 674–688.

Wiens, J.J., Hutter, C.R., Mulcahy, D.G., Noonan, B.P., Townsend, T.M., Sites, J.W., Reeder, T.W., 2012. Resolving the phylogeny of lizards and snakes (Squamata) with extensive sampling of genes and species. Biol. Lett. 8, 1043–1046.

Whelan, N.V., Kocot, K.M., Moroz, L.L., Halanych, K.M., 2015. Error, signal, and the placement of Ctenophora sister to all other animals. PNAS 112, 5773–5778.

Woolrich-Piña, G.A., Smith, G.R., 2012. A new species of *Xenosaurus* from the Sierra Madre Oriental, Mexico. Herpetologica 68, 551–559.

Yang, Z., 2015. The BPP program for species tree estimation and species delimitation. Curr. Zool. 61, 854–865.

Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. USA 107, 9264–9269.

Yang, Z., Rannala, B., 2014. Unguided species delimitation using DNA sequence data from multiple loci. Mol. Biol. Evol. 31, 3125–3135.

Zamora-Abrego, J.G., 2009. Filogenia molecular de las lagartijas del género Xenosaurus (Xenosauridae) y evolución de sus características de historias de vida Ph.D. Dissertation. Universidad Nacional Autónoma de México, p. 118p.

Zamora-Abrego, J.G., Zúñiga-Vega, J.J., Nieto-Montes de Oca, A., 2007. Variation in reproductive traits within the lizard genus *Xenosaurus*. J. Herpetol. 41, 630–637.

Zhang, J., Kapli, P., Pavlidis, P., Stamatakis, A., 2013. A general species delimitation method with applications to phylogenetic placements. Bioinformatics 29, 2869–2876.