MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 909–925 doi: 10.1111/mec.13076

Pan-African phylogeography of a model organism, the African clawed frog '*Xenopus laevis*'

BENJAMIN L. S. FURMAN,* ADAM J. BEWICK,*[1] TIA L. HARRISON,*[2] ELI GREENBAUM,† VÁCLAV GVOŽDÍK,‡ § CHIFUNDERA KUSAMBA¶ and BEN J. EVANS*

*Biology Department, McMaster University, Life Sciences Building, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1, †Department of Biological Sciences, University of Texas at El Paso, 500 West University Avenue, El Paso, TX 79968, USA, ‡Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic, Kvetna 8, 603 65 Brno, Czech Republic, §Department of Zoology, National Museum, 193 00 Prague, Czech Republic, ¶Laboratoire d'Herpétologie, Département de Biologie, Centre de Recherche en Sciences Naturelles, Lwiro, République Démocratique du Congo

Abstract

The African clawed frog *Xenopus laevis* has a large native distribution over much of sub-Saharan Africa and is a model organism for research, a proposed disease vector, and an invasive species. Despite its prominent role in research and abundance in nature, surprisingly little is known about the phylogeography and evolutionary history of this group. Here, we report an analysis of molecular variation of this clade based on 17 loci (one mitochondrial, 16 nuclear) in up to 159 individuals sampled throughout its native distribution. Phylogenetic relationships among mitochondrial DNA haplotypes were incongruent with those among alleles of the putatively female-specific sex-determining gene *DM-W*, in contrast to the expectation of strict matrilineal inheritance of both loci. Population structure and evolutionarily diverged lineages were evidenced by analyses of molecular variation in these data. These results further contextualize the chronology, and evolutionary relationships within this group, support the recognition of *X. laevis sensu stricto*, *X. petersii*, *X. victorianus* and herein revalidated *X. poweri* as separate species. We also propose that portions of the currently recognized distributions of *X. laevis* (north of the Congo Basin) and *X. petersii* (south of the Congo Basin) be reassigned to *X. poweri*.

Keywords: gene flow, phylogeography, population genetics, species limits

Received 20 June 2014; revision received 6 January 2015; accepted 8 January 2015

Introduction

The African clawed frog *Xenopus laevis*, *sensu lato* (Kobel *et al.* 1996), has an unusual connection with humans, having been used in the early 20th century as a pregnancy assay (Shapiro & Zwarenstein 1934; Weisman & Coates 1941) and more recently as a model organism for research (Cannatella & de Sá 1993; Gurdon 1996; Gurdon & Hopwood 2000). Also called the Common Platanna, the native range of this species spans much of sub-Saharan Africa (Tinsley *et al.* 1996). Established invasive colonies exist in portions of Europe, North America and South America (McCoid & Fritts 1980, 1989; Tinsley & McCoid 1996; Measey & Tinsley 1998; Lobos & KJaksic 2005), and nonpersistent populations have been reported in other localities, including parts of Asia (Measey *et al.* 2012). *Xenopus laevis* has been identified as a potential vector for the amphibian pathogens *Batrachochytrium dendrobatidis* (Weldon *et al.* 2004) and ranavirus (Robert *et al.* 2007), although a causal role between *X. laevis* and the dispersal of these pathogens has not been demonstrated (Measey *et al.* 2012). *Xenopus laevis* has potentially harmful consequences for *Xenopus gilli* (Evans *et al.* 2004, 2005; Evans 2007), which is endangered (South African Frog Re-assessment Group (SA-FRoG) ISASG (2013), through

Correspondence: Ben J. Evans, Fax: 905-522-6066;
E-mail: evansb@mcmaster.ca
[1]Current address: Department of Plant Biology, University of Georgia, Miller Plant Sciences Bldg., Athens, GA 30602, USA
[2]Current address: Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario Canada M5S3B2

ecological competition and hybridization (Tinsley 1981; Simmonds 1985; Picker 1993; Picker *et al.* 1996; Evans *et al.* 1997, 1998; Fogell *et al.* 2013).

Xenopus laevis is generally found in slow moving or stagnant water, and occasionally disperses over land (Measey & Tinsley 1998; Eggert & Fouquet 2006). It is a generalist species that does well in disturbed habitat and has a high capacity to tolerate drought conditions, salinity, starvation, anoxia and temperature fluctuations (reviewed in Measey *et al.* 2012). This species (and other frogs in the Family Pipidae) has unusual adaptations in adults for a mostly aquatic lifestyle, including lateral line sensory organs and a complex communication system involving a unique mechanism of sound production, context- and sex-specific vocalizations and female phonotaxis (Tobias *et al.* 1998, 2004, 2011; Kelley & Tobias 1999).

Tetraploidization, sex determination

An ancestor of *Xenopus laevis* experienced genome duplication during its evolution, probably by allopolyploidization between two diploid ancestors with 18 chromosomes, to create a genome with 36 chromosomes (Tymowska 1991). However, its genome is now functionally diploidized in that, during cell division, each chromosome aligns with only one other homologous chromosome (Tymowska 1991). Tetraploidization in the ancestor of *X. laevis* duplicated essentially all genes in its genome, although many of these duplicates were reduced to a single copy as a result of pseudogenization (Chain & Evans 2006; Morin *et al.* 2006; Hellsten *et al.* 2007; Sémon & Wolfe 2008; Chain *et al.* 2011).

Of particular interest is the chromosome W-linked DM-domain containing gene (*DM-W*), which is present as a single allele in female *X. laevis*, but absent in males, and is the sex-determining gene in this species (Yoshimoto *et al.* 2008). *DM-W* originated by partial segmental duplication of one of the two copies (paralogs) of the double sex- and mab-3-related transcription factor 1 gene (*DMRT1*) that arose when an ancestor of *X. laevis* experienced tetraploidization (Bewick *et al.* 2011). If *DM-W* was strictly maternally inherited over evolutionary time, its evolutionary history is expected to match that of mitochondrial DNA, which is also thought to be maternally inherited and nonrecombining in most species.

Taxonomy and phylogeography of X. laevis sensu lato

Xenopus laevis sensu lato (Kobel *et al.* 1996) comprises three currently recognized species (AmphibiaWeb 2014; Frost 2014): *X. laevis sensu stricto* from southern Africa and a disjunct population north of the Congo Basin (Daudin 1802), *X. petersii* from southern Central Africa (du Bocage 1895) and *X. victorianus* from Eastern Africa (Ahl 1924). A fourth species, *X. poweri*, was described based on specimens from the area of the Victoria Falls (Zambia–Zimbabwe border) by Hewitt (1927) but considered a subspecies of *X. laevis* by some authors (Schmidt *et al.* 1959; Poynton 1964), or a synonym with *X.* (*laevis*) *petersii* (e.g. Parker 1936; Poynton & Broadley 1985). *Xenopus laevis sensu lato* also includes two proposed subspecies: *X. l. bunyoniensis* (Loveridge 1932) and *X. l. sudanensis* (Perret 1966). Additional information on the taxonomic history of this clade is provided in Supplementary Information.

Diversity within *X. laevis sensu lato* has been explored in terms of molecular and morphological variation (Carr *et al.* 1987; Grohovaz *et al.* 1996; Evans *et al.* 1997, 2004; Kobel *et al.* 1998; Measey & Channing 2003; Du Preez *et al.* 2009) and variation in vocalization (Tobias *et al.* 2011). In general, these studies consistently found that populations in different parts of Africa, including populations from different portions of South Africa, are differentiated. The distribution of variation within mitochondrial DNA is perhaps best relayed in terms of four geographical zones of sub-Saharan Africa, which we will refer to as 'southern Africa' (including South Africa and Malawi), East Africa (including Tanzania, Kenya, Uganda, Burundi, Rwanda and the eastern portion of the Democratic Republic of the Congo, 'Central Africa' (including Nigeria, Cameroon, western Zambia and northern Botswana), and 'West Central Africa' (including the southern Republic of Congo, the western portion of the Democratic Republic of the Congo and Angola) (Fig. 1). Evans *et al.* (2004) analysed mitochondrial DNA sequences from *X. laevis sensu lato* from each of these zones. Their analysis recovered paraphyly of the group of mitochondrial DNA sequences from southern Africa with relatively weak support, but recovered strong support for monophyly of the group of mitochondrial DNA sequences from East and Central Africa (Evans *et al.* 2004). Mitochondrial DNA from one sample from the Republic of Congo (West Central Africa) was closely related to a clade containing mitochondrial DNA from Central and East Africa (Evans *et al.* 2004). Within the country of South Africa, Grohovaz *et al.* (1996) and Measey & Channing (2003) found a population of *X. laevis sensu lato* sampled near the town of Niewoudtville to be distinct from populations in other parts of the country. Measey & Channing (2003) also identified a zone of admixture of mitochondrial haplotypes from Niewoudtville and haplotypes from the south-western Cape Region in the vicinity of the town of Vredendal (not sampled in the current study), which is ~100 km south-west of Niewoudtville. Du Preez *et al.* (2009) further identified a second zone of admixture

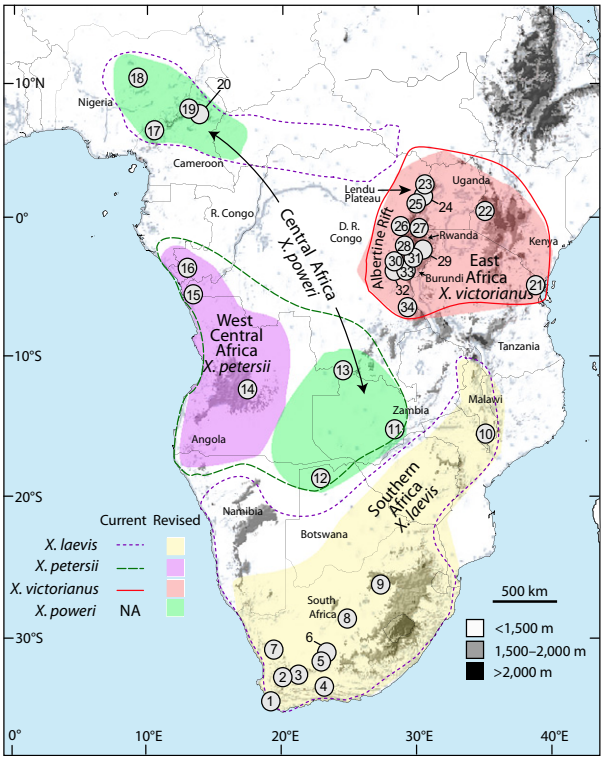

Fig. 1 *Xenopus laevis sensu lato* sampling localities, and currently recognized and revised species ranges. Numbers inside circles indicate locality numbers that correspond with samples listed in Table S1. Unfilled polygons with different lines indicate the currently recognized distributions of *X. laevis*, *X. petersii* (including *X. poweri* as a synonym) and *X. victorianus*, respectively (AmphibiaWeb, 2014; Frost, 2014). Filled polygons indicate four geographical regions that are referred to in the text, each of which corresponds to the distribution of a species (named below the geographical region) that is supported by this study. No locus in this study has data from every sample depicted; mitochondrial DNA has the least missing data (see Table S1 for details). Additional shading refers to meters above sea level as indicated.

near the town of Laingsburg (sampled in the current study), South Africa, between South African populations to the north-east and south-west of this locality based on variation in mitochondrial DNA and two autosomal genes.

The main goal of this study is to further characterize the evolutionary history of *X. laevis sensu lato* in terms of the phylogenetic relationships, divergence times and geographic distributions of diverged evolutionary lineages. We additionally evaluate support for previously proposed species designations within *X. laevis sensu lato* (*X. laevis sensu stricto*, *X. victorianus*, *X. petersi* and *X. poweri*). For evaluating support for the previously proposed species, we adopt the 'General Lineage Concept' (GLC; de Queiroz 1998, 2007) of a species, which defines a species as a 'separately evolving metapopulation lineage' (de Queiroz 2007). The term 'metapopulation' refers to a set of subpopulations that are interconnected by gene flow, and 'lineage' refers to the ancestor–descendant relationship between metapopulations of the same species through time (de Queiroz 2007).

Methods

Samples and molecular data

A total of 183 samples of *X. laevis sensu lato* from 14 countries were used in this study, including 104 samples obtained from South Africa, 37 from Democratic Republic of the Congo (hereafter DRC), 12 from Burundi, 8 from Zambia, 7 from Cameroon, 3 from Nigeria, 3 from Uganda, 2 from Kenya, 2 from Botswana and 1 each from Rwanda, the Republic of Congo, Angola, Malawi and Tanzania (see Table S1, Supporting information for specific locality information). These tissue samples were obtained from field collections, tissue donations from institutional archives (California Academy of Sciences, the Museum of Comparative Zoology at Harvard University, the Natural History Museum of Geneva and the Zoological Research Museum – Alexander Koenig), a collection of live *Xenopus* that was at the University of Geneva, and colleagues (T. Hayes, L. Kalous, R. Tinsley and P. Wagner).

Sequences from a portion of the mitochondrial 12S and 16S rDNA genes and the intervening $tRNA^{Val}$ gene were obtained using primers from Evans *et al.* (2004) for 159 *X. laevis sensu lato* individuals (87% of the samples in this study), with an average of 907 base pairs (bp) per individual (range: 623–2374 bp). Sequences from the female-specific W chromosome gene *DM-W* and flanking regions were obtained using primers detailed in Bewick *et al.* (2011) for 96 female *X. laevis sensu lato* individuals, with an average of 1734 bp per individual (range: 1036–2049 bp). Autosomal DNA sequences were obtained from portions of the protein coding region of 15 loci ranging in length from 341 to 618 bp (Table 1) for 113–136 individuals per locus, using paralog-specific primers detailed in Bewick *et al.* (2011). Sequence data were aligned by eye, and homologies of the aligned characters were unambiguous. Sequences of individual autosomal alleles were inferred using the 'best guess' estimates of allelic states from PHASE v.2.1.1 using default parameters (Stephens *et al.* 2001; Stephens & Donnelly 2003), and both alleles were analysed for the population assignment tests detailed below. DNASP v.5.10.01 (Librado & Rozas 2009) was used to quantify descriptive statistics of the sequence data, and formula 5 of Kimura & Ohta (1972) to calculate 95% confidence intervals (95% CI) for pairwise nucleotide diversity at synonymous sites. All

Table 1 Polymorphism statistics for autosomal loci for *Xenopus laevis*, *X. petersii*, *X. powerii* and *X. victorianus*, including the gene acronym (gene), number of base pairs sequenced (bp), number of alleles sequenced (No. of allelles), number of unique haplotypes (No. of haplotypes), number of synonymous sites (SSites), the number of nonsynonymous sites (NSites), Jukes Cantor corrected pairwise nucleotide diversity for synonymous (πS) and nonsynonmous (πN) sites, the number of segregating synonymous (SS) and nonsynonymous (SN) sites, and Tajima's D based on synonymous sites (D_S), with * indicating significant departure from zero. For some loci, NA indicates that Tajima's D could not be calculated due to insufficient molecular diversity or data

Gene	bp	No. of alleles	No. of haplotypes	SSites	NSites	πS	πN	SS	SN	D_S
Southern Africa (*X. laevis*)										
AR	339	160	9	83	256	0.0015	0.0035	3	6	−1.28
prmt6	612	170	61	146	466	0.0237	0.0033	17	10	0.36
mogA	619	176	26	155	463	0.0031	0.0037	3	17	−0.13
c7orf25	531	184	16	119	412	0.0127	0.0005	11	4	−0.53
nfil3	534	188	21	120	415	0.0172	0.0016	12	7	−0.06
pigo	494	184	28	127	365	0.0230	0.0029	15	12	0.25
Sugp2	438	182	19	108	330	0.0103	0.0027	11	9	−1.16
mastl	537	184	46	119	418	0.0126	0.0073	11	32	−0.53
zbed4	471	170	17	110	361	0.0100	0.0029	9	9	−0.72
Rassf10	486	186	36	98	388	0.0307	0.0070	19	11	−0.29
p7e4	522	186	30	121	401	0.0274	0.0009	16	5	0.45
fem1c	474	146	24	107	367	0.0192	0.0010	17	21	−0.94
znf238.2	531	186	24	117	410	0.0030	0.0073	4	17	0.17
bcl9	489	188	18	114	372	0.0123	0.0010	10	10	−0.45
nufip2	473	156	24	105	363	0.0148	0.0019	27	29	1.97*
West Central Africa (*X. petersii*)										
AR	339	8	6	83	256	0.0065	0.0058	1	4	1.17
prmt6	612	2	2	145	467	0.0210	0.0021	3	1	NA
mogA	619	10	7	155	463	0.0000	0.0139	0	19	NA
c7orf25	531	10	9	118	413	0.0549	0.0099	18	11	−0.10
nfil3	534	10	6	120	414	0.0211	0.0013	6	1	0.72
pigo	494	10	5	127	365	0.0167	0.0006	6	1	−0.06
Sugp2	438	10	4	108	330	0.0052	0.0026	2	3	−0.69
mastl	537	10	4	119	418	0.0056	0.0018	2	3	−0.18
zbed4	471	8	2	109	357	0.0023	0.0000	1	0	−1.05
Rassf10	486	8	4	98	388	0.0118	0.0018	2	2	1.80
p7e4	522	10	3	122	400	0.0089	0.0000	3	0	0.02
fem1c	474	8	3	107	367	0.0175	0.0000	4	0	0.79
znf238.2	531	10	4	117	414	0.0052	0.0027	3	3	−1.56
bcl9	489	10	3	114	372	0.0000	0.0022	0	2	NA
nufip2	473	10	4	106	365	0.0070	0.0018	2	2	0.12
Central Africa (*X. poweri*)										
AR	339	26	5	83	256	0.0071	0.0020	4	2	−1.20
prmt6	612	22	12	145	467	0.0155	0.0017	8	6	0.03
mogA	619	24	8	155	463	0.0000	0.0036	0	10	NA
c7orf25	531	26	7	118	413	0.0043	0.0023	3	3	−0.89
nfil3	534	26	16	120	414	0.0130	0.0023	5	8	0.49
pigo	494	26	7	127	365	0.0087	0.0014	5	4	−0.48
Sugp2	438	24	5	109	329	0.0000	0.0017	0	4	NA
mastl	537	18	14	119	418	0.1052	0.0194	33	27	0.90
zbed4	471	28	7	109	355	0.0095	0.0042	3	6	0.80
Rassf10	486	18	7	95	376	0.0126	0.0021	3	3	0.99
p7e4	522	26	4	121	401	0.0028	0.0005	3	2	−1.29
fem1c	474	26	16	107	367	0.0312	0.0004	13	2	−0.10
znf238.2	531	26	9	118	413	0.0102	0.0033	4	7	0.36
bcl9	489	26	4	114	372	0.0059	0.0016	3	2	−0.36
nufip2	473	22	10	106	365	0.0135	0.0036	7	5	−0.86

Table 1 *Continued*

Gene	bp	No. of alleles	No. of haplotypes	SSites	NSites	πS	πN	SS	SN	D_S
East Africa (*X. victorianus*)										
AR	339	40	7	84	255	0.0007	0.0034	2	4	0.38
prmt6	612	38	10	145	467	0.0104	0.0001	8	1	−0.63
mogA	619	42	10	155	463	0.0009	0.0054	2	9	−1.30
c7orf25	531	40	5	117	413	0.0053	0.0009	3	2	−0.28
nfil3	534	40	9	120	414	0.0052	0.0016	3	5	−0.27
pigo	494	34	7	127	365	0.0052	0.0006	5	2	−1.20
Sugp2	438	42	4	108	330	0.0009	0.0004	2	1	−1.50
mastl	537	34	16	119	418	0.0173	0.0043	11	12	−0.78
zbed4	471	32	10	109	353	0.0039	0.0041	4	8	−1.50
Rassf10	486	42	15	99	387	0.0100	0.0058	4	13	0.12
p7e4	522	40	6	121	401	0.0090	0.0000	6	0	−0.62
fem1c	474	40	11	107	367	0.0137	0.0009	6	4	0.05
znf238.2	531	38	10	117	414	0.0057	0.0034	4	7	−0.73
bcl9	489	40	11	113	370	0.0157	0.0026	7	6	0.18
nufip2	473	38	16	106	365	0.0192	0.0021	11	10	−0.67

new sequence data are deposited in GenBank (Accession nos. KP343951–KP345838), and Accession nos. of other data in these analyses are listed in previous studies (Evans *et al.* 2004, 2005, 2008, 2011; Evans 2007; Bewick *et al.* 2011).

Phylogenetic analyses

We used BEAST v.1.6 (Drummond & Rambaut 2007) to generate time calibrated trees for the mitochondrial and for the *DM-W* data. For each locus, we performed four independent runs, 50 million generations each, using a strict clock. Previously published orthologous sequences from *X. gilli* were used as outgroups. The timing of divergence of *X. laevis* and *X. gilli* was set to 16.7 million years ago (Ma) with a standard deviation of 3.62 Ma (Evans *et al.* 2004), to calibrate these analyses. This divergence time is based on the assumption that the separation of the South Atlantic Ocean triggered the diversification of South American from African pipid frogs ~100 Ma (Pitman *et al.* 1993; Maisey 2000; McLoughlin 2001; Sereno *et al.* 2004; Ali & Aitchison 2008) and was based on analysis of data from mitochondrial DNA (Evans *et al.* 2004). We tested for convergence of the MCMC chains on the posterior distribution by calculating effective sample sizes (ESSs) of post-burn-in likelihoods using TRACER v.1.5 (Rambaut & Drummond 2007), and inspecting traces of parameter estimates. This led us to discard a burn-in of 25% of the generations from each analysis. For each analysis, the model of evolution was selected by the program MrModeltest version 2 (Nylander 2004) based on the Akaike information criterion. The preferred model for the mitochondrial DNA analysis was the general time reversible model (Tavaré 1986), with a proportion of invariant sites, a gamma-distributed heterogeneity in the rate of evolution and estimated base frequencies (GTR+I+Γ+bf). For the *DM-W* data set, the preferred model was the Hasegawa, Kisino and Yano model (Hasegawa *et al.* 1985), with a proportion of invariant sites and estimated base frequencies (HKY+I+bf).

To examine evolutionary relationships among the autosomal genes, three approaches were taken. First, phylogenetic networks were generated among phased autosomal alleles from each locus using SPLITSTREE v.4.13.1 (Huson & Bryant 2006). We used Jukes–Cantor corrected distances between alleles and the Neighbor-Net algorithm (Bryant & Moulton 2004). Support for the splits in the networks was determined with a bootstrap analysis with 1000 replicates. Second, we performed a phylogenetic analysis on concatenated autosomal data using BEAST version 1.7.4 (Drummond *et al.* 2012) including individuals with less than 50% missing data (i.e. the same individuals that were included in the population assignment analyses described below). And third, we estimated a species tree using *BEAST with a reduced data set of 70 individuals and 10 genes that minimized missing data across individuals and loci. For the phylogenetic analyses of concatenated autosomal data, we used a model of evolution selected by MrModeltest, with the same calibration procedure as detailed above for mitochondrial DNA and *DM-W*, and we performed four independent runs for 20 million generations each. For the *BEAST analysis, we assumed a strict molecular clock with an exponentially distributed mutation rate with a mean of 4.7×10^{-10} substitutions/site/

generation following Bewick *et al.* (2012). This mutation rate estimate is based on a multilocus analysis of data from >100 genes from pipid frogs and relied on the same assumption about the geological trigger for diversification of pipid frogs as the mitochondrial DNA analysis above. To achieve convergence, it was necessary to use a simpler model of evolution than that recommended by MrModeltest for the concatenated data set (we used HKY+Γ+bf instead of GTR+I+Γ+bf). For the *BEAST analysis, we linked the model of evolution across all data partitions, and unlinked the phylogeny of each partition. *A priori* species designations were based on eight clades that were observed in the concatenated analysis of autosomal DNA, including: (i) Nigeria and Cameroon, (ii) Botswana and Zambia, (iii) Angola and western DRC, (iv) eastern DRC, Uganda and Burundi, (v) Malawi, and the South African localities Kimberly and Victoria West, (vi) the South African locality Niewoudtville, (vii) the South African localities Betty's Bay, Garden Route National Park and some individuals from Laingsburg and (viii) the South African localities Beaufort West, and other individuals from Laingsburg. The *BEAST analysis was performed with four independent runs, each for 100 million generations. For BEAST and *BEAST analyses, orthologs from *X. gilli* were used as the outgroup, and independent runs were combined using LOGCOMBINER version 1.7.4 (Drummond *et al.* 2012). Similar to the other phylogenetic analyses, 25% of the run was discarded as burn-in, and convergence was assessed based on ESSs of the parameters as calculated by TRACER version 1.6 (Drummond & Rambaut 2007). A maximum clade credibility tree with median node heights was constructed using TREEANNOTATOR version 1.7.4 (Drummond *et al.* 2012).

BP&P analysis

We used BP&P version 2.2 (Yang & Rannala 2010) to test for evidence of species limits. This analysis uses molecular data and a 'guide' phylogeny, which is a hypothesized relationship among populations or species, to evaluate the posterior probability of a species tree. The species tree is assumed to either be the same as the guide tree, or alternatively to be a simplified version of the guide tree that can be obtained by collapsing one or more nodes. We used a guide tree based on clusters obtained from the phylogenetic analysis of concatenated autosomal DNA, and included only the autosomal DNA sequences that were analysed in the *BEAST analyses detailed above. BP&P has two different reversible jump proposal algorithms for species delimitation that influence the probability that nodes within the guide tree are expanded or collapsed during the Markov Chain (Yang & Rannala 2010). We ran two independent chains for each algorithm using a gamma prior G (2, 1000) for both the population size and tree root age priors, with automatic adjustments of step lengths in the MCMC algorithm made by the program. In addition, we explored an alternative prior for both of these parameters in which we calculated a scale parameter (β) for the gamma distribution, by dividing 1 (a diffuse value for the shape parameter (α) of the gamma distribution) by the mutation rate used in the *BEAST analysis (Bewick *et al.* 2012) multiplied by an estimated divergence time from the outgroup taxon *X. gilli* of 16.7 Ma (Evans *et al.* 2004), which resulted in a value of 126. We then ran two independent chains for both algorithms with this new gamma prior distribution G (1, 126) for both the ancestral population size and the tree root age. For each prior setting, the MCMC was run for 100 000 generations, and 20 000 generations were discarded as burn-in, based on visual inspection of the posterior distribution of likelihoods.

Population assignment

The phylogenetic analyses detailed above evaluate evolutionary relationships in the context of a bifurcating phylogeny. However, autosomal DNA relationships may reticulate or be inconsistent among loci as a result of gene flow, lineage sorting and recombination, and this is a particular concern when analysing intraspecific samples. More specifically, use of a phylogeny estimated from autosomal DNA to guide the *BEAST and BP&P analyses comes with the caveat that we did not explore all possible groupings or (for BP&P) all possible relationships among these groups, and therefore the results are contingent on the *a priori* groups and guide tree that we used for *BEAST and BP&P, respectively. Population assignment tests (and also the SplitsTree analysis discussed above) therefore offer a complementary perspective on the nature of multilocus molecular variation among taxa because they do not interpret evolutionary relationships in the context of a bifurcating tree. To assess the degree of population structure and assign individual genotypes to putative populations, we used the programs TESS v.2.3 (Chen *et al.* 2007) and STRUCTURE v.2.3 (Pritchard *et al.* 2000). Both approaches estimate the probability that each individual is assigned to *K* populations, with an aim of minimizing Hardy–Weinberg and linkage disequilibria within the populations (Pritchard *et al.* 2000; Chen *et al.* 2007; François & Durand 2010). Unlike Structure, Tess incorporates spatial information on geographic distances between sampling points (based on GPS coordinates) into the prior distribution when

calculating individual assignment probabilities (Chen *et al.* 2007; François & Durand 2010).

For Tess and Structure analyses, we excluded data from individuals with missing data from more than half of the loci. Data from mitochondrial DNA and *DM-W* were also excluded so that these analyses would provide a perspective on diversification independent from the analyses of the maternally inherited loci. The 135 individuals in the analysis had an average of 6.3% of the loci with missing data. We ran Tess for 100 000 generations with a burn-in of 10 000, using the conditional auto-regressive admixture model (Durand *et al.* 2009), starting from a neighbour-joining tree and using 10 iterations for each value of K ranging from 2 to 10. Because some individuals were sampled from the same location, we used the 'generate spatial coordinates for individuals' option in Tess, with a standard deviation equal to 1.0. Convergence was based on inspection of post-run log-likelihood plots, and support for alternative K values was assessed by inspection of the deviance information criterion (DIC) (Spiegelhalter *et al.* 2002); models with lower DIC values are preferred.

For Structure analysis, we ran 20 million generations with a burn-in of 2.5 million generations for values of K equal to 2–10, with five iterations for each value of K. We specified the 'admixture model' (Falush *et al.* 2003) and assumed no correlation between alleles. The post-run likelihood values were stable and support for alternative K values was evaluated using the ΔK statistic (Evanno *et al.* 2005), as calculated with STRUCTURE HARVESTER WEB v.0.6.93 (Earl & von Holdt 2012), and the *ad hoc* method outlined in Pritchard *et al.* (2000).

The samples used in the population assignment analyses comprise more from South Africa ($n = 107$) than from other portions of the distribution of *X. laevis sensu lato* that are not from South Africa ($n = 41$). To examine whether this uneven geographic sampling affected our results, we reran the Tess analysis with a random subsample of only five individuals from each sampling locality in South Africa, plus all other samples from other countries. This reduced data set had a total of 66 individuals, of which 25 were from South Africa. We ran this analysis for 100 000 generations with 10 000 discarded as burn-in, for K values ranging from 2 to 10, with 10 iterations for each value. Other analytical details were identical to those discussed above.

Tess and Structure runs were post-processed using CLUMPP v.1.1.2 (Jakobsson & Rosenberg 2007), which averages assignment probabilities across iterations. Clumpp offers three separate algorithms that maximize similarity across all of the iterations of a given K; we selected an algorithm as recommended in the program documentation (Jakobsson & Rosenberg 2007).

Results

Phylogenetic incongruence between maternally inherited loci

Estimated phylogenetic relationships from mitochondrial DNA and from *DM-W* each resolve sequences into geographically clustered clades that correspond with one another, and both analyses recover strong and congruent support for paraphyly of the group of haplotypes from individuals in one pond in Laingsburg, South Africa. However, there are strongly supported inconsistencies in the estimated relationships among these clades (Fig. 2; see Fig 1 and insert in Fig. 3 for sampling locations). The mitochondrial DNA phylogeny supports monophyly of the group of sequences from the following South African localities: Niewoudtville, Beaufort West, Laingsburg, De Doorns, Betty's Bay, GRNP, Hoekwil and Cape Town, whereas the *DM-W* phylogeny supports paraphyly of this group of sequences (Fig. 2). Another difference with strong statistical support is seen in relationships among samples from East Africa. In the mitochondrial DNA phylogeny, all East Africa sequences that are not from or near the Lendu Plateau form a strongly supported clade. But in the *DM-W* phylogeny, this group of sequences is inferred to be paraphyletic. Strongly supported inconsistent relationships were also inferred when we restricted both analyses to include only those individuals for whom data were collected from both loci (data not shown).

Molecular variation, evolutionary relationships and species delimitation using autosomal loci

Table 1 presents polymorphism statistics for four diverged lineages of *X. laevis sensu lato* that correspond to previously proposed species within this group as redefined below. All of the loci were polymorphic within *X. laevis sensu lato*. One locus exhibited a Tajima's (1989) D value that was significantly greater than zero within a geographical region depicted in Fig. 1, an observation that could reflect a signature of balancing selection. After weighting individual locus values by the number of synonymous sites at each locus, the largest average pairwise diversity of synonymous sites was similar for individuals from southern Africa (0.0148; 95% CI: 0.0091–0.0206), Central Africa (0.0159; 0.0100–0.0218) and West Central Africa (0.0124; 0.0072–0.0176), but about half as large for individuals from East Africa (0.0081; 0.0039–0.0124).

Phylogenetic analysis of concatenated autosomal data from 135 *X. laevis sensu lato* individuals provided strong support for multiple diverged evolutionary lineages

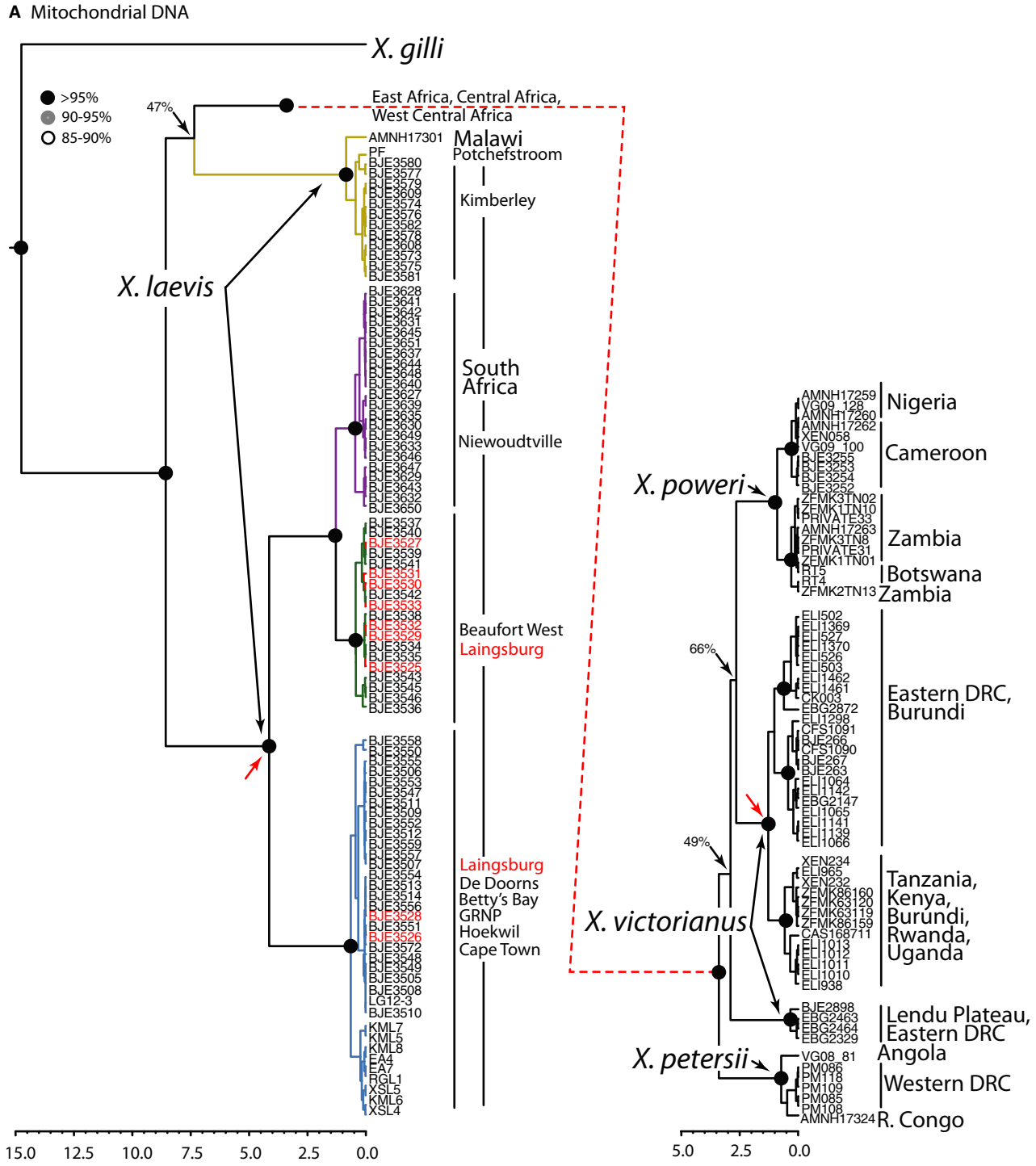

Fig. 2 Chronogram among (A) mitochondrial DNA and (B) *DM–W* haplotypes in *X. laevis sensu lato*. Shaded dots over nodes indicate posterior probabilities, expressed as percentages as indicated; some terminal support values were omitted for clarity, and the posterior probabilities of various poorly supported nodes are indicated. The scale bar indicates divergence time from the present in millions of years. With the exception of the sample from Malawi, shaded branches in southern Africa correspond to sampling localities depicted in Fig. 3. Small arrows indicate relationships that are well supported in each phylogeny but discordant between them.

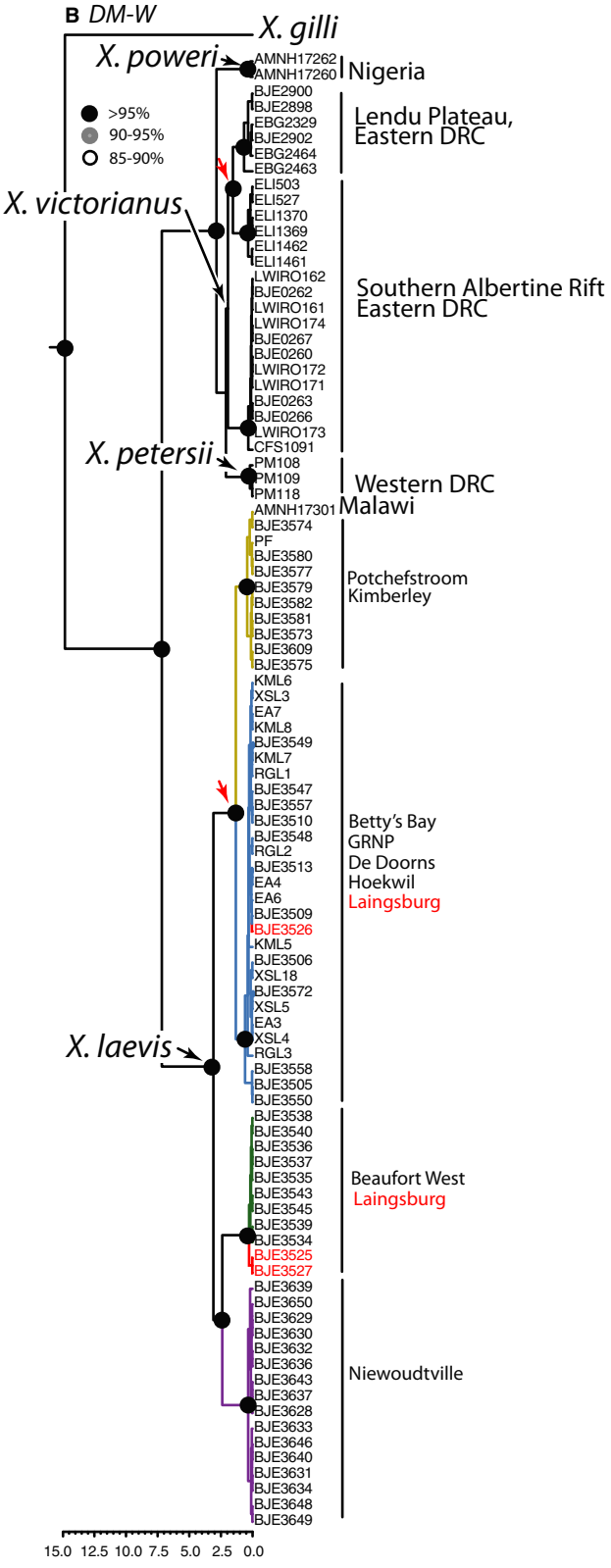

Fig. 2 Continued.

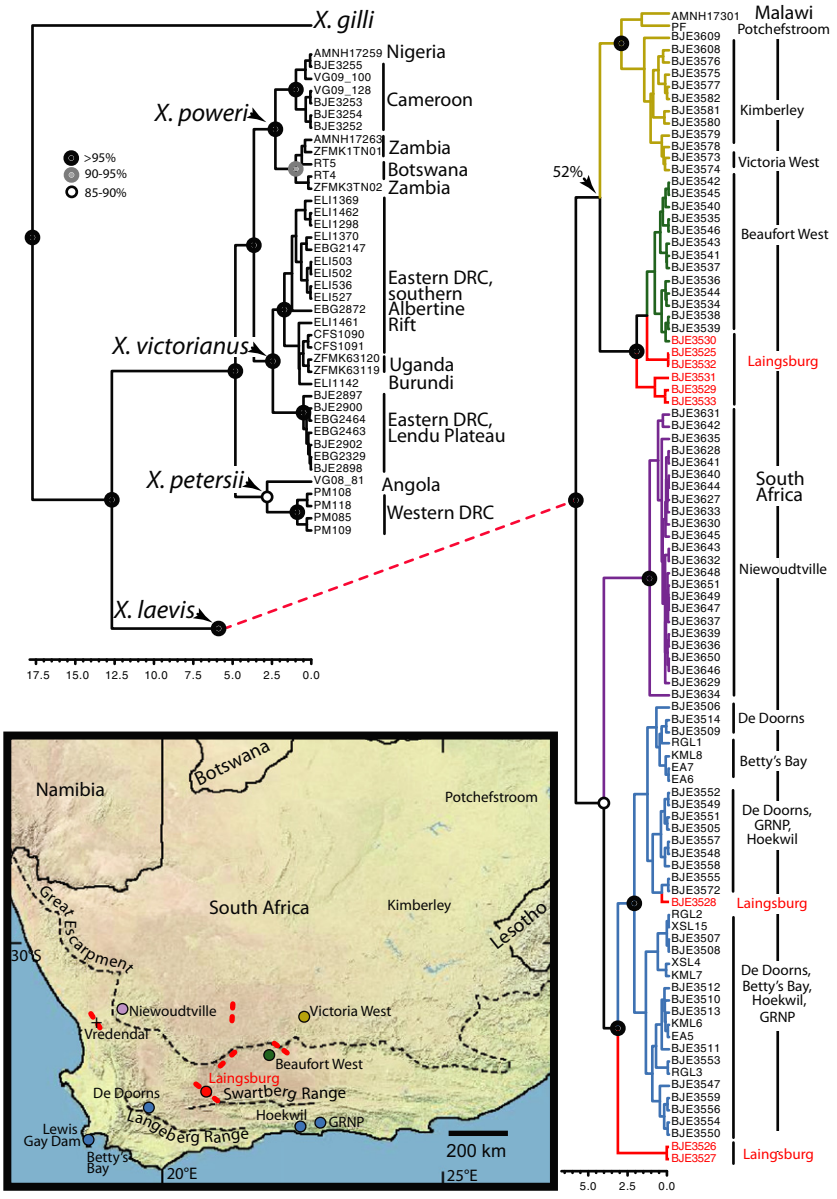

Fig. 3 Phylogenetic analysis of concatenated data from up to 15 autosomal loci per individual. A map shows sampling localities with dots, a plus sign indicates a zone of admixture in Vredendal between mitochondrial DNA lineages from Niewoudtville and the south-western Western Cape Province identified by Measey & Channing (2003), and short dotted lines indicate the approximate locations of confirmed or hypothesized contact zones between *X. laevis* populations. Long dotted lines indicate major geological formations. The scale bars indicates divergence time from the present in millions of years.

(Fig. 3), many of which correspond to those identified in the analyses of mitochondrial DNA and *DM-W*. Diverged lineages in the analysis of concatenated autosomal data include (i) individuals from southern Africa (South Africa and Malawi), (ii) individuals from East Africa (Uganda, Burundi, eastern DRC), (iii) individuals from Central Africa (Nigeria, Cameroon, Zambia, Botswana) and (iv) individuals from West Central Africa (Angola and western DRC). The topology of relationships among the geographically clustered clades was more similar to that inferred from *DM-W* than mitochondrial DNA in the sense that the group of samples from Malawi and South Africa are inferred to be monophyletic. However, it was more similar to the mitochondrial DNA phylogeny than the *DM-W* phylogeny in that the group of samples from East Africa that were not from or near the Lendu Plateau were inferred to be monophyletic. The analysis of concatenated autosomal data differs from the analyses of mitochondrial DNA and *DM-W* in that the former supports monophyly of the group of samples from East Africa (Figs 2 and 3). Both maternally inherited loci supported a close relationship between haplotypes from Niewoudtville, South Africa and those from Beaufort West, South Africa and a few from Laingsburg, South Africa, but this relationship was not observed in the analysis of concatenated autosomal loci.

Geographical clustering of variation was observed in the concatenated analysis. Within Central Africa, for example, samples from Nigeria and Cameroon form a

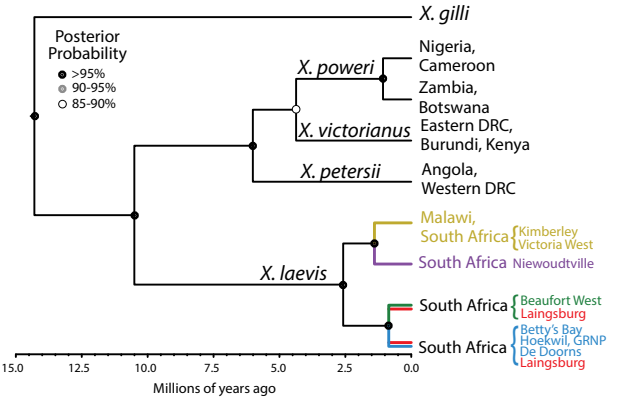

Fig. 4 Species tree analysis by *BEAST. In southern Africa, shaded branches correspond to the sampling localities depicted in Fig. 3.

clade that is most closely related to a clade comprising samples from Botswana and Zambia. Within East Africa, samples from or near the Lendu Plateau form a clade that is most closely related to a clade containing other samples from the rest of the Albertine Rift and samples from Uganda and Burundi. Within South Africa, geographically structured clades were recovered from multiple regions, including (i) samples from Malawi and northern South Africa (Potchefstroom, Kimberley, Victoria West) and (ii) samples from Niewoudtville, South Africa, (iii) samples from Beaufort West and some samples from Laingsburg and (iv) other samples from Laingsburg plus samples from southwestern Western Cape Province.

Species tree analyses with *BEAST (Heled & Drummond 2010) supported the same relationships among most clusters of sequences as the concatenated analysis (Fig. 4). The exception to this is that the species tree analysis infers a monophyletic relationship between the two populations that included individuals from the admixed population in Laingsburg, whereas the concatenated analysis supported a paraphyletic relationship between these two populations with respect to other populations (Figs 3 and 4).

Similar to the phylogenetic analyses discussed above, the network analysis of phased *X. laevis* biparentally inherited alleles reveals strong geographic association of molecular variation. Twelve of 15 networks placed molecular variation from southern Africa and the rest of sub-Saharan Africa on distinct portions of the network (Fig. S1, Supporting information). Variation in East Africa also tended to cluster in portions of these networks that were distinct from variation in other parts of Central Africa or West Central Africa.

Using the species delimitation program BP&P, we recovered strong support (a posterior probably of ~1) for separate species statuses for all of the clusters in the guide tree, with each cluster corresponding to the terminal taxa presented in Fig. 4 for the *BEAST analysis. Results were consistent for both species delimitation algorithms, and for both prior settings that we tried.

Population assignment

Tess and Structure population assignment analyses recovered similar results and support the existence of substantial population structure in *X. laevis sensu lato* (Fig. 5, Figs S2 and S3, Supporting information). The DIC plots suggest that 6–7 populations are preferred by the Tess analysis and the method of Evanno *et al.* (2005) supports 5 populations in the Structure analysis. As a consequence of isolation by distance (identified using Partial Mantel tests, data not shown), we expected the *ad hoc* method of Pritchard *et al.* (2000) to deliver an overestimation of the number of clusters due to departure of the observed data from a model of multiple panmictic populations (Pritchard *et al.* 2000). As expected, this method supported the maximum number of clusters we tested ($K = 10$, $P < 0.001$).

Individuals assigned to each cluster were nearly identical in both analyses at most values of K. Clusters identified by Tess and Structure at higher values of K corresponded to clades identified in the phylogenetic analyses, and to the species identified by BP&P analysis. Similar to the phylogenetic analyses, these assignment tests also highlight genetic uniqueness of the *X. laevis sensu lato* population from or near the Lendu Plateau and that from Niewoudtville, and also distinguish populations in the northern and southern portions of West Central Africa, the former of which corresponds to a proposed subspecies *X. l. sudanensis* (Perret 1966).

Discussion

Phylogenetic incongruence among maternally inherited loci

We observed well supported, discordant relationships among lineages of two putatively maternally inherited genomic regions in the frog *X. laevis sensu lato*: mitochondrial DNA and the female-specific gene *DM-W* (Yoshimoto *et al.* 2008). This observation could reflect error in phylogenetic inference (that is, an incorrect phylogeny may have been inferred in one or both loci) or it could be a 'real' (biological) difference. Missing data, long-branch attraction and model misspecification, for example, may affect phylogenetic inference (Lemmon & Moriarty 2004; Kück *et al.* 2012; Roure *et al.* 2013). A biological difference in these phylogenies could arise if either of these markers was not strictly

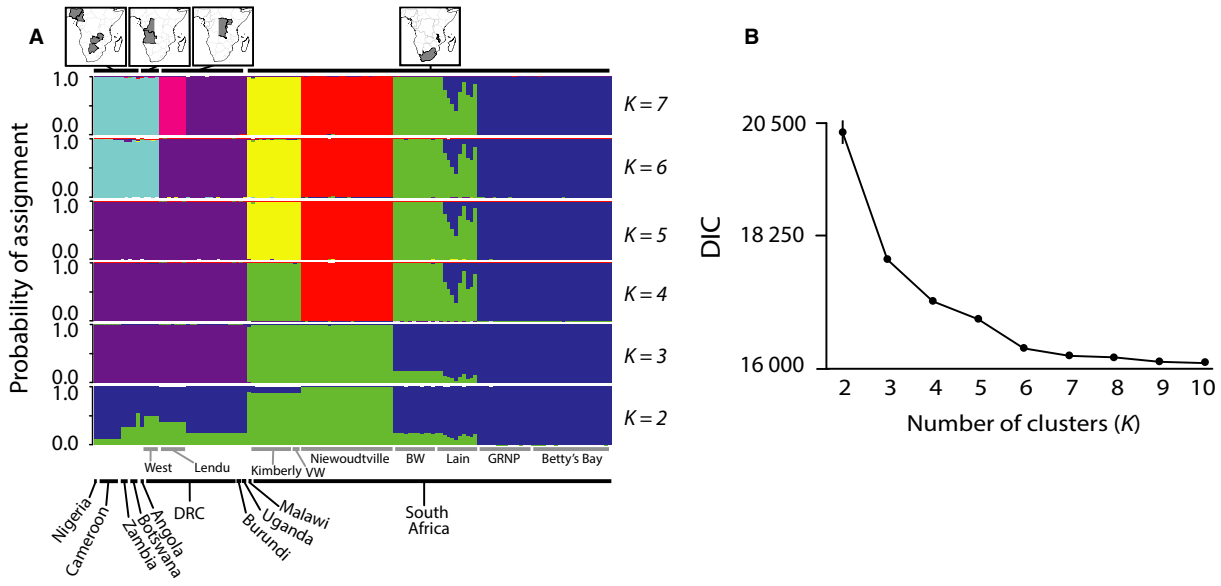

Fig. 5 (A) Results of TESS analysis with population clusters (*K*) ranging from 2–7. (B) The deviance information criterion (DIC) for each value of *K*, with bars (most not visible) indicating the standard deviation of this estimate across iterations. In (A), two localities are labelled within the Democratic Republic of the Congo (DRC) including the western DRC (West DRC) and a region including or near the Lendu Plateau (Lendu). Localities in South Africa include Kimberley, Victoria West (VW), Niewoudtville, Beaufort West (BW), Laingsburg (Lain), Garden Route National Park (GRNP) and Betty's Bay.

maternally inherited, or if either experienced recombination. If individuals carrying *DM-W* occasionally developed as phenotypic males, for instance, this could lead to a mode of inheritance that is not strictly maternal. Periodic phenotypic sex reversal coupled with sex-specific rates of recombination (specifically, a lower recombination in the heterogametic sex) has been proposed as a mechanism for maintaining nondiverged (homomorphic) sex chromosomes in other frogs (Perrin 2009; Stöck *et al*. 2011), and indeed, *X. laevis* has homomorphic sex chromosomes (Tymowska 1991). Further information on rates of recombination in male and female *X. laevis* would be useful to evaluate the applicability of this hypothesis to *X. laevis*. It is also possible that *DM-W* or its flanking region exist in duplicate copies in some females, and that these copies could occasionally undergo ectopic recombination events, even if this locus were strictly maternally inherited. While evidence for recombination in mitochondrial DNA has been reported in various taxa (Piganeau *et al*. 2004; Tsaousis *et al*. 2005), many statistical approaches to detect recombination are prone to false positives (Innan & Nordborg 2002; Galtier *et al*. 2006; Sun *et al*. 2011), and we view this as an unlikely explanation for our observations.

Statuses of previously proposed species

Our results provide novel perspectives on the evolutionary history of *X. laevis sensu lato*, and argue for taxonomic revision from the standpoint of the GLC (de Queiroz 1998, 2007). Within *X. laevis sensu lato*, almost all of our analyses recovered support for at least four evolutionarily diverged lineages in the following geographical regions: (i) southern Africa, including Malawi and South Africa, (ii) Central Africa, including Nigeria, Cameroon, Zambia and Botswana and (iii) West Central Africa, including the Republic of Congo, western DRC and Angola, and (iv) East Africa, including Kenya, Uganda, Rwanda, Burundi, eastern DRC and Tanzania. Each of these groups was identified as a differentiated cluster in the population assignment tests, and lineages i–iii were recovered in each of the phylogenetic analyses we performed (mitochondrial DNA, *DM-W* and concatenated and species tree analysis of autosomal DNA). Lineage iv formed a clade in the phylogenetic analyses of autosomal DNA and *DM-W*, but not in the analysis of mitochondrial DNA. These four lineages, respectively, correspond to four currently or previously recognized species: *X. laevis*, *X. poweri*, *X. petersii* and *X. victorianus*, but we argue for a revised distribution for two of them (*X. laevis* and *X. poweri*). A revision of the distributions of *X. laevis* and *X. poweri* is warranted because individuals from the north of the Congo Basin (Cameroon, Nigeria) are more closely related to individuals from the south of the Congo Basin (Zambia, Botswana) than they are to individuals from other parts of Africa, including southern Africa, which is where *X. laevis* occurs. Thus, we reassign the population of *X. laevis sensu lato* from Nigeria and

Cameroon to *X. poweri* instead of *X. laevis*. We note that a subspecies of *X. laevis*, *X. l. sudanensis*, from the Adamawa Region in Cameroon was described by Perret (1966). Our data potentially support the transfer of *X. l. sudanensis* to the synonymy of *X. poweri* instead of *X. laevis*, although additional data from the type localities or examination of the type specimens is needed. Similarly, another subspecies of *X. laevis*, *X. l. bunyoniensis* (Loveridge 1932), should be tentatively considered a synonym of *X. victorianus*, as evidenced by the inferred phylogeography of *X. laevis sensu lato* and by phylogenetic position of our sample from south-western Uganda. Although again we note that this study lacks samples directly from the type locality of *X. l. bunyoniensis*, which should be investigated in the future. Under our proposed taxonomy, relationships among mitochondrial DNA variants of *X. laevis*, and *X. victorianus* may be paraphyletic within each species; we note also that monophyly is not a requirement of the GLC (de Queiroz 2007).

Although the question of whether further taxonomic division is warranted is beyond the scope of this study, we do note that genetic variation within *X. laevis*, *X. victorianus*, and *X. poweri* is substantial. Within *X. laevis*, differentiated populations were identified in the following regions: (i) south-western Western Cape Province, (ii) Niewoudtville, (iii) Kimberley, Victoria West and Malawi. The south-western Western Cape Province lineage is comprised of two geographically clustered demes with admixture detected at the location of Laingsburg. Individuals from south-western and north-eastern South Africa also differ in body size and in the frequency of naturally occurring testicular oocytes (Du Preez *et al.* 2009). Within *X. poweri*, a population from Cameroon and Nigeria is differentiated from a population from Botswana and Zambia. Clades within *X. laevis* and within *X. poweri* were delimited from one another by the species delimitation program BP&P. However, significant evidence was recovered for isolation by distance using a partial Mantel test (data not shown), and these data therefore violate an assumption (panmixia) of the BP&P analysis. Within *X. victorianus*, the population from or near the Lendu Plateau is differentiated from other populations. The finding of substantial genetic differentiation in these species supports the point made by Du Preez *et al.* (2009) that the geographic provenance of experimental animals is an important experimental variable because of among-population variation in genetic backgrounds.

Our results, which include some of the individuals from Central Africa studied by Evans *et al.* (2004) and Du Preez *et al.* (2009), but a different suite of individuals sampled in South Africa from Du Preez *et al.* (2009), are consistent with the findings from these and other studies (Grohovaz *et al.* 1996; Kobel *et al.* 1998; Measey & Channing 2003). Similar to Du Preez *et al.* (2009), we found evidence for extensive introgression between populations south-west and north-east of the locality of Laingsburg. This was evinced by (i) individuals from this locality having a diversity of evolutionary affinities in the mitochondrial DNA, *DM-W* and concatenated analysis of autosomal DNA and (ii) admixed population affinities that were identified by population assignment tests. We did not recover qualitative evidence for extensive gene flow between other populations of *X. laevis sensu lato* based on the population assignment tests. One possibility is that this could be an artefact of missing genetic information from animals in the contact zones between these lineages, for example in the Congo Basin and south of the Congo Basin, or between differentiated populations in South Africa (Fig. 1). Reciprocal crosses between *X. laevis sensu lato* individuals that were probably from South Africa, and individuals from Uganda or Botswana both produced fertile offspring of both sexes (Blackler *et al.* 1965; Blackler & Fischberg 1968). Thus, gene flow between these species is possible. Analysis of additional material from poorly sampled regions therefore could provide novel insights into the nature of gene flow among species and populations identified here.

Phylogeographic implications

Vegetation in sub-Saharan Africa can be broadly classified into 'savanna' habitat, which is open habitat where a C_4 carbon fixation grass layer exists, and 'non-savanna' (i.e. tropical forest) habitat, which is closed and lacks a C_4 carbon fixation grass layer, with the distribution of each habitat type being largely dependent on the extent and seasonality of rainfall (Jacobs 2004; Lehmann *et al.* 2011). The distributions of these habitat types cycled during climatic oscillations, with savanna habitat becoming more extensive or shifting to lower latitudes during glacial periods (Dupont 2011). Within these habitat types, there is also variation in the seasonality of rain, a factor that may have played a role in the differentiation of *Xenopus laevis* in South Africa (Grohovaz *et al.* 1996). Thus, over the last 15 Myr or so, the evolution of *X. laevis sensu lato* took place on a varied and dynamic ecological and climatic landscape. It is also likely that geological features had an impact on population structure within *X. laevis*. In particular, the Great Escarpment (Fig. 3) lies between the population that ranges from Victoria West to Malawi and another population that ranges from Beaufort West to Laignsburg (Fig. 3). To the south-west of the Great Escarpment, the Cape Fold Belt, including the Swartberge Range and

the Langeberg Range (Fig. 3), lie between the Beaufort West/Laignsburg population and the coastal population in the south-western Western Cape Province, South Africa. The Niewoudtville population is also on top of the Great Escarpment, and has a zone of contact with the south-western Western Cape Province population nearby in Vredendal (Measey & Channing 2003), which is at the bottom of the Great Escarpment.

We present four molecular clock analyses (mitochondrial DNA, *DM-W*, concatenated autosomal DNA and species tree analysis of autosomal DNA) that assumed a strict molecular clock that was calibrated in two different ways (Methods). Despite these different calibration approaches, divergence times were quite similar across these analyses, although this does not necessarily indicate that these estimates are accurate. We resorted to relatively crude models of evolution in these analyses in order to achieve convergence on the posterior distribution of the parameters. Clearly, error in divergence times and evolutionary relationships could arise due to model misspecification, and other model violations. For example, because *BEAST does not account for migration, divergence times may be underestimated in the presence of migration (Leaché *et al.* 2014).

Although our divergence estimates for species within *X. laevis sensu lato* predate the Pleistocene, the geographic locations of proposed Pleistocene savanna refugia (see Fig. 2 in Lorenzen *et al.* 2012) coincide with the distributions of diverged evolutionary lineages in *X. laevis sensu lato*. A possible mechanism for these congruent areas of endemism is that diversification of many of these evolutionary lineages, including *X. laevis sensu lato*, was sculpted by the distributions and connectivity of suitable habitat, which waned and waxed over time. Being mostly aquatic, *Xenopus* presumably are particularly sensitive to ecological factors such as the abundance and seasonality of rainfall that affect opportunities for dispersal over land and time to complete metamorphosis. Regions with consistently habitable habitats potentially acted as 'lifeboats' that sustained divergent lineages that evolved before habitat contraction (Evans *et al.* 2004).

Conclusions

This study reports for the first time, evidence of phylogenetic discordance between two putatively maternally inherited genomic regions, mitochondrial DNA and *DM-W*, in the frog lineage *X. laevis sensu lato*. We do not know of a methodological explanation for this discordance, opening the possibility that there is a biological cause. Results also support the recognition of *X. laevis sensu stricto*, *X. victorianus*, *X. petersii* and newly revalidated *X. poweri*, but with the assignment of populations of *X. laevis sensu lato* from Nigeria and Cameroon to *X. poweri* instead of *X. laevis* and with the assignment of populations of *X. laevis sensu lato* from Botswana and Zambia to *X. poweri* instead of *X. petersii*. In doing so, this study clarifies the evolutionary history of one of the most intensively studied amphibian species in the context of its closely related relatives, and identifies additional differentiated populations that may themselves be meritorious of species status.

Acknowledgements

We thank four external reviewers, each of whom provided extensive and helpful comments on earlier versions of this manuscript. We thank P. Staab, D. Metzler, J. Measey, J. McGuire, and W. Conradie for helpful discussion or comments on this manuscript, Brian Golding for access to computer resources, and T. Hayes, L. Kalous, R. Tinsley, and P. Wagner for providing genetic samples. We thank M. D. Picker, Z. T. Nagy, M. M. Aristote, W. M. Moninga, M. Zigabe, A. M. Marcel, M. Luhumyo, J. F. Akuku, F. I. Alonda, A. M'Mema, F. B. Murutsi, B. Bajope, M. Manunu, M. Collet, and the Institut Congolais pour la Conservation de la Nature for hospitality, assistance with fieldwork, logistical support and permits. Major support for this study was provided by the National Science and Engineering Research Council of Canada (RGPIN/283102-2012 to B.J.E.). Additional support was from IVB institutional support (RVO: 68081766), the Ministry of Culture of the Czech Republic (DKRVO 2015/15, National Museum, 00023272), the Percy Sladen Memorial Fund, an IUCN/SSC Amphibian Specialist Group Seed Grant, K. Reed, M.D., research funds from the Department of Biology at Villanova University, a National Geographic Research and Exploration Grant (no. 8556-08), the University of Texas at El Paso and the National Science Foundation (DEB-1145459).

References

Ahl E (1924) Über eine Froschsammlung aus Nordost-Afrika und Arabien. *Mitteilungen aus dem Zoologischen Museum in Berlin*, **11**, 1–12.

Ali JR, Aitchison JC (2008) Gondwana to Asia: plate tectonics, paleogeography and the biological connectivity of the Indian sub-continent from the Middle Jurrasic through latest Eocene (166–35 Ma). *Earth-Science Reviews*, **88**, 145–166.

AmphibiaWeb (2014) *Information on Amphibian Biology and Conservation [Web Application]*. AmphibiaWeb, Berkeley, California.

Bewick AJ, Anderson DW, Evans BJ (2011) Evolution of the closely related, sex-related genes DM-W and DMRT1 in African clawed frogs (*Xenopus*). *Evolution*, **65**, 698–712.

Bewick AJ, Chain FJJ, Heled J, Evans BJ (2012) The pipid root. *Systematic Biology*, **61**, 913–926.

Blackler AW, Fischberg M (1968) Hybridization of Xenopus *laevis petersi* (*poweri*) and *X. l. laevis*. *Revue Suisse de Zoologie*, **75**, 1023–1032.

Blackler AW, Fischberg M, Newth DR (1965) Hybridisation of two subspecies of *Xenopus laevis* (Daudin). *Revue Suisse de Zoologie*, **72**, 841–857.

du Bocage JVB (1895) *Herpétologie d'Angola et du Congo : ouvrage publié sous les auspices du Ministère de la marine et des colonies*. Impr. Nationale, Lisbonne.

Bryant D, Moulton V (2004) Neighbor-Net: an agglomerative method for the construction of phylogenetic network. *Molecular Biology and Evolution*, **21**, 255–265.

Cannatella DC, de Sá RO (1993) *Xenopus laevis* as a model organism. *Systematic Biology*, **42**, 476–507.

Carr S, Brothers A, Wilson A (1987) Evolutionary inferences from restriction maps of mitochondrial DNA from nine taxa of *Xenopus* frogs. *Evolution*, **41**, 176–188.

Chain FJJ, Evans BJ (2006) Multiple mechanisms promote the retained expression of gene duplicates in the tetraploid frog *Xenopus laevis*. *PLoS Genetics*, **2**, e56.

Chain FJJ, Dushoff J, Evans BJ (2011) The odds of duplicate gene persistence after polyploidization. *BMC Genomics*, **12**, 599.

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, **7**, 747–756.

Daudin FM (1802) *Histoire Naturelle des Rainettes, des Grenouilles et des Crapauds*. Levrault, Paris.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Du Preez LH, Kunene N, Hanner R *et al.* (2009) Population-specific occurrence of testicular ovarian follicles in *Xenopus laevis* from South Africa. *Aquatic Toxicology*, **95**, 10–16.

Dupont L (2011) Orbital scale vegetation change in Africa. *Quaternary Science Reviews*, **30**, 3589–3602.

Durand E, Jay F, Gaggiotti OE, François O (2009) Spatial inference of admixture proportions and secondary contact zones. *Molecular Biology and Evolution*, **26**, 1963–1973.

Earl DA, von Holdt BM (2012) Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Eggert C, Fouquet A (2006) A preliminary biotelemetric study of a feral invasive *Xenopus laevis* population in France. *Alytes*, **23**, 144–149.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Evans BJ (2007) Ancestry influences the fate of duplicated genes millions of years after duplication in allopolyploid clawed frogs (*Xenopus*). *Genetics*, **176**, 1119–1130.

Evans BJ, Morales JC, Picker MD, Kelley DB, Melnick DJ (1997) Comparative molecular phylogeography of two *Xenopus* species, *X. gilli* and *X. laevis*, in the southwestern Cape Province, South Africa. *Molecular Ecology*, **6**, 333–343.

Evans BJ, Morales JC, Picker MD, Kelley DB, Melnick DJ (1998) Absence of extensive introgression between *Xenopus gilli* and *Xenopus laevis laevis* (Anura: Pipidae), in southwestern Cape Province, South Africa. *Copeia*, **1998**, 504–509.

Evans BJ, Kelley DB, Tinsley RC, Melnick DJ, Cannatella DC (2004) A mitochondrial DNA phylogeny of clawed frogs: phylogeography on sub-Saharan Africa and implications for polyploid evolution. *Molecular Phylogenetics and Evolution*, **33**, 197–213.

Evans BJ, Kelley DB, Melnick DJ, Cannatella DC (2005) Evolution of RAG-1 in polyploid clawed frogs. *Molecular Biology and Evolution*, **22**, 1193–1207.

Evans BJ, Carter TF, Hanner R *et al.* (2008) A new species of clawed frog (genus *Xenopus*), from the Itombwe Plateau, Democratic Republic of the Congo: implications for DNA barcodes and biodiversity conservation. *Zootaxa*, **1780**, 55–68.

Evans BJ, Greenbaum E, Kusamba C *et al.* (2011) Description of a new octoploid frog species (Anura: Pipidae: *Xenopus*) from the Democratic Republic of the Congo, with a discussion of the biogeography of African clawed frogs in the Albertine Rift. *Journal of Zoology, London*, **283**, 276–290.

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

Fogell DJ, Tolley KA, Measey GJ (2013) Mind the gaps: investigating the cause of the current range disjunction in the Cape Platanna, *Xenopus gilli* (Anura: Pipidae). *PeerJ*, **1**, e166.

François O, Durand E (2010) Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources*, **10**, 773–784.

Frost DR (2014) *Amphibian Species of the World: an Online Reference*. Version 6 (27 August 2014). American Museum of Natural History, New York.

Galtier N, Enard D, Radondy Y, Basin E, Belkhir K (2006) Mutation hotspots in mammalian mitochondrial DNA. *Genome Research*, **16**, 215–222.

Grohovaz GS, Harley B, Fabran B (1996) Significant mitochondrial DNA sequence divergence in natural populations of *Xenopus laevis* (Pipidae) from South Africa. *Herpetologica*, **52**, 247–253.

Gurdon JB (1996) Introductory comments: *Xenopus* as a laboratory animal. In: *The Biology of Xenopus* (eds Tinsley RC, Kobel HR), pp. 3–6. Oxford University Press, Oxford.

Gurdon JB, Hopwood N (2000) The introduction of *Xenopus laevis* into developmental biology: of empire, pregnancy testing and ribosomal genes. *International Journal of Developmental Biology*, **44**, 43–50.

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **21**, 160–174.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**, 570–580.

Hellsten U, Khokha MK, Grammer TC *et al.* (2007) Accelerated gene evolution and subfunctionalization in the pseudotetraploid frog *Xenopus laevis*. *BMC Evolutionary Biology*, **5**, 31.

Hewitt J (1927) Further descriptions of reptiles and batrachians from South Africa. Record of the Albany Museum. *Grahamstown*, **3**, 371–415.

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

Innan H, Nordborg M (2002) Recombination or mutational hot spots in human mtDNA? *Molecular Biology and Evolution*, **19**, 1122–1127.

Jacobs BF (2004) Palaeobotanical studies from tropical Africa: relevance to the evolution of forest, woodland and savanna biomes. *Philosophical Transactions of the Royal Society of London B*, **359**, 1573–1583.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Kelley DB, Tobias ML (1999) The vocal repertoire of *Xenopus laevis*. In: *The Design of Animal Communication* (eds Hauser M, Konishi M), pp. 9–35. MIT Press, Cambridge.

Kimura M, Ohta T (1972) On the stochastic model for estimation of mutational distance between homologous proteins. *Journal of Molecular Evolution*, **2**, 87–90.

Kobel HR, Loumont C, Tinsley RC (1996) The extant species. In: *The Biology of Xenopus* (eds Tinsley RC, Kobel HR), pp. 9–33. Clarendon Press, Oxford.

Kobel H, Barandun B, Thiebaud CH (1998) Mitochondrial rDNA phylogeny in *Xenopus*. *Herpetological Journal*, **8**, 13–17.

Kück P, Mayer C, Wägele J-W, Misof B (2012) Long branch effects distort maximum likelihood phylogenies in simulations despite selection of the correct model. *PLoS ONE*, **7**, e36593.

Leaché AD, Harris RB, Rannala B, Yang Z (2014) The influence of gene flow on species tree estimation: a simulation study. *Systematic Biology*, **63**, 17–30.

Lehmann CER, Archibald SA, Hoffman WA, Bond WJ (2011) Deciphering the distribution of the savanna biome. *New Phytologist*, **191**, 197–209.

Lemmon AR, Moriarty EC (2004) The importance of proper model assumptions in Bayesian phylogenetics. *Systematic Biology*, **53**, 265–277.

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

Lobos G, KJaksic FM (2005) The ongoing invasion of African clawed frogs (*Xenopus laevis*) in Chile: causes of concern. *Biodiversity and Conservation*, **14**, 429–439.

Lorenzen ED, Heller R, Siegismund HR (2012) Comparative phylogeography of African savannah ungulates. *Molecular Ecology*, **21**, 3656–3670.

Loveridge A (1932) New races of a skink (*Siaphos*) and frog (*Xenopus*) from the Uganda Protectorate. *Proceedings of the Biological Society of Washington*, **45**, 113–116.

Maisey JG (2000) Continental break up and the distribution of fishes of Western Gondwana during the Early Cretaceous. *Cretaceous Research*, **21**, 281–314.

McCoid MJ, Fritts TH (1980) Observations of feral populations of *Xenopus laevis* (Pipidae) in southern California. *Bulletin of the Southern California Academy of Sciences*, **79**, 82–86.

McCoid MJ, Fritts TH (1989) Growth and fatbody cycles in feral populations of the African clawed frog: *Xenopus laevis* (Pipidae), in California with comments on reproduction. *The Southwestern Naturalist*, **34**, 499–505.

McLoughlin S (2001) The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany*, **49**, 271–300.

Measey GJ, Channing A (2003) Phylogeography of the genus *Xenopus* in southern Africa. *Amphibia-Reptilia*, **24**, 321–330.

Measey GJ, Tinsley RC (1998) Feral *Xenopus laevis* in South Wales. *Herpetological Journal*, **8**, 23–27.

Measey GJ, Rodder D, Green SL *et al.* (2012) Ongoing invasions of the African clawed frog, *Xenopus laevis*: a global review. *Biological Invasions*, **14**, 2255–2270.

Morin RD, Chang E, Petrescu A *et al.* (2006) Sequencing and analysis of 10,967 full-length cDNA clones from *Xenopus laevis* and *Xenopus tropicalis* reveals post-tetraploidization transcriptome remodeling. *Genome Research*, **16**, 796–803.

Nylander JAA (2004) *MrModeltest v2*. Evolutionary Biology Centre, Uppsala University, Uppsala.

Parker HW (1936) Reptiles and amphibians collected by the Lake Rudoph Rift Valley Expedition, 1934. *Annals and Magazine of Natural History*, Series 10, **18**, 594–609.

Perret JL (1966) Les amphibiens du Cameroun. *Zoologische Jahrbücher. Abteilung für Systematik, Ökologie und Geographie*, **93**, 289–464.

Perrin N (2009) Sex reversal: a fountain of youth for sex chromosomes? *Evolution*, **63**, 3043–3049.

Picker MD (1993) Embryonic tolerance of *Xenopus* (Anura) to acidic blackwater. *Copeia*, **3**, 574–580.

Picker MD, Harrison JA, Wallace D (1996) Natural hybridization between *Xenopus laevis laevis* and *X. gilli* in the southwestern Cape Province, South Africa. In: *The Biology of Xenopus* (eds Tinsley RC, Kobel HR), pp. 61–71. Clarendon Press, Oxford.

Piganeau G, Gardner M, Eyre-Walker A (2004) A broad survey of recombination in animal mitochondria. *Molecular Biology and Evolution*, **21**, 2319–2325.

Pitman WC III, Cande S, LaBrecque J, Pindell J (1993) Fragmentation of Gondwana: the separation of Africa from South America. In: *Biological Relationships Between Africa and South America* (ed. Goldblatt P), pp. 15–34. Yale University Press, New Haven.

Poynton JC (1964) The Amphibia of southern Africa: a faunal study. *Annals of the Natal Museum*, **17**, 1–334.

Poynton JC, Broadley DG (1985) Amphibia Zambesiaca 1. Scolecomorphidae, Pipidae, Microhylidae, Hemisidae, Arthroleptidae. *Annals of the Natal Museum*, **26**, 503–553.

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

de Queiroz K (1998) The general lineage concept of species, species criteria, and the process of speciation. In: *Endless Forms: Species and Speciation* (eds Howard D, Berlocher S), pp. 57–75. Oxford Press, New York.

de Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology*, **56**, 879–886.

Rambaut A, Drummond AJ (2007) Tracer v1.5, Available from http://beast.bio.ed.ac.uk/Tracer.

Robert J, Abramowitz L, Gantress J, Morales HD (2007) *Xenopus laevis*: a possible vector of ranavirus infection? *Journal of Wildlife Diseases*, **43**, 645–652.

Roure B, Baurain D, Phillippe H (2013) Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Molecular Biology and Evolution*, **30**, 197–214.

Schmidt KP, Inger RF (1959) Amphibians exclusive of the genera *Afrixalus* and *Hyperolius*. *Exploration du Parc National de l'Upemba, Mission G.F. de Witte, en Collaboration avec W. Adam, A. Janssens, L. Van Meel, et R. Verheyen (1946–1949)*, **56**, 1–264.

Sémon M, Wolfe KH (2008) Preferential subfunctionalization of slow-evolving genes after allopolyploidization in *Xenopus laevis*. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 8333–8338.

Sereno PC, Wilson JA, Conrad JL (2004) New dinosaurs link southern landmasses in the Mid-Cretaceous. *Proceedings of the Royal Society of London Series B*, **271**, 1325–1330.

Shapiro HA, Zwarenstein H (1934) A rapid test for pregnancy of *Xenopus laevis*. *Nature*, **133**, 762.
Simmonds MP (1985) Interactions between *Xenopus* species in the southwestern Cape Province, South Africa. *South African Journal of Science*, **81**, 200.
South African Frog Re-assessment Group (SA-FRoG) ISASG (2013) *Xenopus gilli*. In: *IUCN Red List of Threatened Species. Version 2013.1*. IUCN 2013.
Spiegelhalter SD, Best NG, Carlin BP, Lingde AVD (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **64**, 583–639.
Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction. *American Journal of Human Genetics*, **73**, 1162–1169.
Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **86**, 978–989.
Stöck M, Horn A, Grossen C *et al.* (2011) Ever-young sex chromosomes in European tree frogs. *PLoS Biology*, **9**, e1001062.
Sun S, Evans BJ, Golding GB (2011) Heterotachy and false positives for recombination in animal mtDNA. *Molecular Biology and Evolution*, **28**, 2549–2559.
Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
Tavaré S (1986) Some probabilistic and statistical problems on the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, **17**, 277–290.
Tinsley RC (1981) Interactions between *Xenopus* species (Anura Pipidae). *Monitore Zoologico Italiano*, **15**, 133–150.
Tinsley RC, McCoid MJ (1996) Feral populations of *Xenopus* outside Africa. In: *The Biology of Xenopus* (eds Tinsley RC, Kobel HR), pp. 81–94. Clarendon Press, Oxford.
Tinsley RC, Loumont C, Kobel HR (1996) Geographical distribution and ecology. In: *The Biology of Xenopus* (eds Tinsley RC, Kobel HR), pp. 35–59. Clarendon Press, Oxford.
Tobias ML, Viswanathan SS, Kelley DB (1998) Rapping, a female receptive call, initiates male-female duets in the South African clawed frog. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 1870–1875.
Tobias M, O'Hagan R, Horng S, Kelley D (2004) Vocal communication between male *Xenopus laevis*; behavioral context and sexual state. *Animal Behavior*, **67**, 353–365.
Tobias ML, Evans BJ, Kelley DB (2011) Evolution of advertisement call in African clawed frogs. *Behaviour*, **148**, 519–549.
Tsaousis AD, Martin DP, Ladoukakis D, Posada D, Zouros E (2005) Widespread recombination in published animal mtDNA sequences. *Molecular Biology and Evolution*, **22**, 925–933.
Tymowska J (1991) Polyploidy and cytogenetic variation in frogs of the genus *Xenopus*. In: *Amphibian Cytogenetics and Evolution* (eds Green DS, Sessions SK), pp. 259–297. Academic Press, San Diego.
Weisman AI, Coates CW (1941) The "frog" test (*Xenopus laevis*), as a rapid diagnostic test for early pregnancy. *Endocrinology*, **28**, 141–142.
Weldon C, du Preez L, Hyatt AD, Muller R, Speare R (2004) Origin of the amphibian chytrid fungus. *Emerging Infectious Diseases*, **19**, 2100–2105.
Yang Z, Rannala B (2010) Bayesian species delimitations using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 9264–9269.
Yoshimoto S, Okada E, Umemoto H *et al.* (2008) A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 2469–2474.

The goals of this study were developed by B.L.S.F., A.J.B. and B.J.E. Fieldwork for this study was performed by B.L.S.F., A.J.B., E.G., V.G., C.K. and B.J.E. Molecular data were collected by B.L.S.F., A.J.B., T.L.H. and B.J.E. Analysis was performed by B.L.S.F., A.J.B., T.L.H. and B.J.E. The paper was initially written by B.L.S.F., E.G., V.G. and B.J.E., and all authors then edited the paper.

Data accessibility

Sequence data in this study have been deposited in GenBank (new sequences not previously in GenBank have Accession nos. KP343951– KP345838), input files and tree files have been deposited in Dryad (doi:10.5061/dryad.4n2c4) and sampling localities are available in the Table S1.

Supporting information

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

Table S1 Locality information for genetic samples used in this study and details on sequence data collected for each sample.

Fig. S1 Gene networks of the 15 autosomal loci for *X. laevis*.

Fig. S2 Tess analysis with all individuals from North and Central Africa ($n = 41$), with a reduced representation of individuals from South Africa ($n = 25$), including (a) individual assignments and (b) the deviance information criterion (DIC) begins to level off at values of K greater than 5, supporting 6–7 clusters. Labeling follows Fig. 5.

Fig. S3 Structure analysis of (A) 135 *X. laevis* individuals for 15 loci with labeling following Fig. 5. (b) The ΔK statistic (Evanno *et al.* 2005) indicates support for 5 clusters, which corresponds with the value of K where the log-likelihood ($\ln L$) begins to level off.