FINAL REPORT FOR USFWS CONTRACT #F13PX00770

Title – "Using novel genetic markers and multigenic species delimitation methods to resolve the species status of the cave-dwelling spider species *Cicurina wartoni* Gertsch from Travis County, Texas"

Investigator - Dr. Marshal Hedin
Department of Biology, San Diego State University
San Diego, CA 92182
619-251-8674
mhedin@mail.sdsu.edu

Date: 15 March, 2014

Introduction

Cicurina wartoni Gertsch is an eyeless, cave-endemic spider species known only from a single geographic location (Pickle Pit) in Travis County, Texas (Gertsch 1992, Paquin & Duperre 2009). Because of the apparent rarity and endemicity of *C. wartoni*, this species is of conservation concern. The formal taxonomic description of *C. wartoni* was conducted by Willis Gertsch, who recognized this species as distinct based on the epigynal morphology of a **single adult female specimen**. Paquin and Duperre (2009) later redescribed *C. wartoni*, and provided morphological and geographic evidence to suggest that this taxon is a member of the *C. buwata* species complex (including *C. wartoni, C. reddelli, C. buwata*, and *C. travisae*). **Taking into account intraspecific variation,** members of the *C. buwata* species complex are **extremely similar from a morphological perspective, suggesting the possibility that these taxa are actually conspecific** (**see FIG 1**).

The current species hypothesis for *C. wartoni,* based on a single specimen that is not clearly morphologically different from regional congeners, clearly requires further attention. The general goal of this research was to use DNA sequence data from multiple genes (one mitochondrial gene, 8 nuclear genes) to rigorously test the species status of *C. wartoni*. DNA sequence data have long been used to formulate and test species delimitation hypotheses (summarized in Sites & Marshall 2004; Carmargo & Sites 2013). Examples include using population genetic principles to identify distinct genetic clusters (e.g., using STRUCTURE in species delimitation, Weisrock et al 2010; Satler et al 2013), or assessing the monophyly of hypothesized species on gene trees (Avise & Ball 1990; Baum & Shaw 1995). More recently, researchers have developed mutligenic methods that do not require strict gene tree monophyly, but rather, allow for some gene tree heterogeneity while delimiting species using multispecies coalescent models (Leache & Fujita 2012; Fujita et al 2012; McKay et al 2013; Rannala & Yang 2013). Because any single analytical method (or data source) is susceptible to error, most modern species delimitation analyses now include *both* multiple lines of evidence (e.g., mtDNA data, nuclear data, morphology, etc.) *and* comparisons of multiple analyses (Carstens et al 2013; Satler et al 2013). **Robust, data-rich hypotheses are subsequently established using an integration or consensus of analytical methods - this approach to the species delimitation problem was used in this research.**

Methods

Transcriptomics & Targeted Gene Development – Genomic resources were developed specifically for Texas cave-dwelling *Cicurina* using comparative transcriptomic data derived from next-generation sequencing (see Thomson et al 2010). The transcribed portion of the genome (i.e., protein-coding genes, untranslated regions associated with these genes) is called the *transcriptome*. Transcriptomes were characterized for adult female specimens of *C. travisae* (specimens from Tooth Cave) and *C. vibora* (specimens from Temples of Thor Cave). The program FastQC version 0.10.1 (Babraham Bioinformatics) was used to confirm quality of the raw reads prior to assembly, and the perl script Trim Galore! (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used to remove low quality sequences and trim Illumina adaptors. Sequences less than thirty base pairs in length, as well as those consisting of greater than one percent ambiguity characters, were removed with PRINSEQ Lite (Schmieder & Edwards 2011). The remaining raw reads were assembled *de novo* into transcripts using the Trinity software package (Grabherr et al. 2011). Each assembly resulted in over 100,000 transcripts for each taxon, with a mean transcript length of 519 (*C. travisae*) and 407 (*C. vibora*), respectively. Raw reads will be submitted to the Short Read Archive (http://www.ncbi.nlm.nih.gov/sra) upon formal publication of this research.

Cicurina transcripts over 500 basepairs in length were compared using high-stringency BLAST searches, with the goal of finding orthologous, single copy genes that exhibit a large number of basepair differences (in either UTR or exonic regions) between taxa. Based on these comparisons, standard PCR primers were designed to target ~100 different, rapidly evolving gene regions. These primers were tested on a small panel of eyeless *Cicurina* genomics – those gene targets providing consistent, robust PCR results were Sanger sequenced to assay sequence quality and to check for evidence of paralogy (genes that are part of gene families, undesirable for phylogenetic analysis). **After multiple iterations of primer testing and preliminary sequencing, eight nuclear gene regions and one mitochondrial gene region were chosen for further comprehensive specimen sampling.**

Survey of Genetic Variation in Relevant Specimens – Genomic DNA was extracted (using standard methods) from *C. buwata* species complex specimens occurring in **27 regional caves** (**see FIG 2**), plus a handful of outgroup samples from outside the region of interest. One to several spiders were sampled from each cave. Specimen identification within the *C. buwata* complex was based on a combined consideration of morphology, geographic location and/or genetic affiliation. Based on these alternative lines of evidence **specimens were allocated into five *a priori* groups as follows:**

1) *C. buwata* – The sample includes eyeless *immature* specimens from three caves that are **known locations** for *C. buwata* (Buttercup Creek, Marigold Cave, Testudo Tube - Paquin & Duperre 2009). Eyeless immature and/or adult specimens from eight additional caves (No Rent Cave, Weldon Cave, McNeil Bat Cave, Lakeline Cave, Dies Ranch Treasure Cave, Apple Riata Trace, Broken Arrow Cave, Babe Cave) were identified as *C. buwata* based on consistent placement into a *C. buwata* genetic clade (see Results), or based on female epigynal morphology consistent with the description for this taxon.

2) *C. wartoni* – Three **eyeless *immature*** specimens from the type locality of *C. wartoni* (Pickle Pit) were tentatively identified as this species, as no other eyeless *Cicurina* have ever been recorded from this cave.

3) *C. travisae* – The sample includes eyeless immature and/or adult specimens from five caves that are known locations for *C. travisae* (McDonald Cave, Amber Cave, Kretschmarr Double Pit, North Root Cave, Tooth Cave - Paquin & Duperre 2009). Tooth Cave is the type locality of *C. travisae.*

4) *C. reddelli* – An adult female and immature male specimen from Cotterell Cave were identified as *C. reddelli*. This cave is the type and only known locality for *C. reddelli*, and the female specimen available has an epigynal morphology consistent with the description for *C. reddelli* (Paquin & Duperre 2009; see **FIGS 3-4**). Gertsch (1992) also listed Cotterell Cave as a location for *C. buwata* – our samples from Cotterell Cave are clearly not genetically allied with *C. buwata* (see Results).

5) *C. travisae Complex* – Both immature and adult specimens from nine caves geographically intermediate between Cotterell Cave and McDonald Cave (Spider Cave, Beard Ranch Cave, Jester Estates Cave, Jest John Cave, Ken Butler Pit, Two Trunks Cave, Gallifer Cave, Geode Cave, Stovepipe Cave; **see FIG 2**) are genetic members of the ***C. travisae* complex** (see Results), but are from caves not previously known to house eyeless *Cicurina*. Adult females are available from some of these caves, but because *C. travisae*, *C. reddelli*, and *C. wartoni* have **very** similar female epigynal morphologies (Paquin & Duperre 2009; see **FIGS 3-4**), a conservative approach was taken and these specimens were not allocated to any known species *a priori*. Gertsch (1992) also listed Gallifer Cave as a location for *C. buwata* – our samples from Gallifer Cave are clearly not genetically allied with *C. buwata* (see Results).

Digital Imaging of Adult Specimens – Genitalic features of all **adult** specimens used in this analysis were digitally imaged using a Visionary Digital BK plus system (http://www.visionarydigital.com), including a Canon 40D digital camera, Infinity Optics Long Distance Microscope, P-51 camera controller, and FX2 lighting system. Individual images were combined into a composite image using Zerene Stacker V1.04 software; this composite image was then edited using Adobe Photoshop CS3. Female epigyna and male pedipalps were dissected from specimens using fine forceps, immersed for 2-5 minutes in BioQuip specimen clearing fluid (www.bioquip.com) on a depression slide, then imaged directly in this fluid on slides without the use of cover slips.

Sanger Data Collection, Editing, Phasing – After multiple iterations of primer testing and preliminary sequencing, one mitochondrial and eight nuclear genes were chosen for comprehensive specimen sampling. Amplified PCR products were purified using standard techniques, and Sanger sequenced (in both directions) at *Macrogen USA*. DNA sequences were edited using Geneious Pro software (www.geneious.com/), and trimmed to exclude primer sequences. The *C. travisae* transcriptome data were not used for nuclear gene analyses, since heterozygosity could not be assessed (transcriptomes were derived from combined RNA of two individuals). Only one of the matrices included gaps – for the F10F11 matrix, a 9 basepair

insertion found only in *C. buwata* specimens was recoded as a two-state nucleotide transition. Heterozygous nuclear sequences were bioinformatically phased to alleles using the software program PHASE 2.1.1 (Stephens et al. 2001; Stephens & Donnelly 2003). SeqPHASE (Flot 2010) was used to convert matrices for input into PHASE. PHASE analyses were conducting using default settings (phase threshold = 90%, 100 iterations, thinning interval = 1, burn-in = 100), and were repeated multiple times to ensure consistent results. This method of bioinformatic resolution of allelic phase is standard in molecular population biology and systematics (e.g., Stephens & Donnelly 2003 publication with over 2000 citations).

Gene Tree Analyses – Individual gene trees were estimated using maximum likelihood, as implemented in the RAxML_GUI (Stamatakis 2006; Stamatakis et al. 2008; Silvestro & Michalak 2011). All gene trees were estimated using a GTRGAMMA model, with nodal support assessed with 1000 bootstrap replicates.

POFAD Analyses – Multigenic nuclear genetic distances (uncorrected p-distances) among individuals were calculated using POFAD (Joly & Bruneau 2006). To eliminate the potentially confounding influence of extreme female-based population structure, mitochondrial data were not included in POFAD analyses. Also, more distant outgroups were excluded from this analysis. Because all nuclear matrices are similar in aligned length and levels of variation, individual matrices were given the same weight (standardized weights option). Summary distances were used to reconstruct a NeighborNet network in the program SplitsTree4 (Huson & Bryant 2006).

STRUCTURE Analyses – STRUCTURE is a genetic clustering algorithm that detects population structure through the use of allele frequencies, identifying genetically homogeneous groups of individuals that are in both Hardy-Weinberg and linkage equilibrium (Pritchard et al. 2000, 2010). **Although this program was originally developed for use in population structure analyses, STRUCTURE has been used to identify the number of distinct genetic clusters (i.e.*, independently evolving genetic lineages*) in many species delimitation analyses (e.g., Weisrock et al 2010; Rittmeyer & Austin 2012; Satler et al 2013).**

To eliminate the potentially confounding influence of high female-based population structure, mitochondrial data were not included in STRUCTURE analyses. More distant outgroups were also excluded from this analysis. SNAP Map was used (Aylor et al., 2006; Price & Carbone, 2005) to convert nuclear DNA sequences for each gene region to numbered unique alleles (haplotypes), which were then used as input in STRUCTURE analyses (O'Neill et al. 2012). Structure runs were conducted assuming between 1 and 6 genetic clusters ($K = 1$ through $K = 6$), with each K value replicated three times. Analyses used an admixture model, were conducted with a burn-in of 1×10^5 steps (with 1×10^6 MCMC steps after burn-in), and allele frequencies were considered independent among populations. Data were summarized using the *FullSearch* algorithm of CLUMPP (Jakobsson & Rosenberg 2007), and visualized with DISTRUCT (Rosenberg 2004).

Evanno et al (2005) used simulations to show that the maximal value of the log probability of the data given K *(L(K))* does not necessarily provide an accurate estimation of K, but instead ***typically overestimates*** K. Instead, a statistic called ΔK (= rate of change in log probability of data between successive K vales) consistently provides a more accurate estimate of K under the

simulation conditions explored. Here, both approaches were considered to identify an optimal K value – estimates from multiple replicates for multiple K values were calculated in Structure Harvester (Earl & vonHoldt 2012).

BPP Analyses – BPP is a Bayesian coalescent-based method that uses multilocus genetic data in combination with reversible-jump Markov chain Monte Carlo (rjMCMC) algorithms to calculate the posterior probabilities of **different species delimitation *models***. This method was developed by Yang and Rannala (Yang & Rannala 2010, Rannala & Yang 2013), has been used empirically many times (e.g., Leaché & Fujita 2010, Satler et al 2013), and has been evaluated in simulation by Zhang et al (2011). The method is attractive because it allows **alternative species delimitation models to be tested in a probabilistic manner.**

To eliminate the potentially confounding influence of high female-based population structure, mitochondrial data were not included in BPP analyses. Distant outgroups were also excluded. The rjMCMC species delimitation method was used with algorithm 1 (a value = 2, m value = 1), ambiguous data columns were not removed (cleandata =0), and the analysis was set for automatic fine-tune adjustments. Two different combinations for population size (theta, θ) gamma priors were used, including a large theta prior G(alpha = 1, beta = 10), and a more diffuse (uninformative) theta prior G(alpha = 2, beta = 100). For both analyses a diffuse tau (τ) gamma prior of G(alpha = 2, beta = 1000) was used for the age of the root (tau0), while the other divergence time parameters were assigned the Dirichlet prior (Yang & Rannala 2010). Each prior combination was run twice to check convergence and proper mixing. Analyses were run for 100,000 generations, sampling every 5 generations with 10,000 burnin. Species tree nodes with posterior probability values > 0.95 were considered supported species delimitations; values below 0.95 were considered as evidence for collapsing a species tree node.

BPP analyses require *a priori* specimen allocation and an input "guide tree"– following results of several other analyses (see below), two alternatives were considered:

a) Following POFAD, STRUCTURE (K = 3) and geography (see **FIG 2**), four "undetermined" cave populations in the vicinity of *C. travisae* (Two Trunks Cave, Gallifer Cave, Geode Cave, Stovepipe Cave) were allocated to *C. travisae*, and five eastern "undetermined" populations (Spider Cave, Ken Butler Pit, Jester Estates Cave, Jest John Cave, Beard Ranch Cave) were allocated to *C. reddelli*. The guide tree used was as follows: (*C. buwata* sister to (*C. travisae, (C. reddelli* + *C. wartoni*)), which is a topology consistent with POFAD and STRUCTURE (K = 3) results.

b) A more conservative analysis was conducted in which only specimens from known cave locations were used in analyses (i.e., *C. wartoni* specimens from Pickle Pit*; C. travisae* specimens from McDonald Cave, Amber Cave, Kretschmarr Double Pit, North Root Cave and Tooth Cave; *C. reddelli* specimens from Cotterell Cave). The same guide tree as above was used.

Results

Digital Imaging of Adult Specimens – Genitalic structures (female epigyna, male pedipalps) for all adult specimens used in genetic analyses were digitally imaged (**FIGS 3-5**). Although

these structures were not subject to formal quantitative analyses (e.g., morphometric analyses) because of small sample sizes, consideration of patterns of qualitative variation provided important information, summarized as follows:

a) Epigynal morphology of the female specimens examined is consistent with that reported for the *C. buwata* complex (compare **FIG 1** to **FIGS 3-4**).

b) Defining "diagnostic" epigynal features for different species in this complex is very challenging.

c) Epigynal variation ***within*** a species (or population) sometimes exceeds variation ***between*** hypothesized species (e.g., Tooth Cave specimens G1958, G2014; **FIGS 3-4**).

Four adult male specimens were also available for study and imaged (**FIG 5**). Because male specimens have never been described for any species in the *C. buwata* complex (Paquin & Duperre 2009), there is no basis for formal species-level comparison. It is notable however that the male specimens available have **essentially identical palp morphologies**, and that all of these specimens fall into the same preferred (K = 2) STRUCTURE genetic cluster (see **FIG 10**).

Mitochondrial COI gene tree – Mitochondrial sequences are available upon request from M Hedin; these data will be submitted to GenBank upon formal publication of this research. Maximum likelihood phylogenetic analysis of mitochondrial sequences (**FIG 6**) provide evidence for a well-supported (**bootstrap proportions values > 90**) primary separation within the *C. buwata* species complex, corresponding to *C. buwata* versus *C. reddelli/C. travisae/ C. wartoni* (**hereafter called the *C. travisae* complex**).

Specimens of *C. wartoni* form a well-supported subclade on the mitochondrial gene tree, but this gene tree pattern CANNOT be used as conclusive evidence for unique species status, since many single caves with multiple sampled specimens show identical patterns (e.g., McDonald Cave, Tooth Cave, Spider Cave, Jester Estates Cave, Lakeline Cave; **FIG 6**). There is no evidence to hypothesize each of these individual cave populations as a unique species - instead, the mitochondrial data supports limited female-based gene flow at the local geographic level (i.e., population structure), consistent with the natural habitat fragmentation.

Nuclear genetic data & Individual Nuclear Gene Trees – Nuclear sequences are available upon request from M Hedin; these data will be submitted to GenBank upon formal publication of this research. Data were collected and phased for eight nuclear genes, with a combined total aligned length of 5586 basepairs. All nuclear genes correspond to UTR or exonic sequences, and PCR-amplified Sanger data match transcriptomic data (i.e., PCR primers amplified the correct "target" gene region). Final matrices include very little missing data (8 nuclear gene matrices X 34 individuals per matrix – 7 total missing sequences). No single specimen is missing data for more than one gene region.

Each nuclear gene tree is unique, but many ***generalizable patterns*** are apparent (**FIG 7a-h**):

a) **Members of the *C. travisae* complex are recovered as a clade on all nuclear gene trees, and this clade is always genetically distinguishable from *C. buwata*.**

b) Within the *C. travisae* complex, genetic relationships vary from gene to gene, and sequences from individual hypothesized species do NOT fall together on gene trees – instead, **sequences from different hypothesized species within the *C. travisae* complex are intermixed on nuclear gene trees**.

c) **Sequences from *C. wartoni* specimens never form an exclusive group on nuclear gene trees. Instead, Pickle Pit sequences are always intermixed with other members of the *C. travisae* complex.**

NeighborNet network from POFAD distances – The POFAD network, based on nuclear data without distant outgroups, shows an obvious division between *C. buwata* and members of the *C. travisae* complex (**FIG 8**). Within the *C. travisae* complex, ***C. wartoni* specimens are not grouped together on the network**. Specimen (and population) placement on the network coincides roughly with geographic position (western populations basal, eastern populations derived; **FIG 8B**). A group of populations including *C. reddelli* from Cotterell Cave, five eastern "undetermined" cave populations (Beard Ranch Cave, Jest John Cave, Jester Estates Cave, Ken Butler Pit, Spider Cave, see **FIG 2**), and two specimens of *C. wartoni* cluster together – this genetic association is very similar to that recovered by STRUCTURE K = 3 analyses (see below, **FIG 10**).

STRUCTURE Analyses – Results from STRUCTURE suggest two genetic partitions, with ***K* = 2** including the largest ΔK (461.86) as estimated using the Evanno method (**FIG 9A**). The two genetic clusters correspond to *C. buwata* samples versus samples of the *C. travisae* complex (**FIG 10**). **This result is consistent with the hypothesis that members of the *C. travisae* complex represent a single species, rather than three species.**

Although K = 2 is optimal under the Evanno method, other K values were considered, allowing for the possibility that the Evanno method might be underestimating species diversity in this complex. In particular, a K = 3 hypothesis was considered (where *L(K)* is nearly maximal, **FIG 9B**) – this genetic clustering implies three separate genetic groups corresponding to *C. buwata*, *C. travisae (plus western undetermined)*, and *C. reddelli (plus eastern undetermined) + C. wartoni* (**FIG 10**). This hypothesis was further evaluated using BPP (see below). Both K = 4 and K = 5 clustering values are biologically unrealistic, as **all** specimens in the *C. travisae* complex have a high probability of being in multiple genetic clusters under these K values (**FIG 10**).

BPP Analyses – Based on results of POFAD (**FIG 8**) and K = 3 STRUCTURE analyses (**FIG 10**), a (*C. buwata* sister to (*C. travisae, (C. reddelli + C. wartoni*)) guide tree was used. Using both specimen allocations summarized above (i.e., undetermined specimens included versus excluded), there is **no probabilistic support** for a species tree node separating *C. wartoni* from *C. reddelli* (i.e., *C. wartoni, C. reddelli* and *C. travisae* are **collapsed as a single species**), but there is strong support for a *C. buwata* versus *C. travisae* complex species tree node. **To**

summarize, BPP analyses are consistent with the hypothesis that members of the *C. travisae* complex represent a single species, versus three species. This result is concordant with the single *C. travisae* complex POFAD cluster, and the preferred K=2 STRUCTURE results.

Discussion

Modern species delimitation should be a data-rich process, where large amounts of data from several sources are subject to multiple objective analyses to detect distinct evolutionary lineages. The existing species hypothesis for *Cicurina wartoni*, based on **subjective consideration of a single specimen**, is clearly tenuous. The data and analyses presented here were used to rigorously test this existing hypothesis.

Within the geographic region of interest, all genetic data support a clear distinction between northern *C. buwata* and southern *C. travisae* complex members (**FIG 2**). The hypothesis of Gertsch (1992) that *C. buwata* occurs in the southern Cotterell and Gallifer Caves (see Fig 131 of Paquin and Duperre 2009) is not supported by the data presented here. Also, Paquin and Duperre (2009) discussed the potential synonymy of *C. buwata* and *C. travisae* - genetic data showing deep divergences in both mitochondrial and nuclear genomes do not support this synonymy hypothesis.

Three described species comprise the *C. travisae* complex. As shown here, different members of this complex have extremely similar female epigynal morphologies (**FIGS 3,4**). Male palpal morphologies in the *region* of interest are indistinguishable (**FIG 5**). Although mitochondrial data indicate high levels of female-based genetic structuring in this complex (resulting in exclusive mitochondrial clades corresponding to single cave populations – e.g., *C. wartoni*, *C. reddelli*, many other cave populations), this pattern is interpreted as evidence for population structuring, not species-level divergence. It is now well-established that species limits should not be based on mitochondrial evidence alone, and particularly where female-based gene flow is constrained, mitochondrial lineage diversity can greatly exceed species-level diversity (e.g., Satler et al 2013).

The multigenic nuclear perspective developed here shows that *C. wartoni, C. travisae,* and *C. reddelli* are not clearly genetically distinguishable for multiple nuclear loci. Analyses of individual nuclear gene trees reveal that sequences from different *C. travisae* complex members are always intermixed on nuclear gene trees. A POFAD combined analysis of all nuclear data shows a single genetic cluster corresponding to the *C. travisae* complex. There is some signal for WEST (including *C. travisae*) to EAST (including *C. reddelli* but also "western" *C. wartoni*) geographic structuring in this network, mirrored in K=3 STRUCTURE analyses. However, this geographic structure is **not supported as species-level divergence** in BPP analyses, which instead **support the *C. travisae* complex as a single genetic lineage** (consistent with the preferred K=2 STRUCTURE results).

Overall, the combination of data and analyses presented here support moderate geographic genetic structuring in a single widespread species, as might be expected in a system of cave-dwelling taxa. **My recommendation is that *C. wartoni, C. travisae,* and *C. reddelli* be treated as a single species until formal taxonomic changes can be published.** This recommendation

does not negate the potential conservation importance of the Pickle Pit *Cicurina* population – this population is uniquely geographically isolated in the complex, and individuals from this population carry unique genetic variation in both mitochondrial and nuclear genomes.

Acknowledgements:

Cyndee Watson has provided tremendous support for this project, and deserves special thanks. Mark Sanders, Todd Bayless, P. Fushille and Jet Larsen provided specimens. Kristen Emata provided expert lab assistance, with help from David Zezoff. James Starrett extracted RNA, while Shahan Derkarabetian and Dave Carlson helped in the transcriptome assembly process.

REFERENCES

Avise JC, RM Ball. 1990. Principles of genealogical concordance in species concepts and biological taxonomy. Oxford Surv. Evol. Biol. 7:45-67.

Aylor, DL, EW Price, I Carbone. 2006. SNAP: Combine and Map modules for multilocus population genetic analysis. Bioinformatics 22 (11): 1399-1401.

Baum DA, KL Shaw. 1995. Genealogical perspectives on the species problem. Pp. 289-303 *in* P. C. Hoch & A. G. Stephenson, eds. Molecular and experimental approaches to plant biosystematics. Missouri Botanical Garden, St. Louis.

Camargo A, Sites J. 2013. Species delimitation: a decade after the renaissance. In: Pavlinov I, editor. The Species Problem - Ongoing Issues. InTech.

Carstens BC, Pelletier TA, Reid NM, Satler JD. 2013. How to fail at species delimitation. Mol Ecol 22: 4369-4383.

Earl DA, BM vonHoldt 2012 STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4 (2), 359-361.

Ence DD, BC Carstens. 2011. SpedeSTEM: a rapid and accurate method for species delimitation. Molecular Ecology Resources 11, 473-480.

Evanno G., Regnaut S., Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611-2620.

Flot JF. 2010. Seqphase: a web tool for interconverting phase input/output files and fasta sequence alignments. Mol. Ecol. Resources. 10(1): 162-166.

Fujita M.K., Leaché A.D., Burbrink F.T., McGuire J.A., Moritz C. 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol. Evol. 27:480–488.

Gertsch WJ. 1992. Distribution patterns and speciation in North American cave spiders with a list of the troglobites and revision of the cicurinas of the subgenus *Cicurella*. Texas Memorial Museum Speleological Monographs, 3. Studies on the endogean fauna of North America 2, 75–122.

Grabherr MG, + 20 others. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nature Biotechnology 29(7), 644-52.

Huson DH, D Bryant. 2006. Application of Phylogenetic Networks in Evolutionary Studies, Molecular Biology and Evolution 23(2), 254-267.

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

Joly S, A Bruneau. 2006. Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: an example from *Rosa* in North America. Systematic Biology 55(4), 623-636.

Leaché AD, MK Fujita. 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). Proceedings of the Royal Society B: Biological Sciences 277, 3071-3077.

McKay BD, HL Mays Jr, Y Wu, H Li, C-te Yao, I Nishiumi, F Zou. 2013. An empirical comparison of character-based and coalescent-based approaches to species delimitation in a young avian complex. Molecular Ecology 22, 4943-4957.

O'Neill EM + 7 others. 2012. Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. Molecular Ecology, 22: 111-129.
Paquin P, N Duperre. 2009. A first step towards the revision of *Cicurina*: redescription of type specimens of 60 troglobitic species of the subgenus *Cicurella* (Araneae: Dictynidae), and a first visual assessment of their distribution. Zootaxa 2002, 1–67.
Price EW, I Carbone. 2005. SNAP: workbench management tool for evolutionary population genetic analysis. Bioinformatics 21: 402-404.
Pritchard JK, M Stephens, P Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.
Pritchard JK, X Wen, D Falush. 2010. Documentation for *STRUCTURE* software: Version 2.3. Available from http://pritch.bsd.uchicago.edu/structure.html. 1-38.
Rannala B, Z Yang. 2013. Improved reversible jump algorithms for Bayesian species delimitation. Genetics 195, 245-253.
Rittmeyer EN, CC Austin 2012. The effects of sampling on delimiting species from multi-locus sequence data. Mol Phylogenet Evol. 2012 65(2):451-63.
Rosenberg NA. 2004. *Distruct:* a program for the graphical display of population structure. Mol. Ecol. Notes. 4:137-138.
Satler JD, BC Carstens, M Hedin. 2013. Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). Systematic Biology 62, 805–823.
Schmieder R, R Edwards. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27(6), 863-864.
Silvestro D, I Michalak. (2011) RAXMLGUI: A graphical front-end for RAxML. Organisms Diversity & Evolution, DOI: 10.1007/s13127-011-0056-0
Sites Jr, JW, Marshall JC (2004) Operational criteria for delimiting species. Ann Rev Ecol Evol S 35: 199-227.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
Stamatakis A, P Hoover, J Rougemont. 2008. A rapid bootstrap algorithm for the RAxML Web Servers. Systematic Biology 57, 758-771.
Stephens M, N Smith, P Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68, 978-989.
Stephens M, P Donnelly. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. 73, 1162-1169.
Thomson RC, IJ Wang, JR Johnson. 2010. Genome-enabled development of DNA markers for ecology, evolution and conservation. Molecular Ecology 19, 2184-2195.
Weisrock DW, Rasoloarison RM, Fiorentino I, Ralison JM, Goodman SM, Kappeler PM, AD Yoder. 2010. Delimiting species without nuclear monophyly in Madagascar's mouse lemurs. PLoS ONE. 5(3):e9883.
Yang Z, B Rannala. 2010. Bayesian species delimitation using multilocus sequence data. Proc Natl Acad Sci USA 107, 9264-9269.
Zhang C, DX Zhang, T Zhu, Z Yang. 2011. Evaluation of a Bayesian coalescent method of species delimitation. Syst Biol 60 (6): 747-761.

Figure Legends

Fig 1 – Female epigynal drawings for members of the "*C. buwata* complex", from Paquin and Duperre (2009). Left images = ventral view, right images = dorsal view.

Fig 2 – Map showing geographic distribution of sampled populations. Specimens allocated into five primary *a priori* groups based on a combination of morphology, geographic location and/or genetic affiliation (see text for details). Colors used to designate taxa (amber = *C. buwata*, blue = *C. travisae*, green = *C. reddelli*, red = *C. wartoni*, black = *C. travisae species complex*) used for remainder of figures (except FIG 10).

Fig 3 – Epigynal morphologies for adult female specimens used in this study, **ventral view**. Specimen determinations for *C. buwata* based primarily on phylogenetic placement in DNA analyses, except for Buttercup Creek Cave which is a known location for *C. buwata*. Kretcshmarr Double Pit and North Root Cave are known localities for *C. travisae*, and Tooth Cave is the type locality for *C. travisae*. Cotterell Cave is the type locality for *C. reddelli*,

Fig 4 – Epigynal morphologies for adult female specimens, **dorsal view**. Specimen determinations for *C. buwata* are based primarily on phylogenetic placement in DNA analyses, except for Buttercup Creek Cave which is a known location for *C. buwata*. Kretcshmarr Double Pit and North Root Cave are known localities for *C. travisae*, and Tooth Cave is the type locality for *C. travisae*. Cotterell Cave is the type locality for *C. reddelli*,

Fig 5 – Palp morphologies (**left palp, ventral view**) for adult male *C. travisae complex* specimens used in this study. Amber Cave is a known locality for *C. travisae*, although adult males for this species have never been described. All other specimens are from caves where eyeless *Cicurina* have never been recorded.

Fig 6 – RAxML maximum likelihood **mitochondrial gene tree**. Bootstrap values >70 shown.

Fig 7 – RAxML maximum likelihood **nuclear gene trees**. Bootstrap values above 70 shown.

Fig 8 – NeighborNet network reconstructed using standardized POFAD distances derived from nuclear matrices. **A)** Entire network, showing major division between *C. buwata* and the *C. travisae* complex. **B)** POFAD network for *C. travisae* complex. **C)** POFAD network for *C. buwata*.

Fig 9 – Results of STRUCTURE analyses. **A)** ΔK values for K values 1-5. ΔK is maximal at K = 2. **B)** *L(K)* values for K 1-6, maximal at K values 3-5.

Fig 10 – STRUCTURE graphics (resulting from DISTRUCT) for K = 2-5. Each column represents a specimen, grouped by population of origin. Different colors represent different genetic clusters (K); estimated membership coefficients are proportional to bar color height (e.g., for K = 2, the specimen from Apple Riata has a membership probability of 1.0 in the yellow cluster, etc. Figure colors here DO NOT correspond to those in remainder of figures.

Fig 11 – Summary of BPP analyses. **A & B)** Western "undetermined" cave populations allocated to *C. travisae*, eastern "undetermined" populations (Spider Cave, Ken Butler Pit, Jester Estates Cave, Jest John Cave, Beard Ranch Cave) allocated to *C. reddelli*. Guide tree = *C. buwata* sister to (*C. travisae, (C. reddelli* + *C. wartoni*)). **A)** G (alpha = 2, beta = 100) theta prior, **B)** G (alpha = 1, beta = 10) theta prior. **C & D)** Undetermined cave populations removed from analysis, guide tree = *C. buwata* sister to (*C. travisae, (C. reddelli* + *C. wartoni*)). **C)** G (alpha = 2, beta = 100) theta prior, **D)** G (alpha = 1, beta = 10) theta prior.

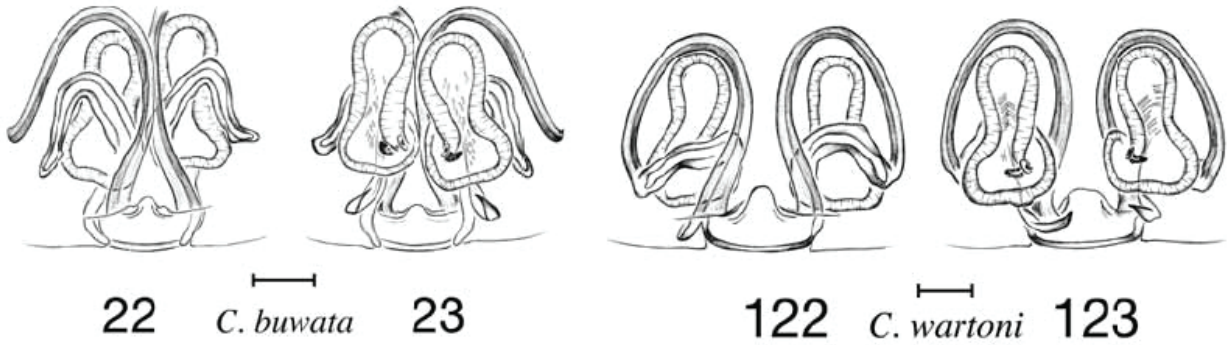
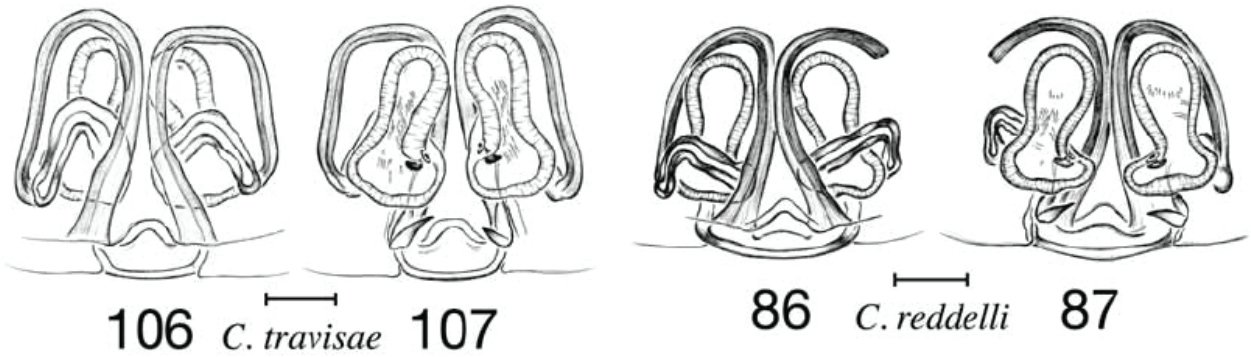


FIGURE 1



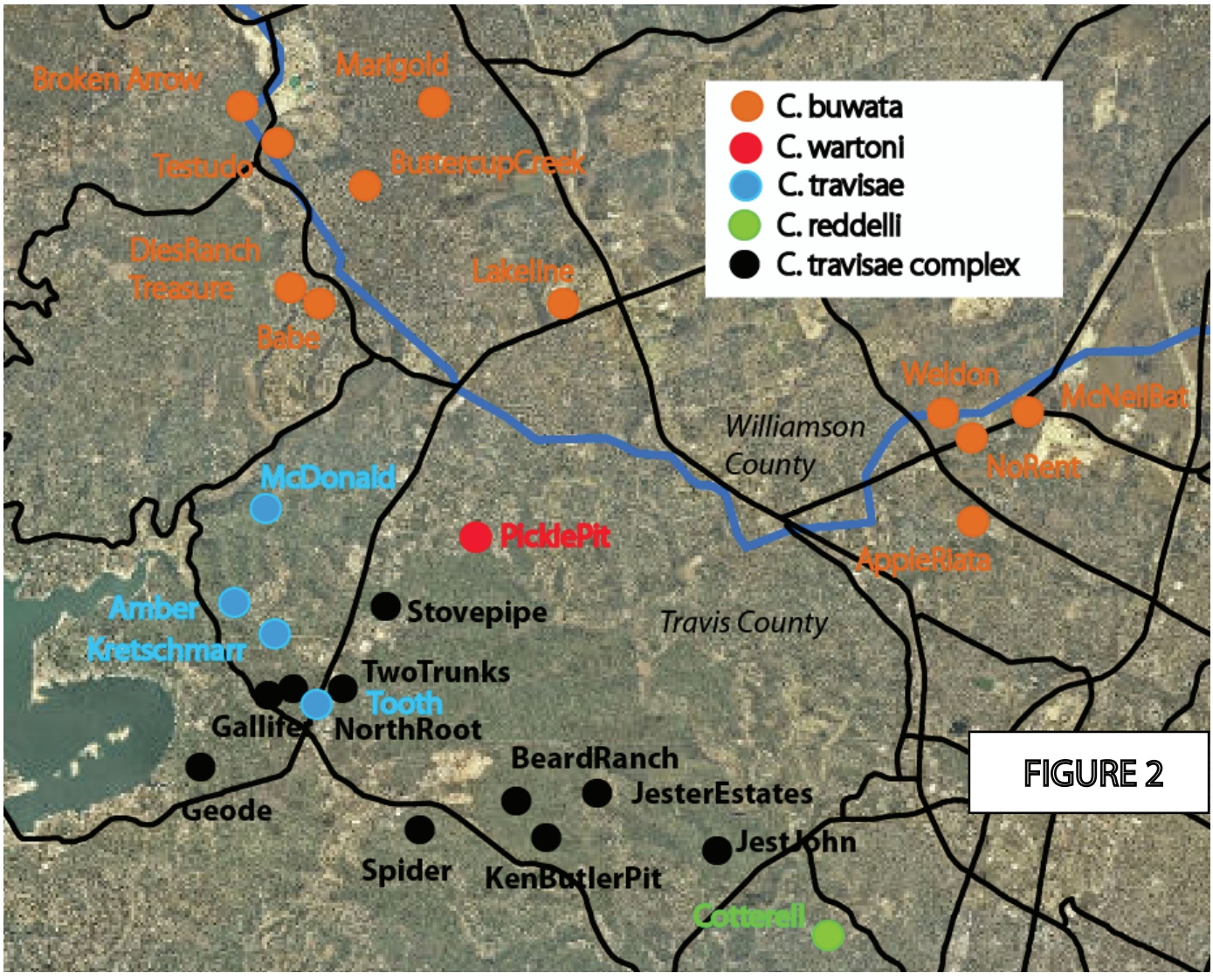


FIGURE 2

FIGURE 3



Broken Arrow Cave G1961

C. buwata "WEST"



Broken Arrow Cave G1959

C. buwata "WEST"



Babe Cave G2011

C. buwata "WEST"



Buttercup Creek Cave G2007

C. buwata "WEST"



Apple Riata G1997

C. buwata "EAST"



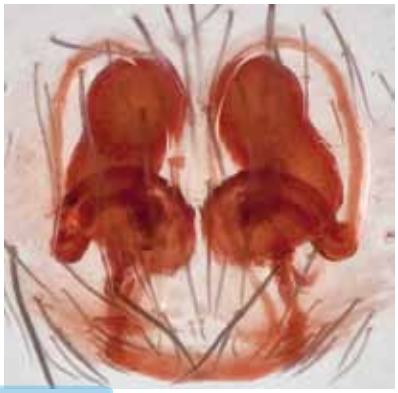
Cotterell Cave G1960

C. reddelli



Jester Estates Cave G1985

C. sp trav complex



Kretschmarr Double Pit G1966

C. travisae



North Root Cave G1970

C. travisae



Tooth Cave G1958



Tooth Cave G2014

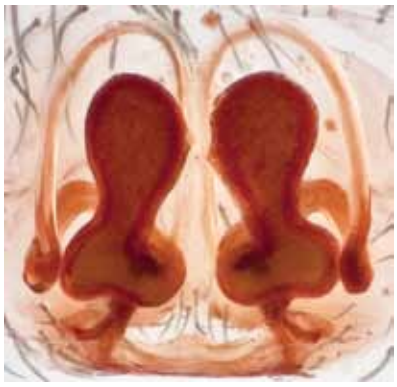
C. travisae

FIGURE 4



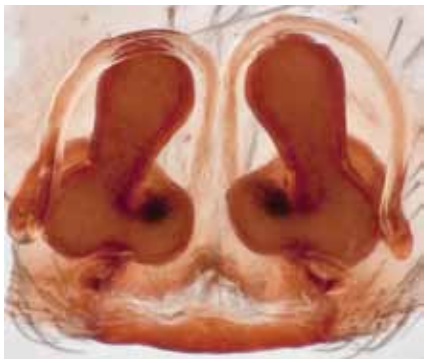
Broken Arrow Cave G1961

C. buwata "WEST"



Broken Arrow Cave G1959

C. buwata "WEST"



Babe Cave G2011

C. buwata "WEST"



Buttercup Creek Cave G2007

C. buwata "WEST"



Apple Riata G1997

C. buwata "EAST"



Cotterell Cave G1960

C. reddelli



Jester Estates Cave G1985

C. sp trav complex



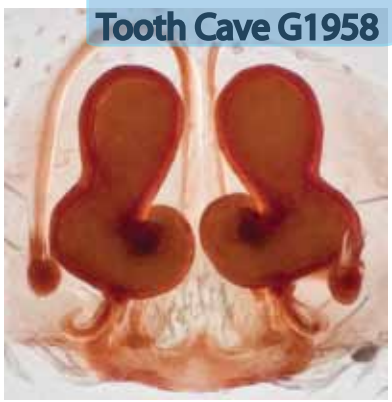
Kretschmarr Double Pit G1966

C. travisae



North Root Cave G1970

C. travisae



Tooth Cave G1958



Tooth Cave G2014

C. travisae

FIGURE 5


Stovepipe
Cave G1977


Spider Cave G1981


Amber Cave
G1998
C. travisae


Jester Estates
Cave G1986

FIGURE 6

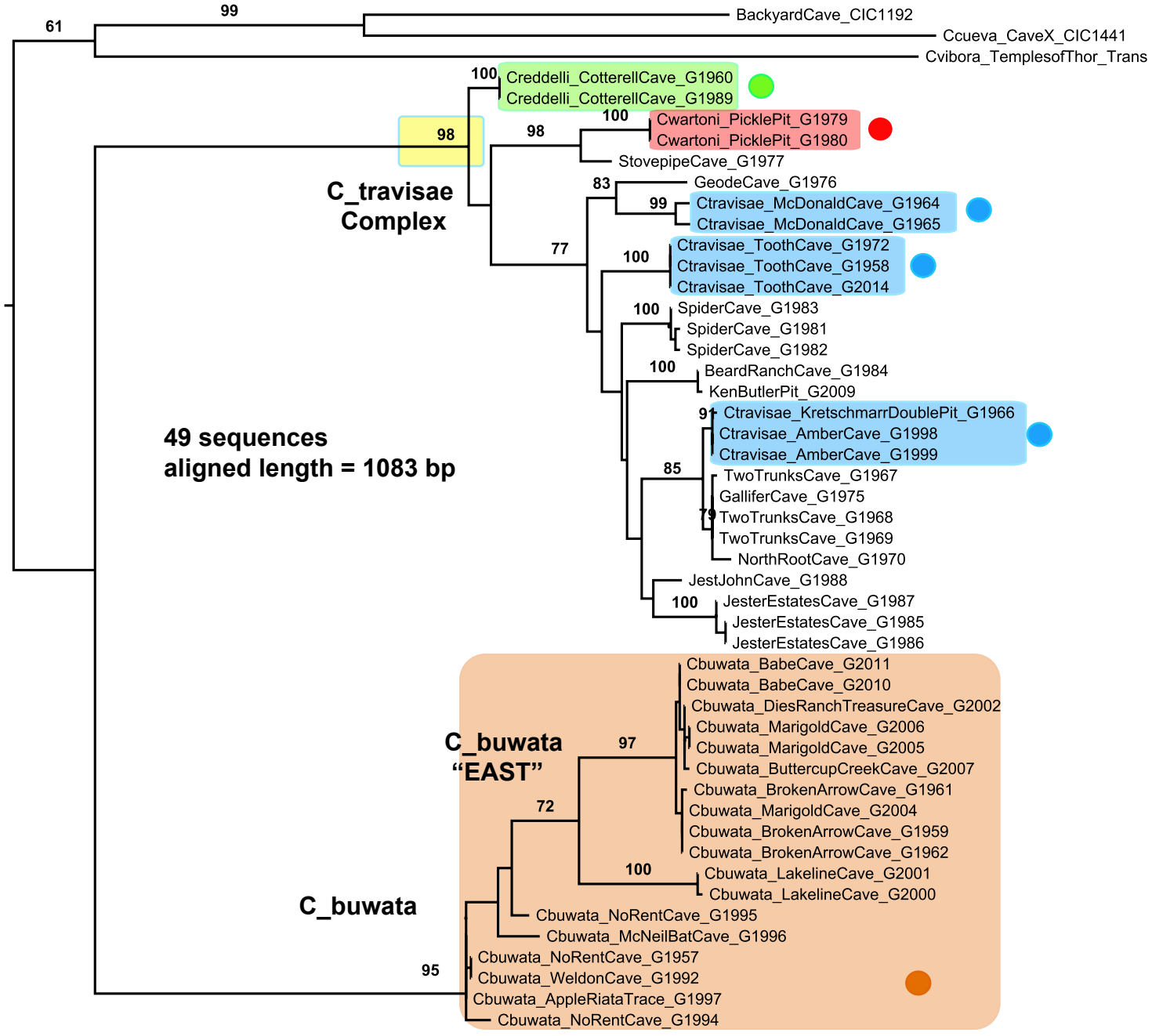

BackyardCave_CIC1192
Ccueva_CaveX_CIC1441
Cvibora_TemplesofThor_Trans
Creddelli_CotterellCave_G1960
Creddelli_CotterellCave_G1989
Cwartoni_PicklePit_G1979
Cwartoni_PicklePit_G1980
StovepipeCave_G1977
GeodeCave_G1976
Ctravisae_McDonaldCave_G1964
Ctravisae_McDonaldCave_G1965
Ctravisae_ToothCave_G1972
Ctravisae_ToothCave_G1958
Ctravisae_ToothCave_G2014
SpiderCave_G1983
SpiderCave_G1981
SpiderCave_G1982
BeardRanchCave_G1984
KenButlerPit_G2009
Ctravisae_KretschmarrDoublePit_G1966
Ctravisae_AmberCave_G1998
Ctravisae_AmberCave_G1999
TwoTrunksCave_G1967
GalliferCave_G1975
TwoTrunksCave_G1968
TwoTrunksCave_G1969
NorthRootCave_G1970
JestJohnCave_G1988
JesterEstatesCave_G1987
JesterEstatesCave_G1985
JesterEstatesCave_G1986
Cbuwata_BabeCave_G2011
Cbuwata_BabeCave_G2010
Cbuwata_DiesRanchTreasureCave_G2002
Cbuwata_MarigoldCave_G2006
Cbuwata_MarigoldCave_G2005
Cbuwata_ButtercupCreekCave_G2007
Cbuwata_BrokenArrowCave_G1961
Cbuwata_MarigoldCave_G2004
Cbuwata_BrokenArrowCave_G1959
Cbuwata_BrokenArrowCave_G1962
Cbuwata_LakelineCave_G2001
Cbuwata_LakelineCave_G2000
Cbuwata_NoRentCave_G1995
Cbuwata_McNeilBatCave_G1996
Cbuwata_NoRentCave_G1957
Cbuwata_WeldonCave_G1992
Cbuwata_AppleRiataTrace_G1997
Cbuwata_NoRentCave_G1994
C_travisae Complex
C_buwata "EAST"
C_buwata
49 sequences
aligned length = 1083 bp

FIGURE 7a

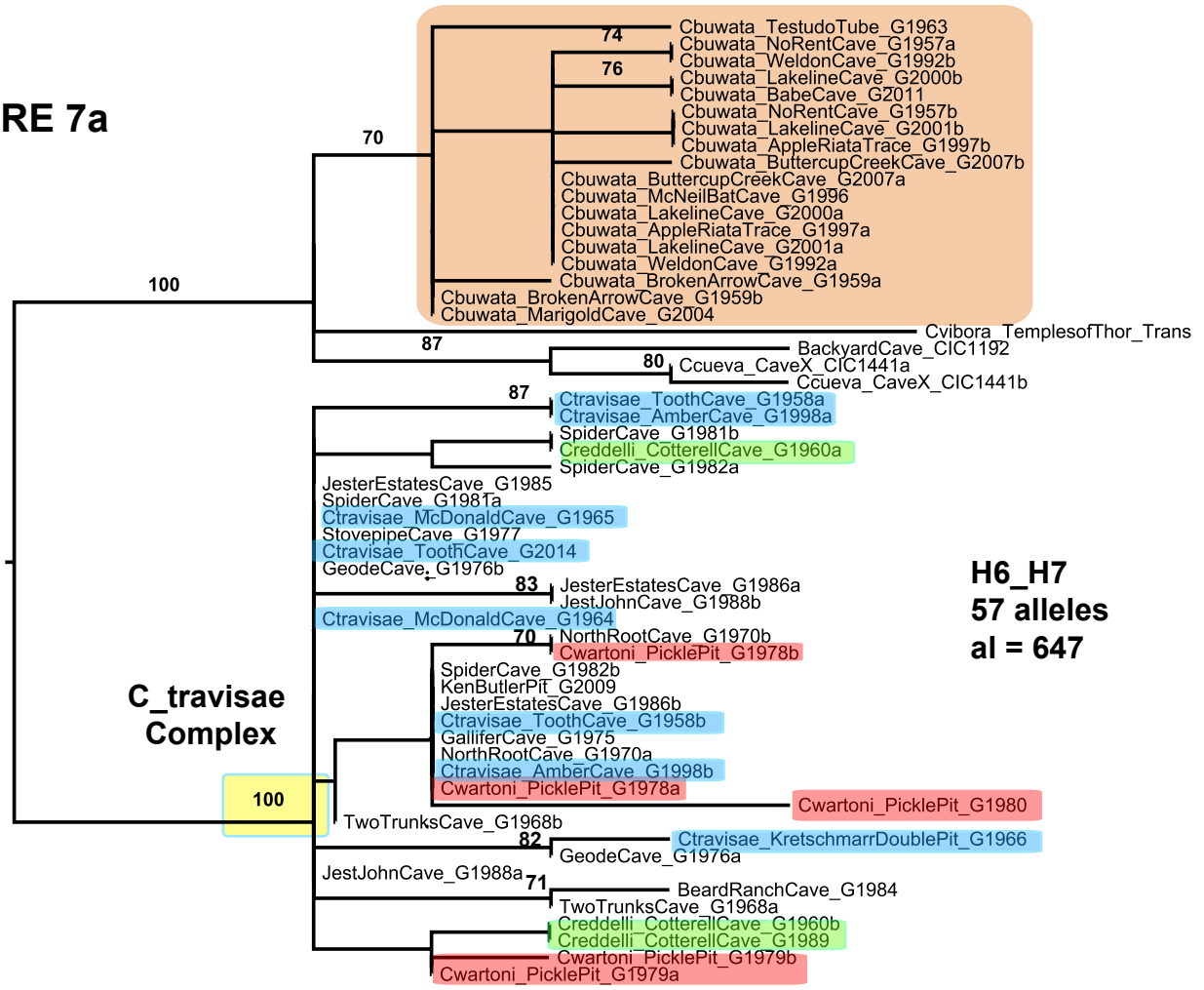
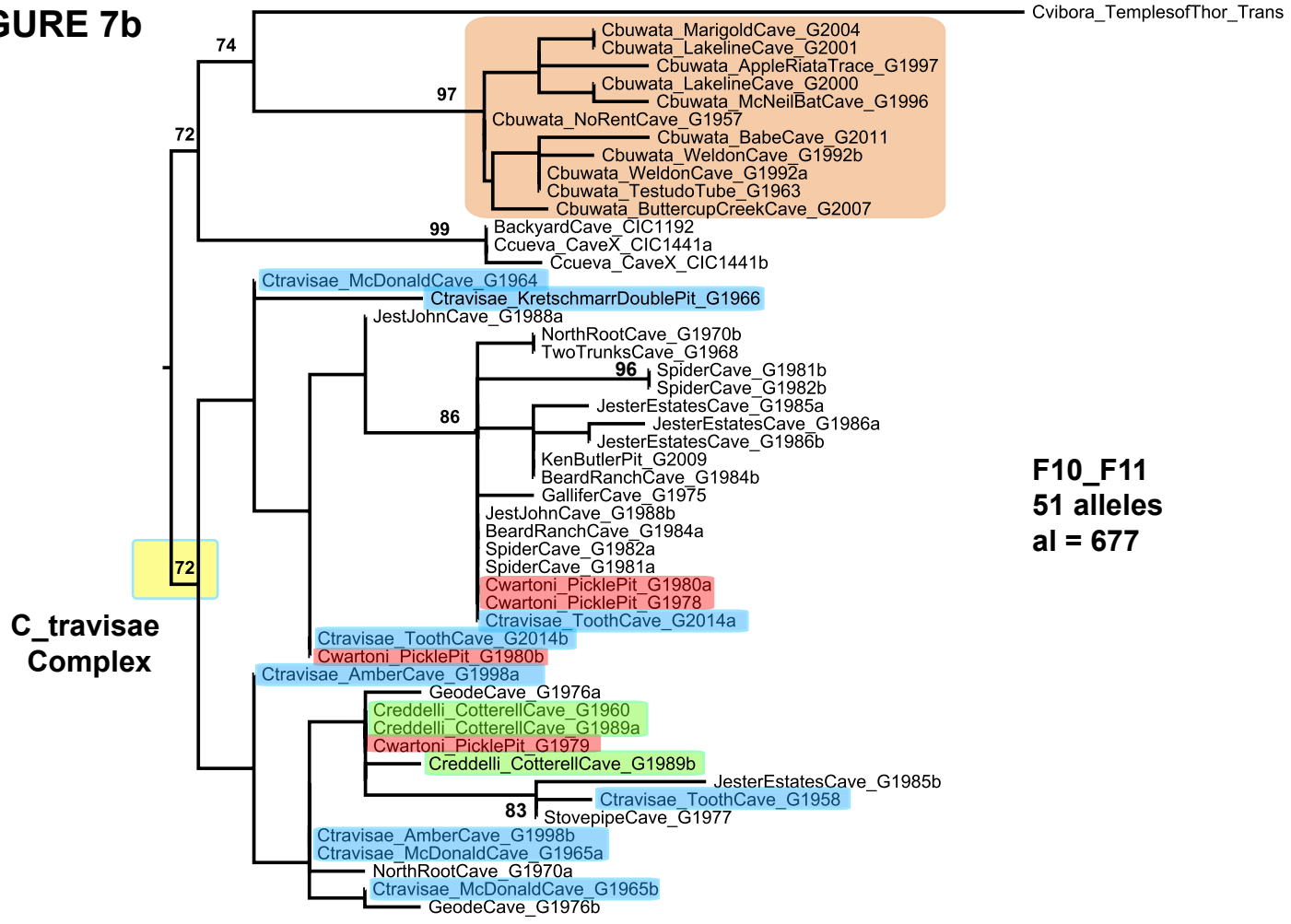


FIGURE 7b



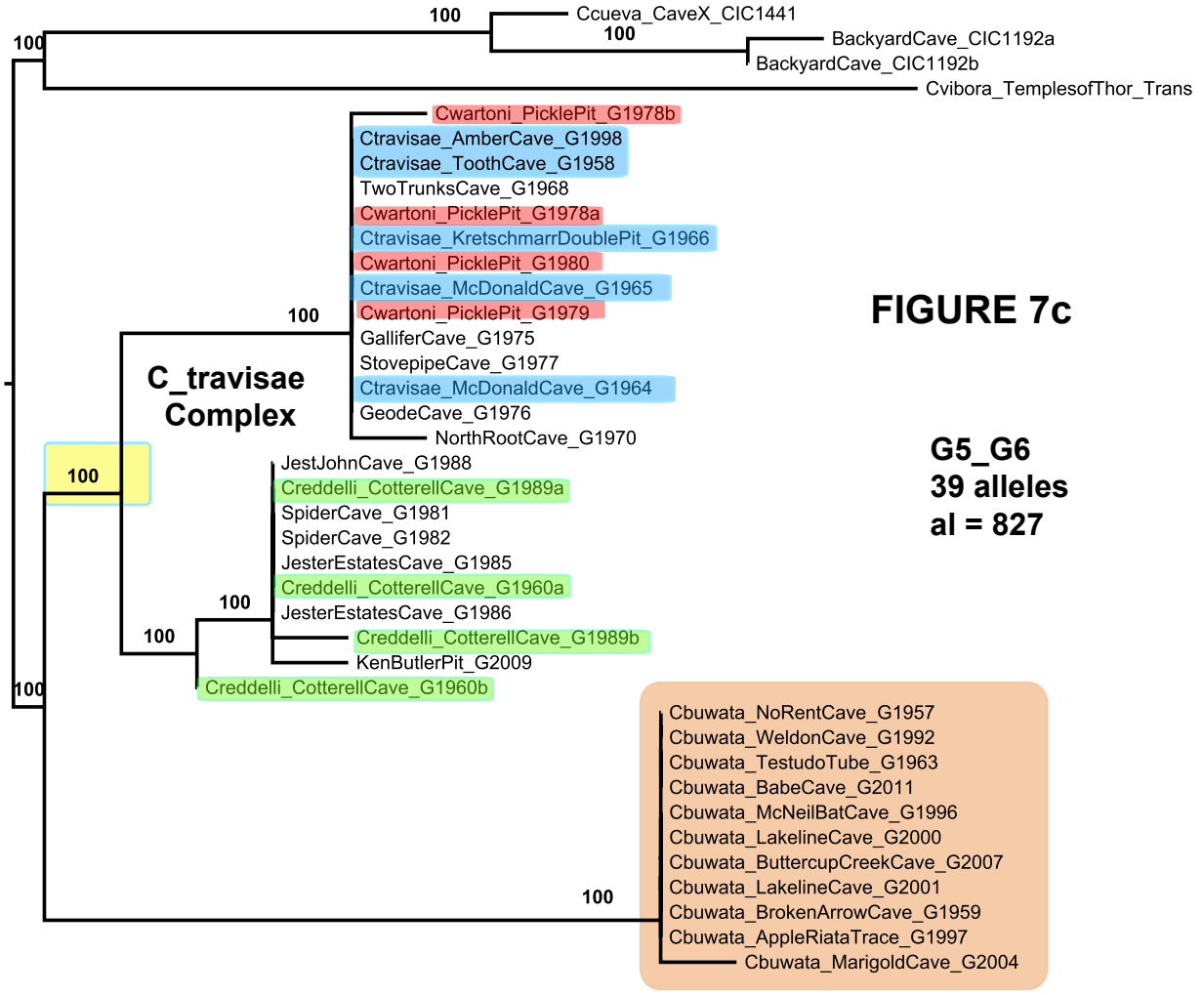


FIGURE 7c

G5_G6
39 alleles
al = 827

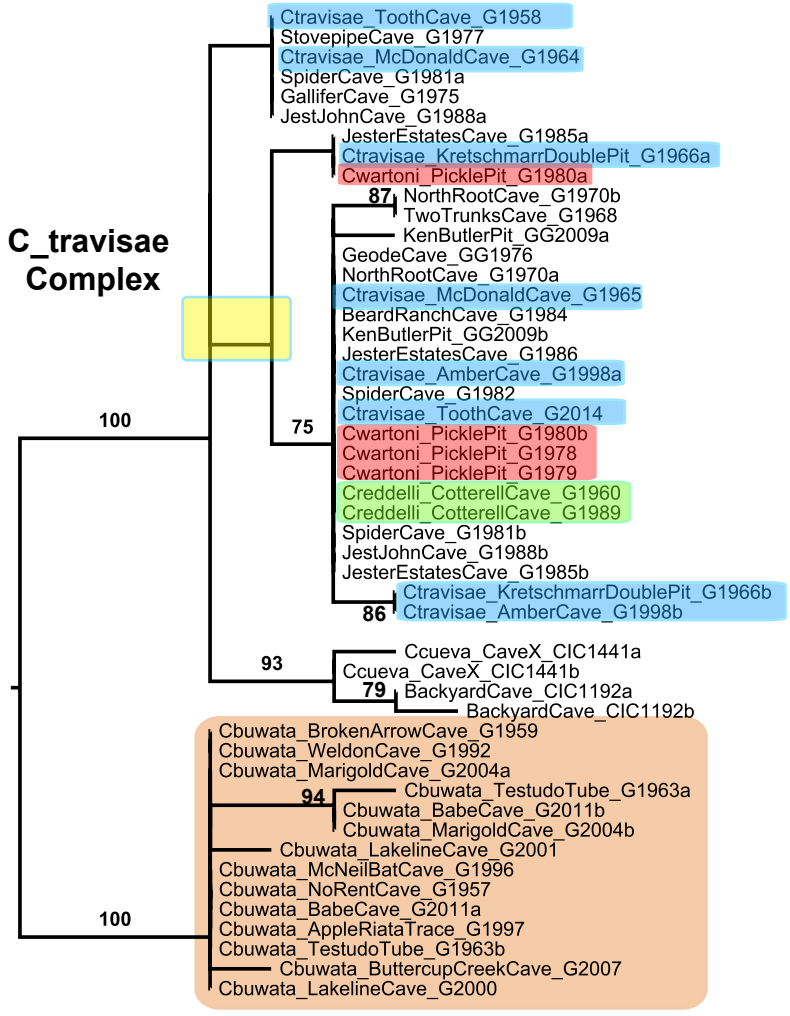


FIGURE 7d

E1_E2
50 alleles
al = 667

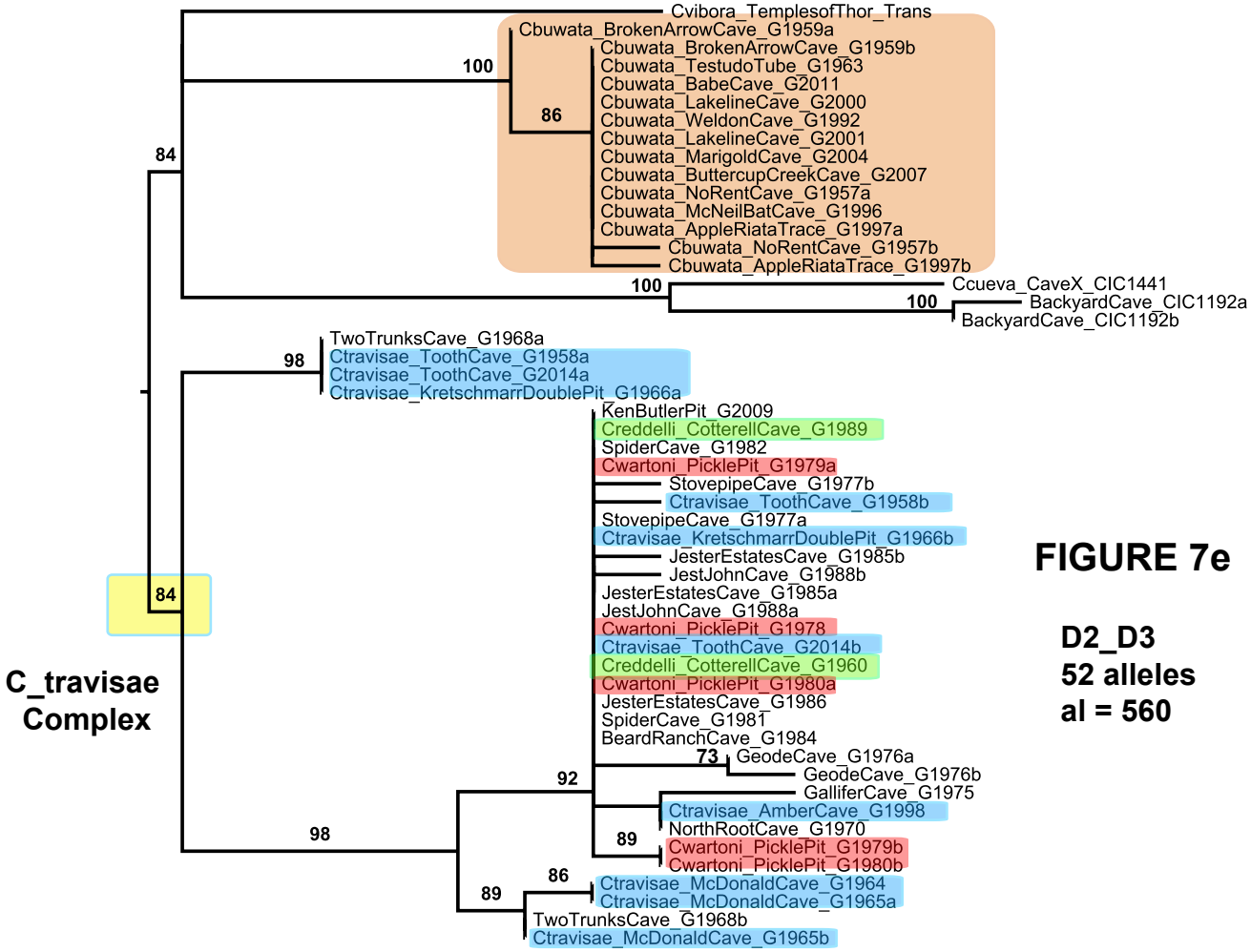
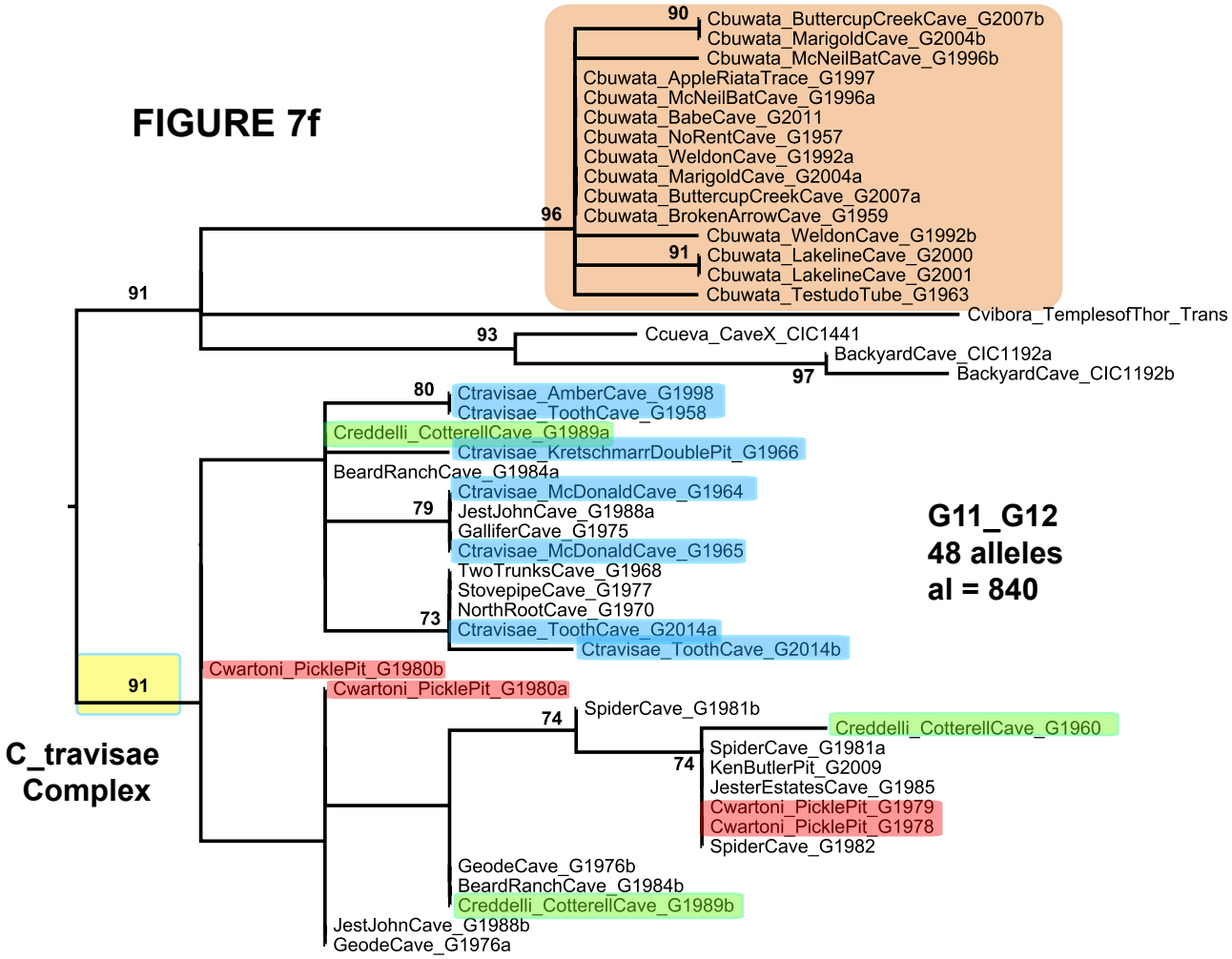


FIGURE 7e

D2_D3
52 alleles
al = 560

FIGURE 7f



G11_G12
48 alleles
al = 840

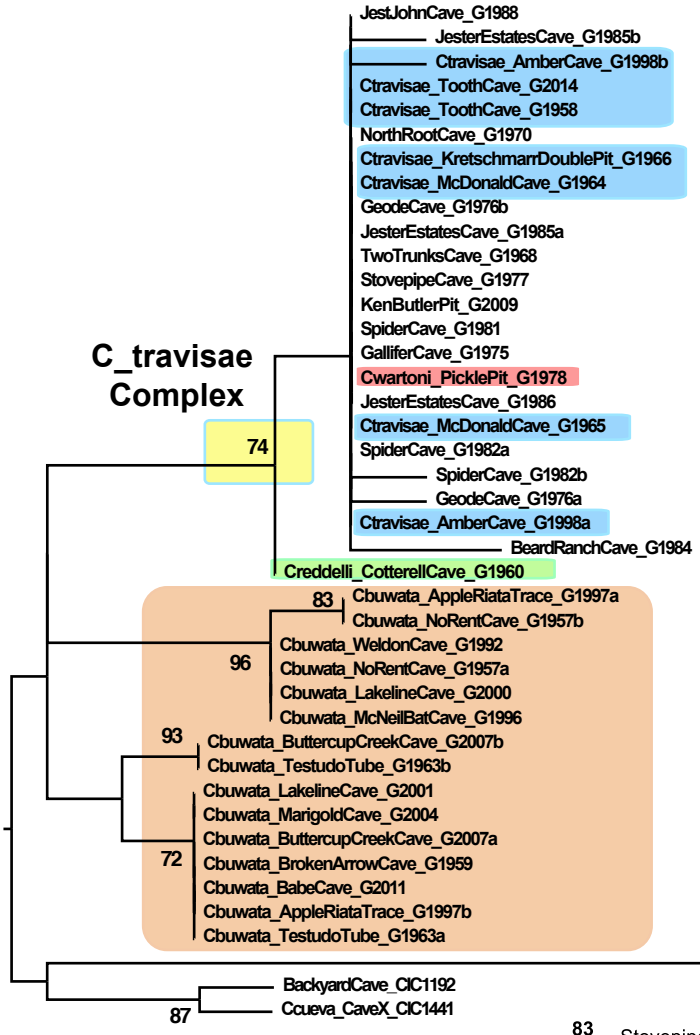


FIGURE 7g

C9_C10
42 alleles
al = 706

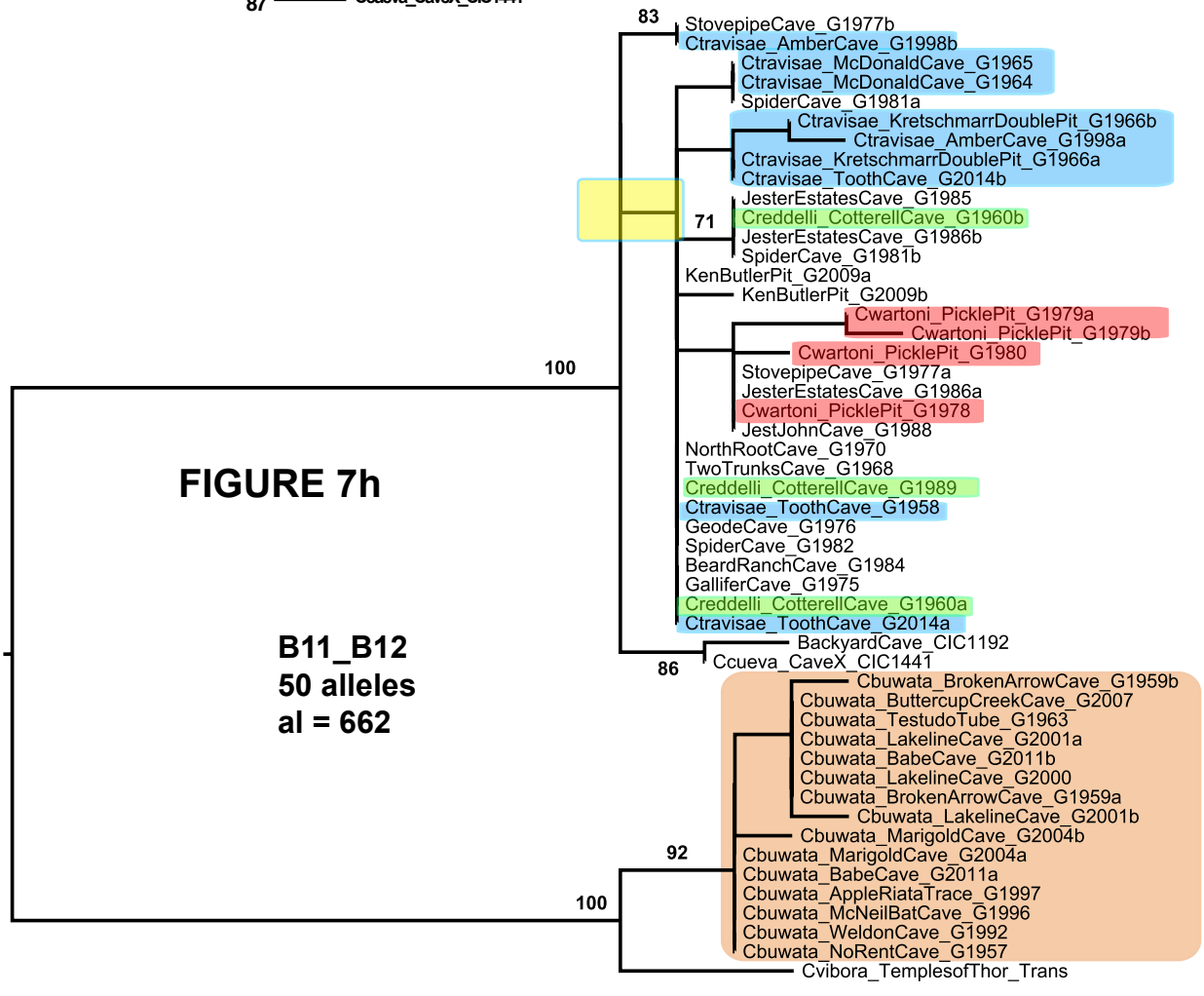


FIGURE 7h

B11_B12
50 alleles
al = 662

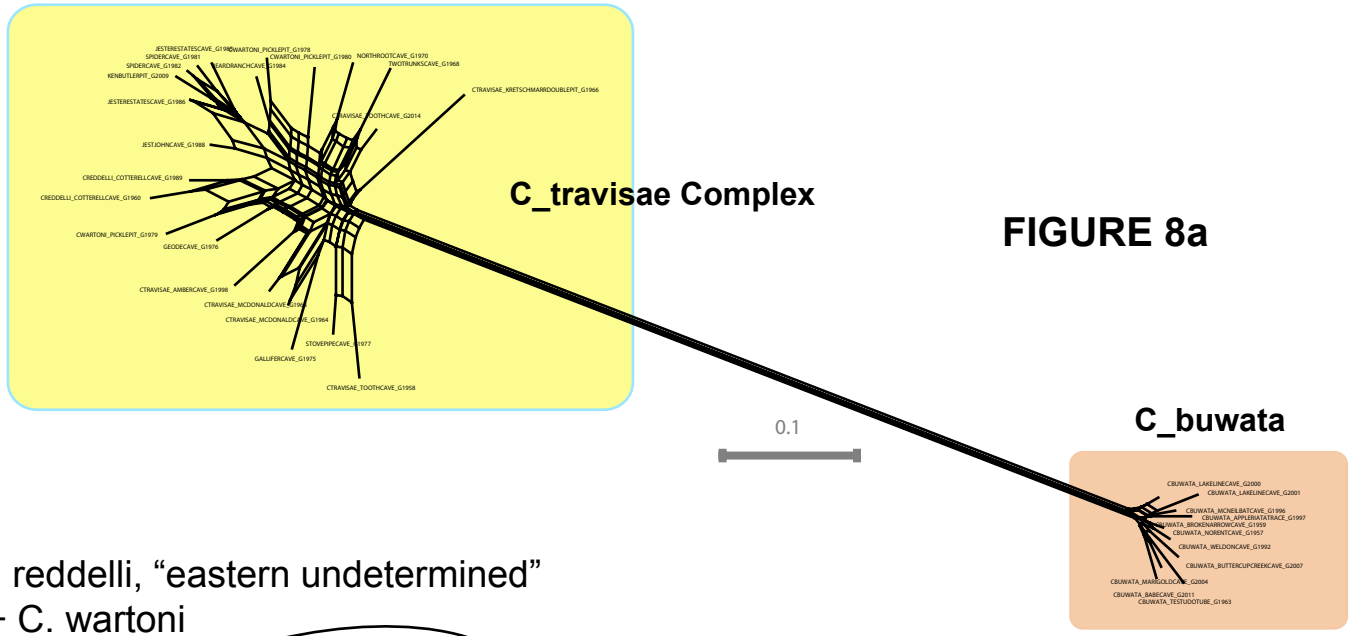


FIGURE 8a

C. reddelli, “eastern undetermined”
, + C. wartoni

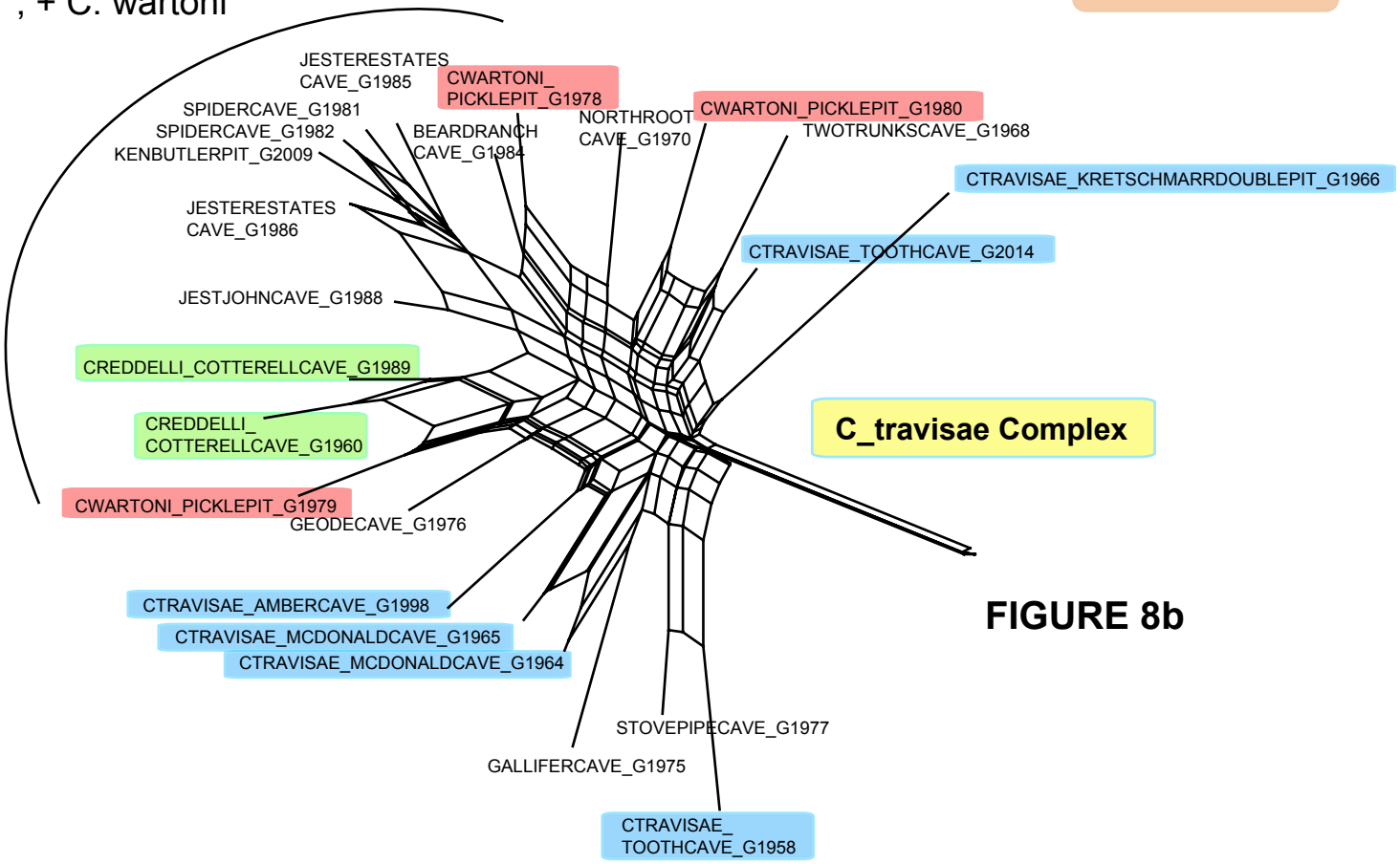


FIGURE 8b

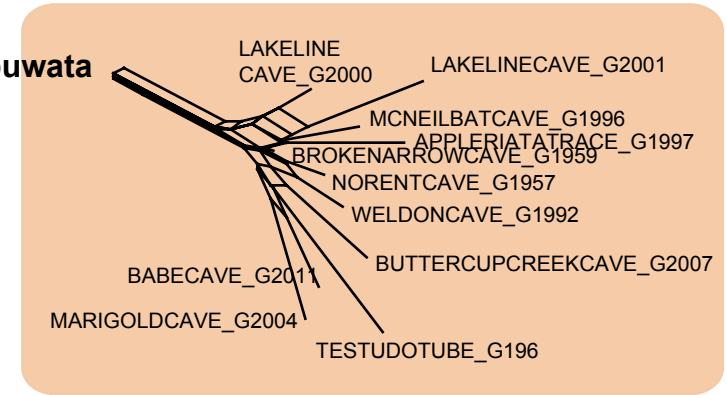
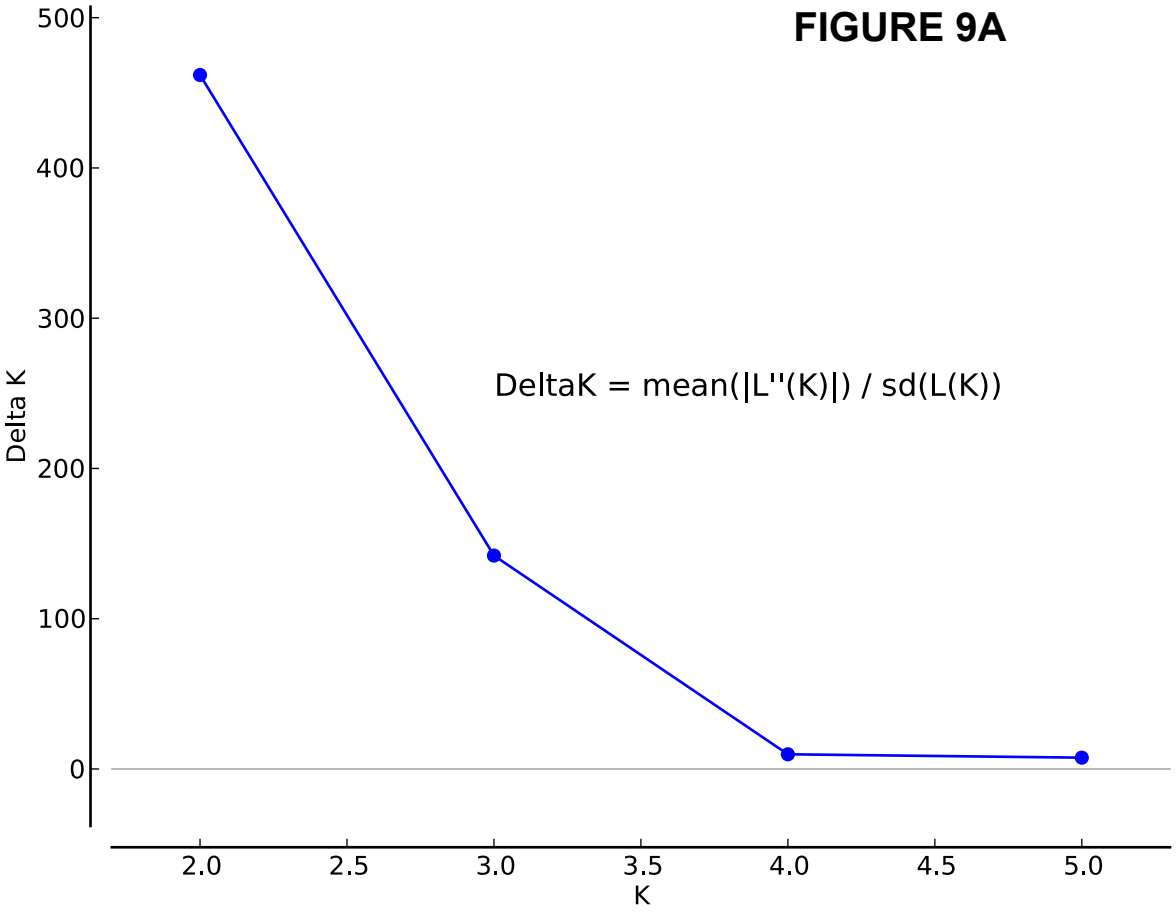


FIGURE 8c

FIGURE 9A



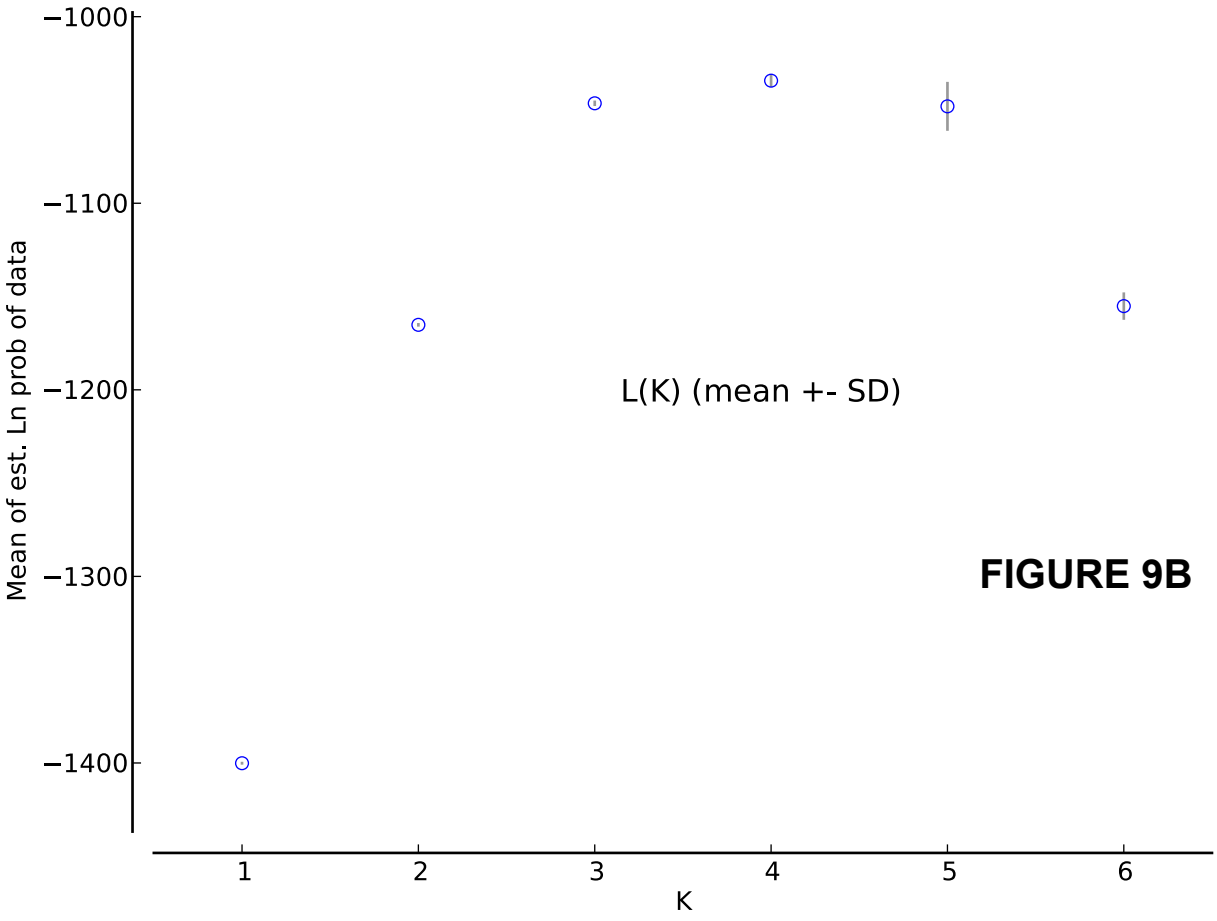


FIGURE 9B

FIGURE 10

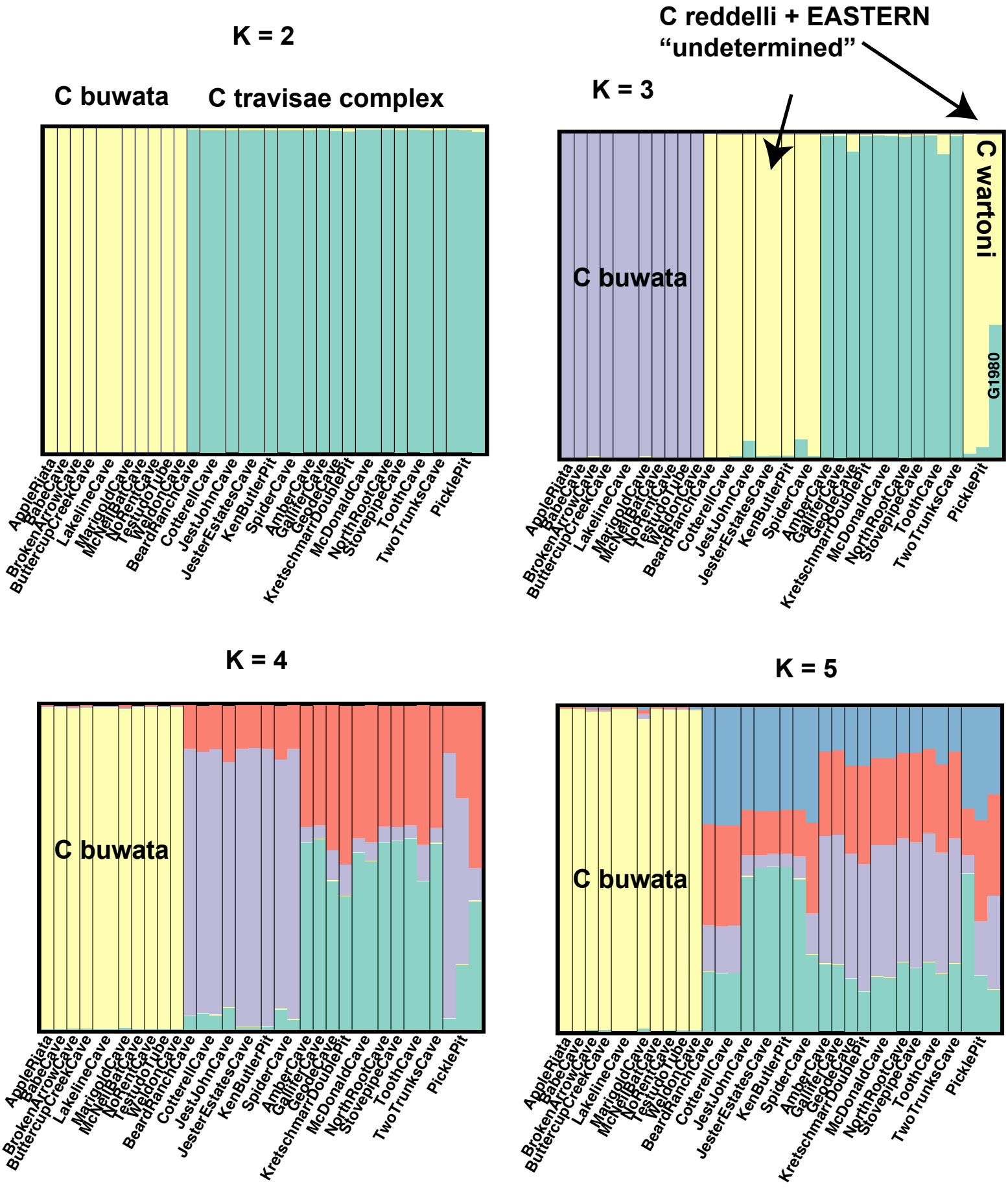
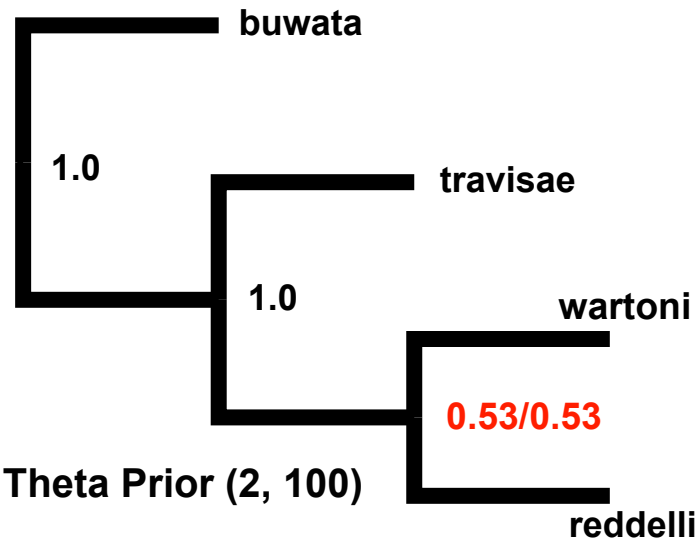
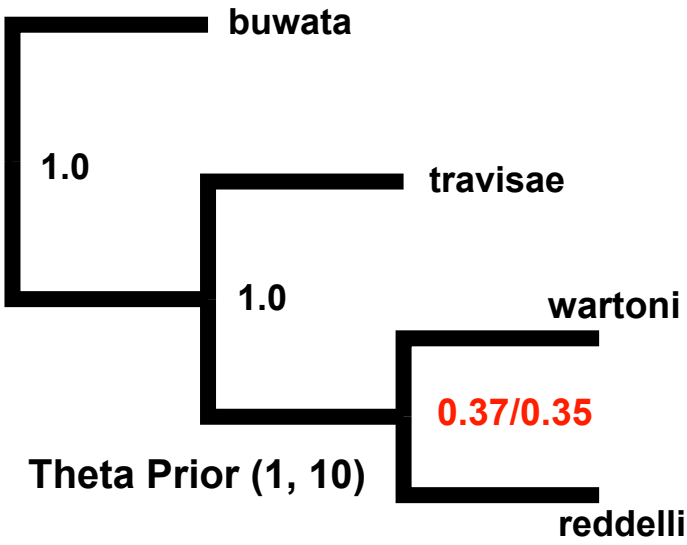

K = 2
C buwata
C travisae complex
K = 3
C reddelli + EASTERN
"undetermined"
C buwata
C wartoni
G1980
K = 4
C buwata
K = 5
C buwata
AppleRiata
BabeCave
BrokenArrowCave
ButtercupCreekCave
LakelineCave
MarigoldCave
McNeilBatCave
NoRentCave
TestudoTube
WeldonCave
BeardRanchCave
CotterellCave
JestJohnCave
JesterEstatesCave
KenButlerPit
SpiderCave
AmberCave
GalliferCave
GeodeCave
KretschmarrDoublePit
McDonaldCave
NorthRootCave
StovepipeCave
ToothCave
TwoTrunksCave
PicklePit

FIGURE 11

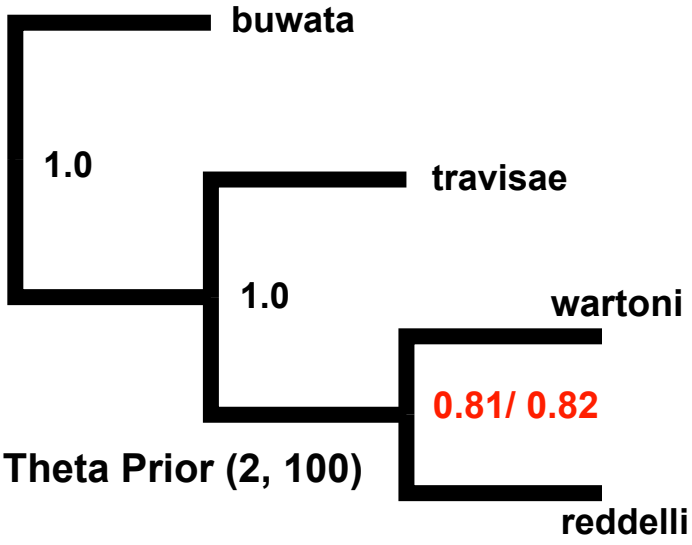
A) Undetermined Pops Removed



B) Undetermined Pops Removed



C) eastern undetermined Pops allocated to C. reddelli



D) eastern undetermined Pops allocated to C. reddelli

