Comparison of genetic variation between rare and common congeners of *Dipodomys* with estimates of contemporary and historical effective population size

Michaela K. Halsey[1,2], John D. Stuhler[2], Natalia Bayona-Vasquez[3,4], Roy N. Platt II[5], Jim R. Goetze[6], Robert E. Martin[7], Kenneth G. Matocha[8], Robert D. Bradley[1,9], Richard D. Stevens[2,9], and David A. Ray[1]

[1] Department of Biological Sciences, Texas Tech University, Lubbock, Texas, United States of America

[2] Department of Natural Resources Management, Texas Tech University, Texas, United States of America

[3] Department of Environmental Health Science, University of Georgia, Athens, Georgia, United States of America

[4] Institute of Bioinformatics, University of Georgia, Athens, Georgia, United States of America

[5] Texas Biomedical Research Institute, San Antonio, Texas, United States of America

[6] Natural Sciences Department, Laredo Community College, Laredo, Texas, United States of America

[7] Department of Biology, McMurry University, Abilene, Texas, United States of America

[8] Department of Biology, South Arkansas Community College, El Dorado, Arkansas, United States of America

[9] Natural Science Research Laboratory, Museum of Texas Tech, Lubbock, Texas, United States of America

* Corresponding Author

Email: david.4.ray@gmail.com (DR)

Abstract

Organisms with low effective population sizes are at greater risk of extinction because of reduced genetic diversity. *Dipodomys elator* is a kangaroo rat that is classified as threatened in Texas and field surveys from the past 50 years indicate that the distribution of this species has decreased. This suggests geographic range reductions that could have caused population fluctuations, potentially impacting effective population size. Conversely, the more common and widespread *D. ordii* is thought to exhibit relative geographic and demographic stability. Genetic variation between *D. elator* and *D. ordii* samples was assessed using 3RAD, a modified restriction site associated sequencing approach. It was hypothesized that *D. elator* would show lower levels of nucleotide diversity, observed heterozygosity, and effective population size when compared to *D. ordii*. Also of interest was identifying population structure within contemporary samples of *D. elator* and detecting genetic variation between temporal samples that could indicate demographic dynamics. Up to 61,000 single nucleotide polymorphisms were analyzed. It was determined that genetic variability and effective population size in contemporary *D. elator* populations were lower than that of *D. ordii,* that there is only slight, if any, structure within contemporary *D. elator* populations, and there is little genetic differentiation between spatial or temporal historical samples suggesting little change in nuclear genetic diversity over 30 years. Results suggest that genetic diversity of *D. elator* has remained stable despite claims of reduced population size and/or abundance, which may indicate a metapopulation-like system, whose fluctuations might counteract any immediate decrease in fitness.

Key words: 3RAD, *Dipodomys elator*, *Dipodomys ordii,* genetic structure, metapopulation, single nucleotide polymorphisms

Introduction

Measuring genetic variation in rare, threatened, endemic, or endangered species has important implications for management and is integral to conservation efforts [1]. Population genetic summary statistics can be used to delimit management units based on significantly different allele frequencies [2], identify population structure [3], or assess connectivity of demographically disparate subpopulations [4]. One such critical measure for small populations is effective population size, N_e [5]. Effective population size can be influenced by fluctuations in census size, mating strategy, biased sex ratios, migration, demographic history, spatial dispersion, and population structure [6-9] and typically is far less than the census size. However, any one N_e value is hard to interpret because it lacks a context. One such context to help understand the potential impacts of fluctuations of N_e is comparison between more restricted, possibly threatened species and a widespread congener, which are presumed to harbor more genetic variation.

The Texas kangaroo rat (*Dipodomys elator*) is a heteromyid rodent that has a limited distribution in north-central Texas [10-14]. Though previously found in two counties in Oklahoma, it appears to have been extirpated from that state [15]. Moreover, *D. elator* has a small geographic range and low dispersal capability [16-17], which increases isolation from nearby subpopulations.

The distribution of the Texas kangaroo rat appears dynamic [18]. For instance, though the species was described from a specimen in Clay County [19], it has not been observed there in more recent surveys. Additionally, resampling of sites where it has been previously documented have failed to detect the species, and new localities of presence have been identified in more recent surveys [20]. Furthermore, previous studies of *D. elator* population genetics [21-22] have been relevant to assess genetic diversity and structure within the species and have established a valuable reference point for *D. elator*. However, these studies relied on few molecular markers (i.e. enzymes, mitochondrial DNA, and microsatellites). Additionally, it is

useful to compare contemporary samples to historical collections to identify potential changes in overall diversity [23-24].

Ord's kangaroo rat, *D. ordii* is a medium sized rodent that occurs from Canada into Mexico [25] Given its large geographic range and preferred commonly available habitat choice (i.e. sandy soils), *D. ordii* is not listed on any state or U.S. federal critically threatened and endangered lists. The population in Canada, however, is listed as endangered [26]. To our knowledge, there has not been a range-wide genetic analysis of *D. ordii*, and the last regional genetic study on *D. ordii* isoenzymes was published by [27].

Here, we compare population genetic parameters of *D. elator* with *D. ordii*, further compare *D. elator* samples from two time periods (pre- and post-2000), and for contemporary samples, investigate differences in genetic diversity across the distribution. We make several predictions: 1) *D. elator* will exhibit a lower effective population size than *D. ordii*, and concomitantly, lower nucleotide diversity, lower observed heterozygosity, and higher inbreeding coefficients; 2) there will be greater genetic diversity among contemporary samples of *D. elator* than in historical samples, as contemporary samples were taken from across the distribution, compared to historical samples collected from three counties in the middle of its distribution; and 3) historical N_e from a coalescent approach for *D. ordii* and *D. elator* will demonstrate that *D. elator* exhibits a lower N_e at present than *D. ordii.*

Methods

Sample collection

Kangaroo rats were captured using Sherman live traps (23x9x8 cm; H.B. Sherman Traps, Inc. Tallahassee, Florida) during surveys within the historical range of *D. elator* (Fig 1) from 2015 to 2017. When a *D. elator* individual was captured, it was either 1) taken as a voucher specimen for deposition at the Natural Science Research Laboratory (NSRL) at the Museum of Texas Tech University or, 2) had between two to four whiskers extracted from either

side of the rostrum [28]. In the latter case, thicker whiskers (i.e., macrovibrissae) were selected with the follicle intact. Whiskers were stored in a sterile vial with 1% sodium dodecyl sulfate (SDS) lysis buffer [29]. A buccal swab was also collected from one *D. elator* individual as described in detail in [28].

Fig 1. Map of contemporary kangaroo rat samples used in this study. Filled stars indicate *Dipodomys elator* samples whereas gray circles represent *D. ordii* samples used in the study. Note the sampling hole located in Foard County, most of Hardman County, and in south Wilbarger County. Trapping restrictions and topography prevented collections in those regions.

Other methods of collecting DNA from rats included tail salvages and from toe clips from museum specimens (Table S1). *D. elator* tail lengths average about 196 mm [30] and at times the end of the tail (i.e., the plume) was severed by the door of an activated Sherman trap. These salvaged tail plumes were placed in sterile vials of 1% sodium dodecyl sulfate (SDS) lysis buffer [29]. Also sampled were toe clips that had been collected from rats from 1986 to 1995 as part of a genetic survey of the species by REM and KGM.

In total, 70 *D. elator* samples were analyzed from five tissue types (i.e, liver, whisker, tail, buccal swab, and toe clips) and two time periods (prior to 2000 and contemporary surveys from 2015 to 2017; Table S1). Additionally, 26 *D. ordii* liver samples were collected in five counties from 2015 to 2017. Contemporary sampling followed guidelines established by the American Society of Mammalogists [31]. Animal handling protocols were approved by the Institutional Animal Care and Use Committee at Texas Tech University (#T14083).

Throughout the manuscript the *D. elator* samples will be referred to using the following descriptors: ‘historical’, collected prior to 2000; ‘contemporary’, collected after 2000; ‘west’, collected from Cottle, Childress, or Hardeman counties; and ‘east’ collected from Baylor, Wilbarger and Wichita counties (Fig 1).

DNA extraction

DNA was extracted using the Qiagen DNeasy Blood and Tissue spin column protocol (Qiagen; Venlo, Netherlands). For liver, toe clips, and tail salvages, the manufacturer's recommendations were followed. For whisker and buccal swab samples, the protocol found in [28] was implemented. In all cases, DNA concentration was fluorometrically quantified using the Qubit 3.0, high sensitivity assay (Invitrogen, Life Technologies, Carlsbad, CA).

3RAD library prep, sequencing, and data husbandry

RADseq libraries were prepared following the 3RAD protocol found in [32] Details of library prep conditions used in this study are provided in Supplemental File S1. In short, restriction enzyme combinations were tested in a subset of samples from both species, and according to digestion patterns and pilot sequence data, the best combination (i.e. MspI, EcoRI, and ClaI), was further used for all samples. Samples were normalized, digested, enzyme-specific adapters were ligated, and ligation products were purified. To generate full- length library constructs, ligated products were amplified using iTru5 and iTru7 primers [33]. For most samples, a molecular ID tag (iTru 5 8N) was incorporated in the first cycle of PCR, to detect PCR duplicates [32, 34]. PCR products were purified, pooled, and size-selected at a range of 550 bp +/- 15%. Size-selected fragments were purified and sequenced using an Illumina HiSeq to generate paired end data at Oklahoma Medical Research Foundation Genomics Core or Novogene Inc.

Stacks v1.48 and v2.01 [35] was employed to demultiplex, analyze, and export data into other formats. After demultiplexing, poor reads were filtered using the AfterQC 'after.py' pipeline [36]. Poor reads were defined as exhibiting a low quality score (PHRED score < 15), bad overlaps (i.e., mismatched reads), too many ambiguous nucleotides (greater than 40% of the read), short read lengths (< 35 base pairs), or homopolymer regions. If a read failed one of these steps, it was removed from downstream analyses. Reads were aligned within Stacks to

the *D. ordii* genome assembly (accession ID GCA_000151885.1) using the Burrows-Wheeler aligner [37].

Data were grouped into putative loci, and polymorphisms were identified with the 'gstacks' module in Stacks. Common population genetic statistics such as observed and expected heterozygosity, nucleotide diversity, and inbreeding coefficients were calculated using the 'populations' module. This step was repeated four times to determine a balance between data used and processing speed. Though the "gold standard" is to include loci where 75 to 80% of the individuals in a population have that locus, known as the -r value [35, 38], this has been shown to bias population genetic measures, especially in cases where data are not plentiful. This influences biological implications [39-42]. The 'populations' module using this 75% rule (-r 0.75), two liberal filters (-r of 0.25 and 0.5) and a more conservative filter (-r 0.95) was run. For most downstream analyses, -r 0.75 was used as the main dataset and for comparison across -r values for any differences. Finally, loci and individuals that had greater than 20% missing data were removed.

Population genetics

Observed and expected heterozygosity were calculated using the 'summary' function in the R package adegenet, version 2.1.1 [43]. F_{IS} and F_{ST} values were calculated using hierfstat [44]. Nei's genetic distances [45] were determined and plotted using the 'aboot' function in the poppr R package [46].

Estimation of effective population size

To determine effective population size using NeEstimator v2.1 [51], the Genepop file [52] generated by Stacks was used on our contemporary dataset. NeEstimator calculates Ne using three methods: linkage disequilibrium, molecular co-ancestry, and a temporal method. The first two methods were used to determine contemporary effective population sizes per species.

For historical Ne of *D. elator*, the Extended Bayesian Skyline Plot (EBSP) coalescent test as implemented in BEAST 2.0 [53] was used. Once loci containing multiple single nucleotide polymorphisms were determined, the protocol outlined in [54] was followed, using a strict molecular clock set to 1.0 and a generation time of 3 years [55]. Plots were constructed with 24 individuals and 47 loci.

This same process was followed for *D. ordii*; however, 15 individuals and 49 loci were used. Only one individual from Dickens County (Fig 1) was included to avoid misinterpretations due to possible inbreeding since many individuals collected from that county were collected from a single location.

Population structure

To infer population structure for each species, the STRUCTURE algorithm was used [47]. All singletons and private doubletons were removed, which have been shown to mask weak population structure [48-49]. Only one randomly selected SNP from each locus was used to minimize possible effects of linked data. For all runs, 50,000 burn-in iterations were executed and 200,000 Markov Chain Monte Carlo (MCMC) repetitions with 3 replicates at each K, which ranged from 1 to 5. The program DISTRUCT v1.1 was used to visualize the final output of structure analyses [50].

Principal Components Analysis

To visualize genetic structure of the population without assigning individuals to clusters *a priori*, a Principal Components Analysis was conducted using the function dudi.pca in the R package 'adegenet' version 2.1.1 [43] on historical and contemporary samples. Only the first two axes were retained for all datasets based on the scree plots generated by glPca.

Results

In all, 96 kangaroo rats were sequenced and analyzed from two species in eight counties in north-central Texas. 3RAD analysis for 70 individual *D. elator* samples produced over 34 million reads. Before filtering within the 'populations' module, there were 330,326 loci suitable for analysis. Approximately 1.5% of reads per sample were removed from the dataset following AfterQC filtering. Similar analysis for the 26 *D. ordii* samples produced over 10 million reads. Prior to 'populations' filtering, approximately 382,514 loci were eligible for further analysis. Fewer than 2% of reads per sample were removed from the dataset. For *D. elator* samples, after removing loci and individuals that had greater than 20% missing data, 3.935 samples remained from 55 individuals.

Summary population genetics

There were as few as 7 single nucleotide polymorphisms (SNPs) to as many as 61,000 SNPs to analyze (Table S2). The general trend was that there were fewer SNPs analyzed as -r value increased and extreme values of -r (i.e. 0.25 and 0.95) yielded stronger deviations between observed and expected heterozygosity across all analyses (Table S2).

When compared to each other, contemporary *D. elator* samples showed lower levels of observed heterozygosity (0.042-0.043) than *D. ordii* (0.128) which suggests lower genetic diversity. F_{IS} was positive in all *D. elator* groups (0.015-0.031) except for in historical samples (-0.007). F_{ST} values show moderately low genetic differentiation (0.035-0.041) across *D. elator* comparisons (Table 1).

Table 1. Genetic diversity summary statistics for *D. elator* and observed heterozygosity and expected heterozygosity values for *D. ordii*. Only one population is assumed for *D. ordii*, so there are no values for F_{ST} or F_{IS}.

	F_{IS}	Observed Heterozygosity	Expected Heterozygosity
Temporal			

Historical	-0.007	0.042	0.038
East	+0.031	0.043	0.043
West	+0.027	0.042	0.040
Spatial			
East	+0.015		
West	+0.031		
Dipodomys ordii		0.128	0.155

Pairwise F_{ST}		
	East	West
Temporal		
Historical	0.035	0.037
East		0.040
Spatial		
East		0.041

Current and historical effective population size

Only the linkage disequilibrium method in NeEstimator v.2.1 produced a value other than 'Infinite' for effective population size for *D. elator*. Estimated N_e of the east group 171.3, with a 95% confidence interval of 158-186.9 using the lowest allele frequency. For the west group, both the linkage disequilibrium and molecular co-ancestry methods returned 'Infinite' for N_e at all allele frequencies. No method with NeEstimator was able to provide an estimate of population size for *D. ordii*, other than 'Infinite.'

The Extended Bayesian Skyline plots generated for the *D. elator* dataset showed a decline in effective population size over the last 10,000 years, to an approximate current N_e of 500. For *D. ordii*, N_e has increased in the last 5,000 years and is estimated to stand at about 10,000 individuals (Fig 2).

Fig 2. Extended Bayesian Skyline Plot for *Dipodomys elator* (top) from 34 individuals and 47 SNPs and for *D. ordii* (bottom) from 15 individuals and 49 SNPs. X-axis is millions of years ago. Y-axis is effective population size (Ne) in millions.

Population substructure

Based on log-likelihood scores (Table S3) and their respective variances from STRUCTURE, the "best" value for k for *D. elator* was 3. When visualizing the PCA bi-polot, PC1 accounts for almost 98% of the variation found in the dataset and shows geographic separation along PC2, which only accounts for 0.1% of the variation (Fig 3). Using Nei's genetic distance, most contemporary samples from the west cluster together and are nested within historical samples (Fig 4). The historical PCA for *D. elator* samples excluding those from the 1960s (Fig S1) largely confirms that all individuals were taken from the same region (Hardeman County).

Fig 3. Principal components analysis on the genotypes for 55 samples (historical and contemporary) using the dudi.pca function in R package 'adegenet'. While there are no clear clusters emerging on PC1, geographic location seems to correspond with PC2.

Fig 4. Nei's genetic distance dendrogram for 55 samples (historical and contemporary). The patchy arrangement of individuals suggests gene flow between the hypothesized east and west populations.

Discussion

This study evaluated changes in genetic diversity across time and space by comparing a rare species with a hypothesized amorphous and restricted distribution to a more common congener with a larger, more defined range. This is only the third population genetic study on *Dipodomys elator* in over 30 years and it is the first to make use of genomic techniques, screening from tens to thousands of markers, making the study valuable for current and future conservation efforts. In [21], allozyme markers were used to conclude that there was moderate genetic differentiation among three *D. elator* localities (Hardeman, Wilbarger, and Wichita counties). This is seemingly incongruent with our results in which we observed little genetic

differentiation (F_{ST} = 0.041), but the difference could simply be the result of the markers used (SNPs versus allozymes).

More recently, [22] observed low mitochondrial DNA variation but high microsatellite diversity within the species. They concluded that genetic drift and not gene flow has had a greater impact on configuring *D. elator* genetic diversity. This result is possible because mitochondrial DNA has a lower effective population size than neutral nuclear markers such as RAD loci [56]. Genetic drift could play a role in structuring mitochondrial DNA diversity, but more time would be needed to detect reduction of diversity in the nuclear genome using older markers such as microsatellites. An insufficient number of polymorphic microsatellite loci limits genetic resolution between individuals with supposed low population-level diversity. Our results suggest that RAD loci, that have a slower rate of mutation than microsatellites, are superior when investigating populations with weak population structure [57]. Finally, genomic data generated from this study can be used for future genomic investigations, such as those examining family structure [58].

Together, these three studies, using allozyme, mtDNA, microsatellite and RAD-Seq markers, offer numerous mean geographic estimates of F_{ST} within this species. In [21], the mean F_{ST} was found to be 0.102, [22] estimated F_{ST} of 0.096 from their late 1960s samples, and our study, at the greatest resolution of all previous studies, reveals a mean F_{ST} value of 0.041. Our lower mean value includes individuals sampled from localities (Cottle and Childress counties) not present in the previous two studies. These results suggest modest population differentiation corresponding with geography.

From an overall genetic diversity perspective, our data suggest that there has not been a substantial loss in genetic diversity over the last 30 years, despite what seems to be a decrease in the distribution (and possibly abundance) of *Dipodomys elator,* similar to what [59] found in *D. ingens*, the giant kangaroo rat. In other words, despite a decline in distribution and census size, the genetic diversity of the species is sufficiently high to offset any short-term effects of reduced

296 fitness. This is supported by our estimate of N_e of between 170 to 500, which exceeds the
297 recommended value to curtail inbreeding depression, as outlined by the 100/1000 rule [60].
298 Within this range, there exist enough individuals to mitigate immediate reduction in fitness but is
299 not sustainable in the longer term (over thousands of years). In [22], it was also found that the
300 N_e of this species was between 65 and 490 individuals.

301 Results from our contemporary samples confirm that subpopulation differentiation is not
302 substantial ($F_{ST} < 0.05$). The STRUCTURE algorithm determined the best value of k to be 3.
303 However, examining the plots suggests that samples represent a single interbreeding
304 population. More clusters (i.e. subpopulations), while possible, are not biologically practical. This
305 may just be an artefact of our sampling scheme (for example, k=5, one for each county).
306 Second, newly colonized subpopulations on the fringes of ranges can exhibit lower levels of
307 genetic diversity than expected [61]. For our contemporary samples, this is not the case; the low
308 mean value of F_{ST} (< 0.05) does not seem to support cluster sizes of k=4 and k=5. Based on
309 climate, vegetation, edaphic, and land use characteristics across the study area [20], our *a priori*
310 assumption was that there are two subpopulations (east and west). However, STRUCTURE, the
311 PCA, and Nei's genetic distances do not clearly support two distinct subpopulations, suggesting
312 there is a fair amount of gene flow in the region.

313 Our *a priori* subpopulations display low levels of inbreeding and very little genetic
314 differentiation, suggesting one large interbreeding population (though not necessarily
315 panmictic). Our samples were collected on opposite sides of a cline, separated by a region of
316 inaccessible private land, so it was difficult to determine if the slight differentiation is due to that
317 distance or if there is true population substructure and isolation from other habitat patches [62].
318 We included additional historic samples from specimens collected in the 1960s from areas
319 within this "sampling hole" to answer this question. We anticipated that if the contemporary east
320 and west subpopulations were indeed distinct, then genetic differentiation would be greater
321 between them than to the samples from the sampling hole. In other words, a STRUCTURE plot

would show the sampling hole samples as intermediate between the two. Alternatively, if the contemporary east and west subpopulations were considered one population then we would expect greater genetic differentiation between them and our "sampling hole" samples. Our results support the second prediction (Fig S2). However, the periods separating the datasets (anywhere between 20 and 50 years) and the relatively short generation times of kangaroo rats (about 3 years; Pacifici et al. 2012) would lead to high genetic turnover, so these results must be interpreted with caution. If there is substantial genetic turnover, this too could indicate small current effective population size, which supports our estimate of approximately 171.

As expected, our *D. ordii* samples exhibited higher genetic diversity estimates in nearly all categories despite our samples being collected from only five counties in north-central Texas. This emphasizes the substantial genetic diversity and evolutionary potential displayed by the common *D. ordii*, compared with a much rarer congener. However, we were surprised to find that *D. ordii* had a greater inbreeding coefficient than *D. elator* across some analyses. This pattern can be attributed to sampling bias, given that we sampled from a small portion of the *D. ordii* range, and half of the *D. ordii* samples were collected from a single ranch in Dickens County, Texas, where most individuals collected may present a certain degree of relatedness by proximity. Comparing between individuals from this ranch and a similarly situated subset of *D. elator* individuals, expected heterozygosity, π, and inbreeding coefficients were largely similar (Table S4). This suggests that potentially related individuals of *D. elator* do not show reduced genetic diversity than similarly related *D. ordii* individuals.

We were unable to generate a value of N_e for the current sample of *D. ordii*, likely a result of samples displaying high degrees of relatedness, so we used the value calculated from EBSP, which was approximately 10,000 individuals. In contrast to *D. elator* N_e, which declined over time, the plot for *D. ordii* increased, perhaps an indication of colonization of new habitat (northward) as glaciers receded after the Last Glacial Maximum 20,000 years ago [63-64].

Coupled with low N_e estimates, and population surveys that recover or fail to locate *D. elator* in different localities, one possibility is that this population exhibits characteristics of a metapopulation [65-66]. Metapopulation theory has been discussed in the context of mammalian conservation biology because it accommodates populations in fragmented habitats [67], but empirical studies to develop metapopulation theory for threatened and endangered mammals are few (see [68-69]). One reason for the difficulty to meet the original metapopulation criteria of [70] is the stringency of the original criteria. In [71], the authors relaxed two criteria, adding that subpopulations, not the colonized habitat patch, are the discrete entity, and that these discrete subpopulations differ in their demography, implying asynchronicity. Based on field surveys, analysis of field notes, museum specimens, and species distribution models [20] there is evidence that the *D. elator* population may benefit from management consideration stemming from metapopulation theory.

However, because this connection to metapopulation theory is still tenuous, the overall population should still be monitored [72]. Perhaps a long-term demographic study is warranted. Managing the metapopulation must be concerned with maintaining dispersal and gene flow and other population dynamics among subpopulations. Should managers elect for extreme measures to manage *D. elator* populations, such as translocations or reintroductions, knowledge that the population is a metapopulation is critical. Lastly, it is important to note that the metapopulation in a conservation context has several assumptions. One assumption is the "equilibrium" between colonization and extinction across long time scales (i.e. if one patch goes extinct, another is colonized). This seems unlikely in many natural populations [73], including that of *D. elator*, but this type of assumption can be used to appropriately model changes in demography and genetics of *D. elator.*

There is no lack of research on habitat associations, mainly those evaluating soil and vegetation changes, as they influence *D. elator.* These studies have greatly improved our understanding of this elusive rodent [16-18, 74-75], but we still do not have answers to many

basic biological questions. We do know, however, that the population of *D. elator* seems to track favorable habitat, albeit in a more restricted range than previously recorded [18].

Overall, the population of *D. elator* exhibits genetic variation lower than that of a species with a predictably greater effective population size. However, contemporary samples show no substantial decrease in genetic diversity from historical samples, suggesting that the *D. elator* population, though small and constantly shifting, has managed to maintain its genetic diversity.

This study demonstrates the effectiveness of using samples from gradations across the range, rather than at two extremes. Sampling from the extremes of a population range could lead researchers to inappropriate conclusions that could wrongly influence management decisions. Though the current effective population size of *D. elator* is estimated to be around 171 to 500 individuals, perhaps small population sizes are the status quo for this species. Increasing population size may be unsustainable for this species (greater competition, reduced resources, delayed or forgoing reproduction).

Conservation Implications

Researchers interested in natural genetic variation and population structure of mammals should consider the possibility the population of their organisms of study could be exhibiting a metapopulation. This is especially important for species that are rarely seen or captured. Our findings suggest that the *D. elator* population could be a metapopulation that must be vigorously monitored so that managers can detect any great losses in genetic diversity and evolutionary potential. Furthermore, given the current advances in molecular techniques and analyses, it is no longer necessary to limit samples in the temporal dimension. Doing so, especially for species that remain understudied, will prove detrimental to any plan long-term plan for management. We advise continued use of reduced representation sequencing (ddRAD, 3RAD) but with inclusion of historic and geographically represented samples to fully encapsulate temporal and spatial genetic variability within a possibly imperiled species.

CRediT authorship contribution statement

Michaela Halsey: Writing – Original Draft, Writing – Review and Editing, Formal Analysis, Investigation, Data Curation, Validation, Methodology, Visualization. **John Stuhler:** Methodology, Investigation, Writing – Review and Editing. **Natalia Bayona-Vasquez:** Methodology, Investigation, Data Curation, Writing – Original Draft, Writing – Review and Editing, Visualization. **Roy Platt II:** Conceptualization, Funding Acquisition, Methodology, Writing – Review and Editing. **Jim Goetze:** Conceptualization, Writing – Review and Editing. **Robert Martin:** Resources. **Kenneth Matocha;** Resources. **Robert Bradley:** Conceptualization, Methodology, Funding Acquisition, Writing - Review and Editing. **Richard Stevens:** Conceptualization, Methodology, Resources, Supervision, Project Administration, Funding Acquisition, Writing – Review and Editing. **David Ray:** Conceptualization, Methodology, Resources, Supervision, Project Administration, Funding Acquisition, Writing – Review and Editing

Acknowledgements

Samples analyzed were loaned from the following collections: Museum of Texas Tech in Lubbock, Texas, Midwestern State University in Wichita Falls, Texas, and the Museum of Southwestern Biology in Albuquerque, New Mexico. Moreover, we would like to acknowledge Travis Glenn for suggesting 3RAD sequencing. This work would not be possible without cooperation from private landowners who allowed us to collect on their land. We sincerely appreciate the efforts of C. Garcia, C. Brothers, A. Kildow, J. Keats, S. de la Piedra, M. Krishnamoorthy, C. Rios-Blanco, G. Langlois, E. Stukenholtz, D. Arenas, L. Lindsay, T. Soniat, E. Roberts, E. Wright, I. Vasquez, J. Francis and many others for their help with fieldwork. Also, we would like to acknowledge the High-Performance Computing Center (HPCC) at Texas Tech University for providing substantial logistical support. We also like to thank H. Wilson and J.

422 Grimshaw and all anonymous reviewers for constructive feedback used to improve earlier
423 versions of this manuscript.

References

1. Frankham R, Ballou SE, Briscoe DA, Ballou JD. Introduction to conservation genetics. Cambridge University Press; 2002 Mar 14.

2. Moritz C. Defining 'evolutionarily significant units' for conservation. Trends Ecol Evol. 1994 Oct 1;9(10):373-5.

3. Rousset F. Genetic structure and selection in subdivided populations (MPB-40). Princeton University Press; 2013 Feb 15.

4. Slatkin M. Gene flow and the geographic structure of natural populations. Science. 1987 May 15;236(4803):787-92.

5. Wright S. Evolution in Mendelian populations. Genetics. 1931 Mar;16(2):97.

6. Franklin IR. Evolutionary change in small populations. Conservation biology: an evolutionary-ecological perspective. 1980;395.

7. Frankham R. Effective population size/adult population size ratios in wildlife: a review. Genet Res. 1995 Oct;66(2):95-107.

8. Harmon LJ, Braude S. 12. Conservation of Small Populations: Effective Population Sizes, Inbreeding, and the 50/500 Rule. In An introduction to methods and models in ecology, evolution, and conservation biology 2010 Jan 4 (pp. 125-138). Princeton University Press

9. Wang J, Santiago E, Caballero A. Prediction and estimation of effective population size. Heredity. 2016 Oct;117(4):193-206.

10. Dalquest WW, Collier G. Notes on Dipodomys elator, a rare kangaroo rat. Southwest Nat. 1964 Aug 20:146-50.

11. Baccus JT. Mammals of Baylor County, Texas. Texas J Sci. 1971 Jan 1;22(2-3):177.

12. Martin RE, Matocha KG. Distributional status of the kangaroo rat, Dipodomys elator. J. Mammal. 1972 Nov 1;53(4):873-7.

13. Carter DC, Jones JK, Jones C, Suttkus RD. Dipodomys elator. Mamm Species. 1985 May 24(232):1-3.

14. Martin RE, Matocha KG. The Texas kangaroo rat, Dipodomys elator, from Motley Co., Texas, with notes on habitat attributes. Southwest Nat. 1991 Sep 1;36(3):354-6.

15. Baumgardner GD. A recent specimen of the Texas kangaroo rat, Dipodomys elator (Heteromyidae), from Oklahoma. Southwest Nat. 1987 Jun 15;32(2):285-6.

16. Garner HW. Population dynamics, reproduction, and activities of the kangaroo rat, Dipodomys ordii, in western Texas. 1970. (Doctoral dissertation, Texas Tech University).

17. Stangl FB, Schafer TS, Goetze JR, Pinchak W Opportunistic use of modified and disturbed habitat by the Texas kangaroo rat (Dipodomys elator). Texas J Sci. 1992 Feb 1;44(1):25-35.

18. Nelson AD, Goetze JR, Watson E, Nelson M. Changes in vegetation patterns and their effect on Texas Kangaroo rats (Dipodomys elator). Texas J Sci. 2009 May 1;61(2):119-30.

19. Merriam CH. Preliminary descriptions of eleven new kangaroo rats of the genera Dipodomys and Perodipus. Proc Bio Soc Wash. 1894;9:109-16.

20. Stuhler JD, Halsey MK, Portillo-Quintero C, Ray, DA, Bradley RD, Stevens, RD. Endangered Species Research: Texas kangaroo rat (Dipodomys elator). Texas Comptroller. 2020.

21. Hamilton MJ, Chesser RK, Best TL. Genetic variation in the Texas kangaroo rat, Dipodomys elator Merriam. J. Mammal. 1987 Nov 27;68(4):775-81.

22. Pfau RS, Goetze JR, Martin RE, Matocha KG, Nelson AD. Spatial and temporal genetic diversity of the Texas kangaroo rat, Dipodomys elator (Rodentia: Heteromyidae). J. Mammal. 2019 Jul 27;100(4):1169-81.

23. Wandeler P, Hoeck PE, Keller LF. Back to the future: museum specimens in population genetics. Trends Ecol Evol. 2007 Dec 1;22(12):634-42.

24. Bradley RD, Bradley LC, Garner HJ, Baker RJ. Assessing the value of natural history collections and addressing issues regarding long-term growth and care. BioSc. 2014 Dec 1;64(12):1150-8.

25. Garrison TE, Best TL. Dipodomys ordii. Mamm Species.1990 Apr 26(353):1-0.

26. Environment Canada. Recovery Strategy for the Ord's Kangaroo Rat (Dipodomys ordii) in Canada. Species at Risk Act Recovery Strategy Series. Environment Canada, Ottawa. 2012

27. Beck ML, Kennedy ML, Biggers CJ. Genic variation in Ord's kangaroo rat Dipodomys ordii in Oklahoma. Am Midl Nat. 1981 Jan 1:77-83.

28. Halsey MK, Stuhler JD, Platt II RN, Bradley RD, Stevens RD, Ray DA. Cheeky Business: Comparing DNA Yield of Buccal, Fecal, and Whisker Samples for Minimally Invasive Genetic Research. Occ. Papers, Mus Tex Tech Univ. 2021;374:1-9

29. Longmire JL, Maltbie M, Baker, RJ. 1997. Use of" lysis buffer" in DNA isolation and its implication for museum collections. Occ. Papers, Mus Tex Tech Univ. 1997;163: 1–3.

30. Schmidly DJ, Bradley RD. The mammals of Texas. University of Texas Press; 2016 Aug 9.

31. Sikes RS, Animal Care and Use Committee of the American Society of Mammalogists. 2016 Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. J. Mammal. 2016 Jun 9;97(3):663-88.

32. Bayona-Vásquez NJ, Glenn TC, Kieran TJ, Pierson TW, Hoffberg SL, Scott PA, et al. Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). PeerJ. 2019 Oct 11;7:e7724.

33. Glenn TC, Nilsen RA, Kieran TJ, Finger JW, Pierson TW, Bentley KE, et al. Adapterama I: universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). BioRxiv. 2016 Jan 1:049114.

34. Hoffberg SL, Kieran TJ, Catchen JM, Devault A, Faircloth BC, Mauricio R, et al. RAD cap: sequence capture of dual-digest RAD seq libraries with identifiable duplicates and reduced missing data. Mol Ecol Resour. 2016 Sep;16(5):1264-78.

35. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013 Jun;22(11):3124-40.

36. Chen S, Huang T, Zhou Y, Han Y, Xu M, Gu J. AfterQC: automatic filtering, trimming, error removing and quality control for fastq data. BMC Bioinformatics. 2017 Mar;18(3):91-100.

37. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinform. 2010 Mar 1;26(5):589-95.

38. Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Piñero D, Emerson BC. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Mol Ecol Resour. 2015 Jan;15(1):28-41.

39. Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhué C, Pudlo P, Cornuet JM, Estoup A. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Mol Ecol. 2013 Jun;22(11):3165-78.

40. Huang H, Knowles LL. Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. Syst Biol. 2016 May 1;65(3):357-65.

41. Hosner PA, Faircloth BC, Glenn TC, Braun EL, Kimball RT. Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). Mol Biol Evol. 2016 Apr 1;33(4):1110-25.

42. Hodel RG, Chen S, Payton AC, McDaniel SF, Soltis P, Soltis DE. Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: comparing microsatellites and RAD-Seq and investigating loci filtering. Sci Rep. 2017 Dec 14;7(1):1-4.

43. Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinform 24:1403-1405.

521 44. Goudet J, Jombart T, Goudet MJ. Package 'hierfstat'. R package version 0.04-22. Retrieved
522 from http://www. r-project. org, http://github. com/jgx65/hierfstat. 2015 Nov 24.

523 45. Nei M. Genetic distance between populations. Am Nat. 1972 May 1;106(949):283-92.

524 46. Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: an R package for genetic analysis of
525 populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. 2014 Mar
526 4;2:e281.

527 47. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus
528 genotype data. Genetics. 2000 Jun 1;155(2):945-59.

529 48. Linck E, Battey CJ. Minor allele frequency thresholds strongly affect population structure
530 inference with genomic data sets. Mol Ecol Resour. 2019 May;19(3):639-47.

531 49. Mynhardt S, Bennett NC, Bloomer P. New insights from RADseq data on differentiation in
532 the Hottentot golden mole species complex from South Africa. Mol Phylogenet Evol. 2020
533 Feb 1;143:106667.

534 50. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. Mol
535 Ecol Notes. 2004 Mar;4(1):137-8.

536 51, Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. NeEstimator v2:
537 re-implementation of software for the estimation of contemporary effective population size
538 (Ne) from genetic data. Mol Ecol Resour. 2014 Jan;14(1):209-14.

539 52. Rousset F. genepop'007: a complete re-implementation of the genepop software for
540 Windows and Linux. Mol Ecol Resour. 2008 Jan;8(1):103-6.

541 53. Heled J, Drummond AJ. Bayesian inference of population size history from multiple loci.
542 BMC Evol Biol. 2008 Dec;8(1):1-5.

543 54. Trucchi E, Gratton P, Whittington JD, Cristofari R, Le Maho Y, Stenseth NC, et al. King
544 penguin demography since the last glaciation inferred from genome-wide data. Pro Royal
545 Soc B: Biol Science. 2014 Jul 22;281(1787):20140528.

546 55. Pacifici M, Santini L, Di Marco M, Baisero D, Francucci L, Marasini GG, et al. Generation
547 length for mammals. Nat Conserv. 2013 Nov 13;5:89.

548 56. Birky Jr CW, Fuerst P, Maruyama T. Organelle gene diversity under migration, mutation,
549 and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells,
550 and comparison to nuclear genes. Genetics. 1989 Mar 1;121(3):613-27.

551 57. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of
552 RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016 Feb;17(2):81-92.

553 58. Lemopoulos A, Prokkola JM, Uusi-Heikkilä S, Vasemägi A, Huusko A, Hyvärinen P, et al.
554 Comparing RADseq and microsatellites for estimating genetic diversity and relatedness—
555 Implications for brown trout conservation. Ecol Evol. 2019 Feb;9(4):2106-20.

59. Statham MJ, Bean WT, Alexander N, Westphal MF, Sacks BN. Historical population size change and differentiation of relict populations of the endangered giant kangaroo rat. J Hered. 2019 Jul;110(5):548-58.

60. Frankham R, Bradshaw CJ, Brook BW. Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biol Conserv. 2014 Feb 1;170:56-63.

61. Eckert CG, Samis KE, Lougheed SC. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. Mol Ecol. 2008 Mar;17(5):1170-88.

62. Audzijonyte A, Vrijenhoek RC. When gaps really are gaps: statistical phylogeography of hydrothermal vent invertebrates. Evolution. 2010 Aug;64(8):2369-84.

63. Jezkova T, Olah-Hemmings V, Riddle BR. Niche shifting in response to warming climate after the last glacial maximum: inference from genetic data and niche assessments in the chisel-toothed kangaroo rat (Dipodomys microps). Glob Chang Biol. 2011 Nov;17(11):3486-502.

64. Jezkova T, Riddle BR, Card DC, Schield DR, Eckstut ME, Castoe TA. Genetic consequences of postglacial range expansion in two codistributed rodents (genus Dipodomys) depend on ecology and genetic locus. Mol Ecol. 2015 Jan;24(1):83-97.

65. Levins R. Some demographic and genetic consequences of environmental heterogeneity for biological control. Am Entomol. 1969 Sep 1;15(3):237-40.

66. Hanski I, Gilpin M. Metapopulation dynamics: brief history and conceptual domain. Biol J Linn Soc Lond. 1991 Jan 1;42(1-2):3-16.

67. Nathan LR, Kanno Y, Vokoun JC. Population demographics influence genetic responses to fragmentation: A demogenetic assessment of the 'one migrant per generation'rule of thumb. Biol Conserv. 2017 Jun 1;210:261-72.

68. Gaona P, Ferreras P, Delibes M. Dynamics and viability of a metapopulation of the endangered Iberian lynx (Lynx pardinus). Ecol Monogr. 1998 Aug;68(3):349-70.

69. Schooley RL, Branch LC. Enhancing the area–isolation paradigm: habitat heterogeneity and metapopulation dynamics of a rare wetland mammal. Ecol Appl. 2009 Oct;19(7):1708-22.

70. Hanski I. Metapopulation ecology. Oxford University Press; 1999 Mar 18.

71. Elmhagen B, Angerbjörn A. The applicability of metapopulation theory to large mammals. Oikos. 2001 Jul;94(1):89-100.

72. Lindenmayer DB, Lacy RC. Metapopulation viability of Leadbeater's possum, Gymnobelideus leadbeateri, in fragmented old-growth forests. Ecol Appl. 1995 Feb;5(1):164-82.

590 73. Akçakaya HR, Mills G, Doncaster CP. The role of metapopulations in conservation. Key
591 topics in conservation biology. 2007:64-84.

592 74. Roberts JD, Packard RL. Comments on movements, home range and ecology of the Texas
593 kangaroo rat, Dipodomys elator Merriam. J. Mammal. 1973 Nov 1;54(4):957-62.

594 75. Goetze JR, Stasey WC, Nelson AD, Sudman PD. Habitat attributes and population size of
595 Texas kangaroo rats on an intensely grazed pasture in Wichita County, Texas. Texas J Sci.
596 2007;59(1):11.

597

Supporting Information

Fig S1. Principal components analysis on the genotypes for historical samples from Hardeman County using the glPCA function in R package 'adegenet'.

Fig S2. STRUCTURE plot of 70 D. elator samples across three time periods (see text for time breakdown). The "sampling hole" individuals (blue) are completely divergent from later samples; however, there is evidence of those loci persisting in the population. These results confirm a) no appreciable decrease in genetic diversity over 30 years and b) one interbreeding contemporary population).

Table S1. Seventy *Dipodomys elator* samples and 26 *D. ordii* samples used in the genetic analysis including temporal (historical, contemporary) subpopulation, spatial (east or west) subpopulation, the specific county the individual was found, tissue type, and the museum where the voucher was received. Museum codes are MSB (Museum of Southwestern Biology), MSU (Midwestern State University), and TTU (Texas Tech University).

Table S2. Summary statistics calculated in Stacks for 26 *D. ordii* and 38 *D. elator* contemporary samples. Private alleles are those alleles not shared with any other subpopulation.

Table S3. Log-likelihood and delta K values used in the Evanno method for *D. elator* STRUCTURE analysis.

Table S4. General summary statistics calculated in Stacks for a comparison between 3 individuals from each species that were collected in proximity (i.e. same tract of land). Private alleles are those alleles not shared with any other subpopulation. Observed and expected heterozygosity are the proportion of loci that are heterozygous based on Hardy-Weinberg frequencies. π is a measure of nucleotide diversity. FIS indicates the inbreeding coefficient.

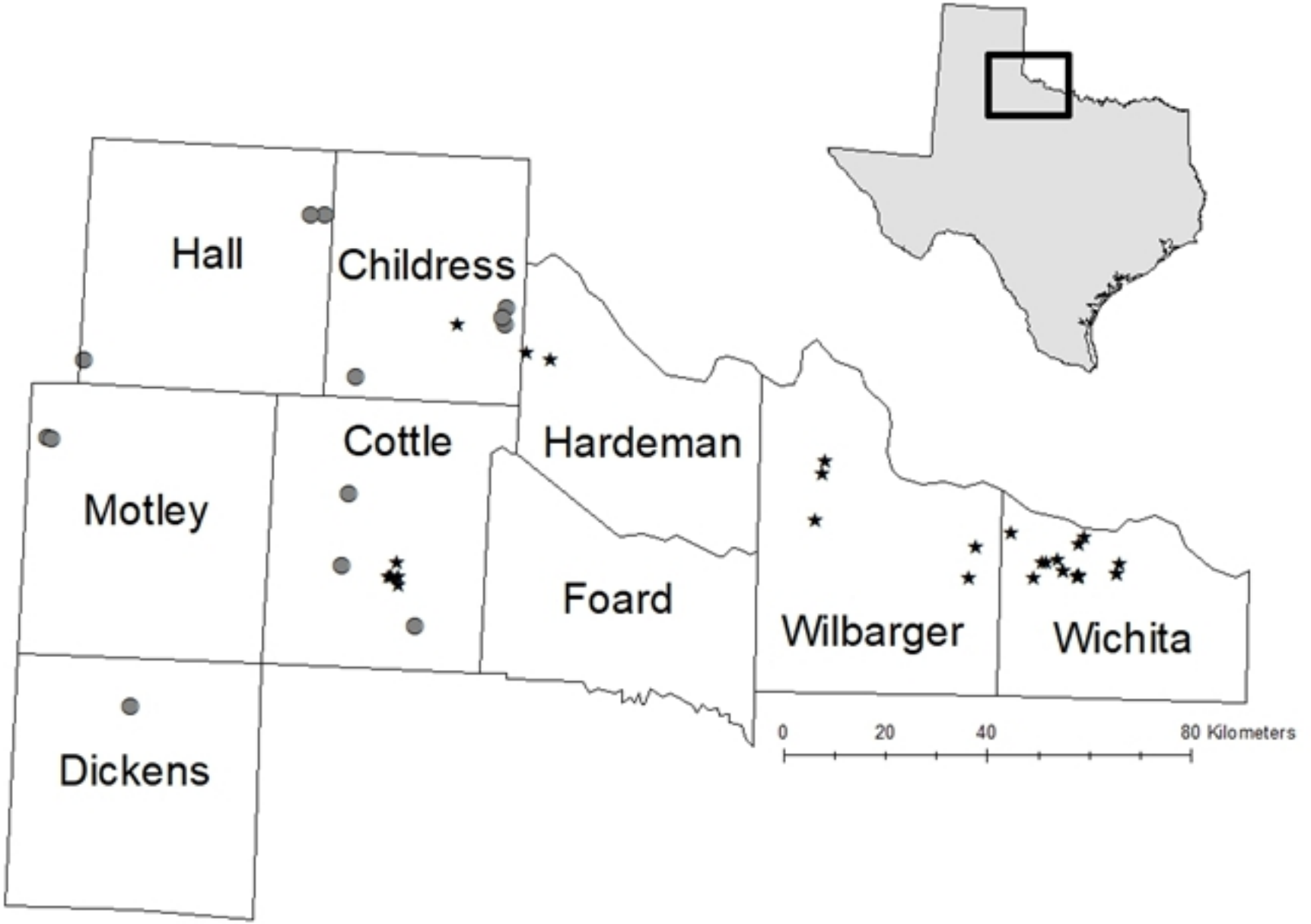
Hall
Childress
Motley
Cottle
Hardeman
Foard
Wilbarger
Wichita
Dickens
0
20
40
80 Kilometers

Figure 1

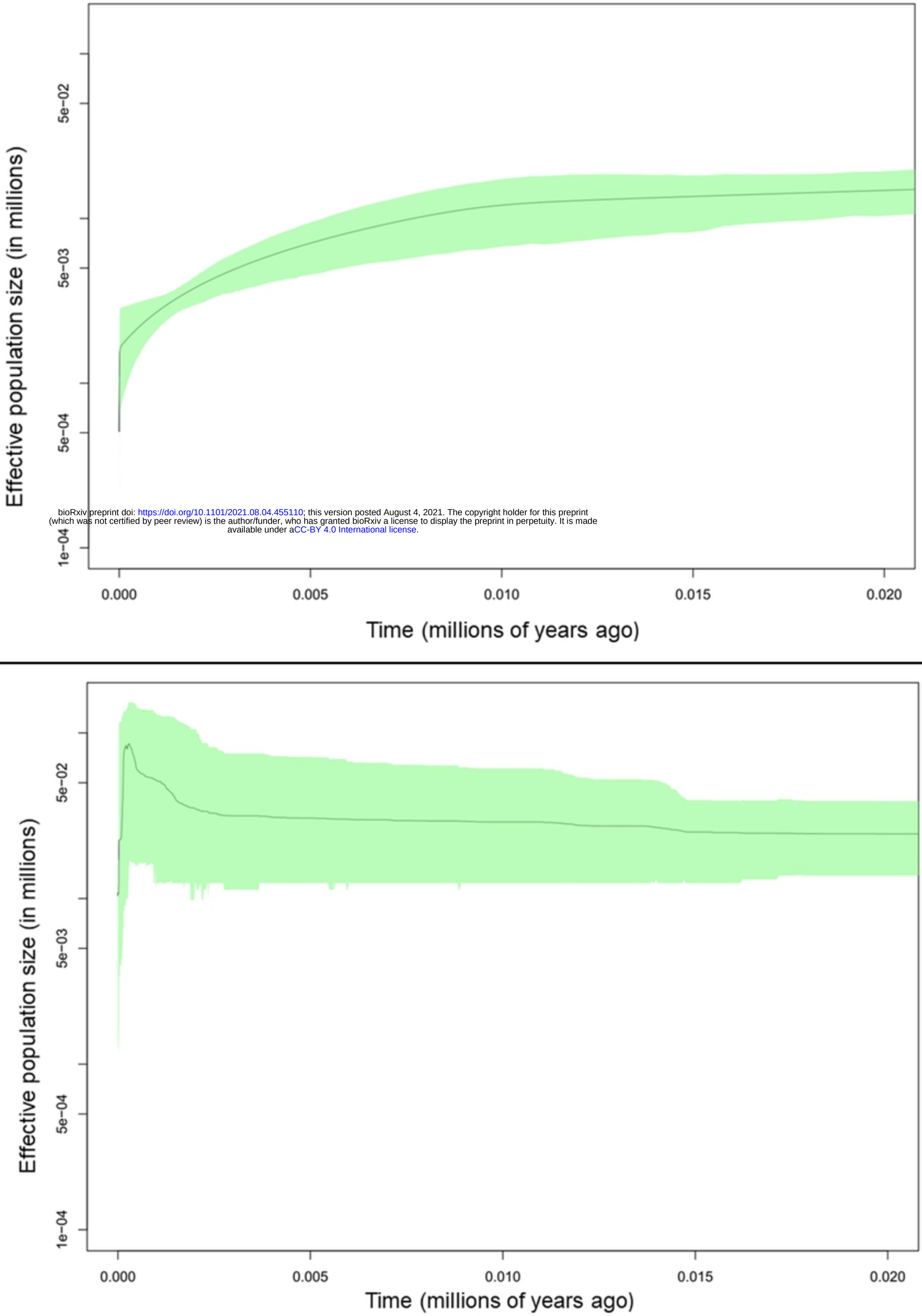

Figure 2

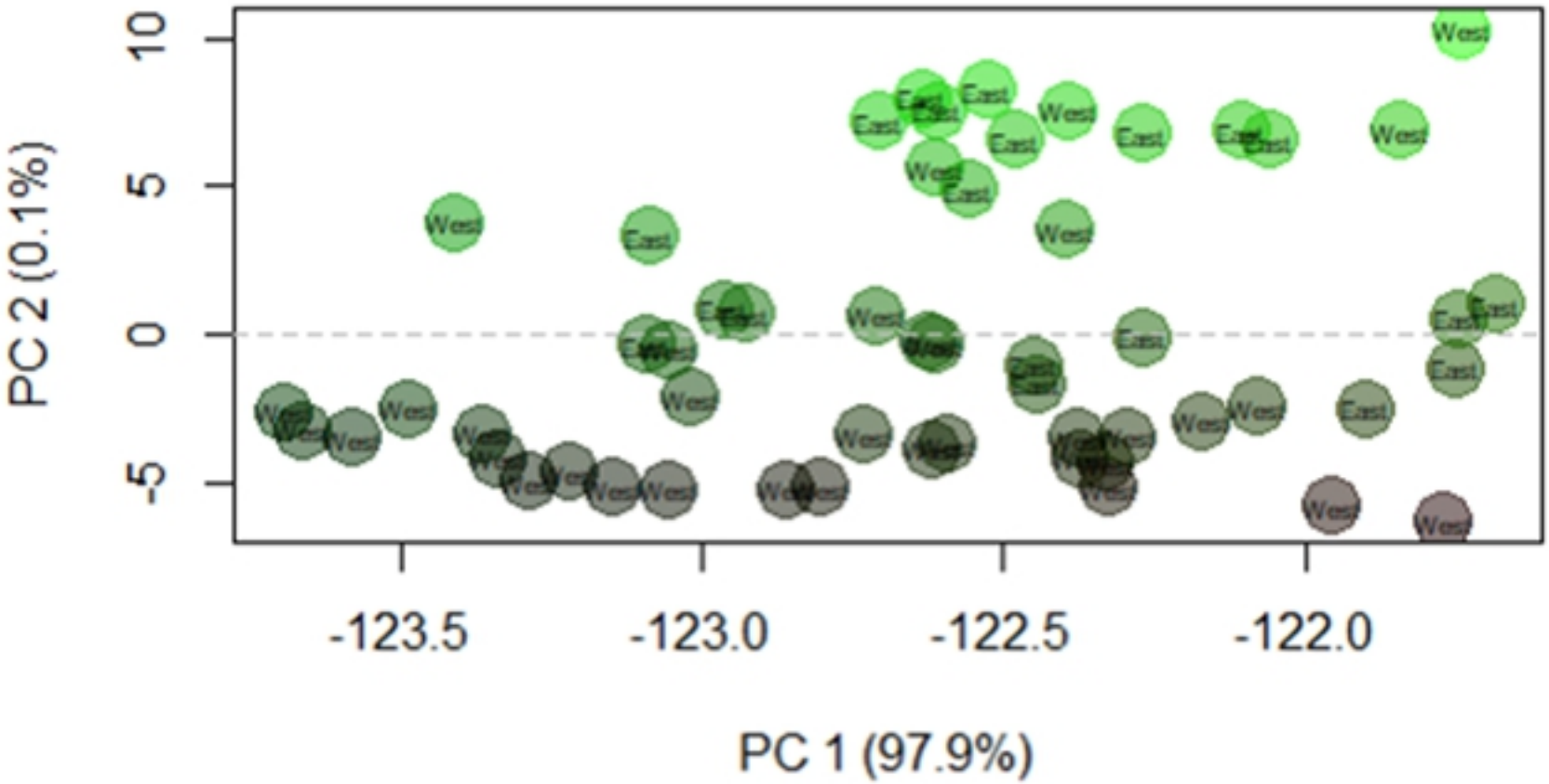

Figure 3

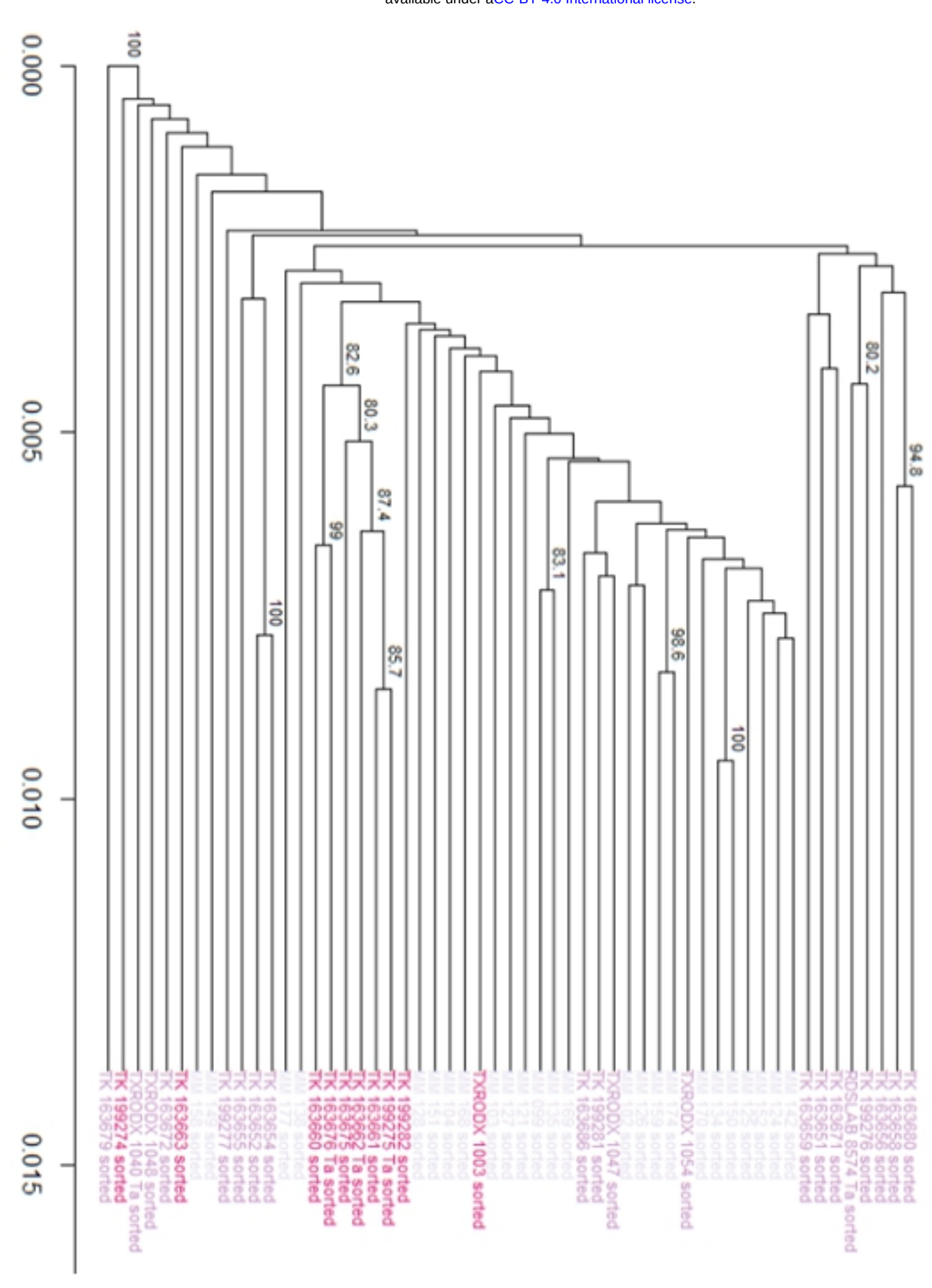

Figure 4