bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Differential introgression reveals thermal adaptation and candidate genes shaping species boundaries in North American box turtles (*Terrapene* spp.)

Running title: Box turtle hybrid zones reveal thermal adaptations

Authors:

Bradley T. Martin[1], Marlis R. Douglas[1], Tyler K. Chafin[1], John S. Placyk, Jr.[2], Roger D. Birkhead[3], Chris A. Phillips[4], Michael E. Douglas[1]

[1]Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas 72701, USA; Email: (BTM) btm002@email.uark.edu (send reprint requests to this address); (MRD): mrd1@uark.edu; (TKC): tkchafin@uark.edu; (MED): med1@uark.edu.

[2]Science Department, Robert E. Lee High School, Tyler, Texas, 75701, USA; Email: japlacyk@gmail.com

[3]Alabama Science in Motion, Auburn University, Auburn, AL 36849, USA; Email: birkhrd@auburn.edu

[4]Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, IL 61820; Email: caphilli@illinois.edu

Disclosure statement: Authors have nothing to disclose.

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ABSTRACT

Hybridization is differentially manifested across the genome, with observed introgression representing a balance between selection and migration. The capacity to quantify introgression and subsequently pinpoint the constituent genetic elements governing cross-species exchange has been promoted by the unprecedented resolution of contemporary sequencing technologies. Furthermore, the availability of annotated reference genomes has allowed genomic patterns to be associated with ecologically relevant phenotypes. We followed this pattern herein by harnessing genomic resources to decipher the role of selection in shaping hybrid zones at the interface of species-boundaries in North American box turtles (*Terrapene*). By so doing, we identified adaptive divergence in genes related to immune system function and intrinsic thermal adaptations. These, in turn, impact temperature-dependent sex determination and hypoxia tolerance. Their patterns were then contrasted among inter- and intra- specific hybrid zones that differed in a temporal and biogeographic context. Our results demonstrate that hybridization is broadly apparent in *Terrapene*, but with varying levels of divergence at loci that impinge upon thermal adaptation. These loci displayed signatures of adaptive introgression across intraspecific boundaries, and do so despite a genome-wide selective trend against intergrades. By contrast, interspecific comparisons at the same loci retained evidence of divergence. Importantly, adaptations that shape species-boundaries in *Terrapene* not only underscore climatic boundaries for these terrestrial ectotherms, but also bookmark their vulnerability to anthropogenic pressures.

Keywords: adaptive divergence; hybrid zone; RADseq; temperature dependence; genomic cline

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1. INTRODUCTION

Hybrid zones may be viewed as natural laboratories within which the genetic architecture of local adaptation or reproductive isolation can be examined. They often reflect underlying environmental gradients, and as such, are often coincident with ecological barriers (Barton & Hewitt 1985; Payseur 2010). Here, divergent selection might promote loci that underpin crucial adaptations, whereas the remainder of the genome essentially homogenizes (Via 2009; Feder *et al.* 2013). Once such genes can be identified, phenotypes that contribute to adaptive divergence can be inferred via a bottom-up, "reverse-ecology" approach (Li *et al.* 2008; Tiffin & Ross-Ibarra 2014). Thus, hybrid zones often provide a 'window' into the speciation process by allowing divergent loci to be associated with various aspects of ecology (Taylor *et al.* 2015).

Diminishing sequencing costs coupled with an upsurge in genomic annotations have facilitated extensive application of this reverse-ecology approach within a variety of ecological contexts. Examples include inference of adaptive divergence in seasonal growth and variability in immune responses (Rödin-Mörch *et al.* 2019), biodiversity response to environmental gradients (temperature, Keller & Seehausen 2012; altitudinal, Guo *et al.* 2016; elevational, Waterhouse *et al.* 2018; Teske *et al.* 2019), and an increased capacity to examine contemporary effects, such as that of anthropogenic modulation of reproductive boundaries (Garroway *et al.* 2010; Taylor *et al.* 2014; Grabenstein & Taylor 2018). A result is an increased resolution with which to gauge how environmental and climatic shifts over both geologic and contemporary timescales shift species distributions and alter pre-existing adaptive gradients (Rosenzweig *et al.* 2008; Taylor *et al.* 2015; Ryan *et al.* 2018). Thus, a genomic perspective enables a refined interpretation of two major patterns: 1) That reproductive boundaries among historically co-existing species often become

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

blurred under substantial environmental change, or contacts among otherwise allopatric taxa are facilitated (Rhymer & Simberloff 1996); and 2) That the inherent selection-migration balance found in hybrid zones (Key 1968; Parmesan *et al.* 1999) may be disrupted, thereby establishing them as particularly vulnerable to rapid and/or exceptional change (Seehausen *et al.* 2008; Kearns *et al.* 2018). Within this context, the application of genomic data to the study of hybrid zones provides not only a window into the phenotypic/genetic underpinnings of species boundaries, but also their temporal dynamics within a changing climate.

Shifting temperatures, a major component of climate change, often coincide with latitudinal changes in species distributions. This, in turn, can impact the location of hybrid zones by either strengthening (Ryan *et al.* 2018) or eroding (Muhlfeld *et al.* 2014) species boundaries. Temperature also has a broad-scale effect on species interactions by altering niche availability and net productivity (Grainger *et al.* 2018; Smith & Amarasekare 2018). It has intrinsic effects on physiological and cellular mechanisms as well (Kingsolver 2009), and directly influences growth, development, reproduction, locomotion, and immune response (Keller & Seehausen 2012). As a result, the manner by which thermal gradients interact with species boundaries has become an overriding imperative, particularly given that elevated average temperatures are a major climatic trend in the Anthropocene (Qin *et al.* 2013). Herein, we attempt to disentangle how climate change shapes species boundaries by quantifying two contrasting geographic and ecological hybrid zones found within the ectothermic North American box turtles (*Terrapene*).

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1.1. *Hybridization in North American box turtles*

North American box turtles (Emydidae, *Terrapene*) are primarily terrestrial, and possess a dome-shaped carapace and a plastral kinesis that contribute to the shape of their rectangular phenotype (Dodd 2001). They exhibit well-known hybridization (*sensu* Rhymer & Simberloff 1996) in both southeastern and midwestern North America (Milstead 1969; Dodd 2001; Cureton *et al.* 2011), and thus offer an excellent model from which to contrast regional patterns of hybridization and introgression. To do so, we evaluated four southeastern taxa [the Eastern (*T. carolina carolina*), Gulf Coast (*T. c. major*), Florida (*T. c. bauri*), and Three-toed (*T. mexicana triunguis*) box turtles (Auffenberg 1958, 1959; Milstead & Tinkle 1967; Milstead 1969; Martin *et al.* 2013)], and two midwestern (the Ornate box turtle, *T. ornata ornata* and *T. c. carolina*; Cureton *et al.* 2011). Each hybridizes regionally. We focus herein on two such regions of inter- and intraspecific contact.

One focal hybrid zone is nested within southeastern North America (Ricketts 1999), where box turtles form but one component of a biodiversity hotspot. Clear-cutting, invasive species, and altered fire regimes are widespread in this region (Stapanian *et al.* 1997, 1998; van Lear & Harlow 2002), and impact numerous endemic species (Lydeard & Mayden 1995). The region also fosters clinal intergradation and hybridization across several taxa (Remington 1968; Swenson & Howard 2004), due largely to coincident ecological and climatic transitions (Swenson & Howard 2005).

By contrast, contact zones in midwestern North America seemingly stem from secondary contact associated with postglacial recolonization and expansion (Swenson & Howard 2005). Here, prairie-grassland habitat has been anthropogenically fragmented such that niche overlap

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

now occurs between grassland and woodland species (Johnson 1994; Samson & Knopf 1994; Rhymer & Simberloff 1996; Samson *et al.* 2004). Furthermore, while overlapping forms in the midwest represent distinct species, those in the southeast are currently recognized as subspecies (Minx 1996), despite previous molecular work suggesting specific status for both *T. m. triunguis* and *T. c. bauri* (Martin *et al.* 2013, 2014).

Here, we used reduced-representation (i.e., ddRAD) next-generation sequencing to contrast genome-wide patterns of clinal introgression and divergence in *Terrapene* hybrid zones that differ markedly in their temporal, biogeographic, and anthropogenic interrelationships. We then considered the function of loci under selection by mapping them to available *Terrapene* genomic and transcriptomic reference data, then deducing their relationship to ecological factors. Hybrid zone dynamics and their functional linkage to ecological gradients provide invaluable insights into the manner by which our study species respond to a changing climate. Additionally, *Terrapene* is of conservation concern throughout its respective range, with many forms protected under federal and/or state mandates (Dodd 2001; NatureServe 2015). In this regard, our research provides the context for a proactive management paradigm, as well as a model that promotes the conservation of co-occurring forms.

2. MATERIALS AND METHODS

2.1. *Tissue and DNA collection*

Tissues for *T. carolina*, *T. ornata*, and *T. mexicana triunguis* were collected by volunteers and agency collaborators (Table S1). Additional samples were provided by numerous museums

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and organizations. Live animals were sampled non-invasively (e.g. blood, toenails, or toe-clips), whereas road-kills were sampled indiscriminately. Individuals from outside the focal contact zones were also sampled as reference populations, with this designation being tested in subsequent analyses. Isolation of genomic DNA was performed using spin-column DNA extraction kits: DNeasy Blood and Tissue Kits (QIAGEN), QIAamp Fast DNA Tissue Kit (QIAGEN), and E.Z.N.A. Tissue DNA Kits (Omega Bio-tek). The presence of genomic DNA was visualized and confirmed via gel electrophoresis using a 2% agarose gel.

2.2. *Library preparation*

In silico double digest restriction-associated DNA sequencing (ddRADseq) was carried out to optimize restriction enzyme selection, using available genomic references [Painted turtle (*Chrysemys picta*), GenBank Accession #: GCA_000241765.2 (Shaffer *et al.* 2013); FRAGMATIC (Chafin *et al.* 2017)]. Empirical verification and optimization were performed by digesting 24 samples and evaluating the distribution of these fragments using an Agilent 4200 TapeStation. Library preparation was conducted per standard protocol (Peterson *et al.* 2012), using *PstI* (5'-CTGCA|G-3') and *MspI* (5'-C|CGG-3') restriction enzymes. We digested ~500-1,000ng of DNA/sample at 37°C, with unique DNA barcode and sequencing adapters subsequently ligated. Prior to sequencing, quality control checks were performed at the core facility, including fragment analysis for confirmation of the correct size range. Fragment concentration was assessed with a 5,000 series Fragment Analyzer (Agilent) and quantitative real-time PCR. Individuals

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(N=96) were pooled per lane of single-end Illumina sequencing at the University of Oregon Genomics and Cell Characterization Core Facility (Hi-Seq 4000, 1 X 100 bp; GC3F).

2.3. *Assembly and quality control*

Read quality was first quantified using FASTQC v. 0.11.5, then demultiplexed using IPYRAD v. 0.7.28, followed by alignment with the *T. mexicana triunguis* reference genome (GenBank Accession #: GCA_002925995.2) at a distance threshold of 0.15. The assembly was restricted to reference-mapped reads to reduce potential contamination, with non-mapping reads discarded. Barcodes and adapters were trimmed in IPYRAD, as were the last five base pair (bp) of each read. Reads exceeding five bases with low PHRED quality score (<33) were discarded, and potential paralogs were filtered by excluding loci with high heterozygosity (>75%) or >2 alleles per individual. Loci were also excluded if they exhibited a sequencing depth of <20X per individual, or <50% presence across individuals.

2.4. *Assessing Admixture and Population Structure*

ADMIXTURE (Alexander *et al.* 2009) was used to assess contemporary hybridization. It employs a model-based maximum likelihood approach that estimates the proportion of ancestry shared across the genome-wide average of each individual. *K*-values ranging from *K*=1-20 were used, with 20 independent replicates per *K* (github.com/stevemussmann/admixturePipeline). SNP data was pre-filtered using VCFTOOLS (Danecek *et al.* 2011). Specifically, SNPs were randomly thinned to one per locus to alleviate linkage bias, and a minimum minor allele frequency filter

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(MAF) of 1.0% was applied to reduce bias associated with erroneous genotypes and singletons (Linck & Battey 2019). Model support across *K*-values was assessed by applying cross-validation (CV) that estimates error by folding random data partitions into missing data, then re-calculating ancestry proportions. Lower values of CV reflect less error, and accordingly, those *K*-values with the three lowest CV scores were chosen, with error summarized using two custom python packages (github.com/stevemussmann/distruct-rerun and github.com/stevemussmann/admixture_cv_sum). Output from each ADMIXTURE run was synopsized using the CLUMPAK server (Kopelman *et al.* 2015), with each individual subsequently plotted as a stacked bar chart (*distruct* v1.1; Rosenberg 2004).

To corroborate ADMIXTURE, we also performed Discriminate Analysis of Principal Components (DAPC) using the *adegenet* v2.0-0 R-package with a 1.0% MAF applied (Jombart *et al.* 2010). The *find.clusters()* function was utilized with 1,000,000 iterations to determine the optimal *K* with the lowest Bayesian Information Criterion (BIC). DAPC cross-validation (100 replicates, 90% training dataset) then evaluated which principle components and discriminant functions to retain, with individuals plotted against the top three DAPC axes.

2.5. *Identifying Hybrids*

NEWHYBRIDS (Anderson & Thompson 2002) was used to assign statistically-supported hybrids to genotype frequency classes (i.e., Pure, F_1, F_2, and backcrosses between F_1 and parental types). In doing so, the program uses Bayesian MCMC sampling to calculate the posterior probability (PP) of assignment to the pre-defined expected genotype frequency classes. The

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

getTopLoc() function in HYBRIDDETECTIVE (Wringe *et al.* 2017a) reduced the data to 300 loci containing the highest among-population differentiation (F_{ST}) and lowest linkage disequilibrium correlation (r^2<0.2). The initial NEWHYBRIDS burn-in was 500,000 MCMC generations followed by 2,000,000 post burn-in sweeps. Seeds were randomized and the analysis was run using the Jeffrey's prior for θ and π. To train the data, individuals sampled outside the focal hybrid zones were pre-assigned as parentals. The following combinations of taxa were employed: *Terrapene carolina carolina* X *T. c. major*, *T. c. carolina* X *T. m. triunguis*, *T. c. major* X *T. m. triunguis*, and *T. c. carolina* X *T. o. ornata*. A PP threshold >0.8 was required for assignment into the genotype frequency classes. A power analysis was also conducted using HYBRIDDETECTIVE and PARALLELNEWHYBRID pipelines to assess convergence in three independent NEWHYBRIDS replicates, and to gauge the statistical power and accuracy of the NEWHYBRIDS results (Wringe *et al.* 2017b; a).

2.6.*Functional Genomic Clines*

For genomic cline analyses, loci were first functionally enriched in IPYRAD v.0.7.28 using the *T. m. triunguis* annotated reference. The same parameters were employed for the full-genome alignment (above), with only mapped reads retained. Genomic clines were then used to assess patterns of locus-specific ancestry, as compared to a neutral expectation derived from 1,000 parametric simulations (INTROGRESS R; Gompert & Buerkle 2010). Outlier loci were then selected following a Bonferroni multiple test correction.

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3. RESULTS

A total of 437 individuals (Tables S1, S2) and 13,338 unlinked reference-mapped loci were retained following quality control steps. This included 134,607 variable and 90,777 parsimoniously informative sites. An additional post-alignment filter was applied to eliminate individuals with >90% missing data, resulting in a final alignment of 392 individuals.

3.1. *ADMIXTURE across taxa*

A post-alignment MAF of 1.0% contained 12,128 SNPs. ADMIXTURE was initially run using all sequenced taxa save those with limited sample size (i.e., *T. m. yucatana* and *T. nelsoni*). Plots with cross-validation (CV) scores are in Fig. 1. The lowest was at K=5 ($\bar{x}$=0.18065, SD=0.00217; Fig. S1), with K=6 ($\bar{x}$=0.18119, SD=0.00333) and K=7 ($\bar{x}$=0.18321, SD=0.00301) but slightly higher.

The all-taxa ADMIXTURE analysis at K=5 yielded population structure for the outgroups (Spotted and Blanding's turtles, *Clemmys guttata* and *Emydoidea blandingii*), *T. ornata*, *T. c. carolina*, *T. c. major*, *T. m. triunguis*, and *T. m. mexicana*. Admixture was evident between *T. o. ornata* X *T. c. carolina* (EAxON), *T. c. carolina* X *T. c. major* (EAxGU), *T. c. carolina* X *T. m. triunguis* (EAxTT), and *T. c. major* X *T. m. triunguis* (GUxTT). Admixture primarily occurred in Illinois for EAxON where their ranges overlap, in Alabama and the Florida panhandle for EAxGU, Georgia and South Carolina for EAxTT, and Mississippi and Alabama for GUxTT. Admixture proportions for *T. c. bauri* were not well resolved due to limited sampling (N=4).

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ADMIXTURE analyses with $K \geq 6$ divided *T. c. major* into Florida and Mississippi subpopulations. Floridian *T. c. major* was primarily admixed with *T. c. carolina,* and those in Mississippi with *T. m. triunguis*. Finally, K=7 split Illinois *T. c. carolina* into a distinct subpopulation. Given these results, ADMIXTURE was then re-run with data partitioned into southeastern and midwestern taxa. Such hierarchical partitioning is recommended because ADMIXTURE often underestimates K by detecting only the uppermost hierarchy of population structure (Evanno *et al.* 2005).

Partitioning the southeastern taxa with a 1.0% MAF and 90% per-individual missing data filter yielded 11,142 SNPs across 259 individuals. The lowest CV score was at K=4 ($\bar{x}$= 0.21851, SD=0.00016), with K=3 ($\bar{x}$=0.22134, SD=0.00015) and K=5 ($\bar{x}$= 0.22519, SD=0.00082) trailing (Fig. 2, 3, Fig. S2). The best-supported southeastern ADMIXTURE analysis depicted two distinct subpopulations in *T. c. major* (Florida and Mississippi) that stood in contrast to the all-taxa analysis. Additionally, K=5 identified South Carolina *T. c. carolina* as a distinct group, with admixture apparent but not geographically meaningful. However, the all-taxon and southeastern analyses agreed that admixture was present between EAxGU (*T. c. major* from Florida and *T. c. carolina* from Alabama), EAxTT (Georgia and South Carolina), and GUxTT (*T. c. major* and *T. m. triunguis* from Mississippi and Alabama).

DAPC also yielded the same four southeastern groups, plus *T. c. bauri* (K=5; Fig. S3). *Terrapene carolina bauri* was highly differentiated along axis 1 (71.9% variance explained), whereas axes 2-3 delineated the remaining southeastern taxa (17.5% and 5.73% variance explained).

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The top three *K*-values for the midwestern analysis (Fig. S4) included *K*=2 ($\bar{x}$= 0.23778, SD= 0.00018) followed by *K*=4 ($\bar{x}$= 0.25210, SD= 0.00343) and *K*=3 ($\bar{x}$=0.25415, SD=0.00272). The *K*=2 groups consisted of *T. c. carolina* and *T. o. ornata*/ *T. o. luteola*. At *K*=3, *T. c. carolina* split as a distinct group from IL, although only a few of the admixture proportions approached 100%. The *K*=4 CV mean was slightly lower (Fig. S5) and split *T. o. luteola* from *T. o. ornata,* with admixture evident between them.

3.2. *Genealogical hybrid classification*

HYBRIDDETECTIVE confirmed convergence for inter- and intra-simulation replicates (Fig. S6) with 500,000 burn-in and 2,000,000 post-burn-in sweeps (the EAxON analysis required 4,000,000 sweeps with 1,000,000 burn-in). Our power analyses suggested 90% assignment accuracy (+/-SD) for all genotype classes at a critical threshold of 0.8 (Fig. S7, S9, S11, S13). The statistical power was also high (≥0.8), although some genotype classes for EAxGU displayed relatively lower power (<0.8) (Fig. S8, S10, S12, S14).

Most assigned individuals were either backcrosses (F_1 hybrids X parental types) or F_2-generation hybrids (Fig. 4, Table 1). Specifically, the EAxGU analysis identified backcrosses with parental *T. c. major* and F_2 hybrids in the Florida panhandle and southern Alabama (Fig. 4A). Similarly, all hybrid-generation EAxTT individuals from Georgia were identified as backcrosses with both parental types, whereas South Carolina EAxTT hybrids were backcrosses with *T. c. carolina* (Fig. 4B). Second-generation hybrids and backcrosses with both parental types were evident among GUxTT (Fig. 4C). Mississippi contained individuals with all three hybrid genotype classes (F_2, B_1, and B_2), but with backcrosses to parental *T. m. triunguis* at the greatest

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

frequency. Alabama and Florida GUxTT were represented by only *T. m. triunguis* backcrosses. Finally, *T. o. ornata* and *T. c. carolina* in Illinois displayed a relatively low frequency of hybrid genotypes (5%) but all were F_1 generation, in contrast to the southeastern analyses (Fig. 4D).

3.3. *Selective signatures at functional loci*

The transcriptome-guided assembly contained 2,829 SNPs across 274 individuals, with subsets generated for EAxGU, EAxTT, and GUxTT. A filter applied to each alignment removed loci with less than 50% missing data per population (e.g., parent 1, parent 2, and admixed), followed by a second filter that retained only bi-allelic sites. The final alignments contained 2,660 (EAxGU), 2,623 (EAxTT), and 2,622 SNPs (GUxTT). Using INTROGRESS, only SNPs with a high allele frequency differential ($\delta > 0.8$) were retained (Andrés *et al.* 2013). One exception was the EAxGU analysis where $\delta > 0.7$ was applied in that no loci were recovered at the higher threshold.

The genomic cline analysis recovered four outlier loci for EAxGU, and seven for EAxTT and GUxTT (Fig. 5). Two loci failed to retain significance after Bonferroni-adjusted correction. All outlier SNPs were directly associated with temperature tolerance, pathogenic resistance, temperature-dependent sex determination (TSD), and anoxia tolerance in turtles and/ or other reptiles (Table 2). Clines were not consistent among all pairwise taxon-comparisons, in that some displayed under-dominance (i.e., selection against heterozygotes), whereas others displayed patterns indicative of adaptively-driven introgression.

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

For example, three of four outlier loci in EAxGU displayed an over-representation of EA alleles in the hybrid zone, concomitant with an under-representation of heterozygotes and GU alleles. The *SULT* locus (associated with hormonal regulation during TSD) was an extreme example (Table 2; Fig. 5), with EA alleles dominant below a hybrid index of ~0.8 (=80% assignment to GU at diagnostic loci). This pattern was replicated to a lesser degree in the *TLR9* and *ZNF236* loci associated with pathogen response and gonad development in relation to TSD. These genotypic proportions, coupled with the non-sigmoidal cline shape, suggest that introgression may be driven by the directional shift towards homozygous P_1 genotypes. In contrast, the genomic trend among putatively non-functional loci was a steep, sigmoidal cline. Taken together, these results suggest an underlying adaptive shift facilitating exchange of EA alleles despite divergence being maintained at the majority of loci (Fig. S15). One locus (*Oacyl*) with an extraordinarily shallow cline was an exception, although we suspect it represents an artifact of low genetic divergence among parental forms at this locus.

Cline shape was not consistent within the EAxTT hybrid zone. Three (of seven) outlier loci (i.e., *SASH3*, *SYPL2*, and *TLR9*) were significantly under-represented with regards to heterozygotes. Their clines displayed steep slopes, suggesting a rapid adaptive transition among parental genotypes. An additional two (of seven) loci (i.e., *CITED4*, *SLCO1A2*) reflected an overrepresentation of P_2 (TT), and a third (*FAM89B*) displayed three equally represented genotypes. Of note, *SLCO1A2* and *FAM89B* were not significant ($P = 0.009$ and $P = 0.036$, respectively).

By contrast, neutral expectations were rejected in all seven of the GUxTT clines ($P = 0$; $\alpha = 0.007$), with five (i.e., *SASH3*, *SYPL2*, *ACAD11, FAM89B*, and *ESPNL*) indicating a pattern of

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

underdominance, or selection against interspecific heterozygotes. A sixth (i.e., *TMEM214*) displayed a pattern of adaptive introgression with the homozygous P_2 (TT) genotype being overrepresented. Both the GUxTT and EAxTT analyses showed a ubiquitous signal of underdominance in non-transcriptomic loci (Fig. S15).

4. DISCUSSION

Our analyses characterized introgression in two North American box turtle hybrid zones (i.e., southeastern and midwestern North America). In the southeast, hybridization was introgressive in nature, as evidenced by numerous backcrosses and F_2 individuals, and a conspicuous lack of F_1 hybrids. We also identified specific functional loci involved in adaptive divergence between three taxa in the region. However, the midwestern hybrid zone did not show evidence of introgression, with hybrids restricted to F_1 at low frequency. Below we consider the functional genomic architecture of reproductive isolation in *Terrapene*, with a particular focus on clinal introgression in the southeastern hybrid zone.

4.1. *Functional Genomic Architecture of the Hybrid Zone in Southeastern North America*

Differential introgression among three southeastern hybrid zone taxa implicated several functional loci as contributing to adaptive divergence in *Terrapene*. Specifically, three loci relate to TSD during embryonic development, while others related to molecular pathways that contributed to anoxia and hypoxia tolerance in skeletal muscle and nervous tissues (N=6),

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

immune response to pathogens (N=2), and fasting blood glucose levels that stimulate feeding behavior [(N=1); see Table 2 for sources].

The implied functions of these genes seemingly relate to thermal adaptation. For example, anoxia/ hypoxia-related genes are associated with freeze tolerance in hibernating turtles (Storey 2006), thus supporting an obvious association with thermal gradients. Here, three loci (*SYPL2, ACAD11,* and *TMEM214*) regulate brain function and metabolism by up-regulating Ca^{2+} concentrations (Takeshima *et al.* 1998; Pamenter *et al.* 2016), inducing lipid metabolism (Gomez & Richards 2018), and initiating stress-induced apoptosis (Milton & Prentice 2007; Kesaraju *et al.* 2009; Li *et al.* 2013). Similarly, the *CITED4* gene (EAxTT) inhibits hypoxia-related transcription factors (Fox *et al.* 2004), whereas *FAM89B* (GUxTT) becomes up-regulated when physiological conditions are hypoxic (Goyal & Longo 2014). Finally, the *Oacyl* gene (EAxGU) regulates blood glucose levels when fasting (e.g. Hawksbill turtles, *Eretmochelys imbricata;* Kojima *et al.* 1999; Goldberg *et al.* 2013). *Terrapene* display elevated activity in warmer thermal conditions (Gienger & Urdiales 2017), and the gene influencing increased feeding behavior may implicate the presence of an underlying thermal gradient. Regulation of immune function is less clearly associated with underlying thermal gradients but may be tied instead to behavioral thermoregulation during infection, given that infection resistance increases at warmer temperatures (Dodd 2001; Agha *et al.* 2017).

Interestingly, a steep, sigmoidal cline was the dominant pattern among inter-specific comparisons (GUxTT and EAxTT), and was most clearly apparent in the EAxTT hybrid zone. This, in turn, reflects selection against interspecific heterozygotes (Fitzpatrick 2013). In contrast, the selective advantage of EA alleles in the EAxGU hybrid zone is intriguing in that it fails to

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

agree with the general genome-wide pattern of underdominance (Fig. S15). This discrepancy seemingly echoes a shifting adaptive landscape within which EA alleles are favored in hybrids under contemporary conditions.

4.2.*Ramifications of Climate Change on Temperature-dependent Hybrid Zones*

Given the positive relationship between outlier genes and thermal adaptations, a natural extrapolation would be that a shifting thermal gradient drives differential introgression. This has multiple implications with regard to the integrity of species boundaries in a changing climate. In one scenario, a shifting adaptive landscape promotes hybridization by contravening long-term reproductive isolation (as with EAxGU), with subsequent introgression observed at specific loci (as herein). Alternatively, rapid environmental change simply outpaces the selective filtering of maladaptive variants, with a subsequent decrease in fitness (Kokko *et al.* 2017). This would be particularly evident when effective population sizes are already depressed following a population bottleneck (Chafin *et al.* 2019). Here, extreme rates of change may also be associated with a genetic swamping effect (Todesco *et al.* 2016). Both scenarios implicate anthropogenic pressures as governing the fates of taxa in tension-zones across diverse taxa (Taylor *et al.* 2015).

Our emphasis is that ectothermic vertebrates are exceptionally vulnerable to such contemporary pressures. They pose a high extinction risk due to their strong reliance on environmental thermoregulation, and their dependence on suitable habitat (Gibbons *et al.* 2000; Sinervo *et al.* 2010; Winter *et al.* 2016). Indeed, ectotherms can exhibit reduced fitness and growth-rates when operating in conditions warmer than their thermal optima (Deutsch *et al.*

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2008; Martin & Huey 2008; Huey *et al.* 2012; Huey & Kingsolver 2019). Increased temperatures can also affect physiological pathways that are comparable with each of the genes described herein, such as increasing metabolic rates (Dillon *et al.* 2010), intensifying hypoxic stress despite higher temperature-driven O_2 demands (Huey & Ward 2005), heightening disease transmission (Pounds *et al.* 2006), or even the over-extension of thermal tolerances (Sinervo *et al.* 2010; Ceia-Hasse *et al.* 2014).

Typical evolutionary responses to insufficient thermal conditions include local adaptation (Holt 1990; Norberg *et al.* 2012; Bush *et al.* 2016), physiological and behavioral mechanisms (i.e. thermoregulation, phenology), plasticity (Urban *et al.* 2014; Sgrò *et al.* 2016), or shifts in species distributions (Parmesan *et al.* 1999; Parmesan & Yohe 2003; Moreno-Rueda *et al.* 2012). On the other hand, local extirpation or extinction reflect an inability to sufficiently respond (Sinervo *et al.* 2010). However, evolutionary responses to environmental change are limited, in that they seemingly operate over much longer timescales than contemporary climate change, particularly when behavioral thermoregulators are considered (Buckley *et al.* 2015). Maintaining phenotypic plasticity also has intrinsic costs (Chevin *et al.* 2010), and the capacity of organisms to acclimate remains quite variable (Seebacher *et al.* 2015). Finally, dispersal requires connectivity to suitable habitat, a particular problem for amphibians and reptiles due to extensive habitat loss and fragmentation (Gibbons *et al.* 2000; Douglas *et al.* 2016). In addition, connectivity can be blocked due to range expansions by generalists, or by hybridization with closely related species

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Hewitt 1999, 2000, 2001; Blois *et al.* 2013). Thus, the rate of anthropogenic change may act to curtail those natural responses typically demonstrated by biodiversity.

4.3. *Regional and Taxon-specific Perspectives*

The southeastern hybrid zone features a far greater abundance of later-generation hybrids (F_2's and backcrosses) and a conspicuous lack of F_1's. The narrow widths of transcriptomic clines suggest selection against hybrids, and both NEWHYBRIDS and the genomic/ transcriptomic cline analyses support our argument that the hybrid zone represents a tension-zone. As such, it is primarily governed by selection against relatively less-fit hybrids (Bazykin 1969; Barton & Hewitt 1985). Given that tension-zones are often temporally and spatially unstable, as well as being environmentally independent (Barton & Hewitt 1985; Buggs 2007), an understanding of their dynamics will be an important consideration for future conservation efforts.

The largest anthropogenic impact worldwide is habitat loss, with deforestation, land-use change, and increasing urbanization and agriculture most important in the southeast. Also, alterations in species composition instigated by invasive species is an important consideration, as are more frequent and extreme weather patterns due to climate change (Wear & Greis 2002; Mourtzinis *et al.* 2016; Park *et al.* 2017). These aspects will only be exacerbated going forward, as species distributions and abiotic habitats are driven northward (Desantis *et al.* 2007; Doyle *et al.* 2010; Davis *et al.* 2015; Raabe & Stumpf 2016). Accordingly, conservation efforts must focus on connectivity at the population-level by providing "stepping stone patches" of habitat along a latitudinal vector. These can allow species to disperse northward in response to climatic

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

conditions, but do not necessitate complete continuity between patches. Recent research in the Gulf Coastal Plain serves as a proof-of-concept, with latitudinal range shifts being successful despite artificially-created refuge habitat that is both sub-optimal and previously uninhabited by the focal species (Cannizzo & Griffen 2019).

Furthermore, our data suggests a capacity for inter- and intraspecific adaptive introgression in this region. However, long term consequences are unknown – introgression may facilitate evolutionary rescue (Oziolor *et al.* 2019), or drive extirpation either through genetic swamping of species or their merger (Todesco *et al.* 2016). Also unknown are the levels of introgression that govern this relationship, as well as the demographic scenarios conducive to each. Similarly, the rate of environmental change and fitness consequences of hybridization will also be indeterminate.

By contrast, the midwestern hybrid zone had a paucity of late-generation hybrids consistent with reproductive isolation accumulated prior to post-glacial secondary contact. Contemporary and predicted climatic pressures in the midwest include higher temperatures and more intense precipitation interspersed by extended drought, with forested habitat being the most severely affected (Pryor *et al.* 2014). Alternatively, agricultural expansion has already fragmented the native prairie grassland habitat on which some endemic species depend (Samson *et al.* 2004; Mussmann *et al.* 2017), potentially reducing biodiversity and eroding species boundaries by reintroducing ecologically divergent species (Haddad *et al.* 2015; Grabenstein & Taylor 2018). The midwest may thus be subject to a bi-directional assault on both forested and prairie-grassland habitat. Our data from the midwestern hybrid zone support such a breakdown in ecological barriers, likely due to fragmentation of the prairie-grassland habitat. Maintaining habitat

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

connectivity is crucial in this sense, and the development of "stepping stone" prairie grassland patches (as above) may prove sufficient (Wimberly *et al.* 2018).

Many turtle species are at elevated risks from climate change due to their long generation times and because they employ TSD during embryonic development. Here, climate change can skew sex-ratios and promote demographic collapse, with male bias initiated by warmer temperatures, and vice versa (Janzen 1994). Similarly, long generation times can also restrict the adaptive capacity of turtles in a rapidly changing climate (Hoffmann *et al.* 2017). Our study suggests that reproductive isolation in turtles involves mechanisms that regulate TSD, among other potential thermal adaptations. Clearly, temperature plays a prominent role in chelonian ecology.

Several *Terrapene*-specific conclusions are evident from our data. First, our ADMIXTURE analyses indicate the presence of discrete *T. c. major* populations in Florida and Mississippi. These implicate the Alabama and Apalachicola river drainages as biogeographic barriers. This variability was markedly absent in previous morphological analyses (Butler *et al.* 2011), which suggested that *T. c. major* may not be a distinct lineage but instead represents an area of admixture between other *Terrapene* in the region. We did indeed detect considerable admixture between *T. c. major*, *T. c. carolina*, and *T. m. triunguis*, but the two putatively pure *T. c. major* populations in the Mississippi/ Alabama panhandles, and the southeastern portion of the Florida panhandle, suggest the presence of cryptic genetic variation (Douglas *et al.* 2009). Given this, the two populations potentially represent distinct evolutionary significant units (ESU's) or management units (MU's). Second, admixture is apparent among multiple southeastern *Terrapene*, excluding *T. c. bauri*, which explained the greatest DAPC axis variation. This is

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

likely attributed to vicariance in the Florida Peninsula, where the Okefenokee Trough divided northern and southern Florida during the Pliocene (Bert 1986; Douglas *et al.* 2009). Third, the presence of selection against interspecific hybrids indicates that intermediate phenotypes are unfavorable in the hybrid zones, but the intraspecific analyses reflect patterns of adaptive introgression favoring EA alleles in hybrids. Each of these aspects will require careful consideration when conservation efforts are planned or implemented, particularly given that *Terrapene* are in decline throughout their range (Dodd 2001).

Specifically, the maintenance of latitudinal habitat connectivity can mitigate climate change by allowing dispersal northward. However, promoting habitat connectivity, yet limiting the homogenizing effects of introgression, will prove more difficult for *T. c. major* given its comparatively narrow range that is surrounded by three other *Terrapene* taxa in the Gulf Coast region. However, the two putatively "pure" populations coincident with the Apalachicola and Alabama river drainages may offer reference populations for conservation focus. Though *Terrapene* is primarily terrestrial, some freshwater turtles display similar phylogeographic patterns associated with these and other regional drainages (Roman *et al.* 1999), and may thus provide additional management foci.

On the other hand, *T. ornata* and *T. carolina* are separated by greater genetic distances than are the southeastern taxa (Martin *et al.* 2013), and accordingly the lack of hybrids beyond F_1 may reflect intrinsic genetic incompatibilities (Barton 2001; Abbott *et al.* 2013). The low frequency of F_1 hybrids in the Illinois ONxEA population may be facilitated by recent degradation of the prairie grassland habitat (Manning 2001; Mussmann *et al.* 2017) that, in turn, initiated increased heterospecific contact. However, this hypothesis cannot be explicitly tested herein. It should be

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

noted that our ONxEA results differ from that of Cureton *et al.* (2011) who depicted potential introgressive hybridization between these two taxa based on mitochondrial DNA and multiple microsatellite markers. Our genome-wide data, on the other hand, recovered only F_1's. We attribute this discrepancy to the increased capacity for assignment in our data, although a regional disparity between our respective sampling areas may also be a possibility.

While our analyses concentrated on *Terrapene*, the focal hybrid zones represent diverse taxa including reptiles, plants, mammals, birds, and amphibians (Remington 1968; Swenson & Howard 2005; Rissler & Smith 2010). For example, a genome-wide assessment of introgression in musk turtles (*Sternotherus*) inhabiting a spatially overlapping hybrid zone corroborates the patterns identified herein (Scott *et al.* 2019). The latter identified southeastern North America as a focal point for ongoing hybridization and introgression, with a similar paucity of F_1 hybrids, with differential introgression relating to ecological gradients and/or biogeographic breaks. Therefore, the changes observed in clinal patterns of *Terrapene* may be a proxy for the manner by which contemporary pressures affect co-distributed taxa.

4.4. *Conclusions*

Two important evolutionary implications are evident in our data. First, we demonstrated differential introgression along an ecological gradient in three taxa inhabiting a North American tension-zone. We then assessed genomic and transcriptomic SNPs to identify several genes whose functions are consistent with physiological processes related to thermal ecology, and thus capable of promoting adaptive divergence. In this sense, they describe ecological gradients

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

related to TSD, anoxia/ hypoxia tolerance, immune response, and feeding behavior in a hybrid zone encompassing three terrestrial turtle taxa. Second, we characterized a tension-zone in southeastern North America that is susceptible to anthropogenic and environmental changes, and which involves a wide variety of taxonomic groups (Remington 1968; Swenson & Howard 2005; Rissler & Smith 2010). It is evident from our results that NGS population genomic methods such as ddRAD can broaden our capacity to identify population structure and detect introgression, whereas traditional Sanger sequencing methods were inadequate to do so (Butler *et al.* 2011; Martin *et al.* 2013). Importantly, the NGS approaches can also link contemporary ecology with evolutionary and speciation processes, and accordingly help bridge the gap between these disciplines.

ACKNOWLEDGEMENTS

The research herein was conducted in partial fulfillment of the Ph.D. degree in Biological Sciences at the University of Arkansas (BTM). We would like to extend our gratitude to the countless volunteers and organizations who collected and/ or provided tissue samples (Table S1). We also thank the current and former members of the Douglas Lab and University of Arkansas faculty who provided advice and support, especially A. Alverson, W. Anthonysamy, M. Bangs, S. Mussmann, J. Pummill, and Z. Zbinden. Sample collections were approved under the University of Texas-Tyler Animal Care and Use Committee (IACUC) permit #113 and University of Illinois IACUC protocols 16160 and 18000. Funding was provided by the Box Turtle Conservation Committee, the American Turtle Observatory (ATO), and the following endowments: The Bruker Professorship in Life Sciences (MRD) and the Twenty-First Century

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Chair in Global Change Biology (MED). Analytical resources were provided by the Arkansas High Performance Computing Cluster (AHPCC) and an NSF-XSEDE Research Allocation (TG-BIO160065) that allowed access the Jetstream cloud service.

5. REFERENCES

Abbott R, Albach D, Ansell S *et al.* (2013) Hybridization and Speciation. *Journal of Evolutionary Biology*, **26**, 229–246.

Agha M, Price S, Nowakowski A, Augustine B, Todd B (2017) Mass mortality of eastern box turtles with upper respiratory disease following atypical cold weather. *Diseases of Aquatic Organisms*, **124**, 91–100.

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160**, 1217–1229.

Andrés JA, Larson EL, Bogdanowicz SM, Harrison RG (2013) Patterns of transcriptome divergence in the male accessory gland of two closely related species of field crickets. *Genetics*, **193**, 501–13.

Auffenberg W (1958) Fossil turtles of the genus *Terrapene* in Florida. *Bulletin of the Florida State Museum*, **3**, 53–92.

Auffenberg W (1959) A Pleistocene *Terrapene* hibernaculum, with remarks on a second complete box turtle skull from Florida. *Quarterly Journal of the Florida Academy of Science*, **22**, 49–53.

Babik W, Dudek K, Fijarczyk A *et al.* (2015) Constraint and Adaptation in newt Toll-Like Receptor Genes. *Genome Biology and Evolution*, **7**, 81–95.

Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology*, **10**, 551–568.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

Bazykin AD (1969) Hypothetical mechanism of speciaton. *Evolution*, **23**, 685–687.

Bert T (1986) Speciation in western Atlantic stone crabs (genus *Menippe*): the role of geological processes and climatic events in the formation and distribution of species. *Marine Biology*, **93**, 157–170.

Blois JL, Zarnetske PL, Fitzpatrick MC, Finnegan S (2013) Climate change and the past, present,

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and future of biotic interactions. *Science*, **341**, 499–504.

Bowden RM, Ewert MA, Nelson CE (2000) Environmental sex determination in a reptile varies seasonally and with yolk hormones. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **267**, 1745–1749.

Buckley LB, Ehrenberger JC, Angilletta MJ (2015) Thermoregulatory behaviour limits local adaptation of thermal niches and confers sensitivity to climate change. *Functional Ecology*, **29**, 1038–1047.

Buggs RJA (2007) Empirical study of hybrid zone movement. *Heredity*, **99**, 301–312.

Bush A, Mokany K, Catullo R *et al.* (2016) Incorporating evolutionary adaptation in species distribution modelling reduces projected vulnerability to climate change. *Ecology Letters*, **19**, 1468–1478.

Butler JM, Dodd Jr. CK, Aresco M, Austin JD (2011) Morphological and molecular evidence indicates that the Gulf Coast box turtle (*Terrapene carolina major*) is not a distinct evolutionary lineage in the Florida Panhandle. *Biological Journal of the Linnean Society*, **102**, 889–901.

Campbell LJ, Hammond SA, Price SJ *et al.* (2018) A novel approach to wildlife transcriptomics provides evidence of disease-mediated differential expression and changes to the microbiome of amphibian populations. *Molecular Ecology*, **27**, 1413–1427.

Cannizzo ZJ, Griffen BD (2019) An artificial habitat facilitates a climate-mediated range expansion into a suboptimal novel ecosystem. *PLOS ONE*, **14**, e0211638.

Ceia-Hasse A, Sinervo B, Vicente L, Pereira HM (2014) Integrating ecophysiological models into species distribution projections of European reptile range shifts in response to climate change. *Ecography*, **37**, 679–688.

Chafin TK, Douglas MR, Martin BT, Douglas ME (2019) Hybridization drives genetic erosion in sympatric desert fishes of western North America. *Heredity*, 1–15.

Chafin TK, Martin BT, Mussmann SM, Douglas MR, Douglas ME (2017) FRAGMATIC: in silico locus prediction and its utility in optimizing ddRADseq projects. *Conservation Genetics Resources*, **451**, 1–4.

Chevin L-M, Lande R, Mace GM (2010) Adaptation, Plasticity, and Extinction in a Changing Environment: Towards a Predictive Theory. *PLOS Biology*, **8**, e1000357.

Cureton JC, Buchman AB, Deaton R, Lutterschmidt WI (2011) Molecular analysis of hybridization between the box turtles *Terrapene carolina* and *T. ornata*. *Copeia*, **2011**, 270–277.

Czerwinski M, Natarajan A, Barske L, Looger LL, Capel B (2016) A timecourse analysis of systemic and gonadal effects of temperature on sexual development of the red-eared slider turtle *Trachemys scripta elegans*. *Developmental Biology*, **420**, 166–177.

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Davis MA, Douglas MR, Webb CT *et al.* (2015) Nowhere to go but up: Impacts of climate change on demographics of a short-range endemic (Crotalus willardi obscurus) in the sky-islands of Southwestern North America. *PLOS ONE*, **10**, e0131067.

Desantis LRG, Bhotika S, Williams K, Putz FE (2007) Sea-level rise and drought interactions accelerate forest decline on the Gulf Coast of Florida, USA. *Global Change Biology*, **13**, 2349–2360.

Deutsch CA, Tewksbury JJ, Huey RB *et al.* (2008) Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences*, **105**, 6668–6672.

Dillon ME, Wang G, Huey RB (2010) Global metabolic impacts of recent climate warming. *Nature*, **467**, 704–706.

Dodd KC (2001) *North American Box Turtles, A Natural History*. University of Oklahoma Press, Norman, OK, USA.

Douglas MR, Davis MA, Amarello M *et al.* (2016) Anthropogenic impacts drive niche and conservation metrics of a cryptic rattlesnake on the Colorado Plateau of western North America. *Royal Society*, **3**, 160047.

Douglas ME, Douglas MR, Schuett GW, Porras LW (2009) Climate change and evolution of the New World pitviper genus *Agkistrodon* (Viperidae). *Journal of Biogeography*, **36**, 1164–1180.

Doyle TW, Krauss KW, Conner WH, From AS (2010) Predicting the retreat and migration of tidal forests along the northern Gulf of Mexico under sea-level rise. *Forest Ecology and Management*, **259**, 770–777.

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

Feder JL, Flaxman SM, Egan SP, Nosil P (2013) Hybridization and the build-up of genomic divergence during speciation. *Journal of Evolutionary Biology*, **26**, 261–266.

Fitzpatrick BM (2013) Alternative forms for genomic clines. *Ecology and Evolution*, **3**, 1951–1966.

Fox SB, Braganca J, Turley H *et al.* (2004) CITED4 inhibits hypoxia-activated transcription in cancer cells, and its cytoplasmic location in breast cancer is associated with elevated expression of tumor cell hypoxia-inducible factor 1α. *Cancer Research*, **64**, 6075–6081.

Garroway CJ, Bowman J, Cascaden TJ *et al.* (2010) Climate change induced hybridization in flying squirrels. *Global Change Biology*, **16**, 113–121.

Gibbons JW, Scott DE, Ryan TJ *et al.* (2000) The global decline of reptiles, déjà vu amphibians:

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

reptile species are declining on a global scale. Six significant threats to reptile populations are habitat loss and degradation, introduced invasive species, environmental pollution, disease, unsustainable use, and global climate change. *Bioscience*, **50**, 653–666.

Gienger CM, Urdiales EM (2017) Influences on Standard Metabolism in Eastern Box Turtles (*Terrapene carolina*). *Chelonian Conservation and Biology*, **16**, 159–163.

Goldberg DW, Leitao SAT, Godfrey MH *et al.* (2013) Ghrelin and leptin modulate the feeding behaviour of the hawksbill turtle *Eretmochelys imbricata* during nesting season. *Conservation Physiology*, **1**, cot016–cot016.

Gomez CR, Richards JG (2018) Mitochondrial responses to anoxia exposure in red eared sliders (*Trachemys scripta*). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, **224**, 71–78.

Gompert Z, Buerkle AC (2010) INTROGRESS: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, **10**, 378–384.

Goyal R, Longo LD (2014) Acclimatization to long-term hypoxia: gene expression in ovine carotid arteries. *Physiological Genomics*, **46**, 725–734.

Grabenstein KC, Taylor SA (2018) Breaking barriers: causes, consequences, and experimental utility of human-mediated hybridization. *Trends in Ecology and Evolution*, **33**, 198–212.

Grainger TN, Rego AI, Gilbert B (2018) Temperature-dependent species interactions shape priority effects and the persistence of unequal competitors. *The American Naturalist*, **191**, 197–209.

Guo B, Lu D, Liao WB, Merilä J (2016) Genomewide scan for adaptive differentiation along altitudinal gradient in the Andrew's toad *Bufo andrewsi*. *Molecular Ecology*, **25**, 3884–3900.

Haddad NM, Brudvig LA, Clobert J *et al.* (2015) Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances*, **1**, e1500052.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnaean Society*, **68**, 87–112.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography—or seeing genes in space and time. *Molecular Ecology*, **10**, 537–549.

Hoffmann AA, Sgrò CM, Kristensen TN (2017) Revisiting adaptive potential, population size, and conservation. *Trends in Ecology and Evolution*, **32**, 506–517.

Holt RD (1990) The microevolutionary consequences of climate change. *Trends in Ecology and Evolution*, **5**, 311–315.

Huey RB, Kearney MR, Krockenberger A *et al.* (2012) Predicting organismal vulnerability to climate warming: roles of behaviour, physiology and adaptation. *Philosophical Transactions*

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of the Royal Society B: Biological Sciences, **367**, 1665–1679.

Huey RB, Kingsolver JG (2019) Climate warming, resource availability, and the metabolic meltdown of ectotherms. *The American Naturalist*, DOI: 10.1086/705679.

Huey RB, Ward PD (2005) Hypoxia, global warming, and terrestrial late Permian extinctions. *Science*, **308**, 398–401.

Janzen FJ (1994) Climate change and temperature-dependent sex determination in reptiles. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 7487–90.

Johnson W (1994) Woodland expansion in the Platte River, Nebraska: patterns and causes. *Ecological Monographs*, **64**, 45–84.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Kearns AM, Restani M, Szabo I *et al.* (2018) Genomic evidence of speciation reversal in ravens. *Nature Communications*, **9**, 906.

Keller I, Seehausen O (2012) Thermal adaptation and ecological speciation. *Molecular Ecology*, **21**, 782–799.

Kesaraju S, Schmidt-Kastner R, Prentice HM, Milton SL (2009) Modulation of stress proteins and apoptotic regulators in the anoxia tolerant turtle brain. *Journal of Neurochemistry*, **109**, 1413–1426.

Key KHL (1968) The Concept of Stasipatric Speciation. *Systematic Biology*, **17**, 14–22.

Kingsolver JG (2009) The well-temperatured biologist. (American Society of Naturalists Presidential Address). *The American Naturalist*, **174**, 755–68.

Kojima M, Hosoda H, Date Y *et al.* (1999) Ghrelin is a growth-hormone-releasing acylated peptide from stomach. *Nature*, **402**, 656–660.

Kokko H, Chaturvedi A, Croll D *et al.* (2017) Can evolution supply what ecology demands? *Trends in Ecology and Evolution*, **32**, 187–197.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources*, **15**, 1179–1191.

van Lear DH, Harlow RF (2002) Fire in the eastern United States: influence on wildlife habitat. In: *Proceedings: the role of fire for nongame wildlife management and community restoration: traditional uses and new directions. General Technical Report 288* (eds Ford W., Russell KR, Moorman CE), pp. 2–10. US Dept. of Agriculture, Forest Service, Northeastern Research Station.

Li YF, Costello JC, Holloway AK, Hahn MW (2008) "Reverse ecology" and the power of population genomics. *Evolution*, **62**, 2984–2994.

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Li C, Wei J, Li Y *et al.* (2013) Transmembrane Protein 214 (TMEM214) mediates endoplasmic reticulum stress-induced caspase 4 enzyme activation and apoptosis. *The Journal of Biological Chemistry*, **288**, 17908–17.

Linck EB, Battey CJ (2019) Minor allele frequency thresholds strongly affect population structure inference with genomic datasets. *Molecular Ecology Resources*, 0–2.

Lydeard C, Mayden RL (1995) A diverse and endangered aquatic ecosystem of the southeast United States. *Conservation Biology*, **9**, 800–805.

Manning B (2001) *Critical Trends in Illinois Ecosystems*. Illinois Department of Natural Resources. Springfield, IL.

Martin BT, Bernstein NP, Birkhead RD *et al.* (2013) Sequence-based molecular phylogenetics and phylogeography of the American box turtles (*Terrapene* spp.) with support from DNA barcoding. *Molecular Phylogenetics and Evolution*, **68**, 119–134.

Martin BT, Bernstein NP, Birkhead RD *et al.* (2014) On the reclassification of the *Terrapene* (Testudines: Emydidae): a response to Fritz & Havaš. *Zootaxa*, **3835**, 292–294.

Martin TL, Huey RB (2008) Why suboptimal is optimal: Jensen's inequality and ectotherm thermal preferences. *The American Naturalist*, **171**, E102-18.

Milstead WW (1969) Studies on the evolution of the box turtles (genus *Terrapene*). *Bulletin of the Florida State Museum, Biological Science Series*, **14**, 1–113.

Milstead WW, Tinkle DW (1967) *Terrapene* of Western Mexico, with comments on species groups in the genus. *Copeia*, **1967**, 180–187.

Milton SL, Prentice HM (2007) Beyond anoxia: the physiology of metabolic downregulation and recovery in the anoxia-tolerant turtle. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **147**, 277–290.

Minx P (1996) Phylogenetic relationships among the box turtles, Genus *Terrapene*. *Herpetologica*, **52**, 584–597.

Moreno-Rueda G, Pleguezuelos JM, Pizarro M, Montori A (2012) Northward shifts of the distributions of Spanish reptiles in association with climate change. *Conservation Biology*, **26**, 278–283.

Mourtzinis S, Ortiz B V., Damianidis D (2016) Climate change and ENSO effects on southeastern US climate patterns and maize yield. *Scientific Reports*, **6**, 29777.

Muhlfeld CC, Kovach RP, Jones LA *et al.* (2014) Invasive hybridization in a threatened species is accelerated by climate change. *Nature Climate Change*, **4**, 620–624.

Mussmann SM, Douglas MR, Anthonysamy WJB *et al.* (2017) Genetic rescue, the greater prairie chicken and the problem of conservation reliance in the Anthropocene. *Royal Society Open Science*, **4**, 160736.

NatureServe (2015) NatureServe Explorer: An online encyclopedia of life [web application].

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Version 7.1. NatureServe, Arlington, Virginia. http://www.natureserve.org/ [date accessed: 13 Oct 2016].

Norberg J, Urban MC, Vellend M, Klausmeier CA, Loeuille N (2012) Eco-evolutionary responses of biodiversity to climate change. *Nature Climate Change*, **2**, 747–751.

Oziolor EM, Reid NM, Yair S *et al.* (2019) Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science*, **364**, 455–457.

Paitz RT, Bowden RM (2008) A proposed role of the sulfotransferase/sulfatase pathway in modulating yolk steroid effects. *Integrative and Comparative Biology*, **48**, 419–427.

Pamenter ME, Gomez CR, Richards JG, Milsom WK (2016) Mitochondrial responses to prolonged anoxia in brain of red-eared slider turtles. *Biology Letters*, **12**, 20150797.

Park WA, Cook BI, Smerdon JE *et al.* (2017) The 2016 southeastern U.S. drought: an extreme departure from centennial wetting and cooling. *Journal of Geophysical Research: Atmospheres*, **122**, 10,888-10,905.

Parmesan C, Ryrholm N, Stefanescu C *et al.* (1999) Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature*, **399**, 579–583.

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.

Payseur BA (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources*, **10**, 806–820.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLOS ONE*, **7**, e37135.

Piferrer F (2013) Epigenetics of sex determination and gonadogenesis. *Developmental Dynamics*, **242**, 360–370.

Pounds AJ, Bustamante MR, Coloma LA *et al.* (2006) Widespread amphibian extinctions from epidemic disease driven by global warming. *Nature*, **439**, 161–167.

Pryor S., Scavia D, Downer C *et al.* (2014) Chapter 18: Midwest. In: *Climate Change Impacts in the United States: The Third National Climate Assessment* (eds Melillo JM, Richmond TC, Yohe GW), pp. 418–440. U.S. Global Change Research Program.

Qin D, Plattner G, Tignor M *et al.* (2013) *Summary for policymakers. Climate change 2013: the physical science basis. Contribution of Working Group I to the fifth assessment report of the Intergovernmental Panel on Climate Change, eds Stocker, TF et al.* Cambridge University Press, Cambridge, UK.

Raabe EA, Stumpf RP (2016) Expansion of tidal marsh in response to sea-level rise: Gulf Coast of Florida, USA. *Estuaries and Coasts*, **39**, 145–157.

Remington CL (1968) Suture-zones of hybrid interaction between recently joined biotas. In:

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Evolutionary Biology (ed Dobzhansky T), pp. 321–428. Springer, New York, NY.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, **27**, 83–109.

Ricketts TH (1999) *Terrestrial ecoregions of North America: a conservation assessment*. Island Press, Washington, DC.

Rissler LJ, Smith WH (2010) Mapping amphibian contact zones and phylogeographical break hotspots across the United States. *Molecular Ecology*, **19**, 5404–5416.

Rödin-Mörch P, Luquet E, Meyer-Lucht Y *et al.* (2019) Latitudinal divergence in a wide-spread amphibian: contrasting patterns of neutral and adaptive genomic variation. *Molecular Ecology*, **28**, 2996–3011.

Roman J, Santhuff SD, Moler PE, Bowen BW (1999) Population structure and cryptic evolutionary units in the alligator snapping turtle. *Conservation Biology*, **13**, 135–142.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Rosenzweig C, Karoly D, Vicarelli M *et al.* (2008) Attributing physical and biological impacts to anthropogenic climate change. *Nature*, **453**, 353–357.

Roth M, Obaidat A, Hagenbuch B (2012) OATPs, OATs and OCTs: the organic anion and cation transporters of the *SLCO* and *SLC22A* gene superfamilies. *British Journal of Pharmacology*, **165**, 1260–1287.

Ryan SF, Deines JM, Scriber JM *et al.* (2018) Climate-mediated hybrid zone movement revealed with genomics, museum collection, and simulation modeling. *Proceedings of the National Academy of Sciences*, 2017–14950.

Samson F, Knopf F (1994) Prairie conservation in North America. *Bioscience*, **44**, 418–421.

Samson FB, Knopf FL, Ostlie WR (2004) Great Plains ecosystems: past, present, and future. *Wildlife Society Bulletin*, **32**, 6–15.

Scott PA, Glenn TC, Rissler LJ (2019) Formation of a recent hybrid zone offers insight into the geographic puzzle and maintenance of species boundaries in musk turtles. *Molecular Ecology*, **28**, 761–771.

Seebacher F, White CR, Franklin CE (2015) Physiological plasticity increases resilience of ectothermic animals to climate change. *Nature Climate Change*, **5**, 61–66.

Seehausen O, Takimoto G, Roy D, Jokela J (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, **17**, 30–44.

Sgrò CM, Terblanche JS, Hoffmann AA (2016) What Can Plasticity Contribute to Insect Responses to Climate Change? *Annual Review of Entomology*, **61**, 433–451.

Shaffer HB, Minx P, Warren DE *et al.* (2013) The western painted turtle genome, a model for the

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biology*, **14**, R28.

Sinervo B, Méndez-de-la-Cruz F, Miles DB *et al.* (2010) Erosion of lizard diversity by climate change and altered thermal niches. *Science*, **328**, 894–9.

Smith DJ, Amarasekare P (2018) Toward a Mechanistic Understanding of Thermal Niche Partitioning. *The American Naturalist*, **191**, E57–E75.

Stapanian MA, Cassell DL, Cline SP (1997) Regional patterns of local diversity of trees: associations with anthropogenic disturbance. *Forest Ecololgy and Management*, **93**, 33–44.

Stapanian MA, Sundberg SD, Baumgardner GA, Liston A (1998) Alien plant species composition and associations with anthropogenic disturbance in North American forests. *Plant Ecology*, **139**, 49–62.

Storey KB (2006) Reptile freeze tolerance: metabolism and gene expression. *Cryobiology*, **52**, 1–16.

Swenson NG, Howard DJ (2004) Do suture zones exist? *Evolution*, **58**, 2391–2397.

Swenson NG, Howard DJ (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *The American Naturalist*, **166**, 581–591.

Takeshima H, Shimuta M, Komazaki S *et al.* (1998) Mitsugumin29, a novel synaptophysin family member from the triad junction in skeletal muscle. *The Biochemical Journal*, **331**, 317–22.

Taylor SA, Larson EL, Harrison RG (2015) Hybrid zones: windows on climate change. *Trends in Ecology and Evolution*, **30**, 398–406.

Taylor SA, White TA, Hochachka WM *et al.* (2014) Climate-mediated movement of an avian hybrid zone. *Current Biology*, **24**, 671–676.

Teske PR, Sandoval-Castillo J, Golla TR *et al.* (2019) Thermal selection as a driver of marine ecological speciation. *Proceedings of the Royal Society B: Biological Sciences*, **286**, 20182023.

Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology and Evolution*, **29**, 673–680.

Todesco M, Pascual MA, Owens GL *et al.* (2016) Hybridization and extinction. *Evolutionary Applications*, **9**, 892–908.

Urban MC, Richardson JL, Freidenfelds NA (2014) Plasticity and genetic adaptation mediate amphibian and reptile responses to climate change. *Evolutionary Applications*, **7**, 88–103.

Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences*, **106**, 9939–9946.

Waterhouse MD, Erb LP, Beever EA, Russello MA (2018) Adaptive population divergence and

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

directional gene flow across steep elevational gradients in a climate-sensitive mammal. *Molecular Ecology*, **27**, 2512–2528.

Wear DN, Greis JG (2002) Southern forest resource assessment: summary of findings. *Journal of Forestry*, **100**, 6–14.

Weilinger N, Tang P, Thompson R (2012) Anoxia-induced NMDA receptor activation opens pannexin Channels via *src* Family Kinases. *The Journal of Neuroscience*, **32**, 12579–12588.

Wimberly MC, Narem DM, Bauman PJ, Carlson BT, Ahlering MA (2018) Grassland connectivity in fragmented agricultural landscapes of the north-central United States. *Biological Conservation*, **217**, 121–130.

Winter M, Fiedler W, Hochachka WM *et al.* (2016) Patterns and biases in climate change research on amphibians and reptiles: a systematic review. *Royal Society Open Science*, **3**, 160158.

Wringe BF, Stanley RRE, Jeffery NW, Anderson EC, Bradbury IR (2017a) HYBRIDDETECTIVE: a workflow and package to facilitate the detection of hybridization using genomic data in R. *Molecular Ecology Resources*, **17**, e275–e284.

Wringe BF, Stanley RRE, Jeffery NW, Anderson EC, Bradbury IR (2017b) *parallelnewhybrid*: an R package for the parallelization of hybrid detection using NEWHYBRIDS. *Molecular Ecology Resources*, **17**, 91–95.

Yatsu R, Miyagawa S, Kohno S *et al.* (2016) RNA-seq analysis of the gonadal transcriptome during *Alligator mississippiensis* temperature-dependent sex determination and differentiation. *BMC Genomics*, **17**, 77.

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DATA ACCESSIBILITY

Upon acceptance for publication, the raw ddRADseq reads will be available as an NCBI sequence read archive (SRA), and the scripts, SNP input files, and metadata will be stored in a DRYAD digital repository.

AUTHOR CONTRIBUTIONS

BTM and TKC conceived the research, laboratory, and analytical tools, scripts and approaches. BTM performed the lab work, analyzed the data, conducted the bioinformatics analyses, and wrote the manuscript. MRD and MED were the study supervisors, guided the study design, and provided funding. JSP facilitated the collection of thousands of *Terrapene* tissues and provided expertise in methodological development. RDB collected hundreds of *Terrapene* tissues from southeastern North America and facilitated the collection of many additional individuals. CAP provided tissues from the midwestern hybrid zone as well as sample site expertise in Illinois. All authors contributed to editing and revising the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1: Population-level genotype frequency proportions derived from four NEWHYBRIDS analyses involving the GU=Gulf Coast (*T. c. major*), EA=Eastern (*T. c. carolina*), TT=Three-toed (*T. m. triunguis*), and ON=Ornate (*T. o. ornata*) box turtles, plus *T. carolina* (TC=subspecies unidentified in the field). The second two letters in the population ID correspond to locality according to U.S. state (AL=Alabama, FL=Florida, LA=Louisiana, SC=South Carolina, GA=Georgia, MS=Mississippi, and IL=Illinois). Columns depict the proportion of assignment to parental (P_1 and P_2), first and second-generation hybrid (F_1 and F_2), backcross (B_1 and B_2), and unassigned (FN) genotype frequency classes.

Population	**P_1**	**P_2**	**F_1**	**F_2**	**B_1**	**B_2**	**FN**
GUxEA							
PureGU	1.00	0.00	0.00	0.00	0.00	0.00	0.00
PureEA	0.00	1.00	0.00	0.00	0.00	0.00	0.00
EAAL	0.00	1.00	0.00	0.00	0.00	0.00	0.00
GUFL	0.46	0.04	0.00	0.08	0.21	0.00	0.21
TCAL	0.02	0.86	0.00	0.02	0.02	0.00	0.08
GUAL	0.20	0.20	0.00	0.00	0.00	0.00	0.60
EAxTT							
PureEA	1.00	0.00	0.00	0.00	0.00	0.00	0.00
PureTT	0.00	1.00	0.00	0.00	0.00	0.00	0.00
TTLA	0.00	1.00	0.00	0.00	0.00	0.00	0.00
EASC	0.47	0.00	0.00	0.00	0.40	0.00	0.13
TCGA	0.80	0.00	0.00	0.00	0.00	0.10	0.10
EAGA	0.91	0.00	0.00	0.00	0.05	0.00	0.05
TCAL	0.96	0.00	0.00	0.00	0.04	0.00	0.00
TTxGU							
PureTT	0.00	1.00	0.00	0.00	0.00	0.00	0.00
PureGU	1.00	0.00	0.00	0.00	0.00	0.00	0.00
GUAL	0.60	0.00	0.00	0.00	0.20	0.00	0.20
TTMS	0.00	0.50	0.00	0.17	0.00	0.06	0.28
TTLA	0.00	1.00	0.00	0.00	0.00	0.00	0.00
TCMS	0.43	0.00	0.00	0.00	0.57	0.00	0.00
GUMS	0.52	0.02	0.00	0.00	0.13	0.04	0.28
GUFL	0.88	0.04	0.00	0.00	0.04	0.00	0.04
ONxEA							
PureON	1.00	0.00	0.00	0.00	0.00	0.00	0.00
PureEA	0.00	1.00	0.00	0.00	0.00	0.00	0.00
ONIL	0.74	0.21	0.05	0.00	0.00	0.00	0.00
EAIL	0.00	0.98	0.00	0.00	0.00	0.00	0.03

Table 2: Potential functions of outlier mRNA loci derived from transcriptomic clines (Fig. 5). The significance threshold (α) was determined using a Bonferroni correction for multiple tests. Gene abbreviations in bold with an asterisk (*) differ significantly from neutral expectation. EA=Eastern (*T. c. carolina*), GU=Gulf Coast (*T. c. major*), TT=Three-toed (*T. m. triunguis*).

Gene Abbr. (P-value)	Genotype† P_1/H/P_2‡	Full Gene Name	Possible Function(s)	Source(s)
EAxGU (α = 0.01)				
SULT (P = 0)*	+/-/-	Amine Sulfotransferase-like	Regulates steroids in yolk during TSD§	(Bowden *et al.* 2000; Paitz & Bowden 2008)
TLR9 (P = 0)*	+/-/-	Toll-like Receptor 9	Immune Response to Pathogens	(Babik *et al.* 2015)
ZNF236 (P = 0)*	+/-/-	Zinc Finger Protein 236	Gonadogenesis involved with TSD§	(Piferrer 2013)
Oacyl (P = 0)*	N/-/+	O-acyltransferase (Ghrelin)	Stimulates feeding behavior, maintains fasting blood glucose	(Kojima *et al.* 1999; Goldberg *et al.* 2013)
EAxTT (α = 0..007)				
SASH3 (P = 0)*	+/-/N	SAM and SH3 Domain Containing 3	Immune signaling; anoxic cell death	(Weilinger *et al.* 2012; Campbell *et al.* 2018)
SYPL2 (P = 0)*	+/-/N	Synaptophysin-like Protein 2	Maintenance of [Ca^{2+}] during anoxia	(Takeshima *et al.* 1998; Pamenter *et al.* 2016)
FAM89B (0.036)	N/N/N	Family with Sequence Similarity 89, member B	Upregulated in hypoxic conditions	(Goyal & Longo 2014)
CITED4 (P = 0)*	N/N/+	Cbp/p300 Interacting Transactivator, Domain 4	Inhibits hypoxia-activated transcription	(Fox *et al.* 2004)
TLR9 (P = 0)*	+/-/N	Toll-like Receptor 9	Immune Response to Pathogens	(Babik *et al.* 2015)
SLCO1A2 (0.009)	N/N/+	Solute Carrier Organic Anion Transporter, family member 1A2	Cellular absorption, distribution, and excretion of xenobiotics	(Roth *et al.* 2012)
GUxTT (α = 0..007)				
SASH3 (P = 0)*	+/-/N	SAM and SH3 Domain Containing 3	Immune signaling; anoxic cell death	(Weilinger *et al.* 2012; Campbell *et al.* 2018)
SYPL2 (P = 0)*	+/-/N	Synaptophysin-like Protein 2	Maintenance of [Ca^{2+}] during anoxia	(Takeshima *et al.* 1998; Pamenter *et al.* 2016)
FAM89B (P = 0)*	+/N/-	Family with Sequence Similarity 89, member B	Upregulated in hypoxic conditions	(Goyal & Longo 2014)
ACAD11 (P = 0)*	+/-/N	Acyl-CoA Dehydrogenase, family member 11-like	Lipid metabolism, responses to anoxia	(Gomez & Richards 2018)
ESPNL (P = 0)*	-/-/+	Espin-like Protein	Gonadal sex determination (TSD§)	(Czerwinski *et al.* 2016; Yatsu *et al.* 2016)
TMEM214 (P = 0)*	-/-/+	Transmembrane Protein 214	Stress-induced apoptosis during anoxia	(Milton & Prentice 2007; Kesaraju *et al.* 2009; Li *et al.* 2013)

†Genotype Proportions: Overrepresented (+), underrepresented (-), or neutral (N)
‡Genotypes = P_1 (parent1; AA), H (heterozygous; Aa), and P_2 (parent2; aa)
§TSD = Temperature-dependent sex determination

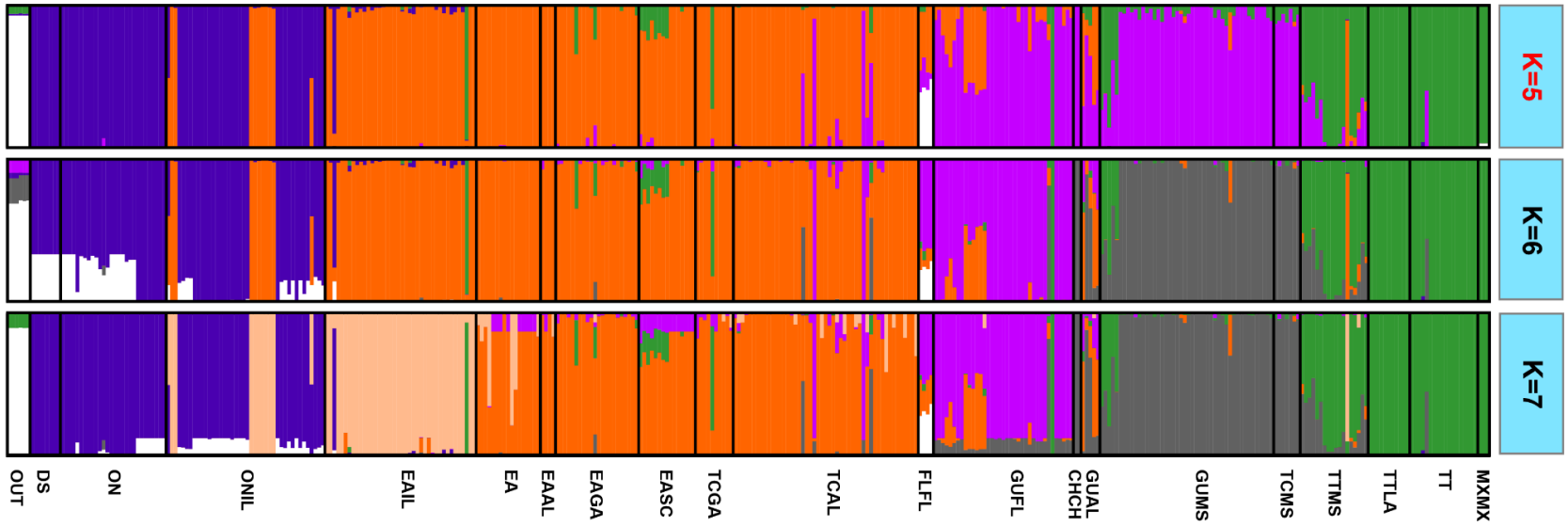

Figure 1: ADMIXTURE plot for K=5, K=6, and K=7 representing 12,128 unlinked SNPs across all sampled populations. The lowest cross-validation score was for K=5 ($\bar{x}$=0.18065, SD=0.00217; depicted at right in red), followed by K=6 ($\bar{x}$=0.18119, SD=0.00333, and then K=7 (mean=0.18321, SD=0.00301). Each bar represents a unique individual, and bars with mixed colors represent admixed ancestry. The first two letters of the populations correspond to subspecific field identification (OUT=outgroups, Spotted and Blanding's turtles, *Clemmys guttata* and *Emydoidea blandingii*; DS=Desert box turtle, *T. o. luteola*; ON=Ornate, *T. o. ornata*; EA=Eastern, *T. c. carolina*; FL=Florida, *T. c. bauri*; GU=Gulf Coast, *T. c. major*; TT=Three-toed, *T. m. triunguis*; MX=Mexican, *T. m. mexicana*; TC=*Terrapene carolina*, with subspecies unidentified in the field). The second two letters (if present) represent locality codes for U.S. or Mexican state (IL=Illinois; AL=Alabama; GA=Georgia; SC=South Carolina; FL=Florida; CH=Coahuila, Mexico; MS=Mississippi; LA=Louisiana; MX=Tamaulipas, Mexico). Populations lacking state locality code consisted of multiple localities sampled outside the hybrid zone.

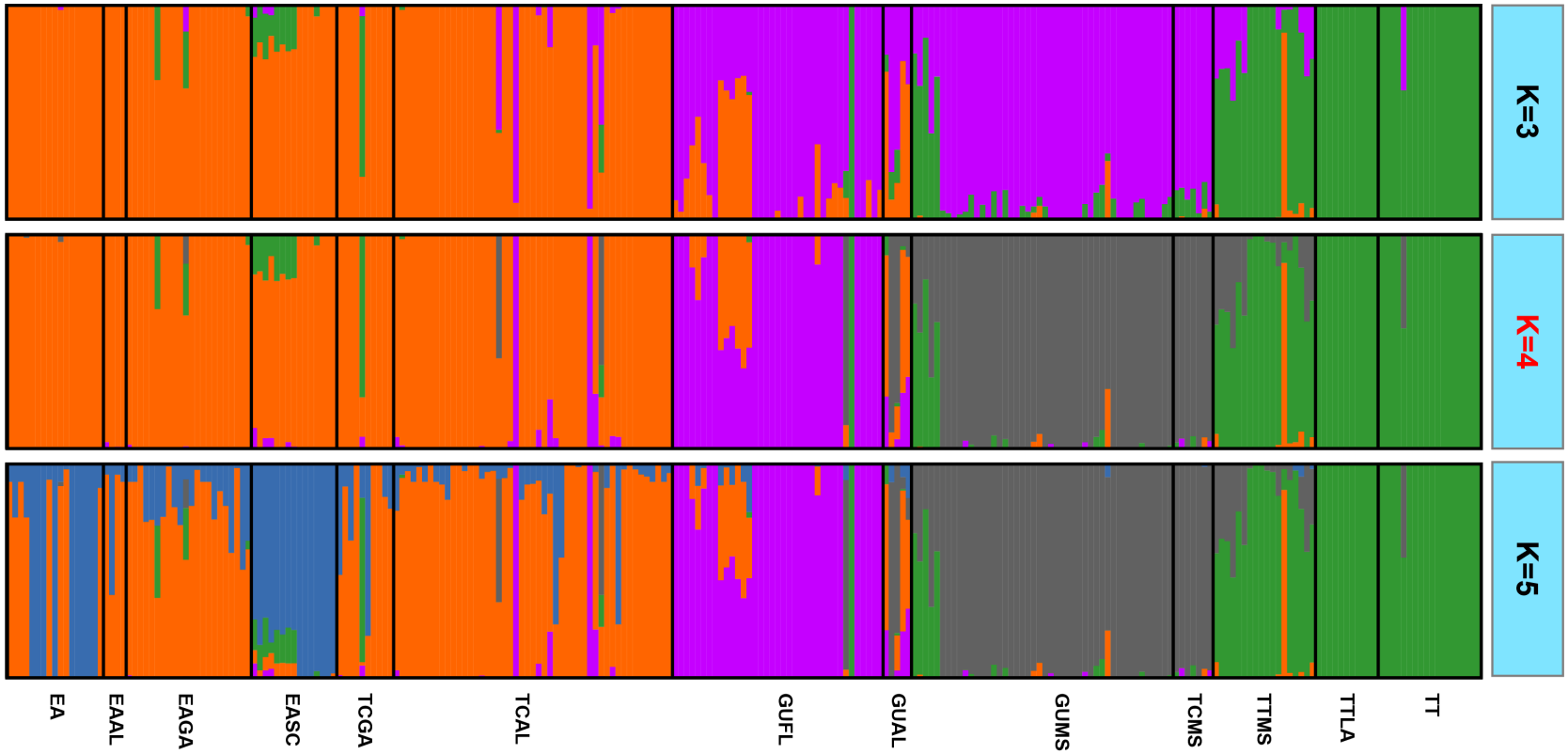

Figure 2: ADMIXTURE plot for K=3, K=4, and K=5 representing 11,142 unlinked SNPs across southeastern taxa. The lowest cross-validation score was for K=4 ($\bar{x}$= 0.21851, SD=0.00016; depicted at right in red), followed by K=3 ($\bar{x}$=0.22134, SD=0.00015), and then K=5 ($\bar{x}$= 0.22519, SD=0.00082). Each bar represents a unique individual, and bars with mixed colors depict admixed ancestry. The first two letters of the population codes correspond to subspecific field identification (EA=Eastern, *T. c. carolina*; GU=Gulf Coast, *T. c. major*; TT=Three-toed, *T. m. triunguis*; TC=*Terrapene carolina*, with subspecies unidentified in the field). The second two letters (if present) represent locality codes for U.S. states (AL=Alabama; GA=Georgia; SC=South Carolina; FL=Florida; MS=Mississippi; LA=Louisiana). Populations lacking a state locality code consisted of multiple localities sampled outside the hybrid zone.

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

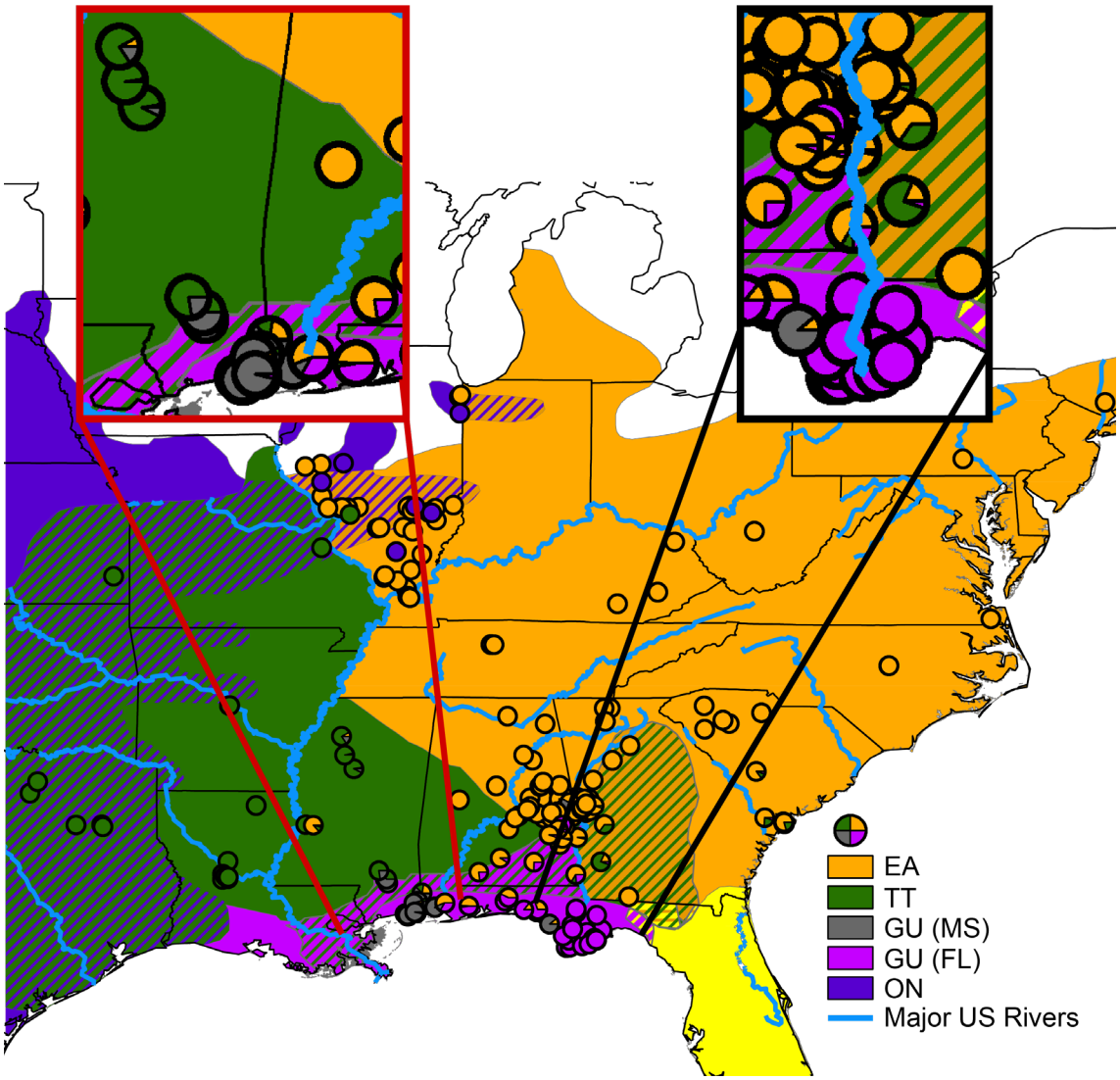

Figure 3: *Terrapene* distribution map. Cross-hatched areas represent contact zones. Circles indicate individual sampling localities, and the accompanying pie charts depict admixture proportions from the all-taxon K=5 (for midwestern individuals) and southeastern *K*=4 ADMIXTURE analyses (Fig. 1, 2). The expanded regions highlight two distinct *T. c. major* populations in the panhandles of Mississippi (red box) and Florida (black box), located in the Alabama and Apalachicola river basins, respectively. EA=Eastern (*T. c. carolina*), TT=Three-toed (*T. m. triunguis*), GU=Gulf Coast (*T. c. major*), ON=Ornate (*T. o. ornata*).

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

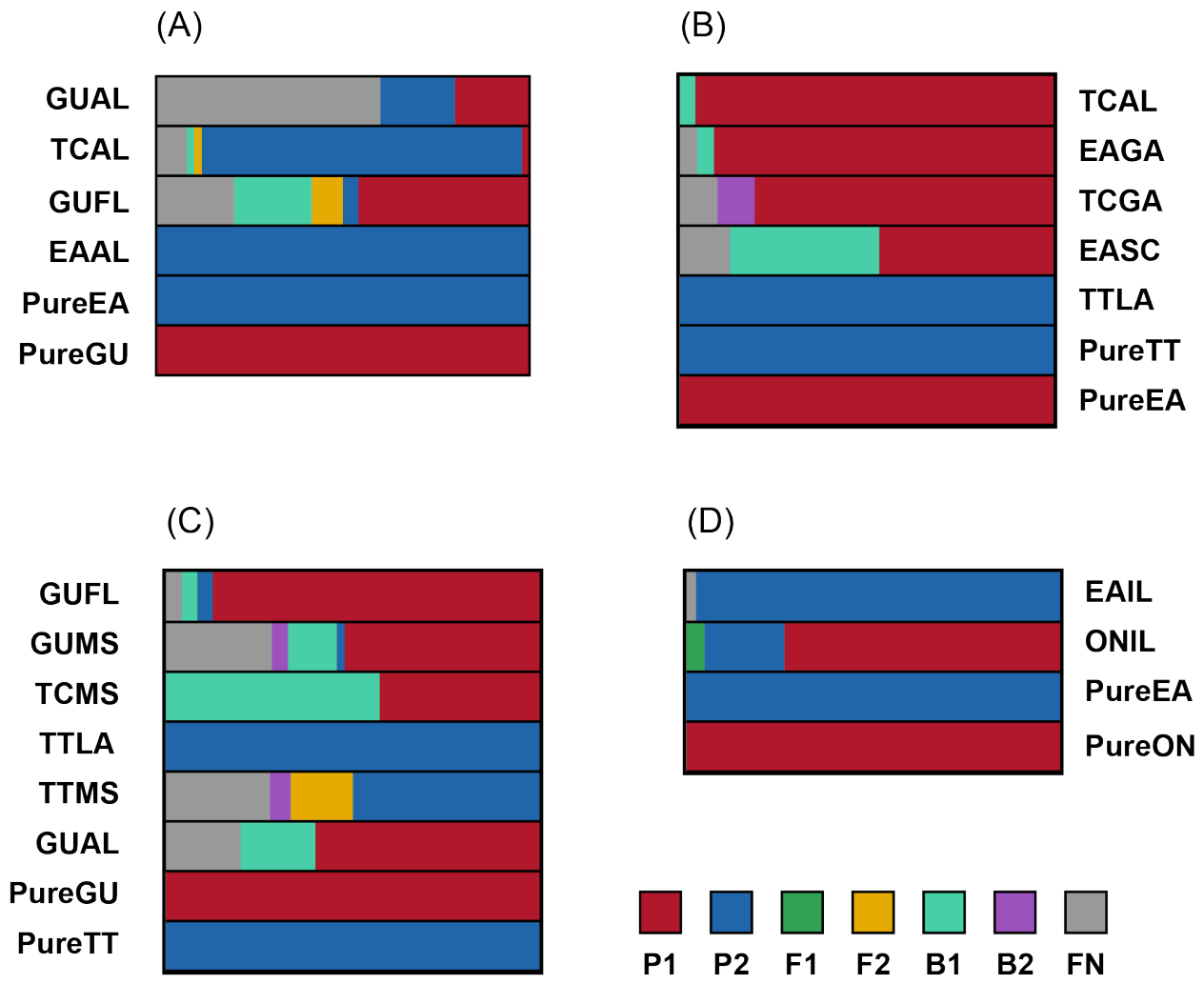

Figure 4: Population-level NEWHYBRIDS plots for four pairs of southeastern and midwestern *Terrapene* taxa. Individuals were collapsed into populations based on field identification at the subspecific level. The first two characters represent: GU=Gulf Coast, *T. c. major*; EA=Eastern, *T. c. carolina*; TT=Three-toed, *T. m. triunguis*; ON=Ornate, *T. o. ornata*; TC=*T. carolina* (limited to species-level field identification). The last two characters represent U.S. state: AL=Alabama, FL=Florida, MS=Mississippi, SC=South Carolina, GA=Georgia, LA=Louisiana, and IL=Illinois). Each plot corresponds to tests between parental groups (A) EAxGU (N=109), (B) EAxTT (N=135), (C) GUxTT (N=139), and (D) EAxON (N=112). A posterior probability threshold of >0.8 was required for genotype frequency class assignments. The genotype classes included P_1 and P_2 (parental types), F_1 and F_2 (first and second-generation hybrids), backcrosses (B_1 and B_2), and FN (unclassified).

bioRxiv preprint doi: https://doi.org/10.1101/752196; this version posted September 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

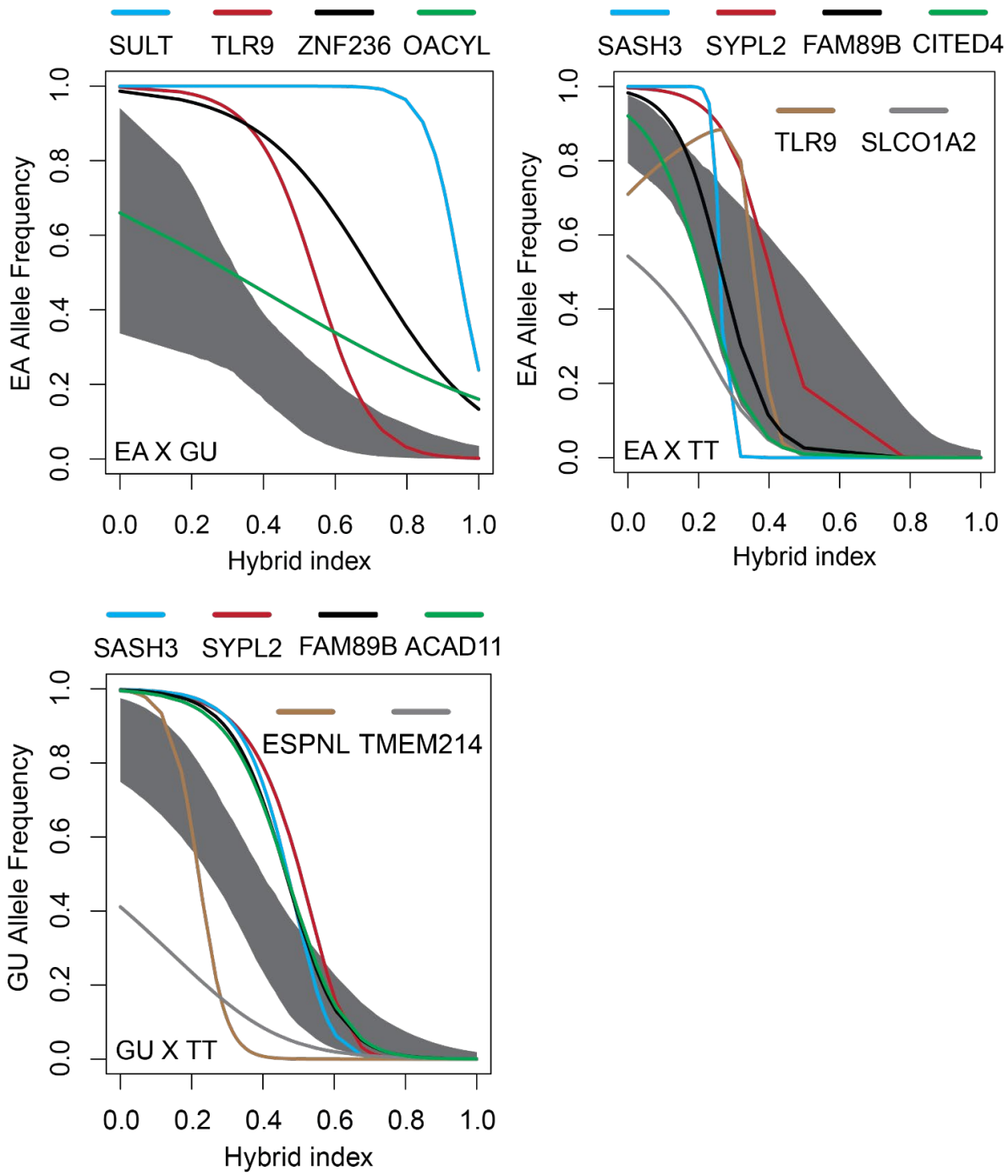

Figure 5: Genomic clines that depict outlier SNPs found in mRNA ddRAD loci. Pairwise comparisons are between EA=Eastern (*T. c. carolina*), GU=Gulf Coast (*T. c. major*), and TT=Three-toed (*T. m. triunguis*) box turtles. The gray area represents neutral expectations based on 2,660 (EAxGU), 2,623 (EAxTT), and 2,622 (GUxTT) transcriptome-aligned SNPs, and each line is a genomic cline for a single outlier locus. Locus abbreviations for each cline are defined in Table 2.