Title: Range-wide pangenomics reveals vulnerability and adaptation in a sedentary bird

Authors: Jong Yoon Jeon[1], Natalie M. Allen[1], Audrey J. Heckel[2], Alberto Macías-Duarte[3], René Jesús Toscano-Meléndrez[4], Jesús Raúl Martínez-Sandoval[5], Angel Montoya[6], John McLaughlin[7], J. Andrew DeWoody[1,8*]

Affiliations:

[1]Department of Forestry and Natural Resources, Purdue University; West Lafayette, Indiana, USA.

[2]Department of Sustainability Engineering and Environmental Engineering, Purdue University; West Lafayette, Indiana, USA.

[3]Cuerpo Académico de Recursos Naturales, Universidad Estatal de Sonora; Hermosillo, Sonora, Mexico.

[4]Maestría en Ciencias Ambientales, Universidad Estatal de Sonora; Hermosillo, Sonora, Mexico.

[5]Programa Académico de Ingeniero en Mecatrónica, Universidad Estatal de Sonora; Hermosillo, Sonora, Mexico.

[6]Las Cruces, New Mexico, USA.

[7]Upland Game Bird Program, Texas Parks & Wildlife Department; Lubbock, Texas, USA.

[8]Department of Biological Sciences, Purdue University; West Lafayette, Indiana, USA.

*Corresponding author. Email: dewoody@purdue.edu

Abstract:

Structural genetic variants are known to underlie evolutionary adaptations, but have never been evaluated across the entire geographic range of a species. We constructed the first range-wide pangenome to evaluate how such variants may influence conservation efforts of the Montezuma quail (*Cyrtonyx montezumae*). Profiles of local adaptation, population structure, and genomic diversity all support a history of isolated glacial refugia and ring-like range expansion. Genomic offset analyses and forward-time simulations consistently indicated that isolated, peripheral populations are particularly vulnerable to climate change. Proactive management (e.g., translocations) could alleviate genetic erosion in fragmented, peripheral populations of this secretive bird. Our study demonstrates that in wild populations, structural variants can be both significant drivers of adaptive potential and genomic sentinels of sustainability.

Main Text:

Structural genetic variants (SVs), including insertions and deletions, represent a substantial yet underexplored source of genomic diversity in natural populations. By altering gene dosage, regulatory landscapes, and even genome structure, SVs often exert disproportionate influence on fitness relative to single nucleotide polymorphisms (SNPs) (*1*, *2*). Such effects may be particularly important for conservation, where evolutionary resilience in the face of rapid environmental change depends on standing functional variation. Nevertheless, SVs remain sparsely characterized in non-model vertebrates for technical reasons such as reference bias (*3*). Genome representations that integrate multiple haplotypes within a unified graph framework (*4*–*6*) offer an innovative path to overcome these technical limitations. Here, we explored the use of such graph-based pangenomes in linking structural variation with environmental adaptation and conservation forecasting (*7*, *8*).

Adaptation to environmentally heterogeneous and fragmented habitats is especially consequential for philopatric species with limited dispersal capacity and climate dependence, where demographic isolation constrains evolutionary potential. The Montezuma quail (*Cyrtonyx montezumae*) is a cryptic and sedentary ground-dwelling Galliform distributed across arid and montane regions (*9*, *10*) in Mexico and the southwestern United States (*11*). Peripheral populations in the US are often isolated, exhibit reduced SNP diversity, elevated inbreeding, and are subject to high genetic drift (*12*). Unknown, however, is whether these patterns arise from demographic history, intrinsic genomic features, local adaptation, or some combination thereof (*12*). The lack of genomic data from core populations in Mexico has prevented meaningful range-wide comparisons as well as the identification of adaptive variants, constraining conservation efforts (*13*). Across the >1000 km latitudinal range of the species, climatic gradients and habitat heterogeneity provide many evolutionary opportunities for adaptive evolution. Given the high mortality of Montezuma quail during winter (*14*), we hypothesized that local adaptation would be especially associated with climatic conditions.

Here we successfully integrate pangenomics, ecological niche modeling, adaptation profiles, and forward-time eco-evolutionary simulations in a conservation framework (Fig. 1). By linking genome structure to population viability and the potential efficacy of genetic rescue, our framework positions structural variation—and the adaptive processes it underpins—as a central component of evolutionary resilience and conservation forecasting under global change, with broad applicability across taxa.

A multi-haplotype pangenome reveals extensive structural variation

We assembled a high-quality linear reference genome (1.01 Gb composed of 287 scaffolds) for Montezuma quail with a N50 of 36.2 Mb and 98.7 % complete BUSCOs (supplementary text, table S1). An additional ten long-read genomes (20 haplotypes), which span much of the species' geographic range, also exhibited high contiguity and low fragmentation for their primary (mean ≈ 1.04 Gb) and alternate (mean ≈ 0.89 Gb) assemblies (table S2). Building upon the reference genome, we constructed a graph pangenome that initially totaled ~1.1 Gb (~10 % larger than the linear reference) and consisted of ~57 million nodes and ~78 million edges. After augmenting the pangenome with sequences from 91 short-read individuals, the graph expanded substantially to 2.02 Gb, comprising ~1.5 billion nodes and ~2.4 billion edges that reflect the deep standing structural diversity across the vast and diverse range of the species.

Using this augmented pangenome, we identified 125,677 variant sites. These sites included 77,727 multiallelic sites across all chromosomes except the W (which was excluded due to coordinate inconsistencies). Across all variant alleles, we observed 98,948 insertions and deletions when compared to the reference genome. After stringent filtering, 30,165 high-confidence variants remained, primarily comprised of 24,195 insertions (mean length 1,055 bp) and 4,792 deletions (mean length 351 bp), with the remainder consisting of 823 SNPs and 355 variants of unknown type (fig. S1, table S3). The predominance of insertions/deletions likely reflects differences in sequencing strategy and variant classification (*2*) relative to long-read-only approaches (*15*). Insertion counts and density increased proportionally with chromosome length whereas deletion density was relatively uniform across chromosomes, suggesting that longer chromosomes provide bigger targets for SV insertions (fig. S2).

Climate adaptation based on structural variation is linked to functional and phenotypic variation

Adaptive SVs are expected to reflect environmental conditions. Using a two-step PCAdapt-Redundancy analysis (RDA) approach for outlier detection (*16*, *17*), we identified 550 environmentally-associated variants among the ~30,000 genome-wide SVs. Of these, >95% (n = 526) were SVs (373 insertions and 153 deletions; table S3, fig. S3). Winter-related climatic variables—the coldest and driest seasonal conditions—emerged as consistent drivers of adaptive genomic differentiation across analytical frameworks. RDA axes were primarily structured by driest-period climate, with driest-quarter mean temperature (Bio9) and driest-month precipitation (Bio14) loading strongly on RDA1, and warmest-month maximum temperature (Bio5) and Bio14 on RDA2. Across ecological niche modeling (ENM), RDA, and functional annotation, driest-month precipitation (Bio14) and precipitation seasonality (Bio15) showed the strongest and most consistent associations (n = 200 and 144 associated outlier SVs, respectively), followed by temperature annual range (Bio7) and driest-quarter mean temperature (Bio9) (fig. S4, table S4). These variables were highly correlated with broader winter climate metrics (e.g., Bio6, Bio11, Bio17 and Bio19; table S5), reinforcing a coherent winter-climate signal.

Environment-associated outlier SVs were enriched on the largest autosomes and were significantly associated with key climatic predictors (Bio7, Bio14, Bio15; fig. S4, table S4), consistent with their importance in ecological niche modeling (table S5). Functional enrichment analyses converged on processes related to DNA damage mitigation, metabolic regulation, and cellular robustness (supplementary text). Because environmental associations do not necessarily imply functional relevance, we examined whether outlier SVs were associated with phenotypic variation. Genotype–phenotype analyses further linked these environmentally associated SVs to plumage variation in these cryptic birds. Plumage lightness (L*) showed the strongest association with the primary genotype axis derived from outlier SVs (fig. S5; supplementary text). This relationship persisted after accounting for neutral population structure but disappeared when environmental structure was removed. This pattern indicates that genotype–phenotype associations are largely mediated by environmentally structured genetic variation rather than direct, environment-independent effects.

Additional analyses corroborate the idea that our suite of outlier SVs underlie adaptation in Montezuma quail. First, the site frequency spectrum of the outlier SVs deviated from neutrality compared with genome-wide SVs (fig. S6). Second, the proportion of deleterious

variants across individuals was highest in outlier SVs (mean = 0.053), followed by genome-wide SVs (mean = 0.045) and SNPs (mean = 0.019) (Fig. 2, tables S6–S8). This pattern was consistent across genetic populations examined (Central Texas + Coahuila, West Texas, and Mexico-wide; see below), indicating that the putatively adaptive SVs harbor a greater proportion of predicted high- and moderate-impact variants than background genomic variation. Together, these results support a winter-driven adaptation framework in which SVs strongly contribute to polygenic responses linking climate, genomic differentiation, and phenotypic variation across the species' range. Furthermore, population-level patterns revealed that the proportion of deleterious variants was lowest in Central Texas + Coahuila, intermediate in West Texas and highest in Mexico-wide populations (Fig. 2). These proportions were inversely correlated with contemporary population size estimates from field surveys (*18–20*), consistent with the idea that more efficient purging in small, isolated populations (*21*, *22*) extends beyond SNPs to include SVs.

SVs define a major east–west adaptive axis and reflect a ring-like northward expansion

As expected by population genetics theory, this SV-based adaptive structure differs from genome-wide neutral patterns. Principal component analysis (PCA) of genome-wide SNPs and SVs, together with admixture analyses (K = 2; fig. S7), primarily separated northern (Texas and Coahuila) populations along neutral axes (Fig. 3, **A** to **C** and **E**). Outlier SVs further separated the northern cluster, representing three distinct clusters: Central Texas + Coahuila, West Texas, and the remaining Mexican individuals (Fig. 3**F**). Individual heterozygosity (range 0.00153–0.00244; mean of 0.00191) declined from south to north, with the lowest values in the northeastern periphery (i.e., Texas; Fig. 3**D**, table S9). The decoupling of neutral and adaptive genomic structure indicates that the observed spatial patterns of local adaptation are unlikely to be explained by genetic drift alone but rather by virtue of natural selection.

Adaptive landscape patterns based on the RDA revealed a pronounced east–west adaptive gradient in local adaptation, with the most distinct profiles identified between the southern and northeastern regions (Fig. 3, **G** to **J**). Central Texas, southeastern Mexico (around Hidalgo and Tlaxcala), and parts of New Mexico formed an "eastern" adaptive regime, whereas mountains in western Mexico (the Sierra Madre Occidental) formed a contrasting "western" regime, with Arizona, New Mexico, and West Texas exhibiting intermediate profiles. Notably, Central Texas + Coahuila shared the same adaptive genomic structure as southeastern Mexico despite harboring the lowest genetic diversity, where genetic drift would be expected to dominate. Collectively, the spatial distribution of heterozygosity and adaptive landscape profiles are consistent with a ring-like northward expansion (*12*).

Climate change threatens peripheral persistence, but targeted genetic rescue mitigates genomic erosion

Climate change can generate genomic mismatch between locally adapted populations and their future environments. Across all three genomic offset approaches (RDA, Generalized Dissimilarity Modeling, and Gradient Forest) (*17*, *23*, *24*), Central Texas consistently exhibited the greatest predicted mismatches between contemporary genomic composition at the adaptive outlier SVs and future environmental conditions (Fig. 4, figs. S8–S11). As expected, projected mismatches intensified under the most extreme climate scenario (SSP585) compared to the most

moderate scenario (SSP126). Reverse-offset analyses (*25*) further indicate that natural dispersal is unlikely to supply sufficient adaptive variants to highly mismatched regions, implying that Central Texas cannot be evolutionarily rescued *in situ* without intervention.

To test whether projected climate risks translate into demographic consequences, we conducted forward-time simulations. Under both climate scenarios, spatial demographic simulations using Geonomics (*26*) predicted that populations will increasingly shift towards the northwestern frontier, where habitats remained relatively stable and connected. In contrast, southern and eastern regions—including Central Texas—retained few or no individuals through time (figs. S12–S13). Landscape genomic simulations using SLiM (*27*) similarly predicted rapid loss of Central Texas and southern Mexico individuals, with Chihuahua emerging as the southernmost persistent region. Across generations, heterozygosity decreased while the fraction of runs of homozygosity (F_{ROH}) also declined, probably due to recombination (fig. S14). Potential load decreased whereas realized load increased (figs. S15–S16), consistent with the combined effects of selection and drift reducing the total number of variants—including deleterious ones—while elevating their homozygosity. These simulations demonstrate that environmental unsuitability and constraints on dispersal jointly heighten extinction risk in peripheral populations.

To evaluate whether assisted gene flow (conducted once annually for three consecutive years) could alter these trajectories, we examined five genetic rescue simulations: a no-rescue baseline, and translocations of either 50 or 100 individuals from either New Mexico or Southeast Mexico. While heterozygosity declined continuously across all observed variant sites (all sites, genome-wide variants, and outlier SVs) in all scenarios (fig. S17), translocations from New Mexico produced immediate increases in heterozygosity across genome-wide and outlier SV sites, accompanied by sustained population growth. Southeast Mexico translocations yielded more limited rescue benefits: increases in heterozygosity were evident primarily at outlier SV sites and only when 100 individuals were introduced, and demographic growth was not sustained. Across rescue scenarios and ROH length classes, F_{ROH} declined sharply following translocation and then gradually increased over time (Fig. 5, and fig. S18). Rescue scenarios involving translocations from New Mexico consistently maintained lower F_{ROH} than the baseline across ROH length classes. In contrast, comparable effects from Southeast Mexico were largely restricted to longer ROH tracts (> 500 kb and > 1 Mb) and only when 100 individuals were introduced.

We further tracked donor ancestry proportion in the recipient population as an indicator of genetic swamping following rescue events. Donor ancestry proportions stabilized following rescue events, reaching ~0.6–0.7 in the 100-individual translocation scenario from New Mexico and ~0.3–0.4 in Southeast Mexico or smaller New Mexico translocations (Fig. 5 and fig. S18). Patterns of potential and realized genomic load were broadly similar (except at outlier SV sites or under high selection coefficients), with consistent load reductions under New Mexico rescue scenarios and more limited effects under Southeast Mexico translocations (Fig. 5, figs. S19–S22). Only when 100 individuals were translocated did Southeast Mexico scenario reduce potential and realized load at outlier SV sites to levels comparable to New Mexico scenarios, particularly for strongly selected variants. Notably, load dynamics (such as stronger purging under higher selection intensity) were most pronounced at outlier SV sites, highlighting the disproportionate contribution of adaptive SVs to genomic fitness trajectories. The consistent load

reductions observed in New Mexico scenarios, coupled with projected demographic growth (fig. S17), suggest that translocations could help forestall inbreeding depression.

Discussion

Our results establish structural variants as central drivers of local adaptation, genomic load, and conservation vulnerability in Montezuma quail—contributions that would remain invisible under neutral, SNP-only frameworks. Winter climate emerges as the primary selective force linking genomic differentiation, functional enrichment, and phenotypic variation across the species' range. Local adaptation to winter climates appears mediated by polygenic SV-based responses that enhance tolerance to cold and arid conditions, including DNA damage mitigation, metabolic regulation, and cellular robustness. Such mechanisms are biologically plausible given that selection is strong in New World quail, where significant winter weather events often lead to mass mortality (*14*). That plumage lightness appears central to climate-mediated adaptation, consistent with previous phenotypic work (*28*), is particularly informative: unlike chromatic axes reflecting hue-based pigmentation, plumage lightness captures structural and condition-dependent properties—such as reflectance, wear and degradation—sensitive to oxidative stress and energetic constraints of cold, dry winters (*29–32*). Cold and dry conditions thus appear to impose differential viability linked to plumage variation (*33*, *34*).

Landscape genomic patterns further illuminate phylogeographic processes. The south–north gradient of heterozygosity, together with the admixture and adaptation patterns, is consistent with a historical ring-like expansion from southern Mexico via the Sierra Madre Oriental and Sierra Madre Occidental mountains (*12*). Eastern and western expansion routes were characterized by demographic and selective histories that are still reflected in the present-day adaptive regimes. These spatial genomic patterns align with the central–marginal hypothesis, whereby genetic diversity is highest in core populations and reduced toward peripheral populations, which may harbor unique adaptive alleles. Elevated diversity in southern Mexico is also consistent with glacial refugia theory (*35*) and long-term persistence within the Mexican highlands, particularly the Trans-Mexican Volcanic Belt (*36*), a region characterized by low climate-change velocity during Pleistocene oscillations. Similar southern refugia have been documented in multiple montane taxa (*36*), supporting a scenario of rear-edge persistence followed by northward expansion. Together, these demographic and adaptive patterns contextualize the climatic inference and underscore the role of SVs in shaping large-scale environmental responses across the species' range.

Implications for conservation under climate change

Montezuma quail exhibit lower genomic diversity than most Galliformes (*37*), implying limited adaptive potential. The IUCN's Green Status framework, designed to evaluate the viability and functional status of a species, can be refined based on genomic metrics such as heterozygosity (*38*). With a mean heterozygosity of 0.00191, the corrected Green Score of Montezuma quail is 0.131. This means that Montezuma quail retains only ~13% of the genomic diversity observed in non-threatened birds from the same taxonomic family, far from full ecological recovery. Multiple lines of evidence (genomic offset projections, ecological niche forecasts, spatial demographic simulations, and landscape genomic simulations) consistently

converge on two conclusions: (1) climate change is expected to reduce habitat suitability and increase genomic mismatch across the species' range, threatening marginal populations; (2) the Central Texas population emerges as a pronounced vulnerability hotspot that may not persist without intervention.

Importantly, Central Texas harbors a distinct SV-mediated adaptive signature. This pattern may reflect either a legacy of historical northward expansion or continued strong natural selection, despite pronounced genomic erosion (*39*). Its loss would therefore represent not merely demographic contraction but erosion of unique adaptive variation within the species. At the same time, its geographic isolation restricts natural gene flow, and large-scale translocations risk genetic swamping of locally adapted variation. These findings highlight a fundamental management trade-off: mitigating genomic erosion through assisted gene flow while preserving adaptive distinctiveness.

Adaptive structural variation as an essential component of conservation genomics

More broadly, our results demonstrate that assessments based solely on neutral diversity or SNP-based metrics risk overlooking critical components of both adaptive potential and extinction vulnerability. Structural variants accounted for the vast majority of environmentally associated loci, disproportionately shaping not only adaptive genomic structure but also genomic load patterns and climate-mediated phenotypic variation—proxies both closely tied to fitness—and drove rescue dynamics in forward-time simulations. Ignoring such variation could therefore underestimate both vulnerability and rescue potential under climate change. Collectively, our study indicates that the integration of SVs into conservation frameworks will be essential for accurately forecasting persistence and designing effective intervention strategies. Our integrative pangenomic approach, linking genome structure to population viability and genetic rescue efficacy, provides a general framework that has substantial advantages in evaluating the landscape of adaptive genomic characteristics and guiding conservation strategies in a rapidly changing world.

References and Notes

1. C. Mérot, R. A. Oomen, A. Tigano, M. Wellenreuther, A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology & Evolution* **35**, 561–572 (2020).

2. K. C. Stuart, R. A. Oomen, A. Tigano, M. Wellenreuther, J. Wold, D. L. Field, C. Mérot, A beginner's guide to structural variants in eco-evolutionary population genomics. *Molecular Ecology* **35**, e70216 (2026).

3. T. M. Pegan, A. H. Sivakumar, C. A. Burns, K. A. Lopez, B. Fang, N. M. Melendez, L. Bartel, S. V. Edwards, The value of structural variants to conservation genomics in the pangenome era. *Journal of Heredity*, esaf098 (2025).

4. E. Garrison, J. Sirén, A. M. Novak, G. Hickey, J. M. Eizenga, E. T. Dawson, W. Jones, S. Garg, C. Markello, M. F. Lin, B. Paten, R. Durbin, Variation graph toolkit improves read

mapping by representing genetic variation in the reference. *Nature Biotechnology* **36**, 875–879 (2018).

5. Y. Zhou, L. Yang, X. Han, J. Han, Y. Hu, F. Li, H. Xia, L. Peng, C. Boschiero, B. D. Rosen, D. M. Bickhart, S. Zhang, A. Guo, C. P. V. Tassell, T. P. L. Smith, L. Yang, G. E. Liu, Assembly of a pangenome for global cattle reveals missing sequences and novel structural variations, providing new insights into their diversity and evolutionary history. *Genome Research* **32**, 1585–1601 (2022).

6. R. M. Sherman, S. L. Salzberg, Pan-genomics in the human genome era. *Nature Reviews Genetics* **21**, 243–254 (2020).

7. J. Y. Jeon, N. M. Allen, A. N. Black, J. A. DeWoody, Short-read pangenomes and their potential utility in population and conservation genomics. *Molecular Ecology Resources* **26**, e70060 (2026).

8. J. Wold, K.-P. Koepfli, S. J. Galla, D. Eccles, C. J. Hogg, M. F. Le Lec, J. Guhlin, A. W. Santure, T. E. Steeves, Expanding the conservation genomics toolbox: Incorporating structural variants to enhance genomic studies for species of conservation concern. *Molecular Ecology* **30**, 5949–5965 (2021).

9. M. R. Stromberg, Habitat, movements and roost characteristics of Montezuma quail in southeastern Arizona. *The Condor: Ornithological Applications* **92**, 229–236 (1990).

10. Bishop, R. A., "The Mearns quail (*Cytronyx montezumae mearnsi*) in southern Arizona," thesis, The University of Arizona (1964).

11. Stromberg, M. R., Montoya, A. B., Holdermann, D., "Montezuma Quail (*Cyrtonyx montezumae*), version 1.0" in *Birds of the World*, Rodewald, P. G., Ed. (Cornell Lab of Ornithology, Ithaca, NY, USA, 2020; https://doi.org/10.2173/bow.monqua.01).

12. S. Mathur, J. M. Tomeček, L. A. Tarango-Arámbula, R. M. Perez, J. A. DeWoody, An evolutionary perspective on genetic load in small, isolated populations as informed by whole genome resequencing and forward-time simulations. *Evolution* **77**, 690–704 (2023).

13. M. Osborne, A. Sharp, J. Monzingo, D. L. Propst, T. F. Turner, Genetic analysis suggests high conservation value of peripheral populations of Chihuahau chub (*Gila nigrescens*). *Conservation Genetics* **13**, 1317–1328 (2012).

14. P. Chavarria, A. Montoya, N. Silvy, R. Lopez, Impact of inclement weather on overwinter mortality of Montezuma quail in southeast Arizona. *National Quail Symposium Proceedings* **7**, Article 124 (2017).

15. S. V. Edwards, B. Fang, D. Khost, G. E. Kolyfetis, R. G. Cheek, D. A. DeRaad, N. Chen, J. W. Fitzpatrick, J. E. McCormack, W. C. Funk, C. K. Ghalambor, E. Garrison, A. Guarracino, H. Li, T. B. Sackton, Multispecies pangenomes reveal a pervasive influence of population size on structural variation. *Science* **390**, eadw1931 (2025).

16. F. Privé, K. Luu, B. J. Vilhjálmsson, M. G. B. Blum, Performing highly efficient genome scans for local adaptation with R package pcadapt Version 4. *Molecular Biology and Evolution* **37**, 2153–2154 (2020).

17. T. Capblancq, B. R. Forester, Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods in Ecology and Evolution* **12**, 2298–2309 (2021).

18. L. H. Palafox, G. D. M. Martínez, J. I. V. Hernández, J. L. A. Carbajal, F. C. Sánchez, M. H. Téllez, Population density and habitat use of Montezuma quail (*Cyrtonyx montezumae*) in Hualtepec, Hidalgo. *Veterinaria México* **38**, 409–417 (2007).

19. Z. J. Pearson, "Montezuma Quail Habitat and Occupancy in the Southern Edwards Plateau of Texas," thesis, Texas A&M University - Kingsville (2018).

20. J. T. Robles, S. I. Zimbrón, G. D. M. Martínez, C. Z. Hernández, F. C. Sánchez, L. A. T. Arámbula, Population density of Montezuma quail (*Cyrtonyx montezumae*) in the northeastern region of the State of Mexico, Mexico. *Veterinaria México* **33**, 255–263 (2002).

21. C. Grossen, F. Guillaume, L. F. Keller, D. Croll, Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nature Communications* **11**, 1001 (2020).

22. S. Mathur, J. A. DeWoody, Genetic load has potential in large populations but is realized in small inbred populations. *Evolutionary Applications* **14**, 1540–1557 (2021).

23. Fitzpatrick, M. C., Mokany, K., Manion, G., Nieto-Lugilde, D., Ferrier, S., gdm: Generalized Dissimilarity Modeling, version 1.6 (2024).

24. N. Ellis, S. J. Smith, C. R. Pitcher, Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93**, 156–168 (2012).

25. A. V. Gougherty, S. R. Keller, M. C. Fitzpatrick, Maladaptation, migration and extirpation fuel climate change risk in a forest tree species. *Nature Climate Change* **11**, 166–171 (2021).

26. D. E. Terasaki Hart, A. P. Bishop, I. J. Wang, Geonomics: Forward-time, spatially explicit, and arbitrarily complex landscape genomic simulations. *Molecular Biology and Evolution* **38**, 4634–4646 (2021).

27. B. C. Haller, P. W. Messer, SLiM 4: Multispecies eco-evolutionary modeling. *The American Naturalist* **201**, E127–E139 (2023).

28. Toscano Meléndrez, R. J., "Local and Geographic Colorimetric Variation of the Montezuma Quail Plumage (*Cyrtonyx montezumae* vigors 1830)," thesis, Sonora State University (2025).

29. M. N. Vitousek, R. A. Stewart, R. J. Safran, Female plumage colour influences seasonal oxidative damage and testosterone profiles in a songbird. *Biology Letters* **9**, 20130539 (2013).

30. I. Galván, A. P. Møller, Pheomelanin-based plumage coloration predicts survival rates in birds. *Physiological and Biochemical Zoology* **86**, 184–192 (2013).

31. M. N. Vitousek, O. Tomášek, T. Albrecht, M. R. Wilkins, R. J. Safran, Signal traits and oxidative stress: A comparative study across populations with divergent signals. *Frontiers in Ecology and Evolution* **4** (2016).

32. S. Losdat, J. D. Blount, V. Marri, L. Maronde, H. Richner, F. Helfenstein, Effects of an early-life paraquat exposure on adult resistance to oxidative stress, plumage colour and sperm performance in a wild bird. *Journal of Animal Ecology* **87**, 1137–1148 (2018).

33. P. Karell, K. Ahola, T. Karstinen, J. Valkama, J. E. Brommer, Climate change drives microevolution in a wild bird. *Nature Communications* **2**, 208 (2011).

34. T. Solonen, Does plumage colour signal fitness in the tawny owl *Strix aluco*? *Journal of Avian Biology* **52** (2021).

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

36. A. Mastretta-Yanes, A. Moreno-Letelier, D. Piñero, T. H. Jorgensen, B. C. Emerson, Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. *Journal of Biogeography* **42**, 1586–1600 (2015).

37. A. Brüniche-Olsen, K. F. Kellner, J. L. Belant, J. A. DeWoody, Life-history traits and habitat availability shape genomic diversity in birds: implications for conservation. *Proceedings of the Royal Society B: Biological Sciences* **288**, 20211441 (2021).

38. J. Y. Jeon, A. N. Black, E. J. Heenkenda, A. J. Mularo, G. F. Lamka, S. Janjua, A. Brüniche-Olsen, J. W. Bickham, J. R. Willoughby, J. A. DeWoody, Genomic diversity as a key conservation criterion: Proof-of-concept from mammalian whole-genome resequencing data. *Evolutionary Applications* **17**, e70000 (2024).

39. S. Mathur, J. M. Tomeček, A. Heniff, R. Luna, J. A. DeWoody, Evidence of genetic erosion in a peripheral population of a North American game bird: the Montezuma quail (*Cyrtonyx montezumae*). *Conservation Genetics* **20**, 1369–1381 (2019).

40. Jong Yoon Jeon, Natalie M. Allen, J. Andrew DeWoody, Pangenome data for: A landscape-level pangenomic framework reveals adaptation based on structural variants and climate-driven vulnerability in a sedentary bird, version 1.0, PURR (2026); https://doi.org/10.4231/VFC3-JQ04.

41. Andrews, S., FastQC: A quality control tool for high throughput sequence data, (2010); http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

42. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

43. B. Bushnell, “BBMap: A fast, accurate, splice-aware aligner” (LBNL-7065E, Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA (US), 2014); https://www.osti.gov/biblio/1241166.

44. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**, 170–175 (2021).

45. D. Guan, S. A. McCarthy, J. Wood, K. Howe, Y. Wang, R. Durbin, Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).

46. D. E. Wood, J. Lu, B. Langmead, Improved metagenomic analysis with Kraken 2. *Genome Biology* **20**, 257 (2019).

47. A. Morgulis, E. M. Gertz, A. A. Schäffer, R. Agarwala, WindowMasker: window-based masker for sequenced genomes. *Bioinformatics* **22**, 134–141 (2006).

48. Smit, A. F. A., Hubley, R., Green, P., RepeatMasker Open-4.0, (2013); http://www.repeatmasker.org.

49. G. Hickey, J. Monlong, J. Ebler, A. M. Novak, J. M. Eizenga, Y. Gao, T. Marschall, H. Li, B. Paten, Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nature Biotechnology* **42**, 663–673 (2024).

50. J. Sirén, J. Monlong, X. Chang, A. M. Novak, J. M. Eizenga, C. Markello, J. A. Sibbesen, G. Hickey, P.-C. Chang, A. Carroll, N. Gupta, S. Gabriel, T. W. Blackwell, A. Ratan, K. D. Taylor, S. S. Rich, J. I. Rotter, D. Haussler, E. Garrison, B. Paten, Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. *Science* **374**, abg8871 (2021).

51. G. Hickey, D. Heller, J. Monlong, J. A. Sibbesen, J. Sirén, J. Eizenga, E. T. Dawson, E. Garrison, A. M. Novak, B. Paten, Genotyping structural variants in pangenome graphs using the vg toolkit. *Genome Biology* **21**, 35 (2020).

52. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

53. E. Garrison, Z. N. Kronenberg, E. T. Dawson, B. S. Pedersen, P. Prins, A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and slivar. *PLoS Computational Biology* **18**, e1009123 (2022).

54. G. Marçais, C. Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).

55. G. W. Vurture, F. J. Sedlazeck, M. Nattestad, C. J. Underwood, H. Fang, J. Gurtowski, M. C. Schatz, GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).

56. D. R. Laetsch, M. L. Blaxter, BlobTools: Interrogation of genome assemblies. *F1000Research* **6**, 1287 (2017).

57. L. Coombe, J. Zhang, B. P. Vandervalk, J. Chu, S. D. Jackman, I. Birol, R. L. Warren, ARKS: chromosome-scale scaffolding of human genome drafts with linked read kmers. *BMC Bioinformatics* **19**, 234 (2018).

58. M. Manni, M. R. Berkeley, M. Seppey, E. M. Zdobnov, BUSCO: Assessing genomic data quality and beyond. *Current Protocols* **1**, e323 (2021).

59. A. Mikheenko, A. Prjibelski, V. Saveliev, D. Antipov, A. Gurevich, Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**, i142–i150 (2018).

60. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 (2013).

61. B. S. Pedersen, A. R. Quinlan, Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).

62. M. Alonge, S. Soyk, S. Ramakrishnan, X. Wang, S. Goodwin, F. J. Sedlazeck, Z. B. Lippman, M. C. Schatz, RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biology* **20**, 224 (2019).

63. J. Oksanen, G. L. Simpson, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, P. Solymos, M. H. H. Stevens, E. Szoecs, H. Wagner, M. Barbour, M. Bedward, B. Bolker, D. Borcard, T. Borman, G. Carvalho, M. Chirico, M. D. Caceres, S. Durand, H. B. A. Evangelista, R. FitzJohn, M. Friendly, B. Furneaux, G. Hannigan, M. O. Hill, L. Lahti, C. Martino, D. McGlinn, M.-H. Ouellette, E. R. Cunha, T. Smith, A. Stier, C. J. F. T. Braak, J. Weedon, vegan: Community ecology package, version 2.7-2 (2025); https://cran.r-project.org/web/packages/vegan/index.html.

64. D. N. Karger, O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. E. Zimmermann, H. P. Linder, M. Kessler, Climatologies at high resolution for the earth's land surface areas. *Scientific Data* **4**, 170122 (2017).

65. U.S. Geological Survey, Shuttle Radar Topography Mission (SRTM) 1 arc-second Global, (2014); https://doi.org/10.5066/F7PR7TFT.

66. J. Josse, F. Husson, missMDA: A package for handling missing values in multivariate data analysis. *Journal of Statistical Software* **70**, 1–31 (2016).

67. R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing (2022); https://www.R-project.org/.

68. A. C. English, V. K. Menon, R. A. Gibbs, G. A. Metcalf, F. J. Sedlazeck, Truvari: refined structural variant comparison preserves allelic diversity. *Genome Biology* **23**, 271 (2022).

69. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).

70. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

71. M. D. Young, M. J. Wakefield, G. K. Smyth, A. Oshlack, Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11**, R14 (2010).

72. W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, F. Cunningham, The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 122 (2016).

73. Lüdecke, Daniel, sjstats: Statistical functions for regression models, version 0.19.1 (2025); https://doi.org/10.5281/zenodo.1284472.

74. Rohart, F., Gautier, B., Singh, A., Le Cao, K.-A., mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Computational Biology* **13**, e1005752 (2017).

75. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).

76. C. Pockrandt, M. Alzamel, C. S. Iliopoulos, K. Reinert, GenMap: ultra-fast computation of genome mappability. *Bioinformatics* **36**, 3687–3692 (2020).

77. B. Gräler, E. Pebesma, G. Heuvelink, Spatio-temporal interpolation using gstat. *The R Journal* **8**, 204–218 (2016).

78. E. A. Fox, A. E. Wright, M. Fumagalli, F. G. Vieira, ngsLD: evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics* **35**, 3855–3856 (2019).

79. J. Meisner, A. Albrechtsen, Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719–731 (2018).

80. L. Skotte, T. S. Korneliussen, A. Albrechtsen, Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).

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

82. K. Caye, T. M. Deist, H. Martins, O. Michel, O. François, TESS3: fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources* **16**, 540–548 (2016).

83. L. Miller, Notes on the Mearns quail. *The Condor: Ornithological Applications* **45**, 104–109 (1943).

84. A. S. Leopold, R. A. McCabe, Natural history of the Montezuma quail in Mexico. *The Condor: Ornithological Applications* **59**, 3–26 (1957).

85. R. Engel-Wilson, W. Kuvlesky, Arizona qquail: Species in jeopardy? *National Quail Symposium Proceedings* **5**, Article 1 (2017).

86. E. A. Oaster, "Ecology of Mearns' quail (*Cyrtonyx montezumae*) in the Capitan Mountains of New Mexico," thesis, Sul Ross State University (2018).

87. A. H. Hirzel, G. Le Lay, Habitat suitability modelling and niche theory. *Journal of Applied Ecology* **45**, 1372–1381 (2008).

88. J. Y. Jeon, Y. Shin, A. J. Mularo, X. Feng, J. A. DeWoody, The integration of whole-genome resequencing and ecological niche modelling to conserve profiles of local adaptation. *Diversity and Distributions* **30**, e13847 (2024).

89. S. A. Chamberlain, C. Boettiger, R Python, and Ruby clients for GBIF species occurrence data. PeerJ Preprints 5:e3304v1 [Preprint] (2017). https://doi.org/10.7287/peerj.preprints.3304v1.

90. Barve, Vijay, Hart, Edmund, rinat: Access "iNaturalist" data through APIs, (2025); https://docs.ropensci.org/rinat/.

91. P. Chavarria, N. Silvy, R. Lopez, D. Davis, A. Montoya, Ranges and movements of Montezuma quail in southeast Arizona. *National Quail Symposium Proceedings* **8**, Article 94 (2017).

92. A. Zizka, D. Silvestro, T. Andermann, J. Azevedo, C. Duarte Ritter, D. Edler, H. Farooq, A. Herdean, M. Ariza, R. Scharn, S. Svantesson, N. Wengström, V. Zizka, A. Antonelli, CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution* **10**, 744–751 (2019).

93. Chamberlain, Scott, scrubr: Clean biological occurrence records, (2024); https://github.com/ropensci/scrubr.

94. Esri Inc., ArcGIS Pro, version 3.2, Esri Inc. (2023).

95. Greene, C. D., "Ecology of Montezuma quail in the Davis Mountains of Texas," thesis, Sul Ross State University (2011).

96. J. L. Brown, J. R. Bennett, C. M. French, SDMtoolbox 2.0: the next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *PeerJ* **5**, e4095 (2017).

97. R. J. Hijmans, S. Phillips, J. Leathwick, J. Elith, dismo: Species distribution modeling, (2017).

98. Posit Team, RStudio: Integrated development environment for R, Posit Software, PBC (2025).

99. M. Guéguen, H. Blancheteau, R. Lemaire-Patin, W. Thuiller, biomod2 - extending presence-absence species distribution models to multiple data types. Authorea [Preprint] (2025). https://www.authorea.com/doi/full/10.22541/au.176606476.60414605/v1?commit=408f8cfc8e680701c4198900fb50cb42fa654729.

100. J. M. Kass, R. Muscarella, P. J. Galante, C. L. Bohl, G. E. Pinilla-Buitrago, R. A. Boria, M. Soley-Guardia, R. P. Anderson, ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. *Methods in Ecology and Evolution* **12**, 1602–1608 (2021).

101. K. Zhu, C. W. Woodall, J. S. Clark, Failure to migrate: lack of tree range expansion in response to climate change. *Global Change Biology* **18**, 1042–1052 (2012).

102. Trapletti, Adrian, Hornik, Kurt, tseries: Time series analysis and computational finance, (2024); https://CRAN.R-project.org/package=tseries.

103. E. T. Chevy, J. Min, V. Caudill, S. E. Champer, B. C. Haller, C. T. Rehmann, C. C. R. Smith, S. Tittes, P. W. Messer, A. D. Kern, S. Ramachandran, P. L. Ralph, Population genetics meets ecology: A guide to individual-based simulations in continuous landscapes. *Ecology and Evolution* **15**, e71098 (2025).

104. C. D. Greene, L. A. Harveson, J. G. Cross, R. S. Luna, Spatial ecology and habitat utilization of Montezuma quail in the Davis Mountains of Texas. *Texas Journal of Science* **72**, 4 (2020).

105. J. A. Robinson, C. C. Kyriazis, S. F. Nigenda-Morales, A. C. Beichman, L. Rojas-Bracho, K. M. Robertson, M. C. Fontaine, R. K. Wayne, K. E. Lohmueller, B. L. Taylor, P. A. Morin, The critically endangered vaquita is not doomed to extinction by inbreeding depression. *Science* **376**, 635–639 (2022).

106. E. Santiago, C. Köpke, A. Caballero, Accounting for population structure and data quality in demographic inference with linkage disequilibrium methods. *Nature Communications* **16**, 6054 (2025).

107. S. Secomandi, G. R. Gallo, M. Sozzoni, A. Iannucci, E. Galati, L. Abueg, J. Balacco, M. Caprioli, W. Chow, C. Ciofi, J. Collins, O. Fedrigo, L. Ferretti, A. Fungtammasan, B. Haase, K. Howe, W. Kwak, G. Lombardo, P. Masterson, G. Messina, A. P. Møller, J. Mountcastle, T. A. Mousseau, J. F. Obiol, A. Olivieri, A. Rhie, D. Rubolini, M. Saclier, R. Stanyon, D. Stucki, F. Thibaud-Nissen, J. Torrance, A. Torroni, K. Weber, R. Ambrosini, A. Bonisoli-Alquati, E. D. Jarvis, L. Gianfranceschi, G. Formenti, A chromosome-level reference genome and pangenome for barn swallow population genomics. *Cell Reports* **42** (2023).

108. B. Fang, S. V. Edwards, Fitness consequences of structural variation inferred from a House Finch pangenome. *Proceedings of the National Academy of Sciences of the United States of America* **121**, e2409943121 (2024).

109. M. D. Rodriguez, R. A. Bay, K. C. Ruegg, Telomere length differences indicate climate change-induced stress and population decline in a migratory bird. *Molecular Ecology* **34**, e17642 (2025).

110. K. Bilham, C. Newman, C. D. Buesching, M. J. Noonan, A. Boyd, A. L. Smith, D. W. Macdonald, Effects of weather conditions on oxidative stress, oxidative damage, and antioxidant capacity in a wild-living mammal, the European badger (*Meles meles*). *Physiological and Biochemical Zoology* **91**, 987–1004 (2018).

111. R. Tian, C. Yang, S.-M. Chai, H. Guo, I. Seim, G. Yang, Evolutionary impacts of purine metabolism genes on mammalian oxidative stress adaptation. *Zoological Research* **43**, 241–254 (2022).

112. Z. Pang, P. Yang, K. Cai, W. Wang, A. E. Afe, Y. Pan, L. Qiao, W. Liu, Integrating whole-genome resequencing data reveals adaptive selection signatures in sheep populations under extreme environments. *BMC Genomics* **26**, 1069 (2025).

113. P. Lipaeva, K. Vereshchagina, P. Drozdova, L. Jakob, E. Kondrateva, M. Lucassen, D. Bedulina, M. Timofeyev, P. Stadler, T. Luckenbach, Different ways to play it cool: Transcriptomic analysis sheds light on different activity patterns of three amphipod species under long-term cold exposure. *Molecular Ecology* **30**, 5735–5751 (2021).

114. M. Al-Abri, K. A. Kharousi, A. A. Hamrashdi, A. G. A. Toobi, M. M. I. Salem, Genome wide association analysis for twinning ability in Jabal Akhdar Omani goats. *Small Ruminant Research* **221**, 106951 (2023).

115. X. Wu, J. Pei, L. Xiong, Q. Ge, P. Bao, C. Liang, P. Yan, X. Guo, Genome-wide scan for selection signatures reveals novel insights into the adaptive capacity characteristics in three Chinese cattle breeds. *BMC Genomics* **26**, 206 (2025).

116. C. Li, J. Han, T. Chang, H. Jafari, Q. Yang, J. Guo, C. Lei, R. Dang, Whole-genome analysis reveals genetic diversity and selection pressure in Sichuan donkey. *BMC Genomics* **26**, 1072 (2025).

117. Y.-F. Ma, X.-M. Han, C.-P. Huang, L. Zhong, A. C. Adeola, D. M. Irwin, H.-B. Xie, Y.-P. Zhang, Population genomics analysis revealed origin and high-altitude adaptation of Tibetan pigs. *Scientific Reports* **9**, 11463 (2019).

118. J. Chen, I. Jakovlić, M. Sablin, S. Xia, Z. Xu, Y. Guo, R. Kuang, J. Zhong, Y. Jia, N. T. T. Tran, H. Yang, H. Ma, N. Šprem, J. Han, D. Liu, Y. Zhao, S. Zhao, Positive selection on rare variants of IGF1R and BRD4 underlying the cold adaptation of wild boar. *Genetics Selection Evolution* **57**, 40 (2025).

119. M. Li, X. Li, Z. Wu, G. Zhang, N. Wang, M. Dou, S. Liu, C. Yang, G. Meng, H. Sun, C. Hvilsom, G. Xie, Y. Li, Z. hui Li, W. Wang, Y. Jiang, R. Heller, Y. Wang, Convergent molecular evolution of thermogenesis and circadian rhythm in Arctic ruminants. *Proceedings of the Royal Society B: Biological Sciences* **290**, 20230538 (2023).

120. Campo, O. A. D., "Using Physiological and Molecular Approaches to Study Micro-and Macro-Evolutionary Patterns of Selected Waterbirds of the High Andes," thesis, University of Miami (2024).

121. M. Jin, H. Wang, G. Liu, J. Lu, Z. Yuan, T. Li, E. Liu, Z. Lu, L. Du, C. Wei, Whole-genome resequencing of Chinese indigenous sheep provides insight into the genetic basis underlying climate adaptation. *Genetics Selection Evolution* **56**, 26 (2024).

122. Z. Han, W. Zhou, L. Zhang, R. Wang, C. Liu, X. Bai, S. Liu, Genetic diversity and runs of homozygosity analysis of Hetian sheep populations revealed by Illumina Ovine SNP50K BeadChip. *Frontiers in Ecology and Evolution* **11** (2023).

123. S. F. Nigenda-Morales, Y. Hu, J. C. Beasley, H. A. Ruiz-Piña, D. Valenzuela-Galván, R. K. Wayne, Transcriptomic analysis of skin pigmentation variation in the Virginia opossum (*Didelphis virginiana*). *Molecular Ecology* **27**, 2680–2697 (2018).

124. Y. Wang, C. Zhu, Z. Wang, W. Song, L. Lu, Z. Tao, W. Xu, S. Zhang, W. Zhou, H. Liu, H. Li, RNA sequencing analysis reveals key genes and pathways associated with feather pigmentation in mule ducks. *Animal Genetics* **56**, e70007 (2025).

125. H. Liu, Y. Xi, Q. Tang, J. Qi, Z. Zhou, Z. Guo, W. Fan, J. Hu, Y. Xu, S. Liang, M. Xie, J. Tang, Y. Zhang, S. Hou, Genetic fine-mapping reveals single nucleotide polymorphism mutations in the MC1R regulatory region associated with duck melanism. *Molecular Ecology* **32**, 3076–3088 (2023).

126. L. Zhang, M. Wan, R. Tohti, D. Jin, T. P. Zhong, L. Zhang, M. Wan, R. Tohti, D. Jin, T. P. Zhong, Requirement of zebrafish Adcy3a and Adcy5 in melanosome dispersion and melanocyte stripe formation. *International Journal of Molecular Sciences* **23** (2022).

127. L. Yang, W. Zhao, S. Chen, L. Xue, J. Tian, H. Xu, H. Zhang, H. Wang, Y. Gu, J. Zhang, Whole genome resequencing reveals genetic markers for plumage colour in Jingyuan Chicken. *Poultry Science* **104**, 105666 (2025).

128. Q. Guo, Y. Jiang, Z. Wang, Y. Bi, G. Chen, H. Bai, G. Chang, Q. Guo, Y. Jiang, Z. Wang, Y. Bi, G. Chen, H. Bai, G. Chang, Genome-wide analysis identifies candidate genes encoding feather color in ducks. *Genes* **13** (2022).

Acknowledgments:

Biological samples used in the study were collected under Texas Parks and Wildlife Department Scientific Permit SPR-0410-139 and under Scientific Collection Authorizations SGPA/DGVS/00787/21 (amended by SPARN/DGVS/09377/24), SGPA/DGVS/00363/22, SGPA/DGVS/05321/22 (amended by SPARN/DGVS/09398/24), SPARN/DGVS/02020/23 (amended by SPARN/DGVS/09396/24), SPARN/DGVS/10711/23 (amended by SPARN/DGVS/09402/24), SPARN/DGVS/11844/24, and SBRA/DGVS/01113/25 issued by the Dirección General de Vida Silvestre of Mexico. Samples collected in Mexico were imported into

the U.S.A. with Wildlife Export Certificates No. 53424 and No. 54484 issued by Dirección General de Vida Silvestre of Mexico. Mexico samples were imported to the United States under USDA APHIS permit 639-23-361-30776.

We are sincerely grateful to the many professionals, academics, landowners, ranchers, falconers, hunters, and hunting guides in Mexico who generously contributed to the collection of specimens. We thank R. Rodríguez at the Universidad Estatal de Sonora. We also thank DeWoody Lab members (X. Guan, B. Kahner, G. Lamka, S. Kreling, A. Tyagi, and J.W. Bickham), as well as P. Schutz (Texas Parks and Wildlife Department) and L. Gibbs (The Ohio State University) for their feedback on an early version of this paper. We are grateful to A. Hernandez for his sequencing advice and K. Walden for assembly advice. D. Terasaki Hart kindly guided the use of Geonomics and updated the software for us. P. Ralph provided practical advice on designing landscape genomic simulations within the SLiM framework. ChatGPT (OpenAI) was used to assist with grammar correction and language refinement of the initial manuscript draft written by the first author, but all subsequent revisions were conducted exclusively by the authors.

Funding:

U.S. Fish and Wildlife Service Award F22AF01479 (AM-D, JAD)

Texas Parks and Wildlife Department Contract CA-0004020 (AM-D, JAD)

Author contributions:

Conceptualization: JYJ, AM-D, JAD

Methodology: JYJ, NMA, AJH

Investigation: AM-D, AM-D, RJT-M, JRM-S, AM, JM

Visualization: JYJ, NMA

Funding acquisition: AM-D, JAD

Project administration: AM-D, JM, JAD

Supervision: JAD

Writing – original draft: JYJ, NMA, AJH

Writing – review & editing: JYJ, NMA, AJH, AM-D, AM, JM, JAD

Competing interests: Authors declare that they have no competing interests.

Data, code, and materials availability:

Newly generated sequencing data (PRJNA1085761, PRJNA1232374, PRJNA1232374) and reference genome assembly (RefSeq GCF_038088225.1) are available in NCBI's GenBank. Long-read-based graph pangenome and short-read-augmented graph pangenome data are available at https://purr.purdue.edu/projects/monqlangen (DOI: 10.4231/VFC3-JQ04)(*40*). All scripts used in the study are available at: https://github.com/jyj5558/monq_adaptive_genomics.

Supplementary Materials

Materials and Methods

Supplementary Text

Figs. S1 to S24

Tables S1 to S13

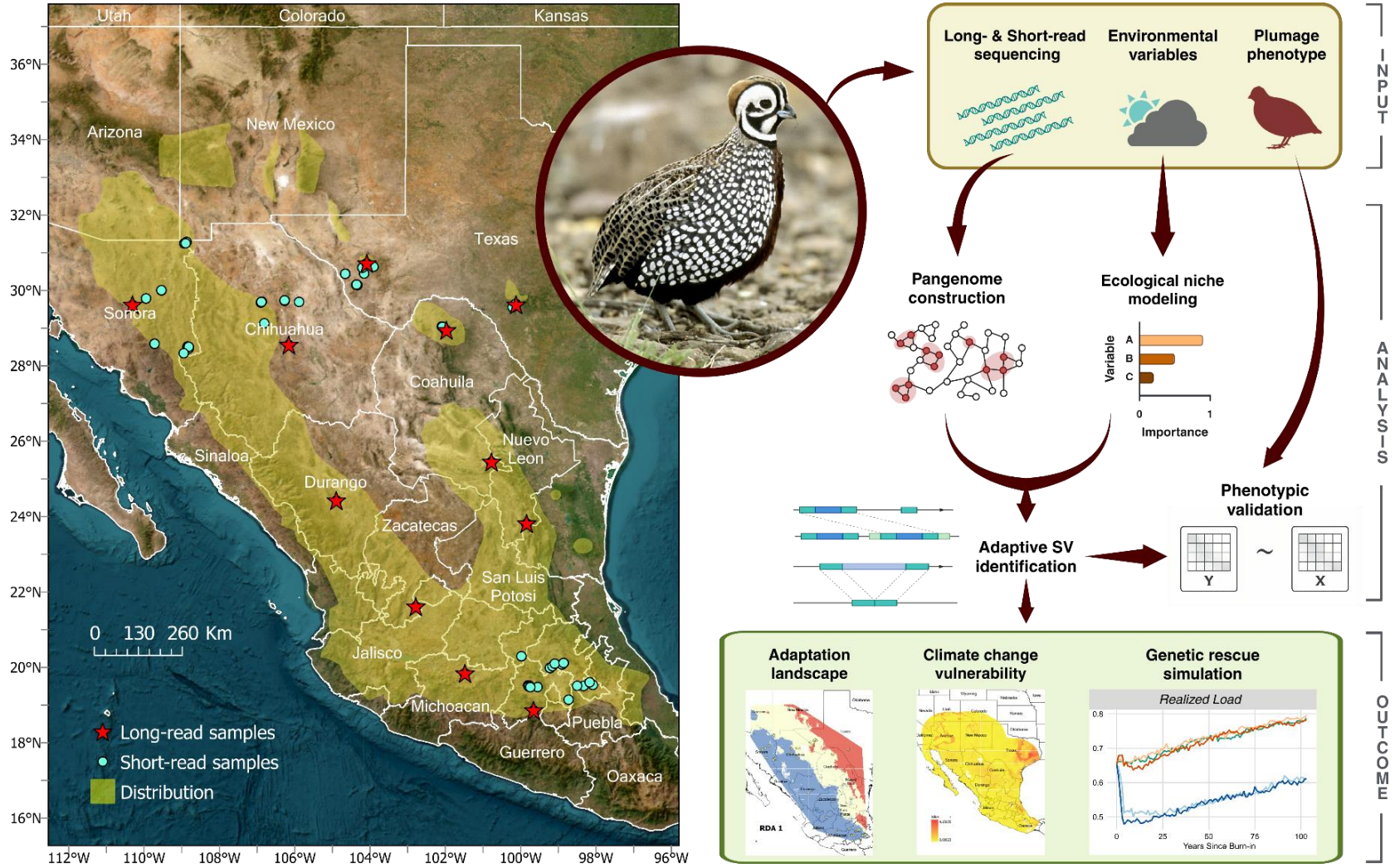

Fig. 1. Geographic distribution of sampled Montezuma quail across the species' range (yellow shaded area denotes the estimated species distribution according to BirdLife International) and overview of the analytical workflow. Red stars and cyan points indicate sampling locations for long-read and short-read sequencing, respectively. Outcome examples are presented in greater detail in subsequent figures.

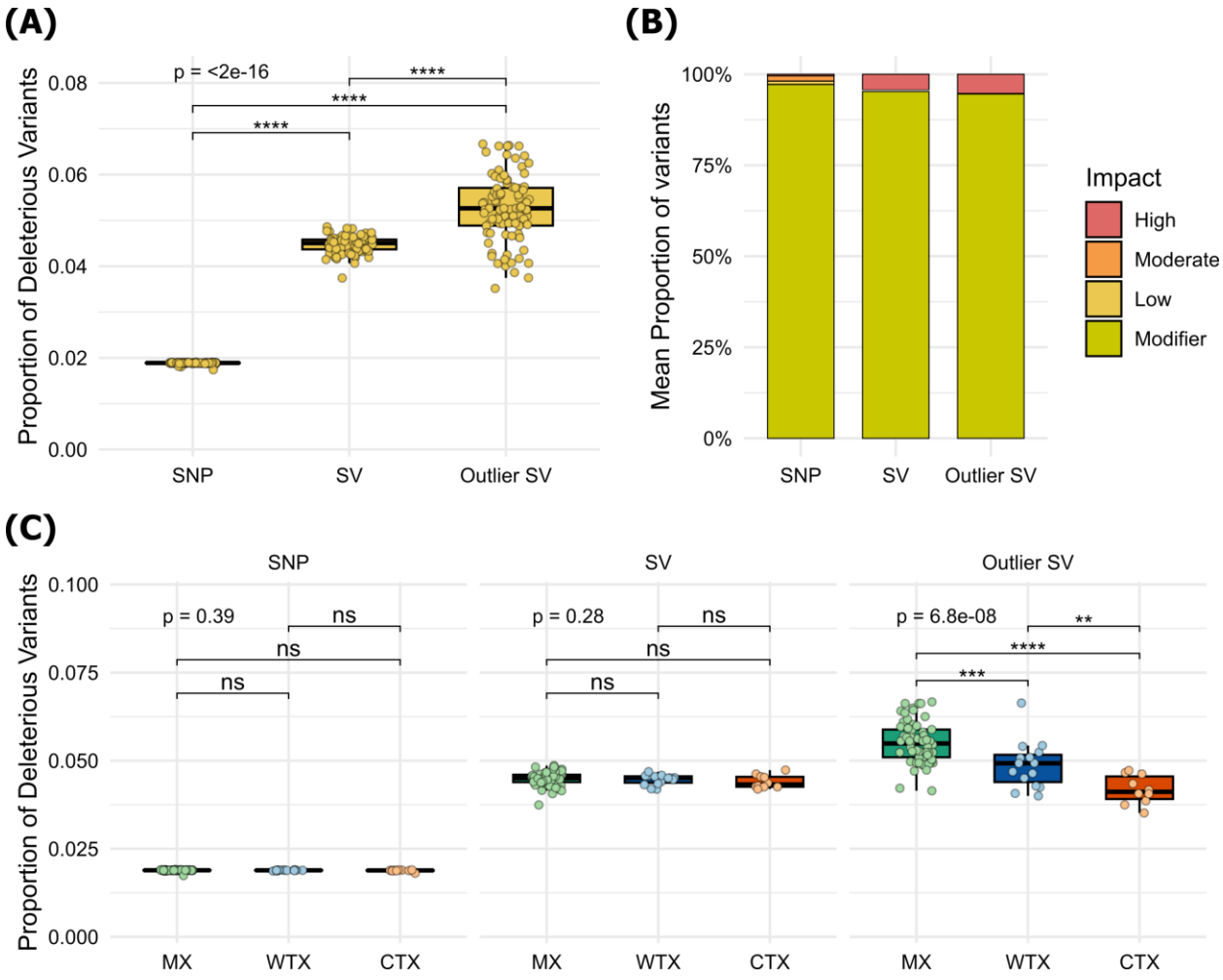

Fig. 2. Proportion of deleterious variants across variant types, impact classes, and populations. (A) Proportion of deleterious variants by variant type. (B) Proportion of deleterious variants by predicted impact class. (C) Proportion of deleterious variants by population and variant type. Global differences were assessed using a Kruskal-Wallis test (*p*-values shown), followed by pairwise Wilcoxon rank-sum tests, with significance indicated by asterisks. Outlier SVs exhibit significantly higher proportions of deleterious variants than genome-wide SNPs and SVs, with a greater representation of high-impact classes. Consistent with more efficient purging in smaller, isolated populations, deleterious proportions at outlier SV loci are lowest in Central Texas (CTX), intermediate in West Texas (WTX), and highest in Mexico-wide (MX) populations, with all pairwise differences significant).

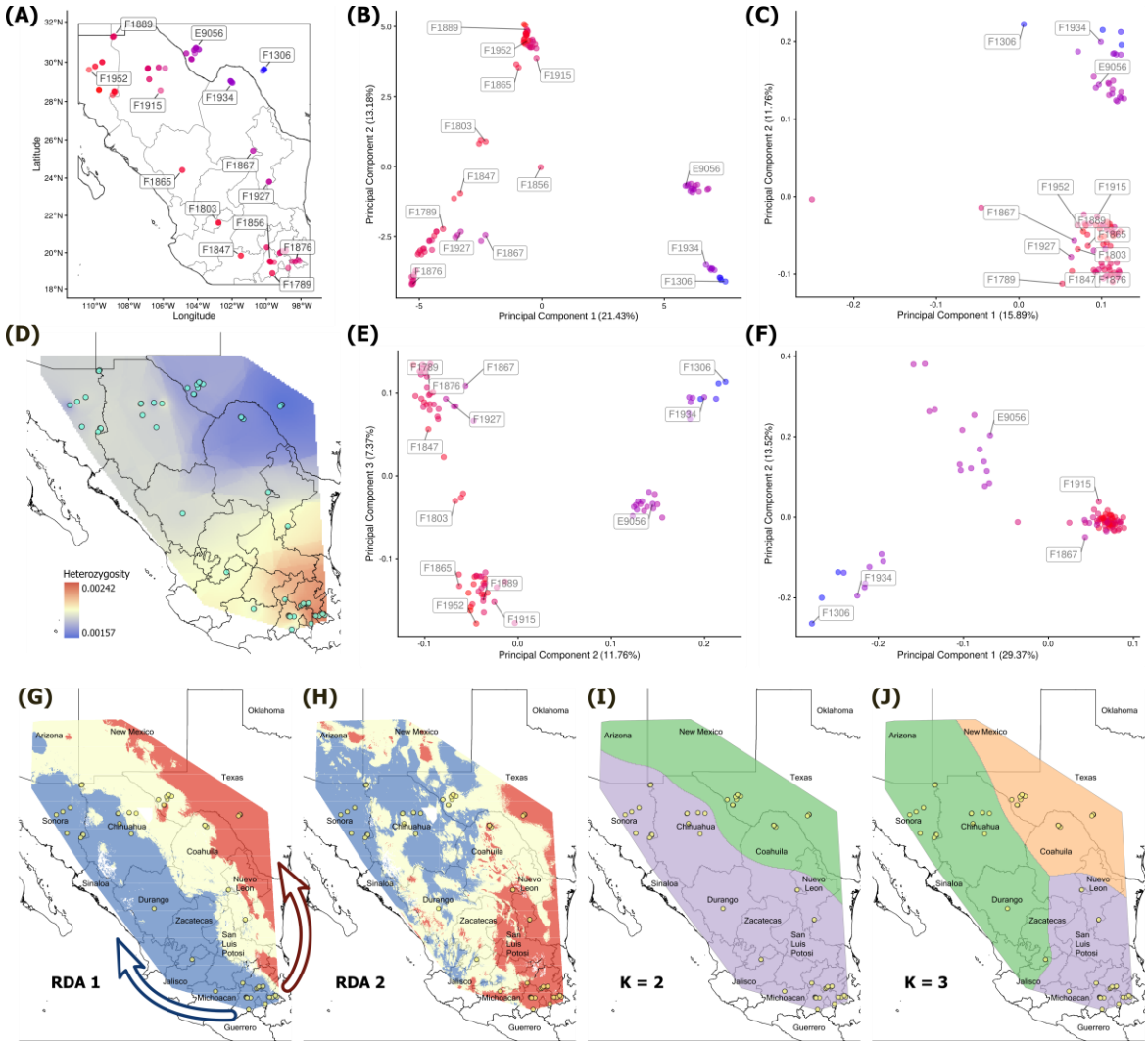

Fig. 3. Spatial patterns of adaptive and neutral genetic structure in Montezuma quail across the species' range. (A) Geographic distribution of labeled samples, with colors encoding a blend of latitude and longitude to facilitate matching across panels. (B) Population structure inferred from genome-wide SNPs (PC1 vs. PC2). (C) Genome-wide SVs (PC1 vs. PC2). (D) Spatial interpolation of individual SNP heterozygosity (range: 0.00157–0.00242) across the study region. (E) Genome-wide SVs (PC2 vs. PC3). (F) Outlier SVs (PC1 vs. PC2). (G) Adaptive landscape pattern based on RDA Axis 1 (RDA1), with arrows indicating the inferred ring-like northward expansion. (H) RDA Axis 2 (RDA2). (I) Neutral population structure based on K=2. (J) K=3. Genome-wide SNPs and SVs revealed broadly concordant population structure, particularly in panels (B) and (E), separating Texas and Coahuila from the remaining Mexican individuals; outlier SVs exhibited a distinct pattern, separating Central Texas + Coahuila, Western Texas, and Mexico-wide individuals. The south-to-north decline in heterozygosity (D) is consistent with predictions of the central–marginal hypothesis. In (G-J), the depicted geographic range represents a minimum convex hull of occurrence records buffered by 5 km to approximate contemporary dispersal limits; distinct colors in (G-H) depict three distinct adaptation profiles across the landscape, and in (I-J) reflect genome-wide population affinities.

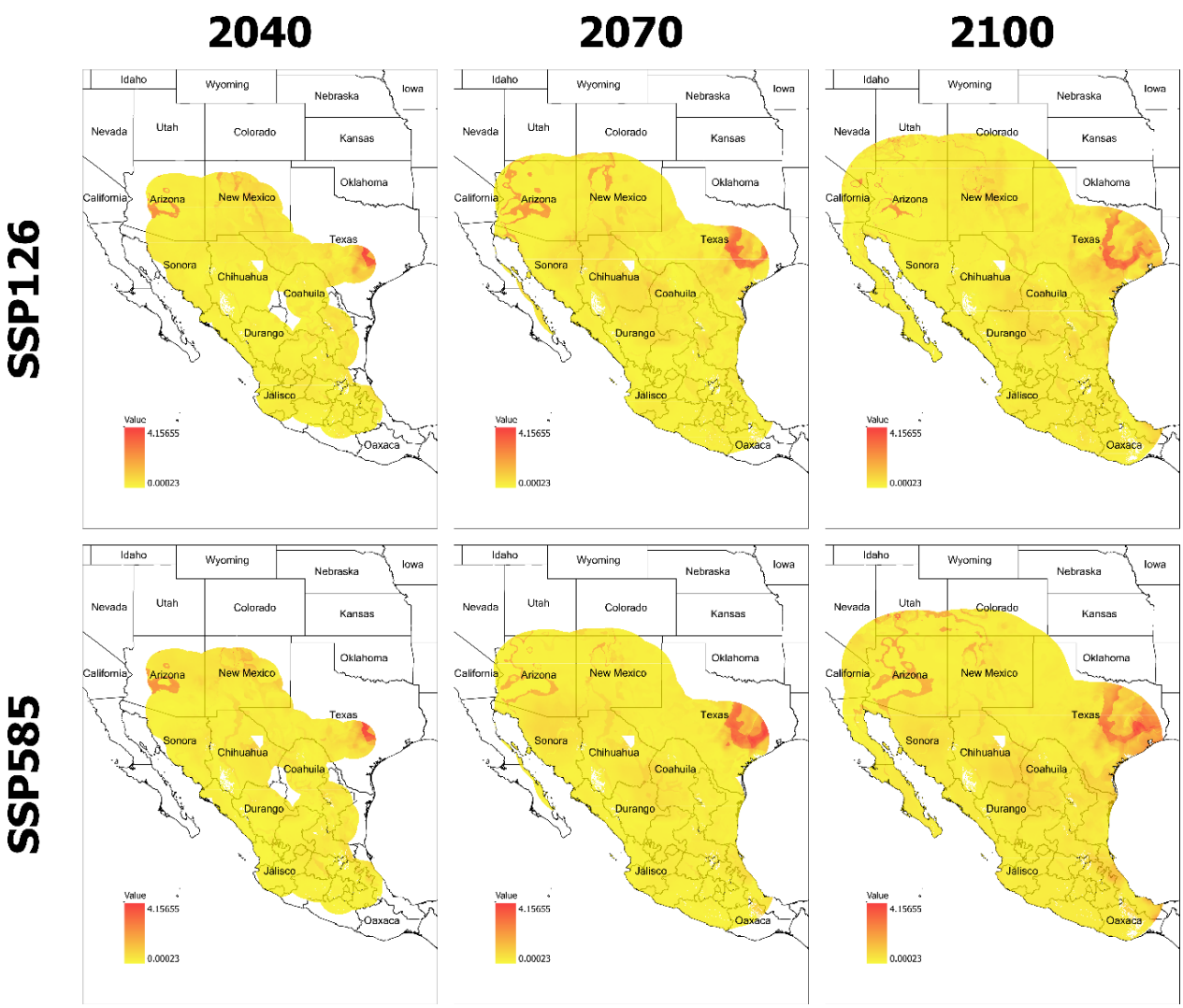

Fig. 4. Genomic offset projected under two climate scenarios and three future time periods. Rows correspond to SSP126 (top) and SSP585 (bottom); columns represent 2040 (left), 2070 (middle), and 2100 (right). The geographic range depicted at each time period reflects the maximum possible extent given projected dispersal capacity. U.S. peripheral populations consistently show greater genomic mismatch over time, with mismatches intensifying under SSP585.

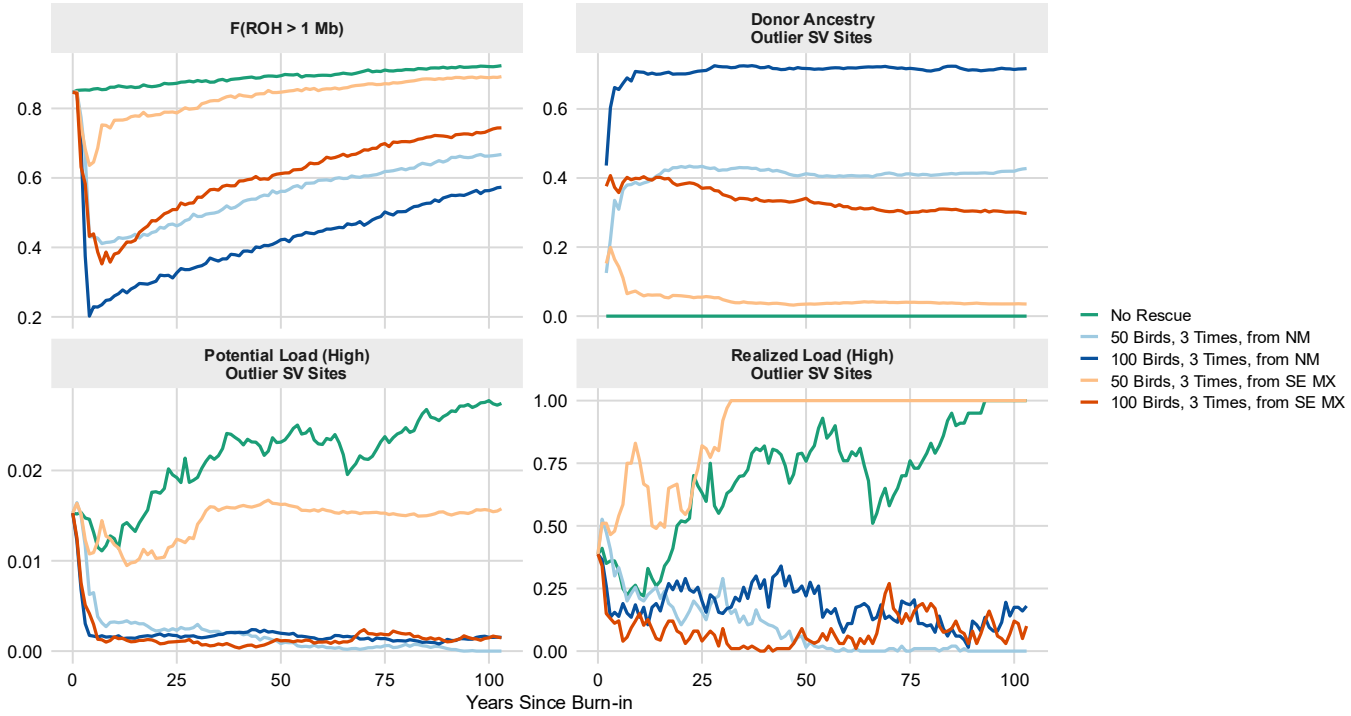

Fig. 5. Genomic forecasts under genetic rescue scenarios targeting Central Texas. Five rescue scenarios are shown (legend): a no-rescue baseline and translocations of 50 or 100 individuals from either New Mexico (NM) or Southeast Mexico (SE MX), conducted once annually for three consecutive years. Metrics include the fraction of runs of homozygosity longer than 1 Mb ($F_{ROH > 1\ Mb}$) across all variant sites, donor ancestry proportion at outlier SV sites, and potential and realized genomic load at outlier SV sites. Load metrics are shown for high-impact variants only, as outlier SV sites were populated exclusively with variants assigned high selection coefficients. In all rescue scenarios, F_{ROH} declined immediately following translocation and subsequently increased over time. Translocations of 100 individuals most effectively reduced F_{ROH} and genomic load, particularly from New Mexico, whereas the 50-individual Southeast Mexico scenario rapidly returned to baseline levels. Donor ancestry proportion increased sharply after immigration and subsequently stabilized, reaching the highest level under the 100-individual New Mexico scenario (~0.6–0.7). Results across additional ROH length classes and donor ancestry proportions at genome-wide and all variant sites are shown in fig. S18. Collectively, rescue simulations forecast reduced inbreeding and genomic erosion at outlier SV sites, while highlighting a trade-off between genetic swamping under high translocation intensity and selection-dependent purging at adaptive SV loci.

Supplementary Materials for

Range-wide pangenomics reveals vulnerability and adaptation in a sedentary bird

Jong Yoon Jeon[1], Natalie M. Allen[1], Audrey J. Heckel[2], Alberto Macías-Duarte[3], René Jesús Toscano-Meléndrez[4], Jesús Raúl Martínez-Sandoval[5], Angel Montoya[6], John McLaughlin[7], J. Andrew DeWoody[1,8]

[1] Department of Forestry and Natural Resources, Purdue University, West Lafayette, Indiana, USA

[2] Department of Sustainability Engineering and Environmental Engineering, Purdue University, West Lafayette, Indiana, USA

[3] Cuerpo Académico de Recursos Naturales, Universidad Estatal de Sonora, Hermosillo, Sonora, Mexico

[4] Maestría en Ciencias Ambientales, Universidad Estatal de Sonora, Hermosillo, Sonora, Mexico

[5] Programa Académico de Ingeniero en Mecatrónica, Universidad Estatal de Sonora, Hermosillo, Sonora, Mexico

[6] Las Cruces, New Mexico, USA

[7] Upland Game Bird Program, Texas Parks & Wildlife Department, Lubbock, Texas, USA

[8] Department of Biological Sciences, Purdue University, West Lafayette, Indiana, USA

Corresponding author: dewoody@purdue.edu

The PDF file includes:

Materials and Methods
Supplementary Text
Figures S1 to S24
References

Other Supplementary Materials for this manuscript include the following:

Tables S1 to S13

Materials and Methods

Library preparation for long-read sequencing

Genomic DNA (gDNA) from tissue samples preserved in lysis buffer was extracted using a MagAttract HMW DNA kit (Qiagen), then the gDNA was sheared to a mean length of 18 kb using a Megaruptor 3 (Diagenode). Sheared fragments were converted to libraries using the SMRTBell Prep kit (3.0) and SMRTbell adaptor index plate 96C. The PacBio HiFi longread sequencing was performed with SPARQ chemistry in a REVIO instrument with a 30hs movie and CCS analysis was done on the instrument. Demultiplexing was conducted using SMRTlink 25.1. The resultant sequencing depth was approximately 10x for each sample. Read quality was assessed using FastQC v0.12.1 (*41*) and adapters were trimmed with fastp v0.23.2 (*42*).

Library preparation for short-read sequencing

Genomic DNA was extracted using Qiagen DNeasy Blood & Tissue kit and the Shotgun genomic libraries were prepared with the Qiagen Magattract HMW DNA kit. The libraries were pooled and quantitated by qPCR before being sequenced on a 25B lane for 151 cycles from both ends of the fragments with V1.0 sequencing kits. Sequencing was performed on the Illumina NovaSeq X Plus platform. Resultant FASTQ files were generated and demultiplexed with the bcl-convert v4.1.7 Conversion Software (Illumina). The resultant sequencing depth was approximately 10x on average. Read quality control was performed with FastQC and Trim Galore v0.6.10 (https://github.com/FelixKrueger/TrimGalore). Contaminated reads were removed using Kraken2. Unpaired reads after this contamination removal were culled using BBmap v38.96 (*43*) ('repair.sh').

Pangenome construction from long-read assemblies

We used long-read sequencing to assemble a high-quality reference genome and construct a multi-haplotype graph pangenome. A single female Montezuma quail (sample ID F1298; collected in Central Texas by JM on 20 February 2023) was selected for the reference genome assembly (see *Reference genome assembly and annotation* section below). Ten additional birds

sampled across the species' geographic range were sequenced for pangenome construction (Fig. 1).

For each individual, *de novo* assemblies were generated with hifiasm v0.19.8 (*44*) (using the –primary option) and redundant haplotigs were removed using purge_dups v1.2.6 (*45*). Assemblies were screened for contamination with Kraken2 v2.1.5 (*46*). Low-complexity regions were identified using WindowMasker v2.2.22 (*47*) and genomes were soft-masked with RepeatMasker v4.1.5 (*48*) with the "aves" repeat library. The Montezuma quail linear reference genome and the primary and alternate assemblies for all ten individuals were integrated into a graph pangenome using the "–pangenome" option in Cactus v2.7.0 (*49*).

Variant detection from short-read data

We used short-read data from 91 birds to capture species-wide genomic variation (Fig. 1). Cleaned short reads were first mapped using 'vg giraffe' and quality-filtered using 'vg filter' of VG Toolkit v1.65.0 (*4*, *50*). Individual graph alignment/map format (GAM) files were sorted and chunked for each chromosome. Next, the chunked GAM files of individuals were combined into a single GAM file per chromosome before integration with the corresponding chromosome of the graph pangenome.

Structural variants were called with 'vg call' (*51*) for each chromosome and each individual, then per-chromosome VCF files were concatenated for each individual. Variant size was capped at 100,000 bp to minimize technical artefacts associated with calling very large structural variants. Low quality variants were culled using vcftools v0.1.16 (*52*) (with options '--minQ 20 --mac 2 --min-meanDP 3 --max-missing 0.8' and excluding loci with missing data proportion > 0.2) and vcfbub v0.1.0 (https://github.com/pangenome/vcfbub; with options '--max-allele-length 100000 --max-level 0') before decomposing complex alleles with vcfwave v1.09 (*53*).

Reference genome assembly and annotation

We extracted genomic DNA (gDNA) from heart tissue (sample F1298 at Purdue University) preserved in lysis buffer using a MagAttract HMW DNA kit (Qiagen). The gDNA was then sheared to an average fragment length of 13 kb using a Megaruptor 3 (Diagenode). Sheared fragments were converted to libraries using the SMRTBell Express Template prep kit (3.0). The library was sequenced on one SMRTcell 8M on a PacBio Sequel IIe using the CCS sequencing mode and a 30hs movie time. The same tissue was used to generate Tell-seq libraries, which were prepared with the TellSeq Whole Genome Sequencing Library Prep kit (Universal Sequencing). The library was quantitated by quantitative polymerase chain reaction (qPCR) and sequenced on one 10B lane for 151 cycles from both ends of the fragments on a NovaSeq X Plus. All benchworks after DNA extraction were conducted at the Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign.

PacBio CCS analysis was conducted with SMRTLink v11.1 using the following parameters: ccs –min-passes 3 –min-rq 0.99. Tell-seq FASTQ files were generated and demultiplexed with bcl2fastq v2.20 Conversion Software (Illumina). The raw sequences were then trimmed with TellRead v1.0.3 (Universal Sequencing) with default parameters. Trimmed sequences were reformatted (interleaved) for ARCS with TellReads (Universal Sequencing). We filtered out short HiFi CCS reads (<1kb) before downstream analyses. We then counted *k*-mers to inform genome assembly with Jellyfish v.2.31 (*54*) using following parameters: jellyfish count -C -m 21 -s 1000000000. A histogram of *k*-mer counts was then used to estimate genome assembly size and heterozygosity with GenomeScope (*55*). PacBio HiFi CCS reads were assembled using hifiasm v0.18.5-r499 (*44*) with default parameters. Following de novo assembly, haplotigs were identified and removed using purge_dups v1.2.5 (*45*) following the developer's guideline. Contaminated contigs were checked using Blobtools v1.1.1 (*56*), but none were detected. Tell-seq reads were then used to scaffold contigs using the *k*-mer approach (k = 60) implemented in ARKS (*57*). We checked assembly quality after each step with BUSCO v5.4.1 (*58*) using 8,338 orthologs in the Avian database and QUAST v5.2.0 (*59*). As a final curation step, we tested whether each contig could be mapped by independent samples. We downloaded and processed short reads (SRR13748625–SRR13748634) from NCBI using SRA-Tools v2.11.0-pl5262 (https://github.com/ncbi/sra-tools/), Trim Galore v0.6.10 (https://github.com/FelixKrueger/TrimGalore), then we mapped them to our new genome

assembly with bwa v0.1.17 using bwa-mem algorithm (*60*). All contigs were validated as mappable by calculating alignment depth using mosdepth v0.3.6 (*61*). Following assembly curation, repeat masking and gene annotation were completed by the NCBI Eukaryotic Annotation Pipeline. Assembly completeness was again evaluated with BUSCO using the Avian database.

Finally, we performed reference-guided scaffolding using RagTag v2.1.0 (*62*) against a chromosome-scale Japanese quail reference genome (GCF_001577835.2) to further improve contiguity by assigning chromosomes and for downstream analyses. Because the Japanese quail reference genome does not include the W chromosome, we extracted the W chromosome from the domestic chicken reference genome (GCF_016699485.2) and appended it to the Japanese quail genome before scaffolding (*15*). Out of 287 contigs of 1,011,858,581 bp in total, 148 contigs of 1,001,962,697 bp were placed in scaffolds. A total of 80 gaps of 830,690 bp were added.

Adaptive landscape genomics

We identified adaptive genomic variants, with a focus on SVs, in a two-step process to resolve spatial adaptation across the species' range. First, we used R package 'PCAdapt' v4.4.1 (*16*) to detect outliers from PCA-based population structure. Because linkage-disequilibrium (LD) pruning had negligible effects on the results, the unpruned SV dataset was used to reduce false negatives. Second, using the adaptive SVs identified by PCAdapt, we further screened adaptive variants with RDA implemented in the R package 'vegan' v2.7-2 (*63*) as a genotype-environment association approach. Bioclimatic (CHELSA-climatologies) (*64*) and topographic variables (i.e., elevation, aspect, and slope; USGS EarthExplorer) (*65*) served as environmental variables (see *Ecological niche modeling* section below). Individuals with > 20 % missing data and loci with > 10 % missing data were removed. To comply with the requirements of the R packages, missing data were imputed using 'estim_ncpPCA' and 'imputePCA' functions from the R package 'missMDA' v1.20 (*66*).

We tested various models and, based on its higher adjusted R^2, proceeded with a partial RDA model controlling for the population genetic structure and geographic locations (table S10).

Outlier SVs were identified using 'rdadapt' function (*17*) (*q*-value <0.05). The correlation between environmental variables and SVs was calculated using the 'cor' function in R v4.2.2 (*67*). A landscape of adaptation profiles was generated using the 'adaptive_index' function (*17*) ('method = loading') based on RDA1 and RDA2, the only two significant axes.

Genomic offset was estimated under the SSP126 and SSP585 climate change scenarios, which represent the most moderate- and extreme-case projections, respectively. We used three different approaches: RDA; generalized dissimilarity modeling (GDM v1.6.0-7) (*23*) based on genetic distances among individuals measured in PC space across loci; and Gradient Forest (GF v0.1-37) (*24*). We used GDM and GF to estimate reverse genomic offset (*25*), the lowest possible genomic offset at each location assuming unlimited individual movement. Whereas genomic offset quantifies the mismatch between contemporary genomic composition and the predicted future optimal composition, reverse genomic offset assesses whether genetic rescue could reduce projected mismatches at each location (*25*).

Structural variant profiling

We compared the outlier SVs with genome-wide SVs to evaluate whether adaptive candidates show any enrichment or deviation from background patterns. Variant type, length and counts were summarized using Truvari v5.3.0 (*68*). Gene ontology (GO) and KEGG pathway enrichment analyses were conducted for outlier SVs. Gene annotation, GO and KEGG datasets from Japanese quail autosomes and the chicken W chromosome were merged prior to analysis. VCF were normalized using bcftools v1.17 (*69*) and overlaps between SVs and gene regions were identified using bedtools v2.31.0 (*70*). To account for length bias in enrichment analyses, we used the R package 'goseq' v1.62.0 (*71*). Multiple testing correction was applied using the Benjamini–Hochberg procedure.

We estimated site frequency spectra for genome-wide and outlier SVs was using vcftools ('--freq2'). Genomic load was quantified for genome-wide SVs, outlier SVs, and SNPs. VEP v108.2 (*72*) was used to annotate and classify variants as modifier or as high-, moderate-, or low-impact. We then calculated (for each population) the proportion of deleterious variants, with high- and moderate-impact variants considered deleterious.

Phenotypic data and statistical analyses

We analyzed plumage phenotypes for 28 males from the genomic dataset (n = 91) using environment-associated (outlier) SVs. Each individual was photographed at 20 equiangular positions spanning 360° around the body (from ventral to dorsal views). Mean color per image was calculated across all pixels in CIELAB space (L*, lightness; a*, green–red opponent axis; b*, blue–yellow opponent axis) and averaged across images to obtain per-individual mean color traits. Detailed phenotypic data collection procedures are described elsewhere (*28*).

To corroborate the biological relevance of the outlier SVs identified in our statistical analyses, we assessed genotype–phenotype associations using mean L*, a*, and b* as response variables (with body mass analyzed separately). To mitigate issues arising from the limited sample size (n = 28) relative to the number of genomic predictors, we performed PCA on the outlier SV genotype matrix and the first *k* 'genotype' PCs were retained as predictors. Climatic variables were similarly summarized by PCA of the bioclimatic dataset used in the genotype–environment RDA, retaining the first two 'bioclim' PCs. Neutral population structure was controlled using two SNP-based 'structure' PCs (see *Population structure and genomic diversity* section below).

We validated genotype–phenotype associations between outlier SVs and plumage traits using three complementary statistical approaches robust to small sample sizes. First, we performed partial RDA with the multivariate phenotype matrix (scaled L*, a*, b*) as the response and the first *k* genotype PCs as predictors (*k* = 1–12), while conditioning on the two structure PCs and two bioclim PCs. Model significance was assessed using permutation tests (n = 999), recording adjusted R^2 and model rank (table S11). For the best-supported specification (largest *k* without overfitting), we inspected which genotype PC axis contributed significantly to the constrained association.

Second, we fit separate linear models for each response trait (L*, a*, b*, and body mass) using genotype PCs as focal predictors while controlling for structure PCs and bioclim PCs. Based on the multivariate RDA results and to avoid overfitting, we used the first three genotype PCs as focal predictors in the main models. When body mass was the response variable, we

included binary ‘age’ (juvenile or adult) as an additional control variable. For each model, we reported overall model fit (p-value and adjusted R^2) and coefficient-level inference for each predictor. We computed partial effect sizes (i.e., partial ω^2) using ‘anova_stats’ function of the R package ‘sjstats’ (*73*) to summarize the relative contribution of individual predictors.

Lastly, we used partial least squares (PLS) regression using R package ‘mixOmics’ v3.22 (*74*) to quantify covariation between the phenotype response matrix (L*, a*, b*) and the genotype predictor matrix (first three genotype PCs). We selected the number of PLS components using repeated M-fold cross-validation (5 folds, 100 repeats) and the predictive Q^2 criterion. In practice, a single component (ncomp = 1) was supported and used for final models. We summarized PLS results by (1) the correlation between phenotype and genotype scores on the first component (PLS1) and (2) cross-validated $Q2$ values for each response trait. To assess whether the observed covariation exceeded that expected by random chance, we additionally conducted a permutation test (n = 999) by permuting sample labels in the predictor matrix and recalculating the PLS1 score correlation. Finally, to evaluate whether genotype–phenotype covariation persisted after accounting for shared climatic and/or population-structure gradients, we repeated the PLS analysis using residualized phenotypes (residuals from linear models regressing each trait on bioclim PCs and/or structure PCs).

Population structure and genomic diversity

To analyze population genetic structure and genomic diversity using SNPs, we first created SNP binary alignment/map (BAM) files from graph alignment/map (GAM) files (see *Variant detection from short-read data* section above). BAM files were generated by projecting per-chromosome GAM files of each individual onto the reference individual’s path within the graph pangenome pruning low-complex regions. The per-chromosome BAM files were then merged for each individual and sorted using samtools v1.22.1 (*69*). Duplicates were marked using picard tools v3.3.0 (http://broadinstitute.github.io/picard).

We employed ANGSD v0.940 (*75*) to estimate genotype likelihood and generate beagle-format files, accounting for low-depth alignments. We also filtered low-quality genomic regions from the BAM files. To do so, we first linearized the reference path in the graph pangenome to

represent a reference linear genome for consistent chromosome assignment. Repeats were masked using RepeatMasker and mappability was computed using genmap v1.3.0 (*76*) with '-K 100 -E 2' options. Chromosomes larger than 100 kb were retained while repeat-masked regions and sites with mappability < 1 were excluded from the downstream ANGSD step with '-GL 1 -doglf 2 -minmaf 0.05 -snp_pval 1e-6 -dopost 1 -minQ 30 -minInd 91/2 -docounts 1 -domajorminor 5 -domaf 1 -only_proper_pairs 1 -remove_bads 1 -uniqueOnly 1 -baq 2 -C 50' options. Based on these genotype likelihoods, we inferred the folded site frequency spectrum for each individual. Individual heterozygosity was then calculated as the proportion of heterozygous sites, defined as the number of heterozygous sites divided by the total number of analyzed genomic sites. We interpolated individual heterozygosity across the species' range using inverse distance weighting implemented in the R package 'gstat' v2.1-3 (*77*).

Population structure was investigated with a principal component analysis (PCA) and admixture analysis using the BEAGLE-format files. The BEAGLE format files were filtered to reduce the number of variants for computational efficiency and performed linkage disequilibrium pruning using ngsLD v1.1.1 (*78*) and prune_graph (https://github.com/fgvieira/prune_graph). We used PCAngsd v1.10 (*79*) for the PCA and NGSadmix v32 (*80*) to generate admixture plots (based on a model of Hardy-Weinberg equilibrium) with five replicates for each K value from 1 to 10. To calculate ΔK and evaluate the most likely number of clusters, we employed the Evanno method (*81*). For each value of K, we used the R package 'tess3r' v1.1.0 (*82*) to generate ancestry barplots and visualize spatial patterns of ancestry using kriging-based interpolation of ancestry coefficients across the sampled region.

We analyzed SV-based population genetic structure using PCA (PCAdapt) for genome-wide and outlier SVs to compare patterns with SNP-based structure. SNPs are expected to reflect genome-wide evolutionary processes (e.g., drift and gene flow) whereas outlier SVs should more effectively capture signatures of selection (*1*, *3*, *15*).

Ecological niche modeling

The Montezuma quail is highly dependent on climate conditions throughout its life cycle, including nesting and brooding (*10*, *14*, *83–86*). Understanding its climate niche and predicting

future habitat suitability are crucial for assessing the species' persistence under climate change. To achieve this, as well as to use as input files for downstream simulations, we employed ecological niche modeling (ENM), also known as species distribution modeling or habitat suitability modeling, depending on the end goal. ENM identifies key environmental conditions, infers species distribution, and estimates habitat suitability by correlating species occurrence data with environmental variables (*87*). When combined with genomic data (*88*), it provides a comprehensive profile of species adaptation at the organismal level.

We first obtained species occurrence records from Global Biodiversity Information Facility (GBIF) and iNaturalist using the R packages 'rgbif' v3.8.1 (*89*) and 'rinat' v0.1.9 (*90*). From GBIF, we extracted 'PRESENT' records and filtered them by record type, removing 'MATERIAL_CITATION', 'MATERIAL_SAMPLE', and 'PRESERVED_SPECIMEN'. From iNaturalist, we collected research-grade records with unobscured coordinates. We included occurrence points recorded since 1970, with an accuracy threshold of ≤1700 m, based on the species' reported home range (*91*). Further filtering was conducted using the R packages 'CoordinateCleaner' v3.0.1 (*92*) and 'scrubr' v0.4.0 (*93*) with specific filters ('capitals', 'centroids', 'gbif', 'institutions', 'outliers', 'seas', and 'zeros') and functions ('coord_incomplete', 'coord_imprecise', 'coord_impossible', and 'coord_unlikely'). The cleaned dataset was supplemented with in-lab geo-coordinates from biological samples. A final visual inspection in ArcGIS Pro v3.2 (*94*) removed three points that fell outside the species' putative range (*11*). To delineate the contemporary species distribution, we generated a minimum convex hull and applied a 5 km buffer, reflecting the species' mean dispersal distance (*95*). To estimate the species' potential range after dispersal, we created geodesic buffered polygons at 150 km (~30-year dispersal), 300 km (~60-year dispersal), and 450 km (~90-year dispersal), with a final 500 km buffer used as a template for data processing to avoid edge effect. Since these buffered polygons included oceanic areas, we clipped them using a 'landmass' polygon derived from the Natural Earth database (https://www.naturalearthdata.com/), combining 'Land' and 'Minor Islands' polygons.

We collected topographic and bioclimatic variables for ENM and processed them in ArcGIS Pro. For topographic data, we downloaded 30-m resolution digital elevation model (DEM) rasters from USGS (SRTM1 DEM) (*65*) covering the USA and Mexico. Nineteen

bioclimatic variable rasters (1-km resolution) were obtained from the CHELSA database (*64*) for four time periods: present (1980–2010), early-century (~2040; 2011–2040), mid-century (~2070; 2041–2070), and late-century (~2100; 2071–2100). We used variables from the most optimistic shared socioeconomic pathway (i.e., SSP126), which targets less than 2 °C increase by the year 2100, and the most pessimistic one (i.e., SSP585), which targets more than 4 °C increase, of the earth system model 'GFDL-ESM4' from the database. To maximize environmental variation for ENM, all environmental rasters were clipped using the 500 km-buffered template polygon. We computed 'Aspect' and 'Slope' from the clipped DEM using the 'Surface' tool, then bilinearly resampled them to match the resolution of bioclimatic rasters. To reduce multicollinearity, we conducted a Pearson correlation test using the 'Band Collection Statistics' tool and selected the least correlated variables relevant to the species' ecological characteristics (*9*, *10*, *14*, *83–86*, *95*): Bio2, Bio5, Bio7, Bio8, Bio9, Bio14, Bio15, Bio18, Aspect, and Slope. The final environmental rasters were clipped using 150 km-, 300 km-, and 450 km-buffered polygons to represent potential future ranges.

Ecological niche modeling can be affected by sampling bias, a bias in input occurrence data due to different accessibility of regions. To minimize the influence of spatial sampling bias in species occurrence data, we implemented two corrective measures using SDMtoolbox v2.6 (*96*). First, spatial thinning was applied by retaining a single occurrence record per 0.5 kilometers (*9*), reducing overrepresentation from frequently surveyed locations. Second, a bias layer was generated to represent the Gaussian kernel density of 0.25-degree distance with rarefied occurrence points. This bias layer was subsequently used to generate 10,000 background points, with point selection weighted by density values using the R package 'dismo' v1.3-14 (*97*).

The entire process of ENM was performed in RStudio 2024.09.1+394 (*98*) using R v4.4.1 (*67*). To choose the best ENM approach, multiple modeling algorithms were tested using the R package 'biomod2' v4.2-5-2 (*99*), with cross-validation parameters set as follows: OPT.strategy = 'tuned', CV.strategy = 'block', CV.nb.rep = 3, CV.perc = 0.7, var.import = 3. The models included 'Random Forest (RF)', 'Generalized Linear Model', 'Artificial Neural Network', 'Flexible Discriminant Analysis', 'Generalized Additive Model', 'Gradient Boosting Machine', 'Multivariate Adaptive Regression Splines', 'Surface Range Envelope', 'Extreme Gradient

Boosting', 'Classification Tree Analysis', and 'Maximum Entropy (MaxEnt)'. A separate MaxEnt model was further evaluated using manual tuning with the R package 'ENMeval' v2.0.4 (*100*). The models were compared based on two evaluation metrics, 'True Skill Statistic (TSS)' and 'Receiver Operating Characteristic curve (ROC)'.

Among these models, RF exhibited the highest predictive performance, achieving a mean True Skill Statistic (TSS) of 0.99 and a Receiver Operating Characteristic (ROC) score of 1.00. Based on this performance, RF was selected as the algorithm for downstream ENM. We then derived a robust ensemble prediction, averaging the results from three cross-validation runs weighted by their TSS scores (the final TSS = 0.928; ROC = 0.994). The final ensemble model was projected across four temporal scenarios: present (1980–2010), early-century (~2040; 2011–2040), mid-century (~2070; 2041–2070), and late-century (~2100; 2071–2100). Finally, the present ENM raster (initially used 500-km buffered data) was clipped by the minimum convex hull from above to represent the current range.

Subsequently, we processed the output rasters in ArcGIS Pro to meet the requirements of downstream simulations. To address regions of missing data, we applied inverse distance weighting interpolation with default parameters to fill 'NA' areas. The habitat suitability rasters were then aggregated by a factor of ten, while preserving the maximum value among input cells, to reduce their resolution to 258 × 291 cells (approximately 10 km × 10 km per cell) for improved computational efficiency. To ensure consistency across all rasters, we used Raster Calculator to expand their extent to match the largest bounding box among them (xmin = -117.0085, xmax = -93.30847, ymin = 15.63319, ymax = 39.36653). Non-land areas within the expanded region were assigned a value of '0'. Additionally, the original scale of habitat suitability values (0–1000) was rescaled to a 0–1 range. The rasters were projected from the geographic coordinate system (WGS 1984) to the projected coordinate system (North America Albers Equal Area Conic) using the 'Nearest Neighbor' resampling methods since the downstream simulations relied on Euclidean geographic distances (e.g., kilometers). Lastly, the land availability rasters were generated by a similar approach from the original habitat suitability rasters, but converting all non-NA values to '1'.

<u>Spatial demographic simulations of climate change (Geonomics)</u>

The Geonomics simulation v1.4.8 (*26*) was conducted to complement the ecological niche models by incorporating species-specific biological constraints. While ecological niche models identify geographic regions where a species can theoretically persist under given ecological conditions, actual occupancy may be constrained by factors such as limited dispersal ability or recruitment failure (*101*). Thus, the main goal here was to see if the species can escape unfavorable habitats and survive. Geonomics is a Python package for spatially-explicit, forward-time, individual-based simulations that integrate biological traits. The simulation framework consists of four sequential stages per generation: movement → mating → mortality → environmental change, allowing for species traits to be parameterized at each stage. Additionally, individuals possess a genome that influences their phenotype, which in turn determines fitness, calculated as the difference between the individual's phenotypic value and the environmental value at its location. Using this approach, we assessed whether the species could adapt to shifting environmental conditions over time. Below we briefly describe the simulation set-up. The detailed scripts can be found at https://github.com/jyj5558/monq_adaptive_genomics.

The Geonomics simulation was run for two iterations, each spanning 150 generations following burn-in period. The initial 60 generations of each iteration were designated for non-neutral burn-in, during which species' mean fitness stabilized over the final 10 generations based on landscape suitability values and individual phenotypic traits. The duration of the non-neutral burn-in period was determined through preliminary runs using the Augmented Dickey-Fuller stationarity test in the R package 'tseries' v0.10-58 (*102*). The habitat suitability raster was used as the input landscape, influencing movement, carrying capacity, and trait-based selection under the assumption that individuals are drawn to more suitable habitats, where larger populations can persist and climate-adapted genotypes are preferentially selected. To simulate future climate change, the landscape raster was updated at generation 60, 90, and 120 to reflect projected environmental shifts.

We aimed to incorporate biological realism into parameter selection to ensure ecologically meaningful simulations. The initial population size was set at 500,000 individuals, representing females only for computational efficiency, under the assumption that each female moves as a pair with a male. This starting population size was derived from an estimated present-day population of 1,000,000 individuals approximately (*12*). Carrying capacity per raster cell

was set at 500 female individuals (*18*). Mating was allowed from age 1 to 3 (out of age 0-3), with an intrinsic rate of population growth set at 0.2, based on an annual mortality rate of 0.8 (*11*). The intrinsic birth rate, defined as the probability of successful mating, was set at 0.5 to account for the fact that in the simulation, a female may encounter another female (failed mating) or an unmodeled male moved with this female as an imaginary pair (successful mating). Potential mates were selected using inverse distance weighting within 1.255 raster cells (approximately 12.55 km), based on the maximum reported home range of 15,751.36 ha (*95*). Offspring per pair were drawn from a Poisson distribution with a lambda (λ) of four (*84*), under the assumption of a 1:1 sex ratio. Mortality probabilities were parameterized with a minimum of 0.4 (*14*) and a maximum of 1.0, while local population density was estimated using a 1.255-cell window.

Movement and dispersal distances were drawn from lognormal distributions fitted to empirical data. Movement distance followed a distribution with parameters $\mu = -0.856$ and $\sigma = 0.420$, based on the mean home range of 2,149.37 ha and maximum of 15,751.36 ha (*95*). Dispersal distance followed a distribution with $\mu = -0.851$ and $\sigma = 0.424$, based on a mean straight-line dispersal distance of 4.67 km and maximum of 12.74 km (*95*).

Although more detailed genomic simulations were conducted using SLiM, we also implemented a basic genomic architecture in Geonomics. The simulated genome consisted of 2,256 loci unequally distributed across 32 chromosomes; for exact locus distribution, see simulation scripts of this study or Mathur et al., 2023 (*12*). To limit the analysis to standing genetic variation, mutations were disabled, and all neutral loci were initialized at a frequency of 0.5. Loci were treated as codominant, with a fixed recombination rate of 0.001. A single universally advantageous trait was subject to strong selection (selection coefficient = 1) based on habitat suitability. This trait was associated with four loci, calculated as $(2{,}256 \times 2.31) / (1{,}000 + 231 + 2.31)$, following the proportion of beneficial mutations derived from Mathur et al. (2023) (*12*). Effect sizes for these loci were drawn from a normal distribution with a mean of 0.1 (*103*). All other parameters were set to default. The individuals' locations, the number of individuals, and mean fitness were collected every 10 years.

<u>Landscape genomic simulations of climate change (SLiM)</u>

SLiM v4.3 (*27*) was used to conduct forward-time, individual-based evolutionary genomic simulations. We used a two-step process: burn-in for (10 × founding population size) generations, followed by 120 generations to trace evolutionary trajectories over ecological timescales relevant to conservation efforts. During the burn-in phase, we employed a non-Wright-Fisher model with sex and set 11,200 individuals (0.01x of the contemporary population size estimate from Mathur et al. (2023) (*12*) for computational efficiency) as both the founding population size and as the total carrying capacity. We used a mutation rate (μ), recombination rate (r), and selection coefficients (s) that were 100x higher than Mathur et al. (2023) (*12*) since our founding population was 100 times smaller to maintain the same $N_e\mu$, N_er, and N_es (with this setting, 112,000 burn-in generations took about six weeks). We used the same 13 mutation types (six deleterious, six beneficial, and one neutral) from Mathur et al. (2023) (*12*). For detailed simulation parameters, we followed Mathur et al. (2023) (*12*) for the number of genes (= 2,256), gene length (= 1,500 bp), sequence length (= 2,256 × 1,500 bp). Age was randomly assigned to individuals by drawing from a binomial distribution (n = 3 trials; p = 0.5 per trial). Mortality rates of [0.8, 0.8, 0.2, 0.2] were assigned to age categories [0, 1, 2, 3] (*11*). Any individuals older than 3 were culled each generation. Extra individuals above the total carrying capacity were also randomly selected and culled for computational efficiency. Each female could mate with multiple males (with probability chosen from a Poisson distribution with $\lambda = 1$) whereas the number of offspring in each pair was randomly chosen from a Poisson distribution with $\lambda = 8$ (i.e., 2 × 4). Mating was allowed only for individuals of age 1–3. We followed Mathur et al. (2023) (*12*) for genomic parameters (e.g., genome architecture, mutation rate (μ), recombination rate (r), and selection coefficients (s))

Based on the same parameters of burn-in simulations, the main simulations added landscape parameters ('xy' dimension), following Chevy et al. (2025) (*103*), such as the interaction distance 'Sx', mating distance 'Sm', and movement distance 'Sv' (= 0.418 cell of the 10 km × 10 km size, such that 3 × Sx = 3 × Sm = 3 × Sv = 12.55 km), and dispersal distance 'Sd' (= 0.425 cell, such that 3 × Sd = 12.74 km) (*104*). For fecundity- and mortality-based spatial selection, selection coefficients (= 0.1; S_FEC and S_MOR) and dominance coefficients (= 0.5; H_FEC and H_MOR) were set for the adaptive mutation. To regulate the population size based on local density dependence, we set a constant 'rho' (*103*) using a local carrying capacity of 1000 in each cell (*18*). The local density was calculated on the land availability raster based on the

maximum interaction distance (3 × Sx). At every 'first' stage, adults moved towards more suitable habitats within the maximum dispersal distance (3 × Sd). Then in each 'early' stage, chicks randomly dispersed less than the maximum dispersal distance in 'reprising' boundaries. Individuals' fitness was adjusted by the age-dependent mortality and habitat suitability at its location. Every 30 years when the habitat suitability map (and the corresponding land availability map; see *Ecological niche modeling* section above for habitat suitability and land availability rasters) was updated, an adaptive mutation was added to a single offspring individual at the location of the highest suitability to model climate-driven selection. During reproduction, females could mate with multiple males, the same as the burn-in period, but mates were chosen from within the maximum mating distance (3 × Sm). In addition, if a female had the adaptive mutation, its fecundity increased according to 2 × S_FEC × (H_FEC × the number of heterozygotes + the number of adaptive homozygotes). The total carrying capacity was rescaled to 112,000 (0.1x of the original estimate) as well as μ, r, and s accordingly.

At every 'late' stage of the main simulations, we recorded various population and genomic metrics using a custom function based on Robinson et al. (2022) (*105*) which are: the population size, mean population fitness, mean individual heterozygosity, mean inbreeding load, mean fraction of runs of homozygosity (F_{ROH}) for various lengths (> 100 kb, 500 kb, 1 Mb, or 10 Mb), the number of deleterious and beneficial mutations under different selection intensities (very strong, $|s| \geq 0.5$; strong, $|s| \geq 0.1$; moderate, $|s| \geq 0.01$; weak, $|s| \geq 0.0001$), mean potential load and mean realized load (see Mathur & DeWoody (2021) (*22*) for equations and formal definitions) of various mutation types (high, moderate, low).

Genetic rescue simulations

We conducted genetic rescue simulations in SLiM to determine how the Central Texas population of conservation concern (*12*) might respond to assisted gene flow from a donor population. Southeast Mexico and New Mexico were evaluated as potential donors because of their comparable adaptive profiles (in the case of southeast Mexico) and for logistical feasibility (New Mexico). For the southeast Mexico samples, we selected 13 individuals from locations closest to the region sharing a comparable adaptive profile with the recipient population and exhibiting the highest heterozygosity. For New Mexico samples, we selected seven samples

previously sequenced by Mathur & DeWoody (2021) (*22*) (NCBI SRA No. SRR11514058–SRR11514065).

For the genetic rescue simulations, we kept most parameters and the model structure the same as the landscape genomic simulations, including the mean number of offsprings per female, recombination rate, age range, culling of oldest individuals, and model type (non-Wright-Fisher model). However, we did not apply 'xy' dimension and landscape features as the simulations are time-intensive and we focused on the dynamics of genetic rescue. We incorporated the empirical variants of the focal donor and recipient populations when rescue simulations were initiated. We assigned a population-specific carrying capacity as $10N_e$ where N_e was estimated using currentNe2 v1.0.1 (*106*) and the SNP data. Because the estimates did not converge for the Central Texas and New Mexico samples when analyzed separately, we pooled each with a genetically and geographically proximate population (Central + West Texas; New Mexico + Mexican Sonora) and estimated N_e for the combined cluster. The cluster-level estimate was then divided by two (i.e., the number of pooled populations) to approximate N_e for each focal population. New mutations were disabled to focus on the dynamics of existing empirical variants in contrast to the climate change simulations. We initialized standing genetic variation using VCF files that contained empirical variants (both SNPs and SVs) and we specified ancestral nucleotide states using a FASTA file. To increase biological realism, we constrained the assignment of fitness effects to initial mutations such that the proportions of deleterious variants among genome-wide SNPs + SVs and outlier SVs matched the empirical proportions of high-impact, moderate-impact, and low-impact variants (with fixed selection coefficients of -0.01, -0.001, and -0.0001; and dominance coefficients of 0, 0.4, and 0.5, respectively) estimated from the VCF data, including New Mexico samples (tables S12–S13). Because the cumulative fitness effects of deleterious mutations impeded simulation persistence under more restrictive demographic settings, we did not apply age-related mortality. Instead, population regulation was implemented via carrying-capacity–based fitness scaling. A relaxed mating system was used in which females sampled a Poisson-distributed number of mates, allowing repeated mating with the same male only under male-limited conditions.

For compatibility with SLiM, we converted SVs to dummy SNPs, recoded all derived states in multiple-allelic polymorphisms as the first derived allele, and replaced missing data

with reference alleles in the VCF files using custom python scripts (https://github.com/jyj5558/monq_adaptive_genomics). To enable tracking of non-variant sites, pre-existing empirical variant sites, and outlier SV positions during the simulations, we encoded their reference alleles in the ancestral FASTA as C, A, and G, respectively. To represent multiple chromosomes in SLiM's single-chromosome framework, we inserted a "T" between chromosomes to enforce a recombination rate of 0.5 at those boundaries.

After population initialization, we ran two short burn-in periods to stabilize population sizes without substantially altering the empirical genetic state. The first burn-in (five generations) allowed all populations to reach carrying capacity from the VCF-derived variation, and the second burn-in (ten generations) allowed demographic recovery following the application of empirical deleterious variant proportions, which induced a transient genomic load shock. The model subsequently implemented three rescue events, each introducing 50 or 100 individuals annually from southeast Mexico or New Mexico. We then tracked the same metrics as in the climate-change simulations and additionally monitored donor ancestry proportions at all variant sites as well as ancestry at pre-existing genome-wide variant sites and outlier SV positions in the recipient population over the next 100 years, plus the genomic loads of these sites in total and respectively.

Supplementary Text

Reference genome assembly and annotation

One SMRTcell of PacBio HiFi CCS sequencing yielded 36.1 Gbps among 2.3 M reads with a mean read length of 15.7 kb. Jellyfish analysis of 21-mers predicted a 0.883 Gb haploid genome length, with a heterozygosity percentage of 0.657% (≈ 1 heterozygous site per 152 bp). Following de novo assembly, 901 contigs were retained with a N50 of 8.1 Mb, L50 of 37, and genome length of 1.027 Gb. No contig was identified as contaminated. Purging haplotigs kept 521 out of 901 contigs. From the de novo assembly before scaffolding, 96.7% complete BUSCOs were identified (consisting of 96.4% sing-copy, 0.3% duplicated, 0.4% fragmented, and 2,9% missing BUSCOs). A total of 974 M Tell-seq read-pairs were yielded by a NovaSeq X Plus run by which the assembly contigs were then scaffolded. All the resultant scaffolds could be mapped and covered by our test data (i.e., independently downloaded short reads). This resulted in 287 scaffolds, with a N50 of 36.2 Mb, a L50 of 7, and haploid genome length of 1.012 Gb (table S1).

The NCBI Eukaryotic Genome Annotation Pipeline identified 16,760 total genes with a mean length of 31,466 bp (16,975 including pseudogenes; GCF_038088225.1). Of these, 15,455 protein-coding genes contained 98.7% of the Avian orthologs, which were classified as complete (98.4% single copy, 0.3% duplicated), 0.2% as fragmented, and 1.1% as missing. Furthermore, 25,270 fully-supported mRNAs and 1,243 fully-supported non-coding RNAs were annotated in the MONQ scaffold level assembly. The annotation pipeline masked 21.50% of the reference genome as repeat regions.

Comparisons with previous non-model bird pangenomes

Considering our data size of 20 haplotypes at ~10x, plus 2 haplotypes at ~36x (reference individual), our initial graph pangenome was comparable to previous non-model bird pangenomes. The number of nodes (~57 million) and edges (~78 million) were two thirds compared to the barn swallow pangenome (92 million nodes, 126 million edges, 1.2 Gb; reconstructed from Secomandi et al. 2023) (*7, 107*) where 12 haplotypes at ~20x depth were

used, 3–4 times lower than the house finch pangenome (*108*) that incorporated 35 haplotypes at ~42x, and 4–5 times lower than the capuchino seedeater pangenome. Augmenting the pangenome with 91 short-read individuals increased the size from 1.1 Gb to 2.02 Gb, which is about two times larger than the linear reference genome (1.01 Gb).

From the augmented pangenome, over five times more insertions were found (24,195 insertions vs. 4792 deletions) with three times longer length (mean 1055 bp of insertions vs. 351 bp of deletions) than deletions after quality filtering. Our two-step outlier detection approach also detected about 2.5 times more insertions (373 insertions vs. 153 deletions) with twice as long (mean 626 bp of insertions vs. 327 bp of deletions) than deletions. However, A higher proportion of selection outliers was observed among deletions (3.19%) than insertions (1.54%), suggesting a greater contribution of deletions to environmental adaptation. Thus, while insertions were a primary SV type of the species, the species adaptation could be driven by deletions. But note that defining insertions or deletions could be flipped based on the reference genome sequence in our backbone-based pangenome approach which is different from backbone-free approach employed by Edwards et al. (2025) (*15*) where deletions were dominant SVs. Yet, the mean lengths of our insertions (1055 bp) and deletions (351 bp) were comparable to theirs (insertions: 826–1201; deletions: 325–331).

Overall, our range-wide pangenome, one of the first in a wild bird, encompasses far more intraspecific variation than any single reference genome and provides a comprehensive basis for downstream adaptive analyses. The adaptive outlier SVs in the pangenome, in particular, aligned with pronounced phenotypic differentiation and broad environmental conditions across the landscape. These SVs were not merely statistical artifacts but were consistently associated with environmental variables and phenotypic traits using independently collected datasets, underscoring their biological relevance.

Functional and phenotypic variation

Gene ontology (GO) and KEGG pathway enrichment analyses revealed no GO or KEGG categories remained significant after statistical corrections for multiple testing. However, several terms were enriched at a relaxed threshold ($p < 0.01$) providing exploratory insight. Enriched GO

terms included “DNA-(apurinic/apyrimidinic site) endonuclease activity,” “cellular response to epidermal growth factor stimulus,” and “cilium” which were associated genes APFL, LEIL3; GAREM1, PDE8A; and TULP3, WDR19, IFT52, ABCC8, respectively. The KEGG pathway “Purine metabolism” contained multiple genes (ENPP3, ADK, PDE1A, PDE3A, ADCY5, PDE8A, NME7, FHIT) intersecting outlier SVs (table S4).

Moreover, the identity of genes associated with outlier SVs suggests a coherent pattern of adaptation to winter-related physiological stress (see the main text). Notably, APLF, associated with the GO term *DNA-(apurinic or apyrimidinic site) endonuclease activity*, was correlated with precipitation seasonality (Bio15), whereas NEIL3, a gene involved in oxidative DNA damage repair, was associated with driest-month precipitation (Bio14). Precipitation variability has been linked to increased oxidative stress and telomere shortening in birds, implying elevated genomic damage under harsh winter environments (*109*). Consistent with this, dry winters are known to intensify oxidative stress, increasing the importance of DNA repair capacity (*110*) for maintaining cellular integrity and fitness. Oxidative stress is also known to drive selection of purine metabolism-related genes, *ADCY5* (associated with Bio14), *ENPP3* (associated with Bio15), and *PDE8A* (associated with Bio7) in cetaceans, artiodactyls, and Tibetan highland mammals by convergent evolution (*111*). These associations suggest that variation in DNA repair pathways may contribute to adaptive responses to winter climatic stress in diverse animals, including Montezuma quail.

Additional candidate genes further support a functional link between winter climate and physiological resilience. Several genes involved in purine metabolism and cyclic AMP signaling (e.g., *ENPP3*, *ADK*, *PDE1A*, *PDE3A*, *PDE8A*, *ADCY5*) were associated with Bio7 and Bio14, pointing to potential modulation of energy metabolism, blood pressure, cardiovascular functions, and cellular signaling under cold and dry conditions. Similarly, cilium-related genes (*IFT52, WDR19, TULP3*), which play roles in cellular sensing and transport, were also associated with winter climate variables. Notably, *PDE8A*, *IFT52*, *ABCC8*, *ADK*, *FHIT*, *PDE1A*, *PDE3A*, *ADCY5*, *NME7* have previously been implicated in cold-temperature or high-altitude adaptation in sheep, amphipods, goats, cattle, donkeys, pigs, boars, muskoxen, and slate-colored coots (*112–120*), while *PDE3A* and *ADCY5* has also been implicated in drought adaptation in many sheep

breeds (*121*, *122*), suggesting that similar physiological pathways may be repeatedly targeted by selection in seasonally stressful environments.

Associations with phenotypic variation revealed that outlier SVs exerted measurable effects on plumage traits. A series of phenotype–genotype RDAs using 1–12 genotype PCs showed that the constrained rank stabilized at three once ≥3 genotype PCs were included, indicating that only three independent genotype dimensions contributed to explaining the observed phenotypic variation (fig. S5, table S11). Model significance was retained up to $k = 4$, and adjusted R^2 values plateaued between 0.176 and 0.195 for $k \leq 3$ (0.143 for $k = 4$). At $k = 3$, RDA1 was the only significant axis, explaining 23.9% of the constrained phenotypic variance. The first genotype PC loaded significantly on RDA1 after permutation testing, identifying it as the primary genetic axis associated with phenotypic variation. Response loadings on RDA1 indicated that L* contributed most strongly to the constrained axis (loading = 1.102), followed by b* (0.837) and a* (−0.488). Univariate multiple regression further supported this pattern. The model with L* as the response variable was highly significant ($p = 1.22 \times 10^{-5}$; adjusted $R^2 = 0.695$), with both genotype PC1 and population-structure PC1 emerging as significant predictors ($p < 0.001$ for both). Their standardized effect sizes ($\omega^2 = 0.549$ and 0.485, respectively) indicate that genetic variation captured by outlier SVs substantially contributes to plumage lightness independently of broad-scale population structure. In contrast, models with a* and b* as response variables were not significant overall, although genotype PC1 remained a significant predictor for b*. Body mass was also significantly explained ($p = 0.008$), but this effect was driven by age and population-structure PC1 rather than genotype PCs. Partial least squares (PLS) analyses corroborated the association between genotype PCs and plumage lightness. In the optimal one-component model, the first PLS axis explained 91.2% of the variance in genotype predictors and 68.2% of the variance in phenotypic responses. The correlation between genotypes and multivariate phenotype scores was 0.60, increasing to 0.63 when considering L* alone, and remained significant after 999 permutations ($p = 0.002$). However, when the effects of population structure and/or climatic variables were removed prior to PLS, predictive power was lost (all $Q^2 < 0$) and permutation tests were no longer significant.

Consistent with the genotype-phenotype association, several vertebrate genes intersecting outlier SVs have previously been implicated in plumage or skin coloration. For example, *ABCC8*

has been linked to variation in skin pigmentation of Virginia opossums (*123*). Moreover, *ADCY5* has been functionally implicated in melanogenesis pathways of ducks and zebrafish (*124–126*) whereas *PDE3A* and *FHIT* have been associated with white-to-black plumage differentiation in chickens and ducks (*127*, *128*). Together, these findings suggest that environment-associated SVs influence plumage traits primarily through mechanisms affecting feather structure, condition, and lightness, rather than through changes in hue-based pigmentation. Importantly, the genes highlighted by functional annotation were not among the highest-loading SVs on genotype PC1 used in the analysis, suggesting that the observed genotype–phenotype associations are influenced by several environment-associated variants that exhibit modest and heterogeneous loadings on genotype PC1 (rather than by a small number of high-loading loci). This pattern is consistent with an underlying process of spatially variable, locally adaptive effects rather than uniform genome-wide shifts.

Population structure based on SNPs

The PCA of SNPs revealed clear population structure aligned with geography (while the genome-wide SV-based PCA yielded similar but not identical patterns; Fig. 4). Texas and Coahuila populations separated from the rest of Mexico along PC1 (21.4% variance), while Mexican individuals formed a linear continuum from Sonora to Mexico City along PC2 (13.2% variance). West and Central Texas populations separated along PC3 (1.8% variance). Admixture analysis supported K = 2 as the optimal clustering (fig. S6), distinguishing U.S. from Mexican individuals, with K = 3 further dividing southeastern Mexico of the Sierra Madra Oriental from western Mexico along the Sierra Madre Occidental.

Notes on southeast Mexico-based genetic rescue simulations

Rescue scenarios using Southeast Mexico birds as a source or donor population revealed pronounced scale dependence. Translocations of 50 individuals failed to alleviate genetic erosion and showed limited benefits across genomic indicators, whereas introducing 100 individuals improved realized genomic load and reduced deleterious homozygosity—particularly at empirically identified outlier SVs—while maintaining substantially lower donor ancestry

proportions than New Mexico rescues. This pattern suggests that larger Southeast Mexico translocations may offer a more balanced genetic rescue strategy, achieving measurable genomic benefits while better preserving the unique adaptive profile of the Central Texas population. Together with the results presented in the main text, these genomic simulations illustrate how various rescue scenarios can be explored and evaluated *in silico*, but until such simulations are groundtruthed they should be viewed like the weekend's weather forecast: useful for planning purposes, but almost certainly wrong to some degree.

Note on ecological niche modeling results

Briefly, the final ensemble model performed strongly (TSS = 0.928; ROC = 0.994). Projected habitat suitability fluctuated through time under both climate scenarios, with sustained decreases (2041–2070) and increases (2011–2040 and 2071–2100) (figs. S23–S24). Habitat suitability generally declined under higher future temperatures, with a decrease in maximum suitability, and differences among scenarios were relatively small. Suitable habitats in Central Texas disappeared, whereas habitats in the Northwestern region—such as Arizona, New Mexico, Sonora, Chihuahua—remained or expanded. West Texas exhibited transient suitability only during periods of overall habitat rebound. We do not describe the ENM in detail here, as it is not the primary focus of this study and is used only as background for the genomic simulations. Readers interested in a more comprehensive and rigorous ENMs are referred to a forthcoming independent paper.

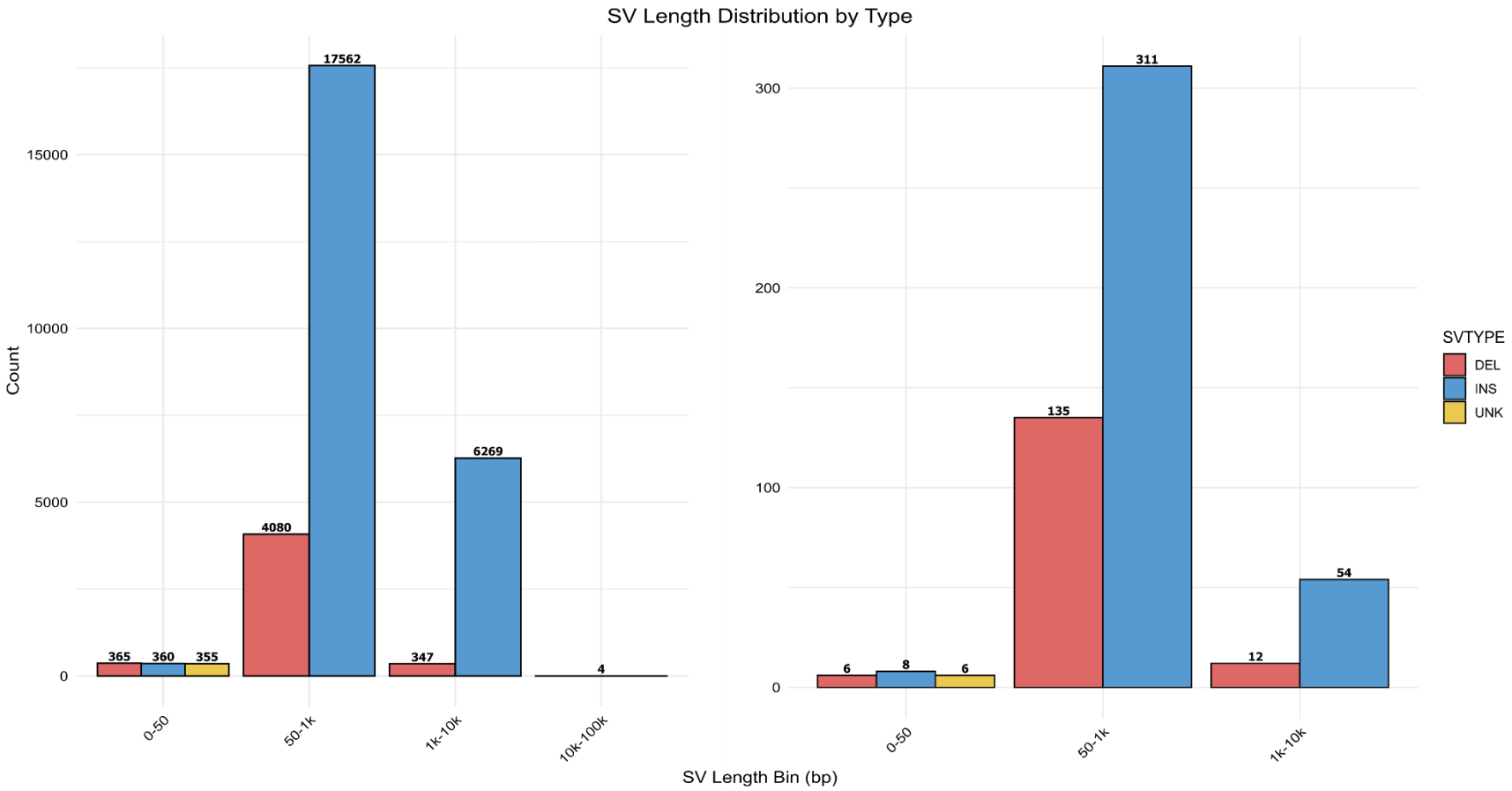

Fig. S1. Length distribution of all structural variants (SVs; left) and outlier SVs (right) by type. The SV type includes deletion (DEL), insertion (INS), and unknown (UNK; mostly multi-nucleotide polymorphisms).

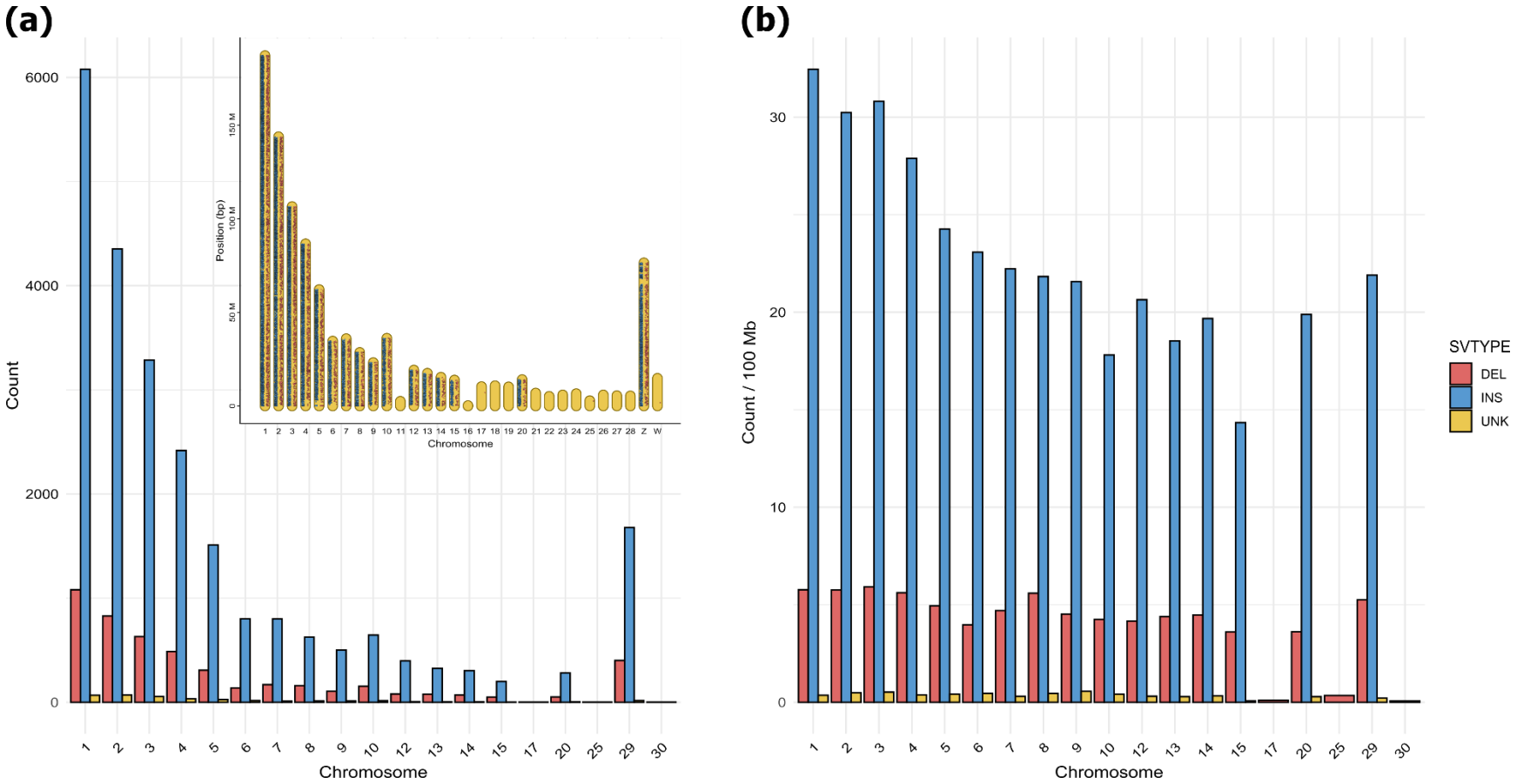

Fig. S2. Chromosomal distribution of genome-wide structural variants (SVs) that mapped to the chicken genome. SV types include deletions (DEL), insertions (INS), and unknown variants (UNK; primarily multi-nucleotide polymorphisms). (a) Number of genome-wide SVs per chromosome. (b) Number of genome-wide SVs per 100 Mb per chromosome. Only chromosomes harboring SVs are shown on the x-axis. The inset displays the distribution of SVs across all chromosomes. Genome-wide SV counts and density scaled with chromosome length and showed no evidence of chromosome-specific clustering.

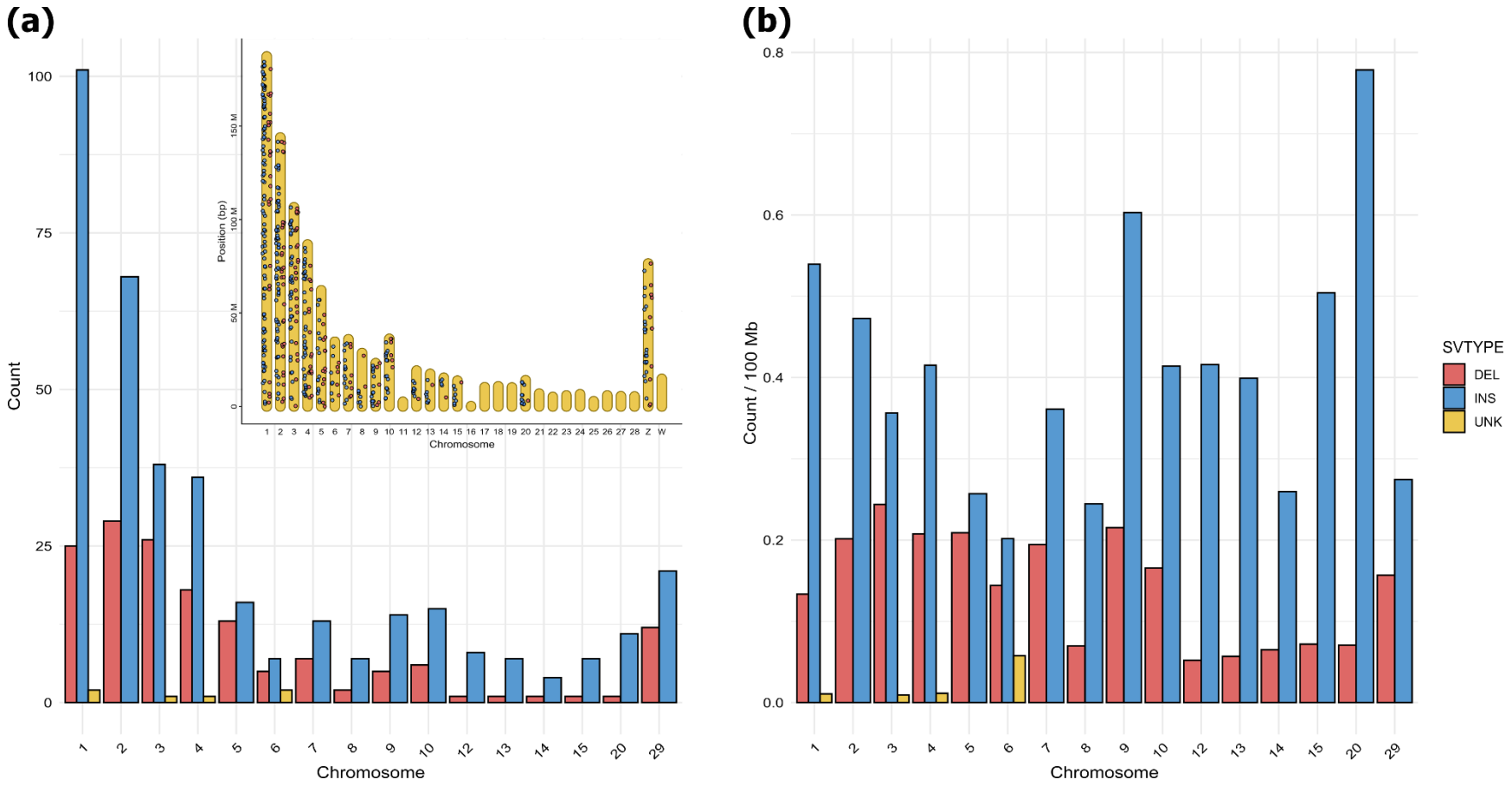

Fig. S3. Chromosomal distribution of outlier structural variants (SVs) that mapped to the chicken genome. SV types include deletions (DEL), insertions (INS), and unknown variants (UNK; primarily multi-nucleotide polymorphisms). (a) Number of outlier SVs per chromosome. (b) Number of outlier SVs per 100 Mb per chromosome. Only chromosomes harboring SVs are shown on the x-axis. The inset displays the distribution of SVs across all chromosomes. Outlier SV counts and density did not scale with chromosome length but showed suggestive evidence of chromosome-specific clustering.

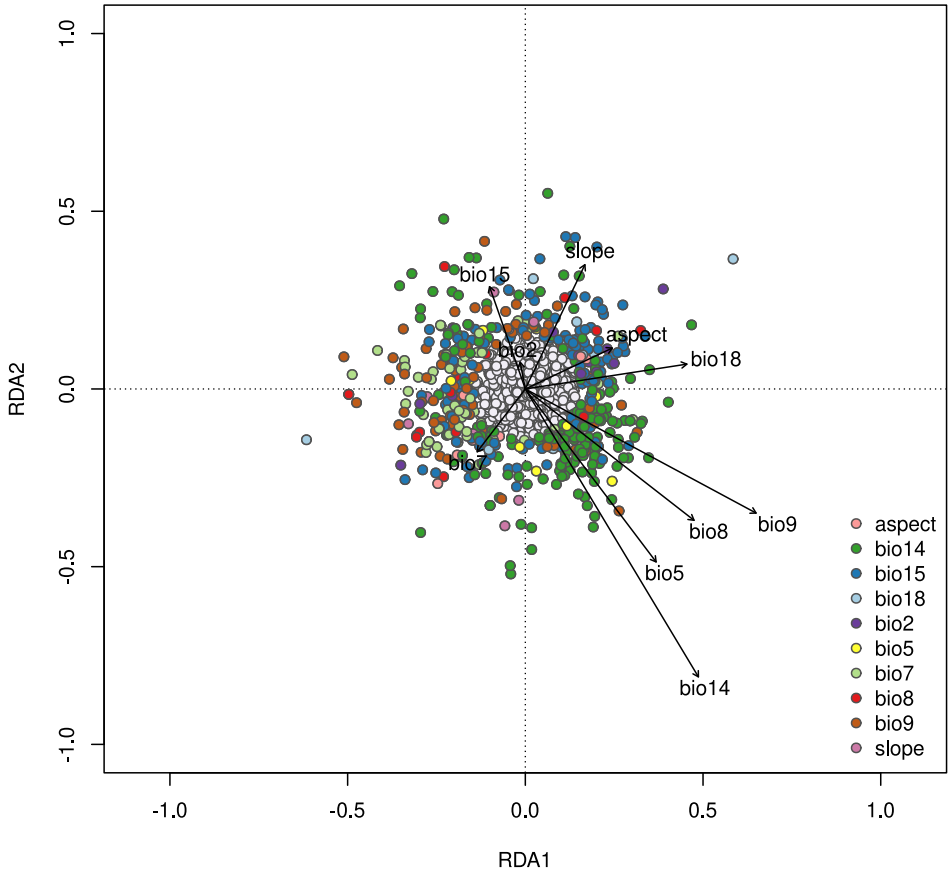

Fig. S4. RDA biplot of environmental variables and structural genomic variants. Neutral variants are shown in grey and outlier variants are color-coded according to the environmental variable of the highest correlation. The environmental arrows were scaled for proper visualization of their relative lengths. The Bio14 and Bio15 variables were the dominant environmental factors associated with the most outliers (n = 200 and 144, respectively). The two most influential environmental variables on RDA1 and RDA2 were Bio9 & Bio14 for the former and Bio5 & Bio14 for the latter.

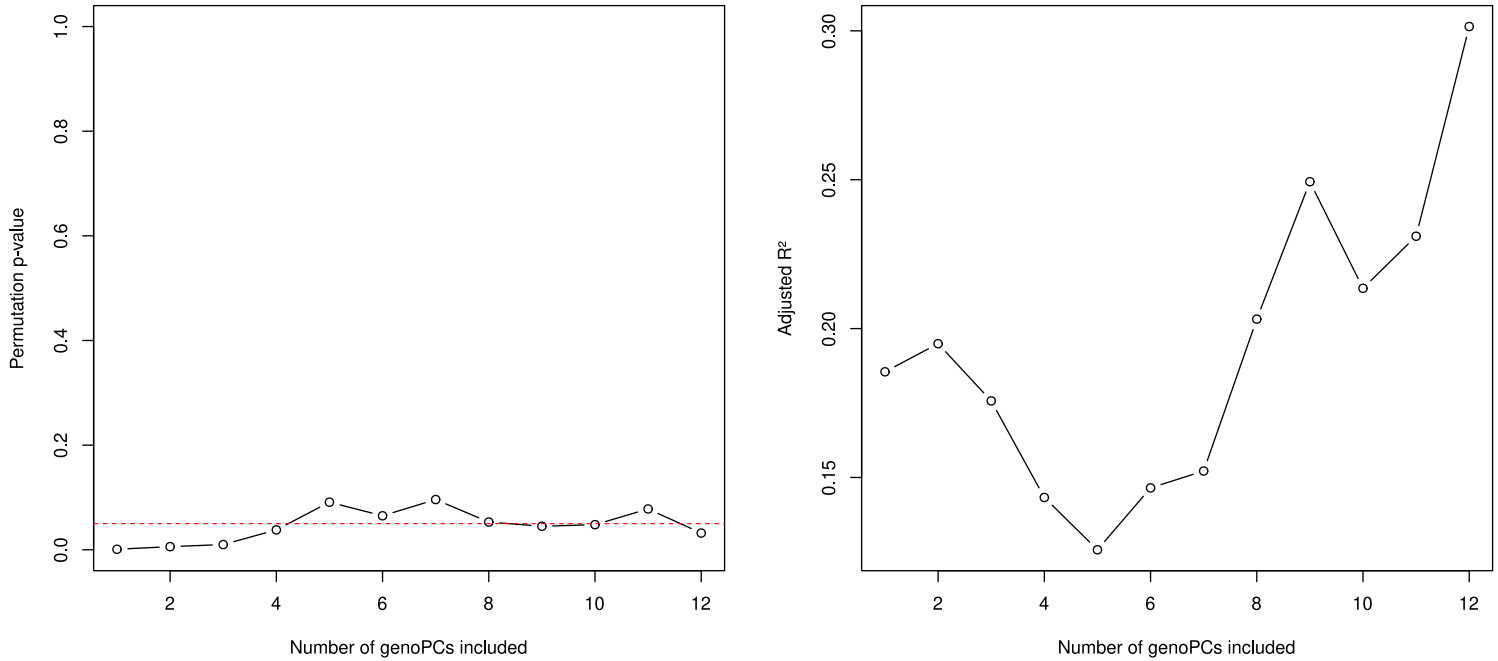

Fig. S5. Permutation *p*-values (left) and adjusted R^2 values (right) according to the number of outlier-SV genotype PCs (genoPCs) in a series of genotype–phenotype RDAs. The red dashed line indicates $p = 0.05$. Interpreting with ttable S11, three genoPCs contributed to explaining phenotypic variation based on the significant *p*-value and stabilized constrained rank.

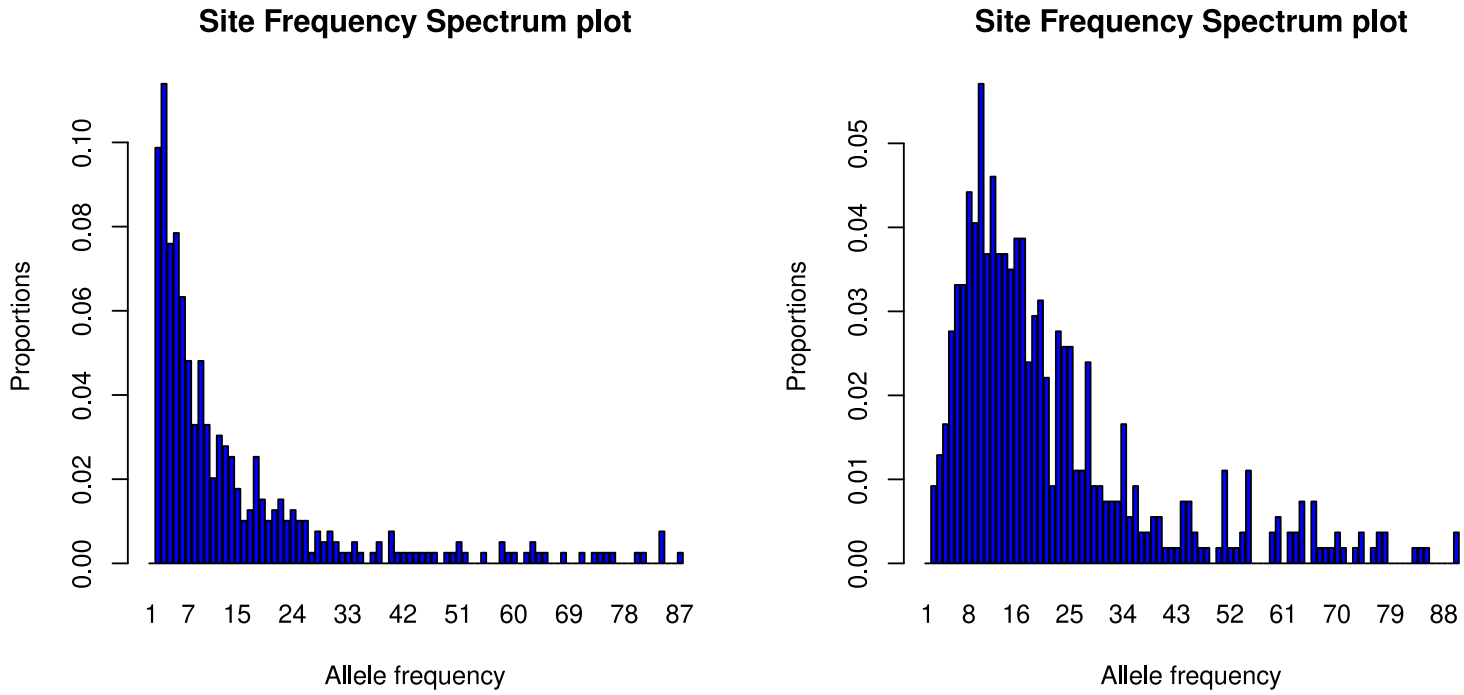

Fig. S6. Site frequency spectrum (SFS) of all (left) and outlier structural variants (right). The overall shape of the SFS is consistent with neutrality of the 'all' variants, but not so with the outlier variants.

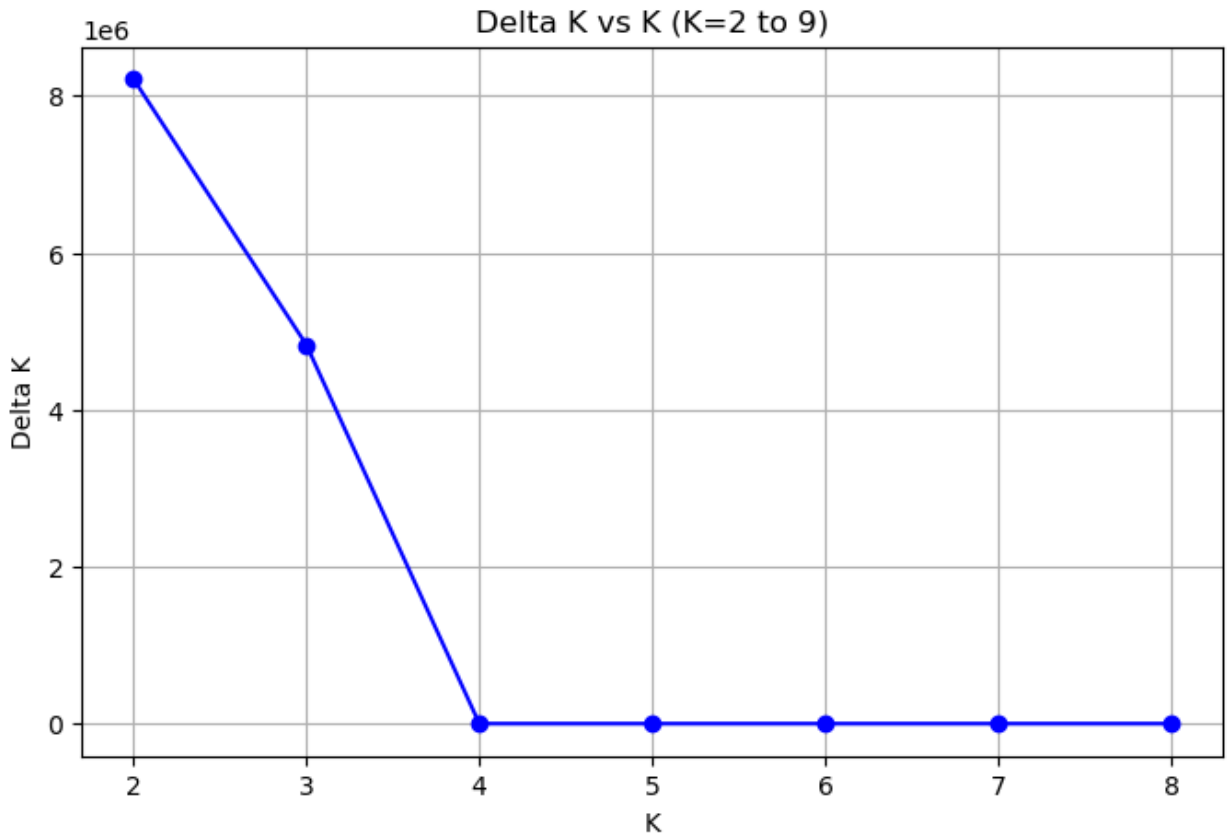

Fig. S7. Delta K values from K = 2 to K= 9 from admixture analysis according to the Evanno et al. (2005) (*81*) method. The best-supported K value was 2.

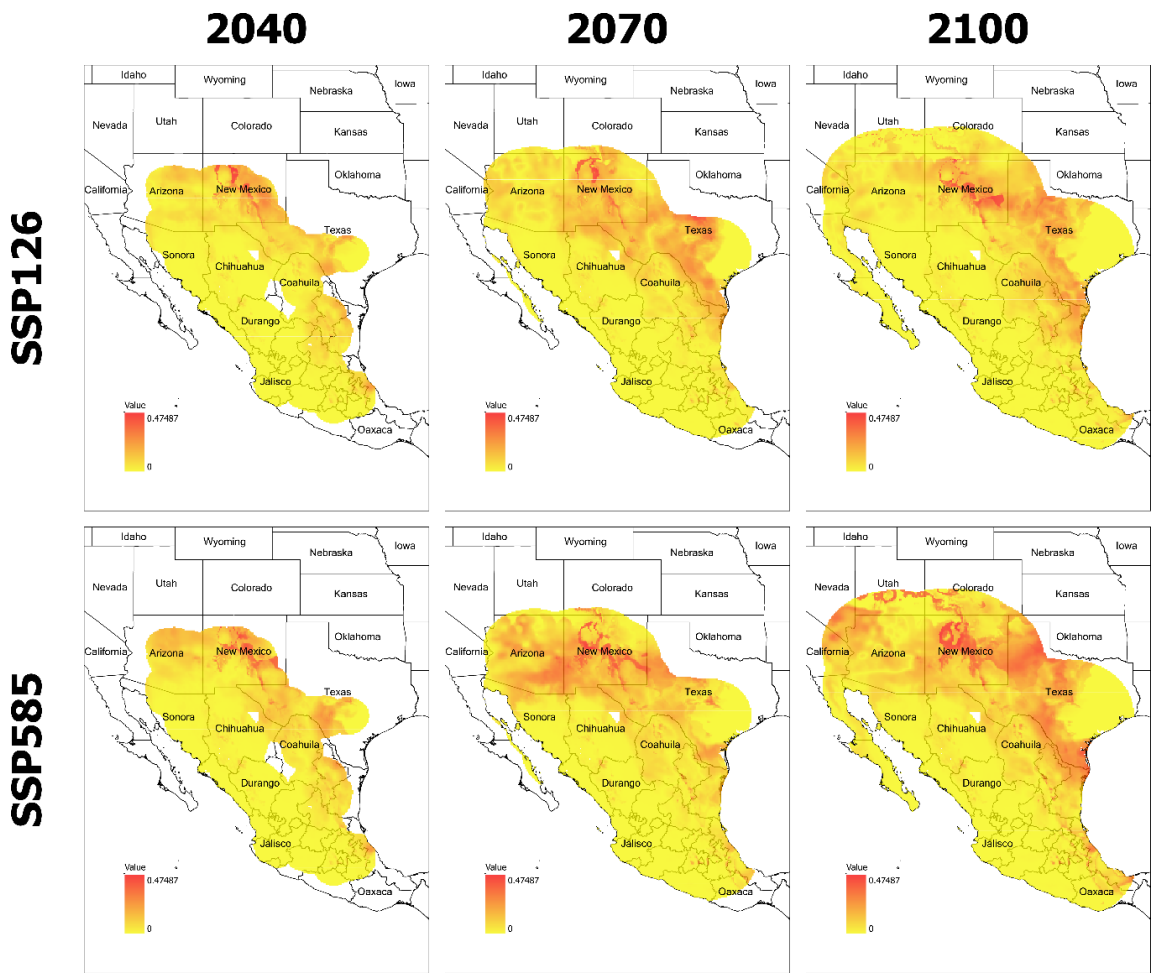

Fig. S8. Projected local genomic offset derived from Generalized Dissimilarity Modeling (GDM). Rows correspond to two climate scenarios: SSP126 (the most extreme scenario; top) and SSP585 (the most moderate scenario; bottom). Columns represent three future time periods: 2040 (left), 2070 (middle), and 2100 (right). The depicted geographic range at each time period reflects the maximum possible extent of the species given its dispersal capacity up to that time. Regardless of the scenario, SSP126 or 585, southwestern USA and eastern Mexico regions show higher genomic mismatches than others.

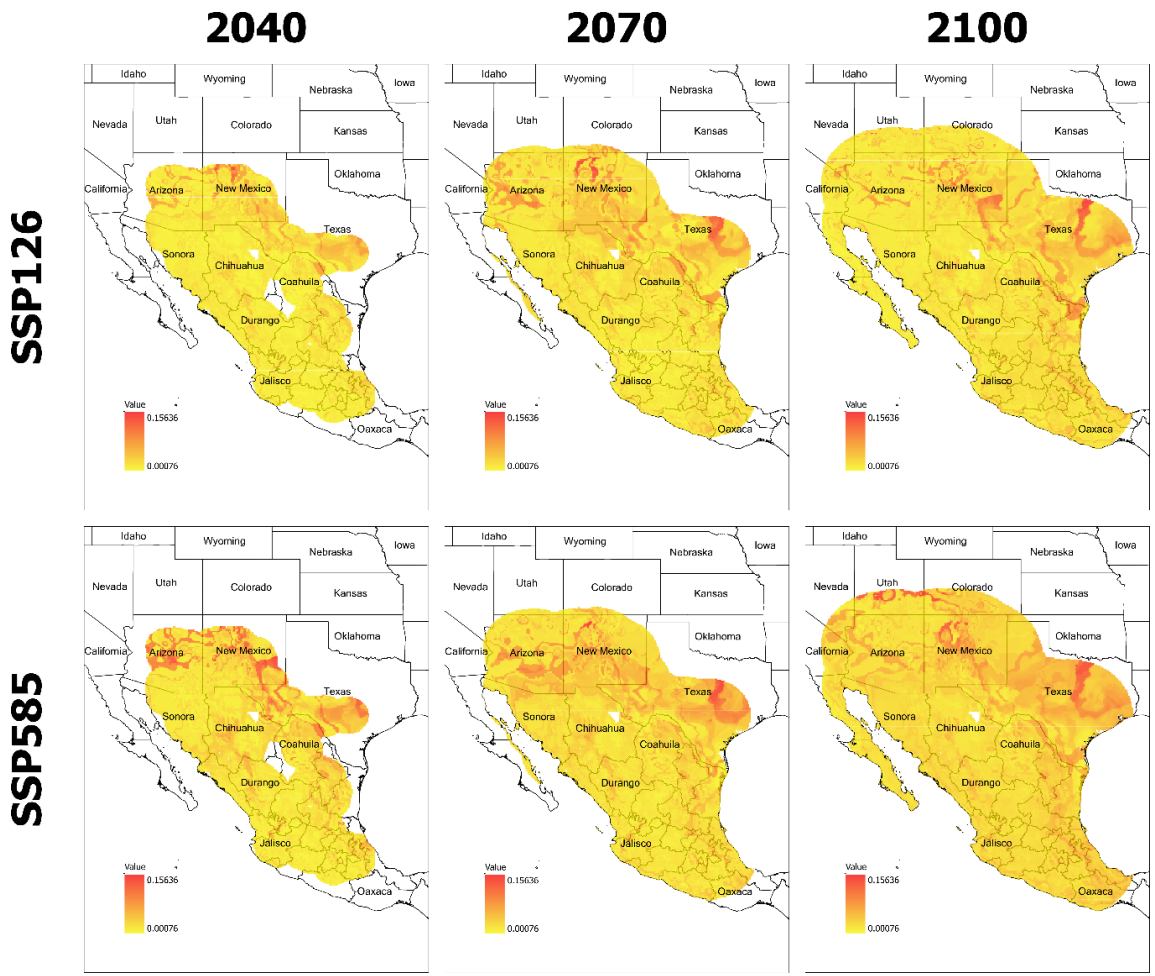

Fig. S9. Projected local genomic offset derived from Gradient Forest (GF). Rows correspond to two climate scenarios: SSP126 (the most extreme scenario; top) and SSP585 (the most moderate scenario; bottom). Columns represent three future time periods: 2040 (left), 2070 (middle), and 2100 (right). The depicted geographic range at each time period reflects the maximum possible extent of the species given its dispersal capacity up to that time. Regardless of the scenario, SSP126 or 585, central Texas regions show increasing genomic mismatches farther into the future and this pattern was more pronounced under SSP585.

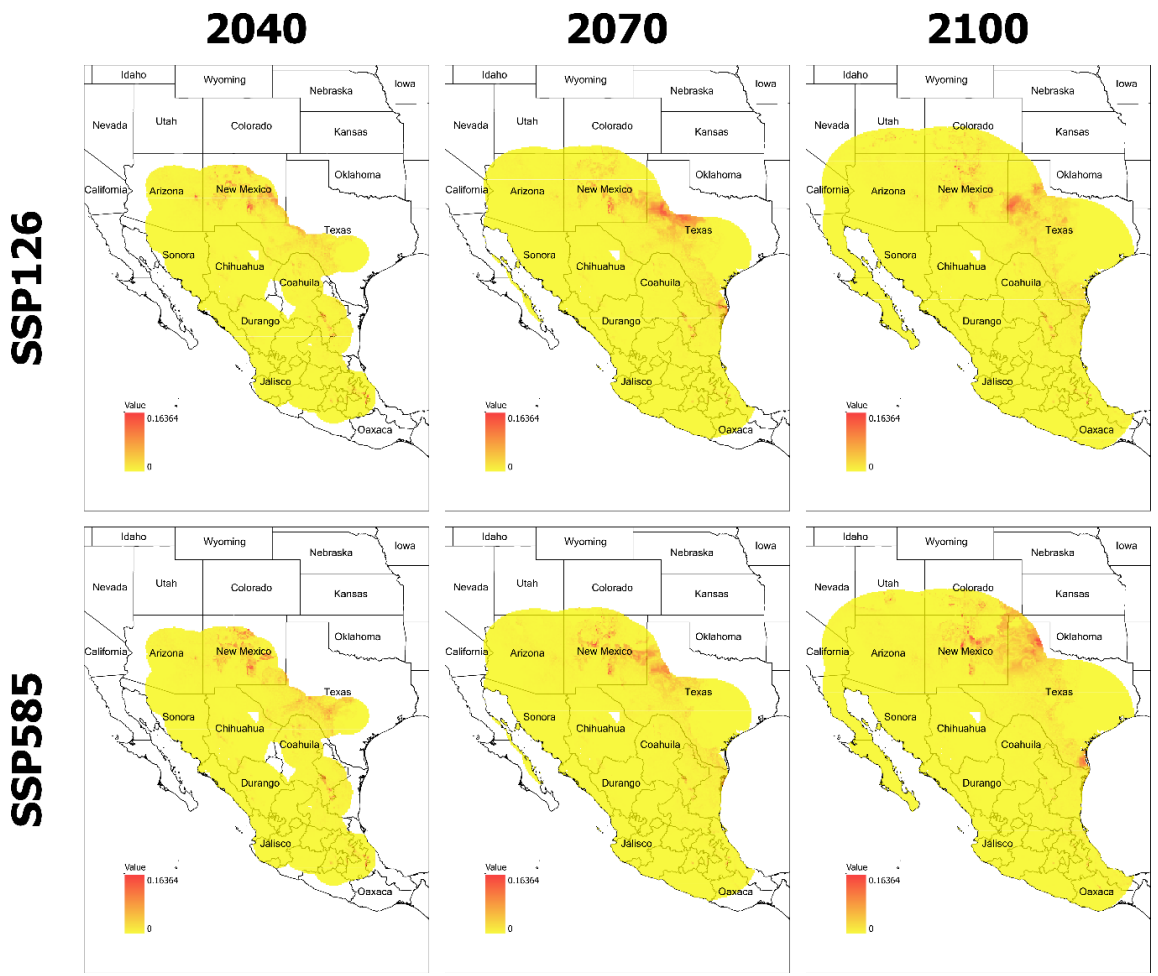

Fig. S10. Projected reverse genomic offset derived from Generalized Dissimilarity Modeling (GDM). Rows correspond to two climate scenarios: SSP126 (the most extreme scenario; top) and SSP585 (the most moderate scenario; bottom). Columns represent three future time periods: 2040 (left), 2070 (middle), and 2100 (right). The depicted geographic range at each time period reflects the maximum possible extent of the species given its dispersal capacity up to that time. Regardless of the scenario, SSP126 or 585, Texas, New Mexico, and eastern Mexico regions show higher genomic mismatches than other regions.

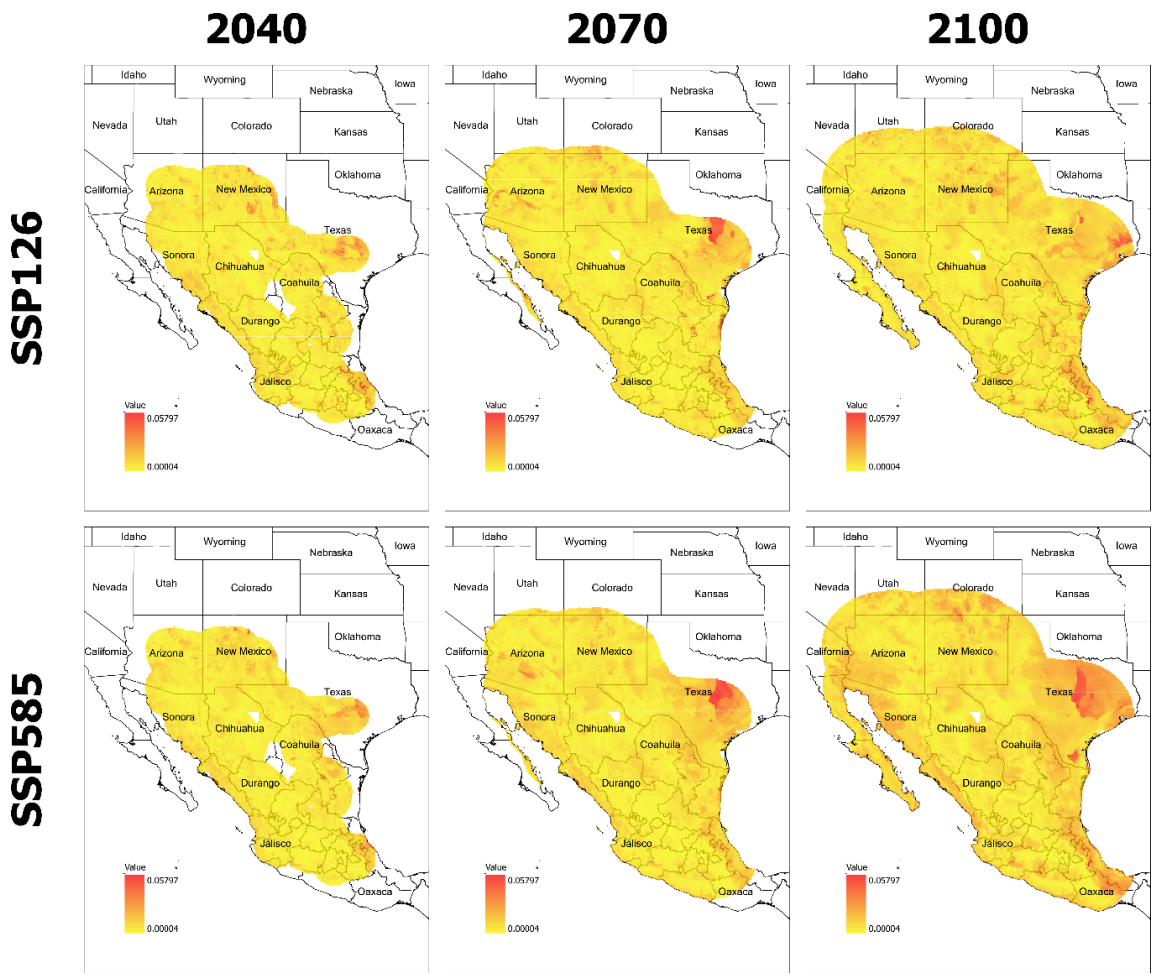

Fig. S11. Projected reverse genomic offset derived from Gradient Forest (GF). Rows correspond to two climate scenarios: SSP126 (the most extreme scenario; top) and SSP585 (the most moderate scenario; bottom). Columns represent three future time periods: 2040 (left), 2070 (middle), and 2100 (right). The depicted geographic range at each time period reflects the maximum possible extent of the species given its dispersal capacity up to that time. Regardless of the scenario, SSP126 or 585, central Texas and southeastern Mexico regions show increasing genomic mismatches farther into the future and this pattern was more pronounced under SSP585.

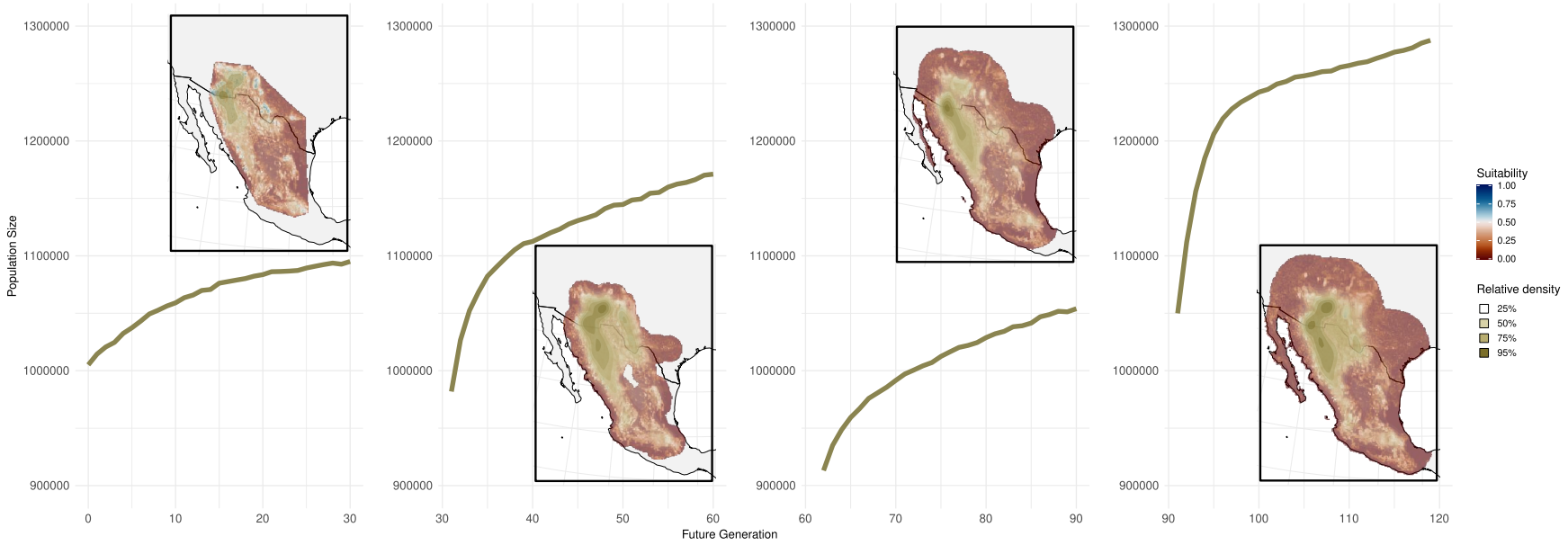

Fig. S12. Population size trend of the Montezuma quail from the present (1st column) to 2040 (2nd column), 2070 (3rd column), and 2100 (4th column) based on the most moderate climate change scenario. Population density on the corresponding habitat suitability map is embedded as an inset in each period. In general, the entire population size is increasing while most individuals kept moving northwest in the future periods, even beyond the suitable habitat range.

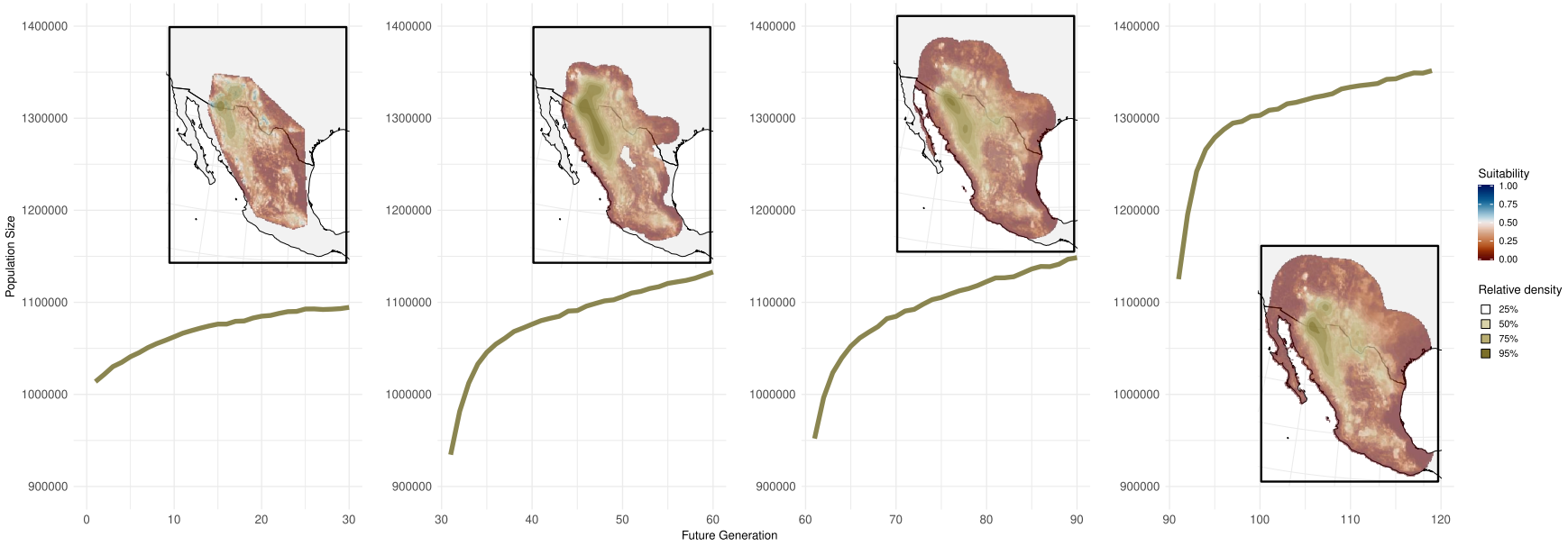

Fig. S13. Population size trend of the Montezuma quail from the present (1st column) to 2040 (2nd column), 2070 (3rd column), and 2100 (4th column) based on the most extreme climate change scenario. Population density on the corresponding habitat suitability map is embedded as an inset in each period. In general, the entire population size is increasing while most individuals kept moving northwest in the future periods, even beyond the suitable habitat range.

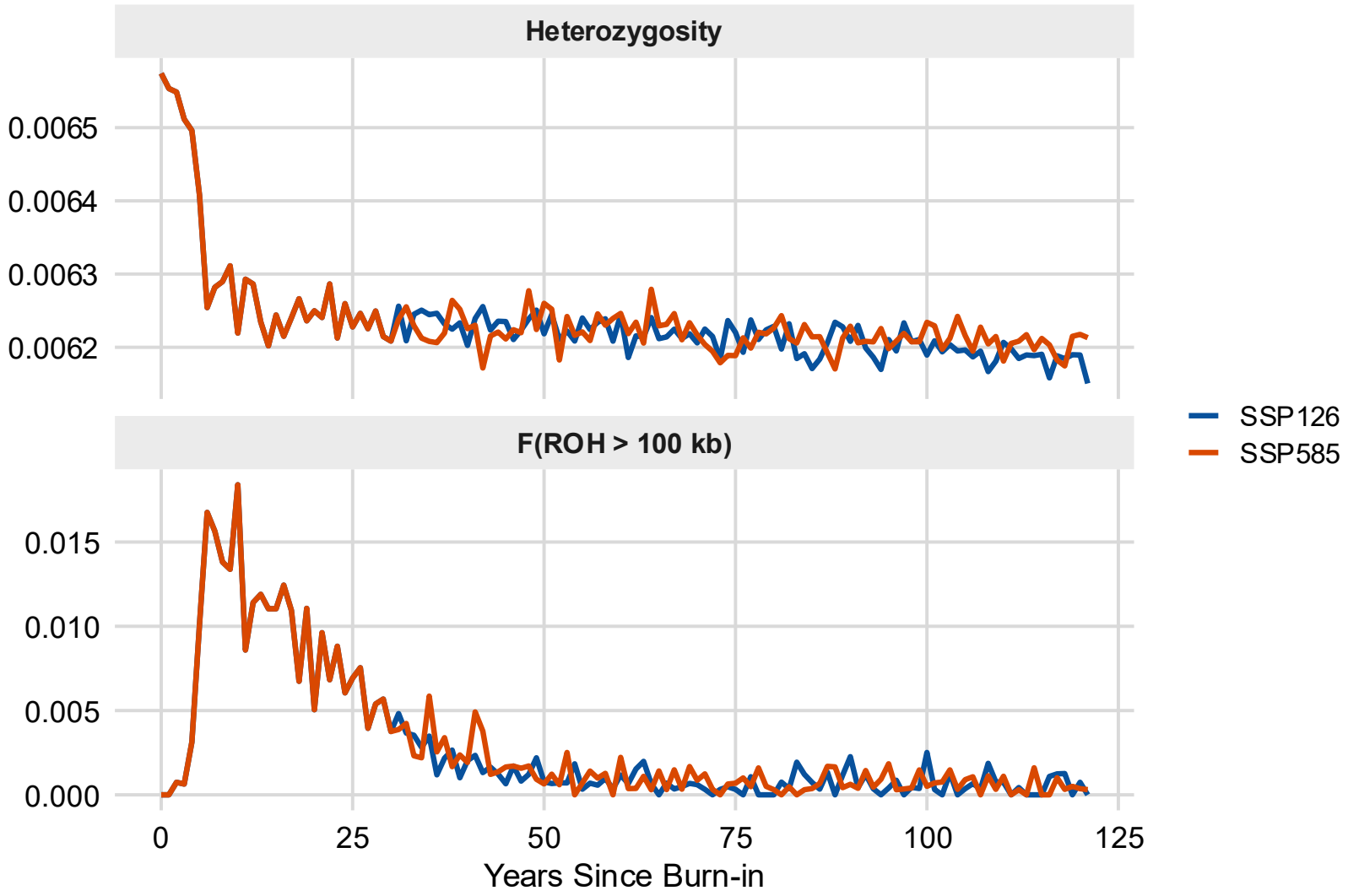

Fig. S14. Fraction of runs of homozygosity (F_{ROH}) longer than 100 kb and heterozygosity under climate change scenarios. Climate change scenarios are color-coded as indicated in the legend. Both genomic metrics show a consistent decreasing trend across scenarios throughout the simulations.

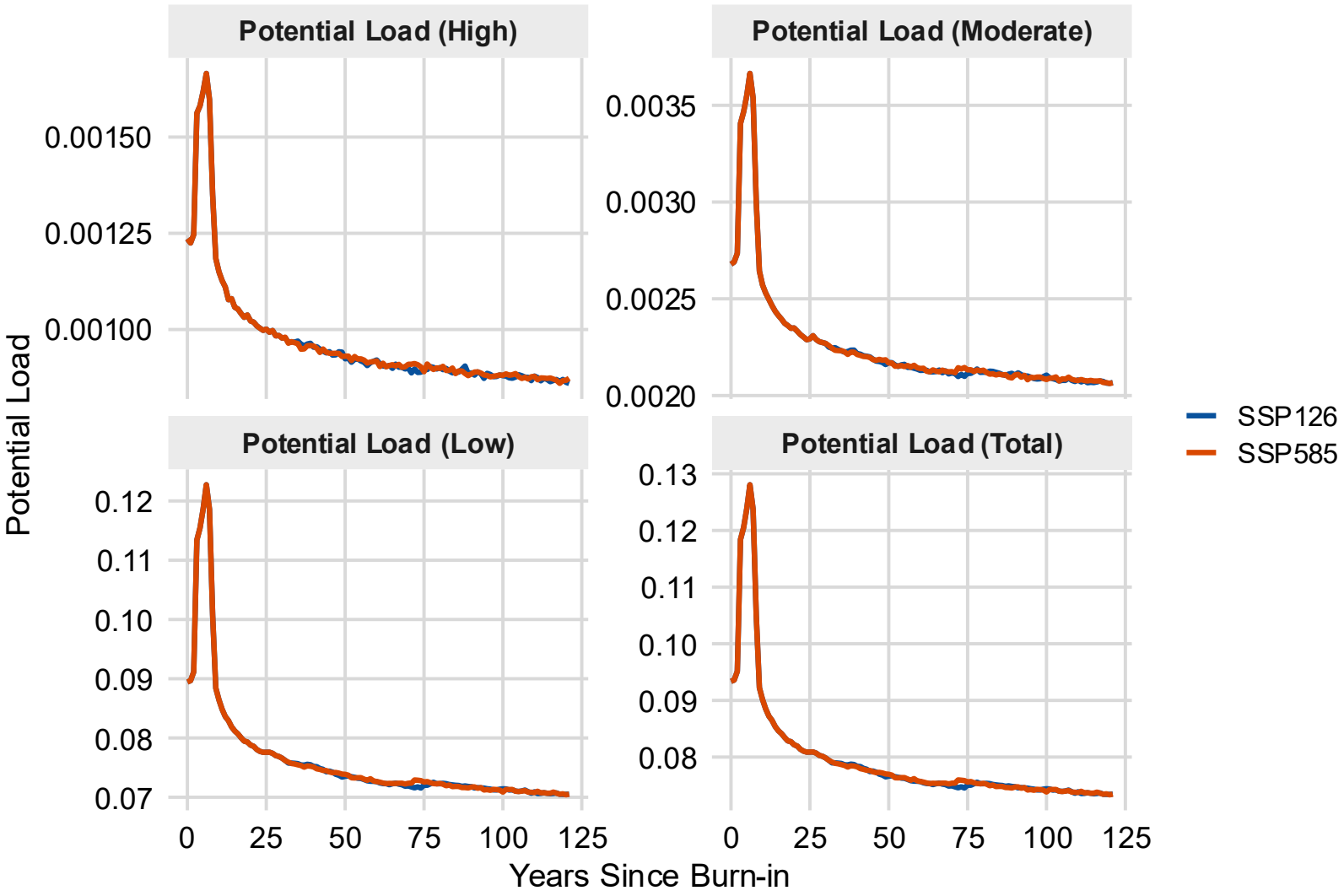

Fig. S15. Potential genomic load under climate change scenarios, partitioned by the intensity of selection coefficients (High, Moderate, Low, and Total). Climate change scenarios are color-coded as indicated in the legend. Both scenarios show an overall decreasing trend of potential genomic load throughout the simulations.

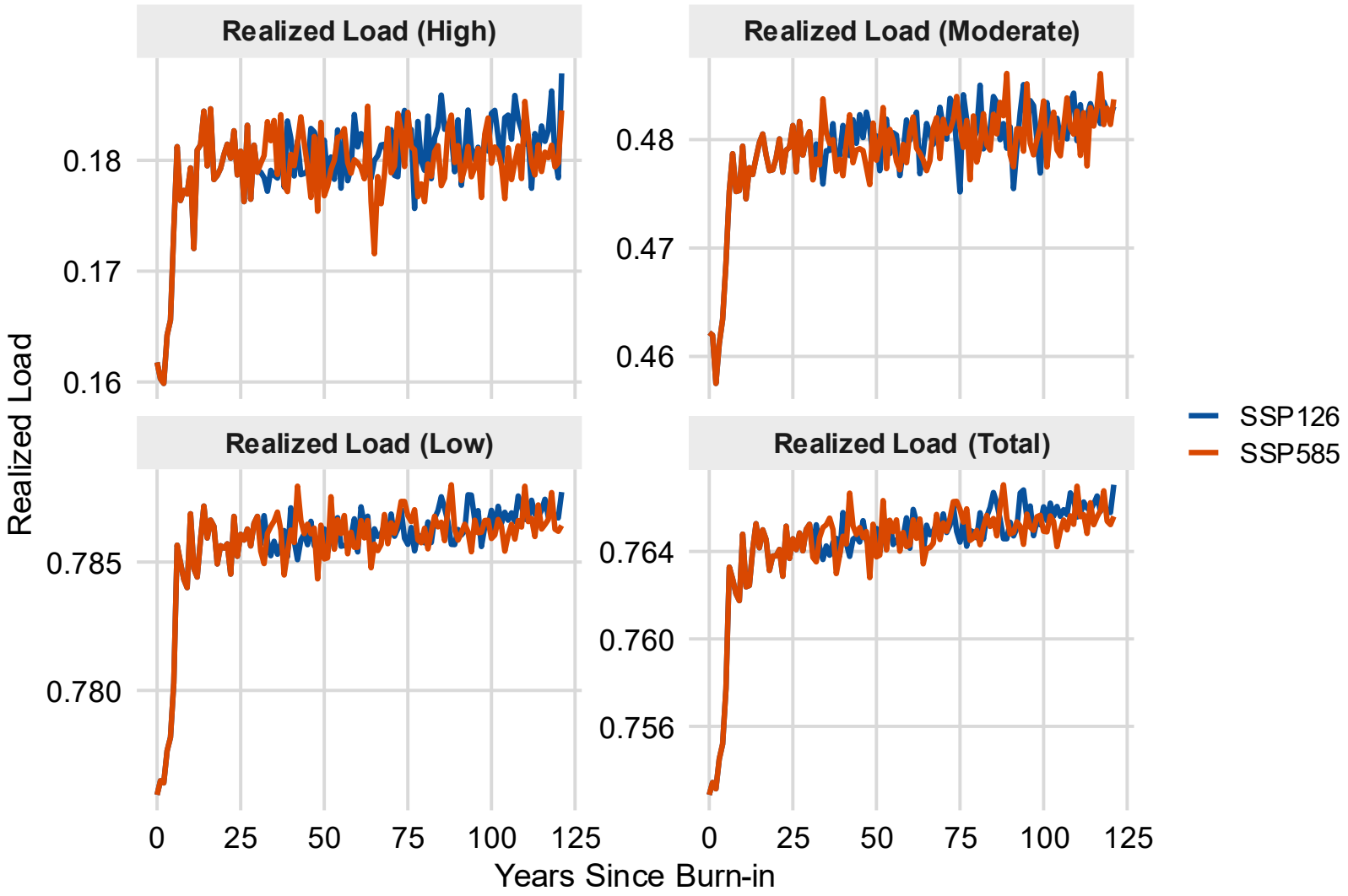

Fig. S16. Realized genomic load under climate change scenarios, partitioned by the intensity of selection coefficients (High, Moderate, Low, and Total). Climate change scenarios are color-coded as indicated in the legend. Both scenarios show an overall increasing trend of realized genomic load throughout the simulations.

Fig. S17. Fitness, heterozygosity, and population size under genetic rescue scenarios. Genetic rescue scenarios are color-coded as indicated in the legend (NM, New Mexico; SE MX, Southeast Mexico). Mean fitness did not show a consistent trend across scenarios throughout the simulations. In contrast, mean heterozygosity at all sites, at sites of empirical genome-wide variants, and at sites of empirical outlier structural variants (SVs) declined continuously over time. However, rescue scenarios involving translocations from New Mexico exhibited an initial, abrupt increase in heterozygosity following rescue events. This increase was most pronounced at outlier SV sites, likely reflecting the smaller number of variant sites. Population size increased only in the New Mexico rescue scenarios. Overall, patterns of heterozygosity and population size indicate that the introduction of either 50 or 100 individuals from New Mexico effectively rescued the recipient population, whereas the introduction of 100 individuals from Southeast Mexico primarily increased variation at outlier SV sites, which may represent putatively adaptive variation.

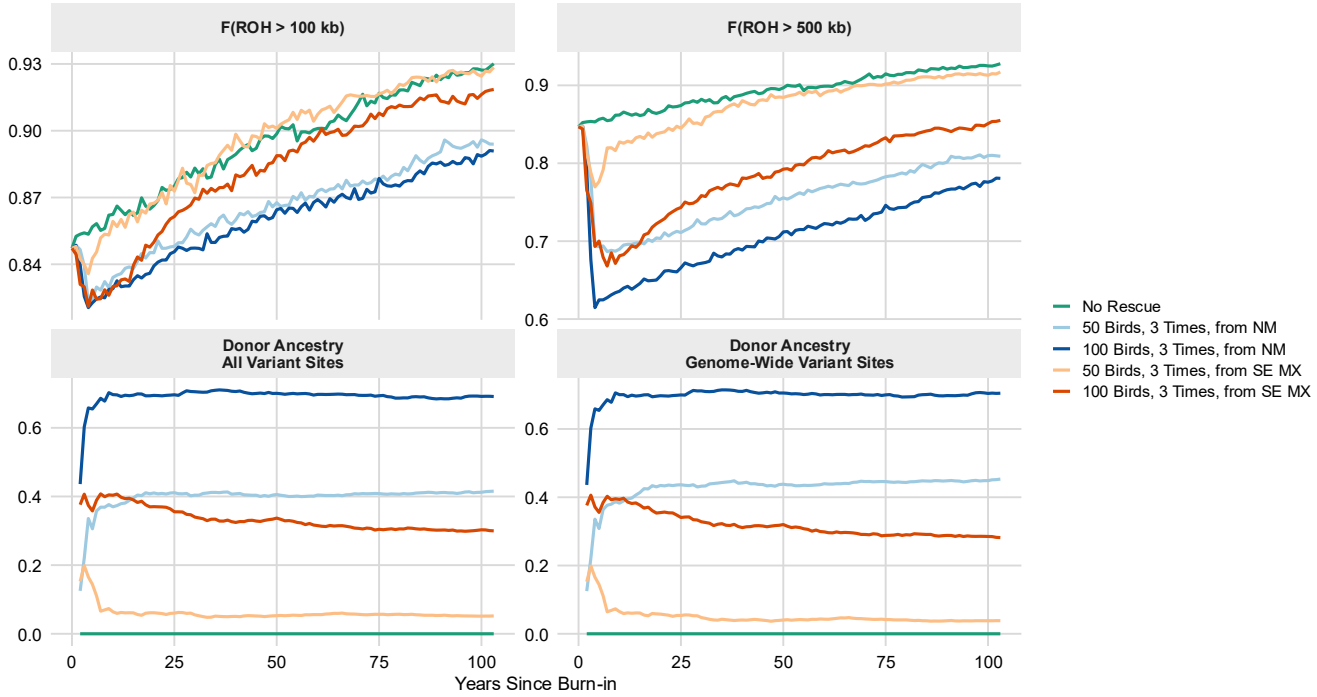

Fig. S18. Fraction of runs of homozygosity (F_{ROH}) and donor ancestry proportion in the recipient population under genetic rescue scenarios. The F_{ROH} was calculated and shown for runs longer than 100 kb and 500 kb, while donor ancestry proportion in the recipient population was quantified and shown across all sites and genome-wide variant sites. Genetic rescue scenarios are color-coded as indicated in the legend (NM, New Mexico; SE MX, Southeast Mexico). F_{ROH} showed a marked initial decrease immediately following rescue events, followed by a gradual increase over time across all scenarios and ROH length classes. Both rescue scenarios from Southeast Mexico rapidly returned to baseline (i.e., no rescue) levels for $F_{ROH} > 100$ kb, whereas the scenario with 100 Southeast Mexico individuals retained lower F_{ROH} for $F_{ROH} > 500$ kb. Rescue scenarios from New Mexico consistently maintained lower F_{ROH} than the baseline across all ROH length classes. Across all site categories, the donor ancestry proportion changed abruptly following rescue events and then remained stable over time. Translocation of 100 individuals from New Mexico resulted in the highest donor ancestry proportion (approximately 0.6–0.7), indicating potential genetic swamping in the recipient population. In contrast, introduction of 50 individuals from New Mexico or 100 individuals from Southeast Mexico resulted in intermediate proportions of approximately 0.4 and 0.3, respectively.

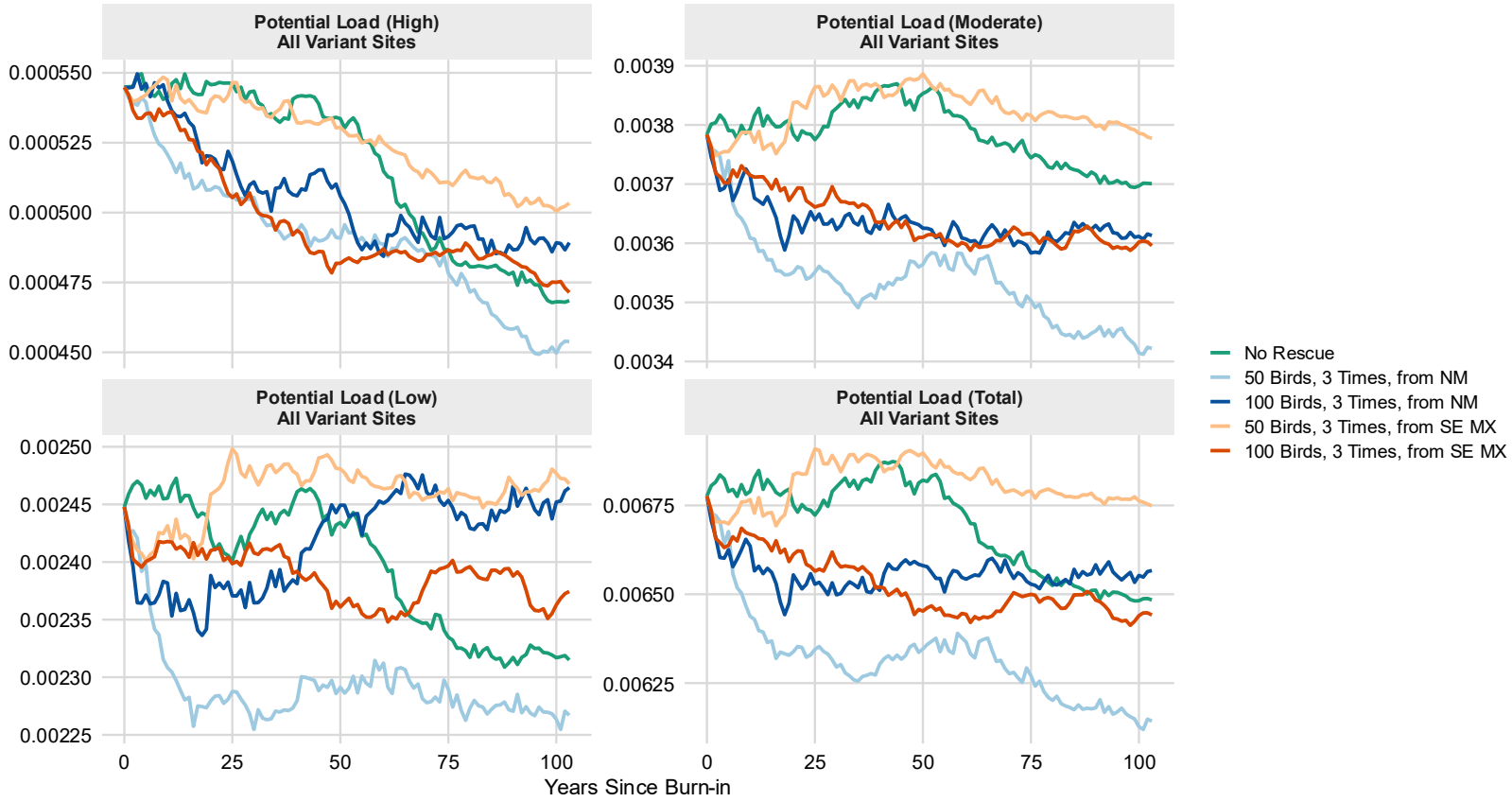

Fig. S19. Potential genomic load across all variant sites under genetic rescue scenarios, partitioned by the intensity of selection coefficients (High, Moderate, Low, and Total). Genetic rescue scenarios are color-coded as indicated in the legend (NM, New Mexico; SE MX, Southeast Mexico). Across all scenarios, potential load associated with variants under strong selection (High) consistently decreased over time. Potential load for Low and Moderate selection coefficients showed little change or a slight decrease across all scenarios, while the overall pattern of Total potential load was largely driven by variants under Moderate selection coefficients. The rescue scenario involving translocation of 50 individuals from Southeast Mexico exhibited the highest potential load for all selection coefficients. Overall, these patterns indicate that purifying selection acted in accordance with selection intensity, with the strongest reduction observed for variants under High selection coefficients.

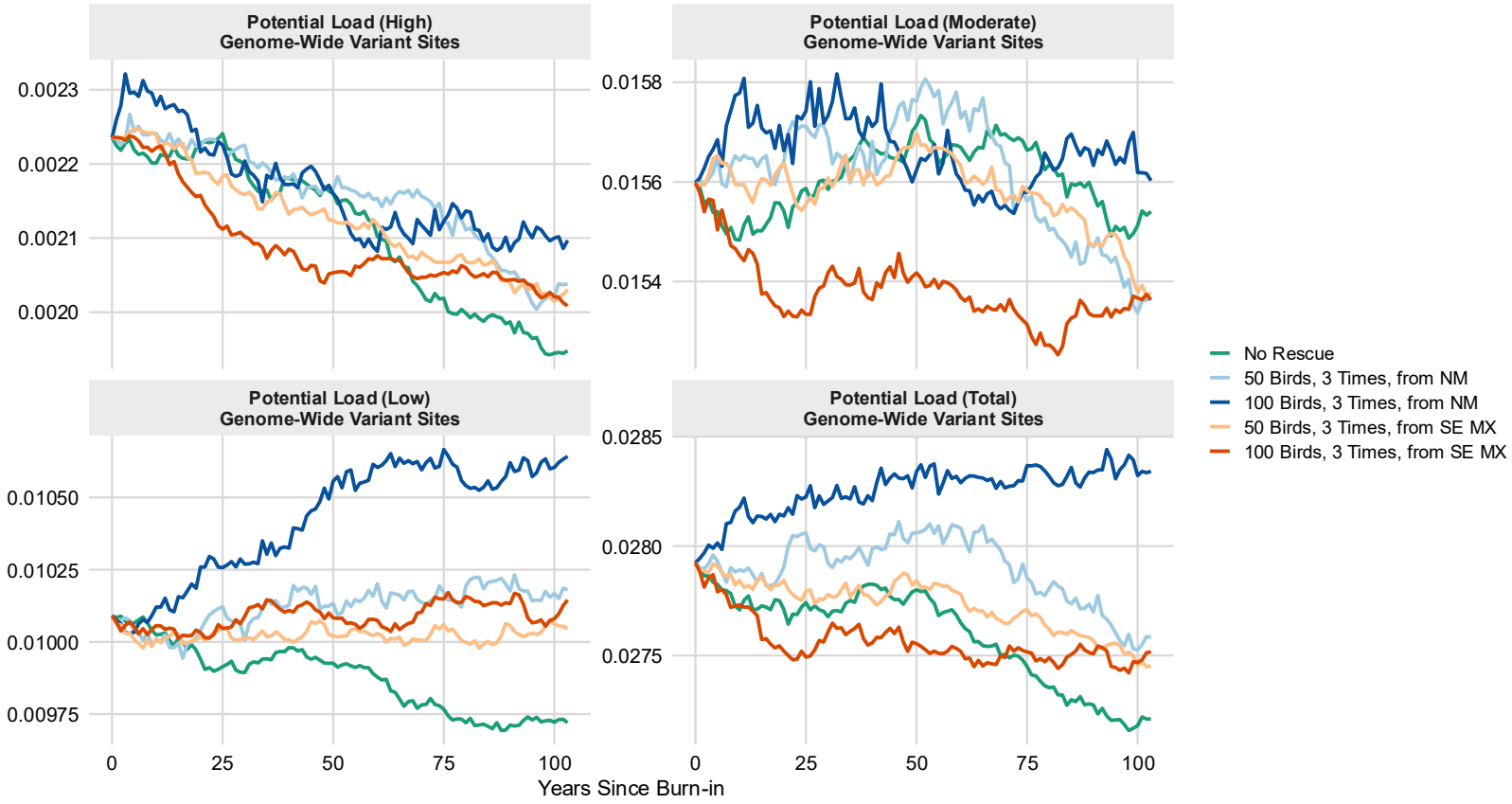

Fig. S20. Potential genomic load at empirical genome-wide variant sites under genetic rescue scenarios, partitioned by the intensity of selection coefficients (High, Moderate, Low, and Total). Genetic rescue scenarios are color-coded as indicated in the legend (NM, New Mexico; SE MX, Southeast Mexico). Across all scenarios, potential load associated with variants under strong selection (High) consistently decreased over time. Potential load for Low and Moderate selection coefficients showed little or stochastic change, while the overall pattern of Total potential load was largely driven by variants under Low selection coefficients. The rescue scenarios involving translocation of 100 individuals from New Mexico exhibited the highest potential load for all selection coefficients. Overall, these patterns indicate that purifying selection acted in accordance with selection intensity.

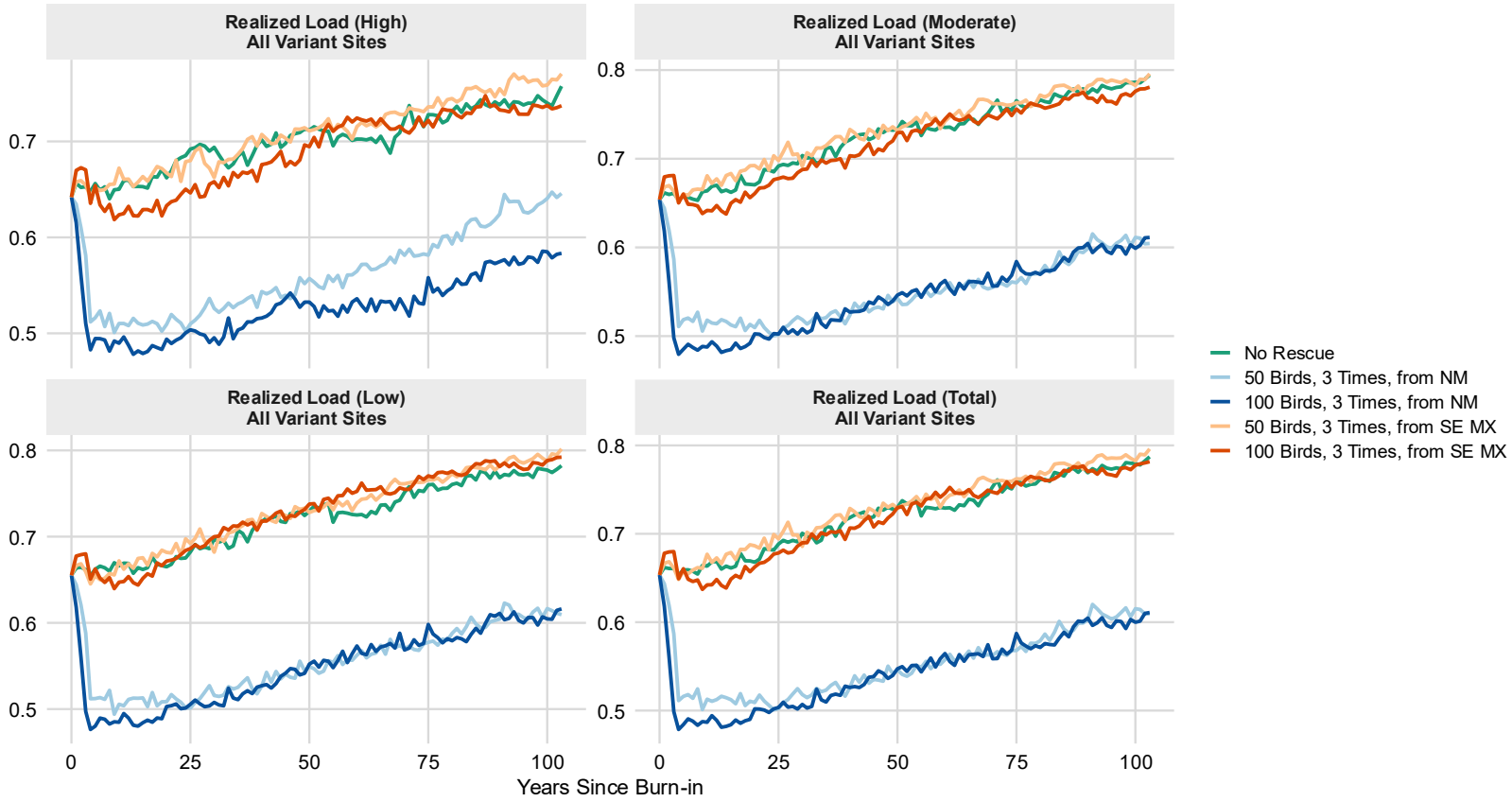

Fig. S21. Realized genomic load across all sites under genetic rescue scenarios, partitioned by the intensity of selection coefficients (High, Moderate, Low, and Total). Genetic rescue scenarios are color-coded as indicated in the legend (NM, New Mexico; SE MX, Southeast Mexico). Across all selection intensity classes, realized load increased over time in all scenarios. Rescue scenarios involving translocation of either 50 or 100 individuals from New Mexico showed an immediate and pronounced reduction in realized load following rescue events across all intensity classes. Overall, these results indicate that although realized load (i.e., the proportion of homozygous deleterious mutations) continued to increase due to the persistently small population size of the recipient population, even after rescue events, translocations from New Mexico effectively mitigated severe genetic erosion.

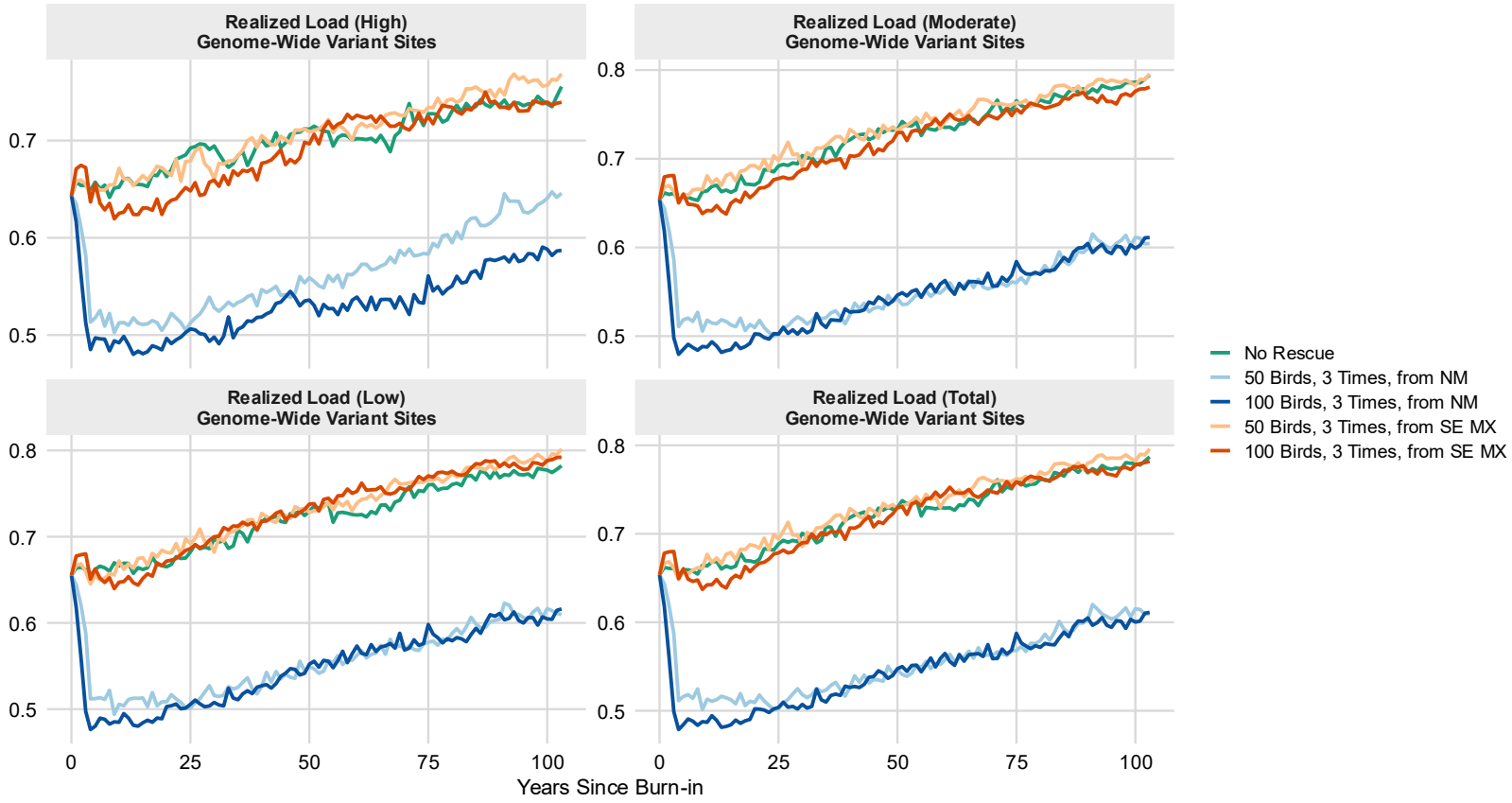

Fig. S22. Realized genomic load at empirical genome-wide variant sites under genetic rescue scenarios, partitioned by the intensity of selection coefficients (High, Moderate, Low, and Total). Genetic rescue scenarios are color-coded as indicated in the legend (NM, New Mexico; SE MX, Southeast Mexico). Across all selection intensity classes, realized load increased over time in all scenarios. Rescue scenarios involving translocation of either 50 or 100 individuals from New Mexico showed an immediate and pronounced reduction in realized load following rescue events across all intensity classes. Overall, these results indicate that although realized load (i.e., the proportion of homozygous deleterious mutations) continued to increase due to the persistently small population size of the recipient population, even after rescue events, translocations from New Mexico effectively mitigated severe genetic erosion.

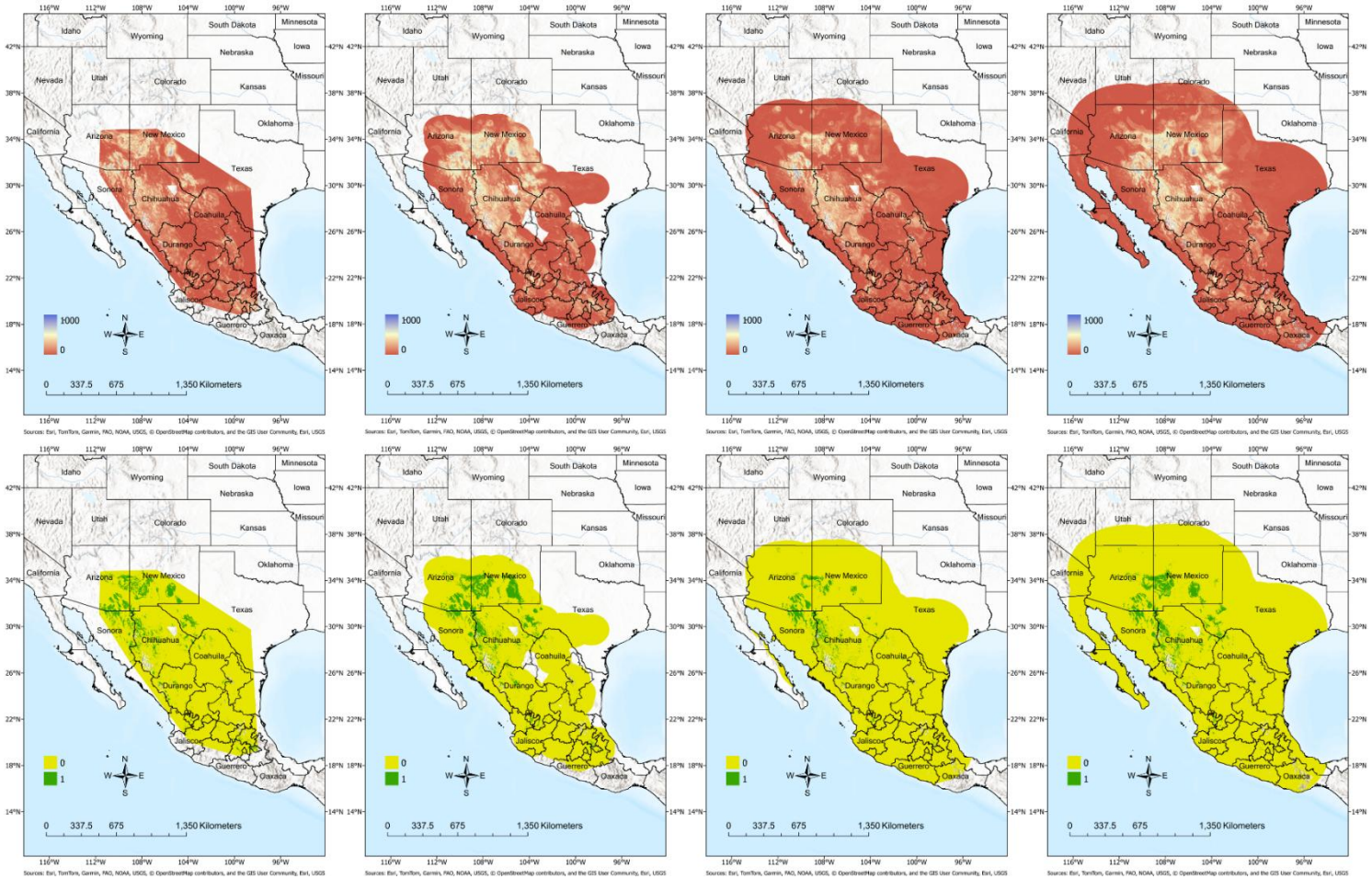

Fig. S23. Habitat suitability trend of the Montezuma quail from the present (1st column), to 2040 (2nd column), 2070 (3rd column), and 2100 (4th column) with continuous (top row) or binary (bottom row) predictions based on the most moderate climate change scenario. Blue and green indicate suitable habitats (continuous and binary, respectively), whereas red and yellow indicate unsuitable habitats. The depicted geographic range at each time period reflects the maximum possible extent of the species given its dispersal capacity up to that time. The suitable habitats fluctuated (increasing in 2040, 2100 and decreasing in 2070) and mostly remained at the northeast of the species' entire range.

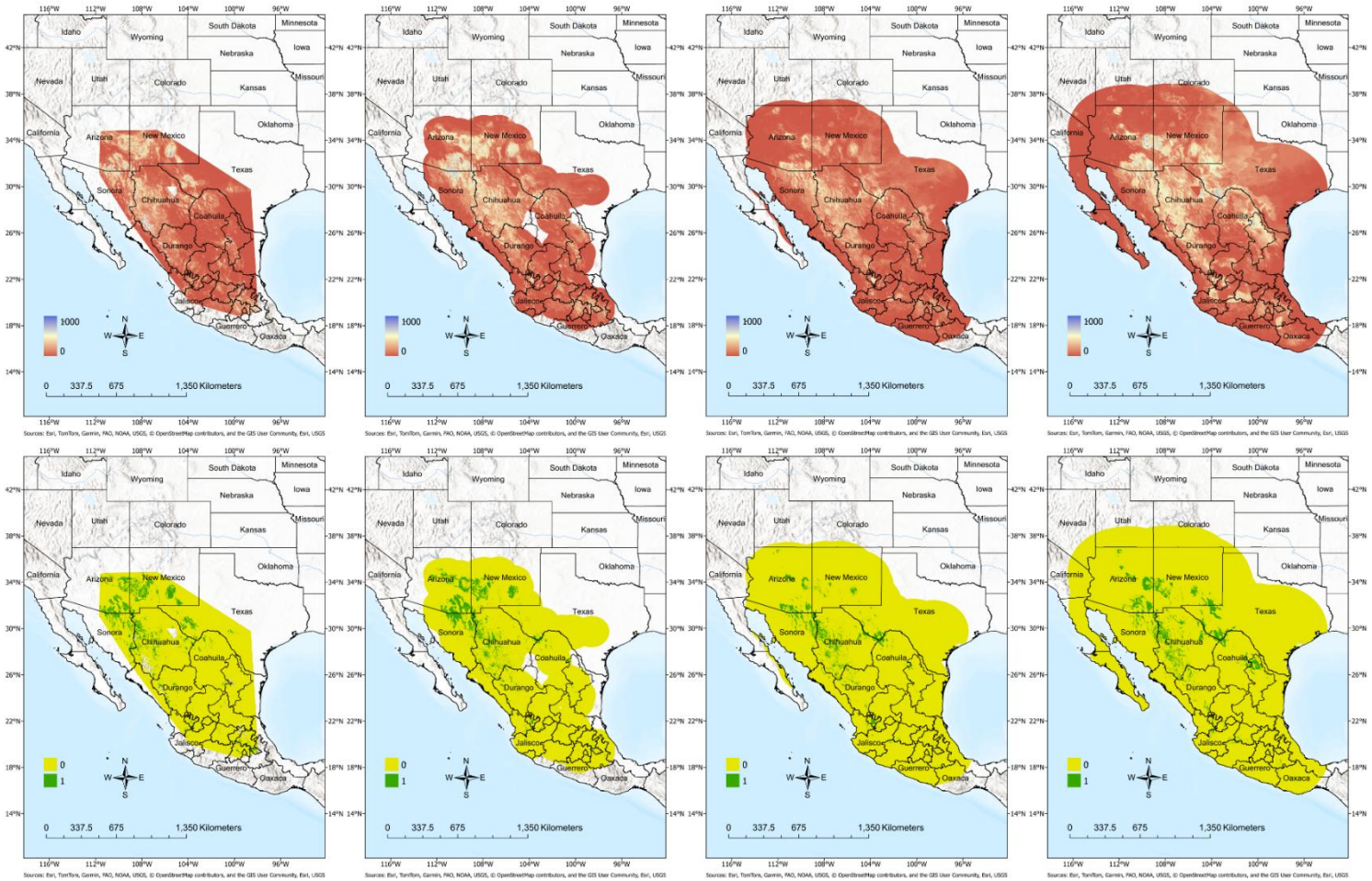

Fig. S24. Habitat suitability trend of the Montezuma quail from the present (1st column), to 2040 (2nd column), 2070 (3rd column), and 2100 (4th column) with continuous (top row) or binary (bottom row) predictions based on the most extreme climate change scenario. Blue and green indicate suitable habitats (continuous and binary, respectively), whereas red and yellow indicate unsuitable habitats. The depicted geographic range at each time period reflects the maximum possible extent of the species given its dispersal capacity up to that time. The suitable habitats fluctuated (increasing in 2040, 2100 and decreasing in 2070) and mostly remained at the northeast of the species' entire range.

Table S1. (separate file)
Montezuma quail genome assembly and annotation summary.

Table S2. (separate file)
Montezuma quail long-read assembly statistics.

Table S3. (separate file)
Summary of all and outlier structural variants.

Table S4. (separate file)
Gene ontology and KEGG enrichment analysis results for outlier structural variants, including associated genes, NCBI gene IDs, environmental predictors, and correlation coefficients.

Table S5. (separate file)
Variable importance from ecological niche modeling, including percent contribution and permutation importance, and correlated bioclimatic variables identified for each predictor.

Table S6. (separate file)
Number of genome-wide SNPs in each impact category and the proportion of deleterious variants (high- and moderate-impact) as classified by VEP, calculated for each sample and each genetic cluster.

Table S7. (separate file)
Number of genome-wide SVs in each impact category and the proportion of deleterious variants (high- and moderate-impact) as classified by VEP, calculated for each sample and each genetic cluster.

Table S8. (separate file)
Number of outlier SVs in each impact category and the proportion of deleterious variants (high- and moderate-impact) as classified by VEP, calculated for each sample and each genetic cluster.

Table S9. (separate file)
Per-individual heterozygosity, neutral and adaptive genetic cluster assignments, and final cluster designations for Montezuma quail samples.

Table S10. (separate file)

Summary of genotype-environment RDA models examining the contributions of environmental, climatic, topographic, and geographic variables to structural variant genotype variation, including model formulas, *R*-squared values, adjusted R-squared values, and model significance.

Table S11. (separate file)

Sensitivity of genotype–phenotype RDA results to the number of genotype principal components (*k*) included as covariates, showing model *p*-values, *R*-squared values, adjusted *R*-squared values, and constrained and residual ranks across $k = 1–12$.

Table S12. (separate file)

Number and proportion of genome-wide variants (SNP + SV) in each impact category as classified by VEP, calculated for each sample and each genetic cluster, with additional NM samples included for SLiM simulations.

Table S13. (separate file)

Number and proportion of outlier SV in each impact category as classified by VEP, calculated for each sample and each genetic cluster, with additional NM samples included for SLiM simulations.