Please cite this article in press as: Taylor et al., Climate-Mediated Movement of an Avian Hybrid Zone, Current Biology (2014), http://dx.doi.org/10.1016/j.cub.2014.01.069

Current Biology 24, 1–6, March 17, 2014 ©2014 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2014.01.069

Report

Climate-Mediated Movement of an Avian Hybrid Zone

Scott A. Taylor,[1,2,*] Thomas A. White,[2,3] Wesley M. Hochachka,[4] Valentina Ferretti,[1,5] Robert L. Curry,[5] and Irby Lovette[1,2]

[1]Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY 14850, USA

[2]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14850, USA

[3]Computational and Molecular Population Genetics, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland

[4]Cornell Lab of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY 14850, USA

[5]Department of Biology, Villanova University, 800 Lancaster Avenue, Villanova, PA 19085-1699, USA

Summary

The interaction between sibling species that share a zone of contact is a multifaceted relationship affected by climate change [1, 2]. Between sibling species, interactions may occur at whole-organism (direct or indirect competition) or genomic (hybridization and introgression) levels [3–5]. Tracking hybrid zone movements can provide insights about influences of environmental change on species interactions [1]. Here, we explore the extent and mechanism of movement of the contact zone between black-capped chickadees (*Poecile atricapillus*) and Carolina chickadees (*Poecile carolinensis*) at whole-organism and genomic levels. We find strong evidence that winter temperatures limit the northern extent of *P. carolinensis* by demonstrating a current-day association between the range limit of this species and minimum winter temperatures. We further show that this temperature limitation has been consistent over time because we are able to accurately hindcast the previous northern range limit under earlier climate conditions. Using genomic data, we confirm northward movement of this contact zone over the past decade and highlight temporally consistent differential—but limited—geographic introgression of alleles. Our results provide an informative example of the influence of climate change on a contact zone between sibling species.

Results

Hybrid Zone Movement and Geographic Introgression

Genomic comparisons show that the chickadee hybrid zone in southeastern Pennsylvania has moved north between historical (2000–2002) and contemporary (2010–2012) periods (Figures 1A–1C; Figure S1 available online). From the 1425 locus dataset, which consisted of 167 individuals, 75 loci showed clinal geographic variation and had cline widths less than 100 km. Of these clinal loci, 23 had significantly different cline center estimates (i.e., nonoverlapping 95% confidence intervals) between periods after Bonferroni correction (Table S1). Averaging across these 23 loci, the center of the hybrid zone shifted north by ~11.5 km over the past decade (Figures 1A and S1).

There were two genetic clusters in both periods (Table S2). Analyses with the program STRUCTURE revealed (1) genetic admixture within historical and contemporary Nolde Forest (NF) and Hawk Mountain (HM) populations and (2) the absence of genetic admixture at Great Marsh (GM), Villanova University (VU), and Tuscarora State Forest (TU) (Figure 1C). Eleven individuals showed admixture in the historical sample, which we determined by examining parental species membership coefficient proportions: proportions < 0.99 in either parental category indicate admixture (Figure 1C, using membership coefficient proportions outlined in [6] and [7]); six (56%) of these had the signature of F1 hybrids. Seven individuals showed genetic admixture in the contemporary sample; four (57%) of these possessed genetic signatures of F1 hybrids.

Geographic introgression of alleles was highly variable across the genome, with individual loci showing either consistency or inconsistency between periods (Figures 1A and 1B; Table S1). A subset of loci exhibited generally concordant cline center estimates clustered around the contact-zone center, whereas other loci showed variable extents of northward introgression (Figure 1B). Loci within these two clusters were partially consistent between time periods (Table S1). The loci exhibiting narrow cline widths and concordant centers were significantly more likely to be located on the Z chromosome than on autosomes and to be identified as interspecific F_{ST} outlier loci in previous analyses (Table S1) (S.A.T., R.L.C., I.L., T.A.W., and V.F., unpublished data).

Distribution and Climate

Independent eBird data allowed us to map contemporary and historical locations of the contact zone between *Poecile atricapillus* and *Poecile carolinensis* across a broader geographic extent than the transect of genetic sampling locations (Figures 2A and 2B) and similarly revealed the hybrid zone movement seen in the genomic data. Changes in proportions of *P. carolinensis* within the region of each sampling site further match anecdotal data from the sites (Figure 2C; Table 1).

Cold winter temperatures appear to limit the northern extent of *P. carolinensis*. The northern range limit of *P. carolinensis* is not coincident with a physical boundary or habitat shift, but instead closely aligns with the mean minimum winter temperature −7°C isotherm [8, 9]. Physiological experiments have shown that *P. atricapillus* tolerates low winter temperatures better than *P. carolinensis* does [10]. We used eBird data and climate records from the PRISM database (2011, PRISM Climate Group; map created July 6, 2013) to evaluate the relationship between mean minimum daily winter temperature ([MDWT]; December to February) and the location of the chickadee contact zone (Figures 3A and 3B). We chose to build the model using MDWT given the aforementioned evidence of a relationship between the northern range limit of *P. carolinensis* and this parameter. By considering evidence for cold temperatures limiting the northward limit of *P. carolinensis*, we believe that MDWT is the best-available temperature indicator from PRISM, reflecting both typical minimum temperatures and extreme fluctuations. Note that mean and minimum temperatures are typically correlated.

*Correspondence: sat235@cornell.edu

Please cite this article in press as: Taylor et al., Climate-Mediated Movement of an Avian Hybrid Zone, Current Biology (2014), http://dx.doi.org/10.1016/j.cub.2014.01.069

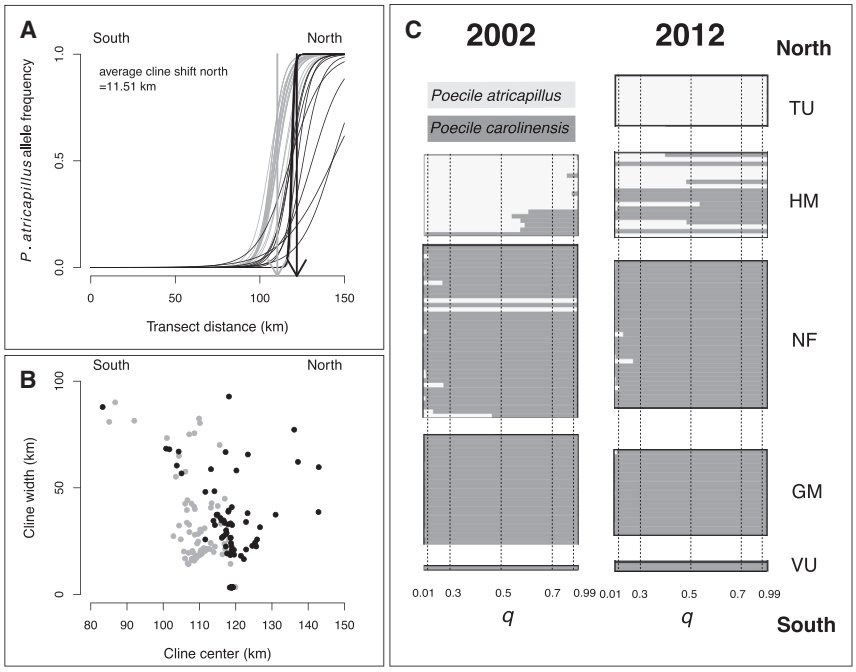


Figure 1. Genomic Evidence for Contact Zone Movement

(A) Locus-specific geographic clines depicting *P. atricapillus* allele frequencies for 23 loci with nonoverlapping 95% confidence intervals for both sampling time periods. Historical samples (2000–2002) in gray, contemporary samples (2010–2012) in black; arrows indicate mean cline center estimates.

(B) Locus-specific geographic cline centers plotted against cline widths. Contemporary samples in black, historical samples in gray.

(C) Bayesian assignment probabilities from STRUCTURE for *P. atricapillus* (light gray) and *P. carolinensis* (dark gray) at K = 2. Each horizontal line represents one individual. q = the probability of assignment to each genetic population. Dashed lines indicate threshold q values used to categorize individuals (see Results). Population acronyms as in Figure 2C.

An eBird observation site was considered to be within the contact zone if at least 5% of the surrounding area was calculated to have both *P. atricapillus* and *P. carolinensis* (Figure 3A). The probability that an eBird observation site was within the contact zone was highly correlated with MDWT (Figures 3C and 3D). This correlation, however, changed slightly from east to west across Pennsylvania: the temperatures that predicted the location of the contact zone were warmer in the interior of Pennsylvania than at the coast (Figure 3D).

Accounting for longitudinal variation in the predictive ability of temperature, the statistical model describing the contemporary location of the contact zone (from eBird) based on current MDWT (from PRISM) was accurately able to predict (hindcast) the location of the contact zone a decade ago based on temperature data from that period (AUC = 0.91). Importantly, in the region of our genetic transect in southeastern Pennsylvania, mean minimum winter temperatures have increased over the past decade: model assessment site MDWT increased by 0.76°C during this time period (PRISM Climate Group). The predictive ability of the model is not an artifact of the absence of temperature change or contact zone movement over the past decade; rather, it captures observed temporal variation and movement in the chickadee contact zone.

Discussion

Climate-Mediated Movement of the Chickadee Hybrid Zone

Climate change is causing northward movement of the chickadee hybrid zone, as seen in both our geographically focused genomic transect in southeastern Pennsylvania and more broadly in the eBird comparisons that span the past decade. Our use of genomic comparisons between sampling periods and our novel approach to contact zone modeling using the eBird database have allowed us to document consistency in observed hybrid zone movement between genomic and distribution data. We have also documented consistency between observed movement from distribution data and predicted movement from our climate model; additionally, MDWT has significant predictive power with respect to the geographic location of the contact zone. We suspect that other features of the environment that potentially ameliorate the effects of winter temperatures may influence regions closer to the coast. For example, the minimum winter temperature records we used do not describe the duration over which cold temperatures are experienced. Coastal sites may experience their lowest temperatures for a shorter duration than inland sites, given the moderating effect of the Atlantic Ocean. The rate of hybrid zone movement exhibited by the chickadee hybrid zone in southeastern Pennsylvania (~1.0 km/year) generally matches the rate of hybrid zone movement of the chickadee hybrid zone recorded in Ohio (1.0–1.6 km/year) [1, 10–13] and in other moving hybrid zones [1].

It has been hypothesized that the chickadee hybrid zone is a tension zone in which maintenance of the narrow zone width is likely caused by strong intrinsic selection against hybrids [11]. Results from our geographic cline analysis support this assertion. Multiple loci, distributed through the chickadee genome, exhibit clinal variation across the hybrid zone and have narrow widths (Figure 1A; Table S1). This pattern is potentially the result of underdominant selection (heterozygote disadvantage) against admixed genomes in hybrids. Furthermore, the increased number of clinal loci that we detected on the Z chromosome follows expectations from Haldane's rule and could reflect selection against heterogametic ZW individuals. Broadly, a pattern of increased differentiation of the sex chromosomes compared to the autosomes is common in birds and other organisms [14, 15]. Other explanations for this pattern include smaller effective population size of sex chromosomes and lower recombination rates, potential for a higher proportion of infertility alleles on sex chromosomes [16], meiotic drive [17], and Z-maternal interactions [18]. Differential response of the parental species to climate change (i.e., expansion of *P. carolinensis* northward), rather than differential survival of hybrids, is likely responsible for the rapid northward shift in hybrid zone location. The predictive ability of our model indicates clearly that climate is playing a role in the northward movement of this hybrid zone.

Northward movement of the chickadee hybrid zone may be facilitated by mate choice and superior competitive ability of

Please cite this article in press as: Taylor et al., Climate-Mediated Movement of an Avian Hybrid Zone, Current Biology (2014), http://dx.doi.org/10.1016/j.cub.2014.01.069

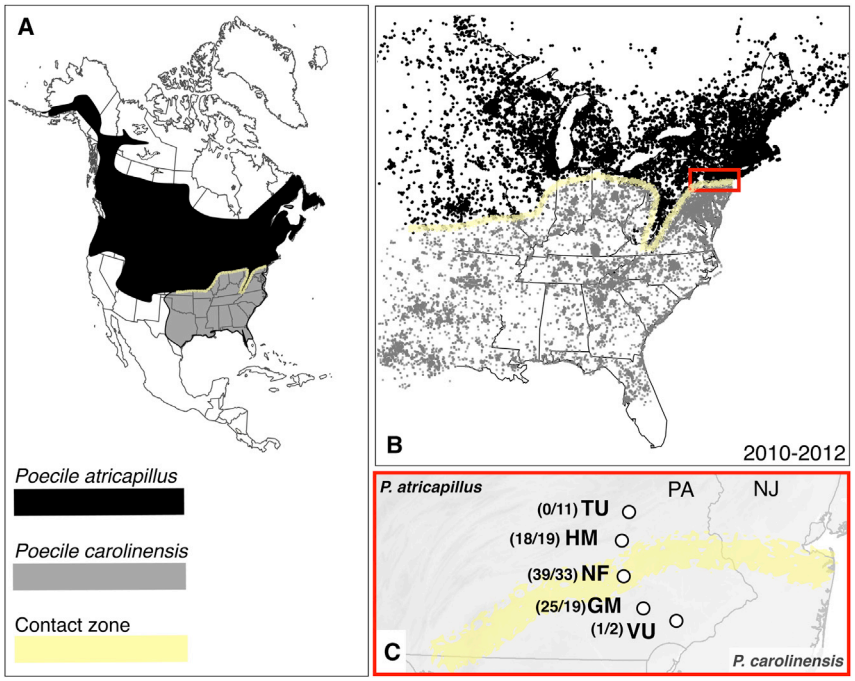


Figure 2. *P. atricapillus* and *P. carolinensis* Distributions, eBird Localities, and Study Sites

(A) Approximate ranges of *P. atricapillus* and *P. carolinensis* and contact zone.

(B) Cumulative distribution of eBird reports for *P. atricapillus* (black), *P. carolinensis* (gray), or both species/hybrids (yellow) in eastern North America during breeding (May to June) in 2010–2012.

(C) Approximate location of contact zone (yellow) and sampling transect in southeastern Pennsylvania. TU = Tuscarora State Forest (40.80N; −76.03W), HM = Hawk Mountain (40.65N; −76.00W), NF = Nolde Forest (40.28N; −75.96W), GM = Great Marsh (40.14N; −75.74W), VU = Villanova University campus (40.04N; −75.34W). Number of samples per site indicated in parentheses (2000–2002 or 2010–2012).

P. carolinensis males [12, 13]. The direction of movement of the chickadee hybrid zone we recorded matches predictions based on mating preferences of *P. atricapillus* females, which preferentially seek out extra pair copulations with *carolinensis*-like males [13], potentially because *carolinensis* tends to be dominant in interspecific male competitions [12]. We currently lack detailed data on differential fertility and/or survival of F1 hybrids and backcrosses that could explain the hybrid zone movement we detected without an influence of climate. However, hatching success of hybrid offspring is significantly lower than either parental species in southeastern Pennsylvania (R.C., unpublished data), like it is in Ohio [11], and multiple allelic clines indicate that selection against hybrids is strong, potentially due to underdominance. The high proportion of hybrids that are F1 and the low number of backcrossed hybrids in our data also suggest that the fitness of hybrid chickadees is low. In the absence of hybridization, we may expect the same directional movement of the hybrid zone, given that *P. carolinensis* appear to be dominant in interspecific interactions, which may aid in territory acquisition at the northern edge of their range [12].

Investigations into physiological tolerance differences between *P. carolinensis* and *P. atricapillus* suggest that *P. atricapillus* are better adapted to colder winter temperatures (to a large extent as a function of body size) and that increases in winter temperatures may facilitate northward expansion of *P. carolinensis* [10]. Our results support this interpretation and indicate that similar responses to climate warming are occurring in discrete regions of the contact zone. Hybrid individuals may be particularly sensitive to winter temperature minima, given their admixed genome and the potential for mitonuclear incompatibilities [10]. Concomitant strong selection against hybrids and increased dispersal of *P. carolinensis* into the contact zone in response to climate change would produce the pattern of a rapid northward movement that we see in both whole-organism and genomic data sets.

Table 1. Proportion of *P. carolinensis* Reported to eBird at Sites Surrounding Genomic Sampling Locations in Both Historic and Contemporary Time Periods

Site	2001	2011
Tuscarora State Forest	0	0
Hawk Mountain	0	0.09
Nolde Forest	0.86	0.91
Great Marsh	1.00	0.99
Villanova University campus	0.98	1.00

Hybrid Zone Movement and Climate

In a recent synthesis, 23 hybrid zones for which primary literature exists were shown to have moved at rates between 0.02 and 5.8 km/year over varying timescales; however, the underlying causes of hybrid zone movement were often unknown or anecdotal [1]. The fastest-moving hybrid zones involved species with high dispersal ability, including butterflies [19], birds [20–22], and invasive ants [23, 24], in hybrid zones thought or known to be tension zones [1, 25]. Similarly, 12 of 39 avian hybrid zones have moved within recorded history at varying rates and for multiple reasons [14]. When zones that are increasing steadily in width (unlike the chickadee hybrid zone) are excluded, 9 of 15 zones in which shifts could have been detected from repeated surveys show movement [14]. The mean rate of movement of these hybrid zones (~1 km/year) matches what we report for the chickadee hybrid zone, and the majority of the zones that show movement without concomitant broadening are moving along north-to-south axes [14]. Climate change was implicated as a causal factor for hybrid zone movement in 2 of 23 overall cases of hybrid zone movement and in 4 of 12 cases of moving avian hybrid zones [1, 14]. Notably, the well-studied carrion crow (*Corvus corone*)/hooded crow (*Corvus cornix*) hybrid zone is only moving where the line of contact runs east to west (allowing north-to-south movement) in Denmark and Scotland. This movement, which may be a response to climate change, is similar to what we report here [1, 14].

Linkages between climate change and hybrid zone movement have been explored rigorously in only a handful of cases. Most recently, climate change and hybrid zone movement

Please cite this article in press as: Taylor et al., Climate-Mediated Movement of an Avian Hybrid Zone, Current Biology (2014), http://dx.doi.org/10.1016/j.cub.2014.01.069

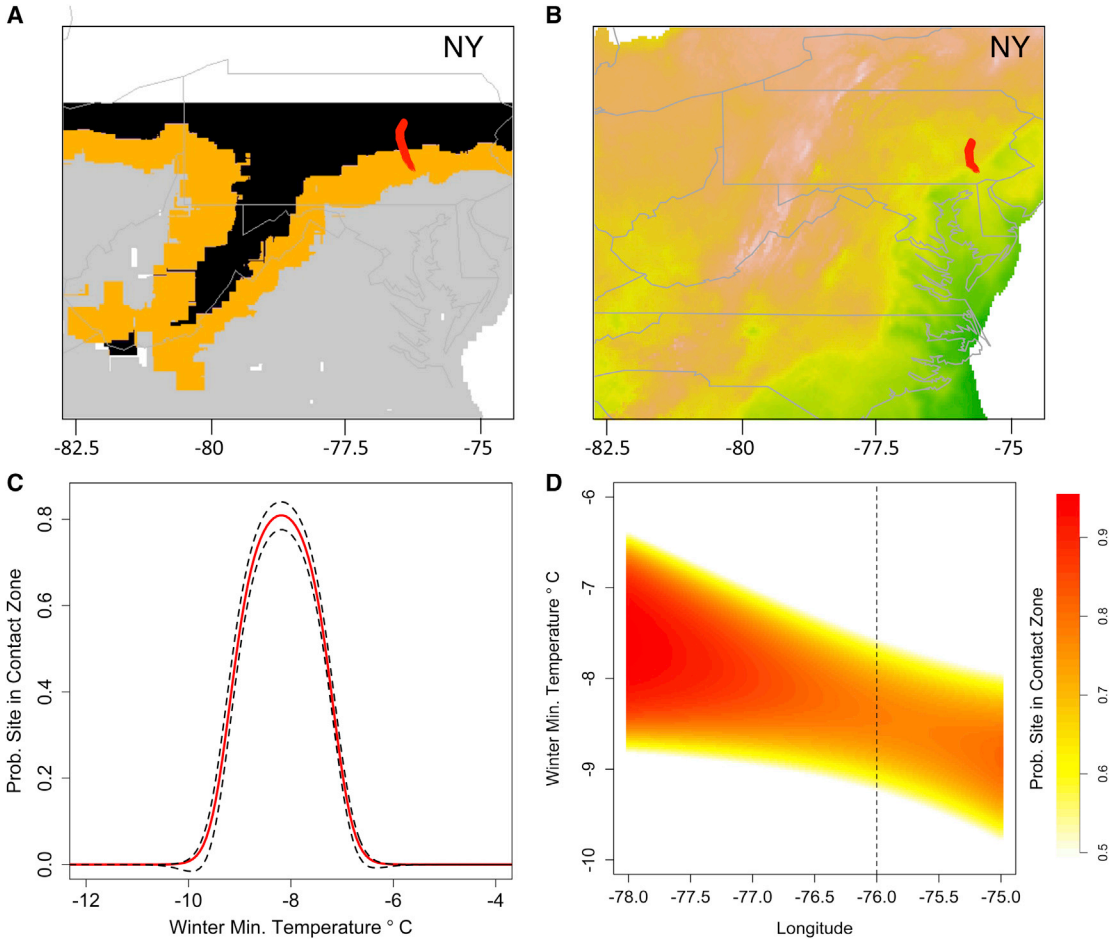


Figure 3. Association between Contact Zone Position and Temperature

(A) Chickadee contact zone mapped from eBird reports from 2010–2012 showing *P. atricapillus* (black shading), *P. carolinensis* (gray shading), and contact zone (minimum 5% of sites predicted to have both species, orange). Red line denotes approximate location of genomic sampling transect.

(B) Mean daily minimum winter temperature in 2011. Red line denotes approximate location of genomic sampling transect.

(C) Quadratic logistic regression of mean minimum daily winter temperature (December 2009 to February 2012) and contact zone position estimated from spring eBird data (2010–2012).

(D) Heat map of complex relationship between predictive ability of daily minimum winter temperature and longitude for contact zone position.

were investigated in a western European avian hybrid zone between migratory melodius warblers (*Hippolais polyglotta*) and icterine warblers (*Hippolais icterina*) using species distribution modeling [2]. Engler et al. sought to determine the relative influence of species interactions and climate on hybrid zone movement, concluding that biotic interactions (i.e., competition with *H. icterina*) are currently limiting range expansion of *H. polyglotta* northward [2]. These biotic limitations may impede the ability of *H. polyglotta*—the apparently less competitive species—to respond to climate change. It seems unlikely that biotic interactions will impede movement of the chickadee hybrid zone in response to climate change, given that the southern species, *P. carolinensis*, appears to dominate interspecific interactions [12]. In fact, mating preferences may even facilitate movement as previously discussed [13]. Not surprisingly, this comparison highlights that the influence of climate on hybrid zone movement will be complex and hard to predict, always influenced by species interactions and ecology. However, in both situations, climate change is having an impact, direct or indirect, on species interactions.

General Conclusions

We document an example of climate-mediated movement of a North American avian hybrid zone using a combination of temporal genomic, distributional, and climatic sampling, making use of the world's largest citizen science database. The rate of hybrid zone movement we report is comparable to movement rate estimates for other hybrid zones, and our exploration of the association of the contact zone with mean minimum winter temperature provides the first robust evidence that climate change is influencing the movement of this hybrid zone. We encourage others to harness the power of comprehensive analyses and data sets for our understanding of the influence of environmental change on speciation and species interactions.

Experimental Procedures

Sample Collection and Preparation

Blood samples were collected as described in [13]. We chose samples from two periods that were 10 years apart (2000–2002 and 2010–2012). Within each period we selected unrelated individuals, with the goal of having as even a sampling as possible across available sampling locations in each

period. In 2000–2002, samples were only available from four locations (Figure 2C). An additional location, TU, was added to the sampling regime in 2006 to ensure that one sampling site remained ahead of the moving hybrid zone. This site is included in the 2010–2012 transect, with five sampling points (Figure 2C).

DNA was extracted from all samples using DNeasy extraction kits (QIAGEN) and standard blood extraction protocols, eluted in water, and concentrated using a vacuum centrifuge. Blood samples are archived at Villanova, and DNA extractions are archived at the Cornell Lab of Ornithology. Villanova University's IACUC approved protocols for all field methods.

Genomic Data

Genotyping-by-sequencing (GBS) libraries were prepared and analyzed at the Institute for Genomic Diversity (IGD) at Cornell [26], using the enzyme PstI for digestion and creating a library with 96 unique barcodes. GBS libraries were sequenced on two lanes of an Illumin HiSeq 2000 (100 base pairs [bp], single-end) at the Cornell University Life Sciences Core Laboratories Center. GBS data were processed as in [27]. See Supplemental Experimental Procedures for data filtering details.

Admixture Analyses

To examine temporal changes in population admixture, we used STRUCTURE version 2.3.1, analyzing each time period separately [28]. We present details of the STRUCTURE analyses in the Supplemental Experimental Procedures.

Geographic Cline Fitting and Concordance

To quantify hybrid zone movement, we used a geographic cline approach. Details can be found in the Supplemental Experimental Procedures.

Distribution Modeling and Climate Association

To map the distribution of the contact zone between *P. atricapillus* and *P. carolinensis*, we used records from the eBird database [29]. We looked for evidence that winter temperature determines the contact zone (Figure 3A) by using information on the contact zone derived from our eBird distribution map (as explained in the Supplemental Experimental Procedures) and information on MDWT from PRISM. We provide details of our distribution modeling and climate association approaches in the Supplemental Experimental Procedures. We conducted all analyses of eBird data using the R statistical language [30], with the supplemental spatial analysis package sp [31] for manipulation of spatial data, the generalized linear model (GLM) function for conducting logistic regression analysis, and the supplemental PresenceAbsence package [32] for calculating AUC statistics.

Accession Numbers

The Dryad DOI for the SNP data reported in this paper is 10.5061/dryad.7gg47.

Supplemental Information

Supplemental Information includes Supplemental Experimental Procedures, one figure, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2014.01.069.

Acknowledgments

We thank the Fuller Lab of Evolutionary Biology at the Cornell Lab of Ornithology and Villanova University for funding; Lovette laboratory group, R.G. Harrison and laboratory group, M.S. Webster and laboratory group, E.L. Larson, N.A. Mason, G. Bradburd, and M.G. Weber for discussion; A. Talaba, L. Stenzler, S. Mitchell, and C. Acharya for laboratory contributions; many Villanova students for fieldwork; and Villanova University, Nature Conservancy (Pennsylvania), Hawk Mountain Sanctuary Association, and Pennsylvania Department of Conservation and Natural Resources (DCNR) for site access.

Received: December 16, 2013
Revised: January 27, 2014
Accepted: January 29, 2014
Published: March 6, 2014

References

1. Buggs, R.J.A. (2007). Empirical study of hybrid zone movement. Heredity (Edinb) *99*, 301–312.
2. Engler, J.O., Rödder, D., Elle, O., Hochkirch, A., and Secondi, J. (2013). Species distribution models contribute to determine the effect of climate and interspecific interactions in moving hybrid zones. J. Evol. Biol. *26*, 2487–2496.
3. Rand, D.M., and Harrison, R.G. (1989). Ecological genetics of a mosaic hybrid zone: mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. Evolution *43*, 432–449.
4. Harrison, R.G. (1993). Hybrid Zones and the Evolutionary Process (Oxford: Oxford University Press).
5. Harrison, R.G. (1990). Hybrid zones: windows on evolutionary process. In Oxford Surveys in Evolutionary Biology, *Volume 7*, D. Futuyma and J. Antonovics, eds. (Oxford: Oxford University Press), pp. 69–128.
6. Aboim, M.A., Mavárez, J., Bernatchez, L., and Coelho, M.M. (2010). Introgressive hybridization between two Iberian endemic cyprinid fish: a comparison between two independent hybrid zones. J. Evol. Biol. *23*, 817–828.
7. Taylor, S.A., Anderson, D.J., Zavalaga, C.B., and Friesen, V.L. (2012). Evidence for strong assortative mating, limited gene flow, and strong differentiation across the blue-footed/Peruvian booby hybrid zone in northern Peru. J. Avian Biol. *43*, 311–324.
8. Root, T. (1988). Atlas of Wintering North American Birds (Chicago: University of Chicago Press).
9. Repasky, R.R. (1991). Temperature and the northern distributions of wintering birds. Ecology *72*, 2274–2285.
10. Olson, J.R., Cooper, S.J., Swanson, D.L., Braun, M.J., and Williams, J.B. (2010). The relationship of metabolic performance and distribution in black-capped and Carolina chickadees. Physiol. Biochem. Zool. *83*, 263–275.
11. Bronson, C.L., Grubb, T.C., Jr., and Braun, M.J. (2003). A test of the endogenous and exogenous selection hypotheses for the maintenance of a narrow avian hybrid zone. Evolution *57*, 630–637.
12. Bronson, C.L., Grubb, T.C., Jr., Sattler, G.D., and Braun, M.J. (2003). Mate preference: a possible causal mechanism for a moving hybrid zone. Anim. Behav. *65*, 489–500.
13. Reudink, M.W., Mech, S.G., and Curry, R.L. (2006). Extrapair paternity and mate choice in a chickadee hybrid zone. Behav. Ecol. *17*, 56–62.
14. Price, T. (2008). Speciation in Birds (Greenwood Village, CO: Roberts and Company).
15. Payseur, B.A., Krenz, J.G., and Nachman, M.W. (2004). Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. Evolution *58*, 2064–2078.
16. Charlesworth, B., Coyne, J.A., and Barton, N.H. (1987). The relative rates of evolution of sex chromosomes and autosomes. Am. Nat. *130*, 113–146.
17. Frank, S.A. (1991). Divergence of meiotic drive suppression systems as an explanation for sex-biased hybrid sterility and inviability. Evolution *45*, 262–267.
18. Badyaev, A.V., Hill, G.E., and Beck, M.L. (2003). Interaction between maternal effects: onset of incubation and offspring sex in two populations of a passerine bird. Oecologia *135*, 386–390.
19. Blum, M.J. (2002). Rapid movement of a *Heliconius* hybrid zone: evidence for phase III of Wright's shifting balance theory? Evolution *56*, 1992–1998.
20. Secondi, J., Faivre, B., and Bensch, S. (2006). Spreading introgression in the wake of a moving contact zone. Mol. Ecol. *15*, 2463–2475.
21. Secondi, J., Bretagnolle, V., Compahnon, C., and Faivre, B. (2003). Species-specific song convergence in a moving hybrid zone between two passerines. Biol. J. Linn. Soc. Lond. *80*, 507–517.
22. Reullier, J., Pérez-Tris, J., Bensch, S., and Secondi, J. (2006). Diversity, distribution and exchange of blood parasites meeting at an avian moving contact zone. Mol. Ecol. *15*, 753–763.
23. Shoemaker, D., Ross, K.G., and Arnold, M.L. (1996). Genetic structure and evolution of a fire ant hybrid zone. Evolution *50*, 1958–1976.
24. Goodisman, M., Shoemaker, D., and Asmussen, M.A. (1998). Cytonuclear theory for haplodiploid species and X-linked genes. II. Stepping-stone models of gene flow and application to a fire ant hybrid zone. Evolution *52*, 1423–1440.
25. Barton, N.H., and Hewitt, G.M. (1989). Adaptation, speciation and hybrid zones. Nature *341*, 497–503.

26. Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE *6*, e19379. http://dx.doi.org/10.1371/journal.pone.0019379.
27. White, T.A., Perkins, S.E., Heckel, G., and Searle, J.B. (2013). Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. Mol. Ecol. *22*, 2971–2985.
28. Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics *155*, 945–959.
29. Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., and Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. Biol. Conserv. *142*, 2282–2292.
30. R Development Core Team (2013). R: A language and environment for statistical computing (Vienna: R Foundation for Statistical Computing).
31. Bivand, R.S., Pebesma, E., and Gomez-Rubio, V. (2013). Applied Spatial Data Analysis with R, Second Edition (New York: Springer).
32. Freeman, E.A., and Moisen, G. (2008). PresenceAbsence: an R package for presence-absence model analysis. J. Stat. Softw. *23*, 1–31.

Current Biology, Volume *24*

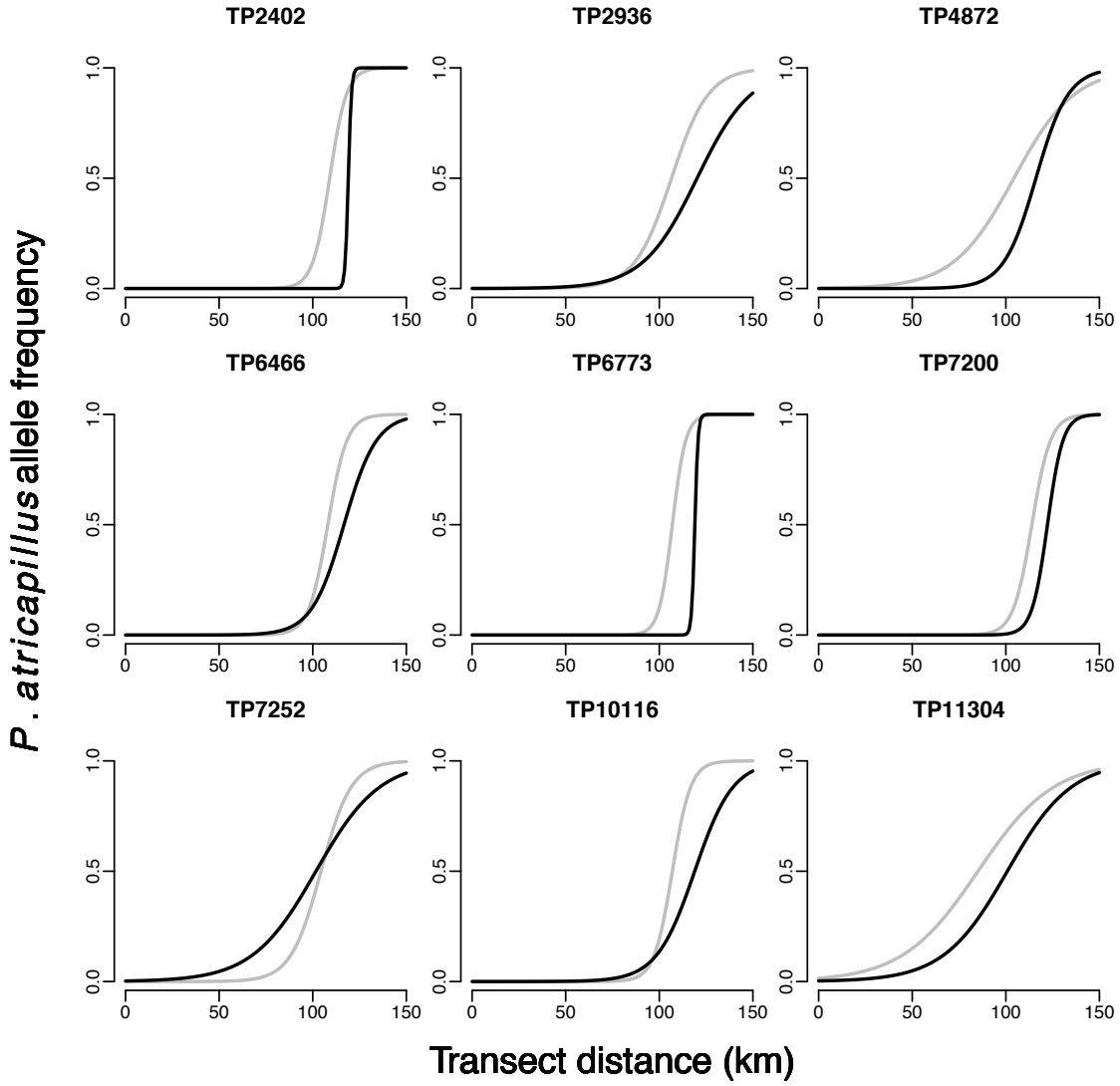
Supplemental Information

Climate-Mediated Movement of an Avian Hybrid Zone

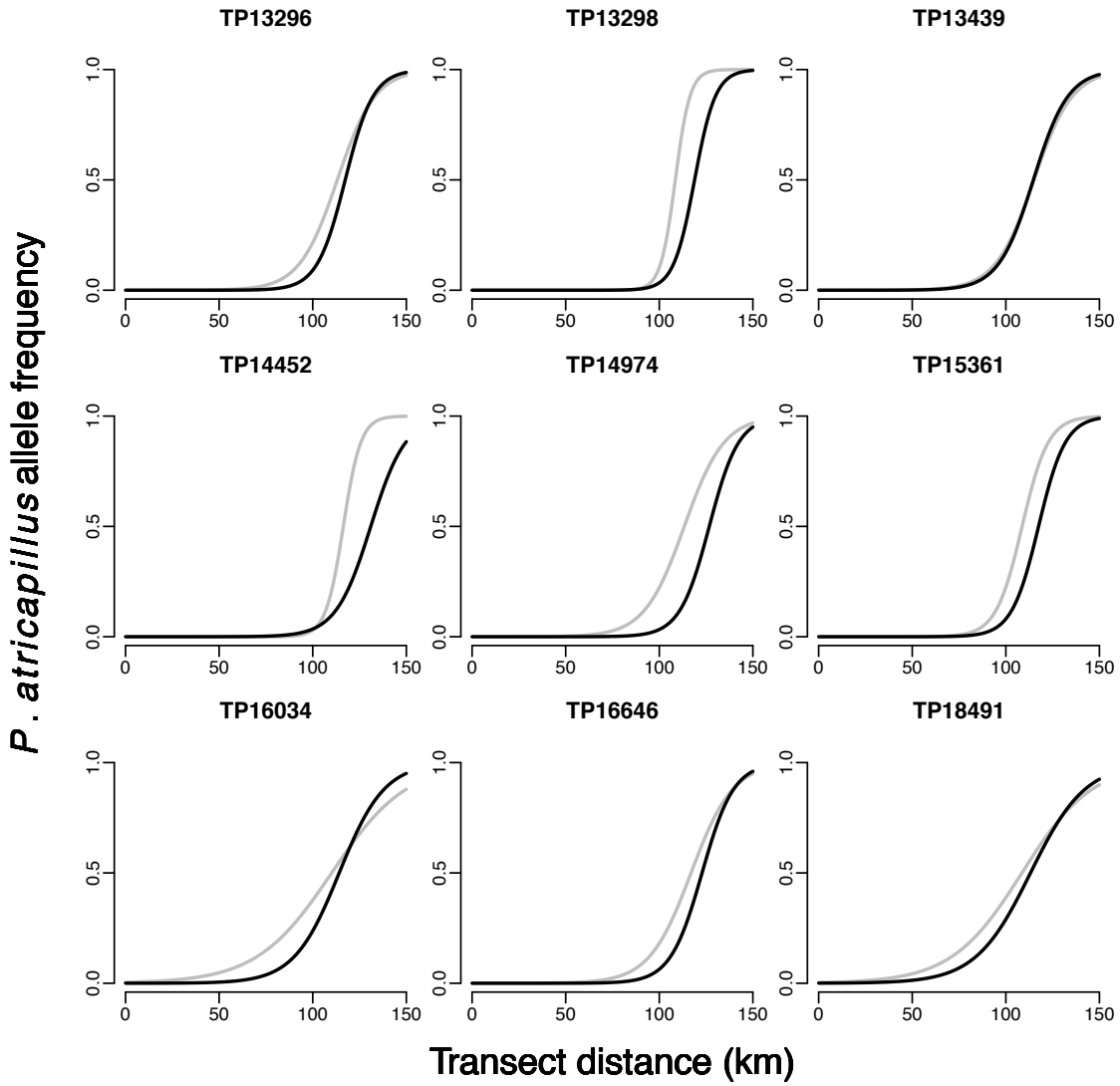
Scott A. Taylor, Thomas A. White, Wesley M. Hochachka, Valentina Ferretti, Robert L. Curry, and Irby Lovette

Supplemental Information

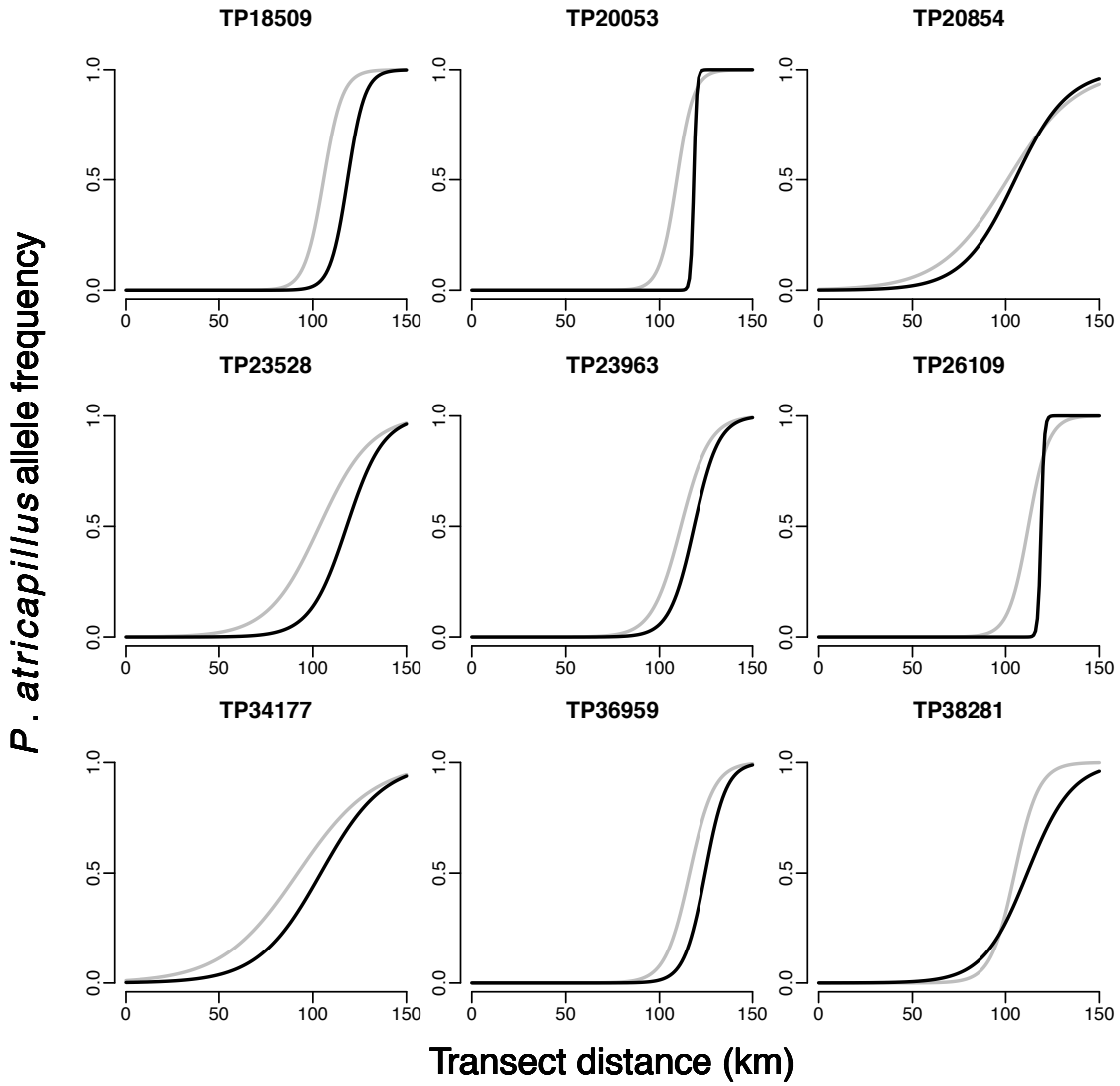
Figure S1, related to Figure 1A and B. Locus-specific geographic clines depicting *P. atricapillus* allele frequencies for the 75 loci exhibiting clinal variation.



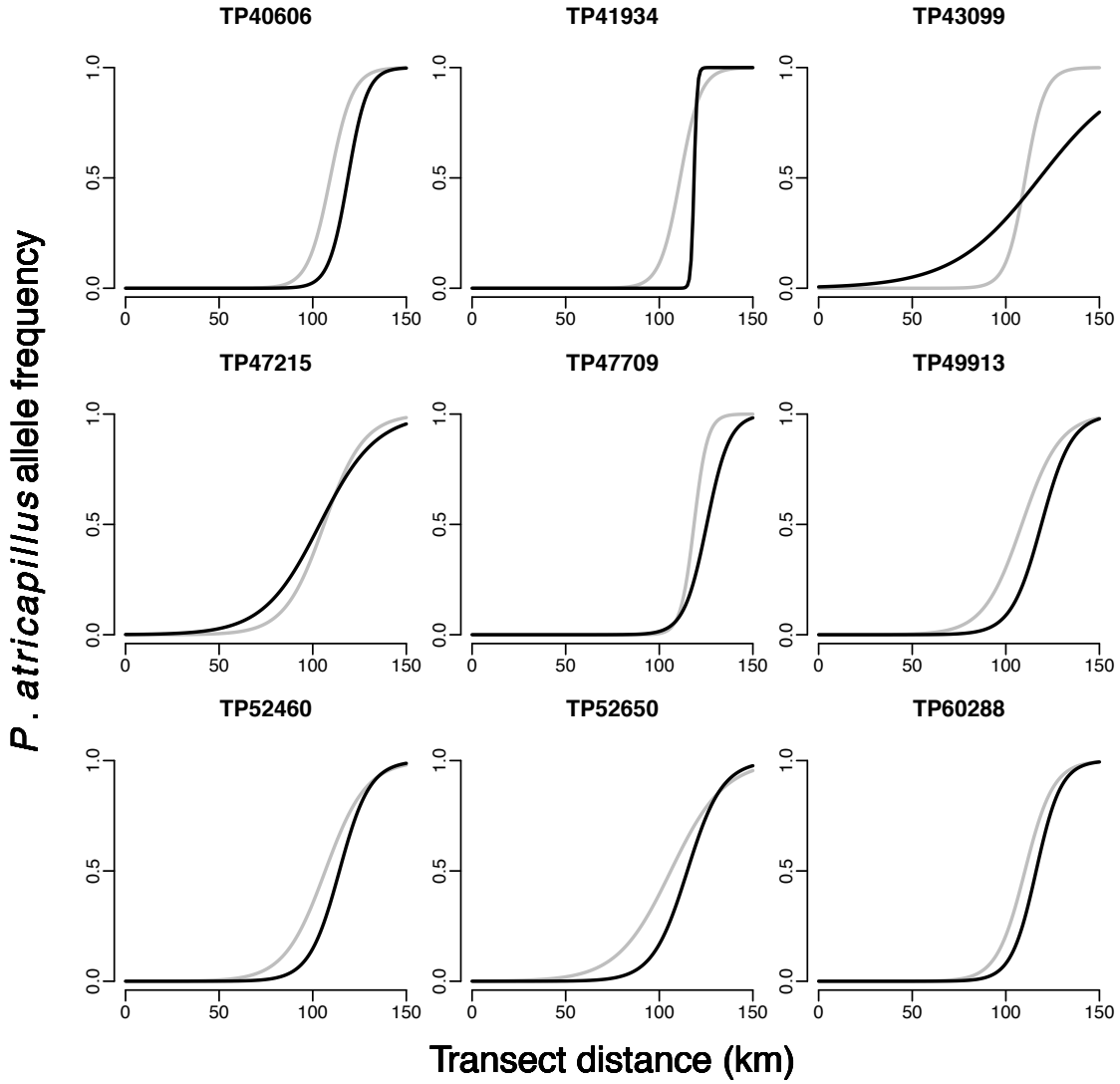
Supplemental Information



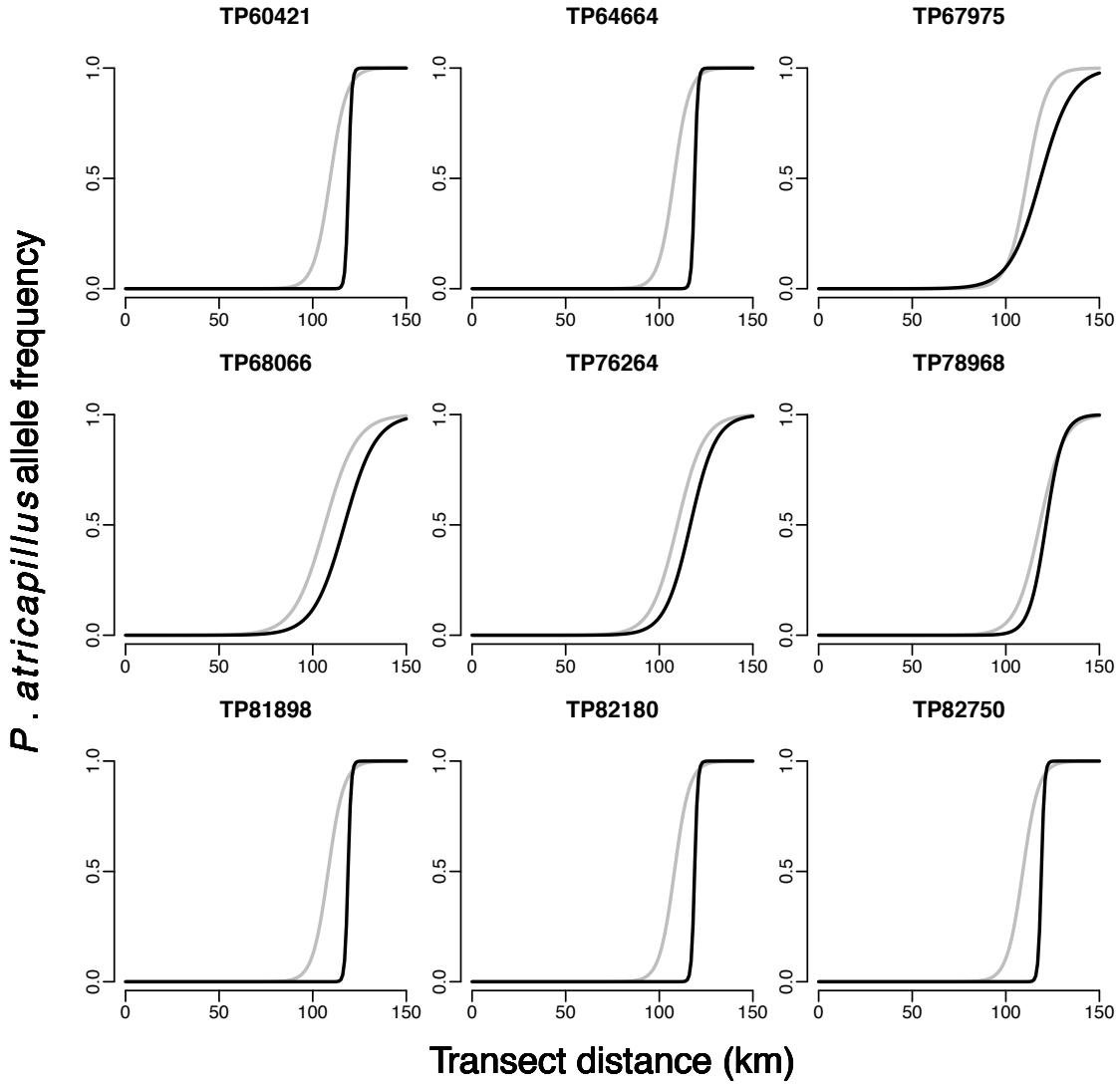
Supplemental Information



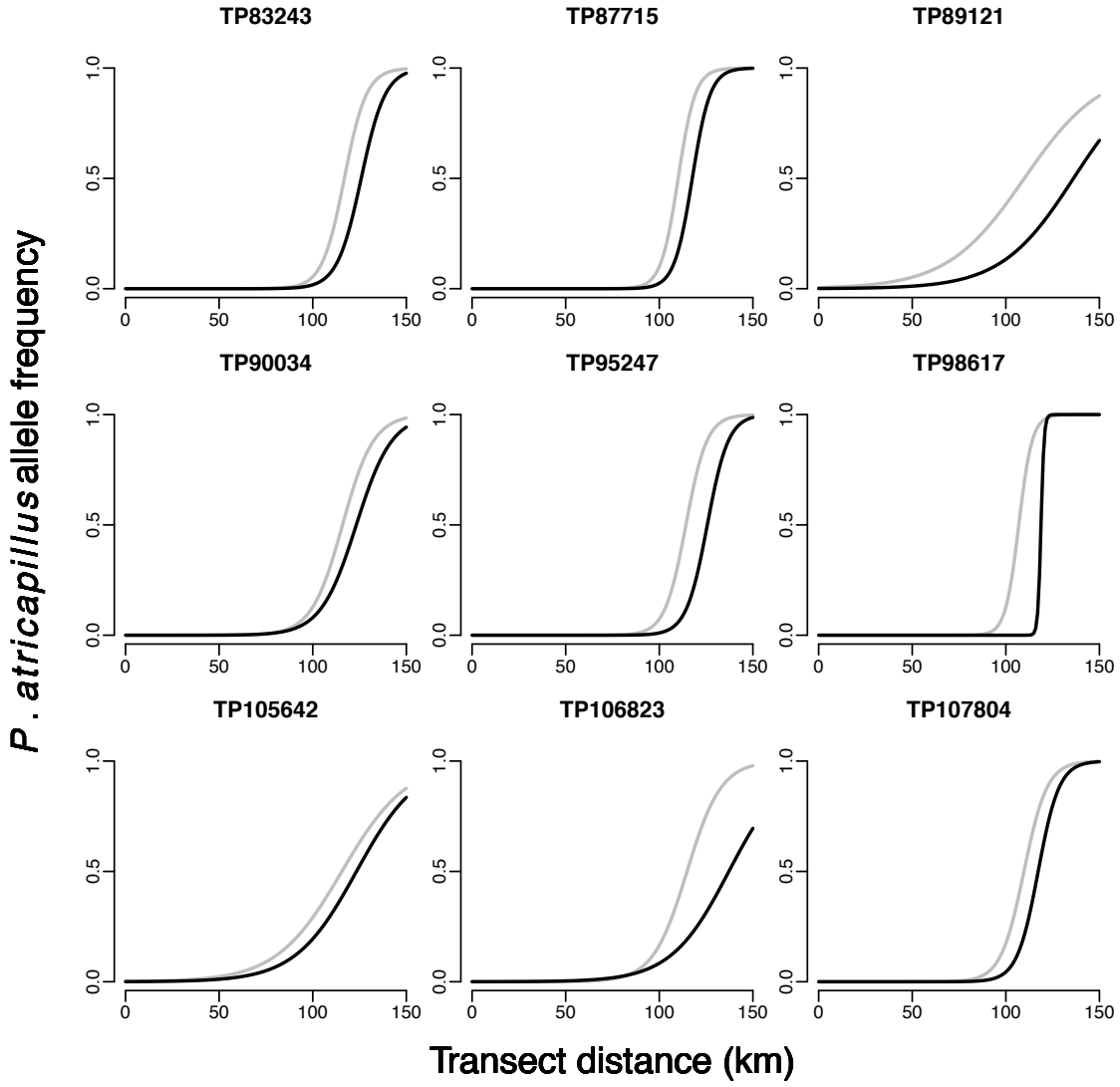
Supplemental Information



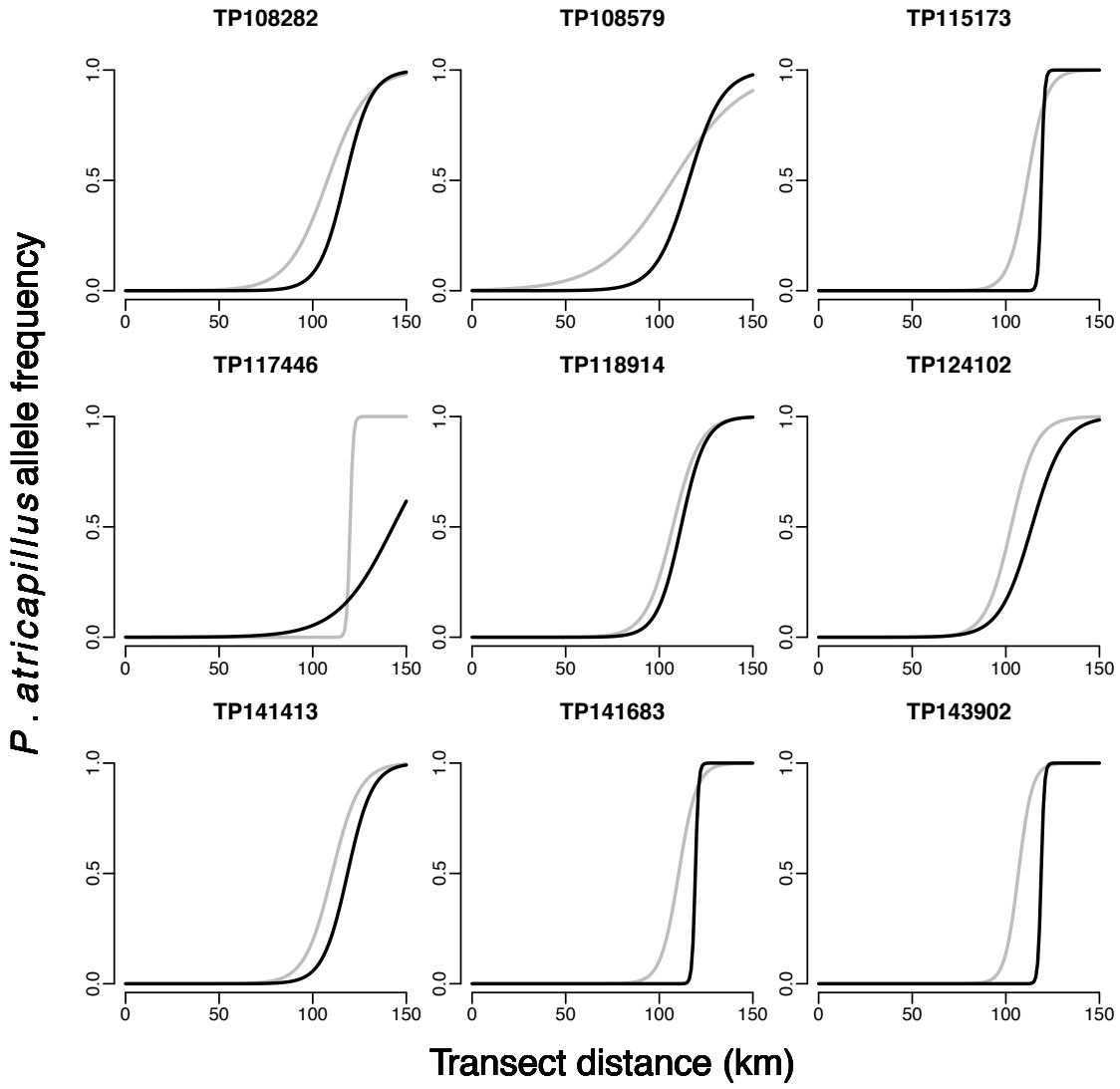
Supplemental Information



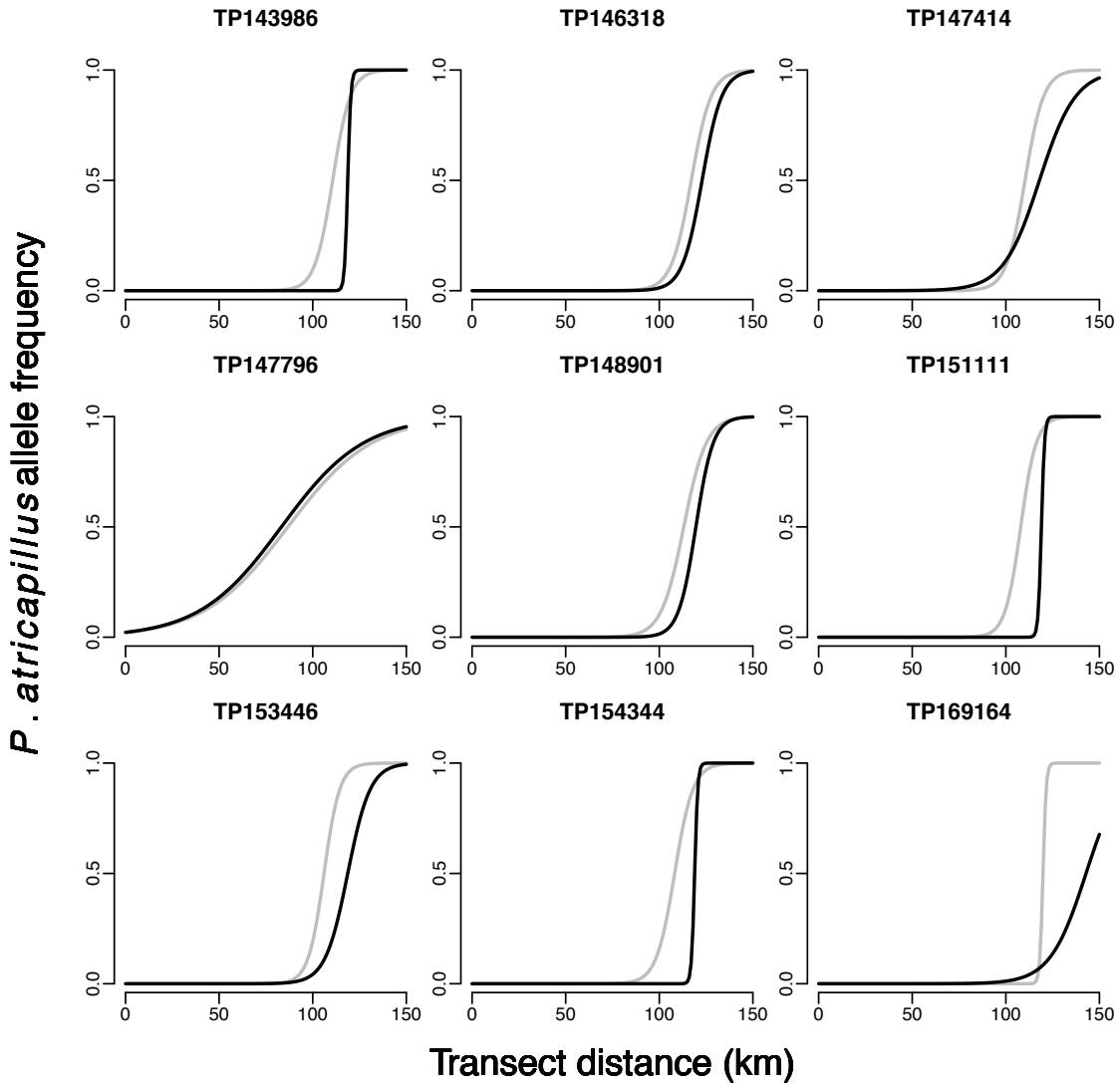
Supplemental Information



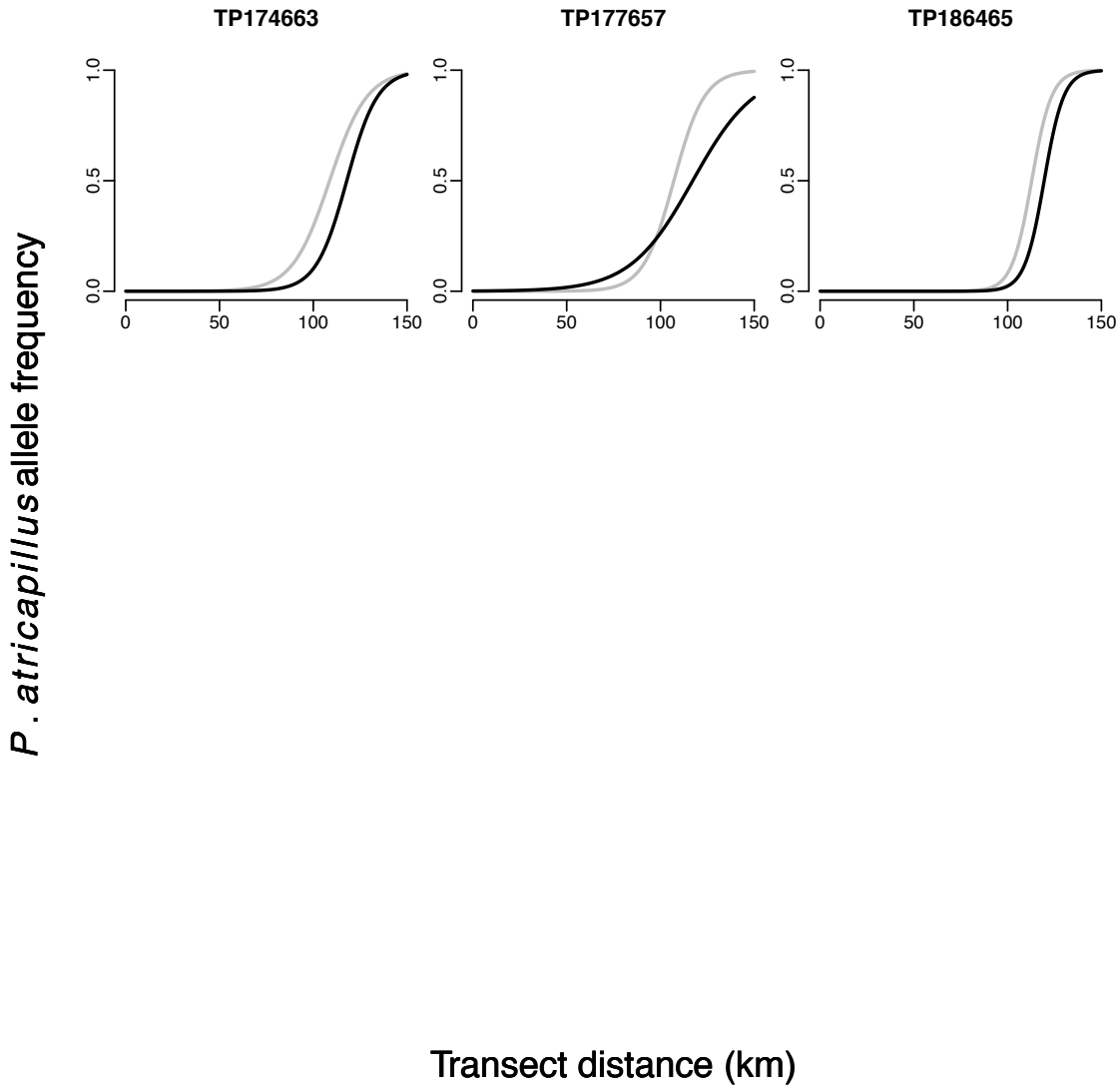
Supplemental Information



Supplemental Information



Supplemental Information



Supplemental Information

Table S1, related to Figure 1A and B. Interspecific F_{ST}, geographic cline parameters for both time periods (2000-2002 shaded grey, 2010-2012 unshaded), and putative chromosomal positions (based on alignment of GBS tags to the Zebra Finch genome) for the 75 loci that show clinal variation across the zone of contact. Interspecific F_{ST} outlier loci (bold + italicized).

Locus	F_{ST}	Center	Width	Centre	Width	Position
2402	0.75	109.04	18.13	119.00	3.33	1
2936	0.44	106.53	39.56	120.19	58.12	3
4872	0.52	104.31	65.00	116.09	34.70	19
6466	0.59	108.06	20.42	116.69	34.70	1
6773	0.76	106.98	14.28	119.00	3.34	Z
7200	0.56	113.92	19.90	122.24	16.56	8
7252	0.51	104.37	32.28	101.70	68.05	10
10116	0.57	106.97	19.52	118.92	40.94	5
11304	0.39	85.11	80.93	100.79	68.37	5
13296	0.5	113.14	40.68	117.36	30.10	4
13298	0.63	108.57	15.81	118.86	22.46	4
13439	0.44	115.03	41.47	114.71	37.36	4
14452	0.24	116.33	18.76	131.00	37.40	1A
14974	0.3	113.35	42.74	126.70	31.53	3
15361	0.55	108.69	28.01	117.48	28.74	1A
16034	0.52	110.12	80.34	114.13	48.37	13
16646	0.49	116.96	44.92	122.85	33.98	5
18491	0.52	108.60	75.56	113.17	58.73	8
18509	0.62	105.70	20.13	118.49	18.43	5
20053	0.63	109.19	18.20	118.34	3.25	5
20854	0.51	101.03	73.37	105.04	56.74	5
23528	0.53	103.46	55.21	118.06	39.39	9
23963	0.57	111.41	30.98	118.56	26.55	2
26109	0.75	112.19	21.38	119.00	3.33	Un
34177	0.46	92.01	81.44	104.25	67.00	1A
36959	0.43	116.13	26.32	124.57	22.78	1
38281	0.56	104.83	25.84	111.62	48.10	3
40606	0.76	109.36	23.90	118.83	20.51	Z
41934	0.76	111.19	21.74	118.58	3.27	Z
43099	0.4	110.43	20.72	118.13	92.81	1
47215	0.5	106.06	42.53	103.73	60.43	1A
47709	0.25	118.57	14.35	125.36	24.26	1A
49913	0.63	108.55	41.18	119.01	32.52	3
52460	0.5	106.68	44.18	114.26	32.53	14
52650	0.54	106.13	57.56	114.99	37.44	19
60288	0.6	110.21	30.42	116.20	26.63	2

Supplemental Information

60421	0.74	109.32	17.61	119.13	3.35	1
64664	0.76	107.82	16.52	118.90	3.33	Z
67975	0.57	111.42	20.68	118.58	33.31	3
68066	0.54	106.52	33.67	116.95	33.54	19
76264	0.58	109.63	28.35	116.59	27.22	2
78968	0.51	118.09	25.43	121.44	18.10	Z
81898	0.76	108.24	16.78	118.87	3.31	Z
82180	0.76	108.09	16.11	118.88	3.31	Z
82750	0.76	109.08	17.00	119.00	3.34	5
83243	0.34	116.89	23.68	125.90	25.77	1B
87715	0.76	110.07	18.34	117.63	19.20	1
89121	0.18	109.89	82.51	136.14	77.26	3
90034	0.29	115.91	33.14	123.27	38.08	1
95247	0.45	114.24	22.47	125.65	22.50	1A
98617	0.76	106.92	14.35	118.79	3.30	Z
105642	0.05	115.57	70.08	123.35	65.63	1A
106823	0.26	114.82	36.58	137.18	62.16	Un
107804	0.59	109.67	24.83	117.27	22.48	5
108282	0.64	107.80	42.60	117.25	28.10	3
108579	0.54	107.17	75.08	115.79	35.82	2
115173	0.75	111.53	20.37	119.00	3.33	1A
117446	0.25	119.93	3.37	142.90	59.68	1
118914	0.5	107.22	29.22	111.61	25.76	4
124102	0.41	102.81	27.29	113.85	34.55	20
141413	0.58	110.37	29.44	118.60	26.55	9
141683	0.76	110.37	19.27	119.26	3.35	1
143902	0.76	106.71	14.60	118.89	3.32	Z
143986	0.76	110.65	19.01	118.70	3.30	1
146318	0.46	117.10	21.30	122.87	21.08	1A
147414	0.76	110.25	19.91	117.96	38.80	5
147796	0.46	86.74	90.07	83.29	87.89	4A
148901	0.63	113.07	24.14	119.56	18.63	1
151111	0.75	107.95	16.75	119.00	3.33	Un
153446	0.45	106.03	16.85	118.71	24.04	1A
154344	0.76	108.21	19.64	118.90	3.31	Z
169164	0.27	119.85	3.37	142.84	38.62	1A
174663	0.64	108.71	40.02	117.87	32.85	3
177657	0.26	107.15	32.69	117.20	66.76	4
186465	0.64	112.84	21.26	119.49	20.88	1A

Supplemental Information

Table S2, related to Figure 1C. Likelihood estimates for different values of K (number of genetic clusters) from STRUCTURE including mean ln likelihood (ln P(X|K)) for 20 iterations and ΔK for each value of K calculated using the method outlined in Evanno et al. (2005).

	2000-2002		2010-2012	
K	ln P(X|K)	ΔK	ln P(X|K)	ΔK
1	-75464		-69687	
2	-69554	5292	-63048	5564
3	-70170	1164	-64123	141
4	-69441	-	-65374	4747
5	-	-	-70273	-

Supplemental Information

Supplemental Experimental Procedures

UNEAK and post-processing methods, and genomic dataset generation

Raw Illumina data files were filtered to individual genotypes using the Universal Network Enabled Analysis Kit (UNEAK) pipeline [S1], which is available as part of TASSEL 3.0 [S2]. The UNEAK pipeline retains reads with a barcode, restriction enzyme cut site and no ambiguous bases ('N's) in the 64 bp of the sequence following the individual barcode, and trims all acceptable reads to 64bp after the barcode. The pipeline then clusters reads into tags and stores counts of the tags present in each barcoded individual. All unique tags are then merged, and their counts in the whole sample of individuals are stored. The pipeline then performs a pairwise alignment of tags. Tag pairs with 1bp mismatches are considered as candidate SNPs. Reciprocal pairs of tags are retained as SNPs according to standard protocols of the Cornell Institute for Genomic Diversity with a user-specified error tolerance rate (0.03 here). After SNP identification, counts of each tag (or allele) are output for each locus and each individual. Following UNEAK filtering, individual genotypes were re-called using a global sequencing error rate of 0.03 and following the method detailed in [S3]. Genotype likelihood was calculated using a binomial sampling distribution. A genotype was called if its AIC value was at least 4 lower than the next best genotype. If this condition was not met, the genotype was coded as 'missing'. Loci with a mean observed heterozygosity greater than 0.75 were discarded as a way to filter out potential paralogs. Discarding loci with heterozygosities between 0.5 and 1 did not significantly impact our results (data not shown). After filtering, we had 1425 loci that could be confidently called in at least 80% of individuals.

Illumina sequencing of 190 individuals on two lanes of the Illumina Hi-Seq platform resulted in 496,872,131 reads. This dataset was trimmed to 400,000,000 reads which contained a

unique barcode and cut site remnant and no ambiguous sites ('N's). Prior to filtering, the data consisted of 103,641 SNP loci with mean coverage 2x (minimum coverage per individual 0.08x, maximum coverage per individual 303x). Of 190 individuals, 167 passed initial filtering. The UNEAK pipeline was run for the remaining 167 individuals, identifying 20,363 biallelic SNP loci, many of which had low coverage or were present in only a handful of individuals. Loci with more than 20% missing data and with observed heterozygosity > 0.75 were excluded, resulting in a dataset of 1,425 loci, with a mean coverage of 22x (min 12x, max 239x). From alignment of the GBS tags to the zebra finch genome the loci in this dataset are putatively distributed throughout the chickadee genome on chromosomes 1 – 28 and the Z chromosome, excepting chromosome 16. (Taylor et al. in review).

Admixture analyses

To examine temporal changes in population admixture we used STRUCTURE Ver. 2.3.1, analyzing each time period separately [S4]. Analyses were performed using an admixture model, correlated allele frequencies, a burn-in period of 50 000 cycles, and 500 000 additional cycles (determined from test runs to be sufficient for parameter stabilization). Using the no admixture model and sampling sites as prior information did not produce significantly different results (data not shown). Analyses were repeated 20 times for K = 1 – 5, where K = the number of genetic populations.

Supplemental Information

Geographic cline fitting and concordance

To quantify hybrid zone movement we used a geographic cline approach. We fitted allele frequencies for each locus to a tanh model of cline shape by MLE [S5], and corrected sample sizes following [S6] as

$$N_e = \frac{2N}{2N * F_{ST} + F_{IS} + 1}$$

where N is the number of individuals sampled in a deme, F_{IS} is the deficit of heterozygotes (zero if not positive), and F_{ST} is the fluctuation of allele frequencies between loci, after accounting for differences in their cline shapes. It is calculated from the residual variation around the regression line fitted during the concordance analysis (see below).

Initially tanh clines were fitted for each locus that allowed both cline centre and width to vary in both time periods. Two further models were fitted to the historic data, constraining either the centre or the width to be equal to that for the same locus in the contemporary sample. Constrained and free models were compared using likelihood ratio tests.

The concordance between mean hybrid index and allele frequency at each individual locus was calculated using the logit-logistic model of [S7]. The predicted allele frequency p in deme i is given by

$$p_i = \frac{S^{v_i}}{S^{v_i} + (1 - S)^{v_i} e^{u_i}}$$

where S is the mean hybrid index over all loci, u gives the relative difference in cline position, and v gives the relative difference in slope. Perfect concordance between a focal locus and the

mean hybrid index would result in $u = 0$ and $v = 1$. Parameters u and v were fitted by MLE, using the R package mle2.

Distribution modeling and Climate association

To complement the genetic information collected from 2010 - 2012 on the transect of intensively-studied locations, we created a model of the current distribution of *P. atricapillus* and *P. carolinensis* over a broader area, based on data from the eBird database [S8]. We used eBird data from the area between 36° and 41°N latitude and 74° and 92° W longitude, and contained 45462 records with observations of the two chickadee species from the months of May and June (the middle of the breeding season) of 2010 through 2012 inclusive. From these data, we created a map of the contact zone by making a spatial interpolation of the proportions of sites within overlapping 0.5° × 0.5° degree grids spaced at 0.015° intervals across the area of interest (Figure 3A). To assess whether the eBird data contained enough information to show contact zone movement over the past decade we examined proportions of Carolina chickadees reported in regions surrounding transect sampling locations and compared these to long-term site-specific data, as well as estimates of population admixture from the genomic data.

We chose not to use the distribution data from eBird to directly describe broad-extent changes in the location of the contact zone between the two species of chickadees, because results from preliminary analyses indicated that the data from eBird were too sparse from the earlier period of 2000-2002 to assess the location of the contact zone in this earlier period. Hence, we used a different approach to examining evidence for shifts in distribution, by testing the hypothesis that changes in range have occurred as a result of climate change and the determination of northern range limit of *P. carolinensis*. The northern range limit of *P.*

carolinensis, and thus contact zone between *P. carolinensis* and *P. atricapillus*, closely aligns to the average minimum winter temperature -7°C isotherm and physiological evidence suggests that *P. atricapillus* in Ohio are better able to tolerate winter temperatures than *P. carolinensis* [9-11]. We first examined whether winter temperature accurately predicted the location of the contact zone in 2010-2012, and then examined whether this relationship also held a decade earlier based on predicting the location of the contact zone in this earlier period, using temperature data from this earlier period, for locations where we did have data from eBird.

We looked for evidence that winter temperature determines the contact zone (Figure 3A) using information on the contact zone derived as explained above and information on mean minimum winter temperature obtained from PRISM (http://www.prism.oregonstate.edu/). We chose mean minimum winter temperature as our climate descriptor recognizing that minima should reflect temperature fluctuations and extremes, and chose not to model variance for the purposes of this study because of an absence of an obvious mechanism for effects of extreme temperatures on chickadee physiology, and the fact that mean and minimum temperatures are typically correlated. The data points were all eBird data locations within the area of interest, and we classified a location as being in the contact zone if at least 5% of eBird checklists within a 0.5° block reported both of the chickadee species (the orange region in Figure 3A). The value for minimum winter temperature at each of these locations was the average minimum daily temperature from the period Dec. 2009 - Feb. 2012 obtained from PRISM GIS data layers. The region within which we assessed the relationship between minimum winter temperature and contact zone location was restricted to central Pennsylvania, between 78° and 75° degrees west longitude, which encompassed the contemporaneously sampled genomic transect. We restricted our analysis to data from this region for two reasons. First, we removed the eastern-most

available data in order to reduce the possible confounding effect of maritime climate causing differences in the meaning of the daily minimum winter temperature values: e.g., minimum temperatures might be experienced for shorter durations closer to the Atlantic ocean. Nevertheless, there might still be a gradient maritime effect that extends well inland, and so we modeled this in our analyses as an east – west gradient. However, this east – west gradient is only interpretable to the east of the Appalachian Mountains that run roughly north – south, which is the reason for our second, western constraint on the extent of the data used. The statistical model that we fit to our data was a logistic regression in which presence in the contact zone was the binary response variable and the predictors were: minimum winter temperature as linear and quadratic predictors, longitude as a linear predictor, and an interaction between longitude and squared minimum temperature. In biological terms, this model predicts the probability that any location will be within the contact zone, allowing for peak probabilities at some intermediate temperature, and with the temperature at which this peak probability occurs being allowed to vary from east to west.

Given the relationship between minimum temperature and presence in the contact zone that we identified for the years 2010-2012, we evaluated whether this relationship was causal or merely a correlation by looking at the accuracy with which this model was able to predict the location of the contact zone for the smaller number of locations at which data from eBird were available in the years 2000-2002. Observed information was created for this earlier period using the same methods outlined above, and the locations of the 2000-2002 points and their associated minimum winter temperatures during this time period were run through the logistic regression model created for the period a decade in the future (see previous paragraph), in order to produce predictions of whether each point was or was not in the contact zone. In order to make estimates

of presence in the contact zone for the earlier period robust, we only used data from 2000-2002 locations if these locations had at least 20 sites within a 0.5° × 0.5° block surrounding each focal location. The accuracy with which predictions matched the observed status of each site was evaluated by calculating the AUC statistic to compare predicted probabilities with observed information. AUC values range from 0.5 (no association between observation and prediction) and 1 (perfect association between ordering of sites from highest to lowest probabilities.

All analyses of data from eBird were conducted using the R statistical language [S12], with the supplemental sp package [S13] used for manipulation of spatial data, the glm function used for conducting the logistic regression analysis, and the supplemental PresenceAbsence package [S14] used to calculate AUC statistics.

Supplemental References

S1. Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler M. D., Buckler, E. S., Costich, D. E. (2013). Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. PLoS genetics. DOI: 10.1371/journal.pgen.1003215

S2. Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., Buckler, E. S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformat *23*, 2633-2635.

S3. Lynch, M. (2009). Estimation of allele frequencies from high-coverage genome-sequencing projects. Genetics *182*, 295-301.

S4. Pritchard, J. K., Stephens, M., Donnelly, P. (2000). Inference of population structure using

multilocus genotype data. Genetics *155*, 945-959.

S5. Szymura, J. M., Barton, N. H. (1991). The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. Evolution. *45*, 237-261.

S6. Raufaste, N., Orth, A., Belkhir, K., Senet, D., Smadja, C. et al. (2005) Inferences of selection and migration in the Danish house mouse hybrid zone. Biol. J. Linn. Soc. *84*, 593-616.

S7. Fitzpatrick, Benjamin M. (2013) Alternative forms for genomic clines. Ecol. Evol. *3*, 1951-1966.

S8. Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D. et al. (2009) eBird: A citizen-based bird observation network in the biological sciences. Biol. Cons. *142*, 2282-2292.

S9. Root, T. (1988). Atlas of Wintering North American Birds (Chicago: University of Chicago Press).

S10. Repasky, R. R. (1991). Temperature and the northern distributions of wintering birds. Ecology *72*, 2274-2285.

S11. Olson, J. R., Cooper, S. J., Swanson, D. L., Braun, M. J., and Williams, J. B. (2010). The relationship of metabolic performance and distribution in black-capped and Carolina chickadees. Physiol. Biochem. Zool. *83*, 263-275.

Supplemental Information

S12. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

S13. Bivand, R. S., Pebesma, E., Gomez-Rubio, V. (2013). Applied spatial data analysis with R, Second edition, (New York, Springer).

S14. Freeman, E. A. and Moisen, G. (2008). PresenceAbsence: An R Package for Presence-Absence Model Analysis. Journal of Statistical Software, *23*, 1-31.