The LINNEAN SOCIETY of London BIOLOGICAL Journal of the Linnean Society

Biological Journal of the Linnean Society, 2014, **111**, 863–868. With 2 figures

Secondary contact followed by gene flow between divergent mitochondrial lineages of a widespread Neotropical songbird (*Zonotrichia capensis*)

LEONARDO CAMPAGNA[1,2,†], CECILIA KOPUCHIAN[2,3], PABLO L. TUBARO[2] and STEPHEN C. LOUGHEED[1*]

[1]*Department of Biology, Queen's University, Kingston, ON, K7L 3N6, Canada*
[2]*División de Ornitología, Museo Argentino de Ciencias Naturales 'Bernardino Rivadavia', Avenida Ángel Gallardo 470, Ciudad de Buenos Aires, Buenos Aires, C1405DJR, Argentina*
[3]*Centro de Ecología Aplicada del Litoral, Ruta 5, Km 2.5, Corrientes, 3400, Argentina*

Received 22 November 2013; revised 27 January 2014; accepted for publication 27 January 2014

Understanding how genetic and phenotypic differences that arise in geographically isolated populations influence the outcome of secondary contact advances our knowledge of speciation. In the present study, we investigate the secondary contact between divergent lineages of a widespread Neotropical songbird, the Rufous-collared sparrow (*Zonotrichia capensis*). *Zonotrichia capensis* is morphologically and behaviourally diverse, and shows a pattern of lineage diversification produced by a Pleistocene expansion and colonization of South America from a probable Central American origin. Consistent with previous results, we find three lineages throughout the species range, showing between 1.5% and 2.5% divergence in mitochondrial control region sequences. These lineages come into secondary contact in the Dominican Republic, La Paz (Bolivia), and North-eastern Argentina. We use DNA microsatellite data to study a broad secondary contact zone in North-eastern Argentina, finding that Bayesian clustering analyses do not assign individuals to their respective mitochondrial lineages. Overall, we did not observe nuclear genetic discontinuities in the study area. We conclude that, if genetic, morphological, and/or cultural differences accumulated among lineages during isolation, they were insufficient to prevent gene flow after secondary contact. *Biological Journal of the Linnean Society*, 2014, **111**, 863–868.

ADDITIONAL KEYWORDS: contact zone – phylogeography – Pleistocene – range expansion – Rufous-collared Sparrow – South America.

INTRODUCTION

Insights into the process of speciation can be obtained when populations within a species (or from closely-related taxa) that diverged in allopatry come into secondary contact (Harrison, 1990). Both the environment and the strength of the pre- and/or postzygotic reproductive isolation barriers that may have arisen between these populations will determine what happens once secondary contact is established (Harrison, 1990). Possible outcomes range from absence of gene flow because of complete reproductive isolation to hybridization and eventually panmixia, with intermediate scenarios, including reinforcement (Sætre *et al.*, 1997), movement of the contact zone (Buggs, 2007), and hybrid speciation (Brelsford, Milá & Irwin, 2011). Inferences about processes that influence speciation can be made by correlating the nature of the differences accumulated between populations and the particular outcome of the secondary contact.

The Rufous-collared sparrow (*Zonotrichia capensis*, Müller 1776) is a widespread, morphologically and behaviourally diverse Neotropical passerine that is a

*Corresponding author. E-mail: steve.lougheed@queensu.ca.
†Current address: Fuller Evolutionary Biology Program, Cornell Laboratory of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY, USA, 14850.

model for the study of avian vocalizations (Nottebohm, 1969; Lougheed & Handford, 1992; Tubaro & Segura, 1994). Phenotypic diversity within this species overlies deep patterns of mitochondrial lineages, interpreted to be the product of the species' colonization history of Central and South America (Lougheed *et al*., 2013). A phylogeographical analysis from across the species' range revealed three mitochondrial lineages that likely originated during the Pleistocene and show signatures of population expansion and low levels of historical gene flow (Lougheed *et al*., 2013). Two localities were found to contain individuals from different mitochondrial lineages suggesting that they lie within contact zones. In the present study, we increase previous sampling efforts and obtain mitochondrial sequence and DNA microsatellite data to study a putative contact zone in North-eastern Argentina in greater detail. Our objective was to obtain a finescale spatial map of the contact zone and use genetic tools to understand the implications of these two lineages coming back into contact.

MATERIALS AND METHODS

Sampling and datasets

We augmented the dataset from Lougheed *et al*. (2013) by adding 94 individuals from 32 localities from the southern half of the species' distribution, with the objective of sampling an area where two mitochondrial lineages had been previously found to come into contact. In total, we sampled 187 individuals at 64 localities from across the species range (Fig. 1A; see also Supporting information, Table S1). Mitochondrial control region sequences (CR) were obtained as described in Lougheed *et al*. (2013). A subset of 151 samples from the southern portion of the *Z. capensis* range (Fig. 1A; see also Supporting information, Table S1) were genotyped for seven DNA microsatellite loci used previously for this species (Moore, Bonier & Wingfield, 2005; Eikenaar *et al*., 2013).

Genetic analysis

We aligned CR sequences and built a Bayesian tree in MrBayes, version 3.1.2 (Ronquist & Huelsenbeck, 2003) as described in detail in Lougheed *et al*. (2013). Inter-individual and inter-lineage *p*-distances were calculated in MEGA, version 5 (Tamura *et al*., 2011), and variation in the inter-individual distance matrix was summarized and displayed graphically using a principal coordinates analysis (PCA) performed in GENALEX, version 6 (Peakall & Smouse, 2006). Pairwise *F*-statistics between lineages or localities were calculated in ARLEQUIN, version 3.5 (Excoffier & Lischer, 2010).

To assess patterns of genetic structure in DNA microsatellite data, we used Bayesian clustering software: STRUCTURE, version 2.3.4 (Pritchard, Stephens & Donnelly, 2000) and GENELAND, version 4.0.3 (Guillot, Mortier & Estoup, 2005). We ran STRUCTURE implementing the admixture ancestry model and correlated allele frequencies, both with and without information on the origin of each individual (e.g., using locality or mitochondrial lineage as priors). We explored values of $K = 1$ through ten, conducting ten iterations per K value, each consisting of 2 000 000 generations following a burn-in of 500 000. The most likely K value was determined in accordance with the methods described in Evanno, Regnaut & Goudet (2005). Different iterations of the optimal K value were combined in CLUMPP, version 1.1.2 (Jakobsson & Rosenberg, 2007) and displayed using DISTRUCT, version 1.1 (Rosenberg, 2004). We used GENELAND to incorporate finescale geographical information into the analysis of population structure, conducting runs of 10 000 000 iterations (thinning of 100) and discarding the initial 20% as burn-in. We implemented the spatial model and correlated allele frequencies, using $K = 5$ as chosen in STRUCTURE. We divided the study area into a 5 × 10 grid of cells (each of approximately five by five degrees) and obtained a map of population membership. Finally, to determine whether genetic variation across the study area was clinal, we tested for isolation-by-distance with a Mantel test performed in GENALEX at both the individual and locality level.

RESULTS

Analysis of mitochondrial control region sequences

Consistent with Lougheed *et al*. (2013), we found three distinct CR lineages within *Z. capensis*. Inter-lineage Φ_{ST} and mean *p*-distances ranged from 0.67 to 0.79 ($P < 0.0001$ for all comparisons) and 1.5% to 2.5%, respectively. These lineages are distributed geographically as shown in Figure 1A (hereafter lineages A, B, and C; colour-coded red, yellow, and blue, respectively). Pairs of lineages come into contact in the Dominican Republic (A/B); La Paz, Bolivia (A/B); and North-eastern Argentina (B/C). The Bayesian tree shows three highly supported clades (posterior probability of 1) with respect to outgroup taxa (Fig. 1B). The remaining nodes were not highly supported (posterior probability < 0.95). Individuals from lineage C form a clade that is, in turn, embedded within a larger clade that contains all of the individuals from B. Finally, this BC clade is embedded within a clade with all of the individuals from A. Previous results (Lougheed *et al*., 2013) show that these lineages are reciprocally monophyletic and the relationship between them is

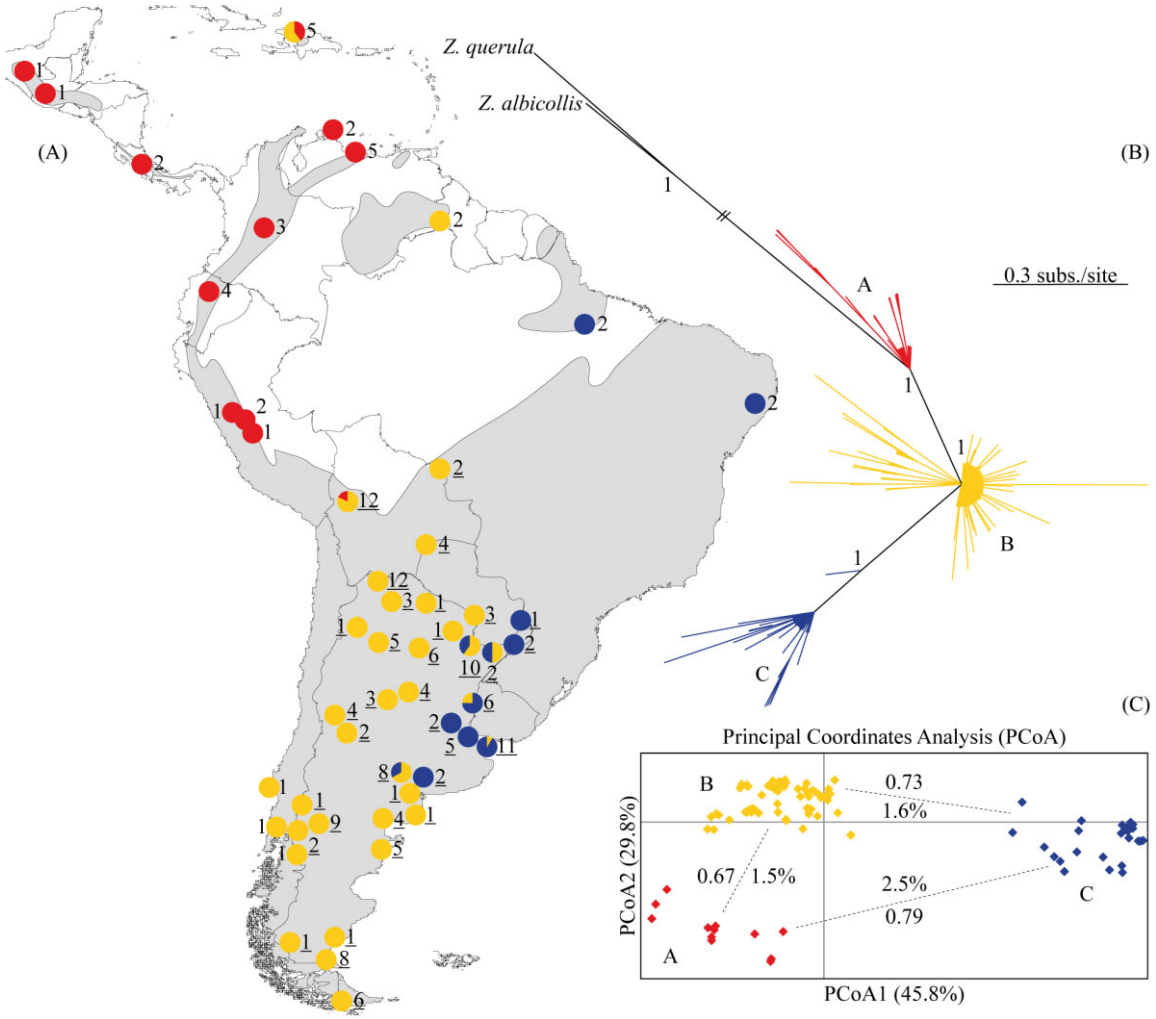

Figure 1. A, circles representing sampling localities superimposed onto a distribution map of *Zonotrichia capensis* in light grey (*sensu* Schulenberg, 2010). Different colours represent the mitochondrial lineage to which individuals from a given locality belong (red, A; yellow, B; blue, C). When more than one lineage was present at a given locality, circles were coloured proportionally to the number of individuals from each lineage. Numbers represent sampling effort and are underlined when both mitochondrial and DNA microsatellite data were obtained for those individuals. B, Bayesian tree with posterior probability showing node support. C, principal coordinates analysis displaying inter-individual *p*-distances (percentage of total variation explained by each axis in parenthesis). Inter-lineage mean percentage *p*-distances and Φ_{ST} values are also shown.

fully resolved by incorporating additional mitochondrial and nuclear sequence data (A is sister to B and C). Finally, individuals from each lineage cluster together in a PCA (Fig. 1C).

Analysis of DNA microsatellite data

Data from DNA microsatellites were used to explore the B/C contact zone in further detail; we did not find strong support for nuclear genetic discontinuities in the study area. STRUCTURE was unable to assign individuals to different genetic clusters: the model with the highest probability was for $K = 1$ (see Supporting information, Fig. S1). Birds sampled in the contact zone did not cluster together, nor did individuals group by mitochondrial lineage (B or C) when this information was incorporated into the analysis as Bayesian priors (see Supporting information, Fig. S1). We observed a pattern consistent with the northern and southern extremes of the study area showing the highest differentiation when individuals from different localities were grouped by latitude (and this information was incorporated into the model). Figure 2A (see also Supporting information, Fig. S1) shows an example where localities were placed into five groups (dashed lines) and the value of K chosen *sensu*

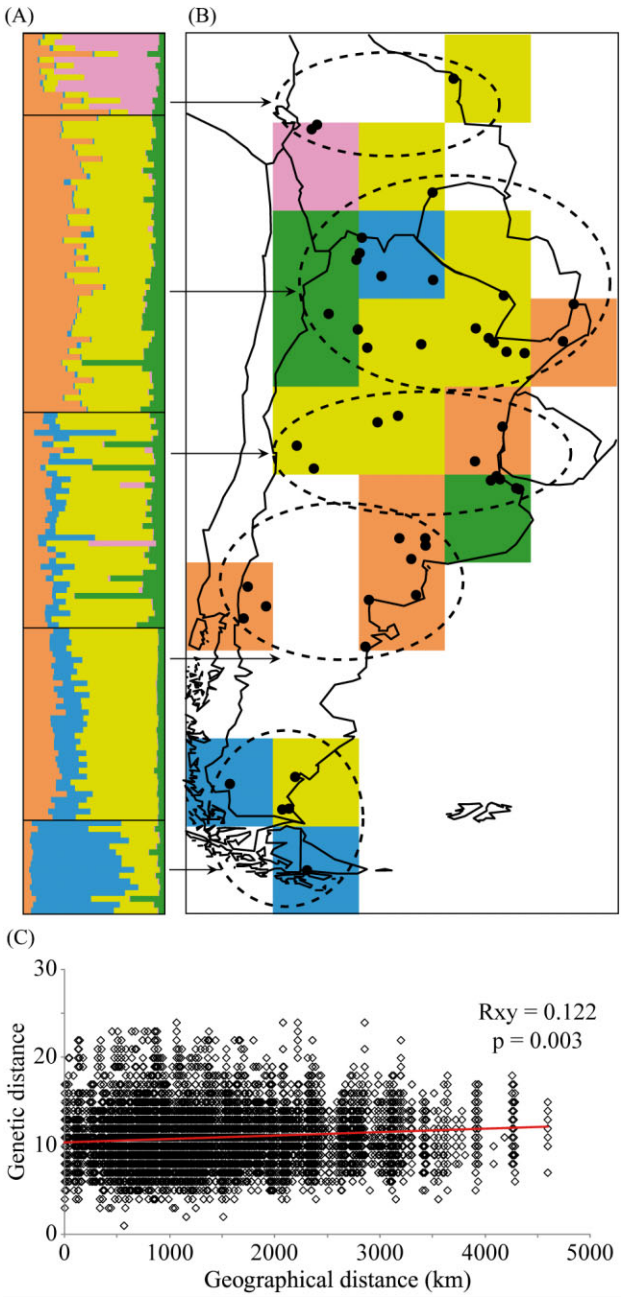

Figure 2. Structure plot (A) and results from a GENELAND analysis (B) of DNA microsatellite data for five genetic clusters. Genetic clusters are shown in different colours, either in the bars of the STRUCTURE plot or the different cells of the 5 × 10 cell grid in which the sampling area was divided for GENELAND analysis (cells on this grid that did not include sampling localities were omitted). Individuals in (A) were grouped by locality as shown with dashed lines in (B) and this information was incorporated as a prior into the model. C, result from an individual-based Mantel test (correlation coefficient and *p*-value) between geographical and genetic distance matrices.

Evanno *et al*. (2005) was five. Consistent with a pattern of isolation-by-distance, the two most genetically distinct clusters ($F_{ST} = 0.08$, $P < 0.0001$) correspond to the most geographically distant sampling localities (Fig. 2A) and intermediate groups of localities show clinal variation in allele frequencies. Because sampling throughout the study area is approximately continuous (i.e. there are no clear groups of localities), we explored alternative ways of grouping localities for analysis using STRUCTURE and obtained similar results for all (not shown). We tested for individual based isolation-by-distance and found a small but statistically significant signal ($P = 0.003$, Fig. 2C), which was not recovered when assessed at the level of localities ($P = 0.246$) or groups of localities as defined in Figure 2A ($P = 0.102$; see Supporting information, Fig. S2).

Finally, we incorporated finescale geographical information into the analysis of population structure using GENELAND, dividing the study area into a 5 × 10 grid. We implemented a five genetic cluster model (clusters are represented by different colours in Fig. 2B). Cells on this grid that did not include sampling localities are not coloured because their assignment to a genetic cluster represents an extrapolation done by GENELAND. Different cells on the grid that include individuals primarily from a given lineage (B or C) did not group together (compare Fig. 2B with Fig. 1A). However, with few exceptions, cells that were closer together tended to belong to the same genetic cluster (Fig. 2B).

DISCUSSION

The stability of a hybrid zone is influenced by factors such as hybrid fitness, rate of dispersal of parental lineages, habitat heterogeneity, and environmental change (Buggs, 2007). In the present study, we find deep mitochondrial lineages in *Z. capensis* that contrast with little nuclear genetic structure in neutral markers. These lineages diverged in the Pleistocene and presumably expanded with negligible gene flow when coming to occupy their current geographical range and establishing secondary contact (Lougheed *et al*., 2013). This has possibly led to extensive gene flow in the contact zone, and erosion of (neutral) nuclear genetic differences that may have existed between individuals of these lineages at the time of secondary contact. This suggests that inter-lineage hybrids do not suffer negative fitness consequences, which is not surprising given that postzygotic isolation may arise very slowly (Arrieta, Lijtmaer & Tubaro, 2013). *Zonotrichia capensis* shows remarkable phenotypic variation not only in morphology, migratory behaviour, and vocal dialects throughout its vast range, but also at a small geographical

scale. Differences in male song, reproductive timing, and genetic differentiation have been observed between two trans-Andean populations separated by only 25 km (Moore *et al*., 2005). It is possible that this is the consequence of local adaptation or phenotypic plasticity: the latter explains the differences in mRNA transcription patterns between high and low altitude Peruvian populations (Cheviron, Whitehead & Brumfield, 2008). Male vocalizations were found to correlate with habitat type consistently for over a decade at a scale of only 7 km^2 (Kopuchian *et al*., 2004), suggesting that cultural evolution can take place at a microgeographical scale. Our DNA microsatellite data do not show genetic discontinuities within Argentine populations, contrasting with the marked phenotypic variation in morphology and song described in this area (Lougheed *et al*., 2013). Taken together, this implies that these characters are either highly plastic and/or have evolved relatively rapidly, overlying deeper genealogical patterns. It is possible that coding regions in the genome are highly differentiated among populations and are responsible for generating some of these traits, in which case neutral loci are not expected to recover this signal unless tightly-linked to these loci. Although the different *Z. capensis* mitochondrial lineages began diverging in the Pleistocene (Lougheed *et al*., 2013), we do not know when they came back into contact. Regardless, we conclude that the genetic and cultural differences that may have accumulated when in isolation were insufficient to act as barriers to gene flow once *Z. capensis* lineages came back into contact.

The present study is the most comprehensive phylogeographical study of *Z. capensis* performed to date and describes the distribution of three mitochondrial lineages found throughout Central and South America. They have come into secondary contact in three areas: North-eastern Argentina, the Dominican Republic, and La Paz (Bolivia). The latter contact zone had not been described previously; however, it is possible that it extends south into Tucumán (Argentina) where two diverged mitochondrial lineages were reported from an altitudinal transect (Lougheed, Handford & Baker, 1993), and thus our sampling in the area was insufficient to identify individuals from both lineages. Increasing sampling across the three contact zones will allow a better understanding of the outcome of secondary contact between the divergent *Z. capensis* lineages.

ACKNOWLEDGEMENTS

The present study was funded by grants to PLT from ANPCyT (Argentina), CONICET (Argentina), UBA (Argentina), IDRC (Canada), and the Richard Lounsbery Foundation (USA), and an NSERC Discovery Grant (Canada) to SCL. We thank three anonymous reviewers and J. A. Allen for helpful comments on the manuscript.

REFERENCES

Arrieta RS, Lijtmaer DA, Tubaro PL. 2013. Evolution of postzygotic reproductive isolation in galliform birds: analysis of first and second hybrid generations and backcrosses. *Biological Journal of the Linnean Society* **110:** 528–542.

Brelsford A, Mila B, Irwin DE. 2011. Hybrid origin of Audubon's warbler. *Molecular Ecology* **20:** 2380–2389.

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

Cheviron ZA, Whitehead A, Brumfield RT. 2008. Transcriptomic variation and plasticity in rufous-collared sparrows (*Zonotrichia capensis*) along an altitudinal gradient. *Molecular Ecology* **17:** 4556–4569.

Eikenaar C, Bonier F, Martin PR, Moore IT. 2013. High rates of extra-pair paternity in two equatorial populations of rufous-collared sparrow, *Zonotrichia capensis*. *Journal of Avian Biology* **44:** 600–602.

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

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10:** 564–567.

Guillot G, Mortier F, Estoup A. 2005. GENELAND: a program for landscape genetics. *Molecular Ecology Notes* **5:** 712–715.

Harrison RG. 1990. Hybrid zones: windows on evolutionary process. In: Futuyma D, Antonovics J, eds. *Oxford surveys in evolutionary biology*. Oxford: Oxford University Press, 69–128.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **14:** 1801–1806.

Kopuchian C, Lijtmaer DA, Tubaro PL, Handford P. 2004. Temporal stability and change in a microgeographical pattern of song variation in the rufous-collared sparrow. *Animal Behaviour* **68:** 551–559.

Lougheed SC, Campagna L, Dávila JA, Tubaro PL, Lijtmaer DA, Handford P. 2013. Continental phylogeography of an ecologically and morphologically diverse Neotropical songbird, *Zonotrichia capensis*. *BMC Evolutionary Biology* **13:** 1–16.

Lougheed SC, Handford P. 1992. Vocal dialects and the structure of geographic variation in morphological and allozymic characters in the rufous-collared sparrow, *Zonotrichia capensis*. *Evolution* **46:** 1443–1456.

Lougheed SC, Handford P, Baker AJ. 1993. Mitochondrial DNA hyperdiversity and vocal dialects in a subspecies transition of the rufous-collared sparrow. *The Condor* **95:** 889–895.

Moore IT, Bonier F, Wingfield JC. 2005. Reproductive asynchrony and population divergence between two tropical bird populations. *Behavioral Ecology* **16:** 755–762.

Nottebohm F. 1969. The song of the chingolo, *Zonotrichia capensis*, in Argentina: description and evaluation of a system of dialects. *The Condor* **71:** 299–315.

Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6:** 288–295.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572–1574.

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

Sætre GP, Moum T, Bures S, Král M, Adamjan M, Moreno J. 1997. A sexually selected character displacement in flycatchers reinforces premating isolation. *Nature* **387:** 589–592.

Schulenberg TS. 2010. *Rufous-collared sparrow (*Zonotrichia capensis*), neotropical birds online.* Ithaca, NY: Cornell Laboratory of Ornithology. Available at: http://neotropical.birds.cornell.edu/portal/species/overview?p_p_spp=657516

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28:** 2731–2739.

Tubaro PL, Segura ET. 1994. Dialect differences in the song of *Zonotrichia capensis* in the southern Pampas: a test of the acoustic adaptation hypothesis. *The Condor* **96:** 1084–1088.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Results from alternative ways of analyzing the DNA microsatellite data in STRUCTURE; the probability of the data and ΔK values ranging from $K = 1$ to 10 are shown to the right of the STRUCTURE plots. Note that in (a), (b), and (c), the highest likelihood is for $K = 1$. (a) The data were analyzed without incorporating information from sampling localities and plots for K values with the highest ΔK (two and five) are shown. (b) The mitochondrial lineage to which each sample belonged to was incorporated into the model; again, plots for models with the highest ΔK are shown (two and five). The only individual belonging to lineage A for which DNA microsatellite data were obtained was excluded from this analysis. (c) The data were analyzed incorporating information on whether the samples belonged to the contact zone between B and C (plots for two and five genetic clusters are shown). (d) Localities were grouped as illustrated in Fig. 2B and this information was incorporated into the model; the STRUCTURE plot for $K = 5$ is shown.

Figure S2. Results from Mantel tests conducted on geographical and genetic distance matrices (correlation coefficient and *p*-value). We tested for isolation-by-distance at the level of (a) individuals, (b) localities, and (c) groups of localities as defined in Fig. 2. In (b) and (c), we used the mean genetic distances between groups of individuals from the same geographical location. The midpoint among groups of localities was used to calculate geographical distances in (c).

Table S1. Details of the *Zonotrichia capensis* individuals sampled in the present study.