

SUPPORTING ONLINE MATERIAL

Speciation by Distance in a Ring Species

Darren E. Irwin, Staffan Bensch, Jessica H. Irwin, and Trevor D. Price

Materials and methods

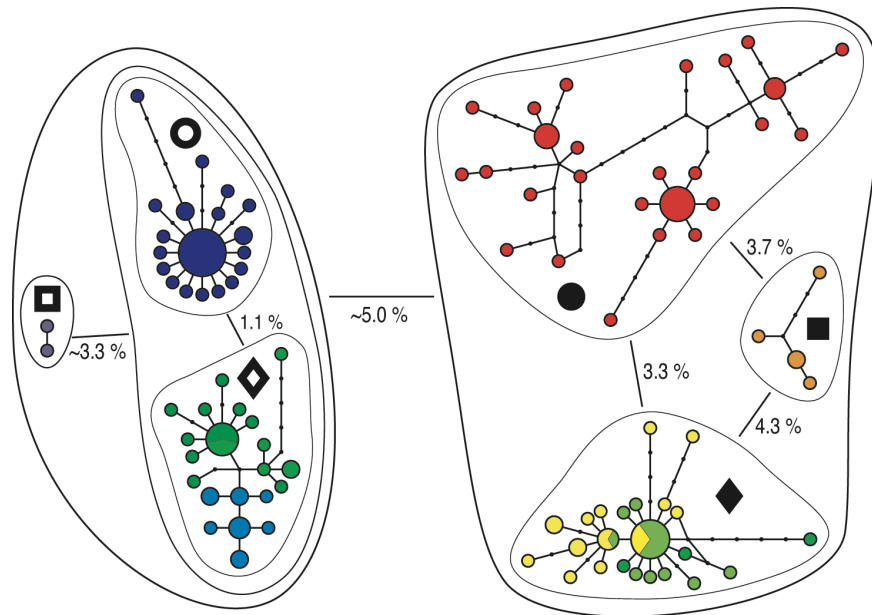
Previously, we examined variation in mitochondrial DNA sequences (~1200 bp from the neighborhood of the control region and ND6 gene) from 149 greenish warblers from 27 sites throughout their breeding range (S1). Here we summarize those data in a haplotype network calculated by the program TCS (S2) and drawn by hand (Fig. S1). In this study we used the same samples for AFLP analysis. Because AFLP requires DNA extracts of a higher concentration and quality than simple PCR amplification of mtDNA requires, only 105 of the samples (from 26 sites) were suitable for AFLP genotyping.

We used the protocol of Bensch et al. (S3), which was based on the method of Vos et al. (S4). We used the restriction enzymes *EcoRI* and *TruI* to digest genomic DNA, and then synthetic oligonucleotides (“adaptors”) were ligated to the fragments. Next, we performed two rounds of PCR using primers corresponding to, in the first round, the adaptor plus 1 arbitrary base pair, and in the second round, the adaptor plus 3 arbitrary base pairs. Fluorescein-labelled primers were used in the second round of PCR (the “selective amplification”). The products were separated in 6% denaturing polyacrylamide gels and visualized using a Vistra FluorImager.

We obtained 62 AFLP markers that were variable and could be scored unambiguously, from three primer combinations (19 markers with $E_{TCT} \times M_{CGA}$, 23 with $E_{TGA} \times M_{CGT}$, 20 with $E_{TAG} \times M_{CAT}$). AFLP bands were scored as absence/presence data. Each of the 105 individuals in the analysis had a unique multilocus genotype. We summarized variation in the resulting 0/1 matrix by performing a principal coordinate analysis (PCO) using the R package (S5). Given a matrix of pairwise genetic distances among individuals, this procedure determines the major axes of variation in the data set. We present results (Fig. 1B) based on Euclidean distances, but similar results were obtained with other distance measures, such as squared Euclidean distances and Jaccard distances. We used the program Arlequin (S6) to calculate corrected average pairwise distances between populations (the mean number of pairwise differences between two populations minus the average distance between individuals within those populations). Arlequin was also used to calculate pairwise F_{ST} values between populations. Note that this method of calculating F_{ST} makes no assumption regarding Hardy-Weinberg equilibrium.

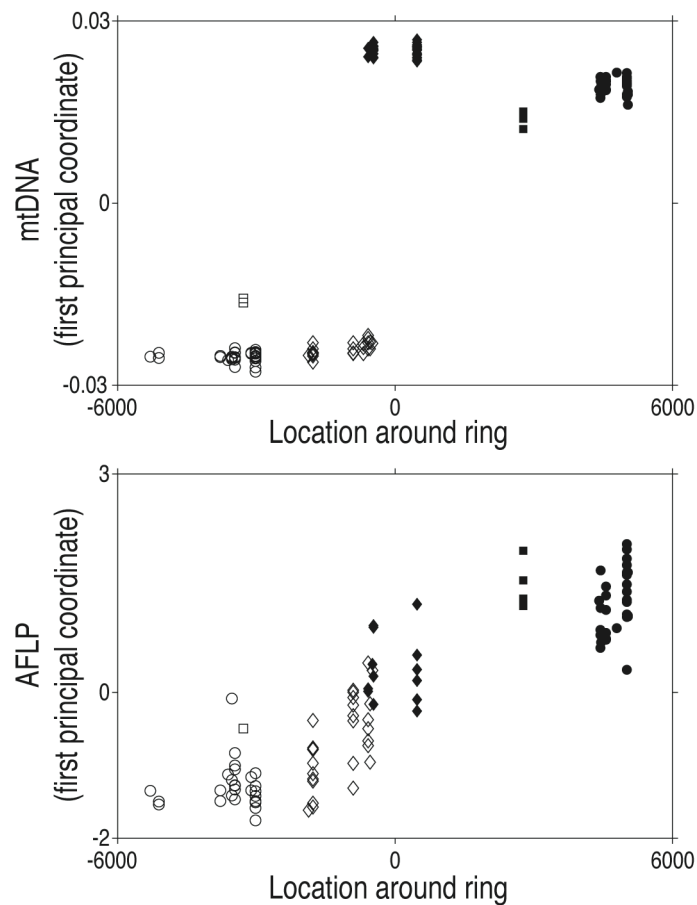
To test for a pattern of isolation by distance, we performed a Mantel test using the R Package (S5) to compare corrected AFLP distances between populations with geographic distances through the ring of greenish warbler populations. These “corrected” geographic distances were measured assuming a barrier to gene flow spanning from central Siberia south into the Tibetan Plateau. Thus corrected distances between west Siberian (*viridanus*) and east Siberian (*plumbeitarsus*) populations were measured through the long chain of populations running to the south of Tibet. To determine a single axis representing geographic location around the ring (e.g. in Fig. S2), we performed a principal coordinates analysis on the matrix of corrected geographic distances. The first principal coordinate axis explained 57% of the variation in the corrected geographic distances, and this was then used as a measure of geographic location around the ring.

Fig. S1. Mitochondrial DNA haplotype network showing six major clades and the divergence between them, based on data described in (*SI*). Colors correspond to the location where each haplotype was sampled according to Fig. 1, and the area of each circle is proportional to the number of samples with that haplotype. Missing haplotypes are indicated by black dots. The symbols next to each clade indicate the symbol used on the map (Fig. 1a) that corresponds to that clade. There are two major mitochondrial clades that overlap both in central Siberia, where they delineate reproductively isolated forms (*SI*), and within the subspecies *ludlowi* in the western Himalayas, where they do not correspond to reproductively isolated forms.



Supplemental results

Comparison of mtDNA and AFLP. Fig. S2. Patterns of variation around the southern side of the ring, measured from west Siberia (on left) around the southern side of Tibet to east Siberia (on right) for mitochondrial haplotypes (top) and AFLP genotypes (bottom). Each vertical axis corresponds to the first principal coordinate axis, with a principal coordinate analysis conducted for the mtDNA distances in a similar way as the AFLP distances. Each point represents a single individual, and the different symbols indicate different mitochondrial clades according to Figs. 1 and S1. See the Methods above for a description of how location was determined for this figure. The much sharper discontinuities in the mtDNA are expected for a single non-recombining molecule (S7).



Analysis based on F_{ST} values. **Fig. S3.** Diagram showing geographic relationships of major sampling sites and pairwise F_{ST} values between geographically close sites based on AFLP variation. Symbols correspond to different mtDNA clades (Fig. S1). The F_{ST} values through the southern chain of populations are much less than those between west Siberian *viridanus* and east Siberian *plumbeitarsus*.

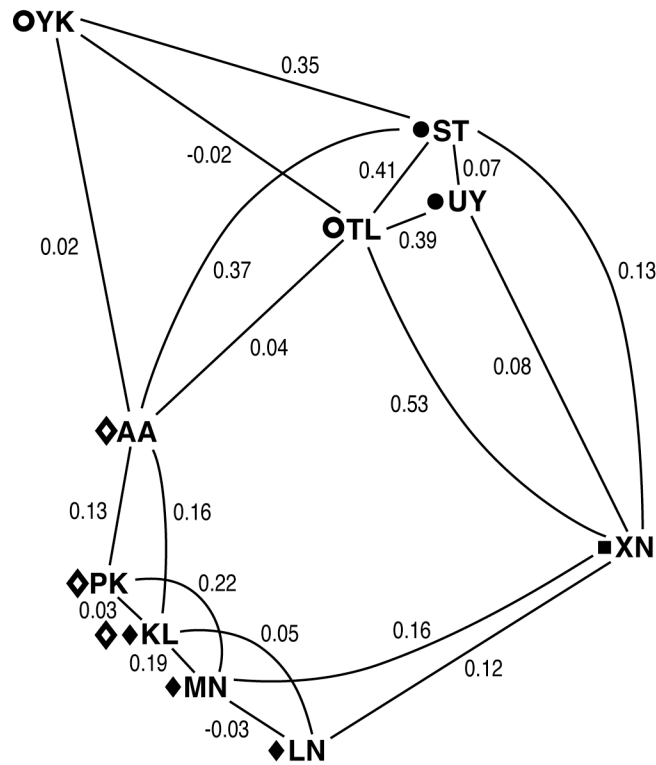


Table S1. Pairwise measures of divergence between major sampling sites.

	YK	TL	AA	PK	KL	MN	LN	XN	ST	UY
YK	7.73	7.04	7.60	10.33	11.05	14.50	13.17	17.00	17.36	15.78
TL	-0.02	6.68	7.20	10.16	11.41	14.94	13.58	17.43	18.16	16.04
AA	0.02	0.04	7.16	10.03	10.46	13.80	12.70	15.64	16.99	14.57
PK	0.11	0.16	0.13	10.43	10.93	14.81	13.00	14.80	16.00	15.04
KL	0.16	0.24	0.16	0.03	10.76	14.57	12.10	14.34	16.95	14.93
MN	0.30	0.38	0.32	0.22	0.19	13.33	12.42	14.10	17.00	15.08
LN	0.24	0.32	0.26	0.13	0.05	-0.03	12.27	13.00	16.44	14.67
XN	0.47	0.53	0.46	0.29	0.26	0.16	0.12	10.60	14.21	13.17
ST	0.35	0.41	0.37	0.24	0.27	0.20	0.21	0.13	13.64	14.52
UY	0.33	0.39	0.33	0.22	0.19	0.11	0.13	0.08	0.07	13.40

Above the diagonal are average AFLP distances between populations. On the diagonal are average AFLP distances within populations. Below the diagonal are F_{ST} between populations. Bold F_{ST} values are significant at the $P < 0.005$ level.

References – Online Supplement

- S1. D. E. Irwin, S. Bensch, T. D. Price, *Nature* **409**, 333 (2001).
- S2. M. Clement, D. Posada, K. A. Crandall, *Mol. Ecol.* **9**, 1657 (2000).
- S3. S. Bensch, A. J. Helbig, M. Salomon, I. Seibold, *Mol. Ecol.* **11**, 473 (2002).
- S4. P. Vos *et al.*, *Nucleic Acids Research* **23**, 4407 (1995).
- S5. P. Casgrain, P. Legendre. *The R Package for Multidimensional and Spatial Analysis* (<<http://www.fas.umontreal.ca/biol/casgrain/en/labo/R/>> Département de sciences biologiques, Université de Montréal, 2001).
- S6. S. Schneider, D. Roessli, L. Excoffier. *Arlequin: A Software for Population Genetics Data Analysis. Ver 2.000.* (<<http://lgb.unige.ch/arlequin/>> Genetics and Biometry Laboratory, University of Geneva, Switzerland, 2000).
- S7. D. E. Irwin, *Evolution* **56**, 2383 (2002).