MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 5726–5739 doi: 10.1111/mec.12977

Mapping migration in a songbird using high-resolution genetic markers

KRISTEN C. RUEGG,*† ERIC C. ANDERSON,‡ § KRISTINA L. PAXTON,¶** VANESSA APKENAS,‡ SIRENA LAO,* RODNEY B. SIEGEL,† † DAVID F. DESANTE,† † FRANK MOORE** and THOMAS B. SMITH*‡ ‡

*Center for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, La Kretz Hall, Suite 300, 619 Charles E. Young Dr. East, Los Angeles, CA 90095, USA, †Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, CA 95060, USA, ‡Southwest Fisheries Science Center, National Marine Fisheries Service, 110 Shaffer Road, Santa Cruz, CA 95060, USA, §Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA 95060, USA, ¶Department of Biological Sciences, University of Southern Mississippi, Hattiesburg, MS 39406, USA, **Department of Biology, University of Hawaii, Hilo, HI 96720, USA, ††The Institute for Bird Populations, PO Box 1346, Point Reyes Station, CA 94956, USA, ‡‡Department of Ecology and Evolutionary Biology, University of California, Los Angles, CA 90095, USA

Abstract

Neotropic migratory birds are declining across the Western Hemisphere, but conservation efforts have been hampered by the inability to assess where migrants are most limited—the breeding grounds, migratory stopover sites or wintering areas. A major challenge has been the lack of an efficient, reliable and broadly applicable method for measuring the strength of migratory connections between populations across the annual cycle. Here, we show how high-resolution genetic markers can be used to identify genetically distinct groups of a migratory bird, the Wilson's warbler (*Cardellina pusilla*), at fine enough spatial scales to facilitate assessing regional drivers of demographic trends. By screening 1626 samples using 96 highly divergent single nucleotide polymorphisms selected from a large pool of candidates (~450 000), we identify novel region-specific migratory routes and timetables of migration along the Pacific Flyway. Our results illustrate that high-resolution genetic markers are more reliable, precise and amenable to high throughput screening than previously described intrinsic marking techniques, making them broadly applicable to large-scale monitoring and conservation of migratory organisms.

Keywords: conservation genetics, neotropical migrant, RAD-sequencing, wildlife management

Received 1 August 2014; revision received 17 October 2014; accepted 20 October 2014

Introduction

Over half of the Neotropic migrant bird species found breeding in North America have shown marked declines in abundance over the last several decades (Robbins *et al.* 1989; Sauer *et al.* 2012). Population declines are thought to relate to stressors encountered by migrants at each stage in the annual cycle—the breeding grounds, the wintering grounds and migratory stopover points (Rappole 1995). At each location, birds are subject to a number of disturbances including habitat loss, collisions with wind turbines and cell phone towers, predation by house cats, exposure to disease and global climate change (Jonzen *et al.* 2006; Altizer *et al.* 2011; Loss *et al.* 2013). Understanding the strength of migratory connectivity, defined as the degree to which individuals from the same breeding site migrate to the same wintering site (Webster & Marra 2004), is critical to identifying the impact of local stressors on population declines because disturbances at one stage of the annual cycle can have carry over effects to the other phases of the annual cycle (Marra *et al.* 1998; Sillett *et al.* 2000; Norris & Taylor 2006). Furthermore, it

Correspondence: Kristen C. Ruegg, Fax: +1 (310) 825 5446; E-mail: kruegg@ucsc.edu

is impossible to fully understand the ecology, evolution and life histories of migratory animals without knowledge of the entire annual cycle (Webster *et al.* 2002).

Historically, efforts to identify the strength of migratory connectivity relied on recovery of individual birds previously tagged with bird bands. However, this approach has met with limited success for small-bodied songbirds because recapture rates of birds away from their original banding sites are often very low (<1 in 10 000; Gustafson & Hildenbrand 1999; Faaborg *et al.* 2010b). More recently, geolocators, small tracking devices that record information on ambient light levels to estimate an individuals location, have increased our knowledge of the movement patterns of many songbird species (Stutchbury *et al.* 2009), but remain impractical for most large-scale applications (1000's of individuals) due to cost, weight restrictions and the need to recover individuals to collect data from the devices (Arlt *et al.* 2013; Bridge *et al.* 2013). Similarly, satellite-tracking devices have revealed a wealth of information on individual movement patterns for larger-bodied animals (Bonfil *et al.* 2005; Croxall *et al.* 2005), but are currently infeasible for small-bodied songbirds due to size and weight restrictions. Alternatively, genetic and isotopic markers that use information contained within the feathers to pinpoint an individuals population of origin have broad appeal because they are cost-effective, non-invasive and do not require recapture (Rubenstein *et al.* 2002; Kelly *et al.* 2005; Rundel *et al.* 2013). However, intrinsic marking techniques have been plagued in the past by low resolution and/or technical issues related to working with feathers (Segelbacher 2002; Lovette *et al.* 2004; Wunder *et al.* 2005). Thus, there remains a need for a broadly applicable method for identifying migratory connectivity at spatial scales that are informative for assessing drivers of regional population declines.

In the last several years, genome sequencing has revolutionized the field of molecular ecology, resulting in new technologies that can be applied to molecular tagging of wild populations (Metzker 2010; Davey *et al.* 2011). Genome reduction techniques, such as Restriction Site Associated DNA sequencing (RAD-seq), can be used to sequence multiple individuals across a large fraction of the genome and identify hundreds of thousands of genetic markers that are useful for distinguishing populations (Baird *et al.* 2008; Hohenlohe *et al.* 2011). One type of genetic marker that can be identified from genomic sequence data is a single nucleotide polymorphism (SNP), DNA sequence variation occurring when a single nucleotide in the genetic code differs between individuals or homologous chromosomes. While a limited number (<400) of randomly selected SNPs may not provide adequate resolving power for identifying migratory populations (Kraus *et al.* 2013), it has been found that a smaller number (<100) of SNPs found within or linked to genes under selection can be targeted to reveal population structure at spatial scales that are critical to regional conservation planning (Nielsen *et al.* 2009, 2012; Hess *et al.* 2011). Furthermore, SNP-specific assays that target short fragments of sequence around the SNP loci of interest and can be advantageous in cases where the DNA is highly fragmented or available only in very small quantities (Kraus *et al.* 2014), such as DNA from a single passerine feather.

Here, we develop high-resolution SNP assays for tracking populations of a migratory bird, the Wilson's warbler, *Cardellina pusilla*, using a combination of Restriction Site Associated DNA Paired-End sequencing (RAD-PE seq) and high throughput SNP screening. The Wilson's warbler, a long-distance Neotropic migratory bird with a cross-continental breeding distribution (Ammon & Gilbert 1999), is particularly appropriate as model for testing the efficacy of high-resolution molecular markers because previous population genetic/connectivity studies on this species provide a solid basis for comparison between methods (Yong *et al.* 1998; Kimura *et al.* 2002; Clegg *et al.* 2003; Paxton *et al.* 2007, 2013; Irwin *et al.* 2011; Rundel *et al.* 2013). By harnessing recent advances in next-generation sequencing, we scan the genomes of Wilson's warblers sampled from across the breeding range and identify a set of highly divergent SNP loci with strong potential for population identification. We then develop SNPtype™ Assays that target the highly divergent loci and use them to screen 1626 samples from across the annual cycle. We illustrate how the resulting map of region-specific migration patterns can be used to help identify drivers of demographic trends at regional scales and inform studies of migrant stopover ecology.

Methods

Sample collection

Collection of 1648 feather and blood samples (22 samples for the SNP ascertainment panel and 1626 for the SNP screening panel) from 68 locations across the breeding, wintering and migratory range was made possible through a large collaborative effort with bird banding stations within and outside the Monitoring Avian Productivity and Survivorship (MAPS), the Landbird Monitoring of North America (LaMNA) and the Monitoreo de Sobrevivencia Invernal (MoSI) networks (Table 1). Genetic samples, consisting of the tip of one outer rectrix or blood collected by brachial vein puncture and preserved in lysis buffer (Seutin *et al.*

Table 1 Number of Wilson's warblers successfully screened at each location across the species breeding, wintering and migratory range. Locations in close proximity were merged on the map in Fig. 1. Uppercase letters are reserved for breeding populations, while lower case letters are reserved for migratory stopover and wintering locations

Location	Latitude	Longitude	*n*	Population
Breeding (June 10–July 31)				
Cantwell_1, Denali National Park, AK	63.449	−150.813	10	A
Cantwell_2, Denali National Park, AK	63.594	−149.611	11	A
Denali, Denali National Park, AK	63.716	−149.088	8	A
Yakutat, AK	59.514	−139.681	21	B
Ugashik_1, AK	57.175	−157.269	10	C
Ugashik_2, AK	57.183	−157.283	16	C
Juneau, AK	58.300	−134.400	10	D
Hardisty Creek, Calgary, AB	53.500	−117.500	2	E
Ram Falls, Calgary, AB	52.000	−115.800	5	E
Benjamin Creek, Calgary, AB	51.500	−115.000	2	E
Beaver Dam, Calgary, AB	51.104	−114.063	16	E
100 Mile House, BC	51.700	−121.300	13	F
Darrington, WA	48.208	−121.576	3	G
Silverton, WA	48.051	−121.433	5	G
Roy, WA	47.056	−122.488	4	G
Harlan, OR	44.506	−123.630	23	H
McKenzie Bridge, OR	44.199	−121.956	22	I
Eureka, CA	40.783	−124.123	18	J
Half Moon Bay, CA	37.506	−122.494	17	K
Big Sur, CA	36.286	−121.842	15	L
San Luis Obispo, CA	35.195	−120.489	23	M
Tennant, CA	41.492	−121.939	25	N
Clio, CA	39.667	−120.600	15	O
Hume, CA	36.799	−118.599	16	P
Hillary Meadow, MT	48.347	−113.976	2	Q
Crow Creek, MT	47.471	−114.279	1	Q
Elgin_1, OR	45.817	−117.865	4	R
Elgin_2, OR	45.679	−118.115	21	R
Pingree Park, Fort Colins, CO	40.550	−105.567	19	S
Grand Mesa, CO	39.000	−107.900	11	T
Camp Myrica, QC	49.700	−73.300	17	U
Hilliardton, ON	47.500	−79.700	4	V
Fredericton, NB	45.800	−66.700	4	W
Migratory stopover (March–May)				
O'Neil Forbay Wildlife Area, CA	37.080	−121.022	75	a
Lower Colorado River, Cibola, AZ	33.300	−114.683	604	b
Buenos Aires National Wildlife Refuge, AZ	31.550	−111.550	71	c
San Pedro Riparian National Cons. Area, AZ	31.583	−110.133	52	c
Albuquerque, NM	35.013	−106.465	12	d
Sierra del Carmen_1, Coahuila, MX	28.909	−102.546	4	e
Sierra del Carmen_2, Coahuila, MX	28.861	−102.650	3	e
Fairview, TX	33.152	−96.600	43	f
Braddock Bay, NY	43.161	−77.611	19	g
Wintering (December—February)				
San Jose del Cabo, Baja California Sur, MX	22.883	−109.900	8	h
Chupaderos, Sinaloa, MX	23.333	−105.500	8	i
Las Joyas, Autlan, Jalisco, MX	19.767	−104.367	25	j
Nevado de Colima, Colima, Jalisco, MX	19.233	−103.717	3	j
U. of Mexico, San Angel, Distrito Federal, MX	19.313	−99.179	9	k
El Cielo Biosphere Reserve, Tamulipas, MX	23.000	−99.100	15	l
Coatapec, Veracruz, MX	19.450	−96.967	13	m
Parque Macuiltepec, Xalapa, Veracruz, MX	19.548	−96.921	7	m
Aeropuerto, Oaxaca, MX	17.100	−96.800	14	n
Tuxtlas, Veracruz, MX	18.400	−95.200	9	o
Chaa Creek, San Ignacio, BE	17.094	−89.069	1	p
Izalco, Sonsonate, SV	13.821	−89.653	17	q
Los Andes National Park, Santa Ana, SV	13.850	−89.620	7	q
	13.943	−89.617	7	q

Table 1 *Continued*

Location	Latitude	Longitude	*n*	Population
Las Lajas, Santa Ana, SV				
Metapan, Santa Ana, SV	14.403	−89.360	9	q
San Salvador Volcano, SV	13.700	−89.200	12	q
Cantoral, Tegucigalpa, HN	14.331	−87.399	11	r
La Tigra National Park, Tegucigalpa, HN	14.100	−87.217	15	r
El Jaguar Cafetal, Jinotega, NI	13.229	−86.053	10	s
Volcan Mombacho, Granada, NI	11.832	−86.008	2	s
Monteverde Cloud Forest, Santa Elena, CR	10.314	−84.825	9	t
San Vito_1, Puntarenas, CR	8.754	−82.926	2	u
San Vito_2, Puntarenas, CR	8.766	−82.943	2	u
San Vito_3, Puntarenas, CR	8.784	−82.975	5	u
San Vito_4, Puntarenaus, CR	8.809	−82.924	1	u
San Vito_5, Puntarenaus, CR	8.822	−82.972	12	u

1991), were purified using Qiagen DNeasy Blood and Tissue Kit and quantified using a NanoDrop™ Spectrophotometer (Thermo Scientific, Inc.; Smith *et al.* 2003). Breeding (June 10–July 31), spring migration (March 1–May 31) and wintering (December 1–February 28) samples were collected and categorized into groups based on collection date, signs of breeding (presence/size of a cloacal protuberance), signs of migration (extent of fat) and life history timetables for the Wilson's warbler (Ammon & Gilbert 1999). To assess migratory stopover site use through time in Cibola, AZ, 686 of the 1648 samples were collected using consistent effort, daily, passive mist-netting from March 22 to May 24, 2008 and 2009 (Table 1).

SNP discovery

To identify SNPs useful for distinguishing genetically distinct breeding groups within the Wilson's warbler, an ascertainment panel of 22 individuals was selected to represent the range of phylogenetic variation known in the species, including all three recognized subspecies (Ammon & Gilbert 1999; Kimura *et al.* 2002). Five individuals from each of five regions were included in the ascertainment panel, except for from the southwestern region where samples were limited to two individuals (Table S1, Supporting information). Purified extractions from blood samples were quantified using Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen Inc.), and Restriction Site Associated DNA Paired-end (RAD-PE) libraries containing individually barcoded samples were prepared at Floragenex, Inc. according to Baird *et al.* (2008) and Ruegg *et al.* (2014; Appendix S1, Supporting information). RAD-PE sequencing made it possible to build longer contigs (~300 bp) from short read, 100-bp Illumina HiSeq2000 (Illumina, San Diego, CA, USA) data in order to improve downstream bioinformatics and provide adequate flanking sequence around SNPs for assay development (Etter *et al.* 2011).

Samples from each isolate were sequenced on an Illumina HiSeq2000 (Illumina) using paired-end 100-bp sequencing reads. Paired-end sequences from each sample were collected, separated by individual, stripped of barcodes, trimmed to 70 bp, scrubbed of putative contaminant and high-copy-number-sequences and filtered to include only those with a Phred score ≥10. The sample with the greatest number of reads passing the initial quality filter was used to create a reference set of RAD-PE contigs against which sequences from other samples were aligned. To create the reference, primary reads were clustered into unique RAD markers and the paired-end sequences associated with each RAD tag were assembled de novo using Velvet (Zerbino & Birney 2008) into contigs ranging from 180 to 610 bp, with an average length of 300 bp. Paired-end reads from the remaining samples were aligned to this reference using Bowtie (Langmead & Salzberg 2012), and SNPs were identified using the SAMTOOLS software (Li *et al.* 2009) with mpileup module under standard conditions.

To narrow our data set to SNPs we could confidently use to assess population structure, we performed a second round of quality filtering and removed: (i) putative SNPs with more than two alleles; (ii) genotypes in individuals with a quality score of <30; (iii) genotypes with <8 reads in a homozygote or <4 reads per allele in heterozygotes; (iv) putative SNPs that had suitable genotypes in <12 of the 17 samples from four western populations or <5 of the 5 samples from the eastern population; and (iv) putative SNPs with <40 bp of flanking sequence on either side. To limit the chances of including linked markers, genomic coordinates were attained by mapping the remaining contigs to the closest, best annotated, songbird genome at the time, the zebra finch (*Taeniopygia guttata*; version 3.2.4; Warren *et al.* 2010) using BLAST+ (version 2.2.25; Camacho 2009).

To avoid the possibility of orthologs and/or erroneous matches, the data were filtered to include

only contigs that aligned to the zebra finch genome with only a single hit and an *E*-value $<10^{-40}$. Because SNPs with large frequency differences are the most effective for identifying populations, all SNPs that passed our second round of quality filters were ranked according to frequency differences between the five regions (Table S2, Supporting information) and 150 SNPs displaying the largest allele frequency differences between each of the 10 pairwise comparisons were selected for conversion to SNPtype™ Assays (Fluidigm Inc.). Before making a final selection, we also considered factors such as: GC content (<65%), number of genotypes per population and average coverage at a SNP across all populations (Table S2, Supporting information). An initial assay prescreening panel was then performed and the assay pool was further reduced to the 96 assays (the number that fit on a single 96.96 Fluidigm Array) that could be genotyped most reliably (Table S2, Supporting information).

SNP screening

The Fluidigm Corporation EP1™ Genotyping System was used to screen 96 SNP loci using 94 individuals per run and two nontemplate controls. To avoid the potential for high-grading bias (i.e. wrongly inflating the apparent resolving power of a group of loci for population identification; Anderson 2010), none of the 22 samples used in our original ascertainment panel were included in the final SNP screening and population structure analyses. To ensure amplification of low quality or low concentration DNA from feathers, an initial pre-amplification step was performed according to the manufacturers protocol using a primer pool containing 96 unlabelled locus-specific SNP type primers (Appendix S1, Supporting information). PCR products were diluted 1:100 and re-amplified using fluorescently labelled allele-specific primers. The results were imaged on an EP1 Array Reader, and alleles were called using Fluidigm's automated Genotyping Analysis Software (Fluidigm Inc.) with a confidence threshold of 90%. In addition, all SNP calls were visually inspected and any calls that did not fall clearly into one of three clusters—heterozygote or either homozygote cluster—were removed from the analysis. As DNA quality can affect call accuracy, a stringent quality filter was employed and samples with >6 of 96 missing loci were dropped. To assess the reliability of SNPtype assays for genotyping DNA from blood and feather extractions, the proportion of samples yielding useable genotype data was calculated. Tests for linkage disequilibrium and conformance to Hardy–Weinberg equilibrium (HWE; Louis & Dempster 1987) were performed using GENEPOP software, version 4.0 (Rousset 2008).

Population structure analysis

While genetic differentiation (F_{ST}) is likely inflated because selected loci were not a random sample from the genome, we calculated F_{ST} here for comparison to previous genetic analysis. F_{ST} between all pairs of populations was calculated as θ (Weir & Cockerham 1984), using the software GENETIX version 4.05 (Belkhir *et al.* 1996–2004) and the data were permuted 1000 times to determine significance. We used the program STRUCTURE version 2.2, to further assess the potential for population structure across the breeding grounds (Pritchard *et al.* 2000). Ten runs at each *K* value (*K* = 1–9) were performed under the admixture model with uncorrelated allele frequencies using a burn-in period of 50 000 iterations, a run length of 150 000. All scripts and data used for the STRUCTURE runs and subsequent population genomic analyses are located at https://github.com/eriqande/wiwa-popgen and archived on dryad (doi:10.5061/dryad.j5d33). To simplify the comparison of results, the program CLUMPP (Jakobsson & Rosenberg 2007) was used to reorder the cluster labels between runs, and individual *q* values (proportion of ancestry inferred from each population within an individual) were plotted using the program DISTRUCT (Rosenberg 2004). Visual inspection of DISTRUCT plots allowed identification of regions where geographic barriers to gene flow exist and/or where admixture is likely.

To identify how population structure was distributed across geographic space, we used the program GENELAND (Guillot *et al.* 2005). Analyses in GENELAND were performed under the spatial model assuming uncorrelated allele frequencies. Inference of population structuring was based on 10 independent runs, each allowing the number of populations to vary between 1 and 10. Each run consisted of 2.2 million MCMC iterations with a thinning interval of 100. Of the 22 000 iterations retained for the MCMC sample after thinning, the first 5000 were discarded as burn-in. Postprocessing of the MCMC sample was performed upon a 250 by 250 point grid that covered the breeding range of the species. Posterior probability of group membership estimates from GENELAND was visualized as transparency levels of different colours overlaid upon a base map from Natural Earth (naturalearthdata.com) and clipped to the Wilson's warbler breeding range using a shapefile (NatureServe 2012), making use of the packages sp, RGDAL, and raster in R (Pebesma & Bivand 2005; Bivand *et al.* 2014; Hijmans 2014; Team RC 2014; see dryad doi:10.5061/dryad.j5d33 and/or GITHUB: https://github.com/eriqande/wiwa-popgen). Thus, within each distinguishable group, the transparency of colours is scaled so that the highest posterior probability of membership in the group according to GENELAND is opaque and the smallest is entirely transparent.

To assess the accuracy of our baseline for identification of individuals from each population to genetically distinct breeding groups, we used the program GSI_SIM (Anderson *et al.* 2008; Anderson 2010). GSI_SIM uses an unbiased leave-one-out cross-validation method to assess the accuracy of self-assignment of individuals to populations. Posterior probabilities were obtained in GSI_SIM by summing the posterior probabilities of the populations within each genetically distinct group and assigning the individual to the group with the highest posterior probability.

Results

SNP discovery

RAD-PE sequencing on 22 individuals from five geographic regions representative of the range of phylogenetic variation known in the species resulted in 123 005 contigs (average length ~300 bp), containing 449 596 SNPs passing our initial quality filters (Table S1, Supporting information). The median depth of sequencing across all contigs within a library was 33×, and the average Phred quality score per library was 35 (Table S1, Supporting information). Overall, 166 268 SNPs passed the second round of quality filters and 19 707 of those were candidates for conversion into SNPtype™ Assays based upon the absence of variation in 40 bp of flanking sequence surrounding the SNPs. Candidate SNPs were ranked according to frequency differences, GC content, the number of genotypes per region and the average coverage, and the final panel was composed of 96 SNPs with pairwise frequency differences between regions ranging from 1 to 0.4 (Table S2, Supporting information). For contigs that could be mapped to the zebra finch genome with high confidence, the minimum distance between SNPs was 41 KB and no two SNPs were selected from the same contig to avoid the possibility of linked markers (Table S2, Supporting information). In this study, we refer to the final panel of 96 highly differentiated SNPs as high-resolution genetic markers.

SNP screening

The resulting high-resolution genetic markers were used to screen 1626 samples collected from 68 sampling locations across the breeding, wintering and migratory range (Table 1), with 117 samples excluded due to low quality genotypes (>6 loci excluded). The samples with the highest proportion of reliable genotypes were from fresh feather extractions ($n_{reliable}$/total = 660/686 or 96% reliable), followed by fresh blood extractions ($n_{reliable}$/total = 100/106 or 94% reliable), and finally feather and blood extractions that were >3 years old ($n_{reliable}$/total = 701/786 or 90% reliable). Tests for conformity to HWE revealed that all but 1 of the 94 loci (AB_AK_20) in 2 of the 23 breeding populations (D and L; Table 1, Fig. 1b) were in HWE after accounting for multiple comparisons ($P < 0.0005$). Deviations from HWE were likely the result of small sample sizes and or the unintentional inclusion of late arriving migrants en route to northern breeding sites. No loci were found to be in linkage disequilibrium after accounting for multiple comparisons using a strict Bonferroni correction ($P < 0.0005$), suggesting that loci were not physically linked even in cases where zebra finch genome coordinates could not be attained.

Population structure analysis

An analysis of population genetic structure on the breeding grounds identified six genetically distinct groups: Alaska to Alberta (purple, A–D), eastern North America (red, U–W), the Southern Rockies and Colorado Plateau (orange, S, T), the Pacific Northwest (green, G–J), Sierra Nevada (pink, N–P) and Coastal California (yellow, K–M; Fig. 1A, B). Pairwise F_{ST}'s between groups ranged from 0 to 0.68 with an overall F_{ST} of 0.179 (95% CI: 0.144–0.218; Table S3, Supporting information). The strongest genetic differentiation was observed between eastern and western groups (F_{ST} = 0.41–0.68) with strong genetic differentiation also seen between the Southern Rockies and Colorado Plateau and all other groups (F_{ST} = 0.09–0.27; Table S3, Supporting information). The number of genetically distinct groups was set at six based upon convergence between results from STRUCTURE ($K = 6$, average ln $P(X|K) = -33\,359$), GENELAND, and GSI_SIM (Fig. 1a, b; Table 2; Fig. S1, Supporting information). While seven genetically distinct groups were also strongly supported by GENELAND and STRUCTURE ($K = 7$, average ln $P(X|K) = -33\,286$; Fig. S1, Supporting information), with sampling locations from British Columbia and Alberta (E and F) forming a seventh group distinct from Alaska, the power to accurately assign individuals to groups at $K = 7$ decreased significantly using both STRUCTURE and GSI_SIM (Fig. S1, Supporting information).

Leave-one-out cross-validation using GSI_SIM indicated that the ability to correctly assign individuals to groups was high, ranging from 80% to 100%. The eastern group had the highest probability of correct assignment (100%), followed by Alaska to Alberta (94%), the Southern Rockies and Colorado Plateau (92%), the Pacific Northwest (84%), the Sierra Nevada (81%) and Coastal California (80%; Fig. 1b; Table 2). The majority of the incorrect assignments were between the Pacific North-

west, Sierra Nevada and Coastal California. Subsequent assignment of migrant and wintering individuals to genetically distinct breeding groups using GSI_SIM indicated that Coastal California, Sierra, and Pacific Northwest breeders winter in western Mexico and southern Baja, and migrate north along the Pacific Flyway, with Coastal California and Sierra breeders found to the west of the Lower Colorado River (Fig. 1b; Table S4, Supporting information). In contrast, Southern Rocky and Colorado Plateau breeders winter from El Salvador to Costa Rica, and migrate north through the central US, while eastern breeders winter primarily in the Yucatan and southern Costa Rica and migrate north through eastern Texas and New York (Fig. 1b; Table S4, Supporting information). Unlike the presence of strong connectivity across much of the range, Wilson's warblers breeding from Alaska to Alberta were identified in all but one of our migratory stopover sites and across all wintering areas, apart from western Mexico and southern Baja (Fig. 1b, all but location g; Table S4, Supporting information).

Assignment of migrants collected in a time series from Cibola, AZ, revealed a strong temporal pattern in stopover site use across the spring migratory period (Fig. 1c; Table 3). Birds en route to coastal California arrived first (week of March 22), followed by birds en route to the Pacific Northwest (week of March 29), the Sierra Nevada (week of April l5) and Alaska to Alberta (week of April 26). Only a few individuals migrating through the stopover site were identified as Sierra Nevada breeders (3 per year), while no populations breeding in the Southern Rocky and Colorado Plateau and Eastern US were identified migrating through the stopover site. Temporal patterns in the arrival of spring migrants were replicated across the years 2008 and 2009 and were consistent regardless of known differences in migration patterns by age and sex (Yong *et al.* 1998).

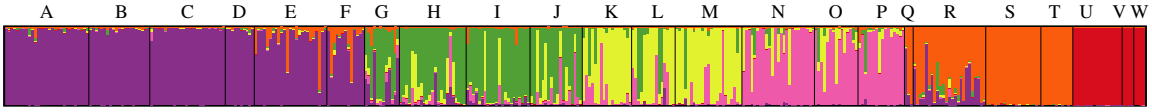
Discussion

Full-life cycle conservation of declining migrant species has been hindered by lack of an efficient and comprehensive method for identifying the strength of migratory connections across the annual cycle. Here, we demonstrate how high-resolution molecular markers can be applied towards full-life cycle conservation of a migrant songbird, the Wilson's warbler, with a degree of reliability and efficiency that has not been demonstrated using previous methods. By harnessing recent advances in next-generation sequencing, we show that 96 highly divergent SNPs selected from a large pool of candidates (~450 000 SNPs) can be used to identify genetically distinct groups on spatial scales that are informative for regional conservation planning. Our analysis indicates that the power to identify individuals to breeding populations is high (80–100%) and that reliable genotypes can be attained from 96% of feathers collected noninvasively from established bird monitoring stations across North and Central America. Because of the biallelic nature of the SNPs in our panel, the markers described here are easier to validate and standardize across laboratories than isotope and other genetic methods (Hobson *et al.* 2012; Kraus *et al.* 2014) and once the SNP assays have been developed, it is possible to genotype ~300 birds per day for ~$10.00/individual in a well-equipped molecular laboratory. Furthermore, the depth of sampling across space and time that is possible using high-resolution genetic markers is currently infeasible using existing extrinsic tracking devices due to cost and weight restrictions and the need to recapture individuals to collect the information (Arlt *et al.* 2013; Bridge *et al.* 2013). Overall, the resolution, efficiency and cost, combined with the ease of feather collection in collaboration with existing bird monitoring/banding infrastructure, make high-resolution genetic markers a broadly applicable method for widespread monitoring of declining migrant species.

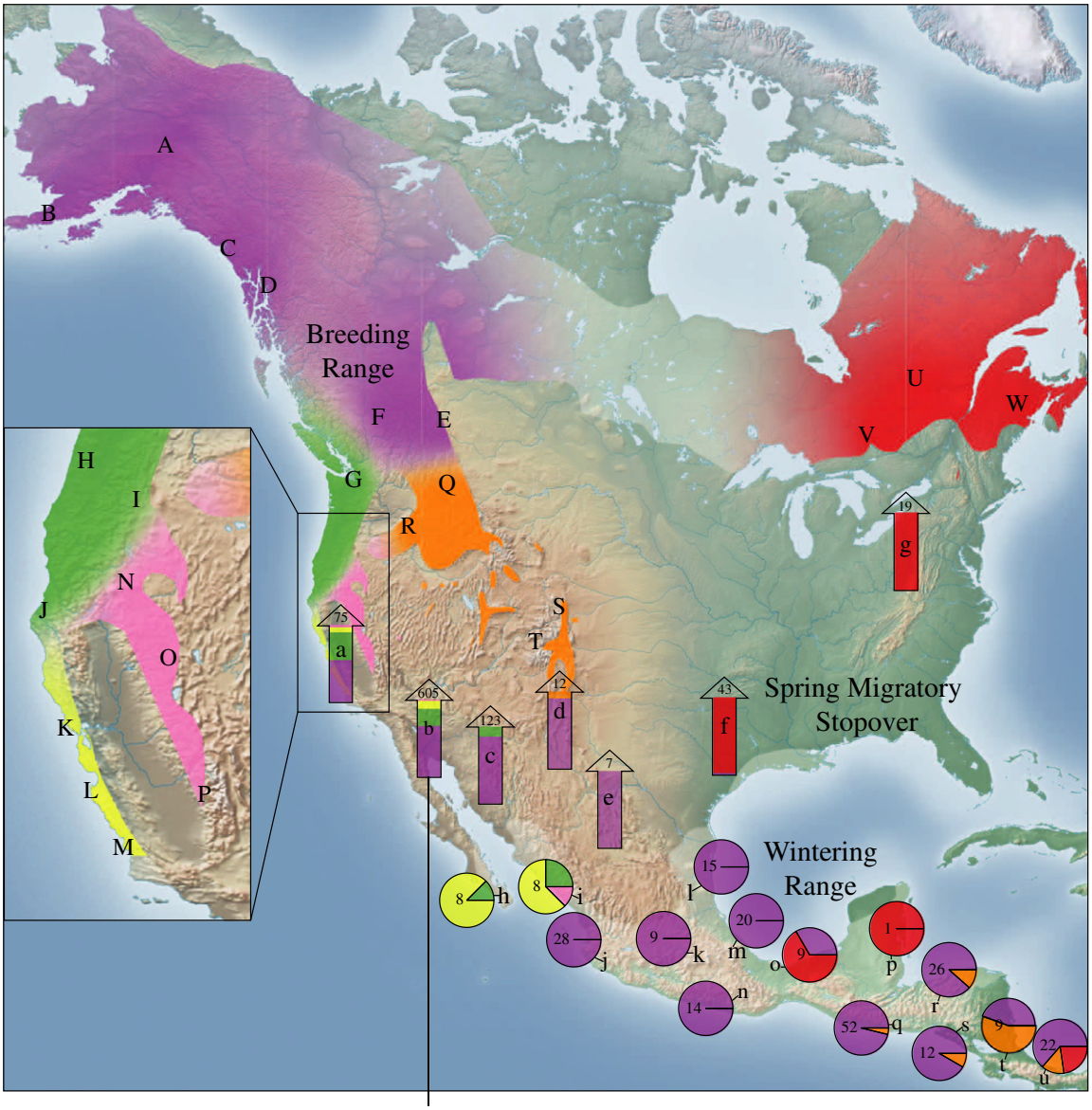
In the last several years, advances in sequencing technology have made it possible to begin characterizing loci under selection or linked to genes under selection rather than neutral genetic variation and such data are transforming our ability to delineate conservation units at finer spatial scales (Bonin *et al.* 2007; Funk *et al.* 2012). The Wilson's warbler provides a good example

Fig. 1 Migratory connections in the Wilson's warbler identified using SNP-based genetic markers. (A) Results from STRUCTURE showing six genetically distinct populations across the breeding grounds. Capital letters (A–W) refer to the location of breeding populations depicted on the map in B as well as listed in Table 1. (B) Spatially explicit population structure across the annual cycle. The colours across the breeding range represent the results from GENELAND which were postprocessed using R so that the density of each colour reflects the relative posterior probability of membership for each pixel to the most probable of the six different genetic clusters (see text). The results were clipped to the species distribution map (NatureServe 2012). Lower case letters (a–g) represent the location of wintering and spring migratory samples (Table 1). Pie charts indicate the proportion of wintering individuals assigned to each breeding group with the number of individuals listed at the centre of each pie. Arrows represent the proportion of migrants assigned to each breeding group with the numbers of individuals listed at the top of the arrows. (C) The proportion of individuals assigned to each breeding population across spring migration of 2008 and 2009. Numbers in the centre of the pies refer to sample sizes and the data are grouped by week with the date representing the midweek date in a nonleap year.

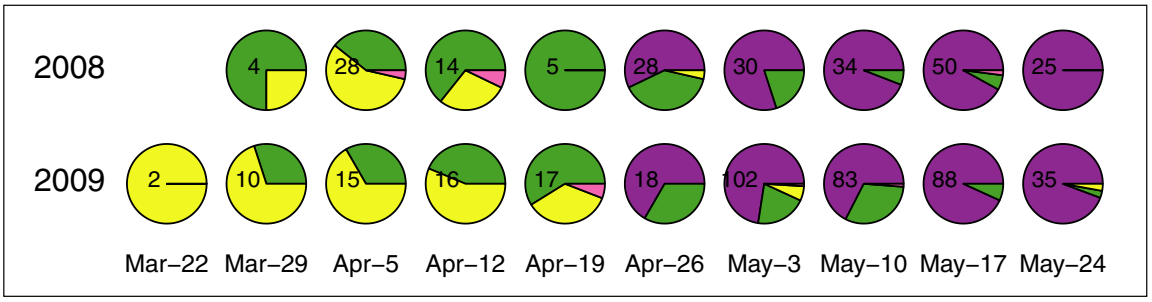
(A) Breeding population structure



(B) Spatially explicit population structure



(C) Population structure across time, Cibola CA (b, above)



of the power of a focusing on highly divergent loci because it has been the focus of numerous population genetic studies in the past decade using a variety of genotyping methods (mtDNA, microsatellites, and AFLPs), but none have yielded the depth and clarity of information on population genetic structure and migratory connectivity that we documented herein (Yong *et al.* 1998; Kimura *et al.* 2002; Clegg *et al.* 2003; Paxton *et al.* 2007, 2013; Irwin *et al.* 2011; Rundel *et al.* 2013). While previous methods identified two genetically distinct groups, an eastern and a western group, with some weak support for a third group in the region of Colorado (Kimura *et al.* 2002; Clegg *et al.* 2003; Irwin *et al.* 2011), we find strong support for six genetically distinct groups across the breeding range (Fig. 1). The distributions of the six groups are approximately concordant with subspecies distributions (*C. p. pileolata*, *C. p. pusilla*, and *C. p. chryseola*; Ammon & Gilbert 1999), with the exception that the western subspecies, *C. p. chryseola*, is further differentiated into three groups, a Sierra Nevada, a Pacific Northwest and a Coastal California group and the Alaska to Rocky Mt. subspecies (Fig. 1, pink, green and yellow), *C. p. pileolata*, is further differentiated into two groups, an Alaska to Alberta group (Fig. 1, purple) and a Southern Rocky Mountain group (Fig. 1b, orange). It is important to note that the predicted breeding distributions of the six groups are based upon a fairly crude spatio-genetic model (GENELAND) and additional sampling is needed along the hypothesized boundaries to clarify the actual shape and location of the genetic transitions (Guillot *et al.* 2008). Furthermore, given that we intentionally biased our marker selection towards highly divergent loci, it is not surprising that our estimates of genetic distance (F_{ST}) between groups are higher than in previous comparisons using putatively neutral genetic markers (global F_{ST} reported here = 0.179; global F_{ST} reported using microsatellites = 0.035; Clegg *et al.* 2003). Overall, this work supports the idea that a genomewide approach focused on high-resolution SNP markers rather than neutral genetic variation can be useful for delineating conservation units at finer spatial scales (Bonin *et al.* 2007; Funk *et al.* 2012).

In addition to resolving breeding populations at finer spatial scales, our results also reveal new patterns of migratory connectivity across time and space that are much richer and stronger than previously recognized. For example, while a strong connection between birds breeding in Coastal California and wintering in Southern Baja, MX was shown previously (Rundel *et al.* 2013), here we show that Wilson's warblers breeding in Coastal California share their wintering area in southern Baja with Pacific Northwest breeders and that both of these groups also winter to the east of Baja in Sinaloa, MX, with Sierra Nevada breeders. Furthermore, our results indicate that western breeders from all three

Table 2 Assignment of Wilson's warblers of known origin back to breeding population using GSI_SIM. Population names are listed in Table 1 and the colours indicate the predicted genetic group of origin (Fig. 1)

Population (Fig. 1, Table 1)	Alaska to Alberta	Pacific Northwest	Coastal California	Sierra	Rocky Mountain	Eastern
A	29	0	0	0	0	0
B	21	0	0	0	0	0
C	26	0	0	0	0	0
D	10	0	0	0	0	0
E	24	0	0	0	1	0
F	9	0	0	0	4	0
G	2	9	1	0	0	0
H	0	20	3	0	0	0
I	0	20	1	1	0	0
J	0	15	0	3	0	0
K	0	2	14	1	0	0
L	0	3	11	1	0	0
M	0	1	19	3	0	0
N	0	2	2	21	0	0
O	0	1	2	12	0	0
P	0	1	0	15	0	0
Q	1	0	0	0	2	0
R	6	0	0	0	19	0
S	0	0	0	0	19	0
T	0	0	0	0	11	0
U	0	0	0	0	0	17
V	0	0	0	0	0	4
W	0	0	0	0	0	4

Table 3 Assignment of Wilson's warblers migrating through Cibola, CA back to genetically distinct breeding groups using GSI_SIM. The data correspond to the information presented in Fig. 1c

Midweek Date*	Week	Alaska to Alberta	Pacific NW	Coastal CA	Sierra Nevada	Rocky Mt.	Eastern
Year 2008							
21-March	11	0	0	0	0	0	0
28-March	12	0	3	1	0	0	0
4-April	13	0	11	16	1	0	0
11-April	14	0	9	4	1	0	0
18-April	15	0	5	0	0	0	0
25-April	16	16	11	1	0	0	0
2-May	17	24	6	0	0	0	0
9-May	18	32	2	0	0	0	0
16-May	19	46	3	0	1	0	0
23-May	20	25	0	0	0	0	0
Year 2009							
22-March	11	0	0	2	0	0	0
29-March	12	0	3	7	0	0	0
5-April	13	0	5	10	0	0	0
12-April	14	0	6	10	0	0	0
19-April	15	0	10	6	1	0	0
26-April	16	12	6	0	0	0	0
3-May	17	74	21	6	1	0	0
10-May	18	56	25	1	1	0	0
17-May	19	82	6	0	0	0	0
24-May	20	33	1	1	0	0	0

*Dates represent the midweek date in a nonleap year.

groups migrate north along the Pacific Flyway, with Coastal California and Sierra Nevada breeders found west of the Lower Colorado River. Our results also reveal that breeders from the Southern Rocky Mountains and Colorado Plateau occupy a restricted El Salvador-to-Costa Rica wintering distribution and migrate North along the Central Flyway, while eastern breeders migrate North through eastern Texas and New York. Overall, our results support the idea that screening large volumes of individuals using high-resolution genetic markers provides a highly effective and comprehensive method for resolving the strength of migratory connections across time and space.

One of the central challenges in migrant conservation is that population declines and conservation planning often occur at regional spatial scales, but our knowledge of population structure and migratory connectivity is often limited to species-wide range maps. Here, we show that high-resolution molecular markers designed for the Wilson's warbler can be used to align the spatial scale of regional population declines with the spatial scale of population genetic structure and migratory connectivity, making it possible for the first time to assess drivers of regional population trends within a full-life cycle framework. For example, in the case of the Wilson's warbler, BBS data suggest that Sierra Nevada breeders are experiencing strong population declines (BBS $\text{Trend}_{\text{sierra}} = 4.71$, 95% CI = −6.41, −2.85), while Pacific Northwest and Coastal California breeders are declining less severely or remaining stable (BBS $\text{Trend}_{\text{Pacific\_Northwest}} = -1.96$, 95% CI = −2.54, −1.31; BBS $\text{Trend}_{\text{Coastal\_California}} = -0.49$, CI = −1.62, 0.84). The fact that the three genetically distinct groups occupy separate breeding ranges, but mix on their wintering grounds and at migratory stopover sites suggests that declines in Sierra Nevada breeders may be mostly driven by factors on the breeding grounds. In the future, our conclusions regarding where migrants are most limited could be strengthened through the use of migratory network models that could incorporate the genetic patterns described herein with information on habitat and demographic change through time (Norris & Taylor 2006; Stanley *et al.* 2014).

Migratory passerines spend approximately a quarter of their year en route between breeding and wintering areas, but relatively little is known about the biology and behaviour of migrants during the migratory phase of their annual cycle (Faaborg *et al.* 2010b). The availability and quality of habitat at stopover sites could have significant effects on populations, but determining the extent to which physiological and ecological demands experienced during migration may limit populations is often contingent upon knowledge of an individual's ultimate destination (Faaborg *et al.* 2010a,b). Here, we genotype samples collected in a time series from a stopover site

near Cibola, AZ, and demonstrate how high-resolution genetic markers can be used to identify the ultimate destination of birds captured en route to their breeding grounds (Fig. 1b, c; location b). Breaking down the results by week revealed distinct waves of migrants, with Coastal California breeders arriving first (March 22–29), followed by Pacific Northwest and Sierra Nevada breeders (March 29–April 5), and Alaska-to-Alberta breeders arriving significantly later (April 19–26). While differences in the timing of migration in Wilson's warblers have been suggested in the past based upon changes in the frequency of haplotypes or isotopic signatures (Paxton *et al.* 2007, 2013), this is the first time that it has been possible to attain individual-level assignments of migrants back to breeding populations, bringing a new level of clarity to our understanding of stopover site use through time. The differences in migratory timing documented herein are particularly interesting in the light of the potential for divergence in migratory behaviour to reduce gene flow across contact zones between distinct migratory forms (Bearhop *et al.* 2005; Ruegg *et al.* 2012, 2014). Future research will focus on the extent to which differences in migratory timing may drive local adaption within distinct eco-geographic regions across the North American breeding range.

While our results suggest that high-resolution molecular markers surpass previous genetic markers in terms efficiency and resolution, our conclusions could be further strengthened by the inclusion of additional data and analyses. For example, the robustness of the patterns described here varies depending upon the sample size at each location and in some locations, such as in Belize and many of the migratory stopover sites (Fig. 1b, locations l, d, e, f, g), additional sampling across time and space is needed. In addition, while our assignment probabilities are very high for an intrinsic marker (80–100%), there is a potential for incorrect assignments, particularly between the three western groups (Coastal California, Pacific Northwest and the Sierras) were admixture is likely (Table 2). Similarly, there are large regions on the breeding grounds that could not be distinguished using our markers, such as birds breeding from Alberta to Alaska (purple, Fig. 1b). In the future, the addition of more genetic loci as well as the addition of isotopic markers and statistical methods for combining both sources of data into a single statistical framework will help further resolve populations across the range (Rundel *et al.* 2013).

It is widely accepted that a better understanding of migratory connections across the annual cycle is needed to fully comprehend the forces that shape the life histories of migratory animals (Webster *et al.* 2002; Webster & Marra 2004). Here, we demonstrate how high-resolution genetic markers can be applied to understand patterns of migratory connectivity in the Wilson's warbler with a level of efficiency and reliability that has not previously been demonstrated using other methods. The resulting information on fine-scale population genetic structure, region-specific migratory connections and migration timing provides a powerful framework from which to base full-life cycle conservation of declining songbird species and opens new opportunities for our ability to understand the ecology and evolution of migratory animals.

Acknowledgements

We would like to thank J.C. Garza at the Southwest Fisheries Science Center for use of laboratory space and equipment and the qb3 Genomics Sequencing Laboratory for their assistance with sequencing. This research was supported by a grant to K.C. Ruegg from the California Institute for the Energy and the Environment (POEA01-Z01), a donation from Margery Nicolson, a grant to T.B. Smith from The Turner Foundation, EPA (RD-83377801), and a grant to F. Moore from the National Science Foundation (IOS-0844703). We would also like to thank J. Boone and T. Atwood for their assistance with the laboratory and bioinformatics components of the assay design, the many LaMNA, MAPS and MoSI station operators who contributed avian tissue samples, C.J. Ralph, L. West, D. Kaschube, P. Pyle, J. Saracco and R. Taylor for coordinating sampling efforts and R. Fleischer at CCEG for contributing genetic material. We are grateful to the USDI Fish and Wildlife Service, the National Park Service, and USDA Forest Service for funding to help operate LaMNA, MAPS and MoSI stations that provided feather samples for this work as well as field crews and staff at Cibola National Wildlife Refuge for help with sample collection and logistics.

References

Altizer S, Bartel R, Han BA (2011) Animal migration and infectious disease risk. *Science*, **331**, 296–302.

Ammon EM, Gilbert WM (1999) Wilson's Warbler (*Cardellina pusilla*). In: *The Birds of North America Online* (ed. Poole A). Cornell Lab of Ornithology, Ithaca, Greece Available from http://bna.birds.cornell.edu.oca.ucsc.edu/bna/species/478.

Anderson EC (2010) Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. *Molecular Ecology Resources*, **10**, 701–710.

Anderson EC, Waples RS, Kalinowski ST (2008) An improved method for predicting the accuracy of genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 1475–1486.

Arlt D, Low M, Part T (2013) Effect of geolocators on migration and subsequent breeding performance of a long-distance passerine migrant. *PLoS One*, **8**, e82316.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, **3**, e3376.

Bearhop S, Fiedler W, Furness RW *et al.* (2005) Assortative mating as a mechanism for rapid evolution of a migratory divide. *Science*, **310**, 502–504.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996–2004) GENETIX *4.05, Logiciel sous WindowsTM pour le Genetique does Populations*. Laboratoire Genome, Populations, Interactions, Universite de Montpellier II, Montepellier, France.

Bivand R, Keitt T, Rowlingson B (2014) RGDAL: *Bindings for the Geospatial Data Abstraction Library*. R *Package Version 0.8-16*. Available from http://www.pwrc.usgs.gov/bbl/homepage/introduction.cfm.

Bonfil R, Meyer M, Scholl MC *et al.* (2005) Transoceanic migration, spatial dynamics, and population linkages of white sharks. *Science*, **310**, 100–103.

Bonin A, Nicole F, Pompanon F, Miaud C, Taberlet P (2007) Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. *Conservation Biology*, **21**, 697–708.

Bridge ES, Kelly JF, Contina A *et al.* (2013) Advances in tracking small migratory birds: a technical review of light-level geolocation. *Journal of Field Ornithology*, **84**, 121–137.

Camacho C (2009) BLAST+: Architecture and applications. *BMC Bioinformatics*, **10**, 421.

Clegg SM, Kelly JF, Kimura M, Smith Thomas B (2003) Combining genetic markers and stable isotopes to reveal population connectivity and migration patterns in a Neotropical migrant, Wilson's warbler (*Wilsonia pusilla*). *Molecular Ecology*, **12**, 819–830.

Croxall JP, Silk JR, Phillips RA, Afanasyev V, Briggs DR (2005) Global circumnavigations: tracking year-round ranges of nonbreeding albatrosses. *Science*, **307**, 249–250.

Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

Etter PD, Preston JL, Bassham S, Cresko WA, Johnson EA (2011) Local de novo assembly of RAD paired-end contigs using short sequencing reads. *PLoS One*, **6**, e18561.

Faaborg J, Holmes RT, Anders AD *et al.* (2010a) Recent advances in understanding migration systems of New World land birds. *Ecological Monographs*, **80**, 3–48.

Faaborg J, Holmes RT, Anders AD *et al.* (2010b) Conserving migratory land birds in the new world: do we know enough? *Ecological Applications*, **20**, 398–418.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, **27**, 489–496.

Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261–1280.

Guillot G, Santos F, Estoup A (2008) Analysing georeferenced population genetics data with GENELAND: a new algorithm to deal with null alleles and a friendly graphical user interface. *Bioinformatics*, **24**, 1406–1407.

Gustafson ME, Hildenbrand J (1999) *Bird Banding Laboratory Homepage, version 07-10-2005*. Available from http://www.pwrc.usgs.gov/bbl/homepage/introduction.cfm.

Hess JE, Matala AP, Narum SR (2011) Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Molecular Ecology Resources*, **11**(Suppl 1), 137–149.

Hijmans RJ (2014) *Raster: Geographic Data Analysis and Modeling*. R *Package Version 2.2-31*. Available from http://R-Forge.R-project.org/projects/ra ster/.

Hobson KA, Van Wilgenburg SL, Wassenaar LI, Larson K (2012) Linking hydrogen (delta2H) isotopes in feathers and precipitation: sources of variance and consequences for assignment to isoscapes. *PLoS One*, **7**, e35137.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**, 117–122.

Irwin DE, Irwin JH, Smith TB (2011) Genetic variation and seasonal migratory connectivity in Wilson's warblers (*Wilsonia pusilla*): species-level differences in nuclear DNA between western and eastern populations. *Molecular Ecology*, **20**, 3102–3115.

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*, **23**, 1801–1806.

Jonzen N, Linden A, Ergon T *et al.* (2006) Rapid advance of spring arrival dates in long-distance migratory birds. *Science*, **312**, 1959–1961.

Kelly JF, Ruegg KC, Smith TB (2005) Combining isotopic and genetic markers to identify breeding origins of migrant birds. *Ecological Applications*, **15**, 1487–1497.

Kimura M, Clegg SM, Lovette IJ *et al.* (2002) Phylogeographical approaches to assessing demographic connectivity between breeding and overwintering regions in a Nearctic-Neotropical warbler (*Wilsonia pusilla*). *Molecular Ecology*, **11**, 1605–1616.

Kraus RH, van Hooft P, Megens HJ *et al.* (2013) Global lack of flyway structure in a cosmopolitan bird revealed by a genome wide survey of single nucleotide polymorphisms. *Molecular Ecology*, **22**, 41–55.

Kraus RH, vonHoldt B, Cocchiararo B *et al.* (2014) A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. *Molecular Ecology Resources*. doi: 10.1111/1755-0998.12307.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Li H, Handsaker B, Wysoker A *et al.* (2009) The sequence alignment/map format and SAMTOOLS. *Bioinformatics*, **25**, 2078–2079.

Loss SR, Will T, Marra PP (2013) The impact of free-ranging domestic cats on wildlife of the United States. *Nature Communications*, **4**, 1396.

Louis EJ, Dempster ER (1987) An exact test for Hardy–Weinberg and multiple alleles. *Biometrics*, **43**, 805–811.

Lovette IJ, Clegg SM, Smith TB (2004) Limited utility of mtDNA markers for determining connectivity among breeding and overwintering locations in three Neotropical migrant birds. *Conservation Biology*, **18**, 156–166.

Marra PP, Hobson KA, Holmes RT (1998) Linking winter and summer events in a migratory bird by using stable-carbon isotopes. *Science*, **282**, 1884–1886.

Metzker ML (2010) Sequencing technologies—the next generation. *Nature Reviews Genetics*, **11**, 31–46.

NatureServe BIa (2012) *Bird Species Distribution Maps of the World*. NatureServe, Arlington, Virginia.

Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D (2009) Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology*, **18**, 3128–3150.

Nielsen EE, Cariani A, Mac Aoidh E *et al.* (2012) Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature Communications*, **3**, 851.

Norris DR, Taylor CM (2006) Predicting the consequences of carry-over effects for migratory populations. *Biology Letters*, **2**, 148–151.

Paxton KL, Van Ripper C III, Theimer TC, Paxton EH (2007) Spatial and temporal migration patterns of Wilson's warbler (*Wilsonia pusilla*) in the southeast as revealed by stable isotopes. *The Auk*, **124**, 162–175.

Paxton KL, Yau M, Moore FR, Irwin D (2013) Differential migratory timing of western populations of wilson's warblers (*Cardellina pusilla*) revealed by mitochondrial DNA and stable isotopes. *The Auk*, **130**, 689.

Pebesma EJ, Bivand RS (2005) Classes and methods for spatial data in R. *R News*, **5**. Avialble from http://cran.r-project.org/doc/Rnews/.

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

Rappole JH (1995) *The Ecology of Migrant Birds—A Neotropical Perspective*. Smithsonian Institution Press, Washington, District of Columbia.

Robbins CS, Sauer JR, Greenberg R, Droege S (1989) Population declines in North American birds that migrate to the Neotropics. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 7658–7662.

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

Rousset F (2008) GENEPOP'007: A complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Rubenstein DR, Chamberlain CP, Holmes RT *et al.* (2002) Linking breeding and wintering ranges of a migratory songbird using stable isotopes. *Science*, **295**, 1062–1065.

Ruegg K, Anderson EC, Slabbekoorn H (2012) Differences in timing of migration and response to sexual signalling drive asymmetric hybridization across a migratory divide. *Journal of Evolutionary Biology*, **25**, 1741–1750.

Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB (2014) A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology*, **23**, 4757–4769.

Rundel CW, Wunder MB, Alvarado AH *et al.* (2013) Novel statistical methods for integrating genetic and stable isotope data to infer individual-level migratory connectivity. *Molecular Ecology*, **22**, 4163–4176.

Sauer JR, Hines JE, Fallon JE *et al.* (2012) *The North American Breeding Bird Survey, Results and Analysis 1966–2011*. USGS Patuxent Wildlife Research Center, Laurel, Maryland.

Segelbacher G (2002) Noninvasive genetic analysis in birds: testing reliability of feather samples. *Molecular Ecology Notes*, **2**, 367–369.

Seutin G, White BN, Boag PT (1991) Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology*, **69**, 82–90.

Sillett TS, Holmes RT, Sherry TW (2000) Impacts of a global climate cycle on population dynamics of a migratory songbird. *Science*, **288**, 2040–2042.

Smith TB, Wayne RK, Marra PP *et al.* (2003) A call for feather sampling. *The Auk*, **120**, 218–221.

Stanley CQ, McKinnon EA, Fraser KC *et al.* (2014) Connectivity of wood thrush breeding, wintering, and migration sites based on range-wide tracking. *Conservation Biology*. [online].

Stutchbury BJ, Tarof SA, Done T *et al.* (2009) Tracking long-distance songbird migration by using geolocators. *Science*, **323**, 896.

Team RC (2014) R: *A Language and Environment for Statistical Computing*. Vienna, Austria. http://www.R-project.org/

Warren WC, Clayton DF, Ellegren H *et al.* (2010) The genome of a songbird. *Nature*, **464**, 757–762.

Webster MS, Marra PP (2004) *Birds of Two Worlds—The Ecology and Evolution of Migration*. The Johns Hopkins University Press, Baltimore, Maryland.

Webster MS, Marra PP, Haig SM, Bensch S, Holmes RT (2002) Links between worlds: unraveling migratory connectivity. *Trends in Ecology & Evolution*, **17**, 76–83.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wunder MB, Kester CL, Knopf FL, Rye RO (2005) A test of geographic assignment using isotope tracers in feathers of known origin. *Oecologia*, **144**, 607–617.

Yong W, Finch DM, Moore FR, Kelly JF (1998) Stopover ecology and habitat use of migratory Wilson's Warblers. *The Auk*, **115**, 829–842.

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, **18**, 821–829.

K.C.R., E.C.A. and T.B.S. conceived of the study and K.C.R. wrote the majority of the manuscript and oversaw the majority of the analyses. E.C.A. wrote the scripts for the population genomic analyses and the R package for the creation of the map. K.L.P. and F.M. contributed ideas and genetic material for the analysis of migrants from Cibola, AZ. V.A. conducted and helped analyse data for the SNP screening and S.L. assisted with feather sample organization, extraction and the analysis of genotyping reliability scores. R.B.S. and D.F.D. facilitated the collection of feather samples in collaboration with the MAPS and the MoSI banding networks.

Data accessibility

The data and scripts for population genomic analysis (from SNP selection to estimation of population genomic parameters) are currently available on GITHUB (https://github.com/eriqande/wiwa-popgen) and are Dryad (doi:10.5061/dryad.j5d33).

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 Methods.

Table S1 RAD sequencing results for SNP ascertainment panel.

Table S2 SNPType Assay information.

Table S3 F_{ST} calculated according to Weir and Cockerham.

Table S4 Assignment of wintering & migrant birds to breeding regions using GSI_SIM.

Fig. S1 Results of the population genetic analysis for Wilson's warblers across the breeding range using STRUCTURE.