SCISPACE
formerly Typeset

Open access • Journal Article • DOI:10.1111/MEC.12977

Mapping migration in a songbird using high-resolution genetic markers

— Source link

Kristen C. Ruegg, Kristen C. Ruegg, Eric C. Anderson, Eric C. Anderson ...+9 more authors

Institutions: University of California, Los Angeles, University of California, Santa Cruz, National Marine Fisheries Service, University of Hawaii at Hilo ...+3 more institutions

Published on: 01 Dec 2014 - Molecular Ecology (Mol Ecol)

Topics: Conservation genetics and Flyway

Related papers:

- Inference of population structure using multilocus genotype data
- Links between worlds: unraveling migratory connectivity
- Linking Winter and Summer Events in a Migratory Bird by Using Stable-Carbon Isotopes
- Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study.
- Harnessing genomics for delineating conservation units

Share this paper:

View more about this paper here: https://typeset.io/papers/mapping-migration-in-a-songbird-using-high-resolution-1f3tqttrtw

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Mapping migration in a songbird using high-resolution genetic markers

Kristen Ruegg[1,2], Eric C. Anderson[3,4], Kristina L. Paxton[5.6], Vanessa Apkenas[3], Sirena Lao[1], Rodney B. Siegel[7], David F. DeSante[7], Frank Moore[6] and Thomas B. Smith[1,8]

[1]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

[2]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, CA 95060, USA; e-mail: kruegg@ucsc.edu; phone: 510-292-5099

[3]Southwest Fisheries Science Center, National Marine Fisheries Service, 110 Shaffer Road, Santa Cruz, CA 95060, USA

[4]Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA 95060, USA

[5] Department of Biological Sciences, University of Southern Mississippi, Hattiesburg, MS 39406

[6] Department of Biology, University of Hawaii, Hilo, HI 96720, USA

[7] The Institute for Bird Populations, PO Box 1346, Point Reyes Station, CA 94956, USA

[8] Department of Ecology and Evolutionary Biology, University of California, Los Angles, CA, 90095, USA

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Neotropical 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 connecting populations across the annual cycle. Here we show how high-resolution genetic markers can be used to identify populations 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 single nucleotide polymorphisms (SNPs) 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, accurate, and amenable to high throughput screening than previously described tracking techniques, making them broadly applicable to large-scale monitoring and conservation of migratory organisms.

Introduction

Over half of the Neotropical migrant bird species found breeding in North America have shown marked declines in abundance over the last several decades (Robbins 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 stage birds are subject to a number of disturbances including habitat loss, collisions with wind turbines and cell

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

54 phone towers, predation by house cats, exposure to disease, and the increasing effects of
55 global climate change (Altizer *et al.* 2011; Jonzen *et al.* 2006; Loss *et al.* 2013).
56 However, without the ability to connect populations across the annual cycle it is difficult
57 to assess the impact of local stressors on population declines. Historically, efforts to map
58 songbird migration patterns relied on recovery of individual birds previously captured
59 and tagged with bird bands; however, this approach has met with limited success for
60 small-bodied songbirds because recapture rates of birds away from their original banding
61 sites are often very low (< 1 in 10,000) (Faaborg *et al.* 2010b; Gustafson & Hildenbrand
62 1999). More recently, geo-locators, small tracking devices that record information on
63 ambient light levels to estimate an individuals location, have increased our knowledge of
64 the migratory pathways in many songbird species (Stutchbury *et al.* 2009), but remain
65 impractical for most large-scale applications (1000's of individuals) due to cost, weight
66 restrictions, and the need to recover individuals to collect data from the devices (Arlt *et*
67 *al.* 2013; Bridge *et al.* 2013). Alternatively, genetic and isotopic markers that use
68 information contained within the feathers to pinpoint an individuals population of origin
69 have broad appeal because they are cost-effective, noninvasive, and do not require
70 recapture (Kelly *et al.* 2005; Rubenstein *et al.* 2002; Rundel *et al.* 2013), but have been
71 plagued in the past by low resolution and/or technical issues related to working with
72 feathers (Lovette *et al.* 2004; Segelbacher 2002; Wunder *et al.* 2005). Thus, there
73 remains a need for a broadly applicable tracking method that can be used to resolve
74 populations on spatial scales that are informative for assessing drivers of regional
75 population declines.
76

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

77 In the last several years, genome sequencing has revolutionized the field of molecular
78 ecology, resulting in new technologies that can be applied to molecular tagging of wild
79 populations (Davey *et al.* 2011). Genome reduction techniques, such as Restriction Site
80 Associated DNA sequencing (RAD-seq), can be used to sequence multiple individuals
81 across a large fraction of the genome and identify hundreds of thousands of genetic
82 markers that are useful for distinguishing populations (Baird *et al.* 2008). One type of
83 genetic marker that can be identified from genomic sequence data is a Single Nucleotide
84 Polymorphism (SNP), DNA sequence variation occurring when a single nucleotide in the
85 genetic code – A, T, C, or G – differs between individuals or homologous chromosomes.
86 In particular, SNPs found within or linked to genes under selection often display elevated
87 allele frequencies and, as a result, can be targeted to reveal population structure at finer
88 spatial scales than is possible using neutral genetic markers (Nielsen *et al.* 2012; Nielsen
89 *et al.* 2009). Furthermore, SNP-specific assays designed to target small fragments of
90 sequence around the SNP loci of interest can be advantageous in cases where the DNA is
91 highly fragmented or available only in very small quantities, such as DNA from a single,
92 small passerine feather.
93
94 Here we develop high-resolution SNP markers for tracking populations of a migratory
95 bird, the Wilson's warbler, *Cardellina pusilla,* using a combination of Restriction Site
96 Associated DNA paired end sequencing (RAD-PE seq) and high throughput SNPtype™
97 Assay screening. The Wilson's warbler, a long-distance neotropical migratory bird with
98 a cross-continental breeding distribution (Ammon & Gilbert 1999), is particularly
99 appropriate as model for testing the efficacy of high-resolution molecular markers

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

100 because previous population genetic/connectivity studies on this species provide a solid
101 basis for comparison between methods (Clegg *et al.* 2003; Irwin *et al.* 2011; Kimura *et*
102 *al.* 2002; Paxton *et al.* 2007; Paxton *et al.* 2013; Rundel *et al.* 2013; Yong *et al.* 1998).
103 By harnessing recent advances in Next-Generation Sequencing we scan the genomes of
104 Wilson's warblers sampled from across the breeding range and identify a set of highly
105 divergent SNP loci with strong potential for population identification. We then develop
106 SNPtype™ Assays that target these highly divergent loci and use them to screen 1626
107 feather and blood samples collected from across the annual cycle in collaboration with
108 bird banding stations located across North and Central America. We illustrate how the
109 resulting region-specific migration map can be used to help identify drivers of regional
110 demographic trends and inform studies of migrant stopover ecology.
111
112 **Methods**
113 *(a) Sample collection*
114 Collection of 1648 feather and blood samples (22 samples for the SNP ascertainment
115 panel and 1626 for the SNP screening panel) from 68 locations across the breeding,
116 wintering and migratory range was made possible through a large collaborative effort
117 with bird banding stations within and outside of the Monitoring Avian Productivity and
118 Survivorship (MAPS), the Landbird Monitoring of North America (LaMNA), and the
119 Monitoreo de Sobrevivencia Invernal (MoSI) networks (Table 1). Genetic samples,
120 consisting of the tip of one outer rectrix or blood collected by brachial vein puncture and
121 preserved in lysis buffer (Seutin 1991), were purified using Qiagen DNeasy Blood and
122 Tissue Kit and quantified using a NanoDrop™ Spectrophotometer (Thermo Scientific,

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

123 Inc) (Smith *et al.* 2003). Breeding (June 10 – July 31), migratory (March 1 – May 31),
124 and wintering (December 1 – February 28) samples were collected and categorized into
125 groups based on collection date, signs of breeding (presence/size of a cloacal
126 protuberance), signs migration (extent of fat) and life history timetables for the Wilson's
127 warbler (Ammon & Gilbert 1999). To assess migratory stopover site use through time,
128 686 of the 1648 samples from a stopover site located on the Lower Lower Colorado
129 River, near the town Cibola, AZ, were collected using consistent effort, daily, passive
130 mist-netting from March 22 – May 24, across the years 2008 and 2009 (Table 1).
131
132 *(b) SNP discovery*
133 To identify SNPs useful for distinguishing genetically distinct regions across the breeding
134 range of the Wilson's warbler, an ascertainment panel of 22 individuals was selected to
135 represent the range of phylogenetic variation known in the species, including all 3
136 recognized subspecies (Ammon & Gilbert 1999; Kimura *et al.* 2002). Five individuals
137 from each of five regions were included in the ascertainment panel, except for from the
138 Southwestern region where samples were limited to 2 individuals (SI Table 1). Purified
139 extractions from blood samples were quantified using Quant-iT™ PicoGreen® dsDNA
140 Assay Kit (Invitrogen Inc), and Restriction Site Associated DNA Paired-end (RAD-PE)
141 libraries containing individually barcoded samples were prepared at Floragenex, Inc.
142 according to Baird *et al.* (2008) and Ruegg *et. al.* (Ruegg *et al.* 2014) (SI Methods).
143 RAD-PE sequencing made it possible to build longer contigs (~300bp) from short read,
144 100bp Illumina HiSeq2000 data in order to improve downstream bioinformatics and

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

145 provide adequate flanking sequence around SNPs for assay development (Etter *et al.*
146 2011).
147
148 Samples from each isolate were sequenced on an Illumina HiSeq2000 (Illumina, San
149 Diego, CA) using paired-end 100 bp sequencing reads. Paired-end sequences from each
150 sample were collected, separated by individual, stripped of barcodes, trimmed to 70 bp,
151 scrubbed of putative contaminant and high-copy-number-sequences and filtered to
152 include only those with a Phred score ≥10. The sample with the greatest number of reads
153 passing the initial quality filter was used to create a reference set of RAD-PE contigs
154 against which sequences from other samples were aligned. To create the reference,
155 primary reads were clustered into unique RAD markers and the paired-end sequences
156 associated with each RAD tag were assembled *de novo* using Velvet (Zerbino & Birney
157 2008) into contigs ranging from 180 – 610 bp, with an average length of 300bp. Paired-
158 end reads from the remaining samples were aligned to this reference using Bowtie
159 (Langmead & Salzberg 2012) and SNPs were identified using the SAMtools software (Li
160 *et al.* 2009) with mpileup module under standard conditions.
161
162 To narrow our dataset to SNPs we could confidently use to assess population structure we
163 performed a second round of quality filtering and removed: (1) putative SNPs with no
164 variants and / or more than two alleles; (2) genotypes in individuals with a quality score
165 of < 30; (3) genotypes with < 8 reads in a homozygote or < 4 reads per allele in
166 heterozygotes; (4) putative SNPs that had suitable genotypes in < 12 out of the 17
167 samples from four western populations or < 5 out of the 5 samples from the eastern

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

168 population and (5) putative SNPs with < 40 bp of flanking sequence on either side. To
169 limit the chances of including linked markers genomic coordinates were attained by
170 mapping the remaining contigs to the closest, best annotated, songbird genome, the zebra
171 finch (*Taeniopygia guttata*) (Version 3.2.4; (Warren *et al.* 2010)) using BLAST+
172 (version 2.2.25).
173
174 To avoid the possibility of erroneous matches, the data was filtered to include only
175 contigs that aligned to the zebra finch genome with only a single hit and an E-value < 10^{-}
176 ^{40}. Because SNPs with large frequency differences are the most effective for identifying
177 populations, all SNPs that passed our second round of quality filters were ranked
178 according to frequency differences between the 5 regions (SI Table 2) and 150 SNPs
179 displaying the largest allele frequency differences between each of the 10 pairwise
180 comparisons were selected for conversion to SNPtype™ Assays (Fluidigm Inc). Before
181 making a final selection, we also considered factors such as: GC content (<65%), number
182 of genotypes per population, and average coverage at a SNP across all populations (SI
183 Table 2). An initial assay pre-screening panel was then performed and the assay pool
184 was further reduced to the 96 assays (the number that fit on a single 96.96 Fluidigm
185 Array) that could be genotyped most reliably (SI Table 2).
186
187 *(b) SNP Screening*
188 The Fluidigm Corporation EP1™ Genotyping System was used to genotype 96 SNP loci
189 using 94 individuals per run and 2 non-template controls. To avoid the potential for high-
190 grading bias (i.e. wrongly inflating the apparent resolving power of a group of loci for

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

191 population identification) (Anderson 2010), none of the 22 samples used in our original
192 ascertainment panel were included in the final SNP screening and population structure
193 analyses. To ensure amplification of low quality or low concentration DNA from
194 feathers, an initial pre-amplification step was performed according to the manufacturers
195 protocol using a primer pool containing 96 unlabeled locus-specific SNPtype primers (SI
196 Methods). PCR products were diluted 1:100 and re-amplified using fluorescently labeled
197 allele-specific primers. The results were imaged on an EP1 Array Reader and alleles
198 were called using Fluidigm's automated Genotyping Analysis Software (Fluidigm Inc)
199 with a confidence threshold of 90%. In addition, all SNP calls were visually inspected
200 and any calls that did not fall clearly into one of three clusters – heterozygote or either
201 homozygote cluster - were removed from the analysis. As DNA quality can affect call
202 accuracy, a stringent quality filter was employed and samples with >90 of 96 missing loci
203 were dropped. To assess the reliability of SNPtype assays for genotyping DNA from a
204 variety of sources (blood and feather extractions), the proportion of samples yielding
205 useable genotype data was calculated. Tests for linkage disequilibrium and conformance
206 to Hardy-Weinberg equilibrium (HWE) (Louis & Dempster 1987) were performed using
207 GENEPOP software, vers. 4.0 (Rousset 2008).
208
209 *(c) Population structure analysis*
210 While genetic differentiation (F_{ST}) is likely inflated because selected loci were not a
211 random sample from the genome, we calculated F_{ST} here for comparison to previous
212 genetic analysis. F_{ST} between all pairs of populations was calculated as θ (Weir &
213 Cockerham 1984), using the software GENETIX vers. 4.05 (Belkhir *et al.* 1996-2004)

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and the data were permuted 1000 times to determine significance. We used the program STRUCTURE ver. 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 correlated allele frequencies using a burn-in period of 50,000 iterations, a run length of 150,000. All scripts used for the STRUCTURE runs and subsequent population genomic analyses are located at https://github.com/eriqande/wiwa-popgen. To simplify 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 5,000 were discarded as burn-in. Post processing of the MCMC sample was done upon a 250 by 250 point grid that covered the breeding range of the species. Posterior probability of group membership estimates from GENELAND were visualized as transparency levels of different colors overlaid upon a base map from

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 (Bivand *et al.* 2014; Hijmans 2014; Pebesma & Bivand 2005; Team 2014) (see https://github.com/eriqande/wiwa-popgen). Thus, within each distinguishable group the transparency of colors 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 2010; Anderson *et al.* 2008). 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 genetically distinct group with the highest posterior probability.

Results

(a) SNP discovery

RAD-PE sequencing on 22 individuals from 5 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 (SI Table 1). The median depth of sequencing across all contigs within a library was 33x and the average Phred quality score per library was 35 (SI Table 1). Overall, 166,268 SNPs passed the second round of quality filters and 19,707 of those were candidates for

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

conversion into SNPtype™ Assays based upon the absence of variation in 40 base pairs 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 – 0.4 (SI Table 2). For contigs that could be mapped to the zebra finch genome with high confidence, the minimum distance between SNPs was 41KB and no two SNPs were selected from the same contig in order to avoid the possibility of linked markers (SI Table 2). In this study we refer to the final panel of 96 highly differentiated SNPs as high-resolution genetic markers.

(b) 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 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 an 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 ($p<0.0005$), suggesting

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

283 that loci were not physically linked even in cases where zebra finch genome coordinates
284 could not be attained.
285
286 *(c) Population Structure analysis*
287 An analysis of population genetic structure on the breeding grounds identified 6
288 genetically distinguishable groups: Alaska (purple, A - D), eastern North America (red,
289 U-W), the Southern Rockies and Colorado Plateau (orange, S, T), the Pacific Northwest
290 (green, G- J), Sierra Nevada (pink, N-P), and Coastal California (yellow, K-M) (Fig. 1a
291 & b). Pairwise F_{ST}'s between groups ranged from 0-0.68 with an overall F_{ST} of 0.179
292 (95% CI: 0.144 – 0.218). The strongest genetic differentiation was observed between
293 eastern and western groups (F_{ST} = 0.41 – 0.68) with strong genetic differentiation also
294 seen between the Southern Rockies and Colorado Plateau and all other groups (F_{ST} = 0.09
295 – 0.27; SI Table 3). The number of genetically distinct groups was set at 6 based upon
296 convergence between results from STRUCTURE (k=6, average ln P(X|K) = -33359),
297 GENELAND, and GSI_Sim (Fig. 1a&b; Table 2). While 7 genetically distinct groups
298 was also strongly supported by GENELAND and STRUCTURE (K=7, average ln P(X|K)
299 = -33286; SI Fig. 1), with sampling locations from British Columbia and Alberta (E and
300 F) forming a seventh group distinct from Alaska, the power to accurately assign
301 individuals to groups at k=7 decreased significantly using both STRUCTURE and
302 GSI_Sim (SI Fig. 1).
303
304 Leave-one-out cross validation using GSI_Sim indicated that the ability to correctly
305 assign individuals to groups was high, ranging from 80 - 100%. The eastern group had

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

306 the highest probability of correct assignment (100%), followed by Alaska to Alberta
307 (94%), the Southern Rockies and Colorado Plateau (92%), the Pacific Northwest (84%),
308 the Sierra Nevada (81%) and Coastal California (80%) (Fig. 1b; Table 2). The majority
309 of the incorrect assignments were between the Pacific Northwest, Sierra Nevada and
310 Coastal California. Subsequent assignment of migrant and wintering individuals to
311 genetically distinct breeding groups using GSI_sim indicated that Coastal California,
312 Sierra, and Pacific Northwest breeders winter in western Mexico and southern Baja, and
313 migrate north along the Pacific Flyway, with Coastal California and Sierra breeders
314 found to the west of the Lower Colorado River (Fig. 1b; SI Table 4). In contrast,
315 Southern Rocky and Colorado Plateau breeders winter from El Salvador to Costa Rica,
316 and migrate north through the central US, while eastern breeders winter primarily in the
317 Yucatan and southern Costa Rica and migrate north through eastern Texas and New York
318 (Fig. 1b; SI Table 4). Unlike the presence of strong connectivity across much of the
319 range, Wilson's warblers breeding from Alaska to Alberta were identified in all but one
320 of our migratory stopover sites and across all wintering areas, apart from western Mexico
321 and southern Baja (Fig. 1b, all but location g; SI Table 4).
322
323 Assignment of migrants collected in a time series from Cibola, AZ revealed a strong
324 temporal pattern in stopover site use across the spring migratory period (Fig. 1c; Table 3).
325 Birds *en route* to coastal California arrived first (week of March 22), followed by birds *en*
326 *route* to the Pacific Northwest (week of March 29), the Sierra Nevada (week of Apri l 5),
327 and Alaska to Alberta (week of April 26). Only a few individuals migrating through the
328 stopover site were identified as Sierra Nevada breeders (3 per year), while no populations

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

329 breeding in the Southern Rocky and Colorado Plateau and Eastern U.S. were identified
330 migrating through the stopover site. Temporal patterns in the arrival of spring migrants
331 were replicated across both the years 2008 and 2009 and were consistent regardless of
332 known differences in migration patterns by age and sex (Yong *et al.* 1998).
333
334 **Discussion**
335 Full life cycle conservation of declining migrant songbirds has been hindered by lack of
336 an efficient tracking technology that is both broadly applicable and high resolution. Here
337 we demonstrate how high-resolution molecular markers can be applied towards full life
338 cycle conservation of a migrant songbird, the Wilson's warbler, with a degree of
339 reliability and efficiency that has not been demonstrated using previous tracking methods.
340 By harnessing recent advances in Next-Generation Sequencing we show that 96 highly
341 divergent SNPs selected from a large pool of candidates (~450,000 SNPs) can be used to
342 identify genetically distinct groups on spatial scales that are informative for regional
343 conservation planning. Our analysis indicates that the power to identify individuals to
344 breeding populations is high (80 - 100%) and that reliable genotypes can be attained from
345 96% of feathers collected non-invasively from established bird monitoring stations across
346 North and Central America. Because of the biallelic nature of the SNPs in our panel, our
347 genetic data are also easier to validate and standardize across labs than isotope and other
348 genetic methods and, once the assays have been developed, it is possible to genotype
349 ~300 birds per day for < $10.00/ individual in almost any well-equipped molecular
350 laboratory. Overall, the resolution, efficiency, and cost, combined with the ease of
351 feather collection in collaboration with existing bird monitoring/banding infrastructure,

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

352 makes high-resolution genetic markers a broadly applicable method for widespread
353 monitoring of declining songbird species.
354
355 One of the central challenges in migratory bird conservation is that population declines
356 and conservation planning often occur at regional spatial scales, but our knowledge of
357 migratory connections is usually limited to species-wide range maps. For example, in the
358 Wilson's warbler, an analysis of Breeding Bird Survey (BBS) data for the years 1966 –
359 2012 suggests that the species is only slightly declining across it's range (BBS Trend = -
360 1.88, 95% CI = 2.97, -1.11), but an analysis of regional trends suggest that populations in
361 the Sierra Nevada and the Southern Rockies/Colorado Plateau are declining more
362 strongly (BBS $\text{Trend}_{\text{sierra}}$ = -4.71, 95% CI = -6.41, -2.85; BBS $\text{Trend}_{\text{rockies}}$ = -2.95, 95%
363 CI = -4.32, -1.42) (Sauer *et al.* 2012). Here we illustrate that by targeting highly
364 divergent SNP loci we can confidently identify a minimum of six genetically distinct
365 groups across the breeding range with a resolution in the western US equivalent to the
366 spatial scale of regional population declines. Furthermore, the spatial scale of our genetic
367 groups is commensurate with many *a priori* defined Bird Conservation Regions,
368 ecologically distinct areas in North America with similar habitats and resource
369 management issues (Millard *et al.* 2012). The ability to align the spatial scale of
370 population genetic structure with the spatial scale of population declines and conservation
371 planning provides a powerful framework from which to base full life cycle conservation
372 (Fig. 1a & b).
373

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

374 The Wilson's warbler has been the focus of numerous population genetic/connectivity
375 studies in the past decade (Clegg *et al.* 2003; Irwin *et al.* 2011; Kimura *et al.* 2002;
376 Paxton *et al.* 2007; Paxton *et al.* 2013; Rundel *et al.* 2013; Yong *et al.* 1998), but none
377 have yielded the depth and clarity of information on migratory connections documented
378 herein. Our results confirm the presence of previously identified connections between
379 birds breeding in Coastal California and wintering in Southern Baja, MX and between
380 birds breeding in eastern North America and wintering in the Yucatan, Belize and Costa
381 Rica (Kimura *et al.* 2002; Rundel *et al.* 2013), but also reveal new patterns across time
382 and space that are much richer and stronger then previously recognized. For example,
383 here we show that Wilson's warblers breeding in Coastal California (Fig. 1b, yellow)
384 share their wintering area in southern Baja with Pacific Northwest breeders (Fig. 1b,
385 green) and that both of these groups also winter to the east of Baja in Sinaloa, MX, with
386 Sierra Nevada breeders (Fig. 1b, pink) (Sauer *et al.* 2012). Samples collected from across
387 the spring migratory period indicate that western breeders from all three groups (Coastal
388 California, Pacific Northwest, and Sierra Nevada) migrate north along the Pacific
389 Flyway, with Coastal California and Sierra Nevada breeders found west of the Lower
390 Colorado River. In addition, we show for the first time that breeders from the Southern
391 Rocky Mountains and Colorado Plateau (Fig. 1b, orange) occupy a restricted El
392 Salvador-to-Costa Rica wintering distribution and migrate North along the Central
393 Flyway, while eastern breeders (Fig. 1b, red) migrate North through eastern Texas and
394 New York. Overall our results indicate that screening high volumes of individuals using
395 high resolution molecular markers can yield a level of clarity in migratory connections

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

396 across time and space that has not been previously demonstrated using other tracking
397 techniques.
398
399 The resulting map for the Wilson's warbler provides an example of how information on
400 region-specific migration patterns can be combined with information on region-specific
401 population declines in order to strengthen predictions about where migrants are most
402 limited. In the case of the Wilson's warbler, BBS data suggests that Sierra Nevada
403 breeders are experiencing strong population declines (BBS $Trend_{sierra}$ = 4.71, 95% CI = -
404 6.41, -2.85), while Pacific Northwest and Coastal California breeders are declining less
405 severely or remaining stable (BBS $Trend_{Pacific\_Northwest}$ = -1.96, 95% CI = -2.54, -1.31;
406 BBS $Trend_{Coastal\_California}$ = -0.49, CI = -1.62, 0.84). The fact that all three groups occupy
407 distinct breeding ranges, but mix on their wintering grounds and at migratory stopover
408 sites suggests that declines in Sierra Nevada breeders are likely driven by factors on the
409 breeding grounds. Alternatively, the migration map as a whole suggests that bottlenecks
410 for Wilson's warblers likely occur in areas where multiple genetically distinct breeding
411 groups funnel through the same stopover site or wintering area such as in Coastal
412 California, Western Mexico, and Costa Rica. These results are supported by work in
413 other taxa and further emphasize the importance of stopover habitat for migrant
414 conservation (Sheehy *et al.* 2011).
415
416 Migratory passerines spend roughly a quarter of their year *en route* between breeding and
417 wintering areas, but relatively little is known about the biology and behavior of migrants
418 during the migratory phase of their annual cycle (Faaborg *et al.* 2010b). The availability

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

419 and quality of habitat at stopover sites could have significant effects on populations, but
420 determining the extent to which physiological and ecological demands experienced
421 during migration may limit populations is often contingent upon knowledge of an
422 individuals ultimate destination (Faaborg *et al.* 2010a; Faaborg *et al.* 2010b). Here we
423 successfully genotype 609 samples collected in a time series from a stopover site near
424 Cibola, AZ and demonstrate how high-resolution genetic markers can be used to identify
425 the ultimate destination of birds captured *en route* to their breeding grounds (Fig. 1b &c;
426 location b). Breaking down the results by week revealed distinct waves of migrants, with
427 Coastal California breeders arriving first (March 22 – 29), followed by Pacific Northwest
428 and Sierra Nevada breeders (March 29-April 5), and Alaska-to-Alberta breeders arriving
429 significantly later (April 19-26). These patterns were replicated across two years and are
430 consistent regardless of known differences in migration patterns by age and sex (Yong *et*
431 *al.* 1998). While differences in the timing of migration in Wilson's warblers have been
432 suggested in the past based upon changes in the frequency of haplotypes or isotopic
433 signatures (Paxton *et al.* 2007; Paxton *et al.* 2013), this is the first time that anyone has
434 attained individual-level assignments of large numbers of migrants collected in a time
435 series, bringing a new level of clarity to our understanding of stopover site use through
436 time. It is important to note, that the depth of sampling across time that we are able to
437 achieve using high-resolution genetic markers would not have been possible using
438 extrinsic tracking devices, such as geolocators, due of cost and weight restrictions and the
439 need to recapture individuals to collect the information (Arlt *et al.* 2013; Bridge *et al.*
440 2013). The resulting information on migratory connections across time can be used to
441 help build timetables of migration along the Pacific Flyway and help to inform when

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

442 particularly vulnerable populations may be migrating through an area. Furthermore,
443 because DNA can be collected from all birds, dead or alive, high resolution genetic
444 markers could be used to identify migrants subject to collisions with wind turbines, cell
445 phone towers and other manmade structures.
446
447 While our results suggest that high-resolution molecular markers surpass previous genetic
448 markers in terms efficiency and resolution, our conclusions could be further strengthened
449 by the inclusion of additional data and analyses. For example, the robustness of the
450 patterns described here varies depending upon the sample size at each location and in
451 some locations, such as in Belize and many of the migratory stopover sites (Fig. 1b,
452 locations l, d, e, f, g), additional sampling across time and space is needed. In addition,
453 while our assignment probabilities are very high for an intrinsic marker (80 - 100%) there
454 is a potential for incorrect assignments, particularly between the three western groups
455 (Coastal California, Pacific Northwest, and the Sierras) were admixture is likely (Table
456 2). Similarly, there are large regions on the breeding grounds that could not be
457 distinguished using our markers, such as birds breeding from Alberta to Alaska (purple,
458 Fig. 1b). In the future, the addition of more genetic loci as well as the addition of
459 isotopic markers and statistical methods for combining both sources of data into a single
460 statistical framework will help further resolve populations across the range (Rundel *et al.*
461 2013). Lastly, it is important to note that the spatially explicit depiction of the genetic
462 results generated in GENELAND may not accurately identify the location of boundary
463 between genetic groups. Additional sampling across the projected boundaries will help
464 clarify the location of the genetic breaks as well as the factors driving differences

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

465 between Wilson's warblers in each region. Such genetic differences are particularly
466 interesting in light of the documented differences in migratory timing for Wilson's
467 warblers described herein and the potential for migration timing to contribute to
468 divergence in migratory birds more generally (Bearhop *et al.* 2005; Ruegg *et al.* 2014;
469 Ruegg *et al.* 2012).
470
471 A review article by Faaborg *et al* (Faaborg *et al.* 2010b) recently identified continuing
472 research needs for Neotropical migrant birds, including identifying migratory pathways
473 and wintering locations, bottlenecks for conservation, and timetables for migration. Here
474 we demonstrate how high-resolution genetic markers designed for Wilson's warblers, can
475 be applied to help address many of these continuing research needs with a level of
476 efficiency and reliability that has not previously been demonstrated. In the last several
477 years there has been a revolution in sequencing technology that has increased by orders
478 of magnitude the amount of sequence data that can be generated, while at the same time
479 reducing the cost of individual-level analysis (Metzker 2010). Our results show that by
480 harnessing recent advances in sequencing technology it is now possible to develop high-
481 resolution genetic markers for tracking populations of migrants on a broad scale. The
482 resulting information on fine-scale population genetic structure, region-specific migratory
483 connections, and timetables of migration provides a powerful framework from which to
484 base full life cycle conservation of declining songbird species.
485

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Literature Cited

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

Ammon E, M., Gilbert W, M. (1999) Wilson's Warbler (*Cardellina pusilla*). In: *The Birds of North America Online* (ed. Poole A), Ithaca: Cornell Lab of Ornithology.

Anderson EC (2010) Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. *Mol Ecol Resour* **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. . Universite de Montpellier II, Montepellier (France), Laboratoire Genome, Populations, Interactions.

Bivand R, Keitt T, Rowlingson B (2014) rgdal: Bindings for the Geospatial Data Abstraction Library. *R package version 0.8 - 16*.

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.

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.

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.

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

Gustafson ME, Hildenbrand J (1999) *Bird Banding Laboratory Homepage, ver 07-10-2005.*

Hijmans RJ (2014) Raster: Geographic data analysis and modeling. *R package version 2.2-31*.

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat 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. *Nat Commun* **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.

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

Millard MJ, Czarnecki CA, Morton JM*, et al.* (2012) A National Geographic Framework for Guiding Conservation on a Landscape Scale. *Journal of Fish and Wildlife Management* **3**, 175-183.

NatureServe BIa (2012) *Bird species distribution maps of the world* NatureServe, Arlington, USA.

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

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.

Paxton KL, Van Ripper III C, 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 mitochondiral DNA and stable isotopes. *Auk* **130, 689**.

Pebesma EJ, Bivand RS (2005) Classes and methods for spatial data in R. *R News* **5**.

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

577 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using
578 multilocus genotype data. *Genetics* **155**, 945-959.
579 Rappole JH (1995) *The ecology of migrant birds - a Neotropical perspective* Smithsonian
580 Institution Press, Washington.
581 Robbins CS, J. R. Sauer, R. Greenberg, and S. Droege (1989) Population declines in
582 North American birds that migrate to the Neotropics. *Proc. Nat. Acad. Sci.* **86**,
583 7658-7662.
584 Rosenberg NA (2004) Distruct: a program for the graphical display of population
585 structure. *Molecular Ecology Notes* **4**.
586 Rousset F (2008) genepop'007: a complete re-implementation of the genepop software
587 for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.
588 Rubenstein DR, Chamberlain CP, Holmes RT*, et al.* (2002) Linking breeding and
589 wintering ranges of a migratory songbird using stable isotopes. *Science* **295**,
590 1062-1065.
591 Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB (2014) A role for migration-linked
592 genes and genomic islands in divergence of a songbird. *Molecular Ecology*.
593 Ruegg K, Anderson EC, Slabbekoorn H (2012) Differences in timing of migration and
594 response to sexual signalling drive asymmetric hybridization across a migratory
595 divide. *Journal of Evolutionary Biology* **25**, 1741-1750.
596 Rundel CW, Wunder MB, Alvarado AH*, et al.* (2013) Novel statistical methods for
597 integrating genetic and stable isotope data to infer individual-level migratory
598 connectivity. *Molecular Ecology* **22**, 4163-4176.
599 Sauer JR, Hines JE, Fallon JE*, et al.* (2012) The North American Breeding Bird Survey,
600 Results and Analysis 1966-2011. USGS Patuxent Wildlife Research Center,
601 Laurel, MD.
602 Segelbacher G (2002) Noninvasive genetic analysis in birds: testing reliability of feather
603 samples. *Molecular Ecology Notes* **2**, 367-369.
604 Seutin G, B. N. White and P. T. Boag (1991) Preservation of avian blood and tissue
605 samples for DNA analyses. *Canadian Journal of Zoology* **69**, 82-90.
606 Sheehy J, Taylor CM, Norris DR (2011) The importance of stopover habitat for
607 developing effective conservation strategies for migratory animals. *Journal of*
608 *Ornithology* **152**, 161-168.
609 Smith TB, Wayne RK, Marra PP*, et al.* (2003) A call for feather sampling. *The Auk* **120**,
610 218-221.
611 Stutchbury BJ, Tarof SA, Done T*, et al.* (2009) Tracking long-distance songbird
612 migration by using geolocators. *Science* **323**, 896.
613 Team RC (2014) *R: A language and environment for statistical computing.*, Vienna,
614 Austria. http://www.R-project.org/
615 Warren WC, Clayton DF, Ellegren H*, et al.* (2010) The genome of a songbird. *Nature*
616 **464**, 757-762.
617 Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population
618 structure. *Evolution* **38**, 1358-1370.
619 Wunder MB, Kester CL, Knopf FL, Rye RO (2005) A test of geographic assignment
620 using isotope tracers in feathers of known origin. *Oecologia* **144**, 607-617.
621 Yong W, Finch DM, Moore FR, Kelly JF (1998) Stopover ecology and habitat use of
622 migratory Wilson's Warblers. *Auk* **115**, 829-842.

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

623 Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using
624 de Bruijn graphs. *Genome Research* **18**, 821-829.
625

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

640

641 **Author Contributions.** K. Ruegg and T.B. Smith conceived of the study and K. Ruegg
642 wrote the majority of the manuscript and conducted and/or oversaw the analyses. E.C.
643 Anderson wrote the scripts for the population genomic analyses and figure creation. K.
644 Paxton and F. Moore contributed ideas and genetic material for the analysis of migrants
645 from Cibola, AZ. V. Apkenas conducted and helped analyze data for the SNP screening.
646 S. Lao assisted with feather sample organization, extraction, and the analysis of
647 genotyping reliability scores. R.B. Siegel and D.F. DeSante facilitated the collection of
648 feather samples in collaboration with bird banding stations within and outside of the

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

649 Monitoring Avian Productivity and Survivorship (MAPS) and the Monitoreo de
650 Sobrevivencia Invernal (MoSI) networks.
651
652 **Figure Ledgend**
653 **Figure 1**. Migratory connections in the Wilson's warbler identified using SNP-based
654 genetic markers. A) Results from STRUCTURE showing 6 genetically distinct
655 populations across the breeding grounds. Capital letters (A-W) refer to the location of
656 breeding populations depicted on the map in B as well as listed in Table 1. B) Spatially
657 explicit population structure across the annual cycle. The colors across the breeding
658 range represent the results from GENELAND which were post-processed using R so that
659 the density of each color relects the relative posterior probability of membership for each
660 pixel to the most probable of the 6 different genetic clusters (see text). The results were
661 clipped to the species distribution map (NatureServe 2012). Lower case letters (a-g)
662 represent the location of wintering and spring migratory samples (Table 1). Pie charts
663 indicate the proportion of wintering indidiviuals assigned to each breeding group with the
664 number of individuals listed at the center of each pie. Arrows represent the proportion of
665 migrants assigned to each breeding group with the numbers of indviduals listed at the top
666 of the arrows. C) The proportion of indivdiuals assigned to each breeding population
667 across spring migration of 2008 and 2009. Numbers in the center of the pies refer to
668 sample sizes and the data are grouped by week with the date representing the mid-week
669 date in a non-leap year.

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 (Jun 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

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Wintering (Dec - Feb)

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
Las Lajas, Santa Ana, SV	13.943	-89.617	7	q
Metapan, Santa Ana, SV	14.403	-89.360	9	q

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 colors indicate the genetic group (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

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 3. Genetic identification of Wilson's Warblers migrating through Cibola, CA by week across the years 2008 and 2009. Results represent the individuals assigned to one of the six genetically distinct groups using the program GSI Sim and the data corresponds to the information presented in Figure 1c.

Mid-week Date*	Week	Alaska to Alberta	Pacific NW	Coastal CA	Sierra Nevada	Rocky Mt.	Eastern
Year 2008							
21-Mar	11	0	0	0	0	0	0
28-Mar	12	0	3	1	0	0	0
4-Apr	13	0	11	16	1	0	0
11-Apr	14	0	9	4	1	0	0
18-Apr	15	0	5	0	0	0	0
25-Apr	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-Mar	11	0	0	2	0	0	0
29-Mar	12	0	3	7	0	0	0
5-Apr	13	0	5	10	0	0	0

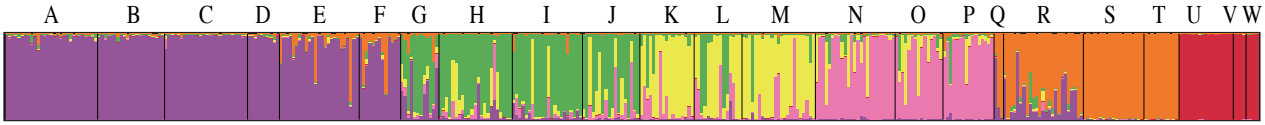
bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

12-Apr	14	0	6	10	0	0	0
19-Apr	15	0	10	6	1	0	0
26-Apr	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

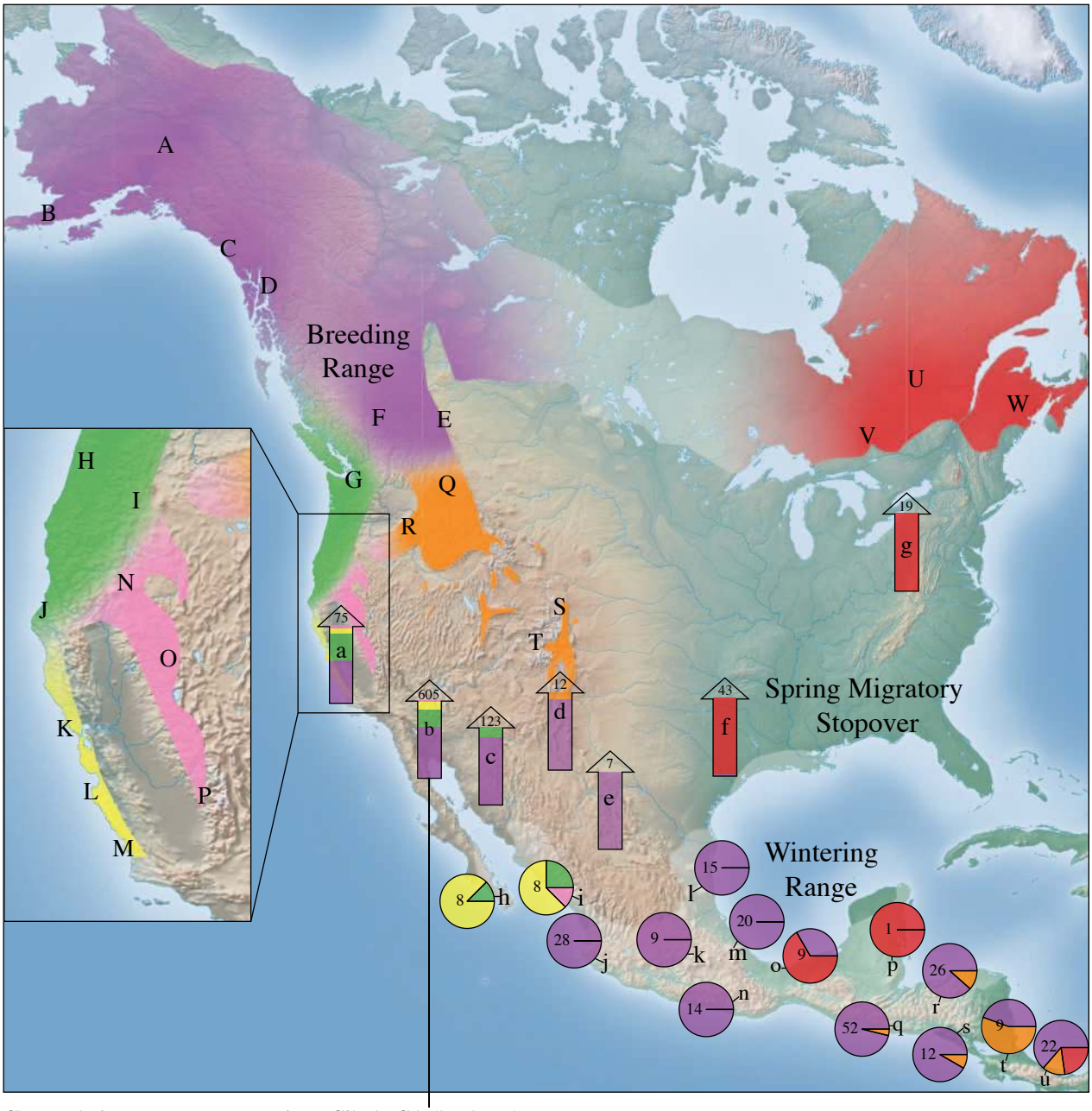
* Dates represent the midweek date in a non-leap year.

bioRxiv preprint doi: https://doi.org/10.1101/007757; this version posted August 8, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A) Breeding population structure



B) Spatially explicit population structure



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