# Identifying Migrant Origins Using Genetics, Isotopes, and Habitat Suitability

Kristen C. Ruegg[1,2,†,*], Eric C. Anderson[3,†], Ryan J. Harrigan[1], Kristina L. Paxton[4], Jeff Kelly[5,6], Frank Moore[7], Thomas B. Smith[1, 8]

[1] Center for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, Los Angeles, CA, USA

[2] Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, CA, USA

[3] Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Santa Cruz, CA, USA

[4] Department of Biology, University of Hawaii, Hilo, HI, USA

[5] Oklahoma Biological Survey, University of Oklahoma, Norman, OK, USA

[6] Department of Biology, University of Oklahoma, Norman, OK, USA

[7] Department of Biological Sciences, University of Southern Mississippi, Hattiesburg, MS, USA

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

[†] **Co-first authors**: Anderson and Ruegg contributed equally.

[*] **Corresponding author**: kruegg@ucsc.edu

**Abstract**

1.  Identifying migratory connections across the annual cycle is important for studies of migrant ecology, evolution, and conservation. While recent studies have demonstrated the utility of high-resolution SNP-based genetic markers for identifying population-specific migratory patterns, the accuracy of this approach relative to other intrinsic tagging techniques has not yet been assessed.

2. Here, using a straightforward application of Bayes' Rule, we develop a method for combining inferences from high-resolution genetic markers, stable isotopes, and habitat suitability models, to spatially infer the breeding origin of migrants captured anywhere along their migratory pathway. Using leave-one-out cross validation, we compare the accuracy of this combined approach with the accuracy attained using each source of data independently.

28  3. Our results indicate that when each method is considered in isolation, the accuracy of genetic assignments

29  far exceeded that of assignments based on stable isotopes or habitat suitability models. However, our joint

30  assignment method consistently resulted in small, but informative increases in accuracy and did help to

31  correct misassignments based on genetic data alone. We demonstrate the utility of the combined method by

32  identifying previously undetectable patterns in the timing of migration in a North American migratory songbird,

33  the Wilson's warbler.

34  4. Overall, our results support the idea that while genetic data provides the most accurate method for tracking

35  animals using intrinsic markers when each method is considered independently, there is value in combining

36  all three methods. The resulting methods are provided as part of a new computationally-efficient R-package,

37  GAIAH, allowing broad application of our statistical framework to other migratory animal systems.

38  **Keywords**

39  Migratory Connectivity — Movement Ecology — Population Assignment — Avian Ecology — Wilson's warbler

40  ## 1. Introduction

41  The ecology and evolution of animals that undergo annual seasonal migration is shaped by events encountered across

42  the entire annual cycle (Sillett *et al.*, 2000; Webster *et al.*, 2002, 2005). It is now well established that habitat conditions

43  during migratory or wintering phases can have significant carry-over effects on breeding ground productivity (Marra

44  *et al.*, 1998; Sillett *et al.*, 2000; Norris & Taylor, 2006). As a result, understanding patterns of migratory connectivity,

45  or the geographic links between breeding, wintering, and stopover sites for a population over the course of an annual

46  cycle, is a critical first step towards studies of migrant ecology, evolution, and conservation.

47  Efforts to identify the strength of migratory connections have relied on a variety of methods for tracking animal

48  movements (Marra *et al.*, 1998; Bonfil *et al.*, 2005; Smith *et al.*, 2005; Stutchbury *et al.*, 2009). Extrinsic devices

49  such as satellite transmitters and geo-locators have increased our knowledge of the movement patterns of individuals

50  of a particular species (Stutchbury *et al.*, 2009), but remain impractical for many large-scale (1000's of individuals)

51  applications due to cost and weight restrictions, and the need to re-capture many individuals to collect the data (Arlt

52  *et al.*, 2013; Bridge *et al.*, 2013). An attractive alternative is the use of genetic and isotopic markers, that capture

53  information contained within the tissue of an animal, to pinpoint an individual's population of origin. These methods

54  have broad appeal because they are minimally invasive, cost-effective when applied at scale, and do not require recapture

55  of individuals (Rubenstein *et al.*, 2002; Kelly *et al.*, 2005; Rundel *et al.*, 2013). Furthermore, intrinsic methods make

56  it possible to trace the origins of animals that have died from both natural and anthropogenic causes (i.e., poaching,

57   collisions, and disease), because genetic and isotopic samples can be collected from carcasses.

58   While some genetic approaches have been limited by a lack of resolution, advances in genome-wide sequencing have
59   resulted in new technologies that can be applied to genetic tagging of wild populations (Allendorf *et al.*, 2010; Metzker,
60   2010; Davey *et al.*, 2011). Even in species with high rates of dispersal, such as birds, fish, and mammals it has been
61   found that a small number ($n < 100$) of single nucleotide polymorphisms (SNPs) found within, or linked to, genes under
62   selection can be targeted to reveal population structure at spatial scales that are critical to regional conservation planning
63   (Nielsen *et al.*, 2009; Hess *et al.*, 2011; Nielsen *et al.*, 2012; Ruegg *et al.*, 2014). For example, Ruegg et al (2014) found
64   that 96 high-resolution SNPs could be used to identify six genetically distinct populations of a migratory songbird, the
65   Wilson's warbler (*Cardellina pusilla*), whereas previous single-marker techniques found support for only two groups
66   (Kimura *et al.*, 2002). Furthermore, SNP assays that isolate short fragments of DNA specific to the taxa of interest
67   make it possible to rapidly screen DNA from a variety of tissue types (i.e., bird feathers, fin clips, animal hair) that can
68   be collected using non-invasive sampling techniques, making this method an attractive choice for conservation genomic
69   studies (Ruegg *et al.*, 2014; Kraus *et al.*, 2015).

70   Despite their appeal, several questions remain as to how genetic tools compare in accuracy with other intrinsic marking
71   methods. In addition, there are clear examples where genetic markers alone fall short of resolving populations across all
72   or large parts of a species geographic range (Gagnaire *et al.*, 2015; Toews *et al.*, 2015). In such situations, the inclusion
73   of additional sources of non-genetic information, such as isotopic data and/or habitat suitability models, can increase
74   the resolution of genetic markers (Kelly *et al.*, 2005; Rundel *et al.*, 2013; Pekarsky *et al.*, 2015). For example, Rundel
75   *et al.* (2013) showed that genetic and isotopic information can be combined to increase the assignment accuracy of
76   individuals (birds) to their population of origin; however this method was not designed to deal with disjunct patterns
77   of genetic variation, like that observed in the Wilson's warbler (Ruegg *et al.*, 2014). More recently, Pekarsky *et al.*
78   (2015) used habitat suitability models to refine their isotopic-based estimates of population assignment, but did not
79   establish whether the use of habitat suitability as a prior actually led to improved estimates of the geographic origin of
80   individuals.

81   Here we combine genetic, stable isotope, and habitat suitability data into a joint assignment procedure that infers the
82   breeding origins of individuals collected anywhere along their migratory trajectory with greater resolution than can be
83   attained using each method individually. We refine previously developed R-code (R Core Team, 2016) for performing
84   each type of assignment alone  (Anderson *et al.*, 2008; Bridge *et al.*, 2013; Vander Zanden *et al.*, 2014), making it
85   computationally feasible to combine assignments and perform statistically rigorous leave-one-out cross validation. To
86   assess the overall accuracy of our method in comparison to existing approaches we examine the results from each
87   type of data—genetics, stable isotopes, and habitat suitability—individually as well as jointly, in order to determine:
88   (1) contributions of each data type to the accuracy of joint assignments, (2) insights that might be gathered from the

89 combined approach, and (3) recommendations for future studies of migratory connectivity based on our results.

90 We assess the accuracy of each method using data from a long-distance migratory bird, the Wilson's warbler (*Cardel-*
91 *lina pusilla*), a particularly appropriate model for testing the efficacy of our approach because previous population-
92 genetic/connectivity studies on this species provide a solid basis for comparison between methods (Kimura *et al.*,
93 2002; Clegg *et al.*, 2003; Paxton *et al.*, 2007; Irwin *et al.*, 2011; Paxton *et al.*, 2013; Rundel *et al.*, 2013). We start by
94 deriving posterior probability rasters (or other scaled "scores") for each method individually (genetics, stable isotopes,
95 and habitat suitability) and then combine those into a joint assignment probability. We then evaluate the gains in
96 accuracy achieved by each method individually as well as by combining multiple data sources using leave-one-out
97 cross-validation with a reference set of Wilson's warblers sampled from known locations during the breeding season.
98 The combined approach is then applied to the assignment of migratory birds of unknown origin at a stopover site in
99 Cibola, AZ during spring migration. Our methods have been implemented in the new R-package, GAIAH (**G**enetic
100 **A**ssignment using **I**sotopes **A**nd **H**abitat suitability), which also includes all the data and scripts required to replicate
101 our results in this paper. It is available on GitHub (`https://github.com/eriqande/gaiah`).

# 2. Methods

## 2.1 Sampling

104 Sampling of Wilson's warblers is detailed in Ruegg *et al.* (2014). Briefly, collection of 357 feathers from 30 locations
105 (average of 12 individuals/site; range: 2 - 25) across the breeding range was made possible through a large collaborative
106 effort with bird banding stations both within and outside of the Monitoring Avian Productivity and Survivorship (MAPS)
107 and the Landbird Monitoring of North America (LaMNA) networks. These samples became the foundation of the
108 subsequent genetic and isotopic analysis of known-origin samples. Breeding samples were collected and categorized
109 into groups based on collection date (June 1 to July 31), signs of breeding (presence/size of a cloacal protuberance), and
110 life history timetables for the Wilson's warbler. To illustrate the efficacy of the combined approach to the assessment
111 of migratory stopover site use-through-time, 686 migrant samples were also collected from Cibola, AZ ($31°18'$N,
112 $114°41'$W), using consistent-effort, daily, passive mist-netting from March 22 to May 24, in both 2008 and 2009.

## 2.2 Genetics

114 The application of genetic markers to the assignment of Wilson's warblers is detailed in Ruegg *et al.* (2014). Briefly,
115 genetic samples, consisting of the proximal end of one rectrix (breeding samples) and whole blood samples (migratory
116 samples) were purified using a Qiagen DNeasy Blood and Tissue Kit and quantified using a NanoDrop$^{TM}$ Spectropho-

117  tometer (Thermo Scientific, Inc) (Smith *et al.* 2003). Genotyping was done using a panel of 96 high-resolution

118  SNP markers ascertained from a genome-wide survey of genetic variation in the species using restriction-associated-

119  digest, paired-end (RAD-PE) sequencing. Analyses using STRUCTURE (Pritchard *et al.*, 2000; Falush *et al.*, 2003)

120  and GENELAND (Guillot *et al.*, 2005) indicated $G = 6$ spatially-distinct population groups that could be reliably

121  distinguished using the 96 SNPs. The spatial extent of these six groups identified by GENELAND was overlaid, and

122  then clipped, by the known range of Wilson's warbler [downloaded from NatureServe, `www.natureserve.org`,

123  (Ridgely *et al.*, 2005)], to create a map of the occurrence of different population groups across the breeding range

124  (See Figure 1, Ruegg *et al.* 2014). Breeding birds collected from locales within each population group were included

125  as reference samples for genetic population assignment to assign posterior probabilities across the 6 genetic groups

126  using the program GSI_SIM (Anderson *et al.*, 2008). For each bird, *i*, this program returns an estimate of the posterior

127  probability,

128  $$P(\mathcal{G}_i = g | y_i^{\text{gen}}, B) \;\;,\;\; g = 1, \ldots, G,$$

129  that *i*'s genetic group ($\mathcal{G}_i$) is equal to $g$ ($g$ denotes one of the six genetic groups), given *i*'s genotype $y_i^{\text{gen}}$, and the

130  reference or baseline samples, $B$.

131  These posterior probabilities refer to group membership of each bird; however, to combine these inferences with stable

132  isotope data requires first converting group membership posteriors into posterior probabilities of spatial location. As the

133  96 SNPs provide only limited ability to resolve local origin of birds *within* each genetic group's geographic area, we

134  converted the genetic posteriors for the $i^{\text{th}}$ bird to a grid, $\mathbf{M}_i^{\text{gen}}$, of spatially-explicit posterior probabilities by distributing

135  the posterior probability $P(\mathcal{G}_i = g | y_i^{\text{gen}}, B)$ uniformly across the raster grid representing the spatial range of group $g$, and

136  ensuring that these are appropriately normalized to sum to one across the whole breeding range. Namely, within group

137  $g$'s range, posterior probability is distributed according to

138  $$\mathbf{M}_i^{\text{gen}}(g) = \frac{P(\mathcal{G}_i = g | y_i^{\text{gen}}, B) \mathbf{U}_g}{C_g}$$

139  where $\mathbf{U}_g$ is a matrix (raster) containing 1's in cells within genetic group $g$'s region and 0's elsewhere, and $C_g$ is the

140  number of raster cells within group $g$'s region (*i.e.*, the sum of the elements of $\mathbf{U}_g$). In most cases within the reference

141  dataset indivdiuals were assigned with high probablity to one genetic group and lower probabity to other groups (see

142  Figure 1, Ruegg et al 2014). To represent the uncertainty in the genetic assignments to different groups (which in

143  most cases was very small, see Supporting Information Document 1, Figure 1) we spread the posterior probability

144  of assignment to each group across the geographic ranges of all the groups to which the bird was assigned and then

145  weighted the values according to their likelihood so that the combined areas summed to 1.

146 Subsequently, the posterior probability of spatial assignment of birds based on genetics and assuming a uniform prior

147 across space is

$$\mathbf{M}_i^{\text{gen}} = \sum_{g=1}^{G} \mathbf{M}_i^{\text{gen}}(g),$$

149 whose elements clearly sum to unity. $\mathbf{M}_i^{\text{gen}}$ is a raster of the same extent and resolution as $\mathbf{M}_i^{\text{iso}}$ (see next section).

## 2.3 Stable Isotopes

151 Predictable continental patterns of stable hydrogen isotopes in rainfall ($\delta^2 H_p$) are highly correlated with stable hydrogen

152 isotopes of animal tissues ($\delta^2 H_f$), allowing for inferences about the origin of where tissues were grown (Hobson &

153 Wassenaar, 2008). When the breeding and molting locations are the same, as is the case for Wilson's warblers, then

154 stable isotope values provide an assessment of the breeding origin of a bird that is independent of a genetic assessment.

155 We expressed all isotope ratios in standard delta notation ($\delta^2 H$) where

$$\delta = \frac{\text{isotope ratio of sample}}{\text{isotope ratio of standard}} - 1,$$

157 with ratios shown as parts per thousand (‰) deviation from Vienna Standard Mean Ocean Water for hydrogen. Prior

158 to analysis, all feathers (breeding and migratory samples) were cleaned with dilute detergent followed by a 2:1

159 chloroform:methanol solution (Paritte & Kelly, 2009). For $\delta^2 H$ analyses, a 0.1–0.2 mg piece of feather was packed into

160 a silver capsule and loaded into an auto-sampling tray. Isotope ratio measurements were performed at the University

161 of Oklahoma with a ThermoFinnigan Delta V isotope ratio mass spectrometer connected to an elemental analyzer (H

162 analyses: TC/EA, Thermo-Finnigan). To control for exchangeable hydrogen, hydrogen isotope ratios were normalized

163 according to Wassenaar & Hobson (2003), using established keratin standards: chicken feathers (-147.4‰), cow (*Bos*

164 *taurus*) hooves (-187‰), and bowhead whale (*Balaena mysticetus*) baleen (-108‰). For additional details on our

165 analysis methods, see Kelly *et al.* (2009) and Paritte & Kelly (2009).

166 We created an isoscape of $\delta^2 H_p$ ratios and its associated variance using IsoMAP (Job Key 54152), an online resource

167 to generate region- and time-specific isoscapes for geographic assignments (Bowen *et al.* 2014, www.isomap.org).

168 In IsoMAP, a geospatial isoscape was generated using precipitation data from 120 stations collected during the time

169 period of 1960–2009, and included CRU-derived climatic variables such as elevation, precipitation, and minimum

170 precipitation in the model (Mitchell & Jones 2005; similar to Hobson *et al.* 2012c). Using a parametric bootstrapping

171 approach we converted the isoscape of $\delta^2 H_p$ values to an isoscape of $\delta^2 H_f$ values based on the relationship between

172 $\delta^2 H_p$ and $\delta^2 H_f$ collected from the 357 known-origin, breeding Wilson's warblers sampled across the breeding range

173 (Appendix 1). We then computed the posterior probability of breeding origin given a feather $\delta^2 H_f$ ratio and three

174 sources of variance following the methods used in Vander Zanden *et al.* (2014). The three sources of variance included

175 variance associated with; 1) the original $\delta^2 H_p$ isoscape generated in IsoMAP, 2) the rescaled precipitation to feather

176 $\delta^2 H$ isoscape, and 3) individual variation in $\delta^2 H_f$ among birds sampled at the same breeding location.

177 Specifically, this approach assumes that $y_i^{\text{iso}}$, the isotope ratio measured in the feather of the $i^{\text{th}}$ bird, is a normal random

178 variable with mean and variance determined by its location in space. If these means and variances, denoted by the

179 matrices $\mathbf{T}^{(\mu)}$ and $\mathbf{T}^{(\sigma^2)}$, respectively, were known across a regular grid of possible origin locations, then it would be

180 straightforward to compute the posterior probability of breeding location. Namely, assuming a uniform prior on spatial

181 origin, the posterior that bird $i$ originated from the $(r,c)^{\text{th}}$ cell in the grid is proportional to the density of observing $y_i^{\text{iso}}$

182 given it was drawn from a normal density with mean $T_{r,c}^{(\mu)}$ and variance $T_{r,c}^{(\sigma^2)}$ (the $(r,c)^{\text{th}}$ elements of $\mathbf{T}^{(\mu)}$ and $\mathbf{T}^{(\sigma^2)}$,

183 respectively).

184 It should be noted that $\mathbf{T}^{(\mu)}$ and $\mathbf{T}^{(\sigma^2)}$ are not known, so we follow Vander Zanden *et al.* (2014) by using $\widetilde{\mathbf{T}}^{(\mu)}$ and

185 $\widetilde{\mathbf{T}}^{(\sigma^2)}$ in their place, computed as

$$
\begin{aligned}
\widetilde{\mathbf{T}}^{(\mu)} &= \bar{a}\mathbf{P}^{(\mu)} + \bar{b} & (1)\\
\widetilde{\mathbf{T}}^{(\sigma^2)} &= \mathbf{R}^{(\sigma^2)} + \mathbf{P}^{(\sigma^2)} + \sigma_{\text{indiv}}^2\mathbf{J}, & (2)
\end{aligned}
$$

188 where $\mathbf{P}^{(\mu)}$ and $\mathbf{P}^{(\sigma^2)}$ are the predictions and associated variances, respectively, for the precipitation isotope ratios from

189 IsoMAP made on the same grid as $\widetilde{\mathbf{T}}^{(\mu)}$ and $\widetilde{\mathbf{T}}^{(\sigma^2)}$; $\mathbf{J}$ is a matrix of 1's of the same dimension as $\widetilde{\mathbf{T}}^{(\sigma^2)}$; $\mathbf{R}^{(\sigma^2)}$, $\bar{a}$, and $\bar{b}$,

190 are determined by a parametric bootstrapping approach described in Appendix 1; and $\sigma_{\text{indiv}}^2$ is a term accounting for

191 individual variation in isotope ratios which was set to the square of the mean of the standard deviation of $y_i^{\text{iso}}$ amongst

192 birds sampled at common locations across the breeding range during the breeding season.

193 Posterior probabilities derived in this manner for bird $i$ are a set of values over a spatial grid denoted by the matrix,

194 $\mathbf{M}_i^{\text{iso}}$.

## 2.4 Habitat Suitability

196 In order for subspecies ranges to be further classified in terms of their utilization by breeding individuals, habitat

197 suitability models were constructed for Wilson's warblers across their breeding range. This was performed by identifying

198 the geographical locations of Wilson's warblers during the breeding season, and then modeling the relationship of

199 these occurrences to the environmental conditions at those sites. This model was then used to predict the probability of

200 occurrence of individuals across the species range.

201  Locations of individuals were extracted from the eBird website (`www.ebird.org`). To ensure that we used only

202  those birds that were breeding (or at least were located on breeding grounds), we filtered the nearly 450,000 records of

203  Wilson's warblers available to include only those sighted on or after June 10, but before August 1, of any given calender

204  year. To avoid redundant observations at the same site artificially influencing models, we used each unique geographic

205  location only once. The final dataset contained 9,984 unique locations of Wilson's warblers on breeding grounds.

206  We used the machine learning algorithm MAXENT (Phillips *et al.*, 2006) to model Wilson's warbler distributions. As

207  this method relies only on presence data (*i.e.*, it does not require records of Wilson's warbler *absence*), it is particularly

208  well-suited for the use of eBird occurrence records to capture complex biological responses to environment (Elith *et al.*,

209  2011), and has performed well in previous statistical comparisons with other species distribution modeling techniques

210  (Harrigan *et al.*, 2014) . Environmental variables were chosen to best reflect unique predictive ability, and variables

211  were removed from final analyses when a Pearson's correlation coefficient was >0.7 among the variable and any other

212  predictor. We ran 3 replicates of a MAXENT model using the Wilson's warbler records as a response variable and

213  14 climate and landscape variables as predictors. These included 8 (among 19 available) bio-climatic layers (Bio

214  1,2,4-6,12,15,19, downloaded from `www.worldclim.org`) as well as variables representing landcover, tree cover,

215  elevation, and vegetation characteristics and heterogeneity [Normalized Difference Vegetation Index (NDVI) mean and

216  standard deviation] downloaded from the Global Land Cover Facility (`www.landcover.org`). Default parameters

217  were used for all MAXENT runs except for the following: clamping turned off, a maximum of 100,000 background

218  points selected, and 20% of data withheld for subsequent testing. In addition, both predictor response curves, as well as

219  jackknifing to assess variable importance, were created as part of the output.

220  Once final models were evaluated and found to converge on similar response curves and probability estimates, the mean

221  output ascii file was exported to ArcGIS and was clipped to the extent of the breeding range of Wilson's warbler, as

222  determined using digital range maps provided by NatureServe (`www.natureserve.org` , (Ridgely *et al.*, 2005)).

223  These final probability estimates were then downsampled in resolution to yield a raster $\mathbf{M}^{\text{hab}}$ with the same extent and

224  resolution as $\mathbf{M}^{\text{gen}}_i$ and $\mathbf{M}^{\text{iso}}_i$.

## 2.5 Combining data types

226  The above approaches yield $\mathbf{M}^{\text{gen}}_i$ and $\mathbf{M}^{\text{iso}}_i$ for the $i^{\text{th}}$ bird, as well as a prior probability of occurrence of a Wilson's

227  warbler in any location, $\mathbf{M}^{\text{hab}}$. Treated as matrices of probabilities, these rasters can be combined using Bayes' rule to

228  obtain a combined estimate of probability of spatial origin for each bird. In order to explore weighting the different

229  sources of information (genetics, stable isotopes, and habitat) differently, we considered incorporating the parameter

230  $\beta = (\beta_{\text{iso}}, \beta_{\text{hab}})$ that determines the relative amount of weight given to the stable isotope and habitat data, and hence

231  compute the combined probability as

$$\mathbf{M}_i^{\mathrm{comb}}(\beta) = \frac{1}{C(\beta)} \mathbf{M}_i^{\mathrm{gen}} \left(\mathbf{M}_i^{\mathrm{iso}}\right)^{\beta_{\mathrm{iso}}} \left(\mathbf{M}^{\mathrm{hab}}\right)^{\beta_{\mathrm{hab}}}, \tag{3}$$

233  where $\left(\mathbf{M}\right)^{\beta}$ denotes the matrix $\mathbf{M}$ with every element raised to the power $\beta$. $C(\beta)$ is a normalizing constant that

234  ensures that the elements of $\mathbf{M}_i^{\mathrm{comb}}(\beta)$ sum to one. It can be easily found for any $\beta = (\beta_{\mathrm{iso}}, \beta_{\mathrm{hab}})$ by simply summing

235  over all the elements of the unnormalized matrix.

236  Among the different genetic groups of birds, the relative utility of isotopic data versus habitat suitability may vary. In

237  theory, one could leverage this to advantage by choosing different values of $\beta_{\mathrm{iso}}$ and $\beta_{\mathrm{hab}}$ for birds that are assigned

238  with genetics to particular genetic groups. However, doing so accurately would require sampling of reference birds

239  proportionally with respect to their true density on the landscape, which may not be the case with the genetic samples

240  used herein, which were obtained opportunistically from bird banding stations. Accordingly, while we explored a range

241  of values of $(\beta_{\mathrm{iso}}, \beta_{\mathrm{hab}})$, for all results in the paper we used values of $\beta_{\mathrm{iso}} = \beta_{\mathrm{hab}} = 1$.

242  **2.6 Assessment of Accuracy**

243  To assess the extent to which the inclusion of stable isotope and habitat data improve the spatial localization of birds on

244  the breeding grounds we test how well birds of known breeding location (those in the reference data set) can be inferred

245  using genetics alone, stable isotopes alone, habitat suitability alone, or a combination of each. When doing so with the

246  reference birds, we use a leave-one-out procedure, removing individual $i$ from the reference data set when computing

247  $\mathbf{M}_i^{\mathrm{gen}}$ and $\mathbf{M}_i^{\mathrm{iso}}$.

248  To assess accuracy we develop an easily interpreted metric, $S_i^{\mathrm{pm}}$, the posterior mean of great circle distance between

249  the true location of $i$ and inferred locations. This is found by averaging the distance between $i$'s true location and the

250  center of every cell in the breeding range, weighted by $\mathbf{M}_i$. In order to assess the accuracy of genetics, stable isotopes,

251  or habitat used separately or together, we compute $S_i^{\mathrm{pm}}$ with $\mathbf{M}_i^{\mathrm{iso}}$, $\mathbf{M}_i^{\mathrm{gen}}$, $\mathbf{M}^{\mathrm{hab}}$, or $\mathbf{M}_i^{\mathrm{comb}}(\beta)$, respectively.

252  **2.7 Assignment of Unknown Migratory Birds - Example Dataset**

253  To illustrate the efficacy of the combined approach we applied the combined approach to a set of migratory birds of

254  unknown origin captured at a stopover site in Cibola, AZ during spring migration. For each bird, we determined the

255  combined probability of origin using equation 3, creating a matrix of probability assignments, $\mathbf{M}_i^{\mathrm{comb}}$, for each migrant.

256  We then calculated the posterior mean migration distance remaining from the stopover site by averaging the distance

257  between the stopover site and the center of every cell in the breeding range, weighted by $\mathbf{M}_i^{\mathrm{comb}}$.

## 3. Results

### 3.1 Genetics

As previously reported (Ruegg *et al.*, 2014), the *maximum-a-posteriori* self-assignments of the reference birds using genetic data showed that a high fraction (88.2%) of birds are correctly assigned to their true region of origin. The rate of correct assignment varied across regions, with the lowest rate of correct assignment occurring in Coastal California (78%), followed by Rocky Mountain (81%), Pacific Northwest (85%), California Sierra (86%), Alaska to Alberta (95%), and Eastern North America (100%). Of the birds that are assigned correctly, their average maximum posterior is 96% and when birds are incorrectly assigned to their true region of origin, the average maximum posterior is 74%. This confirms that when genetic assignments are incorrect, isotopes and suitability have the potential to increase posteriors, and thereby improve localizations in joint assignments.

### 3.2 Stable Isotopes

Across the Wilson's warbler breeding range, $\delta^2 H_p$ values range from -14.3‰ to -190.6‰ with uncertainty across the $\delta^2 H_p$ isoscape varying (SD: 8.53 to 21.64) as a result of the uneven distribution of precipitation stations used to model $\delta^2 H_p$ ratios. Regions with the greatest uncertainty in $\delta^2 H_p$ values included coastal Alaska, Northwest Territories, and Newfoundland. The rescaling equation used to convert the $\delta^2 H_p$ isoscape to a $\delta^2 H_f$ isoscape was $\delta^2 H_f = 0.74(\delta^2 H_p) - 38.01$; that is, in Equation 1, $\bar{a} = 0.74$ and $\bar{b} = -38.01$. The standard deviation associated with the rescaled precipitation to feather $\delta^2 H$ isoscape, $\mathbf{R}^{(\sigma)}$, ranged from 0.61 to 6.56, while the mean, within-site standard deviation of $\delta^2 H_f$ from the 357 Wilson's warblers of known breeding origin sampled across 30 locations was $\sigma_{\text{indiv}} = 12.49$ (Supplement Table 1). The total variability, $\widetilde{\mathbf{T}}^{(\sigma)}$, including the three sources of uncertainty, contained standard deviation values ranging between 15.15 and 24.99, depending on the geographic region of assignment.

### 3.3 Habitat Suitability

Habitat suitability model runs yielded accurate species probability of occurrence maps according to several criteria. First, the mean score of all replicates using only data withheld for testing was high, with an AUC= 0.938. Standard deviations in AUC between each replicate run were minimal (lowest AUC = 0.937, highest AUC = 0.939). Additionally, environmental predictors deemed most important in explaining presence of Wilson's Warblers were consistent across all replicate runs, with Bio 19 (Precipitation of the coldest quarter), Bio 5 (Maximum temperature of the warmest month), and Bio 4 (Temperature Seasonality) contributing to over 70% of the variation in occurrence explained by the MAXENT model. Finally, our final map displaying the point-wise mean probability of occurrence (Fig. SX) closely matched
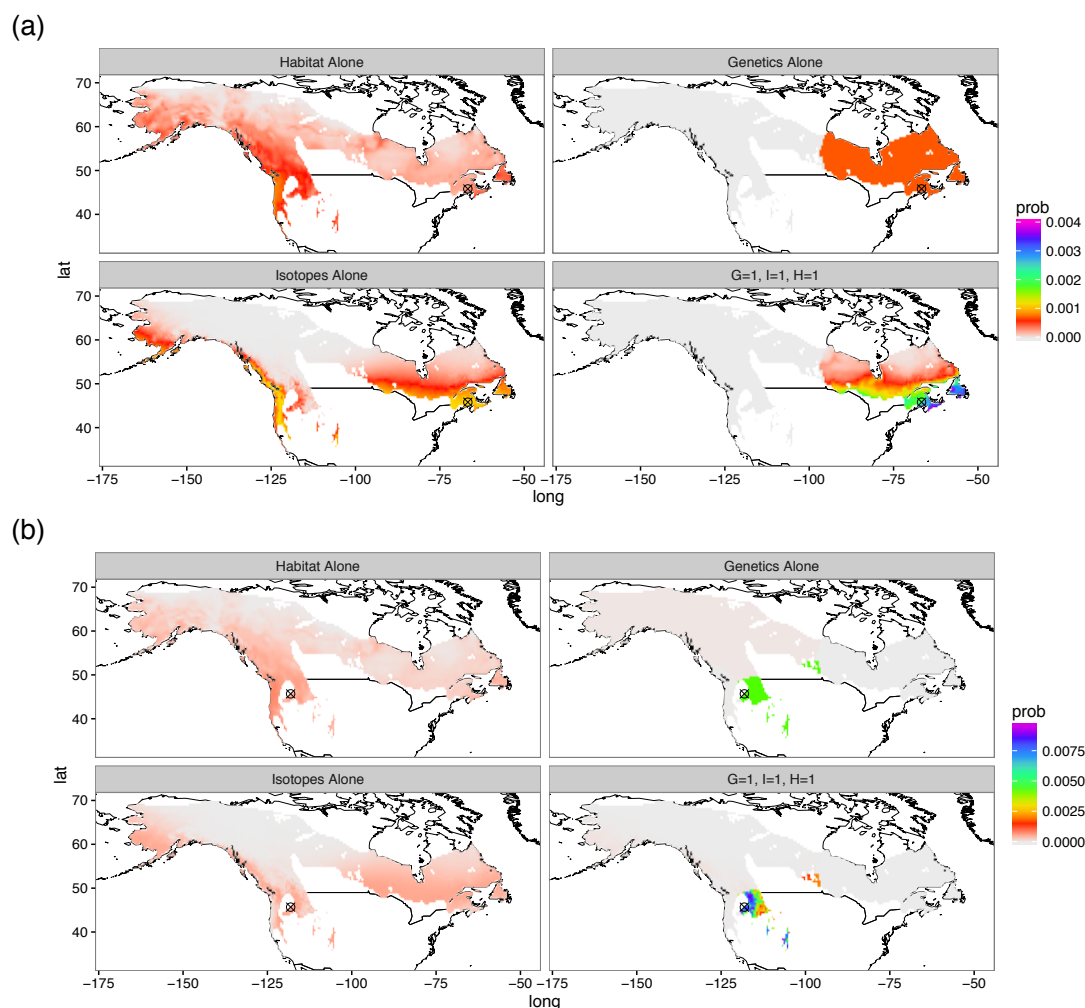
**Figure 1.** Representative posterior probability surfaces for 2 birds: (*a*) a bird from the Eastern region, and (*b*) a bird from the Rocky Mountains region. Each of the four panels shows the result using a different type of data. The first three ("Habitat alone," "Genetics Alone," and "Isotopes alone," show results for each data type applied separately. "$G = 1, I = 1, H = 1$" shows the results from the combined approach with $\beta_{\text{iso}} = \beta_{\text{hab}} = 1$. Maps which showed particularly strong examples of the combined approach are shown here, while maps for all 367 reference birds can be found in Supporting Information and to which we refer those interested in the variability in assignments across methods and individuals. In general, the maps support the idea that combining methods provides the most accurate assignment probabilities.

previously published maps of the species distribution of Wilson's Warbler (for instance, probability of occurrence was consistently <0.23 in regions outside of the species range, and habitat suitability generally identified regions that were also found to harbor high abundance of Wilson's warblers according to previously published data (Status of Birds in Canada, Peter Blancher, based on BBS abundance map estimates).

From predictor response curves, we found that increases in precipitation in the coldest quarter (Bio19) and temperature in the warmest month (Bio5) led to increases in the probability of Wilson's warbler occurrence, whereas increases in temperature seasonality (Bio4) led to decreases in probability of occurrence (Fig. SX). These relationships resulted in several regions of high habitat suitability being identified within the Wilson's warbler range, including much of the Pacific coastline, the Sierra Nevada Mountain range, parts of the Rocky Mountain range, and the Canadian Maritimes for eastern breeding populations.

## 3.4 Comparison of data types

Overall, combining genetics, isotopes, and habitat suitability improved the inference of breeding origin of Wilson's warblers. The posterior mean great circle distance between the true location and inferred location for nearly every bird in the reference dataset was decreased by combining all three sources of data (Figure 2).

Of the three individual data sources, genetic data provided the most accurate localization of individuals, while isotope assignments and the habitat suitability prior used alone resulted in significantly less accurate localization. There were also significant differences in the accuracy of assignments of the reference birds to breeding location based upon geographic region (Figure 3): genetic assignments performed best in all regions; for birds originating from the Pacific Northwest, the Sierras and the Rocky Mountains, more accurate assignments were achieved by using the prior information from habitat suitability than using data on stable isotopes alone; and conversely isotope-only assignments outperformed assignment using just the habitat suitability prior in Coastal California and the Eastern United States.

## 3.5 Example Data - Timing of migration in Pacific Northwest Wilson's warblers

In order to illustrate our combined approach on real-world migratory data, we calculated the posterior mean remaining migratory distance of birds sampled from Cibola, AZ during the spring migrations of 2008 and 2009. A general pattern, previously suggested in Ruegg et al (2014), of birds en route to Coastal California migrating through before birds en route to the Pacific Northwest, the Sierras, and Alaska to Alberta, was reinforced using our combined-data approach (Figure 4*a*). In addition, we found previously undetected patterns in the timing of migrants en route to the Pacific Northwest, with migrants headed to the southern Pacific Northwest arriving earlier than migrants en route to northern Pacific Northwest. These results were concordant across both years (Figure 4*b*).
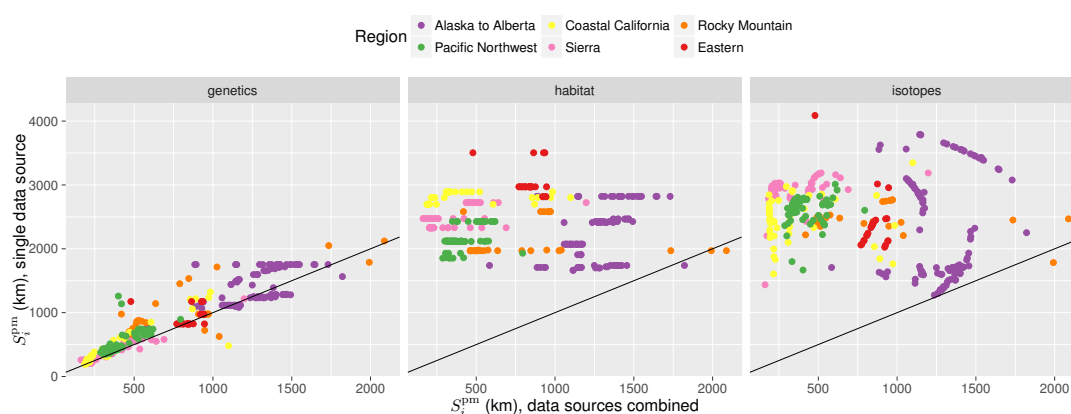
**Figure 2.** A comparison of the posterior mean great circle distances, $S_i^{\mathrm{pm}}$, between individual data types and the combined approach. Each point represents one bird. In all panels the *x*-axis shows $S_i^{\mathrm{pm}}$ using all data sources combined and the *y*-axis shows $S_i^{\mathrm{pm}}$ when using a single data source. The proximity of the points to the $y = x$ line shows how well each method does compared to the combined approach. The results clearly demonstrate that genetic data provide the most accurate inference of the true origin, while habitat suitability and stable isotope data provide more modest improvements. Nonetheless the fact that the majority of points lie above the $y = x$ line in the "genetics" plot confirms that the addition of habitat and stable isotope data improves the inference of breeding origin (as all the points would fall on the $y = x$ line if the inclusion of stable isotopes and habitat did not change the inferences made with genetics alone).
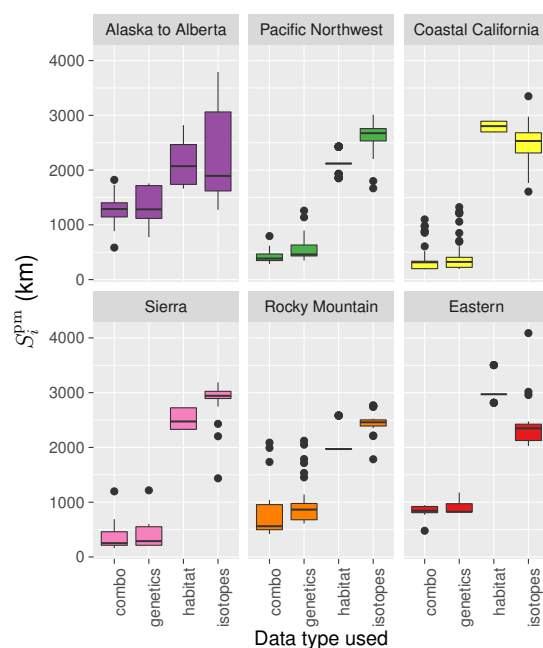
**Figure 3.** The accuracy of each type of data as estimated by the posterior mean great circle distance to the true location by region. The figure demonstrates that genetic data alone contributes the most to the accuracy of assignments to the true breeding location across all regions, while isotopic and habitat suitability data provide more modest improvements that differ in strength by region.
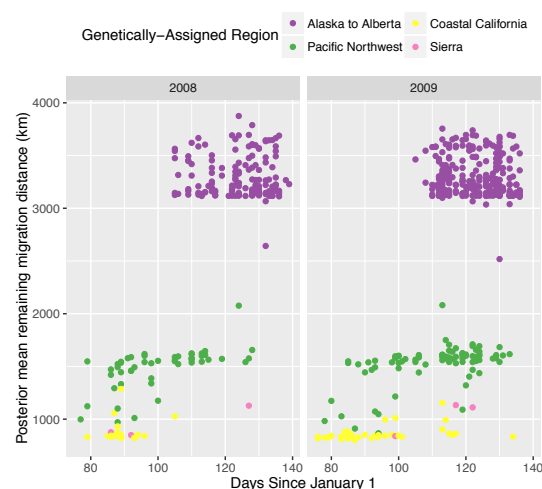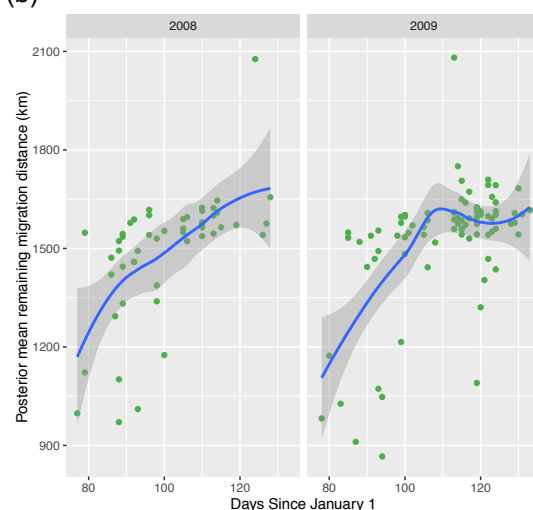
**Figure 4.** Utility of the combined approach demonstrated with spring Wilson's warbler migrants stopping over in Cibola, AZ: (a) The posterior mean remaining migration distance by day of the year for all birds stopping over in Cibola AZ during spring migration across the years of 2008 and 2009. (b) The posterior mean remaining migration distance by day of the year for genetically-assigned Pacific Northwest birds stopping over in Cibola, AZ during spring migration, demonstrating a previously undetected pattern of birds en route to more northern regions of the Pacific Northwest arriving later than birds en route to the southern portion of the Pacific Northwest.

## 4. Discussion

Tracking the origins of migratory animals using non-invasive, intrinsic marking techniques has been a particularly challenging endeavor for movement ecologists. Here we develop a novel method for combining three independent data sources (genetics, stable isotopes, and habitat suitability models) that improves upon the accuracy of each method when used on its own. Using leave-one-out cross validation to compare the relative accuracy of each method independently, we find that genetic data alone provides the most accurate estimation of the true origin of our reference birds, consistently outperforming assignments based upon stable isotopes and habitat suitability models (Figure 1, Figure 3). Improvements to the R-code for batch isotopic and genetic assignments initially developed by Vander Zanden *et al.* (2014) and Anderson *et al.* (2008) respectively, were compiled into an R-package called GAIAH in an effort to make future implementations of the resulting combined approach feasible across a broad range of migratory systems. Below we discuss region- and data-type-specific differences in the assignments, illustrate the utility of our combined method for uncovering new patterns of connectivity across time in Wilson's warblers captured during spring migration, and discuss the implications of our results for future studies considering limitations in both time and resources.

**(1) Accuracy of each data type relative to the combined approach by region**

A comparison between the posterior mean great circle distances using each method alone conclusively demonstrates that genetic-only assignments most closely approximate the true origin of birds in our reference sample (Figure 3, Figure 2). It is important to note that our analysis is based upon a single species and one might wonder whether our results are limited to Wilson's warblers or whether they will be generalizable across other migratory systems. While the accumulation of additional data is needed to more thoroughly answer this question, results from numerous other molecular-based studies suggest that the increase in accuracy gained from using high-resolution genetic tags will have wide utility for population assignment in migratory animals far beyond our single species application (Nielsen *et al.*, 2009; Hess *et al.*, 2011; Nielsen *et al.*, 2012; Ruegg *et al.*, 2014). It is also important to note that in Wilson's warblers the genetic variation was clustered into distinct regions, but in cases where genetic variation changes more gradually as a function of geographic distance (*i.e.*, isolation-by-distance), it would be possible to construct the genetic posterior matrices, $\mathbf{M}_i^{\text{gen}}$, by using an assignment approach designed to deal with continuous spatial data (Wasser *et al.*, 2004; Rañola *et al.*, 2014). Such an approach is implemented in the package ISOSCAT (Rundel *et al.*, 2013), but could be improved by incorporating a genetic model that allows for allele frequencies to change non-linearly with geography (*e.g.*, Rañola *et al.* 2014). Finally, the improvement in assignments based on genetic data alone are not surprising, considering that inferences based upon high-resolution genetic tags are coming from 96 axes that have been selected for informativeness in population level assignments, while inferences based upon stable isotopes are traditionally based upon only one or a few axes.

346  While high-resolution genetic markers alone provide a powerful framework for population-level assignment, the

347  inclusion of stable isotope and habitat suitability data served to increase the precision and accuracy of breeding origin

348  estimates in almost all cases (Figure 3). Surprisingly, for birds in some regions, such as the Rocky Mountains, the

349  Pacific Northwest and the Sierras, simply using the habitat suitability prior alone provided a better inference of the

350  true origin than did stable hydrogen isotope data. This is largely driven by the fact that continental variation in stable

351  hydrogen isotopes in precipitation in North America is greater across latitudinal than longitudinal gradients (Hobson &

352  Wassenaar, 2008), resulting in similar stable hydrogen isotope values between the Pacific Northwest and Sierra regions

353  and comparable longitudes in the Rocky Mountains. Habitat suitability models have more power than stable isotopes

354  in delineating among regions of similar longitudes that have high environmental heterogeneity, such as in the interior

355  western regions of the United States. Whereas, in ecologically more homogeneous regions that vary across latitudes

356  such as Alaska to Alberta, Coastal California, and Eastern North America, stable hydrogen isotope data provide more

357  information about the true region of origin than does the inclusion of habitat suitability models.

358  Despite improvements in assignments using a combined approach, it still remains difficult to accurately identify the

359  origin of individuals from breeding areas within the region from Alaska to Alberta using our methods. Genetic data

360  alone provides only a very coarse inference of the geographic origin, with the spatial scale of assignments to this

361  region being at best on the order of several thousand square kilometers. As a result, the data in Figure 4a appear

362  discontinuous because our assignments are based upon the posterior mean remaining migration distance and the region

363  from Alaska to Alberta is large. The inclusion of habitat suitability and stable isotope data helps little, mostly due to the

364  fact that habitats are equally suitable and stable isotope values are similar in a large area of South-Central Alaska and

365  Southern Interior British Columbia. Preliminary, unpublished results from the genetic data set reported in Ruegg *et al.*

366  (2014) suggest that additional genetic differentiation may exist in this region and it is possible that increased genetic

367  sampling could reduce the spatial scale of assignments. Alternatively, birds breeding from Alaska to Alberta may form

368  a more homogeneous population in general, with birds having more flexibility in the timing of migration, in which case

369  increased genetic sampling may provide limited benefits.

370  **(2) Demonstration of the utility of the combined approach**

371  We demonstrated the utility of the combined method by assigning 686 Wilson's warblers migrating through Cibola, AZ

372  during the spring of 2008 and 2009, back to their most probable breeding destination. The results identify previously

373  undetected patterns in the timing of migrants en route to the Pacific Northwest, with Wilson's warblers en route to

374  more southern locations in Northern California migrating through earlier than birds en route to more northern locations

375  near the Washington, British Columbia boarder (Figure 4b). These results were consistent across years and further

376  corroborate similar patterns of leap-frog migration in western Wilson's warblers identified in previous work on a much

377  coarser continental scale (Clegg *et al.*, 2003; Paxton *et al.*, 2007, 2013; Rundel *et al.*, 2013). Interestingly, similar

378 leapfrog patterns were not seen in migrants en route to breeding regions ranging from Alberta to Alaska (Figure 4), but

379 is not clear whether this is due to a lack of resolution in the markers or the lack of a leap-frog pattern in this region. In

380 addition, similar leapfrog patterns were also not observed within the California population where genetic resolution was

381 high and stable isotopes provided the second most informative measure of the true location of reference birds. Overall

382 these data illustrate the power of the combined approach for identifying fine scale patterns in migratory behavior which

383 have not been detected using any other methodology to date.

384 **(3) Recommendations for future migratory-connectivity studies**

385 Our analyses revealed that if operating under time or cost restraints, assignment accuracy using genetics alone can

386 provide a nearly optimal estimation of accuracy without the need to add additional data sources. This is particularly

387 true in cases where the volume of individuals to be screened is high (1000's of individuals) because at that point the

388 per-individual cost is less. Because of the ease with which high volumes of samples can be screened, intrinsic markers

389 provide a powerful tool for statistically rigorous assessments of migration phenology. Despite this fact, it is important to

390 note that the development of high-resolution genetic markers requires a substantial initial investment in highly-trained

391 personnel able to process genome-wide data, as well as in sequencing and reagent costs. As such, this method may not

392 be feasible in all situations. This initial investment is likely to pay dividends in the long-term, however, and the utility

393 of SNPs for population assignment cannot be understated. Once a basemap of genetic variation across geographic

394 space (i.e. a genoscape) has been produced, it can provide an extremely valuable resource for tracking populations and

395 documenting likely range shifts due to climate change or other anthroprogenic stressors. In addition, the development

396 of these markers can be utilized for a number of other purposes such as the identification of biologically meaningful

397 population boundaries, studies of introgression, hybridization, parentage, kinship and effective population size among

398 others (Andrews *et al.*, 2016; Garner *et al.*, 2016).

399 Despite the broad utility of genetic markers, many research projects will be faced with limitations in either time and/or

400 resources. In these cases, our analyses indicate that a combination of habitat suitability and stable isotope data provides

401 the second best option to genetics alone (SI 1, Fig. 2). Our results suggest that the only region where the inclusion

402 of habitat suitability did not improve the accuracy of our assignments was in Eastern North America. In this region,

403 the more the suitability models were weighted, the less accurate the resulting assignments became. We suspect a

404 number of reasons for this trend; 1) records of our study species were much more limited in this region, leading to less

405 accurate species distribution probabilities in eastern habitats, 2) a single species distribution model was created as a

406 prior across the entire breeding range, which led to a model that was too general to capture nuances in the eastern part

407 of the breeding range, particularly because a separate study using habitat suitability models for the Wilson's warbler

408 found little niche overlap between eastern and western groups, even suggesting that they might be considered cryptic

409 species (Ruiz-Sánchez *et al.*, 2015) and 3) eastern reference individuals were not sampled uniformly across the actual

breeding range, leading to low-deviance (but inaccurate) estimates of origin (see Figure 3).

Even with these limitations, habitat suitability models based on individual occurrence records have proved to be one of the most effective tools for understanding species distributions, ecological requirements, and realized niches (Elith *et al.*, 2011; Harrigan *et al.*, 2014). As such we recommend their inclusion whenever possible. Habitat suitability models are also likely to be the easiest of the three described methods to apply for a given study species: their only required input is records of animal locations and a suite of environmental predictors for the study region. However, ease-of-use should be tempered with caution when utilizing these methods as priors. For instance, habitat suitability priors may purposefully down-weight regions that could represent actual origin locations of individuals, particularly when species are capable of surviving in marginal habitats, or where estimates of habitat suitability and population abundances do not necessarily align. It is vital that occurrence data be sampled adequately and in a manner representative of the entire habitat available to the species, as sampling effort and intensity can greatly affect final model results (as seen in our eastern individuals) (Phillips *et al.*, 2006; Elith *et al.*, 2011). Understanding the ecological mechanisms of the study species, and an informed interpretation of the results of any species distribution model is paramount, and can often lead to surprising insights (Elith *et al.*, 2011; Renner & Warton, 2013; Harrigan *et al.*, 2014). Despite these caveats, our method suggests that the inclusion of habitat suitability model output as a prior in assignment tests could prove valuable to increase the accuracy of estimates.

Well established patterns of stable hydrogen isotopes in precipitation that correspond to isotopic signatures in animal tissues (Cryan *et al.*, 2004; Hobson & Wassenaar, 2008; Hobson *et al.*, 2012b) have facilitated the widespread use of stable isotopes to understand migratory connectivity in birds (Hobson *et al.*, 2010), bats (Cryan *et al.*, 2014), butterflies (Hobson *et al.*, 1999; Brattström *et al.*, 2008), and dragonflies (Hobson *et al.*, 2012a). Like genetic sampling, the collection of tissue samples (e.g., feathers, hair, insect chitin) for isotopic analysis does not require extensive training, and collection and analysis of a large numbers of samples is fairly inexpensive compared to extrinsic tracking devices. While it is important to understand the relationship between the tissue sampled and patterns of stable isotopes in the environment, technical expertise in stable isotope preparation and analysis is not necessary because of the accessibility of reputable isotope labs using a comparative equilibrium approach for the analysis of stable hydrogen isotope samples (Wassenaar & Hobson, 2003). Moreover, analytical tools such as ISoMAP (Bowen *et al.*, 2014) and the R-package developed in this study increase the accessibility of technically challenging stable isotope analyses that incorporate error and prior probabilities to more scientists. These advantages have allowed for the coupling of stable isotope data with a wide range of disciplines, including neural biology (Barkan *et al.*, 2016), endocrinology (Covino *et al.*, 2016), and disease ecology (Gunnarsson *et al.*, 2012; Ogden *et al.*, 2015), greatly advancing our understanding of the ecology of migratory species throughout their annual cycle. Nevertheless, as demonstrated in this study the accuracy of stable isotopes alone to assign individuals to specific locations is much lower than when combined with high-resolution

genetic markers and habitat suitability. The use of stable isotopes alone is most successful when assessing patterns of migratory connectivity at broader spatial scales (e.g., geographic regions), and for migratory species with geographic distributions that span large latitudinal gradients with little longitudinal variation (Hobson & Wassenaar, 2008). The combination of stable hydrogen isotopes with other intrinsic markers that have more longitudinal structure such as other stable isotopes (e.g., Sellick *et al.* 2009), morphometrics (Delingat *et al.*, 2011), and genetic data (Clegg *et al.* 2003; Chabot *et al.* 2012, this study) can increase the accuracy of assignment. Additionally, longitudinal ambiguity can be reduced by constraining assignments to smaller geographic areas within a species overall range (e.g. western breeding locations only) based on known flyways or banding occurrence data that preclude assignment to some areas within a species range (Hobson *et al.*, 2009; Van Wilgenburg & Hobson, 2011).

**(4) Conclusions**

Identifying the population of origin of migratory animals using intrinsic-marker techniques is now feasible at increasingly small spatial scales. Here we show that genetic assignments far outweigh the accuracy of assignments based on isotopes or habitat suitability models alone. When logistically possible, the inclusion of all three data sources (genetics, stable isotopes, and habitat suitability) can, as in the case of the Wilson's warbler, serve to refine genetic-only estimates and reveal previously undetected patterns in the timing of migratory events. Initial (seemingly large) investments in developing high-resolution genetic markers must be weighed on a case-by-case basis, but undoubtedly provide a high return on investment for heavily managed species or species for which high volumes of samples are to be screened. Our results provide methodological recommendations and a framework for analysis that can be used to facilitate future advances in the field of movement ecology including the discovery of new migratory pathways and cryptic migratory species, as well as the tools necessary to manage declining taxa in the face of rapid ecological changes.

## Appendix A. Parametric bootstrap for calculation of isotope posterior probabilities

Here we describe the approach of Vander Zanden *et al.* (2014) for using the $y_i^{\text{iso}}$ values of birds sampled from known breeding locations in a parametric bootstrap procedure to derive values of $\mathbf{R}^{(\sigma^2)}$, $\bar{a}$, and $\bar{b}$. Let $\ell = 1, \ldots, L$ index $L$ different sampling locations from which at least two breeding birds were sampled for isotopes. Nearby locations can be merged into a single location if they are all in the same cell of the grid $\tilde{\mathbf{T}}$. For each location $\ell$ we computed the mean and standard deviation of the $y_i^{\text{iso}}$ values of the breeding birds there, *i.e.*,

$$\bar{y}_{(\ell)}^{\text{iso}} = \frac{1}{n_\ell} \sum_{i \in \ell(i)} y_i^{\text{iso}}$$

$$s_{(\ell)} = \left( \frac{1}{n_\ell - 1} \sum_{i \in \ell(i)} (y_i^{\text{iso}} - \bar{y}_{(\ell)}^{\text{iso}})^2 \right)^{\frac{1}{2}}$$

where $n_\ell$ is the number of breeding birds sampled at $\ell$ and $\ell(i)$ is the set of their indices. Then, to create a parametric bootstrap sample of the isotope data from the reference birds, at each $\ell$, $n_\ell$ values of $y_i$ were simulated from a normal distribution with mean $\bar{y}_\ell^{\text{iso}}$ and standard deviation $s_{(\ell)}$. This bootstrap sample was then regressed as the dependent variable against the values in $\mathbf{P}^{(\mu)}$ corresponding to each location $\ell$, as the independent variable, yielding, for each bootstrap replicate a linear relationship, parameterized by a slope and an intercept, between the predicted isotope ratios from IsoMap and those observed in birds from each location. The average over bootstrap replicates of the these slopes and intercepts, respectively, are used as $\bar{a}$ and $\bar{b}$ in Equation 1. The elements of the matrix $\mathbf{R}^{(\sigma^2)}$ in (2) are the variances over the bootstrapped values of the predicted feather isotope values. That is, for each row $r$ and column $c$:

$$\mathbf{R}_{r,c}^{(\sigma^2)} = \frac{1}{n_B - 1} \sum_{h=1}^{n_B} \left( a_h \mathbf{P}_{r,c}^{(\mu)} + b_h - \tilde{\mathbf{T}}_{r,c}^{(\mu)} \right)^2$$

where $a_h$ and $b_h$ are the slope and intercept estimated from the $h^{\text{th}}$ of $n_B$ bootstrap samples.

## References

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature reviews genetics*, **11**, 697–709.

Anderson EC, Waples RS, Kalinowski ST (2008) An improved method for predicting the accuracy of genetic stock identification. *Can J Fish Aquat Sci*, **65**, 1475–1486.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**, 81–92.

494  Arlt D, Low M, Pärt T (2013) Effect of geolocators on migration and subsequent breeding performance of a long-distance passerine migrant. *PloS*
495    *one*, **8**, e82316.

496  Barkan S, Roll U, Yom-Tov Y, Wassenaar LI, Barnea A (2016) Possible linkage between neuronal recruitment and flight distance in migratory birds.
497    *Scientific reports*, **6**.

498  Bonfil R, Meÿer M, Scholl MC, Johnson R, O'Brien S, Oosthuizen H, Swanson S, Kotze D, Paterson M (2005) Transoceanic migration, spatial
499    dynamics, and population linkages of white sharks. *Science*, **310**, 100–103.

500  Bowen GJ, Liu Z, Vander Zanden HB, Zhao L, Takahashi G (2014) Geographic assignment with stable isotopes in IsoMAP. *Methods in Ecology and*
501    *Evolution*, **5**, 201–206.

502  Brattström O, Wassenaar LI, Hobson KA, Åkesson S (2008) Placing butterflies on the map–testing regional geographical resolution of three stable
503    isotopes in Sweden using the monophagus peacock *Inachis io*. *Ecography*, **31**, 490–498.

504  Bridge ES, Kelly JF, Contina A, Gabrielson RM, MacCurdy RB, Winkler DW (2013) Advances in tracking small migratory birds: a technical review
505    of light-level geolocation. *Journal of Field Ornithology*, **84**, 121–137.

506  Chabot AA, Hobson KA, Van Wilgenburg SL, McQuat GJ, Lougheed SC (2012) Advances in linking wintering migrant birds to their breeding-ground
507    origins using combined analyses of genetic and stable isotope markers. *PLoS One*, **7**, e43627.

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

510  Covino KM, Jawor JM, Kelly JF, Moore FR (2016) Overlapping life-history stages in migrating songbirds: variation in circulating testosterone and
511    testosterone production capacity. *Journal of Ornithology*, 1–10.

512  Cryan PM, Bogan MA, Rye RO, Landis GP, Kester CL (2004) Stable hydrogen isotope analysis of bat hair as evidence for seasonal molt and
513    long-distance migration. *Journal of Mammalogy*, **85**, 995–1001.

514  Cryan PM, Stricker CA, Wunder MB (2014) Continental-scale, seasonal movements of a heterothermic migratory tree bat. *Ecological Applications*,
515    **24**, 602–616.

516  Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using
517    next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

518  Delingat J, Hobson KA, Dierschke V, Schmaljohann H, Bairlein F (2011) Morphometrics and stable isotopes differentiate populations of Northern
519    Wheatears (Oenanthe oenanthe). *Journal of Ornithology*, **152**, 383–395.

520  Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and distributions*, **17**,
521    43–57.

522  Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele
523    frequencies. *Genetics*, **164**, 1567–1587.

524  Gagnaire PA, Broquet T, Aurelle D, Viard F, Souissi A, Bonhomme F, Arnaud-Haond S, Bierne N (2015) Using neutral, selected, and hitchhiker loci
525    to assess connectivity of marine populations in the genomic era. *Evolutionary applications*, **8**, 769–786.

526  Garner BA, Hand BK, Amish SJ, Bernatchez L, Foster JT, Miller KM, Morin PA, Narum SR, O'Brien SJ, Roffler G, *et al.* (2016) Genomics in
527    conservation: case studies and bridging the gap between data and application. *Trends in ecology & evolution*, **31**, 81–83.

528    Guillot G, Mortier F, Estoup A (2005) GENELAND: a computer package for landscape genetics. *Molecular ecology notes*, **5**, 712–715.

529    Gunnarsson G, Latorre-Margalef N, Hobson KA, Van Wilgenburg SL, Elmberg J, Olsen B, Fouchier RA, Waldenström J (2012) Disease dynamics
530      and bird migration—linking mallards Anas platyrhynchos and subtype diversity of the influenza A virus in time and space. *PLoS One*, **7**, e35679.

531    Harrigan RJ, Thomassen HA, Buermann W, Smith TB (2014) A continental risk assessment of West Nile virus under climate change. *Global change*
532      *biology*, **20**, 2417–2425.

533    Hess J, Matala A, Narum S (2011) Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook
534      salmon in the Columbia River Basin. *Molecular Ecology Resources*, **11**, 137–149.

535    Hobson KA, Anderson RC, Soto DX, Wassenaar LI (2012a) Isotopic evidence that dragonflies (Pantala flavescens) migrating through the Maldives
536      come from the northern Indian subcontinent. *PloS one*, **7**, e52594.

537    Hobson KA, Barnett-Johnson R, Cerling T (2010) Using isoscapes to track animal migration, *in: Isoscapes*, 273–298, Springer.

538    Hobson KA, Soto DX, Paulson DR, Wassenaar LI, Matthews JH (2012b) A dragonfly ($\delta 2$H) isoscape for North America: a new tool for determining
539      natal origins of migratory aquatic emergent insects. *Methods in Ecology and Evolution*, **3**, 766–772.

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

542    Hobson KA, Wassenaar LI (2008) *Tracking Animal Migration with Stable Isotopes, 1st Edition*, Elsevier.

543    Hobson KA, Wassenaar LI, Taylor OR (1999) Stable isotopes ($\delta$D and $\delta$13C) are geographic indicators of natal origins of monarch butterflies in
544      eastern North America. *Oecologia*, **120**, 397–404.

545    Hobson KA, Wunder MB, Van Wilgenburg SL, Clark RG, Wassenaar LI (2009) A method for investigating population declines of migratory birds
546      using stable isotopes: origins of harvested lesser scaup in North America. *PLoS One*, **4**, e7915.

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

549    Kelly JF, Bridge ES, Fudickar AM, Wassenaar LI (2009) A test of comparative equilibration for determining non-exchangeable stable hydrogen
550      isotope values in complex organic materials. *Rapid Communications in Mass Spectrometry*, **23**, 2316–2320.

551    Kelly JF, Ruegg KC, Smith TB (2005) Combining isotopic and genetic markers to identify breeding origins of migrant birds. *Ecological Applications*,
552      **15**, 1487–1494.

553    Kimura M, Clegg S, Lovette IJ, Holder KR, Girman DJ, Milá B, Wade P, Smith TB (2002) Phylogeographical approaches to assessing demographic
554      connectivity between breeding and overwintering regions in a Nearctic-Neotropical warbler (*Wilsonia pusilla*). *Molecular Ecology*, **11**, 1605–1616.

555    Kraus RH, Vonholdt B, Cocchiararo B, Harms V, Bayerl H, Kühn R, Förster DW, Fickel J, Roos C, Nowak C (2015) A single-nucleotide polymorphism-
556      based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. *Molecular ecology*
557      *resources*, **15**, 295–305.

558    Marra PP, Hobson KA, Holmes RT (1998) Linking winter and summer events in a migratory bird by using stable-carbon isotopes. *Science*, **282**,
559      1884–1886.

560    Metzker ML (2010) Sequencing technologies—the next generation. *Nature reviews genetics*, **11**, 31–46.

561 Mitchell TD, Jones PD (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids.
562     *International journal of climatology*, **25**, 693–712.

563 Nielsen EE, Cariani A, Mac Aoidh E, Maes GE, Milano I, Ogden R, Taylor M, Hemmer-Hansen J, Babbucci M, Bargelloni L, *et al.* (2012)
564     Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature Communications*, **3**, 851.

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

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

568 Ogden N, Barker I, Francis C, Heagy A, Lindsay L, Hobson K (2015) How far north are migrant birds transporting the tick Ixodes scapularis in
569     Canada? Insights from stable hydrogen isotope analyses of feathers. *Ticks and tick-borne diseases*, **6**, 715–720.

570 Paritte JM, Kelly JF (2009) Effect of cleaning regime on stable-isotope ratios of feathers in Japanese Quail (*Coturnix japonica*). *The Auk*, **126**,
571     165–174.

572 Paxton KL, Van Riper III C, Theimer TC, Paxton EH, Hobson K (2007) Spatial and temporal migration patterns of Wilson's Warbler (Wilsonia
573     pusilla) in the southwest as revealed by stable isotopes. *The Auk*, **124**, 162–175.

574 Paxton KL, Yau M, Moore FR, Irwin DE (2013) Differential migratory timing of western populations of Wilson's Warbler (Cardellina pusilla)
575     revealed by mitochondrial DNA and stable isotopes. *The Auk*, **130**, 689–698.

576 Pekarsky S, Angert A, Haese B, Werner M, Hobson KA, Nathan R (2015) Enriching the isotopic toolbox for migratory connectivity analysis: a new
577     approach for migratory species breeding in remote or unexplored areas. *Diversity and Distributions*, **21**, 416–427.

578 Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological modelling*, **190**, 231–259.

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

580 R Core Team (2016) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

581 Rañola JM, Novembre J, Lange K (2014) Fast spatial ancestry via flexible allele frequency surfaces. *Bioinformatics*, **30**, 2915–2922.

582 Renner IW, Warton DI (2013) Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics*,
583     **69**, 274–281.

584 Ridgely RS, Allnutt TF, T Brooks T, McNicol DK, Mehlman DW, Young BE, Zook JR (2005) *Digital Distribution Maps of the Birds of the Western
585     Hemisphere, version 2.1.*, NatureServe.

586 Rubenstein D, Chamberlain C, Holmes R, Ayres M, Waldbauer J, Graves G, Tuross N (2002) Linking breeding and wintering ranges of a migratory
587     songbird using stable isotopes. *Science*, **295**, 1062–1065.

588 Ruegg KC, Anderson EC, Paxton KL, Apkenas V, Lao S, Siegel RB, DeSante DF, Moore F, Smith TB (2014) Mapping migration in a songbird using
589     high-resolution genetic markers. *Molecular Ecology*, **23**, 5726–5739.

590 Ruiz-Sánchez A, Renton K, Landgrave-Ramírez R, Mora-Aguilar EF, Rojas-Soto O (2015) Ecological niche variation in the Wilson's warbler
591     Cardellina pusilla complex. *Journal of Avian Biology*, **46**, 516–527.

592 Rundel CW, Wunder MB, Alvarado AH, Ruegg KC, Harrigan R, Schuh A, Kelly JF, Siegel RB, DeSante DF, Smith TB, *et al.* (2013) Novel statistical
593     methods for integrating genetic and stable isotope data to infer individual-level migratory connectivity. *Molecular Ecology*, **22**, 4163–4176.

594  Sellick MJ, Kyser TK, Wunder MB, Chipley D, Norris DR (2009) Geographic variation of strontium and hydrogen isotopes in avian tissue:
595      implications for tracking migration and dispersal. *PLoS One*, **4**, e4735.

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

597  Smith TB, Clegg S, Kimura M, Ruegg KC, Mila B, Lovette IJ (2005) Molecular genetic approaches to linking breeding and overwintering areas in
598      five Neotropical migrant passerines, *in: Birds of Two Worlds: The Ecology and Evolution of Migration* (edited by Greenberg R, Marra PP), John
599      Hopkins University Press, Baltimore.

600  Stutchbury BJ, Tarof SA, Done T, Gow E, Kramer PM, Tautin J, Fox JW, Afanasyev V (2009) Tracking long-distance songbird migration by using
601      geolocators. *Science*, **323**, 896–896.

602  Toews DP, Campagna L, Taylor SA, Balakrishnan CN, Baldassarre DT, Deane-Coe PE, Harvey MG, Hooper DM, Irwin DE, Judy CD, *et al.* (2015)
603      Genomic approaches to understanding population divergence and speciation in birds. *The Auk*, **133**, 13–30.

604  Van Wilgenburg SL, Hobson KA (2011) Combining stable-isotope ($\delta$D) and band recovery data to improve probabilistic assignment of migratory
605      birds to origin. *Ecological Applications*, **21**, 1340–1351.

606  Vander Zanden HB, Wunder MB, Hobson KA, Van Wilgenburg SL, Wassenaar LI, Welker JM, Bowen GJ (2014) Contrasting assignment of migratory
607      organisms to geographic origins using long-term versus year-specific precipitation isotope maps. *Methods in Ecology and Evolution*, **5**, 891–900.

608  Wassenaar L, Hobson K (2003) Comparative equilibration and online technique for determination of non-exchangeable hydrogen of keratins for use
609      in animal migration studies. *Isotopes in environmental and health studies*, **39**, 211–217.

610  Wasser SK, Shedlock AM, Comstock K, Ostrander EA, Mutayoba B, Stephens M (2004) Assigning African elephant DNA to geographic region of
611      origin: applications to the ivory trade. *Proc Natl Acad Sci USA*, **101**, 14847–52.

612  Webster MS, Marra PP, Greenberg R, Marra P (2005) The importance of understanding migratory connectivity and seasonal interactions. *Birds of two
613      worlds: the ecology and evolution of migration. Johns Hopkins University Press, Baltimore*, 199–209.

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