Signatures of mitonuclear coevolution in a warbler species complex

Silu Wang[1,2], Madelyn J. Ore[1,3], Else K. Mikkelsen[1,4], Julie Lee-Yaw[5,6], David P. L. Toews[7], Sievert Rohwer[8], and Darren Irwin[1]

[1] Department of Zoology, and Biodiversity Research Centre, 6270 University Blvd, University of British Columbia, Vancouver, BC, V6T1Z4, Canada
[2] Current address: Department of Integrative Biology, University of California, Berkeley, California, 94720, USA
[3] Current address: Cornell Lab of Ornithology, Ithaca, New York, 14850, USA
[4] Current address: Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, M1C 1A4, Canada
[5] Department of Botany, 3200-6270 University Blvd, University of British Columbia, Vancouver, BC, V6T1Z4, Canada
[6] Current address: Biological Sciences, 4401 University Drive, University of Lethbridge, Lethbridge, Alberta, T1K3M4, Canada
[7] Department of Biology, The Pennsylvania State University, University Park, Pennsylvania, 16802, USA
[8] Department of Biology and Burke Museum, Box 353010, University of Washington, Seattle, Washington 98195, USA

Abstract

Mitochondrial (mtDNA) and nuclear (nDNA) genes interact to govern metabolic pathways of mitochondria. When differentiated populations interbreed at secondary contact, incompatibilities between mtDNA of one population and nDNA of the other could result in low fitness of hybrids. In northwestern North America, two hybridizing species of warblers, *Setophaga occidentalis* (abbreviated as SOCC) and *S. townsendi* (STOW), provide an excellent system to investigate the potential co-adaptation of mitochondrial and nuclear DNA. The inland population of STOW (inland STOW) harbors mtDNA haplotype that is half a million years divergent from the SOCC mtDNA, and these populations also differ strongly in a few nDNA regions. Coastal populations of STOW (coastal STOW) have mixed ancestry, consistent with ancient hybridization of SOCC and inland STOW-like population. Of the few highly-differentiated nDNA regions between inland STOW and SOCC, one of these regions (on chromosome 5) is also differentiated between coastal STOW and inland STOW, and covaries with mtDNA among coastal STOW populations. Genes in this 1.2Mb region of chromosome 5 are associated with fatty acid oxidation and energy-related signaling transduction, both of which are closely associated with mitochondrial function. This chromosome 5 region is correlated with mtDNA haplotypes both within and across sampling sites, a pattern consistent with mitonuclear co-adaptation. We show that such mitonuclear coevolution might be maintained by climate-related selection, because mitonuclear ancestry is correlated with climatic conditions among sampling sites. Together, our observation suggests climatic-associated adaptation shaping mitonuclear differentiation and introgression in this species complex.

Introduction

Mitochondrial (mtDNA) and nuclear (nDNA) genomes co-function in maintaining critical functions that influence fitness in nearly all eukaryotes [1–5]. Populations in different areas may harbour distinct mtDNA sequences owing to selection or drift, and because many nuclear genes encode proteins that function within mitochondria, the genomes are expected to co-evolve, each being the target of selection favoring compatibility with the other [4,6,7]. Interbreeding at species boundaries can lead to sub-optimal mitonuclear combinations in hybrids. Specifically, hybrids with nDNA from one species and mtDNA from the other species may have lowered fitness [8,9]. These types of genetic incompatibilities can play a role in keeping hybrid zones narrow and limiting gene flow between species [10,11]. Hence coadaptation of mtDNA and nDNA is increasingly recognized as being important to speciation [10–12].

In at least some species, geographic variation in climate is known to select for different mitochondrial genotypes in different areas [6,13–15]. This may in turn lead to indirect selection on co-functioning nuclear genes. Such climatic mitonuclear coadaptation can lead to genomic differentiation between populations inhabiting different climatic conditions [4,6]. Here we examine the relationship between mtDNA and nDNA variation in a warbler species complex with ancient and ongoing hybridization. In particular, we ask whether there is a signature of mitonuclear coevolution and, if so, could climate-related selection have driven such coevolution?

While secondary contact between differentiated populations sometimes leads to narrow hybrid zones [16], another possible outcome is the formation of a hybrid or mixed population over a broad region [17–19]. Such populations have the potential to reveal strong

71 selection on suboptimal combinations of genes from the two parental species. Despite
72 increasing interest in mitonuclear interactions at species boundaries of natural
73 populations with complex population histories [3,6,20–23], the degree to which mitonuclear
74 interactions are important in the differentiation among lineages is not well understood.

75 The North America warbler species *Setophaga occidentalis* (abbreviated as
76 SOCC) inhabits conifer forests along the states of Oregon, California, and southern
77 Washington, U.S.A. To the north of SOCC, a closely related species *Setophaga*
78 *townsendi* (abbreviated as STOW) consists of an inland population that inhabits areas
79 east of the Coast Mountains of British Columbia, Canada and northern Washington, USA
80 (abbreviated as inland STOW) and a coastal population (abbreviated as coastal STOW)
81 west of the Coast Mountains (Figure 1 A) [24–28]. The SOCC and inland STOW populations
82 demonstrate distinct plumage and mtDNA haplotypes that are separated by ~0.8%
83 sequence divergence, diverged ~0.5 million years ago (Figure 1 BC) [24,26,27,29]. They differ
84 in nDNA mainly at a few small genomic regions, one related to plumage differences
85 (ASIP-RALY gene block), whereas the rest of the genome shows very little
86 differentiation [29].

87 The coastal STOW population is identical to inland STOW in plumage, but
88 harbors both SOCC and inland STOW mtDNA haplotypes (Figure 1 BC). The haplotype
89 network (Figure 1 C) suggests that this population is the product of ancient hybridization
90 between SOCC and inland STOW-like population [27] (Figure 1 A). If so, individuals
91 sampled from coastal STOW should demonstrate a mixture of ancestry between
92 contemporary SOCC and inland STOW, especially detectable in regions that are
93 differentiated between SOCC and inland STOW [30]. In addition, the plumage gene region

that differs between SOCC and inland STOW is expected to remain undifferentiated between coastal STOW and inland STOW, because of their similar appearance. In contrast, we might expect differentiation within coastal STOW with respect to nuclear genes associated with mitochondrial function according to differences in the frequency of inland STOW versus SOCC mtDNA haplotypes in different parts of the range.

Here, we analyze variation at tens of thousands of single nucleotide polymorphisms (SNPs) throughout the nuclear genome of various SOCC, coastal STOW and inland STOW populations. In particular, we ask (1) whether both nuclear and mitochondrial genomic data suggest hybrid origin of coastal STOW; (2) whether genomic differentiation between coastal and inland STOW is related to mitonuclear coevolution; (3) whether spatial variation in the frequency of mitonuclear genotypes is consistent with climate-related selection.

Methods

Museum samples, mtDNA sequences, and nDNA sequencing

As a baseline for understanding the relationships among mtDNA haplotypes and their distributions, sequences of the mtDNA NADH dehydrogenase subunit 2 gene (ND2) for 223 individuals (95 coastal STOW, 81 inland STOW, and 47 SOCC) from the Krosby and Rohwer (2009) [27] study were acquired from GenBank (accession numbers FJ373895-FJ374120). To further understand the relationships among various STOW populations with these mtDNA sequences, we generated a minimum spanning haplotype network [31] with PopART [32]. This network showed two clearly separated haplotype clusters. We then scored each haplotype as 0 and 1 respectively for those that are nested within the SOCC haplotype cluster and the inland STOW cluster [27] (Figure 1 C).

117 Among these individuals with previously-sequenced mtDNA (i.e., from [27]), we
118 selected a subset of tissue samples (70 inland STOW, 57 coastal STOW, and 15 SOCC;
119 obtained from the Burke Museum of Natural History and Culture, University of
120 Washington, Seattle, Washington) for nuclear genomic sequencing. We supplemented
121 this set of genetic samples with 54 blood samples that we obtained directly from birds
122 caught in the field during the breeding season of 2016; these included 47 SOCC from
123 California, 5 inland STOW from Montana, and 2 inland STOW from Washington.
124 Therefore, 196 samples (62 SOCC, 77 inland STOW, and 57 coastal STOW) were
125 included in the nuclear genomic sequencing.

126 **GBS pipeline**

127 We prepared genotyping-by-sequencing (GBS) [33] libraries from DNA samples of
128 the 196 individuals described above as our previous study [34]. Briefly, we digested
129 genomes with the restriction enzyme PstI, then ligated fragments with barcode and
130 adaptors, and amplified with PCR. Amplified DNA was pooled into two libraries which
131 were then paired-end sequenced at GenomeQuebec: the first (115 individuals) were
132 sequenced with an Illumina HiSeq 2500 automated sequencer (read length = 125bp), and
133 the second (81 individuals) were sequenced with an Illumina HiSeq 4000 (read length =
134 100bp) due to equipment updates at the sequencing facility. We randomly assigned
135 samples to different plates and included replicates of three samples among plates.
136 Sequence processing was based on that of a previous study [34], but with some changes
137 included in the summary below. Specifically, the reads were demultiplexed with a custom
138 script and then trimmed using Trimmomatic [35] [TRAILING:3 SLIDINGWINDOW:4:10
139 MINLEN:30]. We aligned reads to a congener *Setophaga coronata* reference [36] with

140 bwa-mem [37] (default settings). Variable sites were identified with GATK [38], then filtered
141 with VCFtools [39] according to the following criteria: 1) removing indels; 2) keeping sites
142 with genotype quality (GQ) > 20; 3) keeping sites with minor allele frequency (MAF) ≥
143 0.05; 4) removing sites with > 30% missing genotypes among individuals; and 5) keeping
144 biallelic single nucleotide polymorphisms (SNPs) only. Thereafter 222,559 SNPs
145 remained.

146 **Population structure and genomic differentiation**

147 To investigate if the coastal STOW is a product of genetic admixture of SOCC
148 and inland STOW, we first examine population structure with principle component
149 analysis (PCA) in the SNPRelate [40] package in R [41], followed by heterozygosity-ancestry
150 proportion analysis [42]. We originally set out to assess the differences between SOCC,
151 coastal STOW, and inland STOW. However, the PCA revealed strong clustering within
152 coastal STOW with Valdez, AK (USA) and Haida Gwaii, BC (Canada) populations
153 distinct from the rest of the coastal STOW populations. In subsequent analysis, we thus
154 compared each of the three coastal STOW groups (Haida Gwaii, Valdez, and other
155 coastal STOW [oc-STOW]) to the inland STOW and SOCC groups. To examine
156 population differentiation across the genome, for each of the 222,559 filtered SNPs we
157 calculated F_{ST} [43] with VCFtools [39] between 1) inland STOW (n = 77) and SOCC (n = 62);
158 2) coastal STOW (n = 57: 10 Haida Gwaii, 15 Valdez, 32 others) and inland STOW; and
159 3) SOCC and each of the three coastal STOW clusters.

160 Ancient hybridization is expected to result in individuals with variable ancestry
161 and reduced heterozygosity, in contrast to recent, early generation hybrids, which will be
162 heterozygous across most loci. To uncover signatures of ancient hybridization, we

163 calculated ancestry and inter-specific heterozygosity of the coastal STOW individuals.
164 We consider SNPs with $F_{ST} > 0.6$ (between inland STOW and SOCC) as being useful for
165 informative parental ancestry. There were 50 SNPs distributed across chr 1, chr 1A, chr
166 4A, chr 5, chr 20, and chr Z. For each SNP, we assigned an ancestry score of 0 and 1 for
167 homozygous SOCC and homozygous inland STOW variants respectively, and 0.5 for
168 heterozygotes. We calculated ancestry (averaged ancestry score of the 50 SNPs) and
169 heterozygosity of the 50 ancestry informative SNPs. To ensure that neutral sites besides
170 selected genomic regions are considered, we repeated heterozygosity-ancestry analysis
171 with a F_{ST} threshold of 0.3, which involves 618 SNPs.
172 The 50 SNPs (from above) at F_{ST} peaks between coastal STOW and inland
173 STOW that are also consistent with the peaks between SOCC and inland STOW were
174 considered candidate loci for further analyses. One possibility is that these loci are linked
175 to genes that have a mitochondrial function and selection maintains their concordance
176 with mtDNA ancestry.
177 **Association of mtDNA and nDNA**
178 Coevolution between genomes is expected to lead to association between mtDNA
179 and mt-associated nuclear genes within and among sites. If individuals with mismatched
180 mt-nDNA genotypes are selected against, there should be an association between mtDNA
181 and nDNA genotypes within each population. Such a force could be counteracted by
182 random mating which breaks down the mt-nDNA association, thus strong selection is
183 required to maintain adaptive mt-nDNA combinations within a single randomly mating
184 population. Over time however, specific geographic regions may favor particular mtDNA

185 variants and thus compatible nDNA variants, increasing mt-nDNA concordance among
186 sampling sites.

187 To examine within-population association between mtDNA and nDNA ancestry,
188 we conducted a genome-wide association study (GWAS) with GenABEL [44] in R to
189 examine if there is association between mtDNA group (0 for SOCC type, or 1 for STOW
190 type) and nuclear genotypes across the genome while controlling for population
191 substructure. We used the *egscore* function that tests association between mtDNA
192 ancestry and nuclear SNP ancestry while controlling for population substructure using the
193 identity-by-state kinship matrix (calculated by the function *ibs*)[44]. We subsequently
194 conducted genomic control assuming that randomly selected sites are not associated with
195 mtDNA. To do so, we calculated inflation factor λ (accounting for both the effects of
196 genetic structure and sample size) and used it to correct for the test statistic χ^2 (and thus
197 *p-values*) of each mt-nDNA association test. A *p-value* cutoff of 10^{-6} was employed to
198 account for multiple hypothesis testing for sites across the genome as a balance of
199 Bonferroni correction (0.05/222,559 SNPs $\approx 2.2\times 10^{-7}$) and dependency among linked
200 SNPs. To examine gene function related to the loci associated with mtDNA haplotypes,
201 we examined known protein-coding genes in vicinity of the candidate SNPs, using the *S.*
202 *coronata* annotation [36]. For each gene, we identified the *Taeniopygia guttata* homolog [45],
203 which was then searched for Gene Ontology molecular and biological functions with
204 UniProt [46].

205 Although within-population mitonuclear association could be largely prevented
206 by random mating, a signature of mitonuclear coevolution might be captured in the
207 mitonuclear ancestry associations among populations. We first calculated the mean

208 mtDNA and nDNA ancestry of each site by averaging the locus-specific ancestry (0 for
209 SOCC and 1 for inland STOW) among individuals. To examine between-population
210 association between mtDNA and nDNA variants of interest, we employed a partial
211 mantel test [47] with the *vegan* package in R to quantify the association between the
212 among-sites (N =19) distance matrices of mtDNA ancestry and the nDNA ancestry, while
213 controlling for spatial autocorrelation. In particular, the partial mantel test examined
214 correlation between the among-sites distance matrix of mtDNA and that of the nDNA
215 locus, conditioned on the among-sites distance matrix of coordinates.

216 **mtDNA sequence analysis**

217 To add additional context to any potential functional interpretation of nuDNA –
218 mtDNA interactions, we compared sequence divergence of protein coding genes in full
219 mitochondrial genomes of SOCC and STOW. We used published whole-genome data
220 from tissues sequenced by Baiz et al. (2021) [36] to extract high quality (>1000X coverage)
221 mtDNA reads (n = 4 SOCC haplotypes, n = 6 STOW haplotypes; see [36] for details on
222 library preparation, sequencing methodology, and sample information). Reads were
223 aligned to the *S. coronata* mitochondrial genome, which we annotated using the MITOS[48]
224 web server annotation platform. To generate individual mtDNA fasta files, we used the
225 "mpileup" command in samtools version 0.1.18 [49], the "call -c" command in bcftools
226 version 1.10.2, and the "vcf2fq" option to generate fastq files. Finally, we used the
227 program seqtk (version 1.3; https://github.com/lh3/seqtk) to translate into fasta format.
228 We generated sequence alignments, calculated the number of fixed SNPs between the
229 divergent haplotypes, and determined number of fixed amino acid substitutions in
230 Geneious (11.0.3).

231 **Climate analysis**

232 To investigate whether there might be selection on mt-nDNA related to climate,
233 we tested association of site-level mt-nDNA ancestry (the averaged site ancestry score of
234 mtDNA, chr 5 differentiation block ancestry) and climate variation. To effectively
235 capture annual climate variation among sites, we extracted data from 26 climate variables
236 (Table S3) from ClimateWNA [50] and used PCA to describe climatic variation among
237 sites. We computed pairwise differences between sites for a) climate based on PC1, b)
238 climate based on PC2 values, c) geographic distance, and d) mt-nuclear ancestry. We
239 then tested for an association between climate differences among sites and differences in
240 mt-nDNA ancestry while controlling for geographic distance using a partial mantel test in
241 R with 10,000 permutations.

242

243 ## Results

244 **Population structure**

245 The mtDNA haplotype clusters are distinct between inland STOW and SOCC,
246 with 0.8% minimum mtDNA sequence divergence between the two clades (Krosby and
247 Rohwer 2009; Figure 1 C). Among these substitutions, there are 5 amino acid changes: 4
248 within NAD2 and 1 within ATP6 (Table S1). Various coastal STOW sampling sites
249 contain a mixture of inland STOW haplotypes and SOCC haplotypes (Figure 1 C),
250 suggesting that these coastal STOW populations are hybrid populations between inland
251 STOW and SOCC (Krosby and Rohwer 2009). Nuclear genomic variation as assessed
252 through variation in the 222,559 SNPs reveals isolated clusters of SOCC and inland
253 STOW, with additional clustering within coastal STOW. The inland STOW and SOCC

254 form two clearly differentiated clusters differing in the first eigenvector (EV1) (Figure 1
255 D), and most individuals from coastal STOW have a somewhat intermediate position.
256 However, we found substructure within coastal STOW with Valdez and Haida Gwaii
257 forming distinct genetic clusters (Figure 1 D). Valdez differs primarily along the second
258 eigenvector (EV2), whereas Haida Gwaii differs by a combination of EV1 and EV2.
259 Moreover, Valdez and Haida Gwaii demonstrate comparable differentiation to the
260 parental populations as the differentiation between the parental populations (SOCC and
261 inland STOW), as measured by genome-wide F_{ST} (Figure S3), possibly a result of drift
262 due to small local population size, especially on Haida Gwaii.

263 Analysis of ancestry and heterozygosity further supports the ancient hybrid status
264 of coastal STOW, as coastal STOW individuals exhibit intermediate ancestry between
265 SOCC and inland STOW but reduced heterozygosity, suggesting many generations of
266 crossing and backcrossing (Figure 2, S2). This pattern is in contrast to what would be
267 expected from recent, early generation hybrids, which would appear at the top of the
268 triangle plot. The c-STOW population falls towards the inland STOW end of the ancestry
269 spectrum, indicating that the ancient mixed population started with more inland STOW
270 than SOCC or that there has been more introgression from inland STOW since the
271 original admixture. Of those divergent SNPs included in the ancestry analysis, we found
272 SOCC-biased introgression in restricted genomic regions on chr1A, chr5, and chrZ
273 (Figure 2 B).

274 **F_{ST} distribution**

275 Genome-wide levels of differentiation show that SOCC and inland STOW are
276 similar (Weir and Cockerham weighted F_{ST} = 0.023) except for a few peaks of

277 differentiation (Figure S1 A). As in the PCA (Figure 1 D), F_{ST} analysis indicates the
278 Valdez and Haida Gwaii coastal STOW populations are more differentiated from both
279 SOCC and inland STOW than other coastal STOW are (see F_{ST} values in Figure S1, S2).
280 The rest of the coastal STOW are more similar to inland STOW (Weir and Cockerham's
281 F_{ST} = 0.008) than to SOCC (Weir and Cockerham's F_{ST} = 0.015) (Figure S3).
282 The inland STOW and SOCC have a number of peaks of differentiation (Figure
283 S1 A) mapping to chromosomes (chr) 1, 1A, 4A, 5, 20, and Z in the *Setophaga coronata*
284 reference. One of these (on chr 20) is in the ASIP-RALY gene block [29], which is known
285 to regulate pigmentation in quail and mice [51,52]. Our earlier study of admixture mapping
286 in the ongoing hybrid zone between inland STOW and SOCC in the Washington
287 Cascades [30] indicated that this locus is highly associated with plumage colour patterns
288 within that zone. As predicted, the present analysis of genomic variation over a much
289 broader geographic region shows high differentiation at the RALY SNP between
290 sampling regions that differ in plumage (i.e., between SOCC and inland STOW, Figure
291 S1 A, F-H) and low differentiation between regions with similar plumage (i.e., between
292 coastal STOW and inland STOW, Figure S1 B-D).
293 Similar to the chr20 RALY peak, the chr5 region also showed extreme
294 differentiation in the comparison of inland STOW and SOCC (Figure S1, 3), but in
295 contrast to the ASIP-RALY region, the chr5 region is also strongly differentiated
296 between coastal STOW and inland STOW and is the most differentiated region between
297 those groups (Figure S1 B-C; 3 B-C). The chr5 peak (Figure 3 A-C) involves a ~1.2Mb
298 region—assayed here with 15 SNPs—and includes 22 genes. These genes are

299 predominantly associated with energy-related signaling transduction (10/22) and fatty
300 acid metabolism (7/22) (Table S2).

301 **Mitonuclear genetic association**

302 We then examined whether any nuclear genomic regions were correlated with
303 mtDNA haplotypes (Figure 1 C), controlling for population substructure. Our GWAS
304 revealed significant ($p_{corrected} < 10^{-6}$) mitonuclear covariation within chr 5 (Figure 4 A).
305 The mtDNA-associated SNP on chr 5 was within the PAPLN gene (Figure 4 B, Table
306 S2), which resides within the island of differentiation between SOCC and inland STOW
307 (Figure 3 A; 4 B), as well as between coastal STOW and inland STOW (Figure 3, 4 B).
308 This ~1.2Mb chr5 differentiation gene block is concentrated with genes with direct
309 functional association with mitochondria (Table S2, energy-related signaling transduction
310 and fatty acid oxidation). We thereafter consider this chr 5 differentiation gene block
311 (Figure 4 C) for subsequent analysis. Among sites, the ancestry of chr 5 differentiation
312 block (Figure 4 D, partial mantel pearson's product-moment $r = 0.9227$, $p < 10^{-4}$) were
313 correlated with the mtDNA ancestry after controlling for spatial autocorrelation (see
314 Methods).

315 **Climatic association**

316 Climate PC1 (Figure 5 AC, S5) explains 64.4% of the variation in climate among
317 sites; this PC demonstrated similar loading among climate variables (Figure S5 AB;
318 Table S3, S4). Climate PC2 explains 23.5% of the variation and was predominantly
319 explained by four climate variables (Figure S5, C, S6; Table S3, S4): Temperature
320 Difference (TD), Climate Moisture Index (CMI), Mean Annual Precipitation (MAP),
321 Winter Precipitation (PPT_wt). The climate in coastal STOW habitat is similar to that of

322 inland STOW along PC1, but more similar to that of SOCC along PC2, although there is
323 great climate variation among various coastal STOW populations (Figure 5 A-C).
324 Overall, coastal STOW habitat is moister and more stable in temperature, which is
325 consistent with the coastal-inland humidity gradient in western North America (captured
326 by PC2, Figure 5 B), and the distribution of mt-nDNA ancestry appears related to this
327 geographical variation in climate. The mt-nDNA ancestry is significantly correlated with
328 climate PC1 (Figure 5 D) (partial mantel test controlling for geographical distance, r =
329 0.307, p = 0.011, Figure 5 D) but not with climate PC2 (r = 0.166, p = 0.081) among 19
330 sites (across SOCC, coastal STOW, and inland STOW habitats).
331

332 **Discussion**

333 *Setophaga occidentalis* (SOCC) and inland *S. townsendi* (inland STOW) are
334 distinct in mtDNA [27] and exhibit three strong regions of differentiation in the nuclear
335 genome [30], whereas coastal *S. townsendi* (coastal STOW) harbor admixed mtDNA and
336 nDNA ancestry from SOCC and inland STOW. One of the regions of strong
337 differentiation between SOCC and inland STOW, a ~1.2Mb region on chr 5,
338 differentiates coastal STOW and inland STOW as well. This region shows strong
339 association with divergent mtDNA haplotypes, after controlling for background
340 population substructure of nDNA, and contains genes with functions that are involved in
341 energy-related signaling transduction and fatty acid metabolism, both of which are strong
342 candidates for cofunction with mtDNA. We found site-level mitonuclear ancestry
343 covaries with the site climate conditions, consistent with adaptation to climate directly, or
344 to other factors that are associated with climatic variation, such as forest characteristics.

345 **Coevolution of mtDNA and nDNA**

346 We found the key nDNA differences between coastal STOW versus inland
347 STOW resides at loci within chr 5 associated with fatty acid metabolism and energy-
348 related signaling transduction (Figure 4 B; Table S2), an intriguing result given that
349 coastal STOW and inland STOW differ so strongly in their mitochondrial haplotype
350 frequencies. This chr5 gene block is concentrated with genes encoding proteins directly
351 involved ATP and NADH activities. Notably, an analysis of the amino differences
352 between whole mitochondrial genomes in this system revealed four amino acid
353 substitutions in the *ND2* gene and one in *ATP6* (Table S1). In addition, the chr5 gene
354 block harbors genes associated with fatty acid oxidation (Table S2), which mainly occurs
355 in mitochondria although it occurs in other cellular compartments as well (see reviews
356 [53,54]). Together, these findings point to the possibility of selection on mitonuclear
357 cofunctions [11,12]. The SOCC nDNA may thus be partially incompatible with the inland
358 STOW mtDNA in multiple functional roles, leading to selection against mismatched
359 mitonuclear ancestries. Such selection maintaining mitonuclear concordance may be
360 counteracted by random mating in admixed populations at each generation and is thus
361 difficult to detect in samples of individuals from a single population. However,
362 mitonuclear ancestry concordance can be more easily detected through comparison of
363 many populations. This association (Figure 4) reveals the potential selection maintaining
364 a functionally compatible mitonuclear fatty acid metabolism over a large temporal and
365 spatial scale.

366 These mitonuclear genotypes are significantly associated with climate variation
367 among sites, suggesting potential selection on the mt-nDNA combinations related to

368 climate or habitat (which is also associated with climate). The climate in the coastal
369 STOW is similar to inland STOW habitat along PC1, but similar to SOCC habitat along
370 PC2. Correlations between any two traits that have large-scale geographic variation are
371 expected, making it difficult to confidently infer causality from such associations alone.
372 However, the climate and habitat differences between SOCC, inland STOW, and coastal
373 STOW are very strong, such that these differences likely cause some selective
374 differences. These patterns are reminiscent of the *Eopsaltria australis* (Eastern Yellow
375 Robin) system in which distinct mt-nDNA combinations are maintained between inland
376 and coastal habitat [6]. Fatty acid metabolic genes have also been shown to be targets of
377 climatic adaptation in humans, within Siberian [55] and Greenlandic Inuit populations [56].
378 Temperature [57] and humidity [58] both influence mitochondrial fatty acid metabolism [57,58].
379 Chr5-mtDNA genotypes might result in functional difference in fatty acid metabolism
380 that is adapted to specific climate (moist and stable versus dry and variable) in the
381 breeding habitat of these warblers.
382 Because SOCC has apparently inhabited coastal areas for a long period of time,
383 the SOCC mt-nDNA gene combination may be more suited for coastal habitats compared
384 to those of inland STOW. If the SOCC mt-nDNA genotype is favored in the coastal
385 habitats, the frequency of SOCC mt-nDNA gene combinations would tend to increase in
386 coastal STOW populations over time. However, ongoing gene flow between inland
387 STOW and coastal STOW would slow down or prevent such increase. The Haida Gwaii
388 island and Valdez population could have escaped from such a balance between selection
389 and gene flow due to their isolation from the rest of the populations respectively by the
390 sea and mountain ranges. Another possibility is that frequency-dependent selection is

391 maintaining long-term mt-nDNA polymorphism in the coastal STOW. Future
392 investigation on the spatial and temporal variation of the strength of association between
393 mtDNA and the chr5 gene block would shed light on the evolutionary forces shaping the
394 present and future of the coastal STOW population. Further examining such mitonuclear
395 relationship in the three hybrid zones between SOCC and inland STOW in Washington
396 and Oregon[59] would shed light on the prevalence of such evolutionary forces.

397 **Genomic architecture of differentiation**

398 In comparisons among these populations, the distribution of F_{ST} across the
399 genome is consistent with the "genic" view of differentiation [30,60,61], in which peaks of
400 differentiation represent genetic targets of selection (divergent selection or selection
401 against hybrids) that are highly distinct between populations despite the rest of the
402 genome being homogenized by gene flow [60–62]. Despite this 'selection with gene flow'
403 scenario exhibiting a similar genomic differentiation landscape as the classic 'divergence
404 with gene flow' model [61,63], the underlying process is different. In this system, there is a
405 known allopatric phase when SOCC and inland STOW were separated by ice sheets
406 (Figure 1 A) [24,27]. Genetic differentiation that accumulated in allopatry (as opposed to
407 gradual build up at sympatry or parapatry under 'divergence with gene flow') can be
408 homogenized by hybridization at secondary contact, while the climate/habitat-related
409 genomic target (on chr 5) of selection remains differentiated.

410 Between coastal STOW and SOCC, there are a number of highly differentiated
411 loci (Figure S1 F), one of which is the ASIP-RALY gene block that was found through
412 admixture mapping to be highly associated with plumage in the narrow Cascades hybrid
413 zone between inland STOW and SOCC [30]. The fact that this marker has now been shown

414 to be strongly associated with plumage both within a local hybrid zone and throughout
415 the whole inland STOW and SOCC species complex is strong evidence for a causal link
416 between the RALY-ASIP genetic region and plumage differences.

417 **Biogeography and semi-parallel introgression**

418 In addition to providing an empirical examination of mitonuclear coevolution, the
419 present study also helps clarify the biogeographic history of this warbler complex. Our
420 genomic evidence is consistent with Krosby and Rohwer's [27] conclusion, based on
421 mtDNA, that coastal British Columbia and Alaska was inhabited by geographically
422 structured SOCC populations before inland STOW expanded from inland areas and
423 mixed with them [27]. The SOCC and inland STOW mtDNA haplotype groups demonstrate
424 many differences (~0.8%), whereas both are common in coastal STOW. It is unlikely that
425 the polymorphisms in mtDNA and nDNA in the coastal STOW were caused by
426 incomplete lineage sorting, as opposed to hybridization (Figure 1 A). In a scenario of
427 incomplete lineage sorting, enough time would have passed following population
428 splitting for both inland STOW and SOCC to have lost the alternative haplotype, while
429 the coastal STOW maintained both. Over such a period of time, sizeable differences
430 would be expected between the SOCC haplotypes found in SOCC population versus
431 coastal STOW, as well as between inland STOW haplotypes found in coastal versus
432 inland STOW population. We did not observe such a pattern, as coastal STOW has some
433 mtDNA haplotypes that are identical to SOCC and some that are identical to inland
434 STOW haplotypes. In addition, the reduced heterozygosity (relative to expected values of
435 recent hybrids) and intermediate ancestry (Figure 2 S2) reflected in the nuclear genome
436 of coastal STOW further supported the ancient hybrid status of coastal STOW.

437 The higher genome-wide differentiation of the Haida Gwaii and Valdez
438 populations (Figure S3) is consistent with at least partially isolated cryptic refugia of
439 SOCC in coastal Alaska and Haida Gwaii during the last glacial maximum [64]. Following
440 expansion of inland STOW from the inland area, presumably after the last glacial period,
441 hybridization between inland STOW and SOCC apparently led to populations of mixed
442 ancestry along the coast of British Columbia and Alaska (Figure 1 A). These coastal
443 populations have the plumage patterns and colors of inland STOW, which is why they
444 have been classified as members of that species. This uniform inland STOW appearance
445 has concealed a more complex history of hybridization with ancient and geographically
446 differentiated populations of SOCC.
447 Following expansion of inland STOW from the interior, gene flow into Haida
448 Gwaii may have been weak due to the expanse of water separating it from the mainland,
449 explaining why that population is more similar to SOCC than other coastal STOW are.
450 Gene flow into Valdez could have also been impeded by geographical barriers, as Valdez
451 is surrounded by mountain ranges (Chugach mountains, Wrangell mountains, and St.
452 Elias mountains). However, both nuclear and genomic data indicate that Valdez has
453 substantial ancestry from both inland STOW and SOCC. Despite genome-wide
454 differentiation among these three coastal STOW genetic clusters, there is an interesting
455 parallelism: all the three populations exhibit the inland STOW-like ASIP-RALY marker
456 (Figure S1) that is associated with plumage [30], and disproportionately higher frequencies
457 of the SOCC-like chr5 nDNA gene block that covaries with mtDNA (Figure 3). Such
458 parallelism might be driven by parallel adaptation to the coastal climate.
459 **Caveats and future directions**

460 While our findings are consistent with mitonuclear coadaptation being important
461 in the pattern of genomic differentiation within this species complex, this being an
462 observational study we cannot definitively conclude that is the case. The strong
463 associations between geographic variation in climate, mitochondrial haplotypes, and
464 highly differentiated regions of the genome, along with the known roles of those
465 divergent regions in mitochondrial-associated function and the abundant evidence for
466 mitonuclear coadaptation in other systems [7,11,65], add up to strong correlative evidence for
467 mitonuclear adaptation in this case. One possibility is that the nuclear and mitochondrial
468 loci are independently selected by the environment, without actual coevolution between
469 the two. Future experimental study should investigate this possibility to distinguish it
470 from actual coevolution. If there is mitonuclear coadaptation, there should be (1) an
471 epistatic effect of mtDNA and nDNA on fatty acid metabolic phenotypes; (2) the high
472 frequency fatty acid metabolic phenotypes (underpinned by mitonuclear epistasis) within
473 each site should be more fit for local climate than foreign climate.
474
475 **Conclusion**
476 Examination of genomic differentiation in this young species group has revealed
477 patterns consistent with climate-related coadaptation among mtDNA and nDNA involved
478 in energy-related signaling transduction and fatty-acid metabolism. Consistent with the
479 mtDNA pattern, the coastal STOW demonstrate both mtDNA and nuclear genomic
480 patterns consistent with ancient admixture between inland STOW and a geographically
481 structured ancient SOCC population. Three genetic clusters of coastal STOW are
482 characterized by a mixed genetic ancestry between the parental populations (SOCC and

483 inland STOW), providing natural replicates for examining the role of selection in shaping
484 genomic differentiation. These three coastal STOW clusters exhibit differentiation from
485 inland STOW at one of the most differentiated genomic regions (on chr 5) between
486 inland STOW and SOCC. The geographic distributions of the mitonuclear genetic
487 combinations related to fatty acid metabolism are associated with geographic variation in
488 climate, suggesting mt-nDNA coevolution may have occurred in response to selection for
489 climate adaptation. Such climate-related mitonuclear selection could be an important
490 force driving population differentiation in this species complex.

491 **Data Accessibility**

492 Sequence data is deposited in GenBank SRA (accession number: PRJNA573930;
493 ID: 573930). Secondary analytical data tables have been deposited in dryad
494 (https://doi.org/10.5061/dryad.44j0zpc9t).

495 **Acknowledgements**

496 In support of the Bird Names for Birds movement, we omitted the existing
497 common name of *Setophaga townsendi*. We are grateful to Sharon Birks (Burke
498 Museum) and Chris Wood (Burke Museum) for access to the tissue samples for
499 sequencing. We thank Geoffrey E. Hill for inspiring ideas to this study. We also thank
500 Gil Henriques for providing digital illustrations of the warblers. For helpful discussion
501 we thank Sally Otto, Dolph Schluter, Loren Rieseberg, Graham Coop, Dahong Chen,
502 Mike Whitlock, Andrea Thomaz, Armando Geraldes, Meade Krosby, Hernan Morales,
503 Xinzhu Wei, Jared Grummer, and Rohit Kolora. We are grateful for research funding
504 provided by the Natural Sciences and Engineering Research Council of Canada (grants
505 311931-2012, RGPIN-2017-03919 and RGPAS-2017-507830 to DEI; and PGS D

331015731 to SW); Pennsylvania State University, the Eberly College of Science, and the Huck Institutes of the Life Sciences to DPLT; and a Werner and Hildegard Hesse Research Award in Ornithology and a UBC Four Year Doctoral Fellowship to SW. For research permits we thank Environment Canada; U. S. Geological Survey; Departments of Fish & Wildlife of Washington, Idaho, and Montana; and the UBC Animal Care Committee.

References

1. Calvo, S. E. & Mootha, V. K. The Mitochondrial Proteome and Human Disease. *Annu. Rev. Genomics Hum. Genet.* **11**, 25–44 (2010).

2. Lane, N. Mitonuclear match: Optimizing fitness and fertility over generations drives ageing within generations. *BioEssays* **33**, 860–869 (2011).

3. Bar-Yaacov, D. *et al.* Mitochondrial involvement in vertebrate speciation? The Case of mitonuclear genetic divergence in chameleons. *Genome Biol. Evol.* **7**, 3322–3336 (2015).

4. Hill, G. E. *Mitonuclear Ecology*. (Oxford University Press, 2019).

5. Ballard, J. W. O. & Whitlock, M. C. The incomplete natural history of mitochondria. *Mol. Ecol.* **13**, 729–744 (2004).

6. Morales, H. E. *et al.* Concordant divergence of mitogenomes and a mitonuclear gene cluster in bird lineages inhabiting different climates. *Nat. Ecol. Evol.* **2**, 1258–1267 (2018).

7. Hill, G. E. *et al.* Assessing the fitness consequences of mitonuclear interactions in natural populations. *Biol. Rev.* **94**, consequences of mitonuclear interactions in natu (2019).

8. Barreto, F. S. & Burton, R. S. Elevated oxidative damage is correlated with reduced fitness in interpopulation hybrids of a marine copepod. *Proc. R. Soc. B Biol. Sci.* **280**, 20131521 (2013).

9. Healy, T. M. & Burton, R. S. Strong selective effects of mitochondrial DNA on the nuclear genome. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 6616–6621 (2020).

535 10. Burton, R. S., Pereira, R. J. & Barreto, F. S. Cytonuclear Genomic Interactions and
536 Hybrid Breakdown. *Annu. Rev. Ecol. Evol. Syst.* **44**, 281–302 (2013).

537 11. Hill, G. E. The mitonuclear compatibility species concept. *Auk* **134**, 393–409
538 (2017).

539 12. Burton, R. S. & Barreto, F. S. A disproportionate role for mtDNA in Dobzhansky-
540 Muller incompatibilities? *Molecular Ecology* **21**, 4942–4957 (2012).

541 13. Innocenti, P., Morrow, E. H. & Dowling, D. K. Experimental evidence supports a
542 sex-specific selective sieve in mitochondrial genome evolution. *Science (80-.).*
543 **332**, 845–848 (2011).

544 14. Harada, A. E., Healy, T. M. & Burton, R. S. Variation in thermal tolerance and its
545 relationship to mitochondrial function across populations of Tigriopus
546 Californicus. *Front. Physiol.* (2019). doi:10.3389/fphys.2019.00213

547 15. Acevedo, P. *et al.* Range dynamics driven by Quaternary climate oscillations
548 explain the distribution of introgressed mtDNA of Lepus timidus origin in hares
549 from the Iberian Peninsula. *J. Biogeogr.* **42**, 1727–1735 (2015).

550 16. Barton, N. H. & Hewitt, G. M. Adaptation, speciation and hybrid zones. *Nature*
551 **341**, 497–503 (1989).

552 17. Elgvin, T. O. *et al.* The genomic mosaicism of hybrid speciation. *Sci. Adv.* **3**,
553 e1602996 (2017).

554 18. Schumer, M., Cui, R., Powell, D. L., Rosenthal, G. G. & Andolfatto, P. Ancient
555 hybridization and genomic stabilization in a swordtail fish. *Mol. Ecol.* **25**, 2661–
556 2679 (2016).

557 19. Rieseberg, L. H. Hybrid origins of plant species. *Annual Review of Ecology and*
558 *Systematics* (1997). doi:10.1146/annurev.ecolsys.28.1.359

559 20. Gagnaire, P. A., Normandeau, E. & Bernatchez, L. Comparative genomics reveals
560 adaptive protein evolution and a possible cytonuclear incompatibility between
561 European and American Eels. *Mol. Biol. Evol.* **29**, 2909–2919 (2012).

562 21. Sambatti, J. B. M., Ortiz-Barrientos, D., Baack, E. J. & Rieseberg, L. H.
563 Ecological selection maintains cytonuclear incompatibilities in hybridizing
564 sunflowers. *Ecol. Lett.* **11**, 1082–1091 (2008).

565 22. Baris, T. Z. *et al.* Evolved genetic and phenotypic differences due to
566 mitochondrial-nuclear interactions. *PLoS Genet.* **13**, e1006517 (2017).

567 23. Boratyński, Z., Ketola, T., Koskela, E. & Mappes, T. The Sex Specific Genetic
568 Variation of Energetics in Bank Voles, Consequences of Introgression? *Evol. Biol.*
569 **43**, 37–47 (2016).

570 24. Weir, J. T. & Schluter, D. Ice sheets promote speciation in boreal birds. *Proc. R.*
571 *Soc. B Biol. Sci.* **271**, 1881–1887 (2004).

572 25. Rohwer, S. & Wood, C. Three Hybrid Zones Between Hermit and Townsend ' S
573 Warblers in Washington and Oregon. *Auk* **115**, 284–310 (1998).

574 26. Rohwer, S., Bermingham, E. & Wood, C. Plumage and mitochondrial DNA
575 haplotype variation across a moving hybrid zone. *Evolution (N. Y).* **55**, 405–422
576 (2001).

577 27. Krosby, M. & Rohwer, S. A 2000 km genetic wake yields evidence for northern
578 glacial refugia and hybrid zone movement in a pair of songbirds. *Proc. R. Soc. B*
579 *Biol. Sci.* **276**, 615–621 (2009).

580 28. Krosby, M. & Rohwer, S. Ongoing movement of the hermit warbler X
581 Townsend's Warbler Hybrid Zone. *PLoS One* **5**, e14164 (2010).

582 29. Wang, S. *et al.* Selection on a small genomic region underpins differentiation in
583 multiple color traits between two warbler species. *Evol. Lett.* **4**–**6**, 502–515 (2020).

584 30. Wang, S. *et al.* Selection on a pleiotropic color gene block underpins early
585 differentiation between two warbler species. *bioRxiv*
586 https://doi.org/10.1101/853390 (2019).

587 31. Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring
588 intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).

589 32. Leigh, J. W. & Bryant, D. POPART: Full-feature software for haplotype network
590 construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).

591 33. Elshire, R. J. *et al.* A robust, simple genotyping-by-sequencing (GBS) approach
592 for high diversity species. *PLoS One* **6**, e19379 (2011).

593 34. Wang, S., Rohwer, S., Delmore, K. E. & Irwin, D. E. Cross-decades stability of an
594 avian hybrid zone. *J. Evol. Biol.* **32**, 1242–1251. (2019).

595 35. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for
596 Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

597 36. Baiz, M. D., Wood, A. W., Brelsford, A., Lovette, I. J. & Toews, D. P. L.
598 Pigmentation Genes Show Evidence of Repeated Divergence and Multiple Bouts
599 of Introgression in Setophaga Warblers. *Curr. Biol.* **31**, 1–7 (2021).

600 37. Li, H. Aligning new-sequencing reads by BWA BWA : Burrows-Wheeler Aligner.
601 *Slides* (2010). doi:10.1002/pssa.200673542

602 38. McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian
603 Cibulskis, Andrew Kernytsky, Kiran Garimella, David Altshuler, Stacey Gabriel,

604 M. D. *et al.* The Genome Analysis Toolkit: A MapReduce framework for
605 analyzing next-generation DNA sequencing data. *Genome Res.* (2010).
606 doi:10.1101/gr.107524.110.20

607 39. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–
608 2158 (2011).

609 40. Zheng, X. *et al.* A high-performance computing toolset for relatedness and
610 principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).

611 41. R Core Team (2017). R: A language and environment for statistical computing. *R*
612 *Found. Stat. Comput. Vienna, Austria.* R Foundation for Statistical Computing
613 (2017). doi:/S0103-64402004000300015

614 42. Raj, A., Stephens, M. & Pritchard, J. K. fastSTRUCTURE: Variational Inference
615 of Population Structure in Large SNP Datasets. *Genetics* **197**, 573–589 (2014).

616 43. Weir, B. S. & Cockerham, C. C. Estimating F-Statistics for the Analysis of
617 Population Structure. *Evolution (N. Y).* **38**, 1358 (1984).

618 44. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: An R
619 library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296
620 (2007).

621 45. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**,
622 W5–W9 (2008).

623 46. Bateman, A. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.*
624 **47**, D506-515 (2019).

625 47. Legendre, P. *Numerical ecology, 2nd English Edition*. *Numerical ecology,*
626 (Elsevier Science, 1998). doi:10.1017/CBO9781107415324.004

627 48. Bernt, M. *et al.* MITOS: Improved de novo metazoan mitochondrial genome
628 annotation. *Mol. Phylogenet. Evol.* **69**, 313–319 (2013).

629 49. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler
630 transform. *Bioinformatics* **26**, 589–595 (2010).

631 50. Wang, T., Hamann, A., Spittlehouse, D. L. & Murdock, T. Q. ClimateWNA-high-
632 resolution spatial climate data for western North America. *J. Appl. Meteorol.*
633 *Climatol.* **51**, 16–29 (2012).

634 51. Michaud, E. J. *et al.* A molecular model for the genetic and phenotypic
635 characteristics of the mouse lethal yellow (Ay) mutation. *Proc. Natl. Acad. Sci.* **91**,
636 2562–2566 (1994).

637 52. Nadeau, N. J. *et al.* Characterization of Japanese quail yellow as a genomic

638 deletion upstream of the avian homolog of the mammalian ASIP (agouti) gene.
639 *Genetics* **178**, 777–786 (2008).

640 53. Console, L. *et al.* The Link Between the Mitochondrial Fatty Acid Oxidation
641 Derangement and Kidney Injury. *Front. Physiol.* **11**, 1–7 (2020).

642 54. Houten, S. M. & Wanders, R. J. A. A general introduction to the biochemistry of
643 mitochondrial fatty acid β-oxidation. *J. Inherit. Metab. Dis.* **33**, 469–477 (2010).

644 55. Clemente, F. J. *et al.* A Selective Sweep on a Deleterious Mutation in CPT1A in
645 Arctic Populations. *Am. J. Hum. Genet.* **95**, 584–589 (2014).

646 56. Fumagalli, M. *et al.* Greenlandic Inuit show genetic signatures of diet and climate
647 adaptation. *Science (80-.).* **349**, 1343–1347 (2015).

648 57. Zoladz, J. A. *et al.* Effect of temperature on fatty acid metabolism in skeletal
649 muscle mitochondria of untrained and endurance-trained rats. *PLoS One* **12**,
650 e0189456 (2017).

651 58. Atkin, O. K. & Macherel, D. The crucial role of plant mitochondria in
652 orchestrating drought tolerance. *Ann. Bot.* **103**, 581–597 (2009).

653 59. Townsend, S., Rohwer, S. & Wood, C. HYBRID ZONES BETWEEN HERMIT
654 AND OREGON trivialized the study of hybrid zones during much of the last half-
655 century . Be- waning before this theoretical literature oped (see reviews in Mayr
656 and Short 1970 , Ris- balancing selection against hybrids with dis. **115**, 284–310
657 (1998).

658 60. Wu, C. I. The genic view of the process of speciation. *Journal of Evolutionary*
659 *Biology* **14**, 851–865 (2001).

660 61. Via, S. Natural selection in action during speciation. *Proc. Natl. Acad. Sci.* **106**,
661 9939–9946 (2009).

662 62. Nosil, P. A Genetic Mechanism to Link Selection to Reproductive Isolation. in
663 *Ecological Speciation* 109–138 (Oxford Univ. Press, 2012).

664 63. Feder, J. L., Flaxman, S. M., Egan, S. P., Comeault, A. A. & Nosil, P. Geographic
665 Mode of Speciation and Genomic Divergence. *Annu. Rev. Ecol. Evol. Syst.* **44**, 73–
666 97 (2013).

667 64. Shafer, A. B. A., Cullingham, C. I., Côté, S. D. & Coltman, D. W. Of glaciers and
668 refugia: A decade of study sheds new light on the phylogeography of northwestern
669 North America. *Molecular Ecology* **19**, 4589–4621 (2010).

670 65. Toews, D. P. L., Mandic, M., Richards, J. G. & Irwin, D. E. Migration,
671 mitochondria, and the yellow-rumped warbler. *Evolution (N. Y).* **68**, 241–255
672 (2014).

Figures

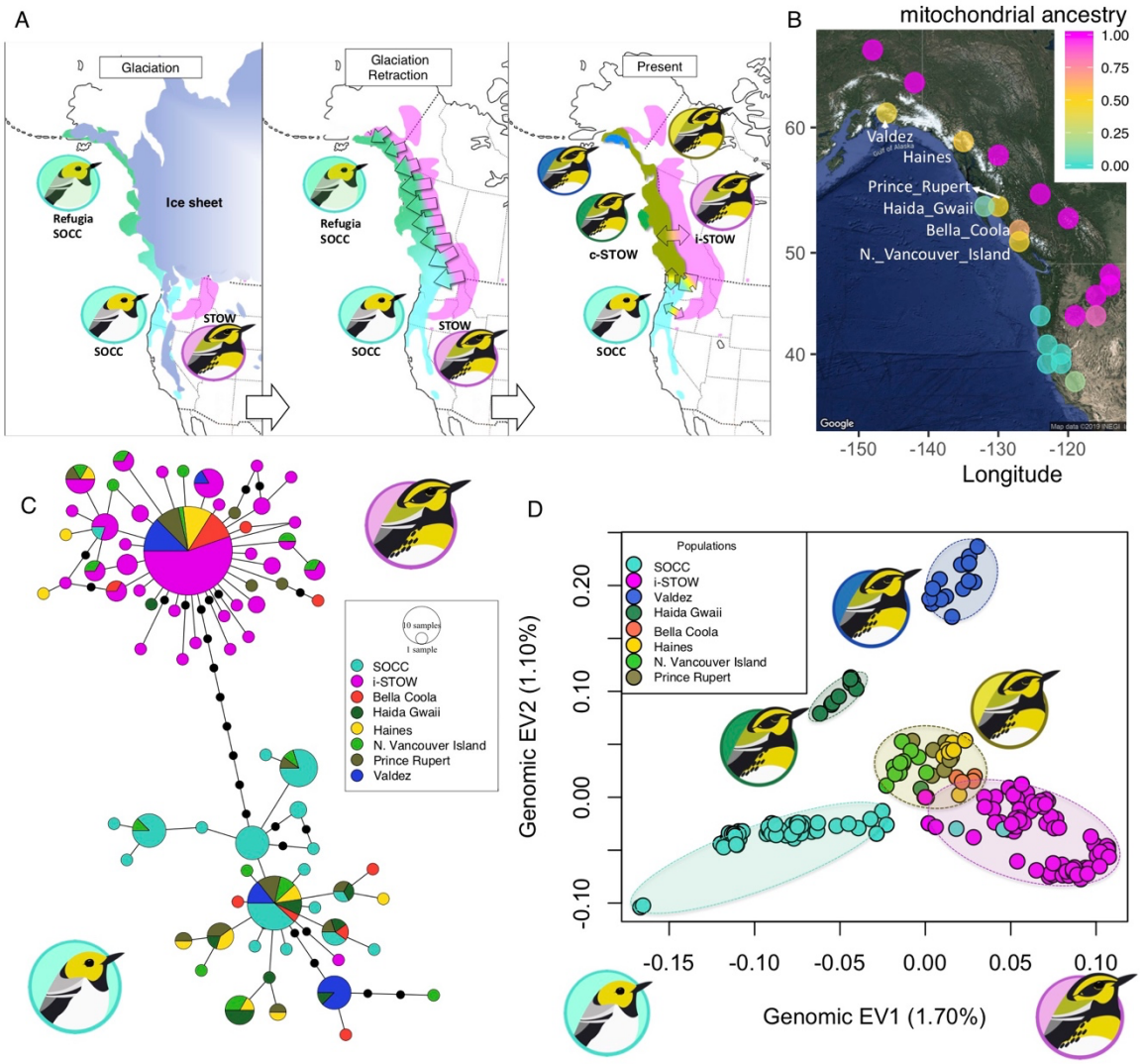

Figure 1 Coastal populations of *S. townsendi* (coastal STOW) demonstrate a mixture of mitochondrial and nuclear ancestry from *S. occidentalis* (SOCC) and inland *S. townsendi* (inland STOW). (**A**) Illustration of the inferred history of differentiation and hybridization between STOW and SOCC during glacial expansion and retraction. Left: During the last glacial maxima, the SOCC and STOW populations resided in isolated glacial refugia. Center: after glacial retraction, the refugial SOCC and inland STOW expanded and hybridized along a broad inland-to-coastal front parallel to the coast. Right: the historical hybridization resulted in coastal STOW populations with admixed ancestry although the plumage resembles that of inland STOW. Population substructure within coastal STOW could be a result of refugial isolation. **(B)** Mitochondrial ancestry of major sampling sites, with colour representing the proportion of individuals belonging to the two major clades shown in: (**C**) Haplotype network of mitochondrial NADH gene sequences from ref. [27], with colours representing major sampling areas. Each circle represents a haplotype and area of the circles are proportional to the number of individuals carrying each haplotype. The lines (regardless of their lengths) between the circles represent one mutation between haplotypes, the black dots on the lines represent

additional mutations among haplotypes. SOCC and inland STOW almost universally belong to divergent mtDNA clades, whereas coastal STOW populations harbor a mixture of these two major clades. (**D**) Principle component analysis of covariance with 222,559 high quality SNPs in the nuclear genome. Populations of coastal STOW are intermediate in Eigenvector (EV) 1 but distinct from inland STOW and SOCC in EV2. EV1 represents admixture between inland STOW and SOCC. Three distinct coastal STOW clusters are evident: Valdez (royal blue), Haida Gwaii (dark green), and other coastal STOW (thereafter abbreviated as ocoastal STOW, gold).

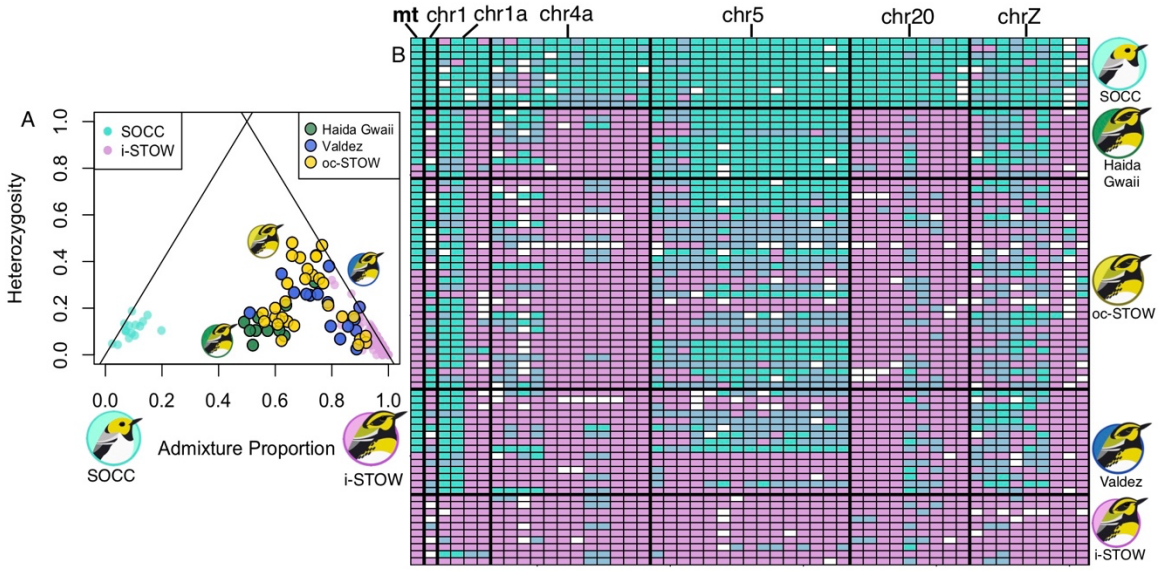

Figure 2 Admixture signature of coastal STOW populations. **(A)** Triangle plot showing relationship between individual admixture proportion vs. heterozygosity based on ancestry-informative SNPs (F_{ST} > 0.6) between SOCC and inland STOW, showing that coastal STOW populations (green: Haida Gwaii, blue: Valdez; gold: oc-STOW) are ancient admixtures between SOCC and inland STOW-like population. **(B)** ancestry score (homozygotes SOCC and inland STOW respectively in turquoise and magenta, heterozygotes in light blue) of mtDNA and 50 ancestry-informative nDNA SNPs (F_{ST} > 0.6) in chr 1, 1a, 4a, 5, 20, and Z of randomly sampled 10 individuals from parental populations (SOCC, inland STOW), and different coastal STOW populations.

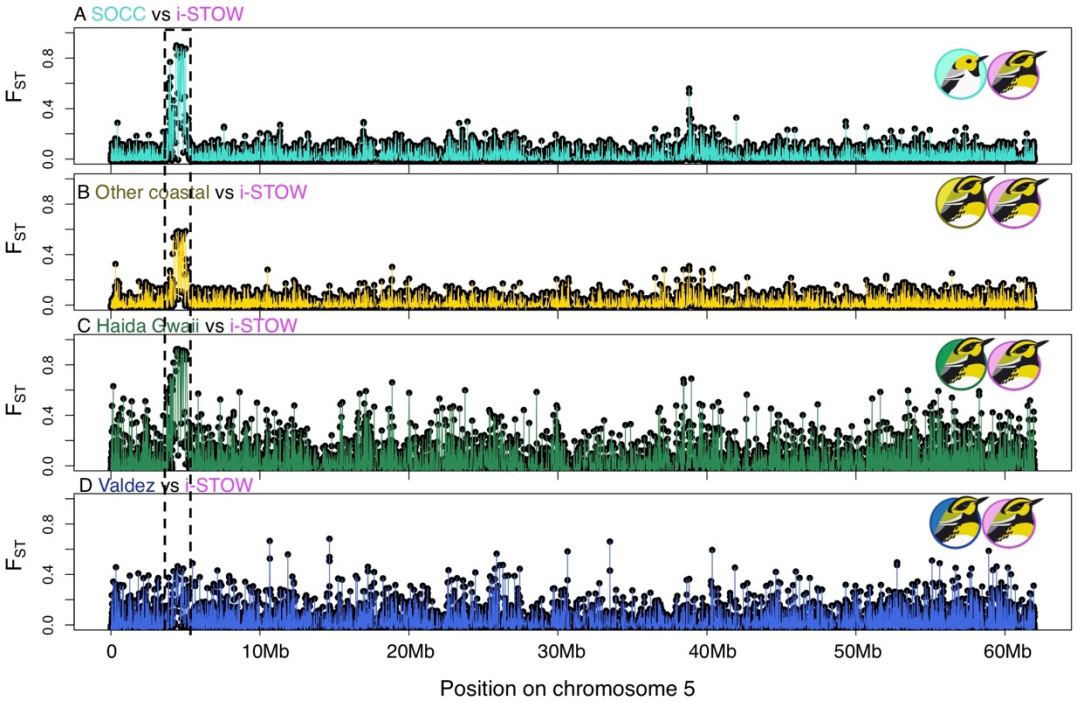

Figure 3 Genetic differentiation (F_{ST}) of variable sites along chromosome 5, between four pairs of populations.

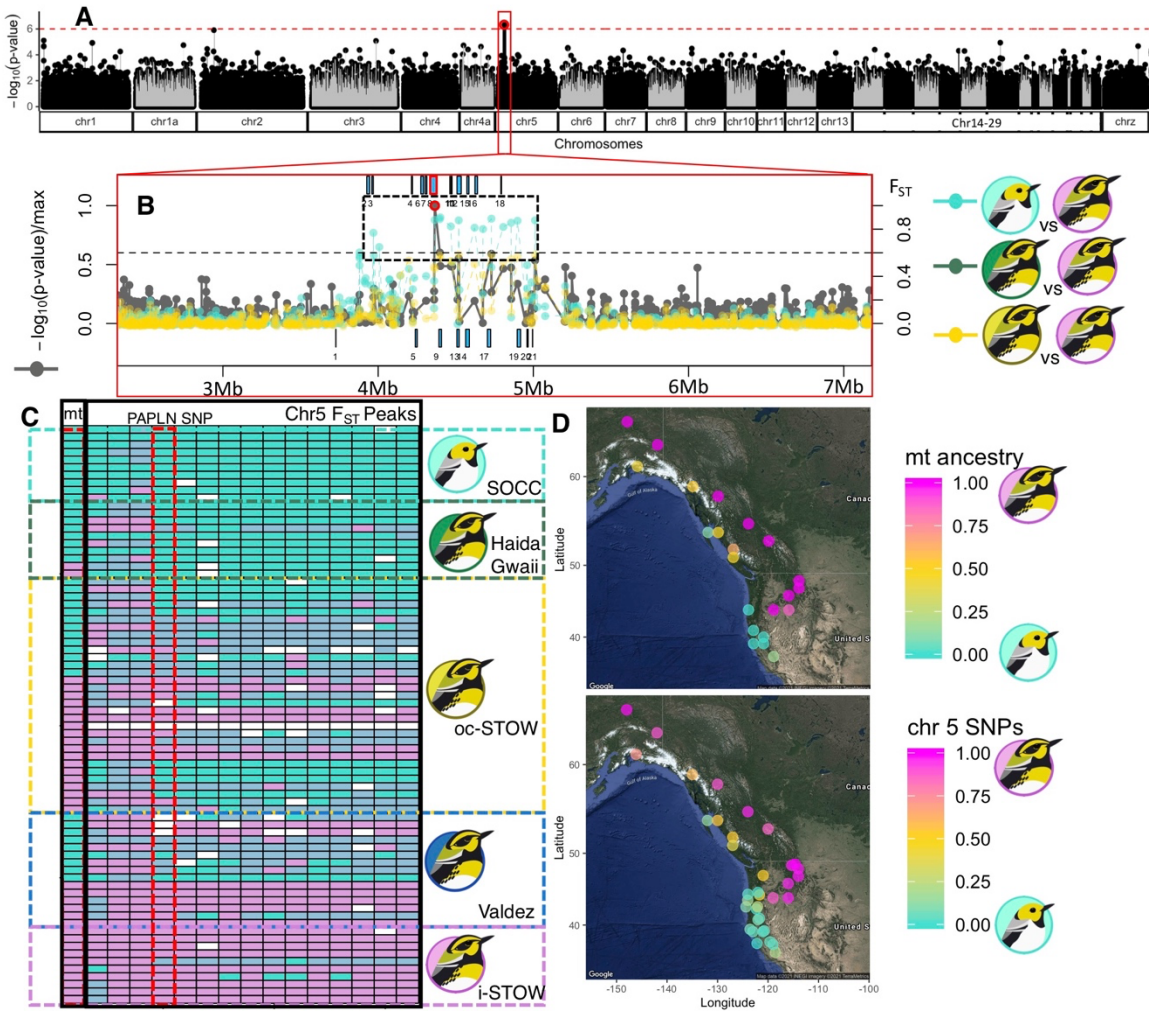

Figure 4 Mitonuclear association across SOCC, coastal STOW, and inland STOW populations. (**A-C**) mitonuclear association coincides with F_{ST} peaks between SOCC and various coastal STOW versus inland STOW. (**A**) GWAS scan of mitonuclear association (controlling for population substructure) across the genome revealed significant associations ($p_{corrected} < 10^{-6}$ indicated by red circles) within chromosome 5. (**B**) The mitonuclear association (red circle) within chromosome 5 coincides with F_{ST} peaks ($F_{ST} >$ 0.6) between SOCC and inland STOW (turquoise), between Haida Gwaii STOW and inland STOW (forest green), as well as between other coastal STOW and inland STOW (yellow). Blue boxes correspond to the positions of genes (top: forward strand; bottom: reverse strand) corresponding to the numbered IDs in Table S2. The mitonuclear association peak is within Papilin, proteoglycan like sulfated glycoprotein (PAPLN) highlighted with red border. (**C**) Ancestry genotypes (turquoise = SOCC; light blue = heterozygous; magenta = inland STOW) of mtDNA and chr5 SNPs correspond to strong genotypic differences between populations. Ten randomly sampled individuals from each parental population were shown along with the individuals from various coastal STOW. The PAPLN SNP haplotypes is red bordered. (**D**) There is significant correlation between mitochondrial (top) and chr 5 (bottom) genetic block (shown in C) after controlling for spatial autocorrelation across sites (Partial Mantel Test, $p < 10^{-4}$).

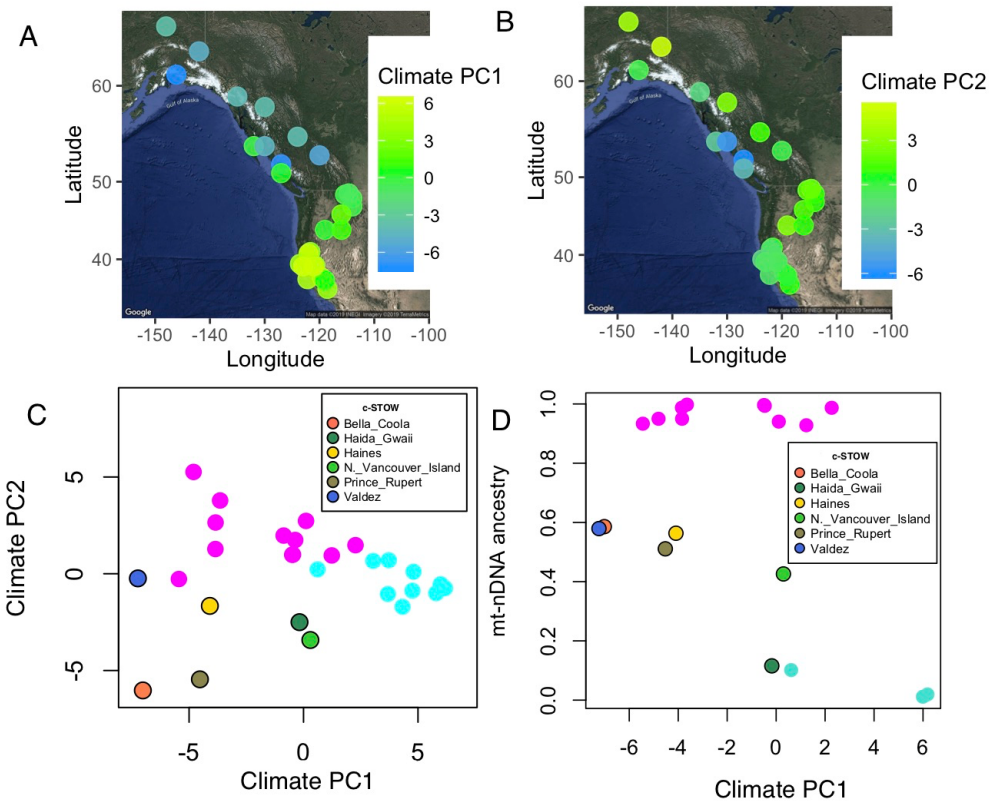

Figure 5 Climate principal component analysis of 26 climate variables from ClimateWNA is associated with mt-nDNA ancestry. (**A**)-(**B**), spatial variation in climate PC1 (**A**) and PC2 (**B**). (**C**) Climate PC1 and PC2, in which SOCC (turquoise), inland STOW (magenta) and coastal STOW habitats are different. (**D**) Site mean mtDNA and chr 5 gene block ancestry is correlated with local climate PC1 (partial mantel test, r = 0.307, p = 0.011).

Supplementary Figures

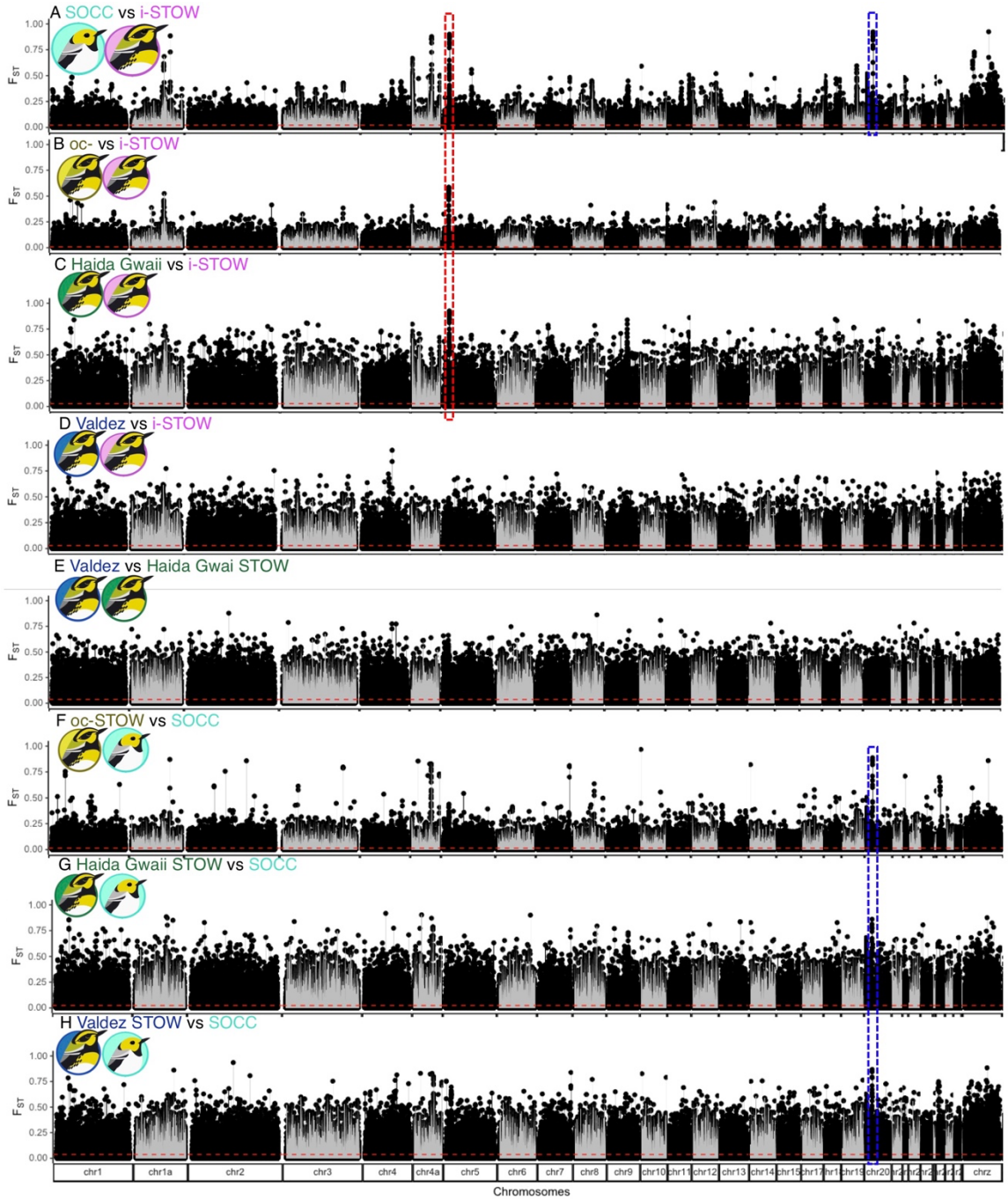

Figure S1 Genetic differentiation (F_{ST}) across the genome, between pairs of *S. occidentalis* (SOCC) and/or *S. townsendi* (STOW) populations. A distinctive differentiation peak was found between coastal STOW (Haida Gwaii, Valdez, and ocoastal STOW) and inland STOW that reside in chromosome 5 (red boxes, **A-C**). The ASIP-RALY peak demonstrates consistent differentiation between SOCC and various STOW (blue boxes, **F-H**). Red horizontal dashed lines reflect Weir & Cockerham weighted average F_{ST}.

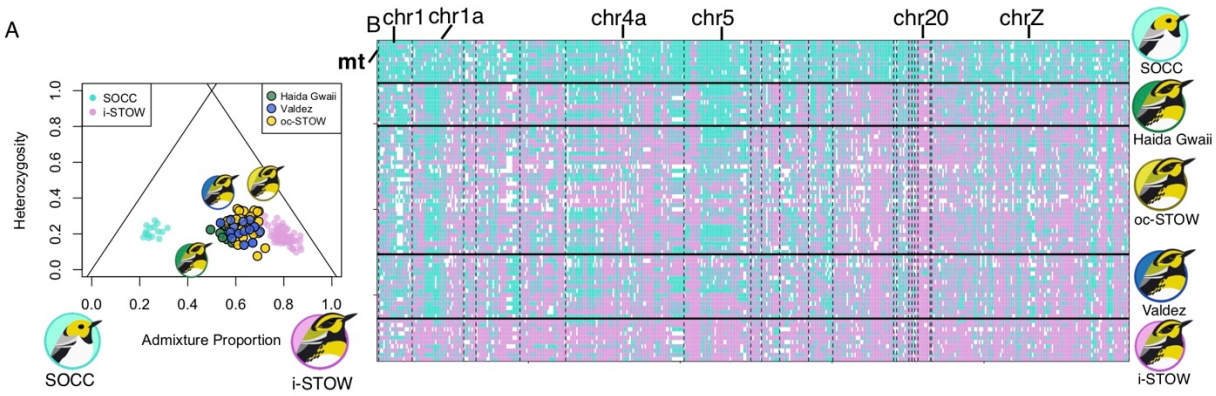

Figure S2 Admixture signature of coastal STOW populations. **(A)** Triangle plot showing relationship between individual admixture proportion vs. heterozygosity based on 618 SNPs (F_{ST}>0.3) between SOCC and inland STOW, showing that coastal STOW populations consistent with ancient admixtures between SOCC and inland STOW-like population. **(B)** ancestry score (homozygotes SOCC and inland STOW respectively in turquoise and magenta, heterozygotes in light blue) of mtDNA and 618 nDNA SNPs (with F_{ST} > 0.3) in randomly sampled 10 individuals from parental populations (SOCC, inland STOW), and different coastal STOW populations.

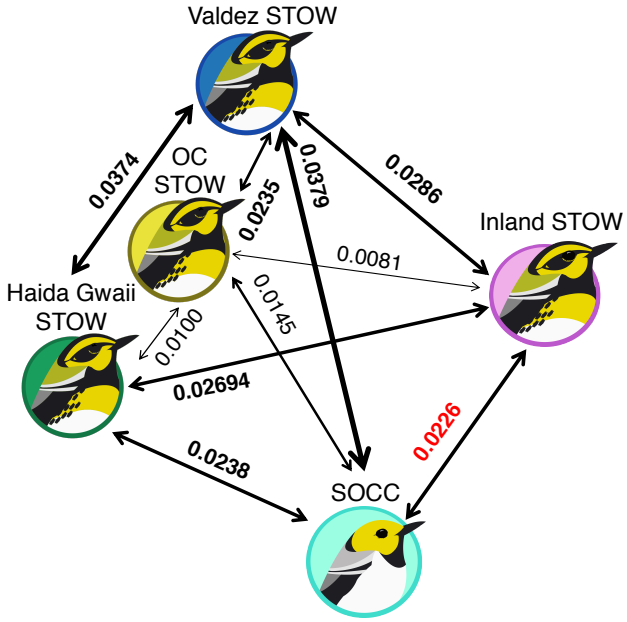

Figure S3 The Weir & Cockerham weighted F_{ST} among SOCC, inland and coastal STOW. Each double-head arrow represents a pairwise comparison among the populations. The populations are oriented according to their relative geographical locations. The widths of the arrows are weighted by the F_{ST} between each pair of populations. Surprisingly some coastal STOW populations demonstrate greater

differentiation from the parental populations than between the parental populations (F_{ST} = 0.0226).

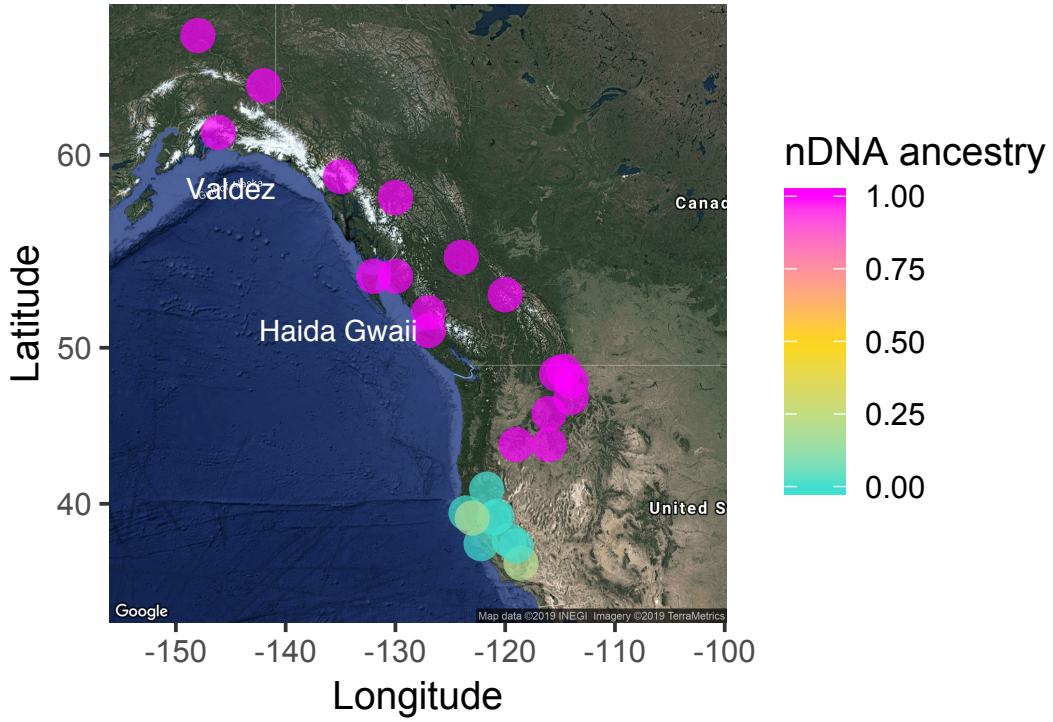

Figure S4 Geographical distribution of ASIP-RALY ancestry.

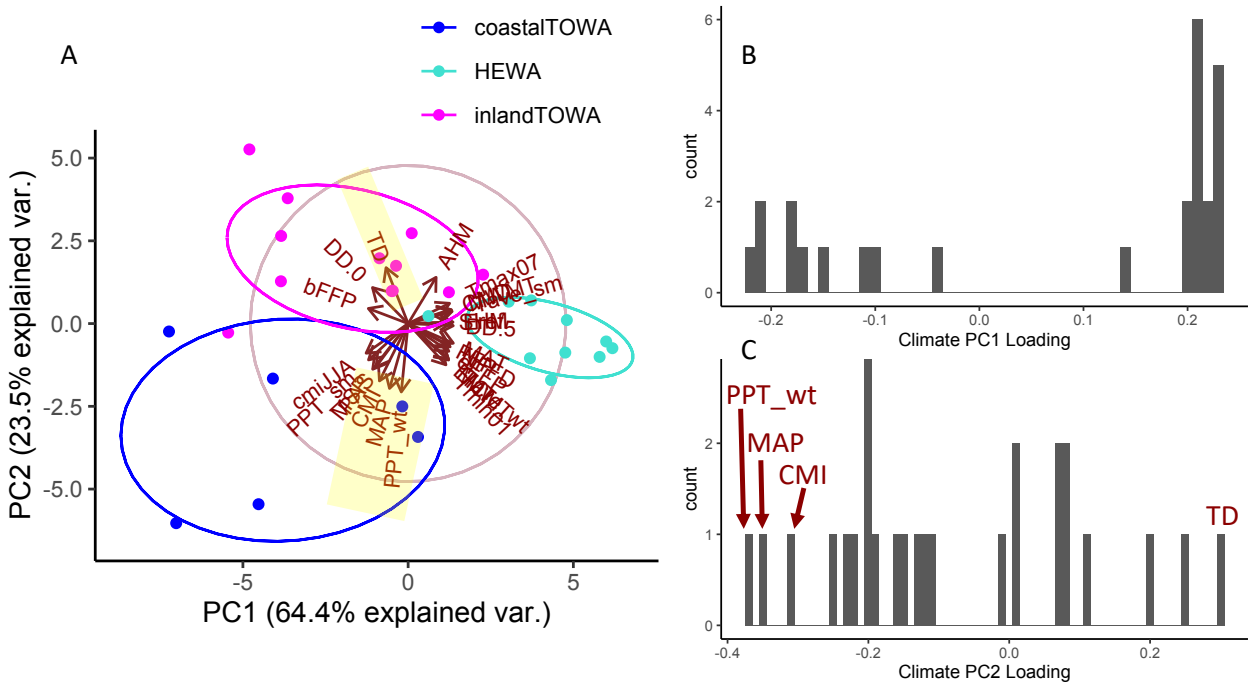

Figure S5 Dissecting climate PCA: **A**, biplot of PCA demonstrating loadings of the 26 climate variables in the PC space. **B**, **C**, histogram of variable loading for PC1 (**B**) and PC2 (**C**). Most of the variables demonstrates strong and even loading along PC1 (**B**), while there are 4 outstanding variables (highlighted in yellow, **A**) explaining PC2 (**C**): Temperature Difference (TD), Climate Moisture Index (CMI), Mean Annual Precipitation (MAP), Winter Precipitation (PPT_wt).

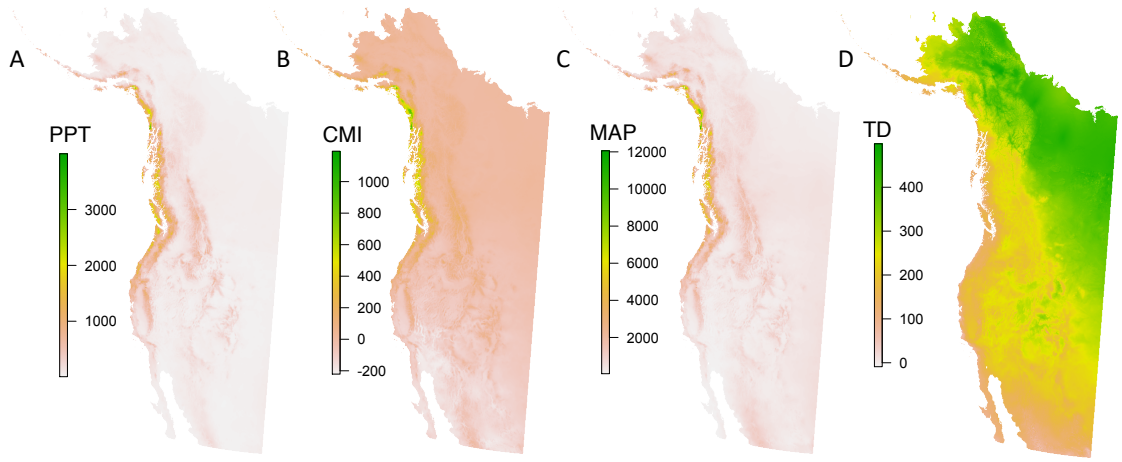

Figure S6 Map demonstrating spatial variation of the 4 key climate variables explaining climate PC2 (see Figure 5, S4): **A**, Winter Precipitation (PPT_wt), **B**, Climate Moisture Index (CMI), **C**, Mean Annual Precipitation (MAP), **D**, Temperature Difference (TD).

Table S1 Divergence of mitochondrial genes between SOCC and inland STOW.

Name	Start	Stop	Length	%Fixed difference	Number of Amino Acid Changes
NAD1	2829	3776	948	0.738	0
NAD2	4009	5043	1035	0.966	4
COX1	5417	6949	1533	0.457	0
COX2	7105	7779	675	0.741	0
ATP8	7861	8022	162	0.617	0
ATP6	8019	8699	681	0.441	1
COX3	8708	9490	783	0.639	0
NAD3	9561	9908	348	1.149	0
NAD4L	9984	10277	294	0.340	0
NAD4	10274	11641	1368	0.439	0
NAD5	11870	13660	1791	0.782	0
COB	13684	14817	1134	0.617	0
NAD6	14996	15511	516	1.357	0

Table S2 names and functions of genes involved in the 1.2 Mb chromosome 5 island of differentiation between inland STOW and coastal STOWs, as well as between inland STOW and SOCC.

ID	Function	start	end	direction	TaeG.ID	Gene name	abbreviation	GO Molecular/biological functions	Related mitochondrial function	Reference
1	Energy-related signaling transduction	3726725	3728106	-	100221946	inositol 1,4,5-trisphosphate receptor-interacting protein-like 1	IP3R1	Protein kinase inhibitaor activity; Ca2+ relase	Mitochondrial voltage-dependent anion channel	Kiviluoto et al. 2013. Molecular Cell Research. 1833: 1612-1624.
2		3909423	3944763	+	100230500	regulator of G protein signaling 6	RGS6	GTPase activity; intracullular signal transduction	Mitochondria-dependent apoptosis	Maity et al. 2011. J. Biol. Chem. 287: 4972-4981.
3		3950193	3969052	+	100230500	regulator of G protein signaling 6	RGS6	GTPase activity; intracullular signal transduction	Mitochondria-dependent apoptosis	Maity et al. 2011. J. Biol. Chem. 287: 4972-4981.
7		4293130	4313915	+	100223971	presenilin 1	PSEN1	ATPase binding; aspartic endopeptidase activity; growth factor receptor binding	Mitochondrial Ca2+ homeostasis, mitochondrial-ER sginaling, mitochondrial respiration	Sarasija et al. 2018. elife. 7:333052.
8		4335909	4376786	+	100222967	papilin, proteoglycan like sulfated glycoprotein	PAPLN	Serine-type endopeptidase inhibitor activity	Mitochondria released serine protease-dependent cell death	Suzuki et al. 2001. Mol. Cell. 8: 613-621.
9		4375633	4408021	-	100226855	NUMB endocytic adaptor protein	NUMB	Alpha-catenin binding; beta-catenin binding; cadherin binding	Drp1-mediated mitochondrial fission	Liu et al. 2019. Antioxid Redox Signal. 30: 1797-1816.
18		4786175	4794879	+	100217582	solute carrier family 5 member 12	SLC5A12	Sodium-glucose transportor; Lactate transmembrane transportor	Lactate uptake, Tricarboxylic Acid (TCA) cycle activation	Pucino et al. 2019. Cell Metabolism. 30: 1055-1074.
19		4875393	4917499	-	100227176	leucine rich repeat containing G protein-coupled receptor 4	LGR4	G protein-coupled peptide receptor activity; activation of adenylate cyclase activity (converting ATP to cAMP)	cAMP/protein kinase A signaling in mitochondrial physiology; Mitochondrion numbers	Amer & Hebert-Chatelain 2018. Biochim Biophys Acta Bioenerg. 1859: 868-877; Wang et al. 2012. J. Am. Soc. Nephrol. 23: 281-293.
20		4951494	4966577	-	100224293	lin-7 homolog C, crumbs cell polarity complex component	LIN7C	Cytoskeletal protein binding; axocytosis; protein transport	Mitochondrial apoptotic Ca2+ signaling	Mendes et al. 2005. J. Biol. Chem. 280: 40892-900.
22		5142823	5172249	-	100224291	kinesin family member 18A	KIF18A	ATPase acitvty, ATP-dependent microtubule motor activity	Mitochondrial transportation	Pereira et al. 1997. J. Cell Biol. 136: 1081-1090.
5	Fatty acid metabolism	4227479	4251606	-	100226854	zinc finger FYVE-type containing 1	ZFYVE1	1-phosphatidylinositol binding; lipid droplet formation; cellular response to starvation; negative regulation of phosphatase activity	Mitochondrial biosynthesis	Cheung et al. 2001. Biochem. 355:113-121.
10		4461621	4463212	+	100221946	acyl-coenzyme A thioesterase 5-like (LOC100221946)	ACOT5	Acyl-CoA metabolic process; long-chain fatty acid metabolic process	Mitochondrial fatty acid beta-oxidation	Hunt et al. 2002. J. Biol. Chem. 277: 1128-1138; Tillander, et al. Trends Endocrinol. Metab. 28: 473-484.
11		4464332	4475990	+	100224870	acyl-coenzyme A thioesterase 1	ACOT1	Acyl-CoA metabolic process; long-chain fatty acid metabolic process	Mitochondrial fatty acid beta-oxidation	Hunt et al. 2002. J. Biol. Chem. 277: 1128-1138; Tillander, et al. Trends Endocrinol. Metab. 28: 473-484.
14		4540115	4587491	-	100190170	prostaglandin reductase 2	PTGR2	Catalyze NAD(P)H to NADP+ Oxidoreductase; prostaglandin metabolic process	Mitochondrial Bax apoptosis	Lalier, et al. 2011. Cell Death. 18: 528-537.
15		4558828	4585008	+	100190170	prostaglandin reductase 2	PTGR2	Catalyze NAD(P)H to NADP+ Oxidoreductase; prostaglandin metabolic process	Mitochondrial Bax apoptosis	Lalier, et al. 2011. Cell Death. 18: 528-537.
17		4683253	4724925	-	100223368	fatty acyl-CoA reductase 1	FAR1	Oxidoreductase activity; ether lipid biosynthetic process; phospholipid biosynthesis	Peroxisome acyl-CoA catalysis that compliments mitochondrial acyl-CoA catalysis	Grevengoed et al. 2014. Annu. Rev. Nutr. 34: 1-30.
21		4993769	4995372	-	751584	brain derived neurotrophic factor	BDNF	Activation of phospholipase C activity; growth factor activity, nerve growth factor binding	Mitochondrial transportation in neurons	Su et al. 2014. J. Biol. Chem. 289: 1213-1226.
4	Ubiquinization	4207849	4221283	+	100223970	DDB1 and CUL4 associated factor 4	DCAF4	Post-translational protein modification; protein ubiquitination		
16		4606924	4640295	+	100220022	ubiquitin protein ligase E3 component n-recognin 1	UBR1	Cellular response to lucine; negative regulation of Target of Rapamycin signaling		
6	Transcription	4257473	4292354	+	100221046	RNA binding motif protein 25	RBM25	mRNA binding; mRNA processing		
12		4481691	4534949	+	100223903	mitotic deacetylase associated SANT domain protein	MIDEAS	Transcription corepressor activity; histone deacetylation		
13		4493333	4522089	-	100223903	mitotic deacetylase associated SANT domain protein	MIDEAS	Transcription corepressor activity; histone deacetylation		

Table S3 Definitions of climate variables from ClimateWNA [50].

Column name	Variable abbreviation	Variable Name
NORM_6190_AHM	AHM	annual heat-moisture index
NORM_6190_bFFP	bFFP	The day of the year on which FFP begins
NORM_6190_CMD	CMD	Climatic moisture deficit
NORM_6190_CMI	CMI	Climatic moisture index
NORM_6190_cmiJJA	cmiJJA	Hogg's summer (June-Aug) climate moisture index
NORM_6190_DD.0	DD.0	degree-days below 0°C, chilling degree-days
NORM_6190_DD.5	DD.5	degree-days above 5°C, growing degree-days
NORM_6190_eFFP	eFFP	The day of the year on which FFP begins
NORM_6190_EMT	EMT	Extreme minimum temperature over 30 years
NORM_6190_Eref	Eref	Hargreaves reference evaporation (mm)
NORM_6190_FFP	FFP	Frost-free period
NORM_6190_MAP	MAP	Mean annual precipitation (mm)
NORM_6190_MAT	MAT	Mean annual temperature
NORM_6190_MCMT	MCMT	Mean coldest month temperature (°C)
NORM_6190_MSP	MSP	May through September precipitation
NORM_6190_MWMT	MWMT	Mean winter temperature
NORM_6190_NFFD	NFFD	number of frost-free days
NORM_6190_PAS	PAS	precipitation as snow
NORM_6190_PPT_sm	PPT_sm	summer precipitation (mm)
NORM_6190_PPT_wt	PPT_wt	winter precipitation (mm)

		ummer heat-moisture index
NORM_6190_SHM	SHM	((MWMT)/(MSP/1000))
NORM_6190_Tave_sm	Tave_sm	summer mean temperature (°C)
NORM_6190_Tave_wt	Tave_wt	winter mean temperature (°C)
NORM_6190_TD	TD	Continentality
NORM_6190_Tmax07	Tmax07	winter mean maximum temperature (°C)
NORM_6190_Tmin01	Tmin01	winter mean minimum temperature (°C)

Table S4 The loading of variables in the climate PCA sorted by their loading along PC1.

Climate Variable	PC1 loading	PC2 loading
cmiJJA	-0.21763162	-0.161032306
bFFP	-0.20625054	0.082056348
PPT_sm	-0.20510277	-0.197710506
DD.0	-0.18472653	0.195940506
MSP	-0.183508	-0.248663084
PAS	-0.1660562	-0.234659637
CMI	-0.1486023	-0.311869513
TD	-0.11435537	0.30417286
MAP	-0.09633503	-0.352869482
PPT_wt	-0.03530235	-0.371886563
AHM	0.14480557	0.251595965
Tmin01	0.19581554	-0.22437128
SHM	0.19978674	0.007101683
MCMT	0.20705781	-0.195744433
Tave_wt	0.20859601	-0.192174784
EMT	0.20990053	-0.198261928
eFFP	0.2107584	-0.150721133
FFP	0.21127471	-0.117676014
Tmax07	0.21473286	0.110168321
NFFD	0.21782119	-0.126257635
CMD	0.22005085	0.076818851
MWMT	0.22507318	0.074770918
Tave_sm	0.22805563	0.066315194
Eref	0.23258512	0.008062124
DD.5	0.23376903	-0.011649806
MAT	0.23420797	-0.106646542