Signatures of mito-nuclear climate adaptation in a warbler species complex

Silu Wang[1], Madelyn J. Ore[1,2], Else K. Mikkelsen[1,3], Julie Lee-Yaw[4,5], Sievert Rohwer[6], and Darren E. Irwin[1]

[1] Department of Zoology, 6270 University Blvd, University of British Columbia, Vancouver, BC, V6T1Z4, Canada
[2] Current address: Cornell Lab of Ornithology, Ithaca, New York, 14850, USA
[3] Current address: Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, M1C 1A4, Canada
[4] Department of Botany, 3200-6270 University Blvd, University of British Columbia, Vancouver, BC, V6T1Z4, Canada
[5] Current address: Biological Sciences, 4401 University Drive, University of Lethbridge, Lethbridge, Alberta, T1K3M4, Canada
[6] 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. Hermit Warblers (*S. occidentalis* abbreviated as HEWA) and inland Townsend's Warblers (*Setophaga townsendi*, abbreviated as i-TOWA) exhibit distinct mtDNA haplotypes and a few nDNA regions of high differentiation, whereas coastal TOWA (c-TOWA) displays a mix of these genetic patterns consistent with ancient hybridization of HEWA and i-TOWA. Of the few highly-differentiated nDNA regions between i-TOWA and HEWA, two of these regions (on chromosome 5 and Z, respectively) are also differentiated between c-TOWA and i-TOWA, similar to the mtDNA pattern. These two nDNA regions are associated with mitochondrial fatty acid metabolism. Moreover, these nDNA regions are correlated with mtDNA ancestries among sites, a pattern consistent with mito-nuclear co-adaptation. Such mito-nuclear coevolution might be driven by climate-related selection, because the mito-nuclear ancestry is correlated with climatic conditions among sampling sites. These results suggest that cryptic differentiation in this species complex has been shaped by climate-correlated adaptation associated with mito-nuclear fatty acid metabolism.

Key Words: *speciation*, *inter-genomic interaction*, *mito-nuclear co-adaptation*, *genetic incompatibility*, carnitine shuttle, climate adaptation, *Setophaga*.

Introduction

Mitochondrial (mtDNA) and nuclear (nDNA) genomes co-function in maintaining critical functions that influence fitness in all eukaryotes (1–5). Populations in different areas may harbour distinct mtDNA owing to selection or drift, and because many nuclear genes encode proteins that function within mitochondria, the two sets of DNA 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 mito-nuclear combinations in hybrids. Specifically, hybrids with nDNA from one species and mtDNA from the other species may have lowered fitness. These types of genetic incompatibilities can play a role in keeping hybrid zones narrow and limiting gene flow between species (8, 9). Hence coadaptation of mtDNA and nDNA is increasingly recognized as being important to speciation (8–10).

Geographic variation in climate is known to select for different mitochondrial genotypes in different areas (6, 11, 12). This may in turn lead to indirect selection on co-functioning nuclear genes. Such climatic mito-nuclear coadaptation can lead to genomic differentiation between population 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 between partially differentiated populations. In particular, we ask whether there is signature of mito-nuclear coevolution. If so, could climate-related selection have driven such coevolution?

While secondary contact between differentiated populations sometimes leads to narrow hybrid zones (13), another possible outcome is the formation of a hybrid or mixed population over a broad region (14–16). Such populations have the potential to reveal

strong selection on suboptimal combinations of genes from the two parental species. Despite increasing interest in mito-nuclear interactions at species boundaries of natural populations with complex population histories (3, 6, 17–20), the degree to which mito-nuclear interactions are important in the differentiation among lineages is not well understood.

Hermit warblers, *S. occidentalis* (referred to as HEWA), inhabit conifer forests along the states of Oregon, California, and southern Washington, U.S.A. To the north of HEWA, Townsend's warblers *Setophaga townsendi* (referred to as TOWA) consist of an inland population that inhabits areas east of the Coast Mountains of British Columbia, Canada and northern Washington, USA (referred as i-TOWA) and a coastal populations (referred as c-TOWA) west of the Coast Mountains (Figure 1A) (21–25). The HEWA and i-TOWA populations demonstrate distinct plumage and mtDNA haplotypes (separated by ~0.8% sequence divergence, diverged ~0.5 million years ago, Figure 1BC) (21, 23, 24, 26), and nDNA difference at a few small genomic regions, one related to plumage differences (ASIP-RALY gene block), whereas the rest of the genome shows very little differentiation (27).

The c-TOWA population is phenotypically identical to i-TOWA, but harbors both HEWA and i-TOWA mtDNA haplotypes (Figure 1BC), suggesting this population is the product of ancient hybridization between HEWA and i-TOWA (24) (Figure 1A). If so, the nuclear genome of c-TOWA should demonstrate a mix of these two ancestries as well. In particular, if the mitochondria-related nDNA regions that are differentiated between HEWA and i-TOWA (26) coevolved with the mitochondrial genome, we expect these nDNA regions to differentiate between i-TOWA and c-TOWA as well. In contrast,

the plumage gene region (represented by the RALY SNP) that differs between HEWA and i-TOWA is expected to remain undifferentiated between c-TOWA and i-TOWA.

The nuclear genome of c-TOWA is thus far unknown. Here, we analyze variation at tens of thousands of single nucleotide polymorphisms (SNPs) throughout the nuclear genome of various HEWA, c-TOWA and i-TOWA populations. In particular, we ask (1) whether the nuclear genomic data is consistent with the mtDNA inference that coastal *townsendi* resulted from admixture; (2) whether the genetic differentiation is related to mitochondrial metabolism, suggesting mito-nuclear coevolution; (3) whether there is climate-related selection on mitonuclear coevolution?

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 c-TOWA, 81 i-TOWA, and 47 HEWA) from the Krosby and Rohwer (2009) study were acquired from GenBank (accession numbers FJ373895-FJ374120). To understand the relationships among the mtDNA sequences, we generated a minimum spanning haplotype network (27) with PopART (28). This network showed two clearly separated haplotype clusters. To understand spatial variation in mtDNA haplotypes, we colored individual c-TOWA sites differently. We then scored each haplotype as 0 and 1 respectively for those that are nested within HEWA haplotype cluster and the i-TOWA cluster (24) (Figure 1C).

Among these individuals with previously-sequenced mtDNA (i.e., from Krosby and Rohwer 2009), we selected a subset of tissue samples (64 i-TOWA, 58 c-TOWA, and 15 HEWA; obtained from the Burke Museum of Natural History and Culture, University of Washington, Seattle, Washington) for nuclear genomic sequencing. We supplemented this set of genetic samples with 30 blood samples that we obtained directly from birds caught in the field during the breeding season of 2016; these included 25 HEWA from California, USA, and 5 i-TOWA from Montana, USA.

GBS pipeline

We prepared genotyping-by-sequencing (GBS) (29) libraries from DNA samples of the 167 individuals described above as our previous study (30). Briefly, we digested genomes with the restriction enzyme PstI, then ligated fragments with barcode and adaptors, and amplified with PCR. Amplified DNA was pooled into two libraries which were then paired-end sequenced: the first (80 individuals) were sequenced with an Illumina HiSeq 2500 automated sequencer (read length = 125bp), and the second (85 individuals) were sequenced with an Illumina HiSeq 4000 (read length = 100bp). After sequencing, we removed two of the HEWA samples because of insufficient read depth and a labeling error respectively, thus 23 HEWA remained for further analysis. To control for plate effects, we randomly assigned samples to different plates and included replicates of three samples among plates. Sequences processing is consistent with our previous study (30). Specifically, the reads were demultiplexed with a custom script and then trimmed using Trimmomatic (31) [TRAILING:3 SLIDINGWINDOW:4:10 MINLEN:30]. Assuming synteny between *Setophaga* and *Taeniopygia guttata* (Zebra Finch) genomes given the evidence of limited rearrangement in avian genomes (32, 33),

we aligned reads to a *T. guttata* reference (34) with bwa-mem (35) (default settings). Variable sites were identified with GATK (36), which resulted in 3,446,846 variable sites among the 165 individuals in the study. We then filtered the variable sites with VCFtools (37) according to the following criteria: 1) removing indels, 2) keeping sites with genotype quality (GQ) > 20, 3) keeping sites with minor allele frequency (MAF) ≥ 0.05, 4) removing sites with > 30% missing genotypes among individuals, and 5) keeping biallelic single nucleotide polymorphisms (SNPs) only. Thereafter 19,083 SNPs remained. To visually evaluate whether there were any plate effects, we compared duplicated samples among different plates, and then visualized samples sequenced on different plates in a principal component analysis (on the covariance matrix). Since plate allocation was random, there should not be a difference in PC space partitioning among the plates. The duplicated samples were compared and then removed from further analysis.

Population structure and genomic differentiation

To examine if the nDNA c-TOWA is a mixture of HEWA and i-TOWA, population structure was examined with principle component analysis (PCA) in the SNPRelate (38) package in R and ancestry assignments in Faststructure with a uniform prior, 10^8 iterations, and K values from 1 to 6 (39). We set out to assess the differences between HEWA, c-TOWA, and i-TOWA. However, the PCA revealed obvious structure within c-TOWA with Valdez, AK (USA) and Haida Gwaii, BC (Canada) populations distinct from the rest of the c-TOWA populations. Thus in subsequent analysis, we compared each of the three c-TOWA groups to the i-TOWA and HEWA groups. We

used the SNPRelate (38) package in R to examine which SNPs were highly correlated with principal component axes.

To examine population differentiation across the genome, for each of the 19,083 filtered SNPs we calculated F_{ST} (40) with VCFtools (37) between 1) i-TOWA ($N = 69$) and HEWA ($N = 38$); 2) c-TOWA ($N = 58$: 10 Haida Gwaii, 15 Valdez, 33 others) and i-TOWA; and 3) HEWA and each of the three c-TOWA clusters.

Candidate genetic regions

The SNPs at F_{ST} peaks between c-TOWA and i-TOWA that are also consistent with the peaks between HEWA and i-TOWA were considered candidate loci for further analyses. One possibility is that these loci are linked to genes that have a mitochondrial function and selection maintains their concordance with mtDNA ancestry. To examine whether these loci are known to be associated with mitochondrial function, we examined what is known about the protein-coding genes in vicinity of the candidate SNPs, using Ensembl (41) and the zebra finch reference genome. If a large region of elevated F_{ST} is involved, Zebrafinch Gene Ontology analysis (42) was conducted to test regional functional enrichment relative to the rest of the genome. While HEWA and i-TOWA differ at the RALY locus that is associated with plumage differences (26), we did not expect this region to differ between coastal and i-TOWA due to their identical plumage features.

Association of mtDNA and nDNA

Coevolution between genomes is expected to lead to association between mtDNA and mt-associated nuclear genes within and among sites. If individuals with mismatched mt-nDNA genotypes are selected against, there should be an association between mtDNA

and nDNA genotypes within each population. Such a force could be counteracted by random mating which breaks down the mt-nDNA association, thus strong selection is required to maintain adaptive mt-nDNA combinations within a single randomly mating population. Over time however, specific geographic regions may favor a particular mtDNA variant and a compatible nDNA variant, increasing mt-nDNA concordance among sampling sites.

To examine within-population association between mtDNA and nDNA ancestry, we conducted permutation tests of independence with the *coin* package (43) in R to examine if there is association between mtDNA group (0 or 1) and nuclear candidate SNP genotype (0 as homozygous HEWA, 0.5 as heterozygous, or 1 as homozygous i-TOWA) within Valdez and the North Vancouver Island populations (c-TOWA sites with $N \geq 10$). The Haida Gwaii population is almost fixed for the HEWA mt haplotype group (Figure 1C), thus mt-nDNA coadaptation would predict that the mitochondrial co-functioning nDNA region is more HEWA-like than the rest of the genome in this population.

Although mito-nuclear coevolution could be masked by random mating within a population, its signature can be captured in the mito-nuclear ancestry associations among populations, which reflects long-term ancestry dynamics. We first calculated the mean mtDNA and nDNA ancestry of each site by averaging the locus-specific ancestry (0 for HEWA and 1 for i-TOWA) among individuals. To examine between-population association between mtDNA and nDNA variants of interest (one on chr 5 and one on chr Z), we employed a partial mantel test (44) with the *vegan* package in R to quantify the association between the among-sites ($N = 19$) distance matrices of mtDNA ancestry and

the nDNA ancestry, while controlling for overall genomic ancestry (i.e., proportion HEWA vs. i-TOWA). In particular, the partial mantel test examined correlation between the among-sites distance matrix of mtDNA and that of the nDNA locus, conditioned on the among-sites distance matrix of admixture index. The admixture index is represented by the PC1 of the genomic PCA with the candidate SNPs (all the SNPs in the 700kb differentiation block on chr5 and the SNP at the peak on chrZ) removed. We employed the genomic PC1 instead of a model-based admixture proportion because the complex admixture history in this case likely violates assumptions of model-based approaches (39). These approaches tend to force admixed individuals into either of the parental clusters, and the output admixture indices for each individual largely depend on the prior distribution (30). In contrast, genomic PC1 naturally represents the admixture between i-TOWA and HEWA (Figure1D).

Climate analysis

To investigate whether there might be selection on mt-nDNA related to climate, we tested association of site-level mt-nDNA ancestry (the averaged site ancestry score of mtDNA, chr 5, and chr Z marker ancestry) and climate variation. To effectively capture annual climate variation among sites, we extracted data from 26 climate variables (Table S1) from ClimateWNA (45) and used PCA to describe climatic variation among sites. We computed pairwise differences between sites for a) PC1 values b) PC2 values c) geographic distance and d) mt-nuclear ancestry. We then looked for an association between climate differences among sites and differences in mt-nDNA ancestry while controlling for geographic distance using a partial mantel test in R with 10000 permutations.

Results

Population structure

The mtDNA haplotype clusters are distinct between i-TOWA and HEWA, with 0.8% minimum mtDNA sequence divergence (Krosby and Rohwer 2009; Figure 1C). Various c-TOWA sampling sites contain a mixture of i-TOWA haplotypes and HEWA haplotypes (Figure 1C), suggesting that these c-TOWA populations are hybrid populations between i-TOWA and HEWA (Krosby and Rohwer 2009). Nuclear genomic variation as assessed through variation in the 19,083 SNPs reveals a pattern broadly consistent with the variation in mtDNA. The i-TOWA and HEWA form two clearly differentiated clusters differing in the first principal component (PC1) of a PCA (Figure 1D, S1), and most individuals from c-TOWA have a somewhat intermediate position. Faststructure analysis further supports the hybrid origin of c-TOWA as k = 2 was most supported and c-TOWA demonstrate admixture between HEWA and i-TOWA ancestry (Figure S2). However, we found substructure within c-TOWA with Valdez and Haida Gwaii forming distinct clusters from the rest (Figure 1D). Valdez differs primarily along the second principal component (PC2), whereas Haida Gwaii differs by a combination of PC1 and PC2. While PC1 is highly correlated with a few strong outlier SNPs, PC2 shows only modest correlations with particular SNPs (Figure S1). Moreover, Valdez and Haida Gwaii demonstrate comparable differentiation to the parental populations as the differentiation between the parental populations (HEWA and i-TOWA) (Figure 1D, S3).

F_{ST} distribution

Genome-wide levels of differentiation show that HEWA and i-TOWA are very similar (Weir and Cockerham's F_{ST} = 0.030) except for a few peaks of differentiation

(Figure 2A). As in the PCA, F_{ST} analysis indicates the Valdez and Haida Gwaii c-TOWA populations are more differentiated from both HEWA and i-TOWA than other c-TOWA are (see F_{ST} values in Figure 2). The rest of the c-TOWA are more similar to i-TOWA (Weir and Cockerham's F_{ST} = 0.009) than to HEWA (Weir and Cockerham's F_{ST} = 0.021).

The i-TOWA and HEWA have a number of peaks of differentiation, with the three highest standing out in particular (Figure 2A) and mapping to chromosomes (chr) 5, 20, and Z in the *T. guttata* reference. One of these (on chr 20) is in the intron of the RALY gene (26), which is known to regulate pigmentation in quail and mice (46, 47). Our earlier study of admixture mapping in the ongoing hybrid zone between i-TOWA and HEWA in the Washington Cascades (26) suggested that this locus is highly associated with plumage colour patterns within that zone. As predicted, the present analysis of genomic variation over a much broader geographic region shows high differentiation at the RALY SNP between sampling regions that differ in plumage (i.e., between HEWA and i-TOWA, Figure 2A, F-H) and low differentiation between regions with similar plumage (i.e., between c-TOWA and i-TOWA, Figure 2B-D).

Mitonuclear genetics

Similar to the chr20 RALY peak, the chr5 and chrZ regions also showed extreme differentiation in the comparison of i-TOWA and HEWA, but opposite to RALY region, these regions also stand out in the comparison between c-TOWA and i-TOWA as the two highest regions of differentiation between those groups (Figure 2A-B). The chr5 differentiation (Figure 2A-C, 3A) involves a ~ 700kb region that is significantly enriched for lipid metabolism (p = 0.0013, $p_{adjusted}$ = 0.021) related to mitochondrial function with

particular relevance to acyl-CoA metabolic process ($p = 0.0027$, $p_{adjusted} = 0.021$), thiolester hydrolase ($p = 0.002$, $p_{adjusted} = 0.021$), and palmitoyl-coA hydrolase activity ($p = 3.7\times10^{-6}$, $p_{adjusted} = 0.0001$), due to the genes ENSTGUG00000011215 and ENSTGUG00000018133 (orthologs of ACOT, acyl-CoA thioesterase). The chr Z SNP (position 66226657 in the *T. guttata* reference) is within the intron of the BBOX1 gene (gamma-butyrobetaine hydroxylase 1) (Figure 3B), which codes for a biosynthesis enzyme of carnitine. Carnitine is the central player in the 'carnitine shuttle' of mitochondria, which activates and transports fatty acid into mitochondria for beta-oxidation (Figure 3C) (1, 48, 49). The other gene associated with this chrZ differentiation is a cytoplasmic-related gene TNP01 that encodes nuclear-cytoplasmic signaling protein, transportin1 (50) (Figure 3B). This chrZ region of differentiation could be narrow, as SNPs flanking this chrZ peak do not demonstrate high F_{ST} (Figure 3B). The chr5 and Z regions are functionally linked to each other as well through the 'carnitine shuttle' of mitochondria (Figure 3C). A moderate peak was found in the F_{ST} scan between the i-TOWA versus other c-TOWA at chr1A (54442413) (Figure 2B), which is in the inter-genic region between golgi gene CHST11 (Carbohydrate Sulfotransferase 11) and the cytoplasmic-functioning gene TXNRD1 (Thioredoxin Reductase 1) (Figure S4).

Mitonuclear association

We then examined whether mtDNA (Figure 1C) and the candidate nDNA markers on chr 5 and Z (Figure 3AB) genotypes were correlated, both among (Figure 4) and within sampling sites (Figure S5) of the admixed c-TOWA population. Among sites, both the chr 5 (Figure 4 AB, partial mantel pearson's product-moment $r = 0.736$, $p < 10^{-4}$) and chr Z marker (Figure 4 AC, partial mantel pearson's product-moment $r = 0.270$, p

296 = 0.03) were correlated with the mtDNA ancestry after controlling for the effect of
297 admixture represented by the distance matrix of genomic PC1 (see Methods). Within
298 sampling sites, partial mitonuclear association was observed. In Valdez, there is an
299 estimated association between mtDNA ancestry and both chr 5 and chr Z marker
300 ancestry, although this was statistically significant only for chr 5 (chr 5 marker: Z = 2.44,
301 N = 14, $p_{(FDR\text{-}corrected)}$ = 0.03, Figure S5A; chr Z marker: Z = 2.14, N = 12, $p_{(FDR\text{-}corrected)}$ =
302 0.065, Figure S5B). In the North Vancouver Island population, neither the chr 5 (Z = -
303 1.41, N = 9, p = 0.157) nor the chr Z (Z = -0.57, N = 11, p = 0.572) marker was
304 significantly associated with mtDNA ancestry (Figure S5A-B). Consistent with the mt-
305 nDNA coadaptation prediction, the *townsendi* homozygotes for the chr 5 and Z markers
306 are missing in the Haida Gwaii *townsendi* population (Figure S5), in which the HEWA
307 mt haplotypes are almost fixed (Figure 1BC).

308 **Climatic association**

309 Climate PC1 (Figure 5AC, S7) explains 64.4% of the variation in climate among
310 sites; this PC was not particularly explained by one or a few climate variables (Figure
311 S7A, B; Table S2). Climate PC2 explains 23.5% of the variation and was predominantly
312 explained by four climate variables (Figure S7C, S8; Table S2): Temperature Difference
313 (TD), Climate Moisture Index (CMI), Mean Annual Precipitation (MAP), Winter
314 Precipitation (PPT_wt). The climate in c-TOWA habitat is similar to that of i-TOWA
315 along PC1, but more similar to that of HEWA along PC2, although there is great climate
316 variation among various c-TOWA populations (Figure 5ACD). Overall, the c-TOWA
317 habitat is moister and more stable in temperature, which is consistent with the coastal-
318 inland humidity gradient (captured by PC2, Figure 5D), and the distribution of mt-nDNA

ancestry appears related to this geographical variation in climate. The mt-nDNA ancestry is significantly correlated with climate PC1 (Figure 5C) (partial mantel test, $r = 0.194$, $p = 0.040$) as well as climate PC2 (partial mantel test, $r = 0.221$, $p = 0.025$, Figure 5B) among 19 sites.

Discussion

Hermit warblers (HEWA) and inland Townsend's warblers (i-TOWA) are distinct in mtDNA (24) and exhibit three strong regions of differentiation in the nuclear genome (26), whereas coastal Townsend's warblers (c-TOWA) harbor admixed mtDNA and nDNA ancestry from HEWA and i-TOWA. Two of the three regions of strong differentiation between HEWA and i-TOWA, on chr 5 and Z, differentiate coastal and i-TOWA as well. Both of these nDNA regions contain genes that are strong candidates for coadaptation with mtDNA, as they are both involved in the mitochondrial carnitine shuttle for fatty acid metabolism. Mitonuclear coadaptation was further supported by mt-nDNA association within Valdez population and among populations. This coadaptation is likely associated with climatic adaptation, because the site-level mito-nuclear ancestry covaries with the site climate conditions.

Coevolution of mtDNA and nDNA

We found the key nDNA differences between c-TOWA versus i-TOWA reside at loci on chr 5 and chr Z associated with mitochondrial fatty acid metabolism (Figure 3), an intriguing result given that coastal and i-TOWA differ so strongly in their mitochondrial haplotype frequencies. This finding points to the possibility of selection on mito-nuclear cofunctions (9, 10). The BBOX1 gene encodes Gamma-butyrobetaine dioxygenase (51),

342 the enzyme that catalyses L-carnitine synthesis (52), which is critical for transporting
343 fatty acids across mitochondrial membranes during beta oxidation (49). Carnitine co-
344 functions with mtDNA and the chr5 region that is enriched for mitochondrial fatty acid
345 metabolism are both involved in 'carnitine shuttle' (Figure 3C). The HEWA nDNA may
346 be partially incompatible with the *townsendi* mtDNA in jointly forming the functional
347 carnitine shuttle leading to selection against mismatched mito-nuclear ancestries. Such
348 selection maintaining mito-nuclear concordance can be counteracted by random mating
349 in admixed populations at each generation and is thus difficult to detect in samples of
350 individuals from a single population. However, mito-nuclear ancestry concordance can be
351 more easily detected through comparison of many populations. This association (Figure
352 4) reveals the potential selection maintaining a functionally compatible mito-nuclear
353 'carnitine shuttle' over a large temporal and spatial scale.

354 These mito-nuclear genotypes are significantly associated with climate,
355 suggesting potential selection on the mt-nDNA combinations related to climate or habitat
356 (which is also associated with climate). The climate in the c-TOWA is similar to i-
357 TOWA habitat along PC1, but similar to HEWA habitat along PC2. Correlations between
358 any two traits that have large-scale geographic variation are expected, making it difficult
359 to confidently infer causality from such associations alone. However, the climate and
360 habitat differences between HEWA, i-TOWA, and c-TOWA are very strong, such that
361 these differences likely cause some selective differences. These patterns are reminiscent
362 of the *Eopsaltria australis* (Eastern Yellow Robin) system in which distinct mt-nDNA
363 combinations are maintained between inland and coastal habitat (6). Fatty acid metabolic
364 genes have also been shown to be targets of climatic adaptation in humans, within

Siberian (53) and Greenlandic Inuit populations (54). Temperature (55) and humidity (56) both influence mitochondrial fatty acid metabolism during beta oxidation, which highly depends on carnitine (55, 56). BBOX1-ACOT-mtDNA genotypes might result in functional difference in fatty acid metabolism that is adapted to specific climate (moist and stable versus dry and variable) in the breeding habitat of these warblers.

Because HEWA has apparently inhabited coastal areas for a long period of time, the HEWA mt-nDNA gene combination may be more suited for coastal habitats compared to those of *townsendi*. If the HEWA mt-nDNA genotype is favored in the coastal habitats, the frequency of HEWA mt-nDNA gene combinations would tend to increase in c-TOWA populations over time. However, ongoing gene flow between i-TOWA and c-TOWA would slow down or prevent such an increase. The Haida Gwaii island and Valdez population could have escaped from such a balance between selection and gene flow due to their isolation from the rest of the populations respectively by the sea and mountain ranges. Another possibility is that frequency-dependent selection is maintaining long-term mt-nDNA polymorphism in the c-TOWA. Future investigation on the spatial and temporal variation of mtDNA-BBOX1-ACOT co-segregation would shed light on the evolutionary forces shaping the present and future of c-TOWA population.

Genomic architecture of differentiation

The distribution of F_{ST} across the genome comparing various c-TOWA to either HEWA and i-TOWA is consistent with the "genic" view of differentiation (26, 57, 58), in which peaks of differentiation represent genetic targets of selection (divergent selection or selection against hybrids) that are highly distinct between populations despite the rest of the genome being homogenized by gene flow (57–59). Despite this 'selection with

388 gene flow' scenario exhibiting a similar genomic differentiation landscape as the classic
389 'divergence with gene flow' model (58, 60), the underlying process is different. In this
390 system, there is a known allopatric phase when HEWA and *townsendi* were separated by
391 ice sheets (Figure 1A) (21, 24). Genetic differentiation that accumulated in allopatry (as
392 opposed to gradual build up at sympatry or parapatry under 'divergence with gene flow')
393 can be homogenized by hybridization at secondary contact, while the climate-related
394 genomic targets (on chr 5 and Z) of selection remain differentiated.

395 Between c-TOWA and HEWA, there are a number of highly differentiated loci
396 (Fig. 2F), one of which is the RALY SNP that was found through admixture mapping to
397 be highly associated with plumage in the narrow Cascades hybrid zone between i-TOWA
398 and HEWA (26). The fact that this marker has now been shown to be strongly associated
399 with plumage both within a local hybrid zone and throughout the whole i-TOWA and
400 HEWA species complex is strong evidence for a causal link between the RALY-ASIP
401 genetic region and plumage differences.

402 **Biogeography and semi-parallel introgression**

403 In addition to being an empirical test of mitonuclear coevolution, the present
404 study also helps clarify the biogeographic history of this warbler complex. Our genomic
405 evidence is consistent with Krosby and Rohwer's (24) conclusion, based on mtDNA, that
406 coastal British Columbia and Alaska was inhabited by geographically structured HEWA
407 populations before *townsendi* expanded from inland areas and mixed with them (24). The
408 HEWA and i-TOWA mtDNA haplotype groups demonstrate many differences (~0.8%),
409 whereas both are common in c-TOWA. It is unlikely that the polymorphisms in mtDNA
410 and nDNA in the c-TOWA were caused by incomplete lineage sorting, as opposed to

411 hybridization (Figure 1A). In a scenario of incomplete lineage sorting, enough time
412 would have passed following population splitting for both i-TOWA and HEWA to have
413 lost the alternative haplotype, while the c-TOWA maintained both. Over such a period of
414 time, sizeable differences would be expected between the HEWA haplotypes found in
415 HEWA population versus c-TOWA, as well as between i-TOWA haplotypes found in
416 coastal versus i-TOWA population. We did not observe such a pattern, as c-TOWA has
417 some mtDNA haplotypes that are identical to HEWA and some that are identical to i-
418 TOWA haplotypes.

419 The higher genome-wide differentiation of the Haida Gwaii and Valdez
420 populations (Figure 2) is consistent with at least partially isolated cryptic refugia of
421 HEWA in coastal Alaska and Haida Gwaii during the last glacial maximum (LGM) (61).
422 Following expansion of i-TOWA from the inland area, presumably after the last glacial
423 period, hybridization between i-TOWA and HEWA apparently led to populations of
424 mixed ancestry along the coast of British Columbia and Alaska (Figure 1A). These
425 coastal populations have the plumage patterns and colors of i-TOWA, which is why they
426 have been classified as members of that species. This uniform i-TOWA appearance has
427 concealed a more complex history of hybridization with ancient and geographically
428 differentiated populations of HEWA.

429 Following expansion of i-TOWA from the interior, gene flow into Haida Gwaii
430 may have been weak due to the expanse of water separating it from the mainland,
431 explaining why that population is more similar to HEWA than other c-TOWA are. Gene
432 flow into Valdez could have also been impeded by geographical barriers, as Valdez is
433 surrounded by mountain ranges (Chugach mountains, Wrangell mountains, and St. Elias

mountains). However, both nuclear and genomic data indicate that Valdez has substantial ancestry from both i-TOWA and HEWA. Despite genome-wide differentiation among these three c-TOWA genetic clusters, there is an interesting parallelism: all the three populations exhibit the i-TOWA-like RALY marker that is associated with plumage (26), and predominantly HEWA-like mitochondria-related markers. Such parallelism might be driven by parallel adaptation to the coastal climate.

Caveats and future directions

While our findings are consistent with mito-nuclear coadaptation being important in the pattern of genomic differentiation within this species complex, this being an observational study we cannot definitively conclude that is the case. The strong associations between geographic variation in climate, mitochondrial haplotypes, and highly differentiated regions of the genome, along with the known roles of those divergent regions in mitochondrial metabolism and the abundant evidence for mito-nuclear coadaptation in other systems (reviewed by (7, 9)), add up to strong correlative evidence for mito-nuclear adaptation in this case. One possibility is that the nuclear and mitochondrial loci are independently selected by the environment, without actual coevolution between the two. Future experimental study should investigate this possibility to distinguish it from actual coevolution. If there is mito-nuclear coadaptation, there should be (1) an epistatic effect of mtDNA and nDNA on fatty acid metabolic phenotypes; (2) the high frequency fatty acid metabolic phenotypes (underpinned by mito-nuclear epistasis) within each site should be more fit for local climate than foreign climate.

Conclusion

Examination of genomic differentiation in this young species group has revealed patterns consistent with climate-related coadaptation among mitochondrial and nuclear genes involved in fatty-acid metabolism. Consistent with the mtDNA pattern, the c-TOWA demonstrate a nuclear genomic pattern consistent with ancient admixture between i-TOWA and a geographically structured ancient HEWA population. Three genetic clusters of c-TOWA are characterized by a mixed genetic ancestry between the parental populations (HEWA and i-TOWA), providing natural replicates for examining the role of selection in shaping genomic differentiation. These three c-TOWA clusters exhibit parallel differentiation from i-TOWA at two of the three most differentiated genomic regions (on chr 5 and chr Z) between i-TOWA and HEWA. Both of these genetic regions are involved in mitochondrial fatty acid metabolism. The geographic distributions of the mito-nuclear genetic combinations related to fatty acid metabolism are associated with geographic variation in climate, suggesting mt-nDNA coevolution may have occurred in response to selection for climate adaptation. Such climate-related mito-nuclear selection could be an important force driving population differentiation in this species complex.

Data Accessibility

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

Acknowledgements

We are grateful to Sharon Birks (Burke Museum) and Chris Wood (Burke Museum) for access to the tissue samples for sequencing. We thank Geoffrey E. Hill for inspiring ideas to this study. We also thank Gil Henriques for providing digital illustrations of the warblers. For helpful discussion we thank Sally Otto, Dolph Schluter, Loren Rieseberg, Dahong Chen, Mike Whitlock, Andrea Thomaz, Armando Geraldes, Meade Krosby, Jared Grummer, and members of the Irwin Lab. We are grateful for research funding provided by the Natural Sciences and Engineering Research Council of Canada (grants 311931-2012, RGPIN-2017-03919 and RGPAS-2017-507830 to DEI; and PGS D 331015731 to SW), 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. S. E. Calvo, V. K. Mootha, The Mitochondrial Proteome and Human Disease. *Annu. Rev. Genomics Hum. Genet.* **11**, 25–44 (2010).

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

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

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

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

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

506 1258–1267 (2018).

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

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

512 9. G. E. Hill, The mitonuclear compatibility species concept. *Auk* **134**, 393–409
513 (2017).

514 10. R. S. Burton, F. S. Barreto, A disproportionate role for mtDNA in Dobzhansky-
515 Muller incompatibilities? *Mol. Ecol.* **21**, 4942–4957 (2012).

516 11. P. Innocenti, E. H. Morrow, D. K. Dowling, Experimental evidence supports a sex-
517 specific selective sieve in mitochondrial genome evolution. *Science (80-.).* **332**,
518 845–848 (2011).

519 12. A. E. Harada, T. M. Healy, R. S. Burton, Variation in thermal tolerance and its
520 relationship to mitochondrial function across populations of Tigriopus
521 Californicus. *Front. Physiol.* (2019) https:/doi.org/10.3389/fphys.2019.00213.

522 13. N. H. Barton, G. M. Hewitt, Adaptation, speciation and hybrid zones. *Nature* **341**,
523 497–503 (1989).

524 14. T. O. Elgvin, *et al.*, The genomic mosaicism of hybrid speciation. *Sci. Adv.* **3**,
525 e1602996 (2017).

526 15. M. Schumer, R. Cui, D. L. Powell, G. G. Rosenthal, P. Andolfatto, Ancient
527 hybridization and genomic stabilization in a swordtail fish. *Mol. Ecol.* **25**, 2661–
528 2679 (2016).

529 16. L. H. Rieseberg, Hybrid origins of plant species. *Annu. Rev. Ecol. Syst.* (1997)
530 https:/doi.org/10.1146/annurev.ecolsys.28.1.359.

531 17. P. A. Gagnaire, E. Normandeau, L. Bernatchez, Comparative genomics reveals
532 adaptive protein evolution and a possible cytonuclear incompatibility between
533 European and American Eels. *Mol. Biol. Evol.* **29**, 2909–2919 (2012).

534 18. J. B. M. Sambatti, D. Ortiz-Barrientos, E. J. Baack, L. H. Rieseberg, Ecological
535 selection maintains cytonuclear incompatibilities in hybridizing sunflowers. *Ecol.*
536 *Lett.* **11**, 1082–1091 (2008).

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

539 20. Z. Boratyński, T. Ketola, E. Koskela, T. Mappes, The Sex Specific Genetic

Variation of Energetics in Bank Voles, Consequences of Introgression? *Evol. Biol.* **43**, 37–47 (2016).

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

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

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

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

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

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

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

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

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

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

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

32. G. Zhang, *et al.*, Comparative genomics reveals insights into avian genome evolution and adaptation. *Science (80-.).* **346**, 1311–1320 (2014).

33. H. Ellegren, Evolutionary stasis: the stable chromosomes of birds. *Trends Ecol. Evol.* **25**, 283–291 (2010).

34. W. C. Warren, *et al.*, The genome of a songbird. *Nature* **464**, 757–762 (2010).

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

36. M. D. McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, David Altshuler, Stacey Gabriel, *et al.*, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* (2010) https:/doi.org/10.1101/gr.107524.110.20.

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

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

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

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

41. S. E. Hunt, *et al.*, Ensembl variation resources. *Database (Oxford).* **2018** (2018).

42. X. Wu, M. Watson, CORNA: Testing gene lists for regulation by microRNAs. *Bioinformatics* **25**, 832–833 (2009).

43. T. Hothorn, K. Hornik, M. A. van De Wiel, A. Zeileis, Implementing a Class of Permutation Tests COIN Package. *J. Stat. Softw.* **28**, 1–23 (2008).

44. P. Legendre, *Numerical ecology, 2nd English Edition* (Elsevier Science, 1998) https:/doi.org/10.1017/CBO9781107415324.004.

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

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

47. N. J. Nadeau, *et al.*, Characterization of Japanese quail yellow as a genomic deletion upstream of the avian homolog of the mammalian ASIP (agouti) gene. *Genetics* **178**, 777–786 (2008).

48. K. Tars, *et al.*, Targeting carnitine biosynthesis: Discovery of new inhibitors against γ-butyrobetaine hydroxylase. *J. Med. Chem.* **57**, 2213–2236 (2014).

49. N. Longo, M. Frigeni, M. Pasquali, Carnitine transport and fatty acid oxidation. *Biochim. Biophys. Acta - Mol. Cell Res.* **1863**, 2422–2435 (2016).

50. J. Brelstaff, *et al.*, Transportin1: A marker of FTLD-FUS. *Acta Neuropathol.* **122**,

591–600 (2011).

51. F. M. Vaz, S. Van Gool, R. Ofman, L. Ijlst, R. J. A. Wanders, Carnitine biosynthesis: Identification of the cDNA encoding human γ-butyrobetaine hydroxylase. *Biochem. Biophys. Res. Commun.* **250**, 506–510 (1998).

52. H. S. Paul, G. Sekas, S. A. Adibi, Carnitine biosynthesis in hepatic peroxisomes: Demonstration of γ-butyrobetaine hydroxylase activity. *Eur. J. Biochem.* **203**, 599–605 (1992).

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

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

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

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

57. C. I. Wu, The genic view of the process of speciation. *J. Evol. Biol.* **14**, 851–865 (2001).

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

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

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

61. A. B. A. Shafer, C. I. Cullingham, S. D. Côté, D. W. Coltman, Of glaciers and refugia: A decade of study sheds new light on the phylogeography of northwestern North America. *Mol. Ecol.* **19**, 4589–4621 (2010).

62. S. Mehta, Oxidation of Fatty Acids - via Beta-Oxidation | Biochemistry Notes | PharmaXChange.info. *Biochem. Notes, Notes* (2013).

63. A. L. Beaudet, Brain carnitine deficiency causes nonsyndromic autism with an extreme male bias: A hypothesis. *BioEssays* **39**, 1–25 (2017).

Figures

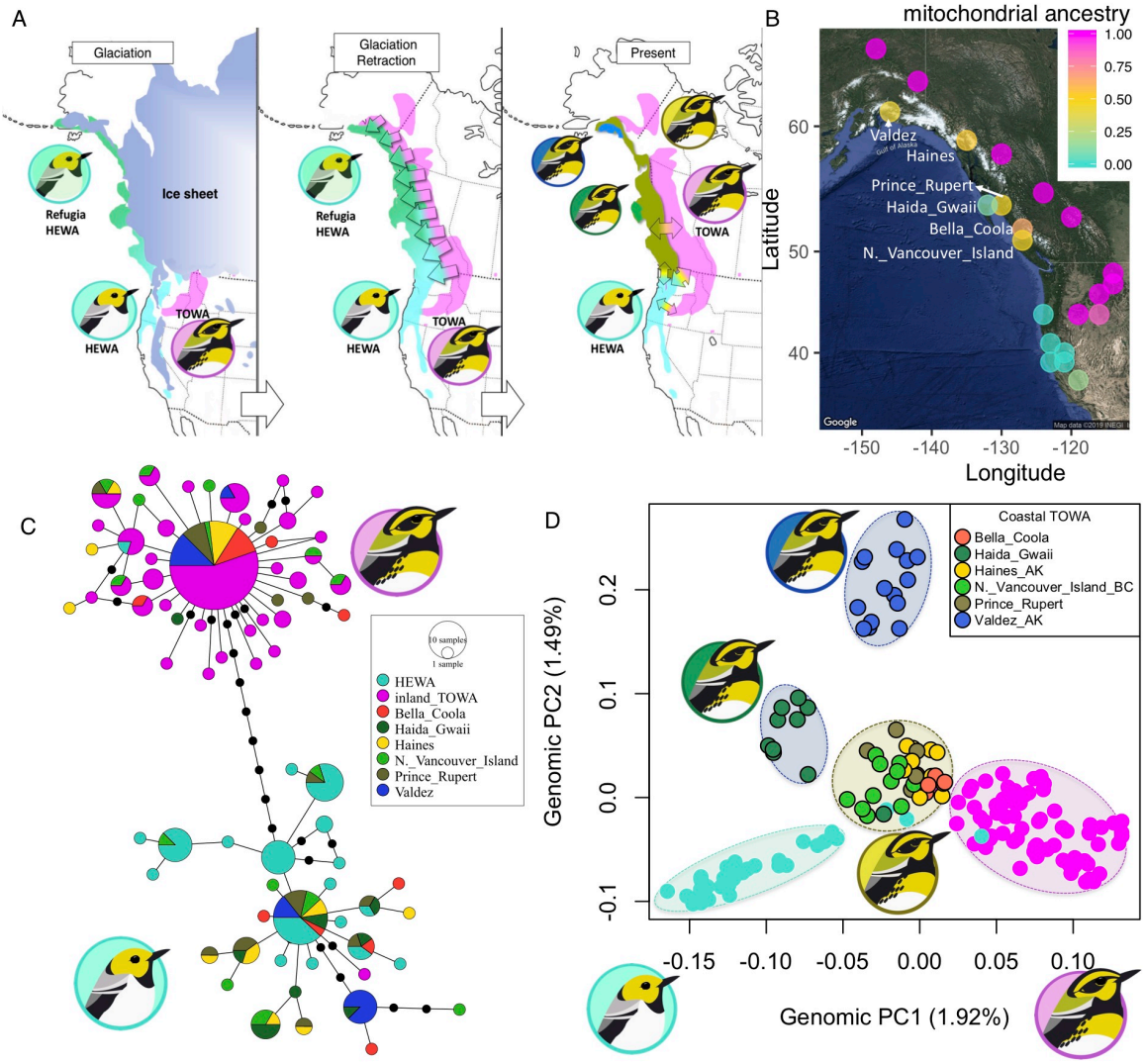


Figure 1 c-TOWA demonstrates admixed mitochondrial and nuclear ancestry of HEWA and i-TOWA. (**A**) Illustration of the history of differentiation and hybridization between TOWA and HEWA during glacial expansion and retraction. Left: During the last glacial maxima, the HEWA and TOWA populations resided in isolated glacial refugia. Center: after glacial retraction, the refugial HEWA and i-TOWA expanded and hybridized along a broad inland-to-coastal front parallel to the coast. Right: the historical hybridization resulted in c-TOWA populations with admixed ancestry although the plumage resembles that of i-TOWA. Population substructure within c-TOWA could be a result of refugia isolation. **(B)** Distribution of mitochondrial ancestry of HEWA, c-TOWA, i-TOWA sites. (**C**) Haplotype network of mitochondrial NADH gene in HEWA, i-TOWA, and various c-TOWA populations, with mtDNA sequences from (24). 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. The c-TOWA (Bella Coola: orange, Haida Gwaii: dark green; Haines:

yellow; North Vancouver Island: light green; Prince Rupert: brown; Valdez: royal blue) populations harbor admixed mtDNA haplotype (some mtDNA haplotypes nested within the turquoise HEWA (banding code: "HEWA") cluster whereas some are in the magenta colored i-TOWA cluster). (**D**) Principle component analysis of 19083 high quality SNPs in the genome. The c-TOWA is intermediate in PC1 but distinct from i-TOWA and HEWA in PC2. PC1 represents admixture between i-TOWA and HEWA, and PC2 represents unique differentiation of c-TOWA.

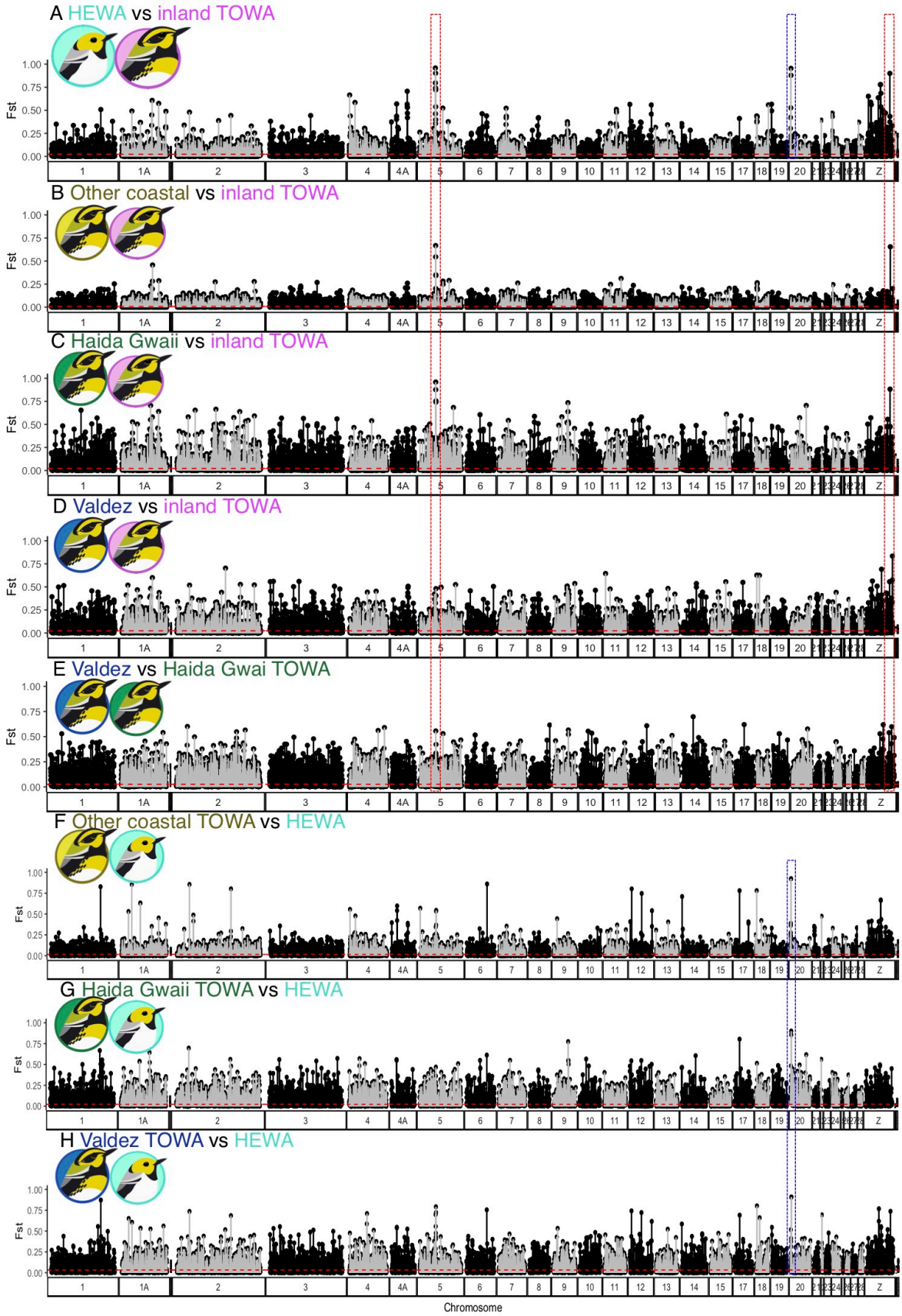


Figure 2 F_{ST} scan between HEWA and i-TOWA (**A**), other coastal and i-TOWA (**B-D**), between Valdez and Haida Gwaii c-TOWA (**E**), as well as HEWA and c-TOWA (**F-G**). Three distinctive differentiation peaks were found between inland and coastal *townsendi* that reside in chromosome 1A, 5 and Z (red boxes, **A-C**). The RALY locus demonstrates consistent differentiation between HEWA and various TOWA (blue boxes, **F-G**). The red horizontal dotted lines represent the genome-wide mean F_{ST}.

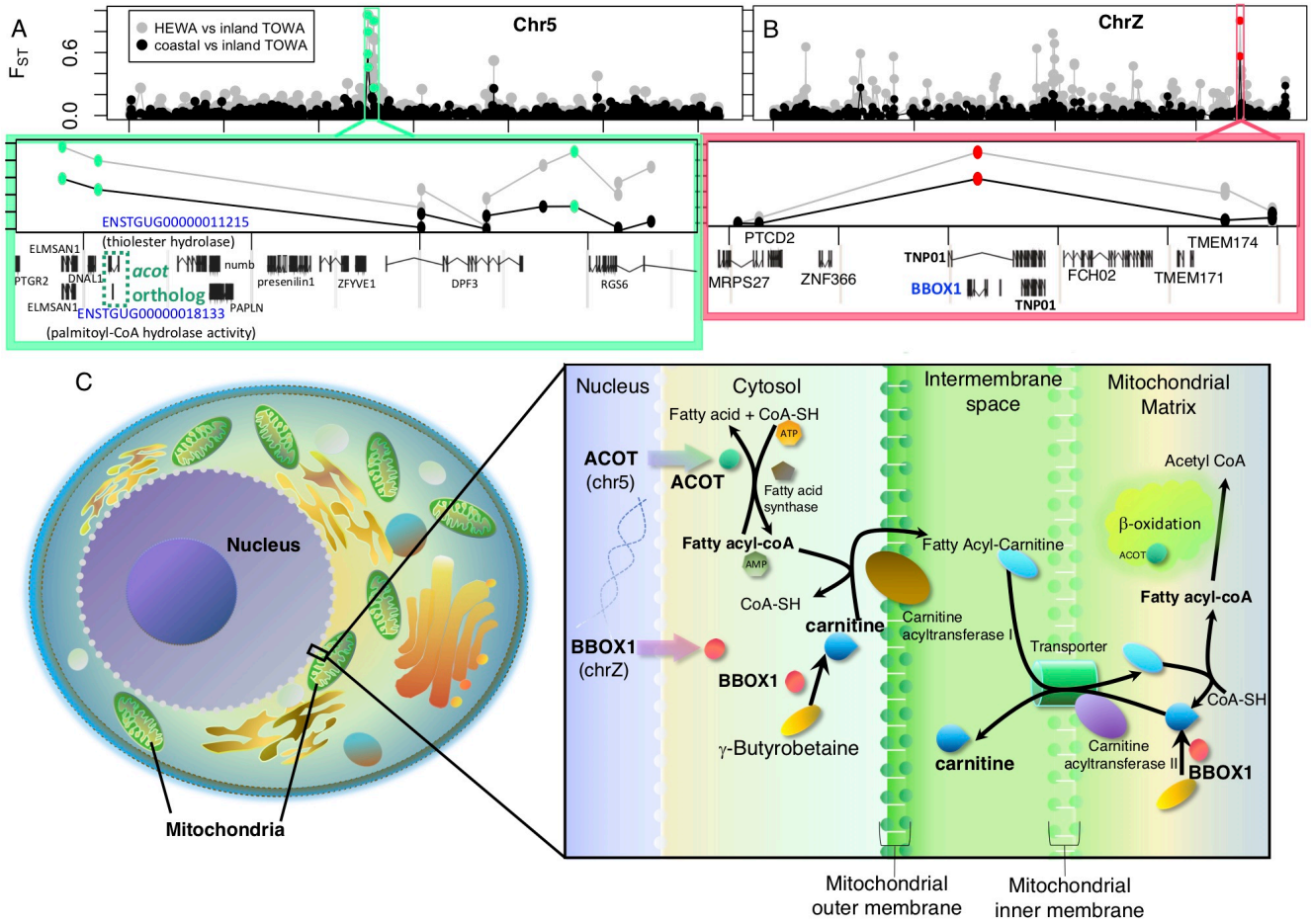
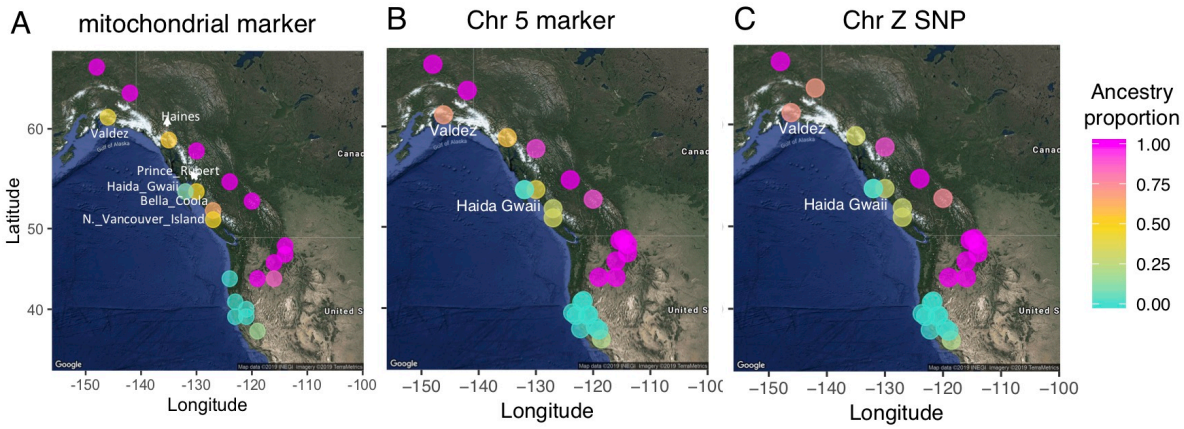


Figure 3 c-TOWA (black) and HEWA (grey) exhibit concordant genetic differentiation from i-TOWA at regions in chr 5 (**A**) and Z (**B**) that are associated with genes involved in mitochondrial fatty acid metabolism (**C**). **A-B**, the window of F_{ST} scan on chr 5 (**A**) and Z (**B**) around the F_{ST} peaks. On the bottom, there is a vertical grey line every 10,000 bases. **A**, the region of differentiation delineated by the jade green F_{ST} peaks are significantly enriched for acyl-CoA metabolism, because of the two orthologs of ACOT (inside the dotted jade green box). **B**, the violet red F_{ST} peak is localized at the Z-chromosome within the intron of gene BBOX1 (involved in fatty acid transportation across mitochondria membranes) and a cyto-nuclear signaling gene TNP01. **C**, Illustration of the mitochondrial carnitine shuttle in which the nuclear genes associated with chr 5 (ACOT) and Z (BBOX1) differentiation were bolded. BBOX1 synthesizes carnitine (bolded), which is essential to transport fatty acyl-coA (bolded) into mitochondrial matrix for beta-

687 oxidation. If not transported into mitochondria, the fatty acyl-coA can be converted back
688 to fatty acid catalyzed by ACOT. This illustration is a synthesis of existing illustrations
689 on carnitine shuttle (62, 63).
690
691



692
693 **Figure 4** Correlated distribution of ancestry proportion (0 = HEWA ancestry, colored in
694 turquoise; 1 = i-TOWA ancestry, colored in magenta) of mtDNA marker (**A**), and nDNA
695 markers (**B**-**C**) related to mitochondrial fatty acid metabolic on chr 5 (**B**) and Z (**C**).
696 There is significant correlation between mitochondrial and nuclear genetic markers after
697 controlling for genome-wide ancestry (Partial Mantel Test, $p < 0.05$).

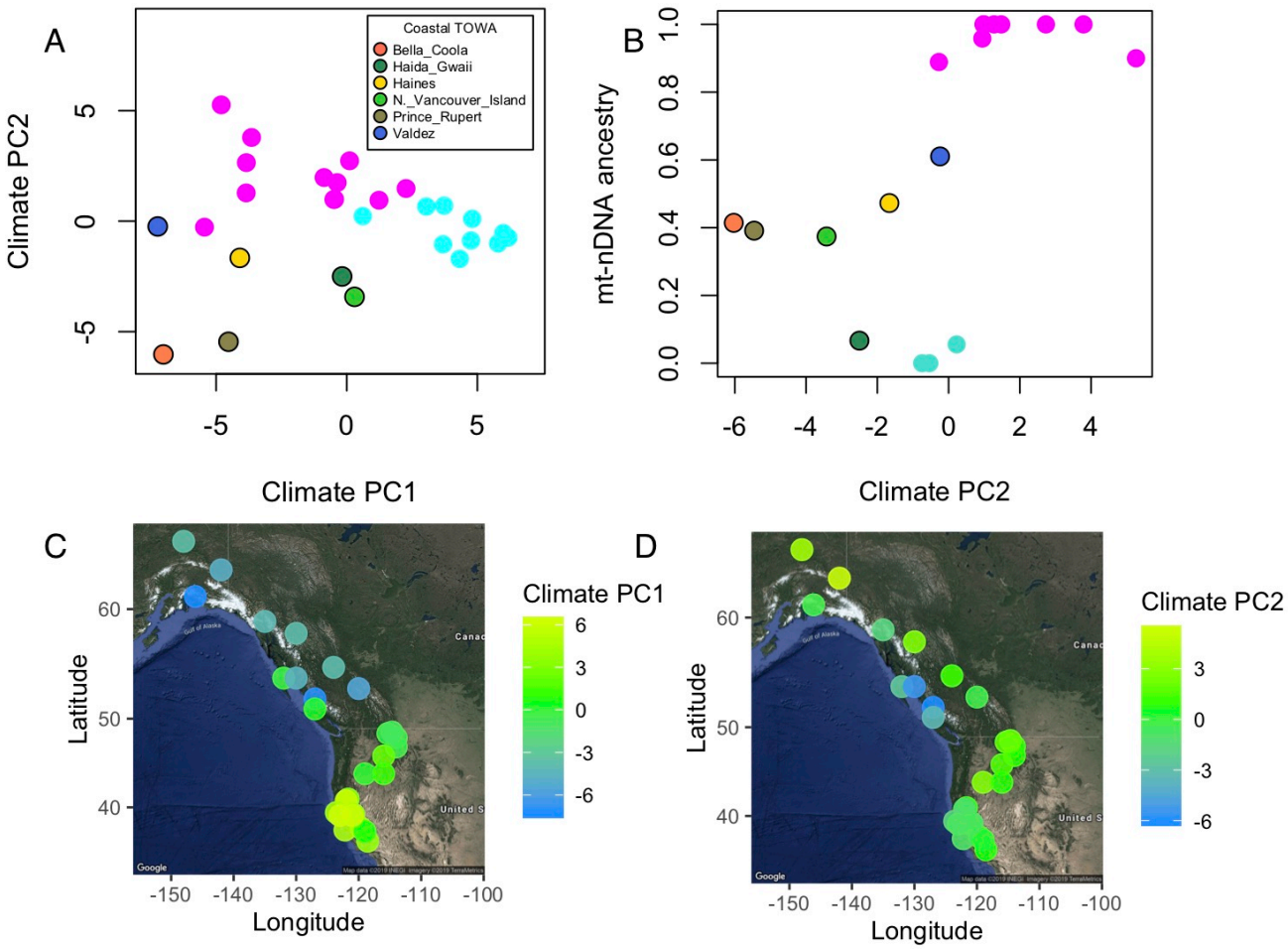


Figure 5 Climate principal component analysis of 26 climate variables from ClimateWNA. (**A**) Climate PC1 and (**B**) PC2, in which HEWA (turquoise), i-TOWA (magenta) and c-TOWA habitats are different. (**B**) Site mean mtDNA and candidate nuclear marker ancestry is correlated with local climate PC2. (**C**)-(**D**), spatial distribution of climate PC1 (**C**) and PC2 (**D**).

Supplementary Figures

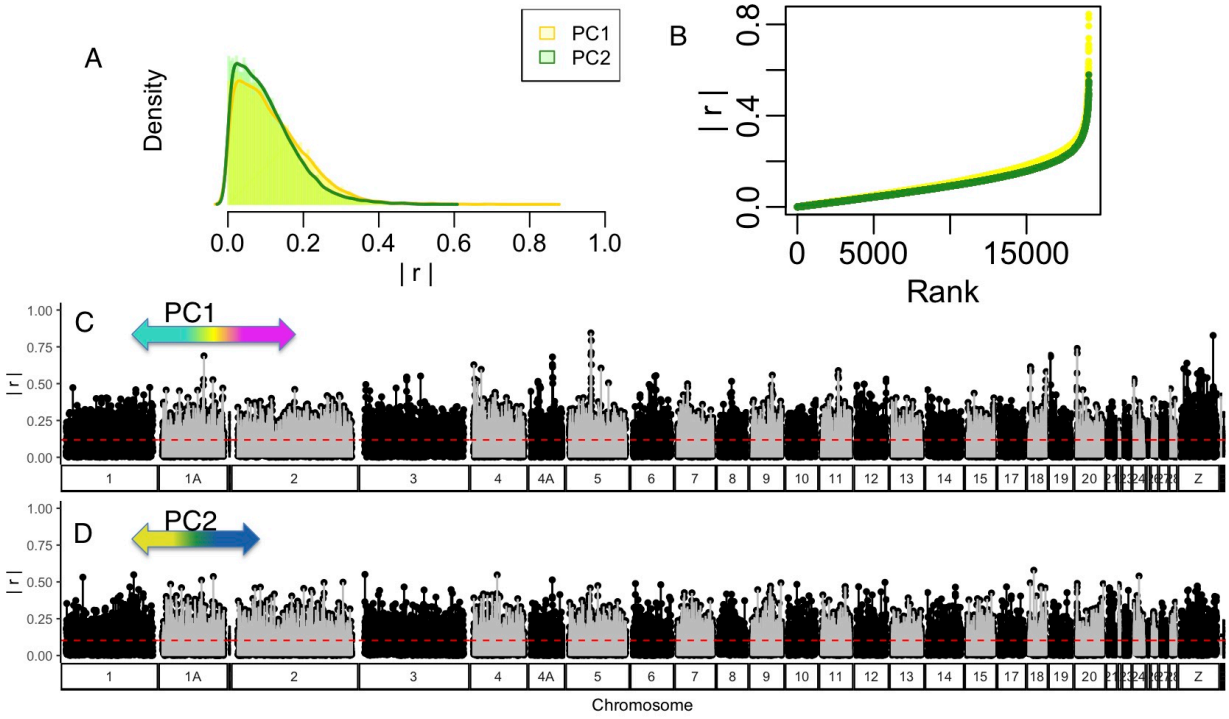


Figure S1 A, Density distribution of absolute correlation coefficient of each SNP with PC1 (yellow shade) and PC2 (green shade). **B**, There were more SNPs that are highly-correlated with PC1 (yellow line) than PC2 (green line). **C**, **D**, Absolute correlation coefficient between SNPs and PC1 (**C**) and PC2 (**D**). The horizontal red dash clines are the mean. **C**, certain regions in chromosome 1A, 5, 18, 20, and Z are highly correlated with PC1. **D**, SNPs are correlated with PC2 similarly across the genome (no obvious peaks).

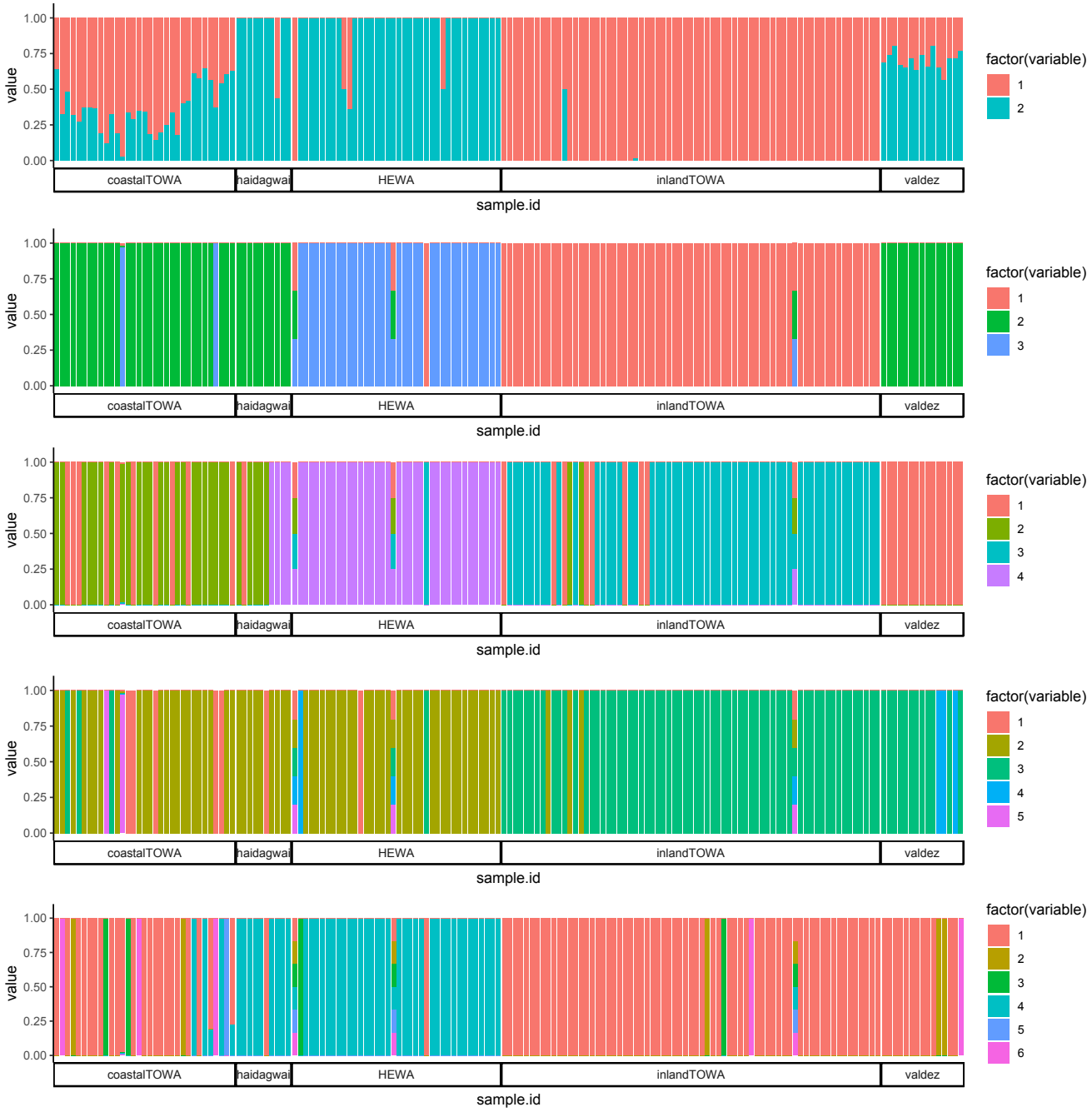


Figure S2 Faststructure analysis with k = 1 to 6 of the 165 individuals collected from HEWA, i-TOWA, Haida Gwaii c-TOWA, Valdez c-TOWA, and other c-TOWA populations.

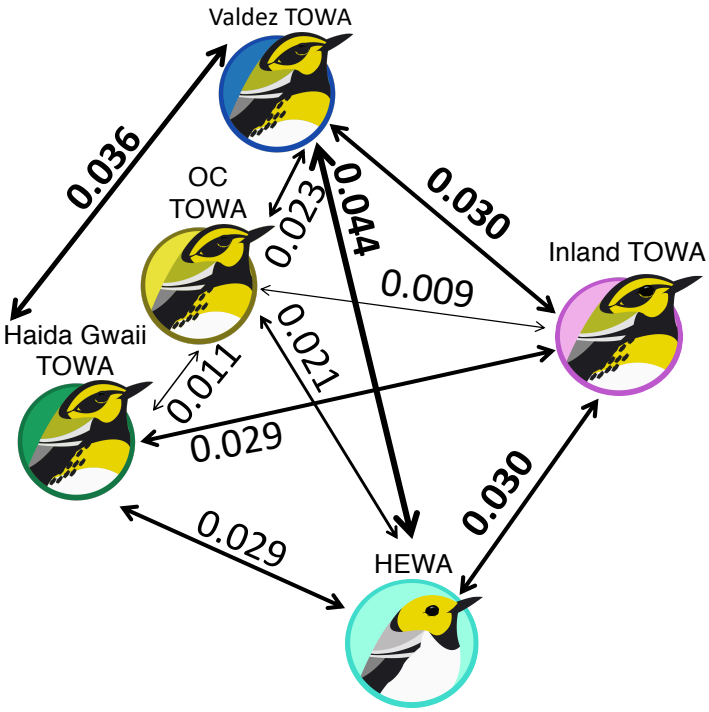


Figure S3 The Weir & Cockerham weighted F_{ST} among HEWA, inland and coastal TOWA(green: Haida Gwaii, blue: Valdez, dark yellow: other coastal (OC) c-TOWA), which demonstrates a gradient of differentiation from the parental populations (i-TOWA, colored as magenta; HEWA, color in turquoise). Each double-head arrow represents a pairwise comparison among the populations. The populations are oriented as their relative geographical location. The widths of the arrows are weighted by the F_{ST} between each pair of populations. Surprisingly some c-TOWA populations demonstrates significantly greater differentiation (paired t-test, $p < 0.001$) from the parental populations than between the parental populations ($F_{ST} = 0.03$).

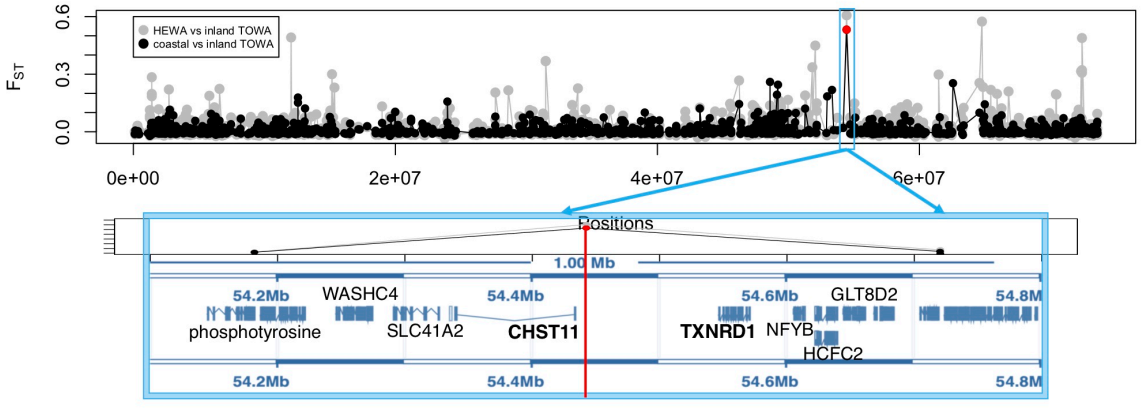


Figure S4 F_{ST} scan between HEWA and i-TOWA (grey) and between i-TOWA and c-

TOWA (black) across chromosome 1A. The strongest F_{ST} peak (red dot) is concordant between the two comparisons (grey versus black). Zooming in around this peak (blue box), this peak is in the intergenic region between gene CHST11 and TXNRD1.

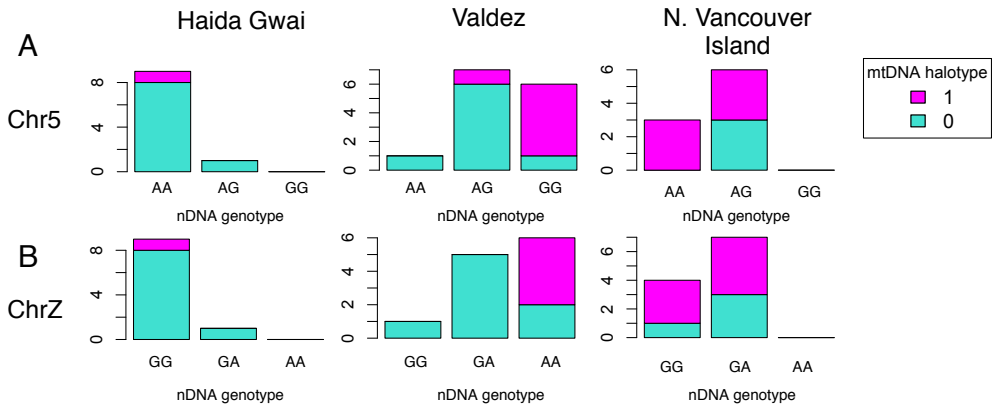


Figure S5 Bar plots showing the mitonuclear ancestry association within Haida Gwaii and Valdez c-TOWA populations (A, chr5 marker; B, chr Z marker).

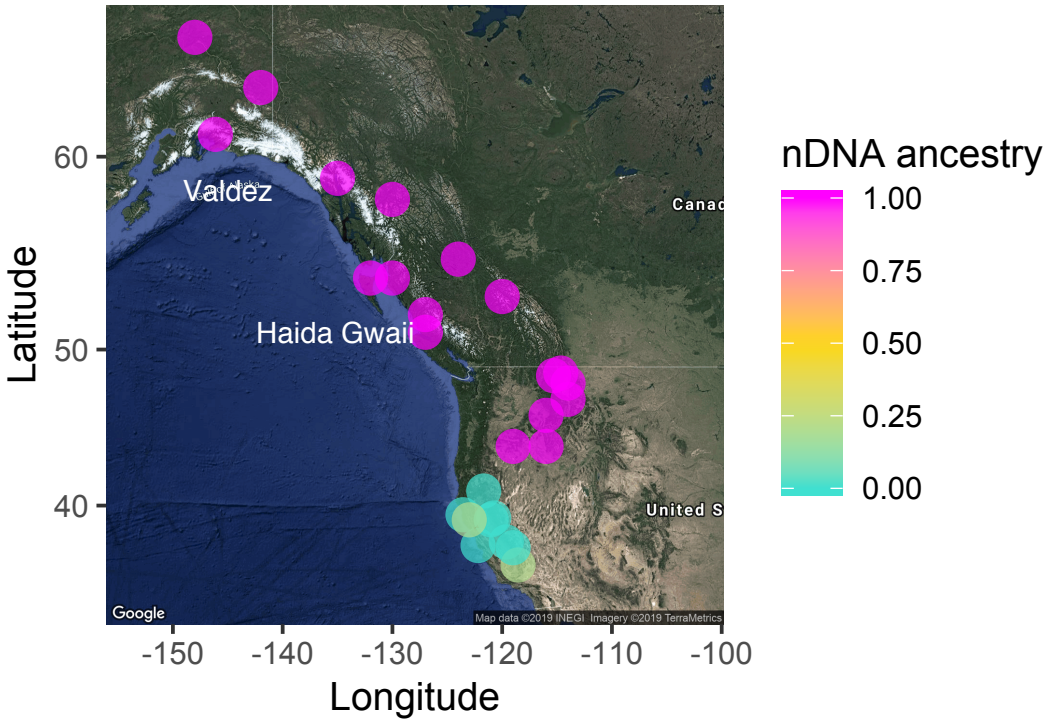


Figure S6 Geographical distribution of RALY ancestry.

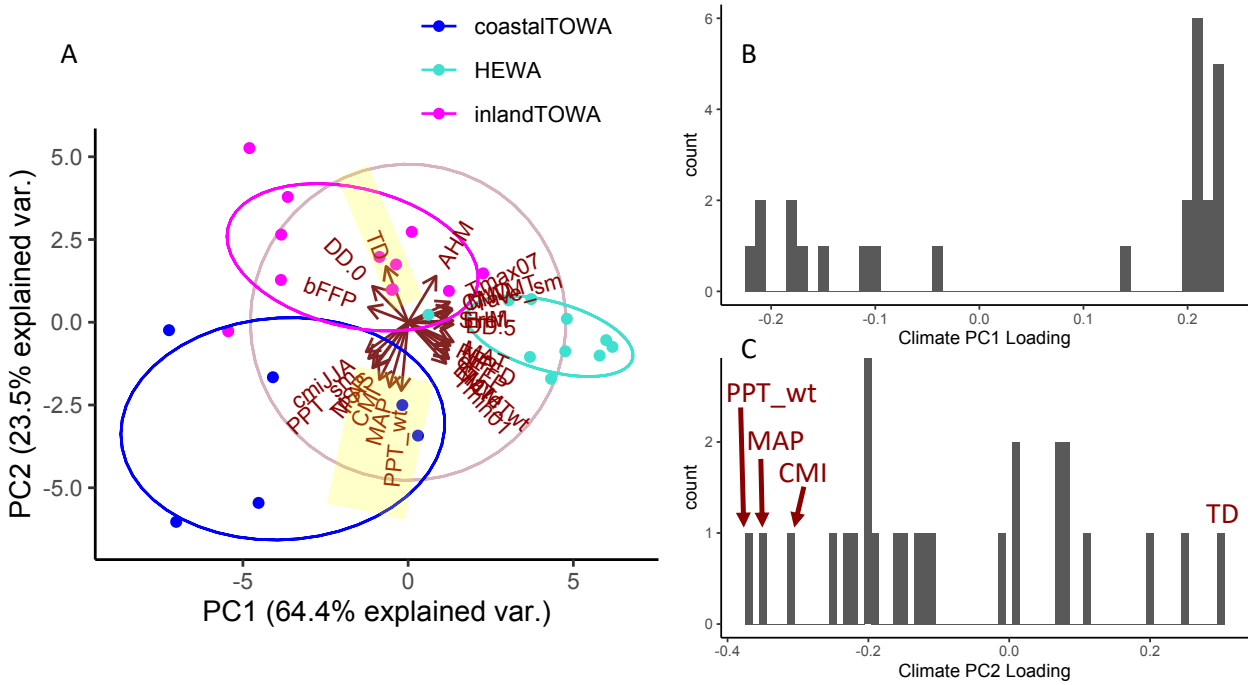


Figure S7 Dissecting climate PCA: **A**, biplot of PCA demonstrating loadings of the 26 climate variables in the PC space. **B**, **C**, histogram of variable loading for PC1 (**B**) and PC2 (**C**). Most of the variables demonstrates strong and even loading along PC1 (**B**), while there are 4 outstanding variables (highlighted in yellow, **A**) explaining PC2 (**C**): Temperature Difference (TD), Climate Moisture Index (CMI), Mean Annual Precipitation (MAP), Winter Precipitation (PPT_wt).

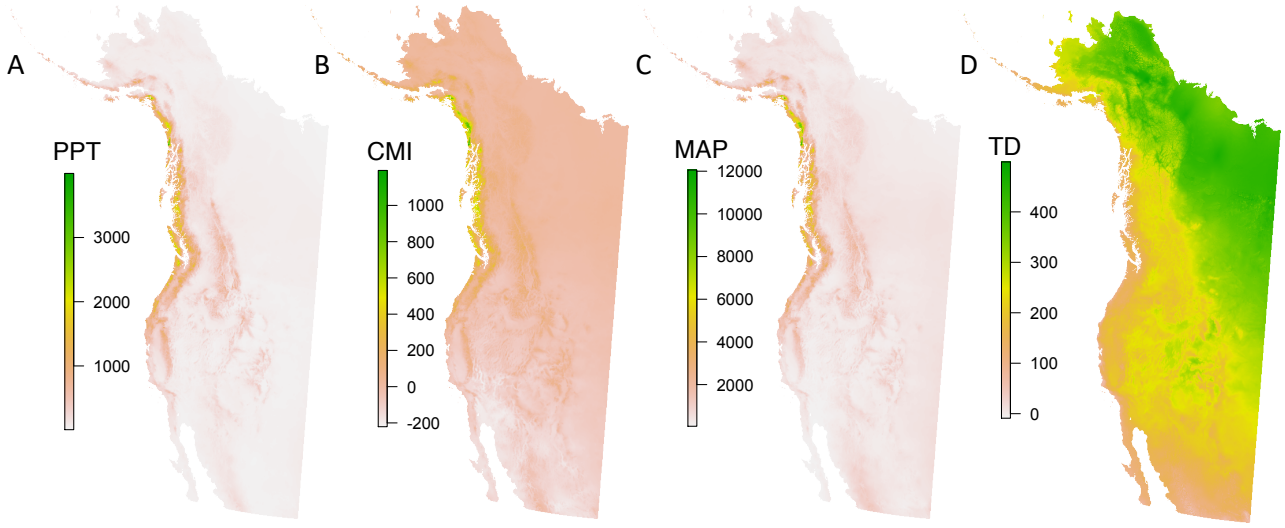


Figure S8 Map demonstrating spatial variation of the 4 key climate variables explaining climate PC2 (see Figure 7, S3): **A**, Winter Precipitation (PPT_wt), **B**, Climate Moisture Index (CMI), **C**, Mean Annual Precipitation (MAP), **D**, Temperature Difference (TD).

Table S1 Definitions of climate variables

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 Septermber previpitation
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)
NORM_6190_SHM	SHM	ummer heat-moisture index ((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 S2 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