Long-distance dispersal, ice sheet dynamics, and mountaintop isolation underlie the genetic structure of glacier ice worms

Scott Hotaling[1], Daniel H. Shain[2], Shirley A. Lang[3], Robin K. Bagley[4], Lusha M. Tronstad[5], David W. Weisrock[6], and Joanna L. Kelley[1]

Affiliations:
[1] School of Biological Sciences, Washington State University, Pullman, WA, USA
[2] Department of Biology, Rutgers University, Camden, NJ, USA
[3] Graduate School of Biomedical Sciences, Rowan University, Stratford, NJ, USA
[4] Department of Biology, University of Iowa, Iowa City, IA, USA
[5] Wyoming Natural Diversity Database, University of Wyoming, Laramie, WY, USA
[6] Department of Biology, University of Kentucky, Lexington, KY, USA

Correspondence: Scott Hotaling, School of Biological Sciences, Washington State University, Pullman, WA, 99164, USA; Email: scott.hotaling@wsu.edu; Phone: (828) 507-9950; ORCID: 0000-0002-5965-0986

Running title: Population genomics of ice worms

Keywords: RAD sequencing, Pacific Northwest, phylogeography, *Mesenchytraeus solifugus*, global change biology, annelid

Abstract:

Disentangling the contemporary and historical factors underlying the spatial distributions of species is a central goal of biogeography. For species with broad distributions but little capacity to actively disperse, disconnected geographic distributions highlight the potential influence of passive, long-distance dispersal (LDD) on their evolutionary histories. However, dispersal alone cannot completely account for the biogeography of any species, and other factors–e.g., habitat suitability, life history–must also be considered. North American ice worms (*Mesenchytraeus solifugus*) are ice-obligate annelids that inhabit coastal glaciers from Oregon to Alaska. Previous studies identified a complex biogeographic history for ice worms, with evidence for genetic isolation, unexpectedly close relationships among geographically disjunct lineages, and contemporary migration across large (> 1,500 km) areas of unsuitable habitat. In this study, we analyzed genome-scale sequence data for most of the known ice worm range. We found clear

support for divergence between populations along the Pacific Coast and the inland flanks of the Coast Mountains (mean F_{ST} = 0.60), likely precipitated by episodic ice sheet expansion and contraction during the Pleistocene. We also found support for LDD of ice worms from Alaska to Vancouver Island, perhaps mediated by migrating birds. Our results highlight the power of genomic data for disentangling complex biogeographic patterns, including the presence of LDD.

Introduction:

For more than a century, long-distance dispersal (LDD) among presumably isolated populations has intrigued biologists [1-4]. Historically considered rare and unpredictable, the idea that LDD can act as a general mechanism influencing the biogeography of presumably dispersal-limited, macroscopic organisms has gained traction in recent years, with examples accumulating for both plants [5] and invertebrates [6-9]. Many animal vectors play an integral role in plant and invertebrate LDD [e.g., 10, 11], however, in most cases, the resulting LDD is limited to less than 10 km. For more extreme LDD events (e.g., greater than 100 km), the most common animal vector is likely migratory birds, as they seasonally move by the millions over broad spatial scales and geographic barriers, visiting similar habitats along the way [12, 13]. Through this mechanism, dispersal units (e.g., whole organisms, eggs, seeds, etc.) may be ingested and dispersed after passing through the digestive tract [9] or by directly adhering to the bird's exterior [12]. Thus, as long as there is an opportunity for migratory birds and dispersal units to interact, the opportunity also exists for LDD. Physically quantifying LDD is difficult, however, because it requires real-time sampling and searching (internal and external) of migrating birds for hitchhiking dispersers. Moreover, because rare migratory events can affect species distributions [14] and influence genetic differentiation among populations [15], even thorough physical surveys of migratory birds that find no evidence for LDD cannot rule out its presence. Therefore, alternative approaches for detecting and characterizing LDD should be employed. Because population genomic tools are well-suited to detecting gene flow and genetic structure among populations [e.g., 16], these tools are also well-suited to the indirect detection of LDD, even in the absence of field observations.

Many mechanisms influence genetic relationships among taxa and a range of factors should be considered when attempting to reconstruct biogeographic patterns. For instance, pulses and contractions of glaciers and ice sheets have shaped the evolutionary histories of populations

67 and species throughout Earth's history [17-20]. These ice sheet dynamics typically affect
68 organisms by separating and reconnecting populations as ice cover changes. However, some
69 species are directly tied to ice sheets [e.g., the meltwater stonefly, 17, 21] and are therefore much
70 more susceptible to ice sheet influence on their evolutionary trajectories. Perhaps no species is
71 more directly tied to ice sheets than glacier ice worms, *Mesenchytraeus solifugus* in North
72 America [22] and *Sinenchytraeus glacialis* in Tibet [23]. The geographic range of *M. solifugus*
73 (hereafter "ice worm") follows a coastal arc from the Chugach Mountains in southeast Alaska to
74 the Cascade Volcanoes of Washington and Oregon [24]. Ice worms cannot tolerate temperatures
75 more than roughly ±7 °C from freezing and require glacier ice for survival and reproduction [25].
76 With such unique ecology and physiology, and a dispersal-limited life history, the evolutionary
77 history of ice worms since diverging from conspecifics [26, 27] should be relatively simple with
78 gene flow occurring during glacial periods and isolation (paired with genetic drift) driving
79 divergence among mountaintop-isolated populations during interglacial periods. Natural systems,
80 however, are often more complex than expected and indeed, the evolutionary history of ice
81 worms challenges general expectations of gene flow and evolutionary dynamics in ice-
82 dominated, mountain ecosystems.
83 Previous genetic studies based on one or two genetic markers identified three ice worm
84 lineages: a "northern" clade in southern Alaska, a "central" clade in southeast Alaska and
85 northern British Columbia, and a "southern" clade ranging over much of British Columbia to
86 southern Oregon [24, 25, 27]. Surprisingly, phylogenetic evidence supported the northern and
87 southern lineages as being most closely related to one another despite the central clade separating
88 them geographically. The most curious aspect of ice worm biogeography, however, has been the
89 repeated discovery of closely related ice worms on glaciers several hundred to thousands of
90 kilometers south of their closest genetic relatives [25, 27; P. Wimberger, unpublished data].
91 These disjunct northern ice worms co-occurred with, but appeared genetically distinct from, their
92 conspecifics (either central or southern clade ice worms) on the same glaciers. Dial *et al.* [25]
93 laid out three possible explanations for this pattern: wind transport, passerine-mediated dispersal,
94 or a more extensive previous range of the northern clade. While wind transport seems unlikely,
95 the potential for passerine-mediated dispersal is reasonable, particularly in light of other
96 examples of bird-mediated LDD [e.g., 7, 28]. The third scenario, a more extensive distribution of
97 the northern clade with holdover lineages inhabiting the same glacier as more recent colonizers

98 could indeed result in more than one distinct lineage on the same glacier. However, this pattern
99 may only apply to mitochondrial DNA (mtDNA) since mtDNA is maternally inherited and does
100 not recombine. For the nuclear genome, unless strong reproductive isolation exists between the
101 holdover lineages and more recent colonizers, genetic differences would be rapidly homogenized
102 by gene flow and recombination. Assuming no selection against migrants, a reasonable
103 expectation given the similarity of habitat across the ice worm range, the frequency of
104 contemporary versus historical mtDNA haplotypes would therefore depend upon time since
105 introduction, scale of migration (i.e., number of introduced haplotypes), and chance.
106 In this study, we leveraged a modern population genomic toolkit to add new perspective
107 to the age-old challenge of identifying LDD in wild populations. We also provide new insight
108 into how multiple factors can interact to shape the evolutionary history of species. We
109 hypothesized that the biogeographic history of ice worms stemmed from a confluence of factors:
110 extreme LDD, glacier dynamics, and mountaintop isolation. To test this hypothesis, we
111 generated a genome-wide single nucleotide polymorphism (SNP) data set to answer three
112 specific questions: (1) How do the clades previously diagnosed from a small number of markers
113 hold up to genome-wide scrutiny? (2) What, if any, genomic evidence exists for LDD in ice
114 worms? (3) How do the evolutionary relationships among ice worm populations and genetic
115 clusters align with glacial history in the region [e.g., 29]? Beyond a refined view of ice worm
116 evolution, our study confirms that LDD does occur in ice worms, providing an example of LDD
117 in an annelid and a rare population genomic exploration of the process. Moreover, while
118 considerable evidence details the existence of refugia in the Pacific Northwest (PNW) during the
119 Pleistocene [30], few studies have explored how ice sheet dynamics influenced the evolutionary
120 history of species directly tied to them [e.g., 17]. Our results reveal the profound impact that ice
121 sheet formation during the Pleistocene (~2.5 million – 11,700 years ago) which flowed
122 episodically from the crest of the Coast Mountains [29] may have had on ice worm evolution,
123 possibly precipitating an ongoing speciation event. Broadly, our findings highlight the power for
124 population genomics to capture contemporary evidence of LDD while also providing
125 biogeographic evidence for reconstructing the glacial history of a region.

126 **Table 1.** Sampling information and summary statistics for all ice worm populations included in this study.
127 Abbreviations: *n* = sample size, π = nucleotide diversity, Het = heterozygosity, F_{IS} = inbreeding
128 coefficient, AK = Alaska, BC = British Columbia, VI = Vancouver Island. π, Het, and F_{IS} were calculated
129 for variable sites only. Asterisks indicate populations included in both Dial *et al.* [27] and this study.

Population	Latitude, longitude	State/Prov.	Elev. (m)	*n*	π	Het	F_{IS}
Learnard* (LEA)	60.806, -148.721	AK	624	3	0.114	0.100	0.026
Davidson* (DAV)	59.067, -135.551	AK	986	15	0.078	0.065	0.037
Treaty* (TRE)	56.586, -130.151	BC	1376	8	0.026	0.033	-0.012
Bear (BEA)	56.096, -129.681	BC	648	6	0.039	0.046	-0.013
William Brown* (WIB)	54.611, -129.129	BC	1260	6	0.056	0.057	0.000
Jacobson* (JAC)	52.050, -126.072	BC	1249	5	0.108	0.106	0.005
White Mantle* (WHM)	50.795, -125.153	BC	1764	8	0.131	0.135	-0.005
Comox* (COM)	49.545, -125.355	BC (VI)	1881	4	0.091	0.072	0.034
Mariner (MAR)	49.460, -125.764	BC (VI)	1754	4	0.098	0.084	0.024

130

131 **Methods:**

132 *Sample collection, library preparation, and SNP calling*

133 During the summer of 2009, ice worms were collected from nine glaciers across most of
134 their geographic range (Figs. 1, S1; Table 1). Samples were stored in > 80% EtOH until DNA
135 was extracted from 59 ice worms using a Qiagen DNEasy Blood and Tissue Kit. Double-digest
136 restriction-site associated DNA (ddRAD) sequencing libraries were prepared following Peterson
137 *et al.* [31] with restriction enzymes EcoRI and NlaIII. During library preparation, samples were
138 divided into two groups and each sample was assigned a unique, variable-length barcode [32]
139 which was incorporated during adapter ligation. Size selection for a 350 bp ± 35 bp window was
140 performed with a Pippen Prep (Sage Science), and both sample groups were subsequently
141 amplified using PCR primers containing a group-specific barcode. The 59-sample library was
142 sequenced on one lane of an Illumina HiSeq4000 at the University of Illinois High-Throughput
143 Sequencing and Genotyping Unit with single-end, 100 bp chemistry.

144 Raw reads were demultiplexed, quality-filtered, and ddRAD loci were assembled *de novo*
145 using the *process_radtags* and *denovo_map* functions of the Stacks v1.46 pipeline [33]. We
146 allowed a maximum distance between stacks of 2 and a minimum read depth of 10. Next, we
147 applied a stringent filtering scheme to identify high-confidence SNPs that were shared among
148 many individuals. We only included SNPs if they were present in ≥ 5 populations, genotyped in
149 ≥ 50% of individuals per population, and were in Hardy-Weinberg equilibrium with a minor
150 allele frequency of ≥ 0.025 overall. We further restricted analyses to one random SNP per locus
151 for all analyses except fineRADstructure (see below). All post-Stacks filtering steps were

performed in PLINK v1.07 [34] and the commands used in this study are provided on GitHub (https://github.com/scotthotaling/ice_worm_ddRAD).

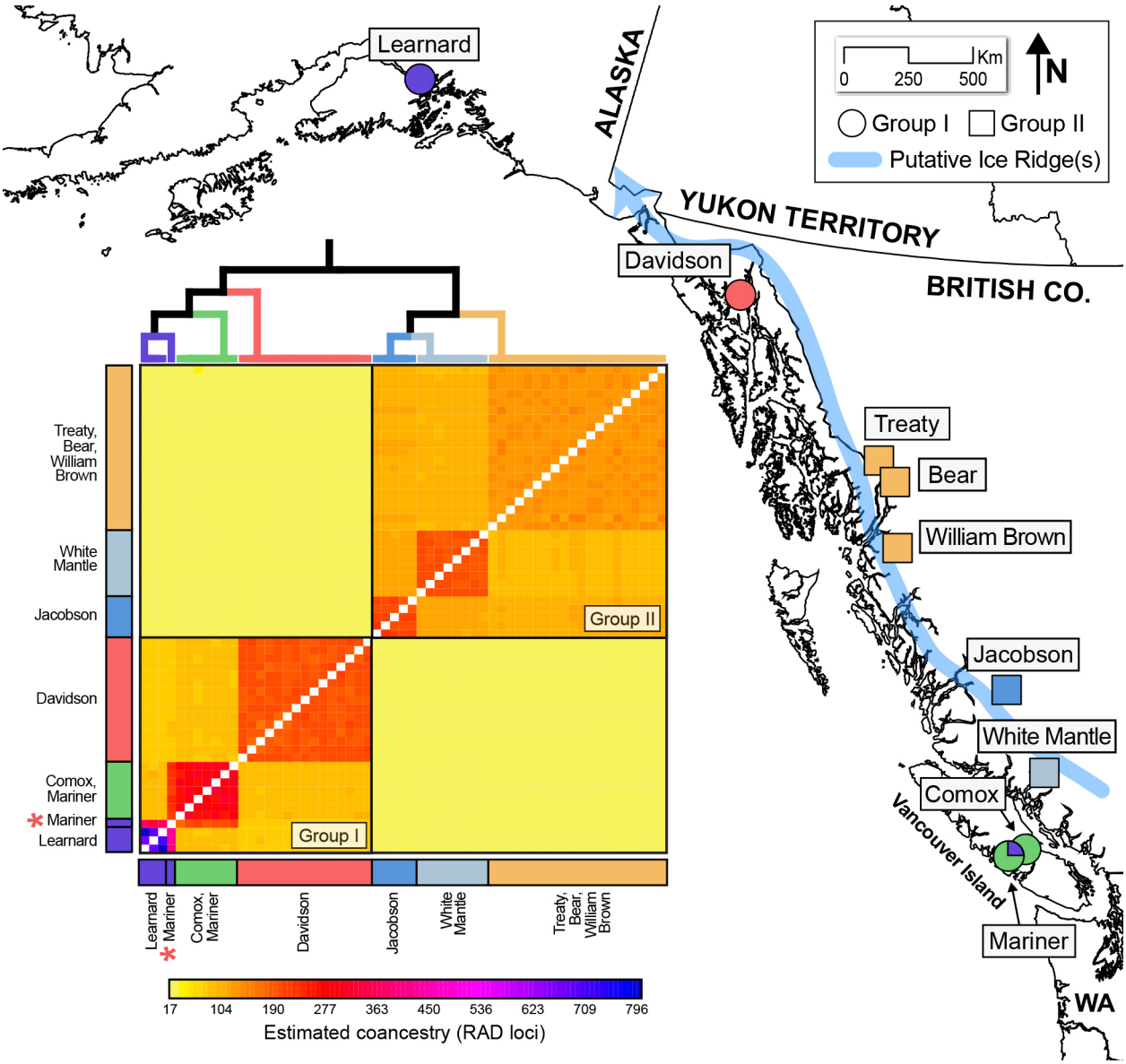

Figure 1. Ice worm populations sampled for this study. Color-coding reflects the results of a fineRADstructure coancestry analysis. Group I (circles) and II (square) populations were generally defined by their presence to the east or west of a key line where ice ridge(s) putatively formed during the Pleistocene > ~25,000 years ago [29] as well as their distance to the Pacific Ocean. The deep divergence between groups I and II is clearly evident along with more recent differentiation within each group. One individual from the Mariner Glacier (asterisk) was admixed between the Mariner/Comox (Vancouver Island) and Learnard (southern Alaska) clusters, indicating recent LDD.

162 *Population genetic and phylogenetic analyses*

163 For each population, we calculated nucleotide diversity (π), heterozygosity (Het), and the
164 inbreeding coefficient (F_{IS}). We also calculated a pair-wise AMOVA F_{ST} for all population
165 combinations [35] with the Stacks *populations* module. To test for a signature of isolation-by-
166 distance [IBD; 36], we estimated Euclidean distances among sites with Google Earth and tested
167 the correlation between geographic distance and F_{ST} with four Mantel tests performed in
168 GenoDive v2.0b27 [37]. The first Mantel test included all nine populations and the second
169 excluded both Vancouver Island populations (Mariner and Comox) to assess whether unique
170 histories for those populations were significantly altering results. The third and fourth Mantel
171 tests focused on signatures of IBD within groups "I" and "II" (see Results). We also measured
172 the Euclidean distance to the Pacific Ocean for each population in Google Earth Pro as an
173 additional spatial comparison of groups I and II. We determined if mean distances to the Pacific
174 Ocean differed between groups with a one-way ANOVA.

175 Population structure was inferred in two ways: a maximum likelihood-based method
176 using ADMIXTURE 1.3.0 [38] and a discriminant analysis of principal components (DAPC)
177 with the R package *adegenet* [39]. ADMIXTURE analyses were performed with default settings,
178 a range of clusters (*K*) from 1-12, and 25 replicates per *K* with the current time as the random
179 seed. The cross-validation (CV) error for each *K* was plotted to identify the best-fit *K* (minimized
180 CV across the mean of all replicates for each *K*). After identifying the best-fit *K*, we considered
181 the replicate that minimized CV across all 25 replicates for all *K*'s to be the best-fit solution
182 overall. However, because all runs did not converge on the same result, we also inspected best-fit
183 solutions for other replicates of $K = 7$ (the best-fit *K* overall) to clarify the distribution of best-fit
184 solutions. For DAPC, we first used the *find.clusters* function to identify the optimal *K* [i.e., the *K*
185 with the lowest Bayesian Information Criterion (BIC)]. Next, to avoid over-fitting of the model,
186 we retained the appropriate number of principal components (PCs) according to the α-score [PCs
187 retained = 6, Fig. S2]. We performed a final DAPC analysis using the best-fit *K* and optimal
188 number of PCs identified in the previous two steps.

189 We extended our population structure analyses to infer both shared ancestry and
190 phylogenetic relationships in two ways: a nearest neighbor haplotype approach to infer
191 coancestry with fineRADstructure [40] and phylogenetic relationships inferred from singular
192 value decomposition estimates for quartets of tips using SVDQuartets [41] as implemented in

193 PAUP* v4.0a159 [42]. For fineRADstructure, we used 100,000 burn-in iterations followed by
194 100,000 iterations sampled every 1,000 steps for the Markov chain Monte Carlo clustering
195 algorithm. Next, we used 10,000 iterations of the tree-building algorithm to assess genetic
196 relationships among clusters. Since fineRADstructure is a haplotype-based approach, analyses
197 were performed using all variable sites for a given ddRAD locus (i.e., a haplotype) rather than
198 randomly selected single SNPs per locus. For SVDQuartets, we performed exhaustive sampling
199 of all possible quartets (every combination of four tips) and branch support was estimated with
200 100 nonparametric bootstrap replicates.
201
202 *Demographic modeling*
203 To test hypotheses of demographic history and estimate divergence times for the two
204 major groups identified in our population genetic and phylogenetic analyses (I and II, see
205 Results), we performed demographic modeling. We used fastsimcoal2 v2.603 [43] which
206 leverages a coalescent-based model to estimate demography from the site frequency spectrum
207 (SFS). We designed and tested the fit of four two-lineage models (Fig. 2A): no gene flow (M1),
208 unidirectional gene flow from group I into II (M2), unidirectional gene flow from group II into
209 group I (M3), and bidirectional gene flow (M4). All four models also included parameters for
210 ancestral (N_{ANC}) and current (N_I and N_{II}) effective population sizes as well as divergence time
211 (T_{DIV}). We maximized the number of shared SNPs between our focal groups by selecting the
212 four individuals with the least missing data from the same population in each group (group I =
213 Davidson; group II = Treaty). We also only retained loci with no missing data, which yielded
214 2,714 SNPs across the eight individuals. For each model, we performed 50 replicate runs, each
215 with 100,000 simulations and 100 cycles of a conditional maximization algorithm. We specified
216 the nuclear mutation rate at 3.5E-9 per site per generation, based on the estimate for *Drosophila*
217 *melanogaster* [44]. To identify the best-fit model, we calculated an Akaike Information Criterion
218 (AIC) score for each model and selected the best model as the one with the lowest AIC relative
219 to the other models. We generated 95% confidence intervals (CIs) by simulating 50 SFS
220 replicates from the best-fit run of the best-fit model. Next, we performed the same 50 replicate
221 analyses described above for each of the 50 newly simulated SFSs.
222 For divergence times, our final results are in generations and we refrain from converting
223 this estimate to years before present because the generation time for ice worms is not known.

Estimates of generation time for the family Enchytraeidae also vary widely, ranging from 21 days at 18°C [45] to a full year at 10°C [46]. Moreover, no estimates of generation time at low temperatures (e.g., ~0°C) are available for the group. Complete details of our demographic analyses are provided in the Supplementary Materials.

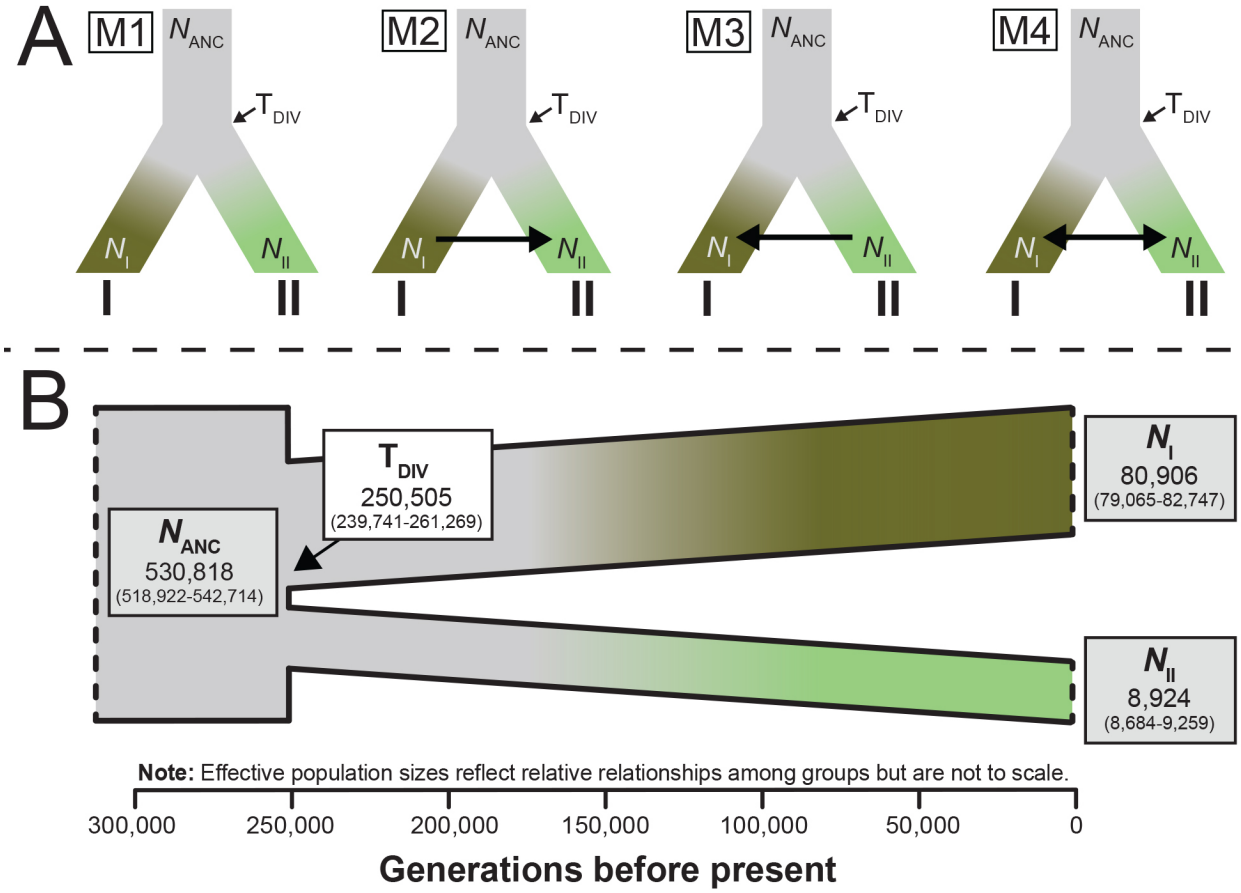

Figure 2. (A) Demographic models tested in this study which included parameters for divergence time (T_{DIV}), effective population sizes [ancestral (N_{ANC}) and for each group (N_I, N_{II})], and migration (arrows) when applicable. (B) The parameterized best-fit demographic model (M1). Numbers in parentheses represent 95% confidence interval around estimates. The white box corresponds to the divergence time (in generations) between groups I and II. Gray boxes correspond to population size parameters.

235 **Table 2.** Above the diagonal: Pair-wise AMOVA F_{ST} values for all populations included in this study.
236 Mean F_{ST} (bottom row) refers to the average pair-wise differentiation for columnar populations versus all
237 others. Below the diagonal (in gray): mean pair-wise shared loci for the fineRADstructure coancestry
238 analysis (see Fig. 1).

	LEA	DAV	TRE	BEA	WIB	JAC	WHM	COM	MAR
LEA	--	0.333	0.634	0.626	0.608	0.586	0.557	0.347	0.344
DAV	74.9	--	0.567	0.557	0.541	0.527	0.515	0.290	0.295
TRE	19.8	20.0	--	0.153	0.207	0.275	0.250	0.656	0.652
BEA	19.0	19.9	127.2	--	0.165	0.250	0.222	0.650	0.645
WIB	20.1	20.5	119.8	123.0	--	0.216	0.201	0.630	0.628
JAC	21.0	20.4	99.7	99.5	104.7	--	0.172	0.604	0.600
WHM	21.6	19.1	90.7	90.9	93.0	107.8	--	0.576	0.572
COM	84.8	86.1	21.8	21.3	23.0	22.9	22.1	--	0.160
MAR	164.5	82.9	21.2	21.3	22.3	21.8	21.4	262.3	--
Mean F_{ST}	0.504	0.453	0.424	0.408	0.399	0.404	0.383	0.489	0.487

239

240 **Results:**

241 *Population genetic and phylogenetic analyses*

242 We generated 343,875,880 reads with an average of 5,828,404 sequences per individual
243 (min. = 446,872 and max. = 40,982,490). Our total RAD data set included 360,534 unique loci.
244 After filtering, we retained 6,019 loci and 10,392 SNPs (mean = 1.73 SNPs per locus). This final
245 data set had genotype calls for ~65% of all SNPs. Nucleotide diversity (π) was highest in the
246 White Mantle and Learnard populations (0.131 and 0.114, respectively) and lowest in the Treaty,
247 Bear, and William Brown populations (0.026-0.056; Table 1). Heterozygosity followed the same
248 pattern as π (Table 1). The inbreeding coefficient (F_{IS}) was highest in the Davidson and Comox
249 populations (0.037 and 0.034, respectively) and lowest in Bear (-0.013) and Treaty (-0.012;
250 Table 1). Mean differentiation (F_{ST}) for all pair-wise comparisons was 0.439. The Learnard
251 population from southern Alaska was, on average, the most differentiated from all others (mean
252 F_{ST} = 0.504) and White Mantle the least differentiated (mean F_{ST} = 0.383; Table 2). We detected
253 no association between genetic and geographic distances in either of the study area-wide Mantel
254 tests (Mantel's *r*, all populations = -0.04, *P* = 0.42; Mantel's *r*, no Vancouver Island populations
255 = 0.15, *P* = 0.36). There was, however, a signature of IBD within group II (Mantel's *r*, group I =
256 0.839, *P* = 0.025) but not within group I (Mantel's *r*, group II = 0.839, *P* = 0.082).

257 Our DAPC analyses supported *K* = 6 as the optimal number of genetic clusters (Fig.
258 3A,C). ADMIXTURE results, however, supported *K* = 7 as the best-fit (Figs. 3B,C) and the
259 SVDQuartets phylogeny largely mirrored both lines of population structure evidence (Fig. 3D).
260 All analyses supported multiple independent genetic clusters of ice worms. Our DAPC and

ADMIXTURE results differed in two ways: (1) the best-fit DAPC result grouped the Treaty, Bear, and William Brown populations into one cluster whereas the best-fit ADMIXTURE result split William Brown into its own cluster. This difference accounted for the $K = 6$ versus $K = 7$ discrepancy between the approaches. (2) While both analyses identified a single individual (MS5) from the Mariner population with genetic assignment to the Learnard cluster, DAPC indicated full assignment of MS5 to the Learnerd cluster whereas Admixture equally assigned it to both the Learnerd and Comox+Mariner clusters (Figs. 1, 3). Finally, because our SNP filtering focused on overarching patterns in the data set, and likely overlooked some degree of population-specific detail, our results are likely conservative estimates of genetic structure in the group.

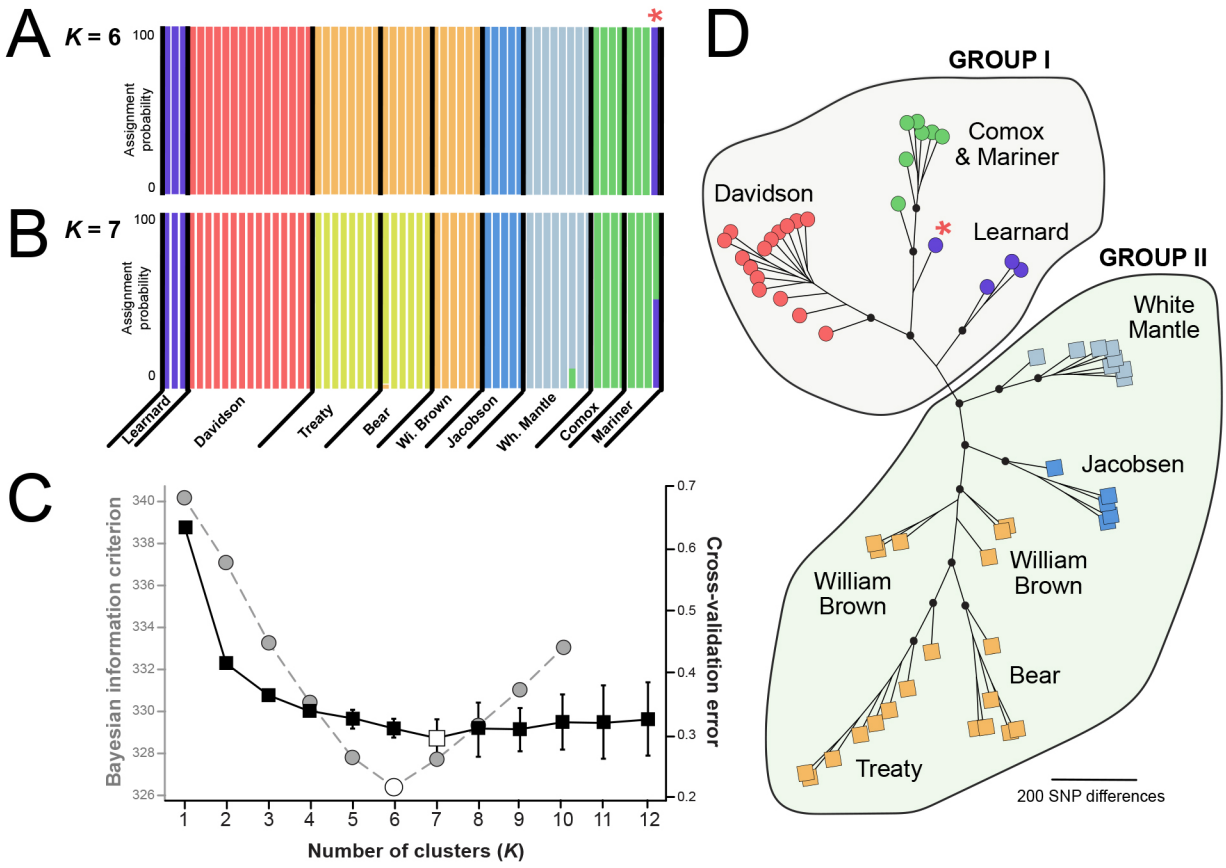

Figure 3. Population genetic structure of ice worms based upon (A) a discriminant analysis of principal components (DAPC) for $K = 6$ and (B) ADMIXTURE results for $K = 7$. (C) Comparisons of support for different values of K for DAPC (Bayesian information criterion, BIC; gray dashed line, left y-axis) and Admixture (cross-validation error, CV; dark line, right y-axis). The best-fit K (white square, ADMIXTURE; white circle, DAPC) corresponds to the K at which CV (ADMIXTURE) or BIC (DAPC) was minimized, respectively. For ADMIXTURE, vertical black bars represent the standard deviation of CV values for each K across 25 replicates. (D) An unrooted phylogeny of ice worms generated with

SVDQuartets. Tip colorations reflect assignments in (A). Dark circles indicate nodes with greater than 95% bootstrap support. Specimens in group I (circles) and II (squares) are denoted with different symbols. Asterisks (A & D) highlight a single specimen, MS5, which showed evidence of shared ancestry across geographically disjunct populations, indicating LDD.

Our fineRADstructure results largely mirrored those from DAPC, identifying the same six genetic clusters. As in the Admixture results, MS5 exhibited evidence for shared ancestry between the Learnard and Comox/Mariner genetic clusters (Fig. 1). On average, MS5 exhibited a ~70/30 split of shared ancestry between Learnard and Comox/Mariner ice worms (Fig. 1). Notably, MS5 also exhibited the highest heterozygosity of any individual in the study (and this result did not stem from outsized coverage, Fig. S3). Our fineRADstructure results also highlighted a primary divergence between two groups of ice worm populations (groups I and II; Fig. 1). This split was corroborated by both SVDQuartets (Fig. 3) and F_{ST} comparisons. Mean pairwise F_{ST} among populations within groups I and II were 0.21 and 0.30, respectively. Between groups, however, mean pairwise F_{ST} was 0.60. Mean distance to the Pacific Ocean also differed by 100.5 km (group I = 73.3 km, group II = 173.8 km; ANOVA, $P < 0.001$).

Demographic modeling

Our tests of demographic models for groups I and II identified a history of divergence without gene flow (model M1) as the best fit to our data (Fig. 2B; Table S1). All other models were rejected with $\Delta AIC \geq 9.1$ (Table S1). The second-best model (M2) included unidirectional gene flow from group I into group II (Fig. 2A), while models M3 and M4 included gene flow from group II into group I, resulting in ΔAIC scores ≥ 118.4. Groups I and II diverged approximately ~250,000 generations ago (Fig. 2B).

Discussion:

Historical and contemporary factors, both biotic and abiotic, interact to shape the present-day genetic structure of populations and species. Disentangling their varied contributions can be difficult, however, particularly when evolutionary histories are muddled by unexpected events (e.g., LDD of an organism with limited potential for active dispersal). The modern population genomic toolkit provides historically unprecedented power to resolve biogeographical complexity by allowing more quantitative perspectives of relatedness and greatly improved resolution of genetic independence or similarity [47]. In this study, we used a population

genomic data set to refine understanding of the evolutionary history of the extremophile, glacier-obligate ice worm, *M. solifugus*. Our results provide a clear genomic perspective of LDD, showing unequivocally that migration has occurred between southern Alaska and the glaciers of Vancouver Island ~1,900 km to the south and across the Pacific Ocean. We also provide an independent line of biological evidence in support of the geological hypothesis that ice ridges formed along the crest of the Coast Mountains during the Pleistocene [29].

Ice worm biogeography and LDD

The recent biogeographic history of ice worms appears to have been shaped by three main factors: i) ice sheet dynamics, ii) mountaintop isolation from conspecifics following the retreat of Pleistocene ice into higher elevations, and iii) LDD.

i. Evidence for the first, overarching factor that has defined the recent evolution of ice worms lies in our overwhelming support for the deepest divergence between two groups (I and II) which fall largely on either side of the Coast Mountains in western North America. During the Pleistocene (~2.5 million–11,700 years ago), western Canada was repeatedly covered by a continental ice sheet [29]. Ice was generated in the high peaks of the Coast Mountains and subsequently flowed west to the Pacific Ocean and east to the interior of British Columbia from the crest of the range [29, 48, 49]. This potential western-eastern divergence in ice worms aligns with this divide, suggesting that each group diverged in allopatry from their conspecifics. The timeline of this divergence, however, is unclear. Our results suggest ~250,000 generations have passed since the initial divergence but with no knowledge of generation times for ice worms, nor related species at very low temperatures, we cannot provide a reliable estimate of years before present. If ice worms develop rapidly, with multiple generations per year (e.g., ≥ 3), the divergence may have occurred less than 100,000 years ago. However, if ice worms develop slowly (e.g., one generation for every five years)—the split may have occurred over a million years ago. Either way, the split appears more recent than estimates from mitochondrial DNA of ~1.7 million years ago [25]. At its maximum, the Pleistocene ice sheet in the region was a ~2,000-3,000 m high convex dish with gentle interior slopes that steepened at its periphery [29, 50].

Given the sensitivity of ice worms to extreme cold [25], populations likely only persisted on the ice sheet margins, as supported by their present-day occurrences on the lower flanks of higher elevation, low-latitude glaciers. It is possible that, as suggested previously [25, 27], the Boundary Ranges (the most northern subrange of the Coast Mountains), are actually the biogeographic barrier that drove the deep divergence in ice worms described in this study. However, without more fine-scale population genomic sampling on both sides of the Coast Mountains (including the Boundary Ranges), this nuance of ice worm biogeography will remain unclear (see Fig. S4 and additional discussion in Supplementary Materials). In the same vein, the strong genetic similarity of White Mantle to populations east of the proposed ice ridge (Fig. 1), despite falling on its western side, indicates that either the ice ridge that precipitated divergence among the two groups actually formed more to the west than previously thought [29], the White Mantle population has migrated west since divergence from conspecifics, the ice ridge itself was not a barrier driving differentiation (as discussed above), or some combination of all three.

ii. In western North American, the formation of the Cordilleran ice sheet was seeded by alpine glaciers [29]. While the specific dynamics of deglaciation on valley and drainage scales are unknown, a safe assumption is that glaciers retreated from valleys into higher elevations, likely with ice worm populations in tow. Increasing mountaintop isolation and subsequent genetic drift likely precipitated the more recent differentiation within groups I and II since their initial split. Evidence for IBD within group II supports this hypothesis. The lack of support for IBD in group I, however, may reflect reality, the reduced power of a smaller sample size, or LDD maintaining connections over larger spatial scales than a purely IBD model would predict.

iii. Despite limited sampling, we were able to identify one instance of recent LDD among geographically disparate ice worm populations. Indeed, with divided ancestry between the Mariner/Comox and Learnard clusters, one specimen (MS5), is likely the progeny of recent migration between the two. This indicates that LDD is both ongoing in ice worms and perhaps not particularly rare. The most plausible mechanism for ice worm LDD is passive dispersal of mucous-coated ice worm

373 cocoons sticking to the feet, beaks, or feathers of southward migrating birds [12].
374 Several passerines (e.g., gray-crowned rosy finches, *Leucosticte tephrocotis*) have
375 been observed feeding on ice worms [51, 52; S.H., personal observation] and the
376 presence of an ice worm clitellum [53] indicates that ice worms, like other
377 *Mesenchytraeus* species, reproduce by egg-laden cocoons [25]. The seemingly
378 exclusive northwest-to-southeast pattern of ice worm LDD also has temporal support
379 from bird migratory behavior. Late autumn ice worm reproduction [at the end of the
380 productive season on mountain glaciers, 54] likely occurs in concert with southward-
381 migrating birds stopping to feed on glaciers free of seasonal snow. In contrast,
382 returning spring migrants pass over the same glaciers when seasonal snowfall still
383 covers overwintering ice worms [25], limiting the potential for LDD in the reverse
384 direction.
385
386 One question remains, however, if birds are precipitating LDD in ice worms, why has it
387 only been observed for populations west of the Coast Mountains? This curiosity ties in to an
388 important question in North American biogeography: to what extent have ice sheets driven
389 present-day patterns of speciation and genetic differentiation among fauna of the northwest? For
390 ice worms, we hypothesize that populations comprising groups I and II have accumulated some
391 degree of reproductive isolation. This inference is supported by our demographic modeling
392 which rejected models involving gene flow from group II into group I. While this may be at least
393 partly linked to directionality in migration (e.g., bird movements), the strength in which these
394 models were rejected suggests that a zygotic barrier may be limiting inter-group migrants from
395 leaving a genomic signature of gene flow. It is also possible, and perhaps likely, that patterns of
396 LDD in ice worms is driven by vector migration patterns. For instance, *L. tephrocotis*, like other
397 songbirds [55], may preferentially follow coastlines during migration. However, until more is
398 known about the specific interactions of ice worms with various bird species, and by proxy, their
399 potential to act as LDD vectors, relating bird migrations to ice worm distributions and
400 demography will remain difficult. Beyond ice worms, ice sheets have been implicated as a key
401 driver of speciation in boreal birds [56] and phylogeographic structure of many taxa, from
402 nematodes to gray wolves [17, 30, 57-60], and our results clearly support these broader
403 implications for biodiversity accumulation and maintenance in North America.

404

405 *Conclusions*

406 In this study, we leveraged population genomic data to unravel the complex evolutionary
407 history of the North American ice worm, *M. solifugus.* Our results add new clarity to previous
408 perspectives on ice worm biogeography while also lending genomic support to the existence of
409 contemporary, likely passerine-mediated LDD in the group. We described two genetic grouping
410 (I and II) which are described with respect to the crest of the Coast Mountains, where ice ridges
411 formed during the Pleistocene [29]. While the phylogenetic data used in this study (i.e., the lack
412 of an outgroup) preclude us from diagnosing groups I and II as monophyletic, given the results
413 of previous studies [24, 25, 27], we predict that future efforts will diagnose them as such,
414 perhaps even representing two nascent species. Finally, our genomic data lend support to the
415 glaciological record in the region, adding a biological line of evidence to a postulated key north-
416 south dividing line along the crest of the Coast Mountains where ice ridges likely formed during
417 the Pleistocene and repeatedly propagated ice flow to the east and west [29]. This potential for
418 genomics to inform the geological record is intriguing and ice worms, as a rare glacier-obligate
419 macroinvertebrate, may be an ideal taxon for similar studies in the future.

420

421 **Acknowledgements:** We thank Joe Giersch for map-making assistance and the University of
422 Kentucky Center for Computational Sciences as well as the Washington State University Center
423 for Institutional Research Computing for high-performance computational resources. This
424 research was partially funded by NSF award #IOS-082050 to D.H.S.

425

426 **Competing interests:** We have no competing interests.

427

428 **Author contributions:** S.H., D.H.S., L.M.T, and D.W.W. conceived of and funded the study.
429 S.H., D.H.S., S.A.L, R.K.B., and D.W.W. collected the data. S.H. and J.L.K. analyzed the data
430 and wrote the manuscript with input from D.H.S., R.K.B., and D.W.W. All authors approved the
431 final version.

432

Data accessibility: Raw sequence data for this study has been submitted to GenBank under BioProject #PRJNA479335 and code to reproduce the analyses is deposited on GitHub (https://github.com/scotthotaling/ice_worm_ddRAD).

Funding: Field sampling for this study was supported by NSF award #IOS-0820505 to D.H.S.

References:

1. Darwin C. 1859 *On the origin of species*, Routledge.

2. Martin C.H., Turner B.J. 2018 Long-distance dispersal over land by fishes: extremely rare ecological events become probable over millennial timescales. *Proc R Soc B* **285**(1881), 20172436.

3. Trakhtenbrot A., Nathan R., Perry G., Richardson D.M. 2005 The importance of long‐distance dispersal in biodiversity conservation. *Diversity and Distributions* **11**(2), 173-181.

4. Helmuth B., Veit R., Holberton R. 1994 Long-distance dispersal of a subantarctic brooding bivalve (Gaimardia trapesina) by kelp-rafting. *Marine biology* **120**(3), 421-426.

5. Viana D.S., Gangoso L., Bouten W., Figuerola J. 2016 Overseas seed dispersal by migratory birds. *Proc R Soc B* **283**(1822), 20152406.

6. Gittenberger E., Groenenberg D.S., Kokshoorn B., Preece R.C. 2006 Biogeography: molecular trails from hitch-hiking snails. *Nature* **439**(7075), 409.

7. Miura O., Torchin M.E., Bermingham E., Jacobs D.K., Hechinger R.F. 2012 Flying shells: historical dispersal of marine snails across Central America. *Proc R Soc B* **279**(1731), 1061-1067.

8. Pearce T.A., Mulvihill R.S., Porter K.A. 2012 Land slugs (Gastropoda: Pulmonata) on birds demonstrate dispersal potential. *Nautilus* **126**, 38-40.

9. Suetsugu K., Funaki S., Takahashi A., Ito K., Yokoyama T. 2018 Potential role of bird predation in the dispersal of otherwise flightless stick insects. *Ecology* **99**(6), 1504-1506.

10. Campos-Arceiz A., Larrinaga A.R., Weerasinghe U.R., Takatsuki S., Pastorini J., Leimgruber P., Fernando P., Santamaría L. 2008 Behavior rather than diet mediates seasonal differences in seed dispersal by Asian elephants. *Ecology* **89**(10), 2684-2691.

11. Anderson J.T., Nuttle T., Rojas J.S.S., Pendergast T.H., Flecker A.S. 2011 Extremely long-distance seed dispersal by an overfished Amazonian frugivore. *Proceedings of the Royal Society of London B: Biological Sciences* **278**(1723), 3329-3335.

12. Viana D.S., Santamaría L., Figuerola J. 2016 Migratory birds as global dispersal vectors. *Trends in ecology & evolution* **31**(10), 763-775.

13. Somveille M., Manica A., Butchart S.H., Rodrigues A.S. 2013 Mapping global diversity patterns for migratory birds. *PloS one* **8**(8), e70907.

14. Simpson G.G. 1940 Mammals and land bridges. *Journal of the Washington Academy of Sciences* **30**(4), 137-163.

15. Slatkin M. 1987 Gene flow and the geographic structure of natural populations. *Science* **236**(4803), 787-792.

16. Hohenlohe P.A., Day M.D., Amish S.J., Miller M.R., Kamps-Hughes N., Boyer M.C., Muhlfeld C.C., Allendorf F.W., Johnson E.A., Luikart G. 2013 Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular ecology* **22**(11), 3002-3013.

17. Hotaling S., Muhlfeld C.C., Giersch J.J., Ali O.A., Jordan S., Miller M.R., Luikart G., Weisrock D.W. 2018 Demographic modelling reveals a history of divergence with gene flow for a glacially tied stonefly in a changing post‐Pleistocene landscape. *Journal of Biogeography* **45**(2), 304-317.

18. Jordan S., Giersch J.J., Muhlfeld C.C., Hotaling S., Fanning L., Tappenbeck T.H., Luikart G. 2016 Loss of genetic diversity and increased subdivision in an endemic alpine stonefly threatened by climate change. *PloS one* **11**(6), e0157386.

19. Antonelli A., Kissling W.D., Flantua S.G., Bermúdez M.A., Mulch A., Muellner-Riehl A.N., Kreft H., Linder H.P., Badgley C., Fjeldså J. 2018 Geological and climatic influences on mountain biodiversity. *Nature Geoscience* **11**(10), 718.

20. Taberlet P., Fumagalli L., Wust-Saucy A., Cosson J. 1998 Comparative phylogeography and postglacial colonization routes in Europe. *Molecular ecology* **7**(4), 453-464.

21. Giersch J.J., Hotaling S., Kovach R.P., Jones L.A., Muhlfeld C.C. 2017 Climate‐induced glacier and snow loss imperils alpine stream insects. *Global change biology* **23**(7), 2577-2589.

22. Emery C. 1898 Diagnosi di un nuovi genere e nuova specie di annelidi della famiglia degli Enchytraeidae. *Atti R Accademei Lincei* **5**, 100-111.

23. Liang Y., Hsu C., Chang T. 1979 A new genus and species of Enchytraeidae from Tibet. *Acta Zootaxonomica Sinica* **4**, 312-317.

24. Hartzell P.L., Nghiem J.V., Richio K.J., Shain D.H. 2005 Distribution and phylogeny of glacier ice worms (Mesenchytraeus solifugus and Mesenchytraeus solifugus rainierensis). *Canadian Journal of Zoology* **83**(9), 1206-1213.

25. Dial R.J., Becker M., Hope A.G., Dial C.R., Thomas J., Slobodenko K.A., Golden T.S., Shain D.H. 2016 The role of temperature in the distribution of the glacier ice worm, Mesenchytraeus solifugus (Annelida: Oligochaeta: Enchytraeidae). *Arctic, Antarctic, and Alpine Research* **48**(1), 199-211.

503 26. Lang S.A., Saglam N., Kawash J., Shain D.H. 2017 Punctuated invasion of water, ice,
504 snow and terrestrial ecozones by segmented worms (Oligochaeta: Enchytraeidae:
505 Mesenchytraeus). *Proc R Soc B* **284**(1864), 20171081.

506 27. Dial C.R., Dial R.J., Saunders R., Lang S.A., Lee B., Wimberger P., Dinapoli M.S.,
507 Egiazarov A.S., Gipple S.L., Maghirang M.R., et al. 2012 Historical biogeography of the North
508 American glacier ice worm, Mesenchytraeus solifugus (Annelida: Oligochaeta: Enchytraeidae).
509 *Mol Phylogenet Evol* **63**(3), 577-584.

510 28. Figuerola J., Green A.J., Michot T.C. 2004 Invertebrate eggs can fly: evidence of
511 waterfowl-mediated gene flow in aquatic invertebrates. *The American naturalist* **165**(2), 274-
512 280.

513 29. Clague J.J., Ward B. 2011 Pleistocene glaciation of British Columbia. In *Developments*
514 *in Quaternary Sciences* (pp. 563-573, Elsevier.

515 30. Shafer A.B., Cullingham C.I., Cote S.D., Coltman D.W. 2010 Of glaciers and refugia: a
516 decade of study sheds new light on the phylogeography of northwestern North America.
517 *Molecular ecology* **19**(21), 4589-4621.

518 31. Peterson B.K., Weber J.N., Kay E.H., Fisher H.S., Hoekstra H.E. 2012 Double digest
519 RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-
520 model species. *PloS one* **7**(5), e37135. (doi:10.1371/journal.pone.0037135).

521 32. Burford Reiskind M., Coyle K., Daniels H., Labadie P., Reiskind M., Roberts N., Roberts
522 R., Schaff J., Vargo E. 2016 Development of a universal double‐digest RAD sequencing
523 approach for a group of nonmodel, ecologically and economically important insect and fish taxa.
524 *Molecular ecology resources* **16**(6), 1303-1314.

525 33. Catchen J., Hohenlohe P.A., Bassham S., Amores A., Cresko W.A. 2013 Stacks: an
526 analysis tool set for population genomics. *Molecular ecology* **22**(11), 3124-3140.
527 (doi:10.1111/mec.12354).

528 34. Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., Bender D., Maller J.,
529 Sklar P., De Bakker P.I., Daly M.J.J.T.A.j.o.h.g. 2007 PLINK: a tool set for whole-genome
530 association and population-based linkage analyses. **81**(3), 559-575.

531 35. Weir B.S., Cockerham C. 1996 Genetic data analysis II: Methods for discrete population
532 genetic data. Sinauer Assoc. *Inc, Sunderland, MA, USA*.

533 36. Wright S. 1943 Isolation by distance. *Genetics* **28**(2), 114-138.

534 37. Meirmans P.G., Van Tienderen P.H. 2004 GENOTYPE and GENODIVE: two programs
535 for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Resources* **4**(4),
536 792-794.

537 38. Alexander D.H., Novembre J., Lange K. 2009 Fast model-based estimation of ancestry in
538 unrelated individuals. *Genome Res* **19**(9), 1655-1664. (doi:10.1101/gr.094052.109).

539 39. Jombart T., Devillard S., Balloux F. 2010 Discriminant analysis of principal components:
540 a new method for the analysis of genetically structured populations. *BMC genetics* **11**(1), 94.

541 40. Malinsky M., Trucchi E., Lawson D.J., Falush D. 2018 RADpainter and
542 fineRADstructure: population inference from RADseq data. *Molecular biology and evolution*
543 **35**(5), 1284-1290.

544 41. Chifman J., Kubatko L. 2014 Quartet inference from SNP data under the coalescent
545 model. *Bioinformatics* **30**(23), 3317-3324.

546 42. Swofford D.L. 1998 Phylogenetic analysis using parsimony.

547 43. Excoffier L., Dupanloup I., Huerta-Sánchez E., Sousa V.C., Foll M.J.P.g. 2013 Robust
548 demographic inference from genomic and SNP data. **9**(10), e1003905.

549 44. Keightley P.D., Trivedi U., Thomson M., Oliver F., Kumar S., Blaxter M.L.J.G.r. 2009
550 Analysis of the genome sequences of three Drosophila melanogaster spontaneous mutation
551 accumulation lines. **19**(7), 1195-1201.

552 45. Reynoldson T. 1943 A comparative account of the life cycles of Lumbricillus lineatus
553 Mull. and Enchytraeus albidus Henle in relation to temperature. *J Annals of Applied Biology*
554 **30**(1), 60-66.

555 46. Standen V. 1984 Production and diversity of enchytraeids, earthworms and plants in
556 fertilized hay meadow plots. *Journal of Applied Ecology*, 293-312.

557 47. Allendorf F.W., Hohenlohe P.A., Luikart G. 2010 Genomics and the future of
558 conservation genetics. *Nat Rev Genet* **11**(10), 697-709. (doi:10.1038/nrg2844).

559 48. Margold M., Jansson K.N., Kleman J., Stroeven A.P., Clague J.J. 2013 Retreat pattern of
560 the Cordilleran Ice Sheet in central British Columbia at the end of the last glaciation
561 reconstructed from glacial meltwater landforms. *Boreas* **42**(4), 830-847.

562 49. Stumpf A.J., Broster B.E., Levson V.M. 2000 Multiphase flow of the late Wisconsinan
563 Cordilleran ice sheet in western Canada. *Geological Society of America Bulletin* **112**(12), 1850-
564 1863.

565 50. Wilson J., Falconer G., Mathews W., Prest V. 1958 Glacial map of Canada. *Geological*
566 *Association of Canada, Toronto, ON.*

567 51. Riedel J.L., Wilson S., Baccus W., Larrabee M., Fudge T., Fountain A. 2015 Glacier
568 status and contribution to streamflow in the Olympic Mountains, Washington, USA. *Journal of*
569 *Glaciology* **61**(225), 8-16.

570 52. Rosvold J. 2016 Perennial ice and snow-covered land as important ecosystems for birds
571 and mammals. *Journal of biogeography* **43**(1), 3-12.

572 53. Shain D.H., Carter M.R., Murray K.P., Maleski K.A., Smith N.R., McBride T.R.,
573 Michalewicz L.A., Saidel W.M. 2000 Morphologic characterization of the ice worm
574 Mesenchytraeus solifugus. *Journal of Morphology* **246**(3), 192-197.

575 54. Hotaling S., Hood E., Hamilton T.L. 2017 Microbial ecology of mountain glacier
576 ecosystems: biodiversity, ecological connections and implications of a warming climate.
577 *Environmental microbiology* **19**(8), 2935-2948.

578 55. Ruegg K.C., Smith T.B. 2002 Not as the crow flies: a historical explanation for circuitous
579 migration in Swainson's thrush (Catharus ustulatus). *Proceedings of the Royal Society of London*
580 *B: Biological Sciences* **269**(1498), 1375-1381.

581 56. Weir J.T., Schluter D. 2004 Ice sheets promote speciation in boreal birds. *Proceedings of*
582 *the Royal Society of London B: Biological Sciences* **271**(1551), 1881-1887.

583 57. Hotaling S., Giersch J.J., Finn D.S., Tronstad L.M., Jordan S., Serpa L.E., Call R.G.,
584 Muhlfeld C.C., Weisrock D.W. 2019 Congruent population genetic structure but differing depths
585 of divergence for three alpine stoneflies with similar ecology, geographic distributions, and
586 climate change threats. *Freshwater Biology* **64**, 335-347.

587 58. Muñoz A.R., Marquez A.L., Real R. 2013 Updating known distribution models for
588 forecasting climate change impact on endangered species. *PloS one* **8**(6), e65462.
589 (doi:10.1371/journal.pone.0065462).

590 59. Burg T.M., Gaston A.J., Winker K., Friesen V.L. 2006 Effects of Pleistocene glaciations
591 on population structure of North American chestnut‐backed chickadees. *Molecular ecology*
592 **15**(9), 2409-2419.

593 60. Koehler A.V., Hoberg E.P., Dokuchaev N.E., Tranbenkova N.A., Whitman J.S.,
594 Nagorsen D.W., Cook J.A. 2009 Phylogeography of a Holarctic nematode, Soboliphyme
595 baturini, among mustelids: climate change, episodic colonization, and diversification in a
596 complex host–parasite system. *Biological Journal of the Linnean Society* **96**(3), 651-663.
597