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

[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 glacier-obligate annelids that inhabit coastal North American 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 evidence for contemporary migration across large (>1,500 km) areas of unsuitable habitat. In this study, we collected genome-scale sequence data for most of the known ice worm range. We found support for a deep divergence between populations along the Pacific Coast and the inland flanks of the Coast Mountains (mean $F_{ST} = 0.60$) as well as support for LDD 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 < 10 km. For more extreme LDD events (e.g., > 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 internally 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 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 and species throughout Earth's history [17-20]. These ice sheet dynamics typically affect organisms as ice traverses the landscape, separating and reconnecting populations as it expands and contracts. However, some species are directly tied to ice sheets [e.g., the meltwater stonefly, 17] and are therefore much more susceptible to ice sheet influence on their evolutionary trajectories. Perhaps no species is more directly tied to ice sheets than the glacier ice worms, *Mesenchytraeus solifugus* in North America [21] and *Sinenchytraeus glacialis* in Tibet [22]. The geographic range of *M. solifugus* (hereafter "ice worm") follows a coastal arc from the Chugach Mountains in southeast Alaska to the Cascade Volcanoes of Washington and Oregon [24]. Ice worms can also not tolerate temperatures more than roughly ±7 °C from freezing and requires glacier ice for survival and reproduction [23]. With such unique ecology and physiology, and a dispersal-limited life history, the evolutionary history of ice worms since diverging from conspecifics [25, 26] should be relatively simple, with gene flow occurring during glacial periods and isolation (paired with genetic drift) driving divergence among mountaintop-isolated populations during interglacial periods. Natural systems, however, are often more complex than expected and indeed, the evolutionary history of ice worms challenges general expectations of gene flow and evolutionary dynamics in mountain systems.

Previous genetic studies based on one or two genetic markers identified three ice worm lineages: a "northern" clade in southern Alaska, a "central" clade in southeast Alaska and northern British Columbia (BC), and a "southern" clade ranging over much of BC to southern Oregon [23-25]. Surprisingly, phylogenetic evidence supported the northern and southern lineages as being most closely related to one another despite the central clade separating them geographically. The most curious aspect of ice worm biogeography, however, has been the repeated discovery of closely related ice worms on glaciers several hundred to thousands of kilometers south of their closest genetic relatives [23, 25; P. Wimberger, unpublished data]. These disjunct northern ice worms co-occurred with, but appeared genetically distinct from, their conspecifics (either central or southern clade ice worms) on the same glaciers. Dial *et al.* [23] laid out three possible explanations for this pattern: wind transport, passerine-mediated dispersal, or a more extensive previous range of the northern clade. While wind transport seems unlikely,

the potential for passerine-mediated dispersal is reasonable, particularly in light of other examples of bird-mediated LDD [see 7, 27]. The third scenario, a more extensive distribution of the northern clade with holdover lineages inhabiting the same glacier as more recent colonizers could indeed result in more than one distinct lineage on the same glacier. However, if this recent colonization occurred at the Last Glacial Maximum when ice sheet extent last peaked in the region, recently colonizing lineages should not remain closely related to the source without at least some degree of contemporary gene flow.

In this study, we leveraged a modern population genomic toolkit to add new perspective to the age-old challenge of identifying LDD in wild populations. We also provide new insight into how multiple factors can interact to shape the evolutionary history of species. We hypothesized that the biogeographic history of ice worms stemmed from a confluence of three main factors: extreme LDD, glacier dynamics, and mountaintop isolation. To test this hypothesis, we generated a genome-wide single nucleotide polymorphism (SNP) data set to answer three more-specific questions: (1) How do the clades previously diagnosed from a small number of markers hold up to genome-wide scrutiny? (2) What, if any, genomic evidence exists for LDD in ice worms? (3) How do the evolutionary relationships among ice worm populations and genetic clusters align with glacial history in the region [e.g., 28]? Beyond a refined view of ice worm evolution, our study confirms that LDD does occur in ice worms, providing an example of LDD in an annelid and a rare population genomic exploration of the process. Moreover, while considerable evidence details the existence of refugia in the Pacific Northwest (PNW) during the Pleistocene [29], few studies have explored how ice sheet dynamics influenced the evolutionary history of species directly tied to them [e.g., 17]. Our results reveal the profound impact that a putative ice ridge that formed along the crest of the Coast Mountains [28] may have had on ice worm evolution, possibly precipitating an ongoing speciation event. Broadly, our findings highlight the power for population genomics to capture contemporary evidence of LDD while also providing novel evidence for reconstructing the glacial history of a region.

Table 1. Sampling information and summary statistics for all ice worm populations included in this study. Abbreviations: *n* = sample size, π = nucleotide diversity, Het = heterozygosity, F_{IS} = inbreeding coefficient, AK = Alaska, BC = British Columbia, VI = Vancouver Island. π, Het, and F_{IS} were calculated for variable sites only.

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

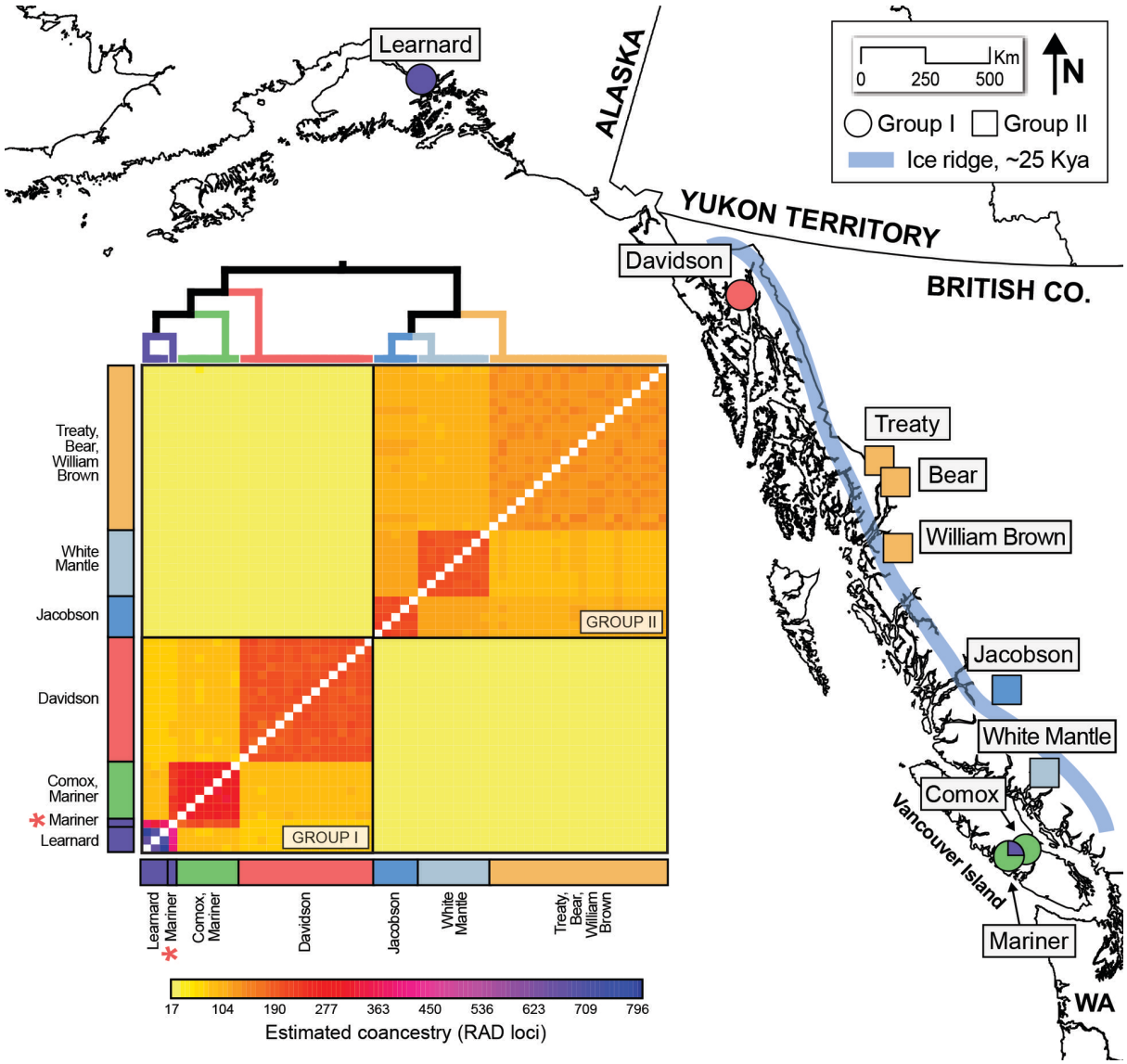
Methods:

Sample collection, library preparation, and SNP calling

During the summer of 2009, ice worms were collected from nine glaciers across most of their range (Figs. 1, S1; Table 1). Samples were stored in > 80% EtOH until DNA was extracted from 59 ice worms using a Qiagen DNEasy Blood and Tissue Kit. Double-digest restriction-site associated DNA sequencing libraries were prepared following Peterson *et al.* [30] with restriction enzymes EcoRI and NlaIII. During library preparation, samples were divided into two groups and each sample was assigned a unique, variable-length barcode [31] which was incorporated during adapter ligation. Size selection for a 350 bp ± 35 bp window was performed with a Pippen Prep (Sage Science), and both sample groups were subsequently amplified using PCR primers containing a group-specific barcode. The 59-sample library was sequenced on one lane of an Illumina HiSeq4000 at the University of Illinois High-Throughput Sequencing and Genotyping Unit with single-end, 100 bp chemistry.

Raw reads were demultiplexed, quality-filtered, and RAD loci were assembled *de novo* using the *process_radtags* and *denovo_map* functions of the Stacks v1.46 pipeline [32]. We allowed a maximum distance between stacks of 2 and a minimum read depth of 10. Next, we applied a stringent filtering scheme to identify high-confidence SNPs that were shared among many individuals. We only included SNPs if they were present in ≥ 5 populations, genotyped in ≥ 50% of individuals per population, and were in Hardy-Weinberg equilibrium with a minor allele frequency of ≥ 0.025 overall. We further restricted analyses to one random SNP per locus for all analyses except fineRADstructure (see below). All post-Stacks filtering steps were

146 performed in PLINK v1.07 [33] and the commands used in this study are provided on GitHub
147 (https://github.com/scotthotaling/ice_worm_ddRAD).



148
149 **Figure 1.** Ice worm populations sampled for this study. Color-coding reflects the results of a
150 fineRADstructure coancestry analysis. Group I (circles) and II (square) populations were generally
151 defined by their presence to the east or west of a putative ice ridge that formed during the Pleistocene
152 ~25,000 years ago [28] as well as their distance to the Pacific Ocean. The deep divergence between
153 groups I and II is clearly evident along with more recent differentiation within each group. One individual
154 from the Mariner Glacier (asterisk) is admixed between the Mariner/Comox (Vancouver Island) and
155 Learnard (southern Alaska) clusters, indicating recent LDD.
156

157 *Population genetic and phylogenetic analyses*

158 For each population, we calculated nucleotide diversity (π), heterozygosity (Het), and the
159 inbreeding coefficient (F_{IS}). We also calculated a pair-wise AMOVA F_{ST} for all population

160 combinations [34] – using the *populations* module in Stacks. To test for a signature of isolation-
161 by-distance [IBD, 35], we estimated Euclidean distances among sites with Google Earth and
162 tested the correlation between geographic distance and F_{ST} with four Mantel tests performed in
163 GenoDive v2.0b27 [36]. The first Mantel test included all nine populations and the second
164 excluded both Vancouver Island (VI) populations from the Mariner and Comox Glaciers to
165 assess whether unique histories for those populations were significantly altering results. The
166 third and fourth Mantel tests focused on signatures of IBD within groups "I" and "II" (see
167 Results). We also measured Euclidean distances for each population to the Pacific Ocean in
168 Google Earth Pro as an additional spatial comparison of groups I and II. We determined if mean
169 distances to the Pacific Ocean differed between the two groups with a one-way ANOVA.
170 Population structure was inferred in two ways: a maximum likelihood-based method
171 using ADMIXTURE 1.3.0 [37] and a discriminant analysis of principal components (DAPC)
172 with the R package *adegenet* [38]. ADMIXTURE analyses were performed with default settings,
173 a range of clusters (*K*) from 1-12, and 25 replicates per *K* with the current time as the random
174 seed. The cross-validation (CV) error for each *K* was plotted to identify the best-fit *K* (minimized
175 CV across the mean of all replicates for each *K*). After identifying the best-fit *K*, we considered
176 the replicate that minimized CV across all 25 replicates for all *K*'s to be the best-fit solution
177 overall. However, because all runs did not converge on the same result, we also inspected best-fit
178 solutions for other replicates of $K = 7$ (the best-fit *K* overall) to clarify the distribution of best-fit
179 solutions. For DAPC, we first used the *find.clusters* function to identify the optimal *K* [i.e., the *K*
180 with the lowest Bayesian Information Criterion (BIC)]. Next, to avoid over-fitting of the model,
181 we retained the appropriate number of principal components (PCs) according to the α-score [PCs
182 retained = 6, Fig. S2; 38]. We performed a final DAPC analysis using the best-fit *K* and optimal
183 number of PCs identified in the previous two steps.
184 We extended our population structure analyses to infer both shared ancestry and
185 phylogenetic relationships in two ways: a nearest neighbor haplotype approach to infer
186 coancestry with fineRADstructure [39] and phylogenetic relationships inferred from singular
187 value decomposition estimates for quartets of tips using SVDQuartets [40] as implemented in
188 PAUP* v4.0a159 [41]. For fineRADstructure, we used 100,000 burn-in iterations followed by
189 100,000 iterations sampled every 1,000 steps for the Markov chain Monte Carlo clustering
190 algorithm. Next, we used 10,000 iterations of the tree-building algorithm to assess genetic

relationships among clusters. Since fineRADstructure is a haplotype-based approach, analyses were performed using all SNPs for a given RAD locus (i.e., a haplotype) rather than randomly selected single SNPs per locus. For SVDQuartets, we performed exhaustive sampling of all possible quartets (every combination of four tips) and branch support was estimated with 100 nonparametric bootstrap replicates.

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

Results:

We generated 343,875,880 reads with an average of 5,828,404 sequences per individual (min. = 446,872 and max. = 40,982,490). Our total RAD data set included 360,534 unique loci. After filtering, we retained 6,019 loci and 10,392 SNPs (mean = 1.73 SNPs per locus). This final data set had genotype calls for ~65% of all SNPs. Nucleotide diversity (π) was highest in the White Mantle and Learnard populations (0.131 and 0.114, respectively) and lowest in the Treaty, Bear, and William Brown populations (0.026-0.056; Table 1). Heterozygosity followed the same pattern as π (Table 1). The inbreeding coefficient (F_{IS}) was highest in the Davidson and Comox populations (0.037 and 0.034, respectively) and lowest in Bear (-0.013) and Treaty (-0.012; Table 1). Mean differentiation (F_{ST}) for all pairwise comparisons was 0.439. The Learnard population from southern Alaska was, on average, the most differentiated from all others (mean F_{ST} = 0.504) and White Mantle was the least differentiated (mean F_{ST} = 0.383; Table 2). We detected no association between genetic and geographic distances in either study area-wide Mantel tests (Mantel's *r*, all populations = -0.04, *P* = 0.42; Mantel's *r*, no VI populations = 0.15,

215 $P = 0.36$). There was, however, a signature of IBD within group II (Mantel's r, group I = 0.839,
216 $P = 0.025$) but not within group I (Mantel's r, group II = 0.839, $P = 0.082$).
217 Our DAPC analyses supported $K = 6$ as the optimal number of genetic clusters (Figs. 2A,C). For
218 ADMIXTURE, our results supported $K = 7$ as the best-fit to the data (Figs. 2B,C) and the
219 SVDQuartets phylogeny largely mirrored both lines of population structure evidence (Fig. 2D).
220 All analyses supported multiple independent genetic clusters of ice worms, many more than three
221 clades identified in previous studies [e.g., 25]. Our DAPC and ADMIXTURE results differed in
222 two ways: (1) DAPC analyses grouped the Treaty, Bear, and William Brown populations into
223 one cluster (as did SVDQuartets) whereas ADMIXTURE split William Brown into its own
224 cluster, and this accounted for the $K = 6$ versus $K = 7$ discrepancy between the two approaches.
225 (2) ADMIXTURE also split Learnard (southern Alaska) and the two Vancouver Island
226 populations – Comox and Mariner – into two clusters with one sample, MS5 (red asterisks in
227 Fig. 2), almost evenly split in terms of ancestry between the two. Conversely, DAPC split
228 Learnard and Comox+Mariner into two clusters with MS5 completely assigned to the Learnard
229 cluster, despite MS5 being sampled from the Mariner glacier far to the south (Figs. 1,2). Aside
230 from the best-fit solutions, we also observed a second common ADMIXTURE solution for $K = 7$
231 which differed from the best-fit solution in one way; rather than splitting Treaty+Bear and
232 William Brown into two clusters (Fig. 2B), Bear specimens were almost evenly admixed
233 between Treaty and William Brown (Fig. S3), its two closest neighbors to the north and south,
234 respectively. Finally, because our SNP filtering focused on overarching patterns in the data set,
235 and likely overlooked some degree of population-specific detail, our results are likely
236 conservative estimates of genetic structure in the group.

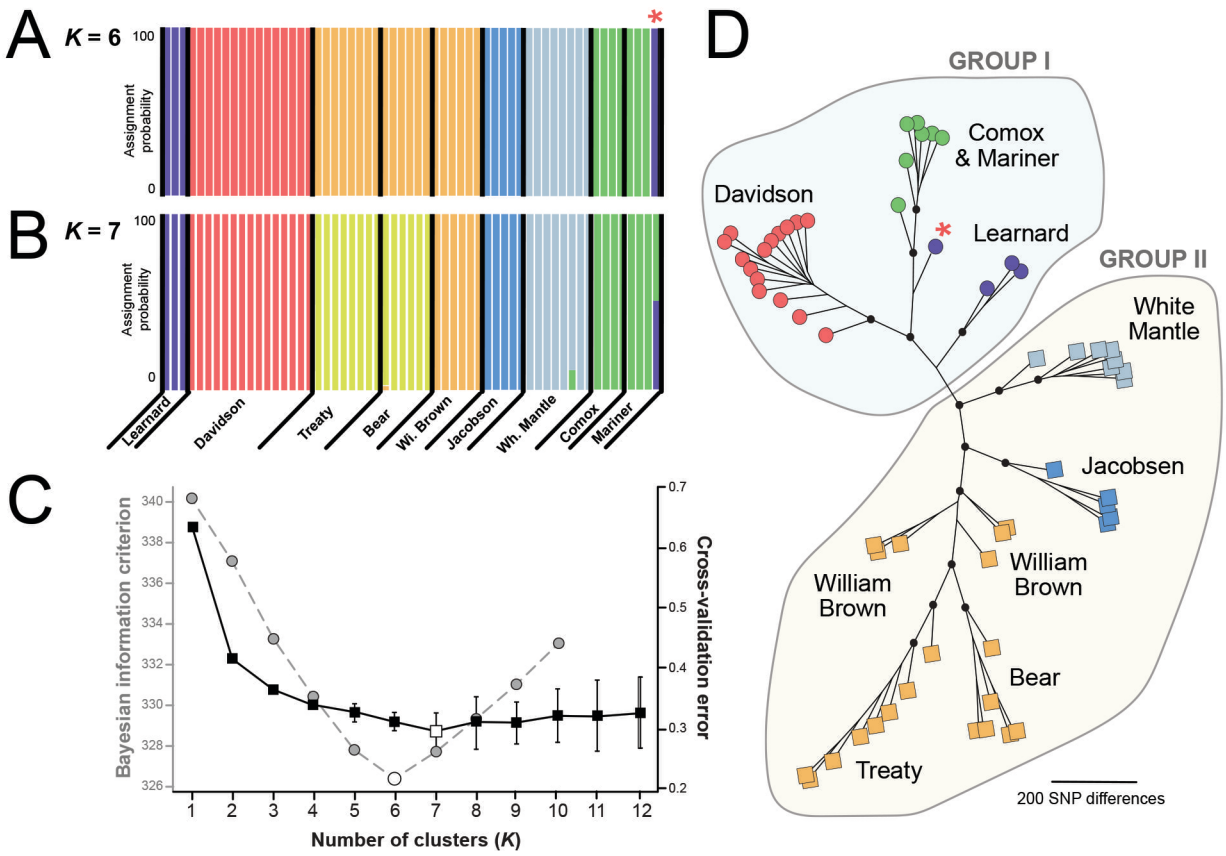


Figure 2. Population genetic structure of ice worms based upon (A) a discriminant analysis of principal components (DAPC) for $K = 6$ and (B) ADMIXTURE results for $K = 7$. (C) Comparisons of support for different values of K for DAPC (Bayesian information criterion, BIC; gray dashed line, left y-axis) 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 white bars represent the standard deviation of 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 > 95% bootstrap support. Specimens in group I (circles) and II (squares) are denoted with different symbols. Asterisks highlight a single specimen, MS5, which showed evidence of shared ancestry across geographically disjunct populations, indicating LDD.

Our fineRADstructure results mirrored those from DAPC, identifying the same six genetic clusters with MS5 assigned to the Learnard cluster rather than the Comox/Mariner cluster. Still, MS5 again exhibited substantial evidence for shared ancestry between the Learnard and Comox/Mariner genetic clusters (Fig. 1). On average, MS5 shared 411 loci with Learnard ice worms and 220 loci with ice worms from the Comox and Mariner ice worms (Fig. 1), a roughly 60/40 split between the two clusters. Notably, MS5 also exhibited the highest heterozygosity of any individual in the study (and this result did not stem from outsized

257 coverage, Fig. S4). Our fineRADstructure analysis also revealed the potential for an overarching,
258 deep divergence between two groups of ice worm populations (groups I and II; Fig. 1). This split
259 was corroborated by both SVDQuartets (Fig. 2D) and F_{ST} comparisons. Mean pairwise F_{ST}
260 among populations within groups I and II were 0.21 and 0.30, respectively. Across groups,
261 however, mean pairwise F_{ST} was 0.60. Mean distance to the Pacific Ocean also differed between
262 groups by 100.5 km (group I = 73.3 km, group II = 173.8 km; ANOVA, $P < 0.001$).
263 To be clear, the geographic groupings (I and II) described in this study are with respect to
264 the Coast Mountains, and specifically a putative ice ridge that formed along the crest of the range
265 during the Pleistocene [28]. All populations, regardless of grouping, followed the known pattern
266 of ice worms inhabiting relatively coastal glaciers (see Fig. S1). Interestingly, while the
267 geographic location of 8 of 9 populations fell on either side of the putative ice ridge, one
268 population did not: White Mantle. Our resolution of a deep, two-group split within ice worms
269 adds new clarification to the three-clade result identified by earlier studies [e.g., 23, 25], showing
270 that the previously recognized northern and central clades should likely be lumped together (now
271 group I).
272
273 **Discussion:**
274 Historical and contemporary factors, both biotic and abiotic, interact to shape the present-
275 day genetic structure of populations and species. Disentangling their varied contributions can be
276 difficult, however, particularly when evolutionary histories are muddled by unexpected events
277 (e.g., LDD of an organism with limited potential for active dispersal). The modern population
278 genomic toolkit provides historically unprecedented power to resolve biogeographical
279 complexity by allowing more quantitative perspectives of relatedness and greatly improved
280 resolution of genetic independence or similarity [42]. In this study, we used a population
281 genomic data set to refine understanding of the evolution history of the extremophile, glacier-
282 obligate ice worm, *M. solifugus*. Our results provide a clear genomic perspective of LDD,
283 showing unequivocally that migration has occurred between southern Alaska and the glaciers of
284 Vancouver Island ~1,900 km to the south and across the Pacific Ocean. We also provide an
285 independent line of biological evidence in support of the geological hypothesis that an ice ridge
286 formed along the crest of the Coast Mountains during the Pleistocene [28].

287 The recent biogeographic history of ice worms appears to have been shaped by three
288 main factors: i) ice sheet dynamics (likely during the Pleistocene), ii) mountaintop isolation from
289 conspecifics following the retreat of Pleistocene ice into higher elevations, and iii) LDD.
290

291 i. Evidence for the first, overarching factor that has defined the recent evolution of ice
292 worms lies in our overwhelming support for divergence between two overarching
293 groups of ice worms (I and II). During the Pleistocene, ice was generated in the high
294 peaks of the Coast Mountains and subsequently flowed west to the Pacific and east to
295 the BC interior [28, 43, 44]. This potential western-eastern divergence in ice worms
296 aligns with this divide, suggesting that each group diverged in allopatry from their
297 conspecifics. At its maximum, the Cordilleran ice sheet was a ~2,000-3,000 m high
298 convex dish with gentle interior slopes that steepened at its periphery [28, 45]. Given
299 the sensitivity of ice worms to extreme cold [23], populations likely only persisted on
300 the margins of the ice sheet, as their present-day occurrences on the lower flanks of
301 higher elevation, low-latitude glaciers suggest. It is possible that, as suggested
302 previously [23, 25], the Boundary Ranges (the most northern subrange of the Coast
303 Mountains), are actually the biogeographic barrier that drove the observed deep
304 divergence in ice worms described in this study. However, without more fine-scale
305 population genomic sampling on both sides of the Coast Mountains (including the
306 Boundary Ranges), this nuance of ice worm biogeography will remain unclear. In the
307 same vein, the strong genetic similarity of White Mantle to populations east of the
308 proposed ice ridge (Fig. 1), despite falling on its western side indicates that either the
309 ice ridge actually formed more to the west than previously thought [28], the White
310 Mantle population has migrated west since the end of the Pleistocene, or the ice ridge
311 itself was not a barrier driving differentiation (as discussed above).

312 ii. The Cordilleran ice sheet was seeded by alpine glaciers and throughout western North
313 America, the ice sheet's signature persists in mountainous regions. While the specific
314 dynamics of deglaciation on valley and drainage scales are unknown, a safe
315 assumption is that glaciers retreated from valleys into higher elevations, likely with
316 ice worm populations in tow. Increasing mountaintop isolation and subsequent
317 genetic drift likely precipitated 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 may simply reflect 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 a roughly 60/40 ancestry split between the Mariner/Comox and Learnard clusters, one specimen (MS5), is likely the progeny of recent hybridization 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 cocoons on southward migrating birds [e.g., 12]. Several passerines (e.g., Gray-crowned rosy finches, *Leucosticte tephrocotis*) have been observed feeding on ice worms [46, 47; S.H., personal observation] and the presence of an ice worm clitellum [48] indicates that ice worms, like other *Mesenchytraeus* species, reproduce by egg-laden cocoons [23]. The seemingly exclusive north-to-south pattern of ice worm LDD also has temporal support from bird migratory behavior. Indeed, late autumn ice worm reproduction [at the end of the productive season on mountain glaciers, 49] likely occurs in concert with southward-migrating birds stopping to feed on glaciers free of seasonal snow. In contrast, returning spring migrants pass over the same glaciers when seasonal snowfall still covers overwintering ice worms [23], limiting the potential for LDD in the reverse direction.

One question remains, however, if birds are precipitating LDD in ice worms, why has it only been observed for populations west of the Coast Mountains? This curiosity ties in to an important question in North American biogeography: to what extent have ice sheets driven present-day patterns of speciation and genetic differentiation among fauna of the northwest? For ice worms, we hypothesize that ice worm populations comprising groups I and II have accumulated some degree of reproductive isolation and are either already species-level lineages or on their way to this conclusion. Thus, we predict a zygotic barrier may be limiting inter-group migrants from leaving a genomic signature of their presence. It is also possible, and perhaps likely, that LDD in ice worms is merely the result of vector migration patterns. For instance, *L.*

349 *tephrocotis*, like other songbirds [50], may preferentially follow coastlines during migration.
350 However, until more is known about the specific interactions of ice worms with various bird
351 species, and thus their potential to act as LDD vectors, relating bird migrations to ice worm
352 distributions will remain difficult. Beyond ice worms, ice sheets have been implicated as a key
353 driver of speciation in boreal birds [51] and phylogeographic structure of many taxa, from
354 nematodes to gray wolves [17, 29, 52-54], and our results clearly support these broader
355 implications for biodiversity accumulation and maintenance in North America.
356
357 *Conclusions*
358 In this study, we leveraged population genomic data to unravel the complex evolutionary
359 history of the North American ice worm, *M. solifugus.* Our results add new clarity to previous
360 perspectives on ice worm biogeography while also lending genomic support to the existence of
361 contemporary, passerine-mediated LDD in the group. While the phylogenetic data used in this
362 study (i.e., the lack of an outgroup) preclude us from diagnosing groups I and II as monophyletic,
363 given the results of previous studies [23-25], we predict that future efforts will diagnose them as
364 such, perhaps even representing two distinct species. Finally, our genomic data lend support to
365 the glaciological record in the region, adding a biological line of evidence to a postulated north-
366 south ice ridge that formed along the crest of the Coast Mountains during the Pleistocene [28].
367 This potential for genomics to inform the geological record is intriguing and ice worms, as a rare
368 glacier-obligate macroinvertebrate, may be an ideal taxon for similar studies in the future.
369
370 **Acknowledgements:** We thank Joe Giersch for map making assistance and the University of
371 Kentucky Center for Computational Sciences for high-performance computing resources. This
372 research was partially funded by NSF award #IOS-082050 to D.H.S.
373
374 **Competing interests:** We have no competing interests.
375
376 **Author contributions:** S.H., D.H.S., L.M.T, and D.W.W. conceived of and funded the study.
377 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
378 and wrote the manuscript with input from D.H.S., R.K.B., and D.W.W. All authors approved the
379 final version.

380

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

384

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

386

387 **References:**

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

389 2. Martin C.H., Turner B.J. 2018 Long-distance dispersal over land by fishes: extremely
390 rare ecological events become probable over millennial timescales. *Proc R Soc B* **285**(1881),
391 20172436.

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

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

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

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

400 7. Miura O., Torchin M.E., Bermingham E., Jacobs D.K., Hechinger R.F. 2012 Flying
401 shells: historical dispersal of marine snails across Central America. *Proc R Soc B* **279**(1731),
402 1061-1067.

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

405 9. Suetsugu K., Funaki S., Takahashi A., Ito K., Yokoyama T. 2018 Potential role of bird
406 predation in the dispersal of otherwise flightless stick insects. *Ecology*.

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

410 11. Anderson J.T., Nuttle T., Rojas J.S.S., Pendergast T.H., Flecker A.S. 2011 Extremely
411 long-distance seed dispersal by an overfished Amazonian frugivore. *Proceedings of the Royal*
412 *Society of London B: Biological Sciences*, rspb20110155.

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. (doi:10.1111/mec.12239).

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

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

23. 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.

24. Hartzell P.L., Nghiem J.V., Richio K.J., Shain D.H. 2005 Distribution and phylogeny of glacier ice worms (Mesenchytraeus solifugusandMesenchytraeus solifugusrainierensis). *Canadian Journal of Zoology* **83**(9), 1206-1213. (doi:10.1139/z05-116).

448 25. Dial C.R., Dial R.J., Saunders R., Lang S.A., Lee B., Wimberger P., Dinapoli M.S.,
449 Egiazarov A.S., Gipple S.L., Maghirang M.R., et al. 2012 Historical biogeography of the North
450 American glacier ice worm, Mesenchytraeus solifugus (Annelida: Oligochaeta: Enchytraeidae).
451 *Mol Phylogenet Evol* **63**(3), 577-584. (doi:10.1016/j.ympev.2012.01.008).

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

455 27. Figuerola J., Green A.J., Michot T.C. 2004 Invertebrate eggs can fly: evidence of
456 waterfowl-mediated gene flow in aquatic invertebrates. *The American naturalist* **165**(2), 274-
457 280.

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

460 29. Shafer A.B., Cullingham C.I., Cote S.D., Coltman D.W. 2010 Of glaciers and refugia: a
461 decade of study sheds new light on the phylogeography of northwestern North America.
462 *Molecular ecology* **19**(21), 4589-4621. (doi:10.1111/j.1365-294X.2010.04828.x).

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

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

470 32. Catchen J., Hohenlohe P.A., Bassham S., Amores A., Cresko W.A. 2013 Stacks: an
471 analysis tool set for population genomics. *Molecular ecology* **22**(11), 3124-3140.

472 33. Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., Bender D., Maller J.,
473 Sklar P., De Bakker P.I., Daly M.J. 2007 PLINK: a tool set for whole-genome association and
474 population-based linkage analyses. *The American Journal of Human Genetics* **81**(3), 559-575.

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

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

478 36. Meirmans P.G., Van Tienderen P.H. 2004 GENOTYPE and GENODIVE: two programs
479 for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Resources* **4**(4),
480 792-794.

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

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

485 39. Malinsky M., Trucchi E., Lawson D.J., Falush D. 2018 RADpainter and
486 fineRADstructure: population inference from RADseq data. *Molecular biology and evolution*
487 **35**(5), 1284-1290.

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

490 41. Swofford D.L. 1998 Phylogenetic analysis using parsimony.

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

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

496 44. Stumpf A.J., Broster B.E., Levson V.M. 2000 Multiphase flow of the late Wisconsinan
497 Cordilleran ice sheet in western Canada. *Geological Society of America Bulletin* **112**(12), 1850-
498 1863.

499 45. Wilson J., Falconer G., Mathews W., Prest V. 1958 Glacial map of Canada. *Geological*
500 *Association of Canada, Toronto, ON*.

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

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

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

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

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

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

517 52. Muñoz-Fuentes V., Darimont C.T., Wayne R.K., Paquet P.C., Leonard J.A. 2009
518 Ecological factors drive differentiation in wolves from British Columbia. *Journal of*
519 *Biogeography* **36**(8), 1516-1531.

520 53. Burg T.M., Gaston A.J., Winker K., Friesen V.L. 2006 Effects of Pleistocene glaciations
521 on population structure of North American chestnut‐backed chickadees. *Molecular ecology*
522 **15**(9), 2409-2419.

523 54. Koehler A.V., Hoberg E.P., Dokuchaev N.E., Tranbenkova N.A., Whitman J.S.,
524 Nagorsen D.W., Cook J.A. 2009 Phylogeography of a Holarctic nematode, Soboliphyme
525 baturini, among mustelids: climate change, episodic colonization, and diversification in a
526 complex host–parasite system. *Biological Journal of the Linnean Society* **96**(3), 651-663.
527