Author for correspondence:
Scott Hotaling
e-mail: scott.hotaling@wsu.edu

THE ROYAL SOCIETY
PUBLISHING

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]

[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

SH, 0000-0002-5965-0986; DHS, 0000-0003-4096-2872; DWW, 0000-0002-6187-5623; JLK, 0000-0002-7731-605X

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 geographical 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 (e.g. greater than 1500 km) areas of unsuitable habitat. In this study, we analysed genome-scale sequence data for individuals from 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.

1. Introduction

For more than a century, long-distance dispersal (LDD) among presumably isolated populations has intrigued biologists [1,2]. 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 [3] and invertebrates [4–7]. Many animal vectors play an integral role in plant and invertebrate LDD [8]; 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 geographical barriers, visiting similar habitats along the way [9]. Through this mechanism, dispersal units (e.g. whole organisms, eggs, seeds, etc.) may be ingested and dispersed after passing through the digestive tract [7] or by directly adhering to the bird's exterior [9]. 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 hitch-hiking dispersers. Moreover, because rare migratory events can affect species distributions [10] and influence genetic differentiation among populations [11], 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. [12]), 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 [13–16]. These ice sheet dynamics typically affect organisms by separating and reconnecting populations as ice cover changes. However, some species are directly tied to ice sheets (e.g. the meltwater stonefly, [13,17]) and their evolutionary trajectories are therefore much more susceptible to ice sheet influence. Perhaps no species is more directly tied to ice sheets than glacier ice worms, *Mesenchytraeus solifugus* in North America [18] and *Sinenchytraeus glacialis* in Tibet [19]. The geographical 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 [20]. Ice worms cannot tolerate temperatures more than roughly $\pm 7^{\circ}$C from freezing and require glacier ice for survival and reproduction [21]. With such unique ecology and physiology, and a dispersal-limited life history, the evolutionary history of ice worms since diverging from conspecifics [22,23] 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 ice-dominated, mountain ecosystems.

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, and a 'southern' clade ranging over much of British Columbia to southern Oregon [20,21,23]. 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 kilometres south of their closest genetic relatives [21,23]. 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. [21] 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 the light of other examples of bird-mediated LDD (e.g. [5,24]). 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, this pattern may only apply to mitochondrial DNA (mtDNA) since mtDNA is maternally inherited and does not recombine. For the nuclear genome, unless strong reproductive isolation exists between the hold-over lineages and more recent colonizers, genetic differences would be rapidly homogenized by gene flow and recombination. Assuming no selection against migrants, a reasonable expectation given the similarity of habitat across the ice worm range, the frequency of contemporary versus historical mtDNA haplotypes would therefore depend upon time since introduction, scale of migration (i.e. number of introduced haplotypes) and chance.

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 factors: extreme LDD, glacier dynamics and mountaintop isolation. To test this hypothesis, we generated a genome-wide single-nucleotide polymorphism (SNP) dataset to answer three specific questions: (i) how do the clades previously inferred from a small number of markers hold up to genome-wide scrutiny? (ii) What, if any, genomic evidence exists for LDD in ice worms? (iii) How do the evolutionary relationships among ice worm populations and genetic clusters align with glacial history in the region (e.g. [25])? 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 [26], few studies have explored how ice sheet dynamics influenced the evolutionary history of species directly tied to them (e.g. [13]). Our results reveal the profound impact that ice sheet formation during the Pleistocene (approx. 2.5 million–11 700 years ago) which flowed episodically from the crest of the Coast Mountains [25] may have had on ice worm evolution, possibly precipitating an ongoing speciation event. Broadly, our findings highlight the power of population genomics to capture contemporary evidence of LDD while also providing biogeographic evidence for reconstructing the glacial history of a region.

2. Methods

(a) Sample collection, library preparation and single-nucleotide polymorphism calling

In 2009, we collected ice worms from nine glaciers across most of their geographical range (figure 1 and table 1; electronic supplementary material, figure S1). We extracted DNA from 59 worms and prepared double-digest restriction-site-associated DNA (ddRAD) sequencing libraries following Peterson et al. [27]. The 59-sample library was sequenced on one lane of an Illumina HiSeq4000 with single-end, 100 bp chemistry. Raw reads were demultiplexed, quality-filtered and ddRAD loci were assembled de novo using Stacks v. 1.46 [28]. For downstream analyses, we only included SNPs if they were present in greater than or equal to five populations, genotyped in greater than or equal to 50% of individuals per population, and were in Hardy–Weinberg equilibrium with a minor allele frequency of greater than or equal to 0.025 overall.
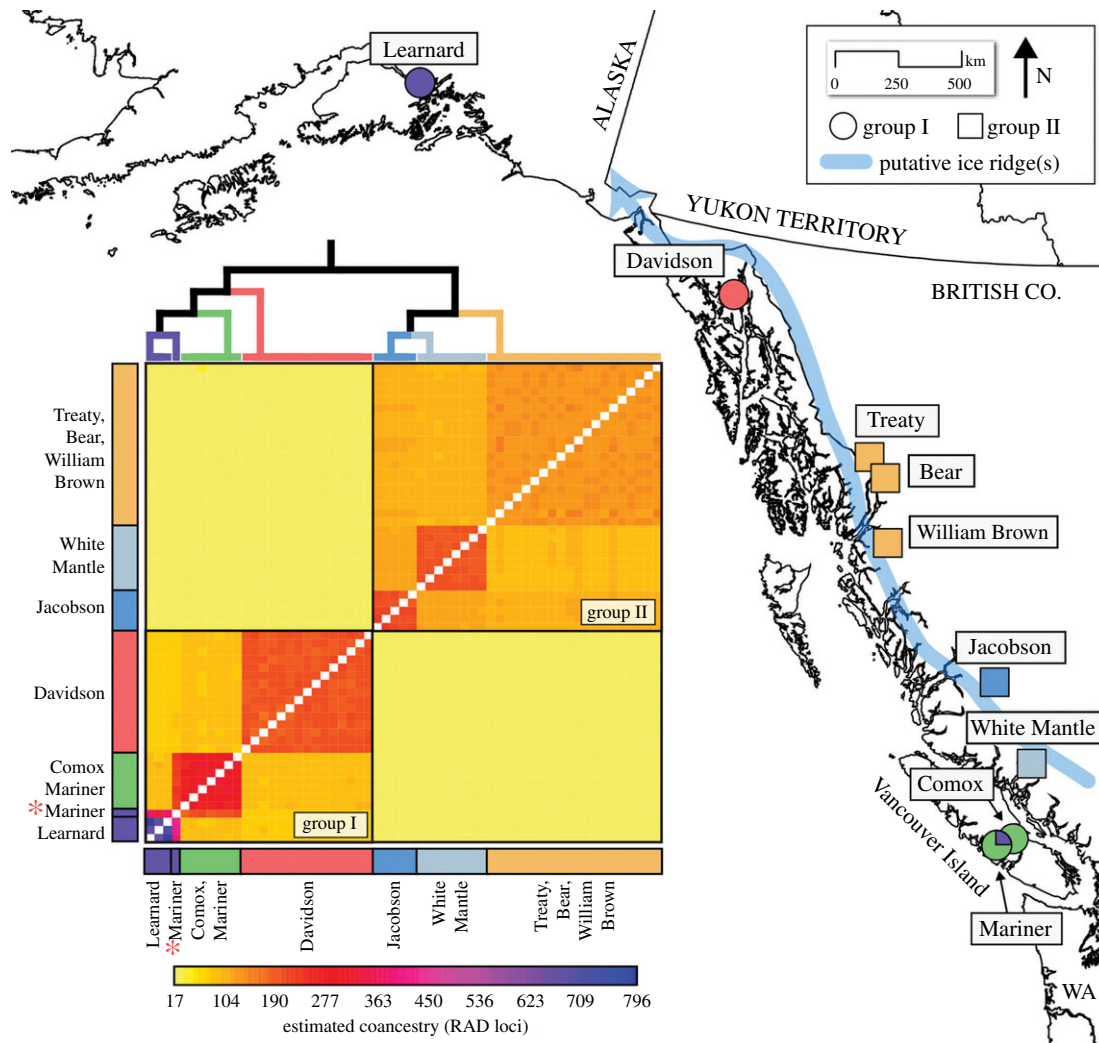
Figure 1. Ice worm populations sampled for this study. Colour-coding reflects the results of a fineRADstructure coancestry analysis. Group I (circles) and II (squares) populations were generally defined by their presence to the east or west of a key line where ice ridge(s) putatively formed during the Pleistocene greater than approximately 25 000 years ago [25] 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. (Online version in colour.)

Table 1. Sampling information and summary statistics for all ice worm populations included in this study. 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[a] (LEA)	60.806, −148.721	AK	624	3	0.114	0.100	0.026
Davidson[a] (DAV)	59.067, −135.551	AK	986	15	0.078	0.065	0.037
Treaty[a] (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[a] (WIB)	54.611, −129.129	BC	1260	6	0.056	0.057	0.000
Jacobson[a] (JAC)	52.050, −126.072	BC	1249	5	0.108	0.106	0.005
White Mantle[a] (WHM)	50.795, −125.153	BC	1764	8	0.131	0.135	−0.005
Comox[a] (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

[a]Populations included in both Dial *et al.* [23] and this study.

We further restricted analyses to one random SNP per locus for all analyses except fineRADstructure (see below). Complete details of SNP calling are provided in the electronic supplementary material and the commands used in this study are provided on GitHub (https://github.com/scotthotaling/ice_worm_ddRAD).

(b) Population genetic and phylogenetic analyses

For each population, we calculated nucleotide diversity (π), heterozygosity (Het) and the inbreeding coefficient (F_{IS}). We also calculated a pair-wise AMOVA F_{ST} for all population combinations [29] with the Stacks *populations* module. To test for a signature of
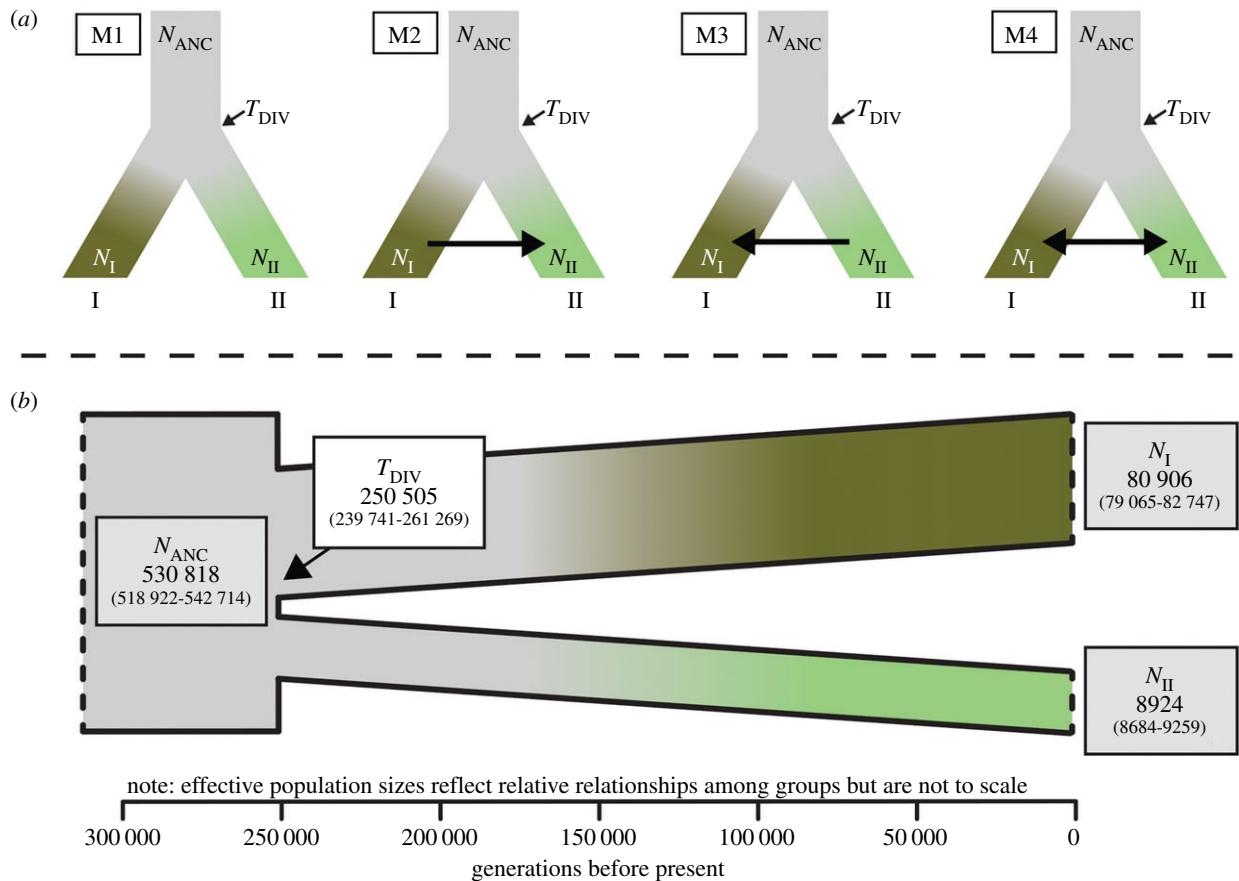
Figure 2. (a) Demographic models tested in this study which included parameters for divergence time (T_{DIV}), effective population sizes [ancestral (N_{ANC}) and for each group (N_I, N_{II})] and migration (arrows) when applicable. (b) The parametrized best-fit demographic model (M1). Numbers in parentheses represent 95% confidence interval around estimates. The white box corresponds to the divergence time (in generations) between groups I and II. Grey boxes correspond to population size parameters. (Online version in colour.)

isolation-by-distance (IBD; [30]), we estimated Euclidean distances among sites and tested the correlation between distance and F_{ST} with four Mantel tests performed in GenoDive v. 2.0b27 [31]. The first Mantel test included all nine populations and the second excluded both Vancouver Island populations (Mariner and Comox). The third and fourth Mantel tests focused on signatures of IBD within groups 'I' and 'II' (see Results). We also assessed if the mean distances to the Pacific Ocean differed between groups with a one-way ANOVA.

Population structure was inferred in two ways: a maximum-likelihood-based method using ADMIXTURE 1.3.0 [32] and a discriminant analysis of principal components (DAPC) with the R package *adegenet* [33]. ADMIXTURE analyses were performed with default settings, a range of clusters (K) from 1 to 12, and we identified the best-fit solution as the replicate that minimized the cross-validation score across all replicates for all Ks. For DAPC, we first used the *find.clusters* function to identify the optimal K, selected the appropriate number of principal components (PCs) according to the α-score and performed a final DAPC analysis using the best-fit K and optimal number of PCs identified in the previous two steps.

We extended our population structure analyses to infer shared ancestry and phylogenetic relationships with fineRADstructure [34] and SVDQuartets [35] as implemented in PAUP* v. 4.0a159 [36]. Since fineRADstructure is a haplotype-based approach, analyses were performed using all variable sites for a given ddRAD locus (i.e. a haplotype) rather than randomly selected single SNPs. For SVDQuartets, we performed exhaustive sampling of all possible quartets (every combination of four tips) and branch support was estimated with 100 non-parametric bootstrap replicates.

Complete details of these analyses are provided in the electronic supplementary material.

(c) Demographic modelling

To test hypotheses of demographic history and estimate the timing of divergence for the two groups identified in our population genetic and phylogenetic analyses (I and II), we performed demographic modelling. We used fastsimcoal2 v. 2.603 [37] which leverages a coalescent-based model to estimate demography from the site frequency spectrum (SFS). We designed and tested the fit of four two-lineage models (figure 2a): no gene flow (M1), unidirectional gene flow from group I into II (M2), unidirectional gene flow from group II into group I (M3) and bidirectional gene flow (M4). We included parameters for ancestral (N_{ANC}) and current (N_I and N_{II}) effective population sizes as well as divergence time (T_{DIV}). We maximized shared SNPs between groups by selecting the four individuals with the least missing data from the same population in each group (group I = Davidson; group II = Treaty). We also only retained loci with no missing data, which yielded 2714 SNPs across the eight individuals. We selected the best-fit model using an Akaike information criterion (AIC) and generated 95% confidence intervals by simulating 50 SFS replicates from the best-fit run of the best-fit model. Next, we performed the same analyses described above for each of the newly simulated SFSs.

We report T_{DIV} in generations but refrain from converting to years before present because the generation time for ice worms is not known. Estimates for the family Enchytraeidae also vary widely, ranging from 21 days at 18°C [38] to a full year at 10°C [39], with no estimates for low temperatures (e.g. approx. 0°C)

Table 2. Above the diagonal: pair-wise AMOVA F_{ST} values for all 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 grey): mean pair-wise shared loci for the fineRADstructure coancestry analysis (figure 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

available. Complete details of our demographic analyses are provided in the electronic supplementary material.

3. Results

(a) Population genetic and phylogenetic analyses

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 dataset included 360 534 unique loci. After filtering, we retained 6019 loci and 10 392 SNPs (mean = 1.73 SNPs per locus). This final dataset had genotype calls for approximately 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). The mean differentiation (F_{ST}) for all pair-wise 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 the least differentiated (mean $F_{ST} = 0.383$; table 2). We detected no association between genetic and geographical distances in either of the study area-wide Mantel tests (Mantel's r, all populations = −0.04, $p = 0.42$; Mantel's r, no Vancouver Island populations = 0.15, $p = 0.36$). There was, however, a signature of IBD within group II (Mantel's r, group I = 0.839, $p = 0.025$) but not within group I (Mantel's r, group II = 0.839, $p = 0.082$).

Our DAPC analyses supported $K = 6$ as the optimal number of genetic clusters (figure 3a,c). ADMIXTURE results, however, supported $K = 7$ as the best-fit (figure 3b,c) and the SVDQuartets phylogeny largely mirrored both lines of population structure evidence (figure 3d). All analyses supported multiple independent genetic clusters of ice worms. Our DAPC and ADMIXTURE results differed in two ways: (i) 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. (ii) 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 Learnard cluster, whereas Admixture equally assigned it to both the Learnard and Comox + Mariner clusters (figures 1 and 3). Finally, because our SNP filtering focused on overarching patterns in the dataset, and likely overlooked some degree of population-specific detail, our results are likely conservative estimates of genetic structure in the group.

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 (figure 1). On average, MS5 exhibited approximately 70/30 split of shared ancestry between Learnard and Comox/Mariner ice worms (figure 1). Notably, MS5 also exhibited the highest heterozygosity of any individual in the study (and this result did not stem from outsized coverage; electronic supplementary material, figure S3). Our fineRADstructure results also highlighted a primary divergence between two groups of ice worm populations (groups I and II; figure 1). This split was corroborated by both SVDQuartets (figure 3) and F_{ST} comparisons. The mean pair-wise F_{ST} among populations within groups I and II were 0.21 and 0.30, respectively. Between groups, however, the mean pair-wise F_{ST} was 0.60. The mean distance to the Pacific Ocean also differed by 100.5 km (group I = 73.3 km, group II = 173.8 km; ANOVA, $p < 0.001$).

(b) Demographic modelling

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 (figure 2b; electronic supplementary material, table S1). All other models were rejected with $\Delta AIC \geq 9.1$ (electronic supplementary material, table S1). The second-best model (M2) included unidirectional gene flow from group I into group II (figure 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 (figure 2b).

4. Discussion

Historical and contemporary factors, both biotic and abiotic, interact to shape the present-day genetic structure of
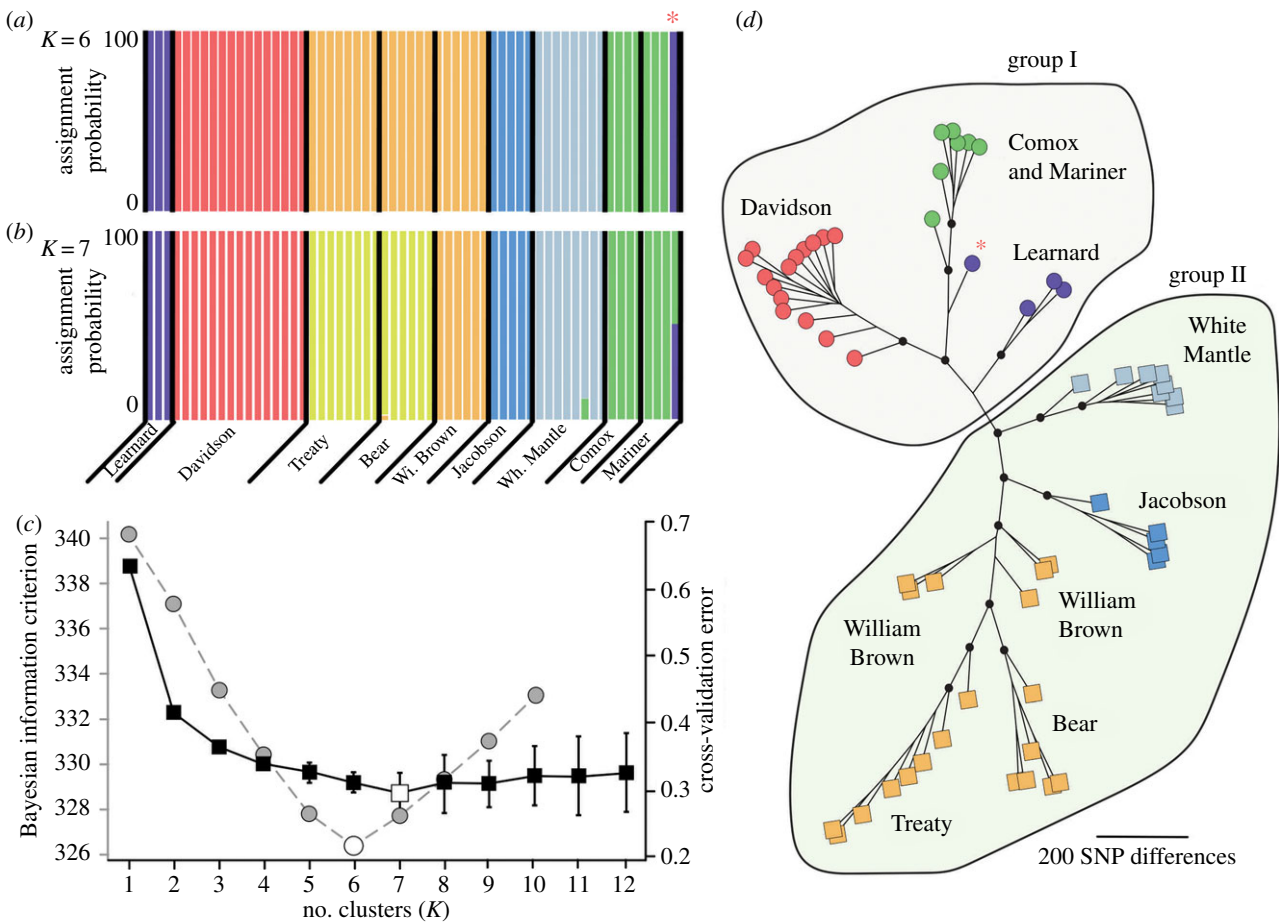
Figure 3. Population genetic structure of ice worms based upon (a) a DAPC for $K = 6$ and (b) ADMIXTURE results for $K = 7$. (c) Comparisons of support for different values of K for DAPC (Bayesian information criterion, BIC; grey dashed line, left y-axis) and Admixture (cross-validation error, CV; dark 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. (Online version in colour.)

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 biogeographic complexity by allowing more quantitative perspectives of relatedness and greatly improved resolution of genetic independence or similarity [40]. In this study, we used a population genomic dataset 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 approximately 1900 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 [25].

(a) Ice worm biogeography and long-distance dispersal

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 deep divergence between two groups (I and II) which fall largely on either side of the Coast Mountains in western North America. During the Pleistocene (approx. 2.5 million–11 700 years ago), western Canada was repeatedly covered by a continental ice sheet [25]. 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 [25,41,42]. 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 approximately 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. greater than or equal to 3), the divergence may have occurred less than 100 000 years ago. However, if ice worms develop slowly (e.g. one generation for every 5 years)—the split may have occurred over a million years ago. Either way, the split appears more recent than estimates from mtDNA of approximately 1.7 million years ago [21]. At its maximum, the Pleistocene ice sheet in the region was approximately 2000–3000 m high convex dish with gentle interior slopes that steepened at its periphery [25,43]. Given the sensitivity of ice worms to extreme cold [21], 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 [21,23], 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 electronic supplementary material, figure S4 and additional discussion). In the same vein, the strong genetic similarity of White Mantle to populations east of the proposed ice ridge (figure 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 [25], 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 [25]. 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 cocoons sticking to the feet, beaks or feathers of southward-migrating birds [9]. Several passerines (e.g. grey-crowned rosy finches, *Leucosticte tephrocotis*) have been observed feeding on ice worms [23] and the presence of an ice worm clitellum [44] indicates that ice worms, like other *Mesenchytraeus* species, reproduce by egg-laden cocoons [21]. The seemingly exclusive north-west-to-southeast pattern of ice worm LDD also has temporal support from bird migratory behaviour. Late autumn ice worm reproduction (at the end of the productive season on mountain glaciers, [45]) likely occurs in concert with southward-migrating birds stopping to feed on glaciers free of seasonal snow. By contrast, returning spring migrants pass over the same glaciers when seasonal snowfall still covers overwintering ice worms [21], 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 populations comprising groups I and II have accumulated some degree of reproductive isolation. This inference is supported by our demographic modelling which strongly rejected any model that included gene flow from group II into group I. While this may be at least partly linked to directionality in migration (e.g. bird movements), the strength in which these models were rejected suggests that a zygotic barrier may be limiting inter-group migrants from leaving a genomic signature of gene flow. It is also possible, and perhaps likely, that patterns of LDD in ice worms are driven by vector migration patterns. For instance, *L. tephrocotis*, like other songbirds [46], may preferentially follow coastlines during migration. However, until more is known about the specific interactions of ice worms with various bird species, and by proxy, their potential to act as LDD vectors, relating bird migrations to ice worm distributions and demography will remain difficult. Beyond ice worms, ice sheets have been implicated as a key driver of speciation in boreal birds [47] and phylogeographic structure of many taxa, from nematodes to grey wolves [13,26,48,49], and our results clearly support these broader implications for biodiversity accumulation and maintenance in North America.

5. Conclusion

In this study, we leveraged population genomic data to unravel the complex evolutionary history of the North American ice worm, *M. solifugus*. Our results add new clarity to previous perspectives on ice worm biogeography while also lending genomic support to the existence of contemporary, likely passerine-mediated LDD in the group. We described two genetic groupings (I and II) which are described with respect to the crest of the Coast Mountains, where ice ridges formed during the Pleistocene [25]. While the phylogenetic data used in this study (i.e. the lack of an outgroup) preclude us from diagnosing groups I and II as monophyletic, given the results of previous studies [20,21,23], we predict that future efforts will diagnose them as such, perhaps even representing two nascent species. Finally, our genomic data lend support to the glaciological record in the region, adding a biological line of evidence to a postulated key north–south dividing line along the crest of the Coast Mountains where ice ridges likely formed during the Pleistocene and repeatedly propagated ice flow to the east and west [25]. This potential for genomics to inform the geological record is intriguing and ice worms, as a rare glacier-obligate

macroinvertebrate, may be an ideal taxon for similar studies in the future.

References

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

2. Martin CH, Turner BJ. 2018 Long-distance dispersal over land by fishes: extremely rare ecological events become probable over millennial timescales. *Proc. R. Soc. B* **285**, 20172436. (doi:10.1098/rspb.2017.2436)

3. Viana DS, Gangoso L, Bouten W, Figuerola J. 2016 Overseas seed dispersal by migratory birds. *Proc. R. Soc. B* **283**, 20152406. (doi:10.1098/rspb.2015.2406)

4. Gittenberger E, Groenenberg DS, Kokshoorn B, Preece RC. 2006 Biogeography: molecular trails from hitch-hiking snails. *Nature* **439**, 409. (doi:10.1038/439409a)

5. Miura O, Torchin ME, Bermingham E, Jacobs DK, Hechinger RF. 2012 Flying shells: historical dispersal of marine snails across Central America. *Proc. R. Soc. B* **279**, 1061–1067. (doi:10.1098/rspb.2011.1599)

6. Pearce TA, Mulvihill RS, Porter KA. 2012 Land slugs (Gastropoda: Pulmonata) on birds demonstrate dispersal potential. *Nautilus* **126**, 38–40.

7. 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**, 1504–1506. (doi:10.1002/ecy.2230)

8. Anderson JT, Nuttle T, Rojas J.SS, Pendergast TH, Flecker AS. 2011 Extremely long-distance seed dispersal by an overfished Amazonian frugivore. *Proc. R. Soc. B* **278**, 3329–3335. (doi:10.1098/rspb.2011.0155)

9. Viana DS, Santamaría L, Figuerola J. 2016 Migratory birds as global dispersal vectors. *Trends Ecol. Evol.* **31**, 763–775. (doi:10.1016/j.tree.2016.07.005)

10. Simpson GG. 1940 Mammals and land bridges. *J. Wash. Acad. Sci.* **30**, 137–163.

11. Slatkin M. 1987 Gene flow and the geographic structure of natural populations. *Science* **236**, 787–792. (doi:10.1126/science.3576198)

12. Hohenlohe PA *et al.* 2013 Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Mol. Ecol.* **22**, 3002–3013. (doi:10.1111/mec.12239)

13. Hotaling S, Muhlfeld CC, Giersch JJ, Ali OA, Jordan S, Miller MR, Luikart G, Weisrock DW. 2018 Demographic modelling reveals a history of divergence with gene flow for a glacially tied stonefly in a changing post-Pleistocene landscape. *J. Biogeogr.* **45**, 304–317. (doi:10.1111/jbi.13125)

14. Jordan S, Giersch JJ, Muhlfeld CC, Hotaling S, Fanning L, Tappenbeck TH, Luikart G. 2016 Loss of genetic diversity and increased subdivision in an endemic alpine stonefly threatened by climate change. *PLoS ONE* **11**, e0157386. (doi:10.1371/journal.pone.0157386)

15. Antonelli A *et al.* 2018 Geological and climatic influences on mountain biodiversity. *Nat. Geosci.* **11**, 718. (doi:10.1038/s41561-018-0236-z)

16. Taberlet P, Fumagalli L, Wust-Saucy A, Cosson J. 1998 Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.* **7**, 453–464. (doi:10.1046/j.1365-294x.1998.00289.x)

17. Giersch JJ, Hotaling S, Kovach RP, Jones LA, Muhlfeld CC. 2017 Climate-induced glacier and snow loss imperils alpine stream insects. *Glob. Change Biol.* **23**, 2577–2589. (doi:10.1111/gcb.13565)

18. Emery C. 1898 Diagnosi di un nuovi genere e nuova specie di annelidi della famiglia degli Enchytraeidae. *Atti R. Accademei Lincei* **5**, 100–111. (doi:10.1080/11250003509435624)

19. Liang Y, Hsu C, Chang T. 1979 A new genus and species of Enchytraeidae from Tibet. *Acta Zootaxon. Sin.* **4**, 312–317.

20. Hartzell PL, Nghiem JV, Richio KJ, Shain DH. 2005 Distribution and phylogeny of glacier ice worms (*Mesenchytraeus solifugus* and *Mesenchytraeus solifugus* rainierensis). *Can. J. Zool.* **83**, 1206–1213. (doi:10.1139/z05-116)

21. Dial RJ, Becker M, Hope AG, Dial CR, Thomas J, Slobodenko KA, Golden TS, Shain DH. 2016 The role of temperature in the distribution of the glacier ice worm, *Mesenchytraeus solifugus* (Annelida: Oligochaeta: Enchytraeidae). *Arct. Antarct. Alp. Res.* **48**, 199–211. (doi:10.1657/AAAR0015-042)

22. Lang SA, Saglam N, Kawash J, Shain DH. 2017 Punctuated invasion of water, ice, snow and terrestrial ecozones by segmented worms (Oligochaeta: Enchytraeidae: Mesenchytraeus). *Proc. R. Soc. B* **284**, 20171081. (doi:10.1098/rspb.2017.1081)

23. Dial CR *et al.* 2012 Historical biogeography of the North American glacier ice worm, *Mesenchytraeus solifugus* (Annelida: Oligochaeta: Enchytraeidae). *Mol. Phylogenet. Evol.* **63**, 577–584. (doi:10.1016/j.ympev.2012.01.008)

24. Figuerola J, Green AJ, Michot TC. 2004 Invertebrate eggs can fly: evidence of waterfowl-mediated gene flow in aquatic invertebrates. *Am. Nat.* **165**, 274–280. (doi:10.1086/427092)

25. Clague JJ, Ward B. 2011 Pleistocene glaciation of British Columbia. In *Developments in quaternary sciences*, pp. 563–573. Amsterdam, The Netherlands: Elsevier.

26. Shafer AB, Cullingham CI, Cote SD, Coltman DW. 2010 Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Mol. Ecol.* **19**, 4589–4621. (doi:10.1111/j.1365-294X.2010.04828.x)

27. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012 Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**, e37135. (doi:10.1371/journal.pone.0037135)

28. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013 Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140. (doi:10.1111/mec.12354)

29. Weir BS, Cockerham C. 1996 *Genetic data analysis ll: methods for discrete population genetic data*. Sunderland, MA: Sinauer Assoc. Inc.

30. Wright S. 1943 Isolation by distance. *Genetics* **28**, 114–138.

31. Meirmans PG, Van Tienderen PH. 2004 GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Mol. Ecol. Resour.* **4**, 792–794. (doi:10.1111/j.1471-8286.2004.00770.x)

32. Alexander DH, Novembre J, Lange K. 2009 Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664. (doi:10.1101/gr.094052.109)

33. Jombart T, Devillard S, Balloux F. 2010 Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94. (doi:10.1186/1471-2156-11-94)

34. Malinsky M, Trucchi E, Lawson DJ, Falush D. 2018 RADpainter and fineRADstructure: population inference from RADseq data. *Mol. Biol. Evol.* **35**, 1284–1290. (doi:10.1093/molbev/msy023)

35. Chifman J, Kubatko L. 2014 Quartet inference from SNP data under the coalescent model. *Bioinformatics* **30**, 3317–3324. (doi:10.1093/bioinformatics/btu530)

36. Swofford DL. 1998 PAUP*: phylogenetic analysis using parsimony (and other methods). Sunderland, MA: Sinauer Associates.

37. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll MJPG. 2013 Robust demographic inference from genomic and SNP data. PLoS Genet. 9, e1003905. (doi:10.1371/journal.pgen.1003905)

38. Reynoldson T. 1943 A comparative account of the life cycles of Lumbricillus lineatus Mull. and Enchytraeus albidus Henle in relation to temperature. J. Ann. Appl. Biol. 30, 60–66. (doi:10.1111/j.1744-7348.1943.tb06155.x)

39. Standen V. 1984 Production and diversity of enchytraeids, earthworms and plants in fertilized hay meadow plots. J. Appl. Ecol. 21, 293–312. (doi:10.2307/2403055)

40. Allendorf FW, Hohenlohe PA, Luikart G. 2010 Genomics and the future of conservation genetics. Nat. Rev. Genet. 11, 697–709. (doi:10.1038/nrg2844)

41. Margold M, Jansson KN, Kleman J, Stroeven AP, Clague JJ. 2013 Retreat pattern of the Cordilleran Ice Sheet in central British Columbia at the end of the last glaciation reconstructed from glacial meltwater landforms. Boreas 42, 830–847. (doi:10.1111/bor.12007)

42. Stumpf AJ, Broster BE, Levson VM. 2000 Multiphase flow of the late Wisconsinan Cordilleran ice sheet in western Canada. Geol. Soc. Am. Bull. 112, 1850–1863. (doi:10.1130/0016-7606(2000)112<1850:MFOTLW>2.0.CO;2)

43. Wilson J, Falconer G, Mathews W, Prest V. 1958 Glacial map of Canada. Toronto, ON: Geological Association of Canada.

44. Shain DH, Carter MR, Murray KP, Maleski KA, Smith NR, McBride TR, Michalewicz LA, Saidel WM. 2000 Morphologic characterization of the ice worm Mesenchytraeus solifugus. J. Morphol. 246, 192–197. (doi:10.1002/1097-4687(200012)246:3<192::AID-JMOR3>3.0.CO;2-B)

45. Hotaling S, Hood E, Hamilton TL. 2017 Microbial ecology of mountain glacier ecosystems: biodiversity, ecological connections and implications of a warming climate. Environ. Microbiol. 19, 2935–2948. (doi:10.1111/1462-2920.13766)

46. Ruegg KC, Smith TB. 2002 Not as the crow flies: a historical explanation for circuitous migration in Swainson's thrush (Catharus ustulatus). Proc. R. Soc. Lond. B 269, 1375–1381. (doi:10.1098/rspb.2002.2032)

47. Weir JT, Schluter D. 2004 Ice sheets promote speciation in boreal birds. Proc. R. Soc. Lond. B 271, 1881–1887. (doi:10.1098/rspb.2004.2803)

48. Hotaling S, Giersch JJ, Finn DS, Tronstad LM, Jordan S, Serpa LE, Call RG, Muhlfeld CC, Weisrock DW. 2019 Congruent population genetic structure but differing depths of divergence for three alpine stoneflies with similar ecology, geographic distributions, and climate change threats. Freshwater Biol. 64, 335–347. (doi:10.1111/fwb.13223)

49. Burg TM, Gaston AJ, Winker K, Friesen VL. 2006 Effects of Pleistocene glaciations on population structure of North American chestnut-backed chickadees. Mol. Ecol. 15, 2409–2419. (doi:10.1111/j.1365-294X.2006.02957.x)