

RESEARCH ARTICLE

# Quantitative Phylogenomics of Within-Species Mitogenome Variation: Monte Carlo and Non-Parametric Analysis of Phylogeographic Structure among Discrete Transatlantic Breeding Areas of Harp Seals (*Pagophilus groenlandicus*)

Steven M. Carr<sup>1\*</sup>, Ana T. Duggan<sup>1‡</sup>, Garry B. Stenson<sup>2</sup>, H. Dawn Marshall<sup>3</sup>

**1** Genetics, Evolution, and Molecular Systematics Laboratory, Department of Biology, Memorial University of Newfoundland, St John's, NL A1B 3X9, Canada, **2** Wildlife Genetics and Genomics Laboratory, Department of Biology, Memorial University of Newfoundland, St John's, NL A1B 3X9, Canada, **3** Marine Mammals Section, Science Branch, Dept. of Fisheries and Oceans, PO Box 5667, St. John's, Nfld., A1C 5X1, Canada

‡ Current address: McMaster Ancient DNA Centre, McMaster University, Hamilton ON, L8S 4L8, Canada

\* [scarr@mun.ca](mailto:scarr@mun.ca)



CrossMark  
click for updates

OPEN ACCESS

**Citation:** Carr SM, Duggan AT, Stenson GB, Marshall HD (2015) Quantitative Phylogenomics of Within-Species Mitogenome Variation: Monte Carlo and Non-Parametric Analysis of Phylogeographic Structure among Discrete Transatlantic Breeding Areas of Harp Seals (*Pagophilus groenlandicus*). PLoS ONE 10(8): e0134207. doi:10.1371/journal.pone.0134207

**Editor:** Axel Janke, BiK-F Biodiversity and Climate Research Center, GERMANY

**Received:** June 26, 2014

**Accepted:** July 7, 2015

**Published:** August 24, 2015

**Copyright:** © 2015 Carr et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** Lab work was supported by an NSERC Discovery Grant [RGPIN-2010-05265][<http://www.nserc-crsng.gc.ca>] and DFO University Partnership grants (including funding from the Northern Cod Research Program) to SMC and GBS. Funding for field collection was provided by the Department of Fisheries and Oceans (Canada), the Institute of Marine Resources (Norway), and the Polar Research of Marine Fisheries and Oceanography (Russia).

## Abstract

Phylogenomic analysis of highly-resolved intraspecific phylogenies obtained from complete mitochondrial DNA genomes has had great success in clarifying relationships within and among human populations, but has found limited application in other wild species. Analytical challenges include assessment of random versus non-random phylogeographic distributions, and quantification of differences in tree topologies among populations. Harp Seals (*Pagophilus groenlandicus* Erleben, 1777) have a biogeographic distribution based on four discrete trans-Atlantic breeding and whelping populations located on “fast ice” attached to land in the White Sea, Greenland Sea, the Labrador ice Front, and Southern Gulf of St Lawrence. This East to West distribution provides a set of *a priori* phylogeographic hypotheses. Outstanding biogeographic questions include the degree of genetic distinctiveness among these populations, in particular between the Greenland Sea and White Sea grounds. We obtained complete coding-region DNA sequences (15,825 bp) for 53 seals. Each seal has a unique mtDNA genome sequence, which differ by 6 ~ 107 substitutions. Six major clades / groups are detectable by parsimony, neighbor-joining, and Bayesian methods, all of which are found in breeding populations on either side of the Atlantic. The species coalescent is at 180 KYA; the most recent clade, which accounts for 66% of the diversity, reflects an expansion during the mid-Wisconsinan glaciation 40 ~ 60 KYA.  $F_{ST}$  is significant only between the White Sea and Greenland Sea or Ice Front populations. Hierarchical AMOVA of 2-, 3-, or 4-island models identifies small but significant  $\Phi_{SC}$  among populations within groups, but not among groups. A novel Monte-Carlo simulation indicates that the observed distribution of individuals within breeding populations over the phylogenetic tree requires

**Competing Interests:** The authors have declared that no competing interests exist.

significantly fewer dispersal events than random expectation, consistent with island or *a priori* East to West 2- or 3-stepping-stone biogeographic models, but not a simple 1-step trans-Atlantic model. Plots of the cumulative pairwise sequence difference curves among seals in each of the four populations provide continuous proxies for phylogenetic diversification within each. Non-parametric Kolmogorov-Smirnov (K-S) tests of maximum pairwise differences between these curves indicates that the Greenland Sea population has a markedly younger phylogenetic structure than either the White Sea population or the two Northwest Atlantic populations, which are of intermediate age and homogeneous structure. The Monte Carlo and K-S assessments provide sensitive quantitative tests of within-species mitogenomic phylogeography. This is the first study to indicate that the White Sea and Greenland Sea populations have different population genetic histories. The analysis supports the hypothesis that Harp Seals comprises three genetically distinguishable breeding populations, in the White Sea, Greenland Sea, and Northwest Atlantic. Implications for an ice-dependent species during ongoing climate change are discussed.

## Introduction

Analyses of multiple complete intraspecific mtDNA genomes were first applied to humans to clarify the historical emergence of modern humans and their subsequent migrations Out of Africa [1]. The accumulation of many thousands of such genomes has clarified more recent population movements in great detail, including those influenced by successive Holocene glaciations [2]. For example, high-resolution mitogenomic sampling of Iberian refugial lineages effectively discriminates postglacial dispersal and ecological events [3]. Only a few other wild species have been the subject of extensive mitogenomic phylogeographic analysis. Atlantic Cod (*Gadus morhua*) have a much more ancient species structure than previously suspected, with coalescence during the Wisconsinan glacial, and major clade differentiation at the peak of the Sangamon / Würm interglacial rather than subsequent to the last Recent glacial [4]. The major clades are phylogeographically mixed, and well-separated geographic samples coalesce only towards the base of the gene tree. The sister species of Atlantic Cod, the Walleye Pollock (*Gadus chalcogrammus*), includes geographically isolated populations originally supposed to be of recent origin that are shown by mitogenomic analysis to be ancient [5]. In contrast, separate geographic populations of freshwater whitefish (*Coregonus*) are discrete genomic clades [6]. Major phylogenetic lineages in Killer Whales (*Orcinus*) correspond to discrete geographic populations including some assignable to subspecies [7], and in Fin Whales (*Balaenoptera*) to what appear to be recognizable species [8].

Standard methods of phylogenetic analysis and (or) quantitative assessment of within-species mtDNA population structure founder when every individual is a distinct branch. Hierarchical analysis of molecular variance (AMOVA) among haplotypes *per se* is uninformative, as the entire variance occurs among individuals with respect to the total ( $F_{ST} = 1.0$ ). AMOVA based on nucleotide divergences between haplotypes ( $\phi_{ST} < 1.0$ ) may partition variance among populations ( $\phi_{SC} > 0.0$ ), or groups ( $\phi_{CT} > 0.0$ ), but does not capture phylogenetic structure within and among populations. Similarly, row by column tests of the relative abundances of major clades across populations cannot take into account phylogenetic relationships either within or among those clades [9]. Evaluation of hypotheses that phylogeographic sub-components within species have different phylogenetic tree topologies is elusive, not least

because it is difficult to define precisely what may be meant by ‘different’. Simple phylogenetic clustering [7,8] or nested clade analysis is informative for more overt structures wherein discrete clades are nested within others [10], but NCA and other coalescence analyses are less informative when geographic population components are of mixed genetic composition and clades coalesce only at the most basal levels [4]. We require novel methods that can be tested on species with phylogeographies that provide *a priori* hypotheses.

Harp seals (*Pagophilus groenlandicus* Erxleben 1777) are endemic to the North Atlantic and adjacent Arctic waters (Fig 1), where they are the most abundant pinniped [11]. The genus name refers to their “ice-loving” habit, and their whelping (pupping) and molting are associated with drifting pack ice. The species comprises exactly four populations associated with discrete whelping and breeding areas in the White Sea, in the Greenland Sea near Jan Mayen Island, and in the Northwest Atlantic along the east coast of Canada. The latter has been separated into two major sub-areas, one on the pack ice off the coast of northeast Newfoundland and southern Labrador and the other in the southern Gulf of St. Lawrence. During the whelping period, seals in each area may form one or several concentrations (“patches”), which may be split, mixed, and reformed by current- and wind-driven ice movements. There is an extremely rapid, two-week nursing period, and breeding occurs immediately after the pups are weaned. Harp seals undertake extensive migrations between winter whelping areas and summer feeding areas, and the ranges of adult and immature seals from different populations are known to overlap during the non-breeding season [11,12,13,14,15]. Animals are exchanged between the Gulf and Front populations through the Strait of Belle Isle between the Northern Peninsula of Newfoundland and the Quebec mainland. Exchange between the White and Greenland Seas and the Northwest Atlantic occurs along pack ice attached to the Norwegian and southern Greenland coasts, and through the Davis Strait between Greenland and Canada. Tag return data indicate a high degree of breeding site fidelity of adults, although there are rare reports of pups tagged in one whelping area, and recovered as adults from other areas shortly after the breeding period [11,12,13,16,17,18,19,20,21]. It is therefore possible that there are genetically effective exchanges of reproductive adults among breeding areas.

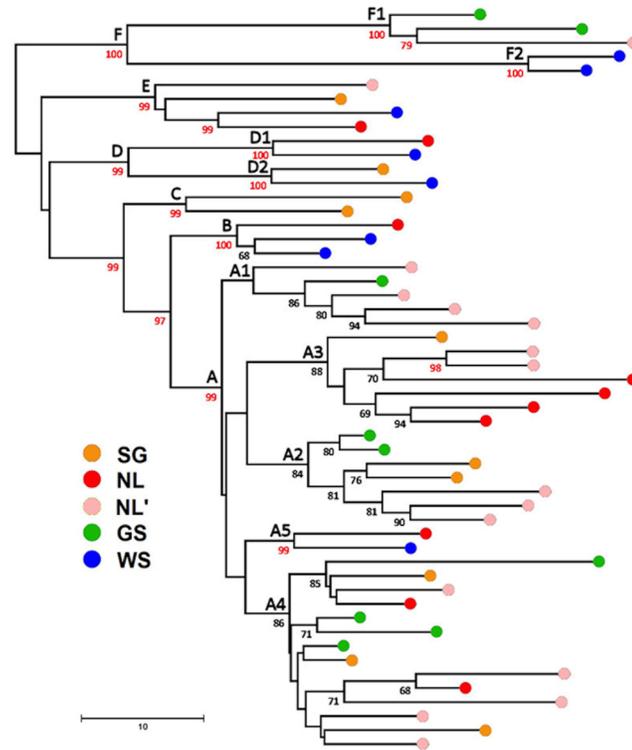
Relationships among the four populations have been examined by means of cranial measurements, underwater vocalizations, and a variety of genetic methods including allozyme and minisatellite analysis (reviewed in [22]). Collectively, these studies suggest that the Northwest Atlantic breeding areas are to some degree reproductively isolated from the Greenland Sea and White Sea populations, but there have been no clear indications of genetic differentiation between the latter two populations. Microsatellite studies of other fast-ice seal species have not identified significant population structure [23], including Hooded Seals (*Cystophora cristata*) at Jan Mayen Island *versus* those in the Northwest Atlantic [24].

Besides a conventional AMOVA, we introduce here two novel methods for analysis of intra-specific gene trees, as applied to an examination of the mitogenomic structure of Harp Seal breeding areas. These are a Monte Carlo simulation of phylogeographic distribution, and a non-parametric analysis of cumulative pairwise mismatch distributions between populations. The discrete, ordered geographic distribution of the Harp Seal breeding and whelping areas provides several *a priori* hypotheses of phylogeographic relationships. With mitogenomic data that distinguish each individual as a discrete “twig” in the intraspecific family tree, it is possible to evaluate these hypotheses quantitatively. The results show that fine distinctions among populations can be made with mitogenomic data, which with respect to AMOVA methods offer an improved view of relationships among breeding areas that bears on biological and ecological questions of genetic connectivity among populations.



**Fig 1. Breeding and whelping areas of Harp Seals (*Pagophilus groenlandicus*) and the origins of samples used in this study (from [22]).** The Dominion of Newfoundland was a separate nation until union with Canada in 1949, and its postage stamps often depicted wildlife, including Harp Seals. The stamp shows a “White Coat”, a neonatal seal prior to weaning at ca. 12 days of age.

doi:10.1371/journal.pone.0134207.g001



**Fig 2. Mitochondrial DNA genome phylogeography of 53 Harp Seals.** Neighbor-joining tree rooted with respect to the next most closely related species, the Ribbon Seal (*Histriophoca fasciata*) [25]. Clusters supported by >67% of 10,000 bootstrap replications are indicated by numbers below the corresponding vertex; support >95% is indicated in red. Origins of samples from the four breeding areas (including separate Gulf and Front samples) are indicated by colored circles. The six primary groups / clades and their inter-relationships ((F+ [(E + D) + [C + [B+A]]])) are all supported by >97% of NJ and MP bootstraps.

doi:10.1371/journal.pone.0134207.g002

## Results

Each of the 53 seals had a unique 15,825 bp genome sequence, which differed by 7 to 106 substitutions (mean =  $50.9 \pm 19.8$ ) among 588 variable sites (S1 Table). We designate the southern Gulf, Newfoundland and Labrador ice front, Greenland Sea, and White Sea population samples as SG, NL, GS, and WS, respectively, and the Northwest and Northeast Atlantic sample population pairs as NWA and NEA, respectively. The sequences have been submitted to GenBank and assigned the accession numbers KP942529—KP942581.

### Topological structure

Bootstrap analysis of the neighbor-joining tree identified six primary groups (designated A, B, C, D, E, and F) and a further nine sub-groups or pairs within these at  $\geq 95\%$  (Fig 2; Table 1). Within the majority A group (66% of all seals), five subgroups designated A1–A5 are supported in 84 ~ 99% of replicates, as are two subgroups each in D and E. Parsimony analysis based on 243 phylogenetically informative sites identified  $n = 57$  minimum-length trees of length  $L = 808$ , and bootstrap analysis identified the same 34 groups in the NJ analysis as clades, including the six primary groups in 87 ~ 99% of replicates (results not shown). Unresolved relationships in the topology of the phenetic and cladistic trees occur only within the A clade, in particular basal relationships within A3 and A4 and the relative branching order of A2 and

**Table 1. Summary of genetic diversity within major groups and clades of Harp Seals.**

Clade	Mean	s.d.	Min	Max	N
F	55.0	27.3	13	82	5
E	33.3	4.1	27	39	4
D	39.7	12.3	23	51	4
C	32.0	-	-	-	2
B	20.7	4.2	16	224	2
A	36.3	10.1	7	65	35
A1	23.8	6.1	14	35	5
A2	22.2	5.3	7	29	7
A3	30.7	7.7	15	49	7
A4	25.4	8.4	8	49	2
A5	36.6	-	-	-	14
Total	51.6	19.8	7	106	53

doi:10.1371/journal.pone.0134207.t001

A3 with respect to A4 / A5. In general, NJ relationships supported in <67% are unresolved in the strict-consensus of the 57 L = 808 MP trees. A Bayes analysis also produces the same topology, with posterior probabilities > 0.95 supporting the nodes above, as well as among designated subgroups within clade A (Fig 3). Relationships within A4 remain unresolved and by the method collapse as a polytomy. A maximum likelihood tree based on a clock-constrained general linear model (see below) again identifies the same topology as the NJ and MP analyses, with nodes A-F and A1-A5 all supported at >95%, again with unresolved relationships only within A4 (Fig 4). With whole-mitogenome sequences, identification of well-supported groups is largely method-independent. The coincident groups identified by NJ, ML, MP, and Bayes analysis are hereinafter referred to as ‘clades’.

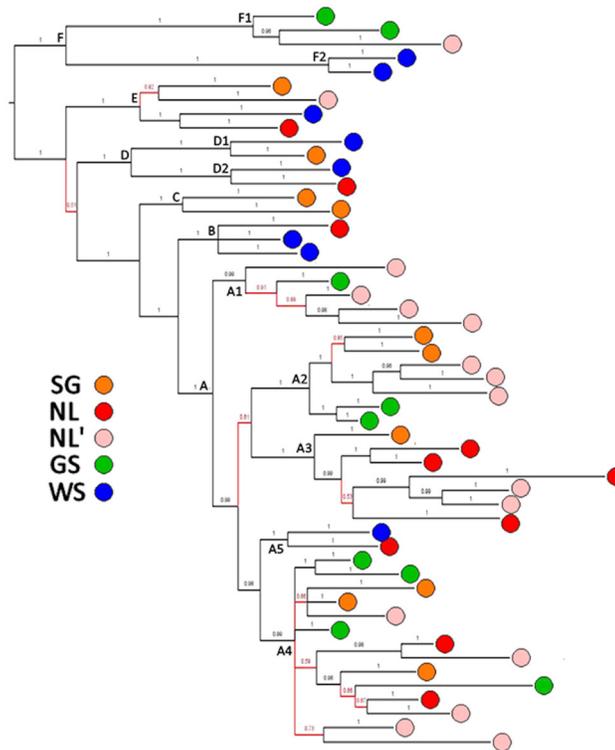
### Molecular clock and population history

Fig 5 shows a power curve fit of the ML-corrected substitutions for the MRCA of the four hominines to the reference mitogenome, with the equation  $T = 1.886 \times (d)^{1.21}$  ( $r^2 = 0.9983$ ), where T = time in KY and d = ML-corrected substitutions (Fig 5A). Use of the same curve to estimate intra-specific *Pagophilus* coalescents gave the results shown in Fig 5B and as compared to glacial history in Fig 4. The species ancestral node [ABCDE + F] occurs in the middle of the Illinoian / Saale glaciation ~180 KYA. The ancestral node of the AB+C and ABC+ DE superclades occur during the Sangamonian interglacial 120 ~ 130 KYA. The ancestral node of Clade A, which comprises 66% of genomes examined, occurs at ~80 KYA, and its five component subclades alongside three of the four in D and F coalesce ca. 40 ~ 60 MYA during the mid-Wisconsinan.

Analysis of the pairwise mismatch distribution rejects the hypothesis of constant population size (Fu & Li’s [26] and Tajima’s [27] D statistics both  $p \ll 0.01$ ) in favor of a population expansion at 53 KYA, as estimated from  $\tau = 37.4$  (Fig 6). This coincides with the origin and diversification of subclades within the modal clade A, which includes 66% of the samples.

### Analysis of molecular variance

Because every seal has a distinct haplotype sequence, hierarchal AMOVA based on haplotypic diversity *per se* assigns the entire variance as among individuals, rather than to any among-population-within-group or among-group structure ( $F_{ST} = 1.00$ ,  $F_{SC} = F_{CT} = 0.00$ ). We calculated AMOVAs based on nucleotide distances ( $\phi$  statistics) among haplotypes for all pairwise



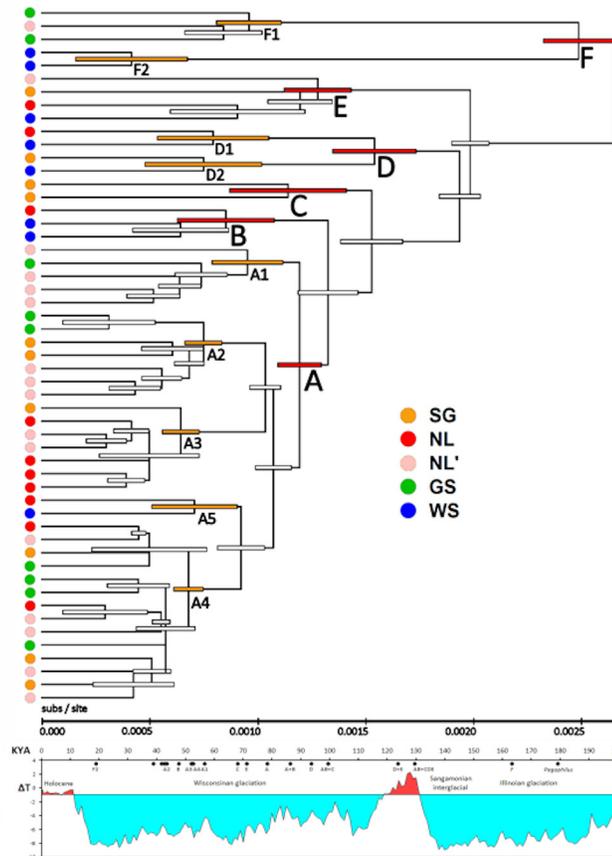
**Fig 3. Bayes analysis of mtDNA genome phylogeography of Harp Seals.** Tree rooted as in Fig 2. Nodes supported by posterior probabilities of  $p \geq 0.95$  are shown in black, those supported by  $0.50 \leq p < 0.95$  are shown in red. All groups identified by  $> 95\%$  of bootstraps in Fig 2 are supported with posterior probabilities  $p \sim 1.00$ , as are relationships within A. Differences in relationships within subgroup A4 are not strongly differentiated by bootstrap or posterior probabilities.

doi:10.1371/journal.pone.0134207.g003

combinations of breeding grounds (Table 2), and for four partitions of breeding grounds based on three *a priori* and one *a posteriori* biogeographic models (Table 3). These are (I) a non-hierarchical four-island model [without an among-group  $\phi$  component], (II) a two-island Trans-Atlantic model with groups NWA and NEA, (III) a three-island model with group NWA, and (IV) based on Table 2, a two-island model with WS contrasted with the other three breeding grounds grouped. In all models,  $0.88 \sim 0.94$  of the variance occurs among haplotypes within populations ( $p \ll 0.01$ ) as expected. Variance among populations within groups  $\phi_{SC}$  [=  $\phi_{ST}$  in Model I] is small ( $0.02 \sim 0.06$ ) but significant ( $p < 0.05$ ), except in Model IV. None of the hierarchal Models II ~ IV has a significant among-group  $\phi_{CT}$ . Among-group variance components are dominated by the relative distinctness of WS, but even in Model IV with  $\phi_{CT} > 0.10$  this does not approach significance ( $p > 0.25$ ).

### Quantitative phylogenomics: Monte Carlo and non-parametric methods

Because the hierarchal groups within AMOVA models are intrinsically unordered, they can capture island models but not linear stepping-stone models based on *a priori* hypotheses corresponding to the east-to-west arrangement of Harp Seal breeding grounds. Where the AMOVA is based on pairwise distances, it also cannot capture phylogenetic differences among breeding groups in their distributions as haplotypes across a gene tree topology. We therefore developed a novel Monte Carlo analysis that accommodates island and stepping-stone models, as well as topological heterogeneity among groups. The logic of our approach is illustrated in Fig 7: the

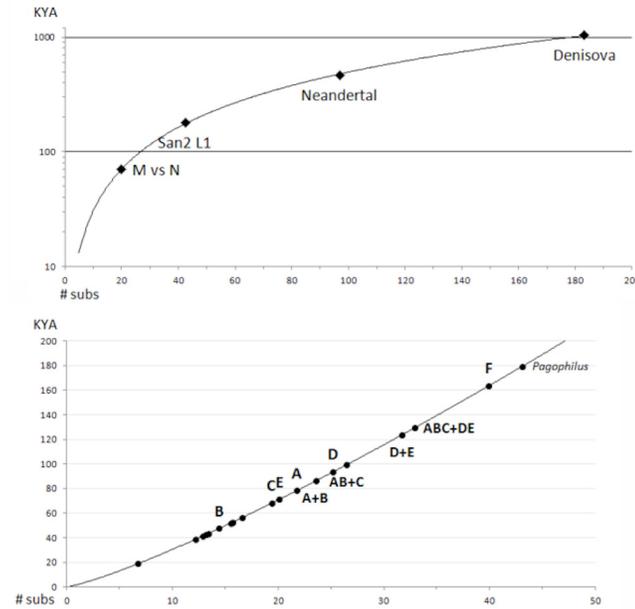


**Fig 4. Linearized maximum likelihood tree for mtDNA genome phylogeography of Harp Seals.** (a) Rates of substitution per site are estimated from the clock model in Fig 5A; confidence intervals ( $\pm 95\%$ ) are indicated by bars. (b) Correlation of Holarctic temperature and glacial histories with times of clade origins. The temperature trend over the last 200 Kyr is taken from the Vostok ice core data, which are tied to analogous data from Greenland [4].

doi:10.1371/journal.pone.0134207.g004

more highly structured a species is with respect to clade distribution among breeding areas, the smaller the number of dispersal events required to explain an observed distribution. The null hypothesis of an unstructured species may then be tested by comparing, for the topology of an observed tree (Fig 2), the length  $L$  of the observed distribution of breeding area assignments with the random distribution of  $L$  for assignments as generated by Monte Carlo simulation, for any phylogeographic model. The null hypothesis of no structure is rejected when the observed  $L$  falls in the lower (one-tailed) 5% of the random distribution. The smaller the observed  $L$  and the smaller the area of the tail in which it occurs, the greater the departure from the model-specific random distribution, and the stronger the structure. For example, in Harp Seals, an  $L = 3$  is possible if the species phylogeny were to show four clades corresponding exactly to the four breeding patches, with three of the breeding patches derived from the fourth each by single events (cf. Fig 7A).

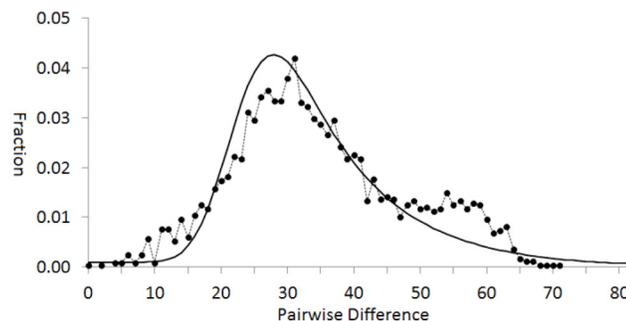
For the Monte Carlo dispersal analysis, models and matrices for each of four main models as shown graphically in Fig 8, and summary data for these and four differentially-weighted models are given in Table 4. The range of each graph includes the maximum possible length and one length class lower than the observed lower minimum. A random distribution among breeding patches is rejected in favor of a simple island model (Fig 8, Model A: 3.0% tail). A



**Fig 5. Calibration of intra-specific divergences among Harp Seal clades.** (a) ML clock-corrected estimates of the molecular distances to the MRCA for four hominines with established divergence dates < 1 MYA follow a power curve relationship  $T = 1.886 \times (d)^{1.21}$  ( $r^2 = 0.9983$ ), where  $T = \text{KYA}$  and  $d = \text{linearized ML substitution rates}$  (semi-log plot). Sources of homine mtDNA genomes are given in Methods: San2 is in the L1 haplogroup that is the basal divergence from other extant *Homo* [1], M vs N is the divergence between haplotypes in the U and A haplogroups, respectively [9]. (b) Dates of the MRCA of the major intra-specific *Pagophilus* clades estimated from the same power curve (linear plot).

doi:10.1371/journal.pone.0134207.g005

random distribution is also rejected in favor of linear, East ◀► West trans-Atlantic stepping-stone models with two or three steps (Models D and C, 3.2% and 1.3% tails, respectively), though not for a simple one-step model (B, 8.3%). (Replicate randomizations give essentially the same results). Note that Model A is essentially equivalent to AMOVA Model I (among population  $\phi_{ST} = 0.0580$ ,  $p \ll 0.01$ ). The one-step Trans-Atlantic Model B is roughly equivalent to the group partition in AMOVA Model II ( $\phi_{CT} = 0.0057$ ,  $p \gg 0.05$ ), except that Model B pools seals within the NWA and NEA groups, whereas Model II separates them within groups.



**Fig 6. Population expansion as inferred from the pairwise mismatch distribution of Harp Seals (DnaSP: [54]).** The null hypothesis of constant population size is rejected (Fu & Li's [26] and Tajima's [27] D statistics both  $p \ll 0.01$ ) in favor of a population expansion with an estimated  $\tau = 33.4$  corresponding to the upward inflection of the peak, at 57 KYA by the calibration in Fig 4.

doi:10.1371/journal.pone.0134207.g006

**Table 2. AMOVA results.** Pairwise  $\phi_{ST}$  values for the four breeding grounds [lower triangular matrix] and their significance values [upper triangular matrix]. Abbreviations: GS = southern Gulf, NL = Newfoundland and Labrador ice front, GS = Greenland Sea, WS = White Sea, NWA = GS and NL, NEA = GS and WS.

	SG	NL	GS	WS
SG	-	0.11761	0.41441	0.05871
NL	0.02007	-	0.09019	0.00089**
GS	0.00357	0.02775	-	0.03366*
WS	0.05799	0.12979	0.09921	-

\* =  $p < 0.05$   
 \*\* =  $p < 0.01$ .

doi:10.1371/journal.pone.0134207.t002

Results of the non-parametric Kolmogorov-Smirnov (K-S) test are given in Fig 9 and Table 5. The cumulative curves for the GS and WS samples are significantly offset ( $D = 0.522$ ,  $p << 0.01$ ), whereas those for the NL and SG samples are not ( $D = 0.200$ ,  $p > 0.05$ ), although those for the replicate Front sample and SG are ( $D = 0.259$ ,  $p < 0.05$ ). The curve for the NL +GS samples combined (Table 6) is significantly differentiated from all other samples, including the replicate Front sample ( $D = 0.211$ ,  $p < 0.01$ ). All four trans-Atlantic comparisons are significant. The same comparisons are significant, with slightly reduced p values, if  $D_a$  is calculated from  $(n-1)(n-2)/2$ . None are significant if  $D_a$  is calculated from  $n$ .

### Discussion

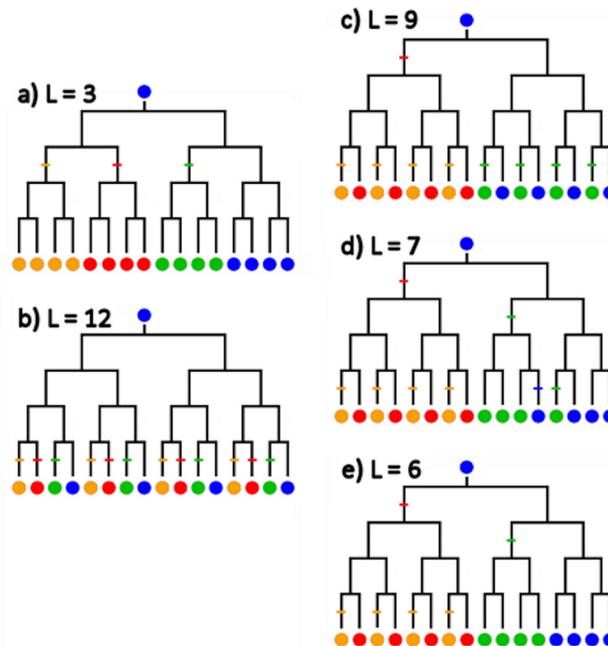
The AMOVA models identify small but significant structures among the four breeding grounds within 2-, 3-, or 4-island models, but not among these model groups. In contrast, the

**Table 3. AMOVA within and among breeding grounds of Harp Seals: alternative island models.** Four models were evaluated (I-IV), as described in Methods.

Model	Groups	Source	df	Variance component	% variance	Fixation indices	$p$		
I	Four-Island: SG NL GS WS	among groups	-	-	-				
		among pops	3	1.5303	5.80	$F_{ST}$	0.0580	0.0020	**
		w/i pops	49	24.8392	94.20				
II	Trans-Atlantic: (SG NL) (GS WS)	among groups	1	0.1502	0.57	$F_{CT}$	0.0057	0.6721	
		among pops w/i groups	2	1.4330	5.42	$F_{SC}$	0.0545	0.0188	*
		w/i pops	49	24.8392	94.01	$F_{ST}$	0.0599	0.0022	**
III	Three-Island: (SG NL) GS WS	among groups	2	1.2170	4.56	$F_{CT}$	0.0456	0.2994	
		among pops w/i groups	2	0.9471	3.55	$F_{SC}$	0.0372	0.0133	*
		w/i pops	48	24.5247	91.89	$F_{ST}$	0.0811	0.0010	**
IV	Two-Island: (SG NL GS) WS	among groups	1	2.8901	10.26	$F_{CT}$	0.1026	0.2575	
		among pops w/i groups	2	0.4282	1.52	$F_{SC}$	0.0170	0.0880	
		w/i pops	49	24.8392	88.22	$F_{ST}$	0.1179	0.0020	**

\* =  $p < 0.05$   
 \*\* =  $p < 0.01$ .

doi:10.1371/journal.pone.0134207.t003



**Fig 7. Dispersal tree length decreases as phylogeography becomes more structured.** Consider a dichotomously-branching mitogenomic tree of 16 individuals distributed among four geographic areas (red, orange, green, and blue), with one population (blue) designated as ancestral. (A) The most highly structured model, in which each of the three alternative areas corresponds to a separate clade originating by a single dispersal event, requires just  $L = 3$  events. (B) The least structured model, in which clades are distributed uniformly among areas such that there is no correspondence between genetic relationship and population, requires  $L = 12$  events. [Model drawn with delayed events, and terminal red / orange and green / blue pairs: alternative event distributions are possible]. Intermediate structures include: (C) Separate clades in the red and orange areas and in the green and blue areas, but uniform geographic distribution within those two clades ( $L = 9$ ), (D) as in (C) but with minimal exchange between blue and green areas ( $L = 7$ ), and (E) as in (C) but with complete separation of blue and green ( $L = 6$ ). The models are drawn with delayed dispersal events: other event distributions are possible. Models C, D, & E are analogous respectively to a two-population Harp Seal model, a three-population mixed model with a uniform single Northwest Atlantic population and limited exchange between the Greenland and White Seas, and a pure three-population model with separate Greenland and White Sea populations (cf. Fig 8).

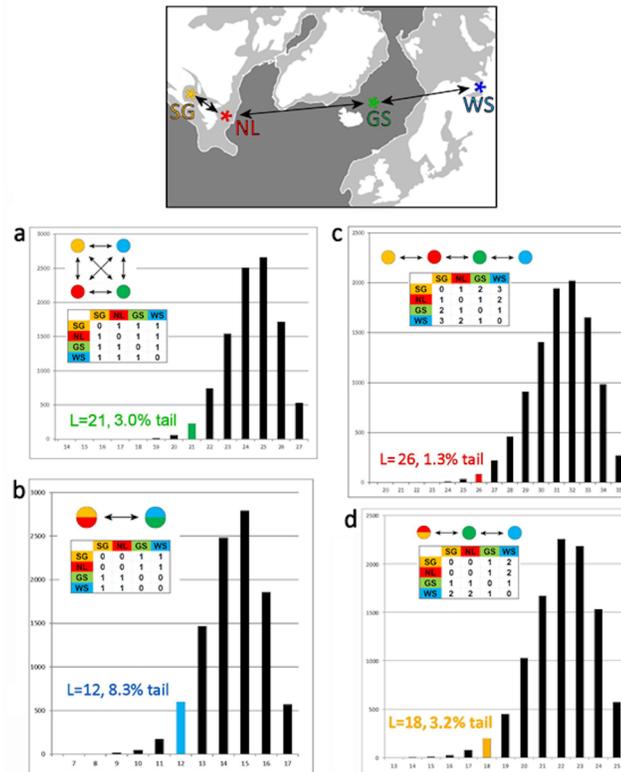
doi:10.1371/journal.pone.0134207.g007

Monte Carlo simulation and K-S tests taken together support a non-random distribution of seals, across stepping-stone as well as among island phylogeographic models in which the breeding areas comprise three distinct populations. These correspond to separate populations in the White Sea and Greenland Sea, each with distinctive phylogeographic and temporal structures, and a more homogeneous population in the Northwest Atlantic with intermediate but variable temporal structure. This is the first demonstration of genetic differentiation between the two breeding grounds in the Northeast Atlantic.

### Symplesiomorphy of cytochrome B single-locus haplotypes

Analysis of a 307bp segment of the mtDNA cytochrome *b* gene identified a common haplotype “A” in four out of ten seals from each of the two Northeast Atlantic areas and eight out of ten from each of the two Northwest Atlantic areas [22]. A row by column frequency test indicated significant differentiation of the Northwest and Northeast Atlantic pairs.

In Atlantic Cod (*Gadus morhua*), inspection of whole-genome sequences of the most common haplotype for the same gene region was a symplesiomorphic assemblage of unrelated lineages from different distantly related clades [4]. Here, inspection of the whole-genome



**Fig 8. Results of Monte Carlo simulations of genomic phylogeography.** Diagrams and matrices show the four main models graphically: matrix elements are the cost of dispersal events required by the tree. (a) Two-population model, SG+NL versus GS+WS; (b) Island model, equal cost for any movement among populations; (c) Four-population east / west linear stepping-stone model; (d) Three-population linear stepping stone model, SG+NL combined. Summary results for these and four differentially weighted models are given in Table 4. The range of each graph includes the observed maximum length and one length class lower than the observed lower minimum.

doi:10.1371/journal.pone.0134207.g008

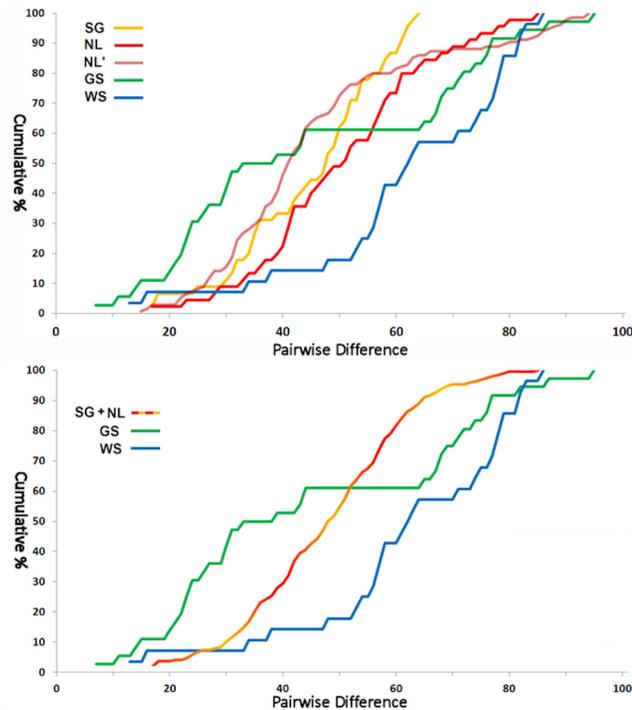
sequences of the seals with the modal Cytochrome b haplotype in [22] also identifies it as a symplesiomorphic, paraphyletic assemblage that arose with the E clade and comprises seals from the D, C, B (both seals) and A clades as well. We previously suggested that focal genotypes in many other ‘star phylogenies’ [28] based on relatively short DNA sequences and hypothesized as indicative of recent origins may in fact be symplesiomorphic groups in much older phylogenies, which would be evident in whole-mtDNA genome analyses.

**Table 4. Results of Monte Carlo simulation of alternative phylogeographic models of Harp Seal dispersal.** Abbreviations as in Table 2. For each model in Fig 8, the table gives the dispersal cost (length L) required by the topology of the observed tree (Fig 2), the frequency of that L class among 10,000 randomizations, the frequency of the tail of the distribution that includes that class, and the maximum, mode, and observed minimum among the randomizations.

Model	Description	L	L freq	L-inclusive freq	Random Min	Random Mode	Random Max
A	Island	21	2.25%	2.97%*	17	25	27
B	Trans-Atlantic, WS/GS—NL/SG	12	5.99%	8.30%	8	15	17
C	Stepping stone, WS—GS—NL—SG	26	0.84%	1.34%*	21	32	35
D	Stepping stone, WS—GS—NL/SG	18	2.07%	3.16%*	14	22	25

\* =  $p < 0.05$ .

doi:10.1371/journal.pone.0134207.t004



**Fig 9. Cumulative pairwise distribution curves for the Kolmogorov-Smirnov test.** (a) Curves for five samples analyzed separately; abbreviations as in Fig 8(B) Curves for the summary three-population model: SG+NL (not including replicate NL' samples) versus GS versus WS. Note that the former is not the simple sum of the curves for the component samples, as each includes between-sample distances as well.

doi:10.1371/journal.pone.0134207.g009

### Rates of mitogenomic evolution in *Pagophilus*

In the absence of a suitably recent biological or geological reference point for calibration of an intraspecific molecular clock in *Pagophilus* or for other phocid seals over the appropriate genetic divergences and (or) time range [29], we constructed a whole-mtDNA-genome molecular clock based on sequence data from three extant human lineages along with those of Neanderthal and a Denisovan. Sequence differences among the hominine lineages span the range

**Table 5. Kolmogorov-Smirnov tests of differences among cumulative pairwise distributions.** *D* values for pairwise comparisons among four breeding populations. *D* expresses the maximum difference between distributions weighted for sample size [60]. Significance was evaluated with the critical value *D<sub>a</sub>* calculated by two alternative criteria, either the total number of pairwise comparisons in each samples ((*n*)(*n*-1)/2), or the sample size of each samples (*n*) (see Materials and Methods). The table gives the value of *D* and its significance based on the first criterion. None of the *D* values were significant by the second criterion.

	SG	NL	GS	WS
SG	-	ns	**	***
NL	0.200	-	***	*
GS	0.389	0.389	-	**
WS	0.533	0.377	0.522	-

\* =  $p < 0.05$   
 \*\* =  $p < 0.01$   
 \*\*\* =  $p < 0.001$ .

doi:10.1371/journal.pone.0134207.t005

**Table 6. Kolmogorov-Smirnov tests of differences among cumulative pairwise distribution. *D* values for pairwise comparisons, with SG and NL samples combined. See Table 5 for considerations of *D*, *D<sub>a</sub>*, and *n*.**

<b>SG+NL</b>	<b>NL'</b>	0.211	**
<b>SG+NL</b>	<b>GS</b>	0.356	***
<b>SG+NL</b>	<b>WS</b>	0.437	***
<b>SG+NL</b>	<b>GS+WS</b>	0.335	***

\* =  $p < 0.05$   
 \*\* =  $p < 0.01$   
 \*\*\* =  $p < 0.001$ .

doi:10.1371/journal.pone.0134207.t006

observed in *Pagophilus*, and the lineage divergence times are relatively well dated though highly debated. The best linear fit has a slope of 4.5 Kyr / substitution, as compared with coding-region rates in *Homo* of 5.1 Kyr / substitution [29], a linearized Hominoid rate of 6.1 Kyr / substitution (Fig 5), and 5.7 Kyr / substitution for coding-regions in *Gadus* spp. codfish [4]. Part of the variation may be explained by use of slightly different partitions and (or) regions of homology within the coding region [cf. 30]. Other calibrations suggest that phocid mitogenomes may evolve more slowly than those of hominids [31]. Extrapolation of this curve (and variants) to date the MRCA of *Pagophilus* and its closest relative, the endemic Pacific ribbon seal (*Histriophoca fasciata*) gives a date of 3.7 ~ 4.1 MYA, which slightly antedates the most recent opening of the Bering Strait 3.5 MYA [32]. This geological event was used to calibrate the intra- [4] and interspecific [33] clocks for *Gadus* spp. The Strait has opened and closed several times over the Pleistocene [32], and the *Pagophilus* x *Histriophoca* divergence may date to one of those earlier episodes.

### Temporal implications of pairwise mismatch analysis of population expansion

Genetic biodiversity in *Pagophilus* is dominated by clade A, which originated and diversified as part of a population expansion 53 KYA (Figs 4B and 5). The estimated population expansion time immediately precedes the coalescent times of all five subclades in clade A, along with three of the other four major clades. Sequence diversity thus accompanies lineage diversification within the currently predominant lineages. The whole-genome data indicate a deep phylogeographic structure, in which the age of the deepest within-species coalescent F is similar to that of the exclusively African L lineages in *Homo*. The F1 and F2 lineages have been separate almost as long as the species coalescence. The exceptional seal from the replicate NL' sample in F1 is well separated from its closest relative, as evidence of ancient temporal divergence rather than recent dispersal, similar to what was observed in isolated populations of *Gadus chalcogrammus* [5].

### Implications of AMOVA and Monte Carlo biogeographic models

The outstanding biogeographic question in Harp Seals is the extent of movement among breeding and whelping areas, and in particular the question of whether seals from the White Sea and Greenland Sea areas are distinguishable populations. The degree of separation of the Front and Gulf populations also affects questions of management of numbers in those two areas [34,35].

The observed phylogeographic distribution is significantly structured with respect to random expectation (a simple 'island' model). All stepping-stone models fall within the shortest

5% tail of all distributions, except for the one-step, two-population Model B. The models that depart mostly strongly from random are the four- and three-step trans-Atlantic stepping-stones, where the observed distribution falls within the 1.3% and 2.1% tails, respectively. The four-island AMOVA Model I also identifies significant  $\phi_{ST}$  among populations. Thus the phylogeographic distribution is significantly structured, though the significance of the smaller number of dispersal events  $L$  with respect to the random distribution as measured by the  $p$ -value ranking of island *versus* two- or three-step stepping-stone models cannot be tested at present. The rankings of stepping-stone models seem to be dominated by the marked differentiation of WS and GS, as is also evident in the K-S tests, below. Differentiation of WS from other breeding grounds is also shown by significant pairwise  $\phi_{ST}$  and suggested by AMOVA Model IV, though the  $\phi_{CT}$  is non-significant.

### Implications of cumulative pair-wise mismatch for phylogeographic structure

The program DnaSP (see [Methods](#)) uses the distribution of pairwise mismatches among individuals within species to test whether the species showed evidence of constant size or expansion, and if so when. Prompted by this approach, we developed a novel analysis that uses a cumulative pair-wise mismatch curve calculated for each population segment as a proxy for its branching phylogenetic structure. Differences between the structures of any two populations can then be evaluated quantitatively by a non-parametric Kolmogorov-Smirnov (K-S) test of the maximum vertical offset between curves (see [Methods](#)). Here, K-S identifies the greatest disparity in the time required to reach a comparable level of phylogenetic diversity. The position of this disparity on the horizontal (cumulative pairwise difference) axis is heuristically valuable, as is the relative ranking of curves at 50% cumulative pairwise differences.

In [Fig 9A](#), the significant vertical offset of curves for the GS and WS samples reflects the different phylogenetic histories of the two populations. As noted in [Figs 2](#) and [3](#), seals from WS occur almost exclusively in the basal clades, so that the modal pairwise difference is larger and the curve shifted rightward (>50% of differences >60). Seals from GS occur predominantly in the relatively younger A clade, such that the modal difference is smaller and the curve shifted leftward (>50% of differences < 40), with the exception of the basal subclade F1, which shifts the last ~40% of differences closer to the East curve. The structure of GS (allowing for small sample size) is consistent with an older population with a relatively recent bottleneck in numbers such that the numbers of the minority clades B, C, D & E have been severely diminished, and only the plurality A clade survives alongside the basal F1. The historical catch records indicate that the population was severely depleted by the 1870s and remained low until the 1960s when it began to recover [[11](#)]. In contrast, the basal mtDNA haplotype composition as quantified by the displacement of the cumulative curve to the right indicates a much older population. Finally, lack of significant differentiation between the cumulative curves for NL and GS (though the latter is consistently to the left of the former) suggests similar population dynamics in the two, consistent with their treatment as a single biological entity.

Comparison of relative ranks among populations across the horizontal pairwise difference axis clarifies the phylogeography of the replicate Front samples. Differences among the four main samples are consistently ranked GS < SG < NL < WS, but the replicate NL (NL') curve crosses the others at several points. Inspection of [Fig 2](#) suggests a broad qualitative similarity between the phylogenetic distributions of NL' and GS samples, *i.e.*, almost all component seals are in the A clade, with outliers in the more basal clades. However, the NL' and GS curves are distinct, the former with a cumulative 50% reached at 42 differences, and the latter with a cumulative 50% at only 33 differences, and the two curves crossing at ~55% / 45 differences.

This seems to indicate, within the **A** clade, a more recent modal relationship among GS seals than among NL seals. The Front curve also crosses the Gulf curve at 30% and again at 80%; the latter seems to reflect its inclusion of basal individuals that are absent in the Front and Gulf samples. Fig 9B summarizes the three-population model as a combined NL+SG curve for comparison with the separate GS and WS curves from Fig 9A. The combined NWA curve is clearly distinct from both of the latter, and crosses the GS curve at about 60% at ~ 50 differences. Thus, the method of cumulative mismatch curves as applied to intra-specific genomic trees seems to be sensitive to subtle differences in phylogeographic structure of populations of mixed phylogenetic relationship, even with relatively small sample sizes. A limitation of K-S is that only the maximum *D* is evaluated, although this maximum is typically part of a continuous trend in cumulative differences: it would be useful to identify significant ranges.

## Implications for Harp Seal ecology and management

An understanding of the relationships among the various Harp Seal whelping areas is critical for proper science-based management during the present period of climate change in the North Atlantic [36,37,38,39,40]. Allowable catches are based upon abundance estimates derived primarily from pup production surveys in each of the whelping areas. Pup production in the White Sea declined unaccountably by more than 50% between 2003 and 2006 [41]. This decline has in part been attributed to changes in maternal body condition [42], but may also be a consequence of the movement of seals from their traditional whelping locations. Harp seals whelp along the southern limits of pack ice: as ice conditions in the North Atlantic continue to deteriorate due to climate change [43,44,45,46], it is likely that seals will be forced to move to new areas to give birth. In 2010, seals in the Northwest Atlantic were observed whelping north of their traditional area, which was attributed to a lack of suitable ice for pupping [47]. Similarly, seals have recently been reported whelping on pack ice off Cape Farewell at the southern tip of Greenland [48], an area of overlap between seals from the Northwest Atlantic and Greenland Sea in the non-whelping hunting season [49], but with no historical records of pupping.

Currently, the Northwest Atlantic is managed as a single population with quotas allocated between the Front and Gulf based upon the relative historical number of pups in each area. However, in some years, very little pupping occurs in the southern Gulf and Gulf animals move northward where they may mix with animals that normally breed at the Front [11,47]. A pattern of general patch fidelity in combination with inter-annual variation in degrees of intermixing may explain the significant differentiation of the replicate Front and SG samples, and the degree of departure from random of the three- and four-population models, with NL and SG combined or not. As ice conditions in the Gulf of St Lawrence continue to deteriorate, demographic and genetic relationship between the two Northwest Atlantic sub-populations may change, so as to require a change in the management approach [47].

## Opportunities and challenges

Whole-mitogenome sequences offer fully-resolved phylogenetic trees within species that are literal “family trees” for maternal lineages. This raises new opportunities and challenges for data analysis. Assessments that differences among phylogenetic trees drawn from different populations or geographic segments are ‘different’ is elusive, not least because of the multiple ways in which such trees may be ‘different’. AMOVA partitions nucleotide variance but does not capture phylogenetic structure. We address two aspects of differentiation, first whether phylogeographic distribution is structured or random, and second whether temporal phylogenetic patterns among populations are distinguishable. Harp Seals are an ideal species in which to test ideas about biogeography and intraspecific mitogenomic phylogeny, because the discrete

geographic distribution among a limited number of breeding areas provides several *a priori* hypotheses. A novel Monte Carlo method shows that different biogeographic models have greater or lesser departures from random expectation, and thus represent stronger or weaker hypotheses of structure. In other, more continuously distributed species, mitogenomic data offer the potential for population structures to be detected in more dispersed collections of individual genome-types. Our non-parametric analysis assesses cumulative mismatch curves as proxies for intra-specific genomic trees, by converting a complex branching topology to a continuous curve approximation of phylogenetic diversity over time. Although the K-S  $D$  statistic assesses the single maximum difference between curves, the maxima seen here are typically the high point on trending differences between curves. The method is shown here to be sensitive to subtle differences in temporal structure of populations of mixed phylogenetic composition.

## Materials & Methods

### Samples and DNA sequencing

Seal DNA samples include 37 out of 40 of those analyzed in [22] from the three breeding areas, including separate samples from the Newfoundland and Labrador coastal Front and the southern Gulf of St Lawrence, plus 16 new samples from the Front. All tissue samples were collected from seals taken on the respective whelping areas under scientific permits issued by the Norwegian and Russian governments, and by Department of Fisheries and Oceans scientists under permission granted by Section 52 of the Fisheries Act of Canada. All seals were killed according to the humane killing methods legally required in the respective countries, and were taken as part of ongoing research on population dynamics. No animals were killed specifically for the present study. DNA was also extracted from samples obtained by punch biopsy from the hind flippers of 16 additional seals on the Front in 1994 and 1999, by DFO scientists, under the terms and conditions of the Fisheries Act.

Complete 15,825 bp mitogenomes (without the D-Loop Control Region) were generated by the polymerase chain reaction with a set of 20 primer pairs that amplify the mitochondrial genome in fragments of 750~1400 bp, with overlaps between adjacent fragments of 80~300 bp. Of the 53 genomes, 45 were sequenced on a custom-designed multi-species iterative re-sequencing microarray, the “ArkChip” (Affymetrix) [50]. Our base-calling algorithm integrates differential fluorescence intensities of base-specific oligonucleotide binding to bases on either DNA strand with empirical rules for identification of SNPs [51]. The remaining eight genomes were sequenced with the BigDye chemistry v.2.0 (Applied Biosystems) on an ABI377 Prism automated sequencer. The entire set of 53 sequences was assembled with the Sequencher 4.9 program (GeneCodes, Ann Arbor MI). Both DNA strands were sequenced for all individuals.

### Phylogenetic analysis and calibration of the molecular clock

For the genomic data, a neighbor-joining (NJ) tree was identified from the absolute number of nucleotide differences (10,000 bootstrap replications) (MEGA5 [52]). A maximum parsimony (MP) tree was identified by heuristic search of equally-weighted SNP differences among haplotypes (TBR branching swapping and 100 taxon additions for identification of the minimum tree, 10,000 bootstrap replications with SPR swapping and 5 additions each) (PAUP\*, v.4.10 [53]). A maximum likelihood (ML) tree from a general time-reversible model was used to correct substitution rates for calculation of absolute times to MRCAs either by clock-constrained (PAUP\*) or linearized (MEGA5) models. A Bayes tree was calculated with the GTR substitution model and gamma-distributed rate variation over sites with allowance for invariable sites. A total of 2,000,000 generations were run, with a final standard deviation of the split

frequencies  $< 0.008$  (MrBayes, v.3.2 [54]). Pairwise mismatches and tests of the mismatch distribution were calculated with DnaSP v.5.1.10 [55].

In the absence of any well-defined recent biological or geological reference point for calibration of an intra-specific *Pagophilus* molecular clock [25], we estimated a clock-constrained maximum likelihood model by means of four hominines with complete mtDNA genome sequences and known MRCAs in the interval 70 ~ 1,000 KYA with respect to the senior author's mitogenome in Haplogroup U of macrohaplogroup N. These are Fr2 in Haplogroup A1 of macrohaplogroup M [9], San2 in Haplogroup L1 [1], a Neandertal [56], and a Denisovan [57]. The calibration was calculated for a 15,383 bp region homologous with the *Pagophilus* mitogenome alignment.

## Analysis of molecular variance

We used Arlequin v.3.5 [58] to calculate all pairwise  $\phi_{ST}$  values among populations based on nucleotide distances among haplotypes. We tested three partitions of within- and among-group variance ( $\phi_{SC}$  and  $\phi_{CT}$ , respectively) [59], corresponding to two-, three-, and four-island models of Harp Seal structure based on *a priori* east-to-west groupings of breeding grounds. We also tested a fourth *a posteriori* two-island model with the eastern-most White Sea breeding ground separated from the other three, based on the observation of significant  $\phi_{ST}$  between WS *versus* NL, and *versus* GS.

## Phylogeographic analysis by Monte Carlo simulation

We developed two new analytical methods and applied to the intraspecific phylogeographic tree. We first sought to determine how well the observed phylogeography fit various *a priori* models suggested by the geographic distribution of breeding populations, relative to the random distribution expected for such models. We performed a series of Monte Carlo randomizations of population assignments over the tree topology in Fig 2, as follows. We first used an Excel spreadsheet to generate 10,000 sets of 53 random numbers. For each set, we coded each number according to its rank in the set into a new matrix as a character state a, g, c, or t, in successive blocks with the same proportions as the observed population assignments (SG = a = 10, NL = g = 26, GS = c = 9, & WS = t = 8). Use of a, g, c, & t as character states allows the computationally efficient datatype = dna option of PAUP to be used. The resulting 53 x 10,000 matrix then corresponds to a PAUP file where each of 10,000 characters is a tree with the observed topology, but with population assignments randomized over the 53 branch tips. Given the topology of the observed tree (Fig 2) as a treefile, we next use PAUP to ascertain the number of dispersal events (L) required to explain the observed tree, *versus* each of the random assignments over that same topology. We finally ask where L of the observed tree falls relative to the distribution of L over the random assignments. Different phylogeographic models can be represented encoded as a set of 4x4 Character Type matrices that count the stepwise cost of dispersal events among populations.

We tested four models (Table 4): (a) an island model with unit cost for any event between populations [unordered character matrix], (b) a two-population model with a unit cost only for a trans-Atlantic event, (c) a four-population east-west stepping-stone model with unit cost for each step in the series SG  $\leftrightarrow$  NL  $\leftrightarrow$  GS  $\leftrightarrow$  WS, and (d) a three-population variant of Model (c), with a cost of 0 for an SG  $\leftrightarrow$  NL event. All matrices were symmetrical (bidirectional dispersal). Diagrams and matrices for each of the models are shown graphically in Fig 8.

## Non-parametric analysis of phylogenetic structure

Differences between population-specific cumulative pairwise difference curves were evaluated with a non-parametric, two-sample Kolmogorov-Smirnov (K-S) test procedure [60]. For a continuous two-population comparison, the method identifies the largest unsigned observed difference ( $D$ ) between the cumulative frequency distribution over the range of observed differences. Here, the cumulative curves are the cumulative numbers of pairwise differences in each sample, and  $D$  is the vertical offset between curves. The significance of  $D$  is evaluated by the critical value  $D_\alpha$ , which calculation includes counts of the number of comparisons within each population as a measure of the degrees of freedom. For pairwise data this would ordinarily be  $(n)(n-1)/2$ , however because the contributions of each individual to the set of pairwise difference comparisons are not independent,  $D_\alpha$  may more accurately be estimated from the total number of individuals  $n$  within both populations (J. Rohlf, pers. comm.) or by  $(n-1)(n'-1)/2$  for an  $n \times n'$  matrix. We evaluated  $D$  on all three criteria.

## Supporting Information

**S1 Table. Pairwise differences among mtDNA genomes of 53 Harp Seals (*Pagophilus groenlandicus*).**

(DOCX)

## Acknowledgments

We thank S. Amanda Marsh for assistance with microarray preparation, M. Hammill, T. Øritsland, and T. Haug for providing samples from the northeast Atlantic and Gulf of St. Lawrence, W. Penney, D. McKinnon and D. Wakeham for samples from the Front, Jim Rohlf for advice on the statistical analysis, and Linda Lait and several anonymous reviewers for comments on this and previous drafts of the MS.

## Author Contributions

Conceived and designed the experiments: SMC HDM GBS. Performed the experiments: SMC ATD. Analyzed the data: SMC HDM. Contributed reagents/materials/analysis tools: SMC GBS HDM. Wrote the paper: SMC GBS HDM.

## References

1. Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708–13. PMID: [11130070](#)
2. Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, et al. (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Amer J Hum Genet* 75: 910–918. PMID: [15382008](#)
3. Olivieri A, Pala M, Gandini F, Hooshiar Kashani B, Perego UA, Woodward SR, et al. (2013) Mitogenomes from two uncommon haplogroups mark late glacial/postglacial expansions from the near east and neolithic dispersals within Europe. *PLoS One* Jul 31; 8(7):e70492. doi: [10.1371/journal.pone.0070492](#) PMID: [23936216](#)
4. Carr SM, Marshall HD (2008a) Intraspecific phylogeographic genomics from multiple complete mtDNA genomes in Atlantic Cod (*Gadus morhua*): Origins of the “Codmother,” trans-Atlantic vicariance, and mid-glacial population expansion. *Genetics* 108: 381–389.
5. Carr SM, Marshall HD (2008b) Phylogeographic analysis of complete mtDNA genomes from Walleye Pollock shows an ancient origin of genetic biodiversity. *Mito DNA* 19: 490–496.
6. Jacobsen MW, Hansen MM, Orlando L, Bekkevold D, Bernatchez L, Willerslev E, et al. (2012) Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (*Coregonus* spp.). *Molec Ecol* 21: 2727–2742.

7. Morin PA, Archer FI, Goote AD, Vistrup J, Allen EE, Wade P, et al. (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Res* 20: 908–916. doi: [10.1101/gr.102954.109](https://doi.org/10.1101/gr.102954.109) PMID: [20413674](https://pubmed.ncbi.nlm.nih.gov/20413674/)
8. Archer FI, Morin PA, Hancock-Hanswer BL, Robertson KM, Leslie MS, Bérubé M, et al. (2013) Mitogenomic phylogenetics of Fin Whales (*Balaenoptera physalus* spp.): Genetic evidence for revision of sub-species. *PLoS ONE* doi: [10.1371/journal.pone.0063396](https://doi.org/10.1371/journal.pone.0063396)
9. Pope AM, Carr SM, Smith KN, Marshall HD (2011) Mitogenomic and microsatellite variation in descendants of the founder population of Newfoundland: high genetic diversity in an historically isolated population. *Genome* 54: 110–119. doi: [10.1139/G10-102](https://doi.org/10.1139/G10-102) PMID: [21326367](https://pubmed.ncbi.nlm.nih.gov/21326367/)
10. Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Mol Ecol* 7: 381–397. PMID: [9627999](https://pubmed.ncbi.nlm.nih.gov/9627999/)
11. Sergeant DE (1991) Harp seals, man and ice. *Can Spec Pub Fish Aquat Sci* 114, 153pp.
12. Sergeant DE (1973) Transatlantic migration of a harp seal, *Pagophilus groenlandicus*. *J Fish Res Board Canada* 30: 124–125.
13. Øien N, Øritsland T (1995) Use of mark-recapture experiments to monitor seal populations subject to catching. In: Blix L, Walle L, Ulltang Ø, editors. *Whales, Seals, Fish and Man*. Amsterdam: Elsevier Science BV. pp. 35–45.
14. Kapel FO (1996) Recoveries in Greenland, 1949–94, of tagged or branded harp and hooded seals. *Northw Atl Fish Org Sci Coun Stud* 26: 87–99.
15. Stenson GB, Sjare B (1997) Seasonal distribution of Harp seals, *Phoca groenlandica*, in the Northwest Atlantic. International Council for the Exploration of the Sea Council Meeting document 1997/CC: 10 23pp.
16. Sergeant DE (1965) Migrations of harp seals *Pagophilus groenlandicus* (Erleben) in the northwest Atlantic. *J Fish Res Board Canada* 22, 433–464.
17. Sergeant DE (1976) History and present status of populations of harp and hooded seals. *Biol Cons* 10: 95–118.
18. Bowen WD, Sergeant DE (1983) Mark-recapture estimates of harp seal pup (*Phoca groenlandica*) production in the northwest Atlantic. *Can J Fish Aquat Sci* 40: 728–742.
19. Bowen WD, Sergeant DE (1985) A mark-recapture estimate of 1983 harp seal pup production in the Northwest Atlantic. NAFO SCR Document 85/1/1, Serial Number N935, 14pp.
20. Kjellqwist S, Heide-Jorgensen C, Øritsland T (1995) Trends in age-composition, growth and reproductive parameters of Barents Sea harp seals, *Phoca groenlandica*. *ICES J Mar Sci* 52: 197.
21. Nilssen K, Haug T, Øritsland T, Lindblom L, Kjellqwist S (1998) Invasions of harp seals *Phoca groenlandica* Erleben to coastal waters of Norway in 1995: ecological and demographic implications. *Sarsia* 83: 337–345.
22. Perry EA, Stenson GB, Bartlett SE, Davidson WS, Carr SM (2000) DNA sequence analysis identifies genetically distinguishable populations of harp seals (*Pagophilus groenlandicus*) in the northwest and northeast Atlantic. *Mar Biol* 137: 53–58.
23. Davis CS, Stirling I, Strobeck C, Coltman DW (2008) Population structure of ice-breeding seals. *Molec Ecol* 17:3078–94.
24. Coltman DW, Stenson G, Hammill MO, Haug T, Davis CS, Fulton TL (2007) Panmictic population structure in the hooded seal (*Cystophora cristata*). *Molecular Ecology*, 16: 1639–48. PMID: [17402979](https://pubmed.ncbi.nlm.nih.gov/17402979/)
25. Árnason U, Gullberg A, Janke A, Kullberg M, Lehman N, Petrov EA, et al. (2006) Pinniped phylogeny and a new hypothesis for their origin and dispersal. *Mol Phylo Evol* 41: 345–54.
26. Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, 133: 693–709. PMID: [8454210](https://pubmed.ncbi.nlm.nih.gov/8454210/)
27. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595. PMID: [2513255](https://pubmed.ncbi.nlm.nih.gov/2513255/)
28. Avise J (2000) *Phylogeography: The History and Formation of Species*. Cambridge: Harvard University Press.
29. Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, et al. (2003) Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* 100:171–176. PMID: [12509511](https://pubmed.ncbi.nlm.nih.gov/12509511/)
30. Perelman P, Johnson WE, Roos C, Seuánez HN, Horvath JE, Moreira MS, et al. (2011) A molecular phylogeny of living primates. *PLoS Genet* 7(3): e1001342. doi: [10.1371/journal.pgen.1001342](https://doi.org/10.1371/journal.pgen.1001342) PMID: [21436896](https://pubmed.ncbi.nlm.nih.gov/21436896/)
31. Fulton TA, Strobeck CS (2010) Multiple fossil calibrations, nuclear loci and mitochondrial genomes provide new insight into biogeography and divergence timing for true seals (Phocidae, Pinnipedia). *J Biogeog* 37: 814–829.

32. Vermeij GJ (1991) Anatomy of an invasion: the trans-Arctic interchange. *Paleobiology* 17: 281–307.
33. Coulson MW, Marshall HD, Pepin P, Carr SM (2006) Mitochondrial genomics of gadine fish: implications for taxonomy and biogeographic origins from whole-genome data sets. *Genome* 49: 1315–1320.
34. Hammill MO, Stenson GB, Kingsley MCS (2011) Historical Abundance of Northwest Atlantic harp seals (*Pagophilus groenlandicus*): influence of harvesting and climate. Canadian Science Advisory Secretariat Research Document 2011/100. iv + 16 pp.
35. Hammill MO, Stenson GB, Doniol-Valcroze T, Mosnier A (2012) Estimating carrying capacity and population trends of Northwest Atlantic harp seals, 1952–2012. Canadian Science Advisory Secretariat Research Document 2012/148. iii + 31 p.
36. Vinnikov K, Robock A, Stouffer R, Walsh JE, Parkinson CL, Cavalieri DJ, et al. (1999) Global warming and northern hemisphere sea ice extent. *Science* 286: 1934–1937. PMID: [10583952](#)
37. McCarthy JJ, Canzani OF, Leary NA, Dokken DJ, White KS, editors (2001) Climate change 2001: Impacts, Adaptation, and Vulnerability. Contribution of working group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press.
38. International Panel on Climate Change (2007) Fourth assessment report. International Panel on Climate Change (IPCC Secretariat, c/o World Meteorological Organization, Geneva, Switzerland. Available: <http://www.ipcc.ch>.
39. Walsh JE (2008) Climate of the Arctic marine environment. *Ecol Appl* 18 (Suppl.): S3–S22. PMID: [18494360](#)
40. Blunden J, Arndt DS, editors (2012) State of the Climate in 2011. *Bull Amer Meteorol Soc* 93: S1–S264.
41. International Council for the Exploration of the Sea (ICES) (2011) Report of the Working Group on Harp and Hooded Seals (WGHARP). ICES Council Memorandum 2011/ACOM: 22.
42. Øligård TA, Lindstrøm U, Haug T, Nilssen KT, Smout S (2013) Functional relationship between harp seal body condition and available prey in the Barents Sea. *Mar Ecol Prog Ser* 484: 287–301.
43. Wilkinson JP, Wadhams P (2005) A method of detecting change in the ice conditions of the central Greenland Sea by the whelping locations of harp seals. *J Climate* 18: 1216–1226.
44. DFO (2007) A review of ice conditions and potential impact on harp seal neonatal mortality in March 2007. DFO Canadian Science Advisory Secretariat Science Response 2007/008.
45. Bajzak CE, Hammill MO, Stenson GB, Prinsenber S (2011) Drifting away: Implications of changes in ice conditions for a pack-ice breeding phocid, the harp seal (*Pagophilus groenlandicus*). *Can J Zool* 89: 1050–1062.
46. Johnston DW, Bowers MT, Friedlaender AS, Lavigne DM (2012) The effects of climate change on harp seals (*Pagophilus groenlandicus*). *PLoS One* 7: e29158. doi: [10.1371/journal.pone.0029158](https://doi.org/10.1371/journal.pone.0029158) PMID: [22238591](#)
47. Stenson GB, Hammill MO (2012) Living on the edge: Observations of Northwest Atlantic harp seals in 2010 and 2011. *Can Sci Adv Sec Res Doc* 2011/108. iv + 12 p.
48. Rosing-Asvid A (2008) A new harp seal whelping ground near South Greenland. *Mar Mammal Sci* 24: 730–736.
49. Stenson GB (2009) Total removals of Northwest Atlantic harp seals (*Pagophilus groenlandicus*), 1952–2009. *Can Sci Adv Sec Res Doc* 2009/112.
50. Carr SM, Marshall HD, Duggan AT, Flynn SMC, Johnstone KA, Pope AM, et al. (2008c) Phylogeographic genomics of mitochondrial DNA: patterns of intraspecific evolution and a multi-species, microarray-based DNA sequencing strategy for biodiversity studies. *Comp Biochem Physiol D: Genom Proteom* 3: 1–11.
51. Flynn SMC, Carr SM (2007) Interspecies hybridization on DNA resequencing microarrays: efficiency of sequence recovery and accuracy of SNP detection in human, ape, and codfish mitochondrial DNA genomes sequenced on a human-specific MitoChip. *BMC Genomics* 8: 339. PMID: [17894875](#)
52. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28: 2731–2739. doi: [10.1093/molbev/msr121](https://doi.org/10.1093/molbev/msr121) PMID: [21546353](#)
53. Swofford DL (1998) PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods) Version 4. Sinauer Associates, Sunderland, Massachusetts.
54. Ronquist F, Teslenko M, van der Mark P, Ayres D, Darling A, Höhna S, et al. (2011) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61: 539–542.

55. Rozas J, Rozas R (1995) DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *Comp Appl Biosci* 11: 621–625. PMID: [8808578](#)
56. Krause J, Fu Q, Good JM, Viola B, Shunkov MV, Derevianko AP, et al. (2010) The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* 464: 894–897. doi: [10.1038/nature08976](#) PMID: [20336068](#)
57. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, et al. (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468, 1053–60. doi: [10.1038/nature09710](#) PMID: [21179161](#)
58. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resources* 10: 564–567.
59. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491. PMID: [1644282](#)
60. Sokal RS, Rohlf JF (2011) *Biometry*, 4th ed. New York: WH Freeman.