SC/69A/NH/02

Sub-committees/working group name: NH

Determining the number of breeding populations of humpback whales (Megaptera novaeangliae) in the North Atlantic and levels of mixing on the Central and Northeast Atlantic summer range

Palsbøll, Bérubé, Gabualdi, Robbins, Berrow, Chalifour, Clapham, Gose, Kovacs, Lopez-Suarez, Lydersen, Magnúsdóttir, Pace, Pampoulie, Rasmussen, Ryan, Schweinsberg, Sigurjónsson, Kizildas, Tollrian, Vely, Vikingsson, Wenzel, Van Der Zee, Øien

INTERNATIONAL
WHALING COMMISSION

Determining the number of breeding populations of humpback whales (*Megaptera novaeangliae*) in the North Atlantic and levels of mixing on the Central and Northeast Atlantic summer range

Per J. Palsbøll[1,2$], Martine Bérubé[1,2], Mario Gabualdi[1], Jooke Robbins[2], Simon Berrow[3], Julien Chalifour[4], Phillip J. Clapham[5], Marc-Alexander Gose[6], Kit M. Kovacs[7], Pedro Lopez-Suarez[8], Christian Lydersen[7], Edda E. Magnúsdóttir[9], Richard M. Pace III[10], Christophe Pampoulie[11], Marianne Rasmussen[9], Conor Ryan[3], Maximilian Schweinsberg[6], Jóhann Sigurjónsson[11], Yakamoz Kizildas[1], Ralph Tollrian[6], Michel Vely[12], Gisli Vikingsson[11], Frederick Wenzel[10], Jurjan van der Zee[1], and Nils Øien[13]

[1]University of Groningen, Groningen, The Netherlands

[2]Center for Coastal Studies, Massachusetts, U.S.A.

[3]The Irish Whale and Dolphin Group, Kilrush, Co. Clare, Ireland and Marine & Atlantic Technological University, Galway, Ireland

[4]Réserve Naturelle Nationale de Saint-Martin, Hope Estate, Saint-Martin, French West Indies

[5]Seastar Scientific, Vashon Island, Washington, USA.

[6]Ruhr-Universität Bochum, Bochum, Germany

[7]Norwegian Polar Institute, Tromsø, Norway

[8]Bios, Cape Verde

[9]University of Iceland, Reykjavik, Iceland

[10]NOAA Northeast Fisheries Science Center, Woods Hole, Massachusetts, U.S.A

[11]Marine and Freshwater Research Institute, Hafnarfjörður, Iceland

[12]Megaptera, Saint Barth, Saint-Martin, French West Indies

[13]Norwegian Marine Research Institute, Bergen, Norway

[$]Corresponding author: p.j.palsboll@rug.nl

Abstract

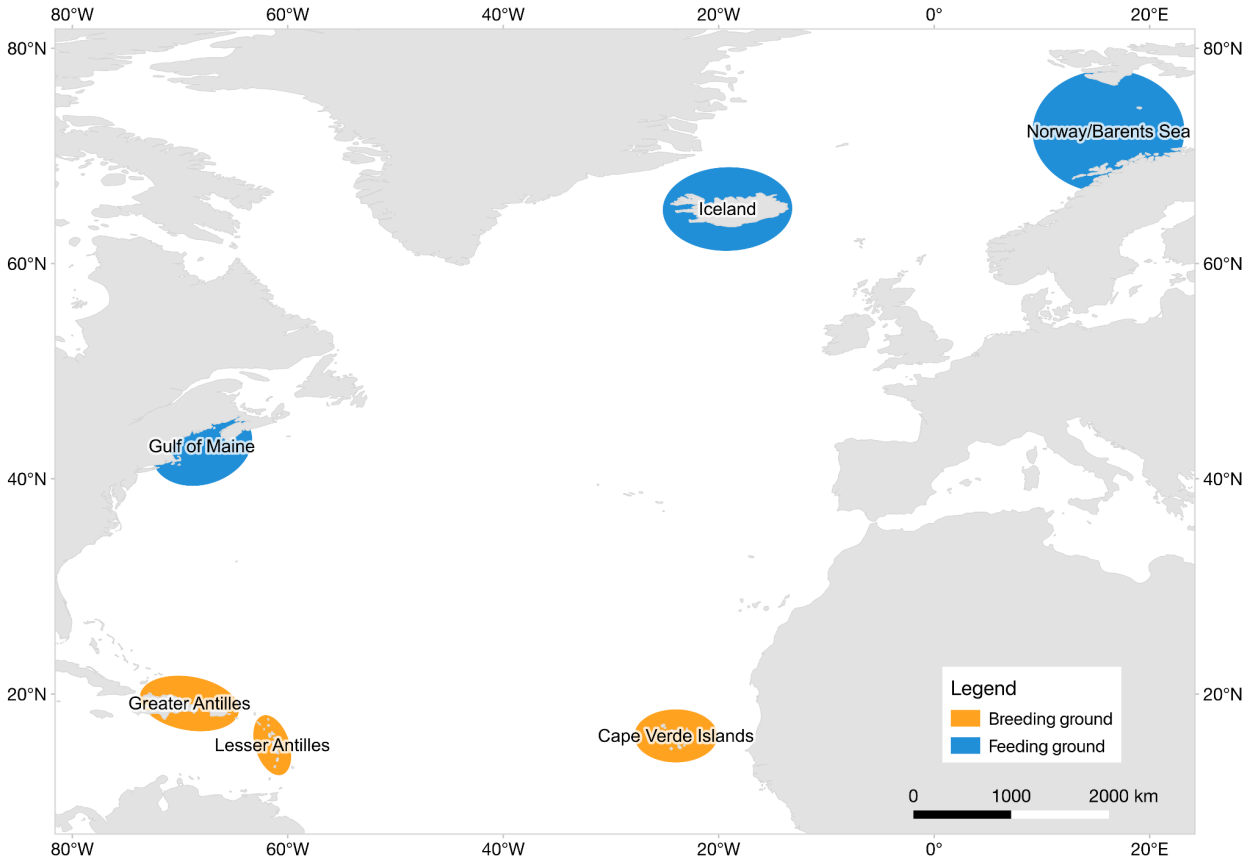
We estimated the degree of divergence and inferred the rate of contemporary connectivity among humpback whale breeding populations in the North Atlantic. Analysis was based on traditional genetic markers (simple tandem repeat loci, STRs and mitochondrial control region DNA sequences) as well as genome data (10,308 SNPs across the genome and entire mitochondrial genome DNA sequences). A single predominantly eastern North Atlantic cluster appears to be confined to winter and summer ranges within the eastern North Atlantic and likely represents the remnants of what used to be a historically isolated eastern North Atlantic breeding population. There is a 10-25% rate of immigration of individuals of western origin into the eastern breeding population. The magnitude of this introgression suggests that it is a recent phenomenon because such a high immigration rate over the long-term would have erased any signal of west-east genetic structure. Individuals with a 50 - 100 % eastern North Atlantic origin can be identified easily by genotyping 20 STR loci. Hence, it is possible to monitor if mortalities are from this much smaller population, and potentially to determine whether certain areas are critical to individuals with an eastern origin.

Introduction

North Atlantic humpback whales feed in temperate and subarctic waters ranging from the eastern North American coast to northern Norway. In the winter, they congregate to breed at low latitudes in the western North Atlantic (the Greater and Lesser Antilles of the Caribbean) and the eastern North Atlantic (the Cape Verde Islands, CVI, off western Africa, Fig, 1). Commercial whaling for North Atlantic humpback whales peaked in the early 20th century and took place primarily on their winter breeding grounds. Whaling records and population assessment modeling suggest that the pre-whaling abundance of humpback whales was ~25,000 in the Caribbean and ~5,000 off western Africa (Punt et al., 2006). By contrast, the latest capture-mark-recapture estimates of humpback whale abundance in these areas differ by a far greater extent: 11,570 (95% CI: 10,290 - 13,390, Stevick et al., 2003) in the Greater Antilles versus only 272 (SE 10, Wenzel et al., 2020) at CVI. These differential recovery rates suggest that there is limited connectivity between the western and eastern North Atlantic breeding areas. However, some photographically-identified individual humpback whales have been sighted both at CVI and in the Lesser Antilles (Stevick et al., 2016). In addition, photo-id data as well as tracks of radio satellite-tagged humpback whales suggested some exchange between the Greater and Lesser Antilles (Kennedy et al., 2014; Mackay et al., 2019; Stevick et al., 1999). There is also evidence of behavioral differences between the Greater Antilles and Lesser Antilles, with the latter exhibiting later breeding timing and preferential exchange with eastern North Atlantic feeding grounds (Stevick et al., 2018). The implications of these findings for North Atlantic breeding population structure have not previously been investigated. Here, we estimate the degree of genetic divergence and infer the rate of contemporary connectivity between the

western and eastern North Atlantic breeding grounds. We also evaluate whether there is genetic evidence for more than one breeding population in the eastern North Atlantic (see discussion in Smith & Pike, 2009).

Figure 1. Feeding and breeding grounds in the North Atlantic where samples were collected.



Materials and Methods

Sampling areas

This study focused on samples collected at three primary feeding grounds and three breeding areas shown in Figure 1. The samples from the Lesser Antilles were collected in the waters off St. Martin, and from the Greater Antilles on Silver Bank, north of the Dominican Republic.

Samples, preservation, storage and DNA extraction

Tissue samples were collected as skin biopsies from free-ranging or stranded (few) humpback whales as previously described (Palsbøll et al., 1991). Samples were preserved in liquid nitrogen,

70% ethanol or 5M NaCl with 25% DMSO (Amos & Hoelzel, 1991) and stored at -80 or -20 degrees Celsius (℃). Total-cell DNA was extracted either using phenol-chloroform extraction (Sambrook & Russell, 2001) or columns (e.g., QIAGEN DNEasy™).

Data generation

The first 400 nucleotides of the mitochondrial control region (the most variable part) were PCR-amplified (polymerase chain reaction, Mullis & Faloona, 1987), using the primers MT4F (Arnason et al., 1993) and BP16071R (Drouot et al., 2004). The initial PCR amplifications were performed in a 10 µL volume comprising 0.2 µM of each dNTP, 67mM Tris-HCl (pH 8.8), 2mM $MgCl_2$, 17mM NH_3SO_4, 10mM β-mercaptoethanol, 0.1µM of each primer, 0.4 units of Taq DNA polymerase and approximately 10 - 20 ng of DNA extraction. The thermocycling conditions were: 2 min at 94° C, followed by 25 cycles each consisting of 15 sec. at 94°C, 30 sec. at 54°C and 120 sec. at 72°C. After amplification, unincorporated nucleotides and excess primers were enzymatically removed using shrimp alkaline phosphatase and exonuclease I as described by Werle et al. (1994). The cleaned PCR amplification products were sequenced using fluorescently labeled ddNTPs according to the manufacturer's instructions (Big Dye™ v3.1 Terminator Ready Reaction Mix, Life Technologies Inc.), using either the primers MT4F or BP16071R. Excess dideoxy-terminator nucleotides were removed by ethanol/EDTA precipitation and re-suspended in 10 µL deionized formamide (Calbiochem Inc.). The order of sequencing products was resolved by capillary electrophoresis using an ABI Prism™ 3730 sequencer (Applied Biosystems Inc.).

Samples were genotyped at 20 microsatellite loci Table 2: AC087 (Bérubé et al., 2005), EV001, EV037, EV094, EV096 (Valsecchi & Amos, 1996), GATA028, GATA098, GATA053, GATA417, TAA031 (Palsbøll et al., 1997), GT011 (Bérubé et al., 1998), GT015, GT023, GT101, GT195, GT211, GT271, GT575 (Bérubé et al., 2000), GATA43950, GATA97408 and a Y-chromosome specific marker (Suárez-Menéndez, Bérubé, unpublished).

Samples were genotyped in multiplex PCR reactions (between six to eight microsatellite loci per amplification), using the MM2X™ Multiplex kit Plus (Qiagen Inc.) in 5µL reaction volumes. The thermocycling conditions were: 2 minutes (min.) at 94°C, followed by 35 cycles each of 30 seconds (sec.). at 94°C, 90 sec. at 57°C and 30 sec. at 72°C followed by a final cycle of 10 min. at 68°C. The PCR products were separated by capillary electrophoresis using an ABI Prism™ 3730 (Life Technologies Inc.). The size of the amplification products was estimated against a Genescan™ ROX-500 size standard (Life Technologies Inc.) in the software GENEMAPPER™ (v.4.0; Life Technologies Inc.).

Double-digest Restriction Aided Digestion sequencing (ddRADSeq, Franchini et al., 2017; Peterson et al., 2012) was employed to generate genome-wide data for SNP genotyping. The generation of the ddRADSeq libraries was previously described by Cabrera *et al* (2022). Briefly, genomic DNA was digested with *Msp*I and *Hind*III, i5 and i7 Illumina™-compatible adapters, each with inline i7 and i5 barcodes of 6-8 nucleotides, were ligated (in unique combinations) to

the digested DNA. Using a PippinPrep™ (Sage Science Inc.), fragments between 300 - 400 base pairs were isolated and unique i7 indexes added to each sub-library by PCR amplification. Pooled sub-libraries were sequenced on Illumina™ sequencers model HISEQ™ 2500 or NOVASEQ™, in paired-end mode at 125 or 150 cycles, respectively, at The Center for Applied Genomics (Toronto, Ontario, Canada).

The mitogenome library preparation, sequencing and bioinformatic analyses were performed as described by Suárez-Menéndez *et al.* (2022). Briefly sequencing libraries were generated using *in solution* DNA capture with custom RNA baits (Arbor Biosciences Inc.). Genomic DNA was degraded with Fragmentase, blunt-end repaired and Illumina™ compatible adapter ligated onto fragments. Unique dual combinations of i7 and i5 were added by PCR amplification. Sequencing was conducted as described above. Mitochondrial genome sequencing libraries were assembled by first aligning against the blue whale (*Balaenoptera musculus*) assembly (NCBI assembly GCA_009873245.3, Bukhman et al., 2022) in order to remove reads that aligned against regions in the nuclear genome. Subsequently, unmapped reads were aligned against the humpback whale mitochondrial genome (NCBI assembly NC_006927.1, Arnason et al., 1993).

Data analyses

The genome coordinates of each microsatellite locus were inferred by aligning the oligo-nucleoitides employed to genotype each locus against the blue whale genome assembly using BOWTIE2 (v. 2.5.0, Langmead & Salzberg, 2012).

The minimum number of identical microsatellite locus genotypes necessary to identify samples from the same individual was determined from the probability of identity (*I*, Paetkau & Strobeck, 1994) for full-siblings as the unique combination of loci with the highest value of *I* yieldeding n number of expected chance matches (assuming all pairs of samples were related as full siblings) below 0.01. Full-siblings have the highest probability of genotype identity, although they constitute a very small fraction of the total number of possible pairs (Rew et al., 2011).

The software MLRELATE (v. 04/2008, Kalinowski et al., 2006) was employed to identify putative parent and offspring dyads (from 20 STR loci) sampled in close proximity (i.e., with close serial sampling numbers in the same year and sampling area). Such dyads were likely associated (and hence sampled) in a non-independent manner, and one member of each pair was excluded from the downstream analyses. However, the putative parent and offspring dyads detected where the two individuals were sampled in different years or allocated to different sample partitions were retained (Waples & Anderson, 2017).

All but one sample inferred to originate from the same individual (identical genotypes at >= 16 STR loci, based on *I*, see Results section) within the same sample partition were removed. SNP genotypes from the ddRADSeq data were generated as outlined in Table 1.

Table 1: Overview of generation of SNP data from *fastq* files (delivered demultiplexed to index)

No.	Task	Software & scripts	Parameter settings
1	Removal of PCR duplicates	STACKS[1] *clone_filter*	*--inline_inline --oligo_len_1 4 --oligo_len_2 4*
2	Assigning indexed *fastq* files to sample IDs	STACKS[1] *process_radtags* Unique i5 and i7 inline barcode combinations	Default (no RAD-site rescue)
3	Reference alignment	BOWTIE2[2] and *Balaenoptera musculus* genome assembly[3]	*–very-sensitive*
4	Remove samples with few reads/poor alignment		Samples with < 2.5 million reads or < .85 alignment rate
5	Genotype catalog	STACKS[1] *gstacks*	*–min-mapq 30 –model marukilow*
	Generation of an initial blacklist of RAD loci		
6	Initial SNP dataset	STACKS[1] *populations*	*-p 1 -R 0.5 -r 0.5 --min-mac 3 -write-single-snp* Samples with > 25 million reads from two initial libraries (from CV[4] and WI[5] as one population)
7	Blacklist proximate loci	PYTHON[7] custom code with *populations.sumstats.tsv* as data source	Blacklist RAD loci with: a) > 2 alleles b) closer than 10,000 bps[6]
	Removal of samples (from population map)		

8	Individual genotyping rates	VCFTOOLS[8]	*--missing-indv*
9	Remove duplicates samples (in same sample partition)	PYTHON[7] custom code using microsatellite genotypes (20 loci)	*Samples with identical genotypes at >= 16 loci*. Sample with lowest genotyping rate added to blacklist
10	Remove proximately sampled close relatives	MLRELATE[9] using microsatellite genotypes (20 loci)	*Parent-offspring pairs.* Sample with lowest genotyping rate added to blacklist
11	Update SNP data	STACKS[1] *populations* (filtering out blacklisted proximate loci and samples)	*-write-single-snp -p 1 -R 0.85 -r 0.85 --min-mac 3 --blacklist <blacklist file>* Remaining samples from two initial libraries (from CV[4] and WI[5] as one population)
12	Remove samples high missing genotype rate	VCFTOOLS[8]	*--missing-indv* Remove samples with missing rate >15%
13	Identify possible batch effect loci from initial whitelist	STACKS[1] *populations*	*–write-single-snp -p 2 -R 0.85 -r 0.85 --min-mac 3 –whitelist <whitelist file> -M <popmap file>* Popmap file: Samples with > 25 million reads from all libraries (CV[4] and WI[5] as one and the remaining libraries, with samples from IL[10] and NE[11], as the other population)
14	Subtract high-F_{ST} loci from whitelist	PYTHON[7] custom code with *populations fst* file as data source	Remove RAD loci with $F_{ST} > 0.3$

15	Update SNP data	STACKS[1] *populations*	*-write-single-snp -p 1 -R 0.85 -r 0.85 --min-mac 3 –whitelist <whitelist file>* All samples as part of one population
16	Removal of duplicate and proximately sampled parent-offspring pairs	As described in steps 8 - 11	All samples as part of one population
	Final data set after filtering for whitelisted loci		
17	Final SNP dataset	STACKS[1] *populations*	*-write-single-snp -p 1 -R 0.85 -r 0.85 --min-mac 3 –whitelist* All samples as part of one population (sample locations added later during analysis)

Notes: [1]v. 2.62 (Catchen et al., 2011); [2]v. 2.5.0 (Langmead & Salzberg, 2012); [3] NCBI assembly GCA_009873245.3; [4]Cape Verde; [5]West Indies; [6]base pairs; [7]v. 3.11; [8]v. 0.16 (Danecek et al., 2011); [9]v. 04/2008 (Kalinowski et al., 2006); [10]Iceland; [11]Northeast Atlantic (Barents Sea).

The degree of genetic divergence was estimated as Weir and Cockerham's θ (θ_{WC}, Weir & Cockerham, 1984) for nuclear ddRADSeq SNP and STR genotypes using the CRAN R (v. 4.2.2) package *diversity* (v. 1.9.90, Keenan et al., 2013) or Hudson's H_{ST} (Hudson et al., 1992) using DNASP (v. 6, Rozas et al., 2017) in case of mitochondrial DNA sequence data. The 95% confidence interval of θ_W was estimated from 1,000 bootstrap samples (resampling individuals) for nuclear loci as implemented in *diversity*. The probability of the observed value, assuming the data were collected from the same panmictic population, was estimated as implemented in the CRAN R package *genepop* (nuclear loci, exact G test using 10,000 dememorizations, 100 batches and 1,000 iterations, v. 1.2.2, Rousset, 2008) and DNASP (mitochondrial DNA sequences, χ^2 with 1,000 permutations).

The software STRUCTURE (v. 2.3.4, Falush et al., 2007) was employed to assess the presence of multiple distinct "gene pools" among the STR and the ddRADSeq SNP genotypes. Estimates were conducted assuming the admixture model and correlated allele frequencies from 15 replicates for each prior value of *K*. The sampling locations were employed as population priors. The initial one-third of the Markov Chain Monte Carlo (MCMC) iterations were discarded as burn-in. The number of MCMC iterations was increased until three separate estimations yielded similar estimates of *K* from the likelihood and Evanno's ΔK (Evanno et al., 2005) as well as similar admixture proportions among estimations. The likelihood of *K*, Evanno's ΔK and admixture proportions were estimated and depicted using *pophelperShiny* (v. 2.1.1, Francis, 2017). λ was estimated as the mean of three estimations conducted as described above separately for ddRADSeq and STR genotype-based data sets, as per the STRUCTURE manual.

BAYESASS (v. 3.0.4, Wilson & Rannala, 2003) and BA3-SNPs (v. 1.1, Mussmann et al., 2019) were employed to estimate current dispersal rates (*m*, the fraction of immigrant individuals in the targeted population(s)), for STR and SNP genotypes, respectively between western Caribbean and western Africa. Preliminary estimations based on 10,000 - 100,000 iterations were employed to adjust the three mixing parameters (Δ_M, Δ_A, Δ_F, for migration, allele frequencies and inbreeding coefficients, respectively) to be within the range between 0.35 - 0.45 as recommended by Mussman *et al.* (2019). The number of iterations were increased until three estimations initiated with different random seed numbers yielded very similar estimates of *m*.

Results

Mapping microsatellite loci

The 20 microsatellite loci mapped to 13 different scaffolds (Table 2). The loci aligned to the same scaffold were mapped between 9 million and 150 million base pairs apart.

Table 2. Mapping coordinates of microsatellite loci to the blue whale genome assembly

Assembly scaffold	Locus	Position	Base pairs to next locus
CM020941.2	EV096	51,709,452	
CM020942.2	GATA97408	166,113,841	
CM020943.2	GATA43950	2,789,293	150,060,452
	GATA417	152,849,745	
CM020944.2	EV001	56,211,193	
CM020945.2	EV037	7,650,033	113,851,625
	GT023	121,501,658	16,393,434
	GATA053	137,895,092	
CM020946.2	GT575	31,067,623	37,862,610
	GT271	68,930,233	
CM020947.2	AC087	78,883,199	31,626,126
	GT015	110,509,325	
CM020948.2	GATA028	40,438,231	
CM020950.2	GATA098	67,162,158	
CM020954.2	GT011	8,682,630	9,436,468
	EV094	18,119,098	
CM020958.2	GT195	24,140,172	39,767,629
	GT211	63,907,801	
CM020959.2	GT101	22,160,101	
CM020960.2	TAA031	37,325,733	

Mitochondrial DNA data and divergences

The mean read depth was 4,026 reads (95 percentile range: 7,246 - 819). The final size of the mitochondrial genome assembly was 16,400 base pairs including insertions and deletions. These latter sites were excluded from the estimates of diversity and divergence, which was based upon 16,385 base pair long sequences. The final data set comprised 32 sequences from CVI, 32 from the Greater Antilles, 12 from the Lesser Antilles as well as 68 from Iceland and 71 from

Norway/Barents Sea. H_{ST} among all sampling areas was estimated at 0.098 ($P < 0.001$). Estimates of divergence, diversity and the probabilities of homogeneity were identical in terms of trends between entire mitochondrial genome sequences and the first 450 base pairs of the mitochondrial control region, mainly because 20% (34/173) of the segregating sites are located within the first 450 base pairs of the mitochondrial control region, resulting in 22 haplotypes, whereas 33 haplotypes were detected among the whole mitochondrial genome sequences. The trend in both mitochondrial DNA sequence datasets was an increase in genetic divergence with increasing geographic distance for the winter and summer ranges, with Iceland in an intermediate position. The levels of diversity in the two data sets are tabulated in Table 3, and estimates of genetic divergence and sampling area diversity in Tables 4 and 5. In general (apart from the more extreme values) the estimated degree of divergence was similar between the two data sets.

Table 3. Diversity of mitochondrial data for whole genome sequences and the control region

	Entire mitochondrial genome	Mitochondrial control region (450 base pairs)
Number of sequences	215	215
Number of segregating sites	173	34
Number of haplotypes	33	22
Haplotype diversity	0.84	0.75
Avg. number of differences	27.1	5.9
Nucleotide diversity[1]	0.0017	0.013

Notes: [1](Nei & Li, 1979)

Table 4. Divergence (F_{WC}) and diversity between and within sampling areas based on whole mitochondrial genome sequences

	Winter breeding areas			Summer feeding areas	
	Greater Antilles	Lesser Antilles	CVI	Iceland	Norway/ Barents Sea
Greater Antilles	32/13/112/ 0.75/0.0027	0.072*	0.13***	0.075***	0.18***
Lesser Antilles		12/9/86/ 0.94/0.0011	0.0058	-0.0051	0.032***
CVI			32/10/129 /0.81/0.0011	0.013*	0.028**
Iceland				68/25/100/ 0.90/0.0016	0.042**
Norway/Barents Sea					71/11/81/ 0.60/0.0007

Notes: Numbers along the diagonal represent intra-area diversity, [sample size]/[no. haplotypes]/[no. segregating sites\/[haplotype diversity]/[nucleotide diversity]. No asterisk denotes P>0.05, * 0.01<P<0.05, ** 0.001<P<0.01, and *** P<0.001.

Table 5. Divergence (H_{ST}) and diversity between and within sampling areas estimated using the first 450 base pairs of the 5' end of the mitochondrial control region

	Winter breeding areas			Summer feeding areas	
	Greater Antilles	Lesser Antilles	CVI	Iceland	Norway/ Barents Sea
Greater Antilles	32/8/21/ 0.71/0.018	0.078*	0.13***	0.084***	0.18***
Lesser Antilles		12/7/20/ 0.83/0.012	0.0092	-0.0055	0.023**
CVI			32/9/28/ 0.75/0.010	0.025**	0.059**
Iceland				68/16/24/ 0.79/0.013	0.029**
Norway/Barents Sea					71/6/ 0.49/0.0060

Notes: Numbers along the diagonal represent intra-area diversity, [sample size]/[no. haplotypes]/[no. segregating sites/[haplotype diversity]/[nucleotide diversity]. No asterisk denotes $P>0.05$, * $0.01<P<0.05$, ** $0.001<P<0.01$, and *** $P<0.001$.

Nuclear DNA data and divergences

ddRADSeq was conducted for a total of 278 skin biopsy samples; 72 from Norway/Barents Sea, 72 from Iceland, and 49 from the Gulf of Maine, as well as 35 sampled in CVI, 12 from the Lesser Antilles and 38 from the Greater Antilles. After the bioinformatic analysis the final data set comprised 203 samples; 54 from Norway/Barents Sea, 59 from Iceland, and 38 from the Gulf of Maine, as well as 19 from CVI and 33 from the Greater Antilles. These samples were genotyped at 10,308 SNPs, situated at least 10,000 base pairs apart and genotyped in minimum 85% samples.

STR-based analyses were performed on a total of 793 unique and complete 20-locus (Table 2) genotypes including; 42 from CVI, 95 from the Gulf of Maine, 209 from Iceland, 344 from Norway/Barents Sea, 12 from the Lesser Antilles and 91 from the Greater Antilles.

The expected number of pairs of different individuals matching at 16 loci between individuals related as full siblings (the relationship with the highest *I*) was estimated at less than 0.01 pairs (Table 5). Accordingly, a match at 16 loci was employed as the threshold for identifying individuals. Thus, all specimens with identical microsatellite genotypes at a minimum of 16 loci were inferred as duplicate specimens collected from the same individual.

The overall picture, i.e., the spatial configuration of genetic divergence for the STR and ddRADSeq genotype data was similar between the two datasets as well as the observed in the mitochondrial sequence data sets (Tables 7 & 8). Contrary to the mitochondrial DNA data, the values of divergence estimates differed in most cases.

Table 6. Probability of identity for 793 samples estimated from the STR genotypes

# Loci[1]	Locus	I_{FS}[1]	$\prod I_{FS}$	Expected FS matches[2]	I_{PO}[1]	$\prod I_{PO}$	Expected PO matches[2]	I_{UR}[1]	$\prod I_{UR}$	Expected UR matches[2]
1	GATA 028	0.512	5.13E-01	1.61E+05	0.525	5.25E-01	1.65E+05	0.3	3.00E-01	9.43E+04
2	EV 001	0.471	2.41E-01	7.58E+04	0.442	2.32E-01	7.29E+04	0.239	7.17E-02	2.25E+04
3	GT 271	0.472	1.14E-01	3.58E+04	0.444	1.03E-01	3.24E+04	0.228	1.64E-02	5.14E+03
4	GT 195	0.423	4.82E-02	1.51E+04	0.346	3.57E-02	1.12E+04	0.19	3.11E-03	9.78E+02
5	EV 094	0.407	1.96E-02	6.16E+03	0.314	1.12E-02	3.52E+03	0.15	4.67E-04	1.47E+02
6	GATA 098	0.406	7.97E-03	2.50E+03	0.312	3.49E-03	1.10E+03	0.125	5.85E-05	1.84E+01
7	GT 575	0.4	3.19E-03	1.00E+03	0.3	1.05E-03	3.30E+02	0.112	6.57E-06	2.06E+00
8	GT 101	0.377	1.20E-03	3.77E+02	0.254	2.66E-04	8.36E+01	0.101	6.63E-07	2.08E-01
9	GATA	0.355	4.26E-04	1.34E+02	0.21	5.58E-05	1.75E+01	0.07	4.67E-08	1.47E-02

	97408									
10	TAA 031	0.353	1.50E-04	4.72E+01	0.205	1.15E-05	3.60E+00	0.069	3.21E-09	1.01E-03
12	GT 015	0.343	1.78E-05	5.58E+00	0.187	4.04E-07	1.27E-01	0.057	1.10E-11	3.45E-06
13	GATA 053	0.341	6.06E-06	1.90E+00	0.182	7.33E-08	2.30E-02	0.055	6.00E-13	1.89E-07
14	GT 211	0.337	2.04E-06	6.40E-01	0.173	1.27E-08	3.99E-03	0.054	3.21E-14	1.01E-08
15	EV 096	0.339	6.90E-07	2.17E-01	0.177	2.25E-09	7.06E-04	0.053	1.72E-15	5.39E-10
16	AC 087	0.34	2.35E-07	7.37E-02	0.18	4.04E-10	1.27E-04	0.051	8.76E-17	2.75E-11
17	GT 011	0.334	7.83E-08	2.46E-02	0.168	6.78E-11	2.13E-05	0.05	4.38E-18	1.38E-12
18	GATA 417	0.319	2.50E-08	7.84E-03	0.138	9.34E-12	2.93E-06	0.034	1.48E-19	4.63E-14
19	EV 037	0.315	7.86E-09	2.47E-03	0.13	1.21E-12	3.81E-07	0.03	4.45E-21	1.40E-15
20	GATA 43950	0.3	2.36E-09	7.41E-04	0.1	1.22E-13	3.82E-08	0.018	8.01E-23	2.52E-17

Notes: [1]Pr(*I*) printed in descending order. [2]Expected no. of chance matches based on 793 samples (equal to 314,028 pairwise comparisons) assuming all of the specific relationship. UR = unrelated, FS = full siblings, PO = parent and offspring

Table 7. Estimates of genetic divergence (θ_{WC}) estimated from STR genotype data

	Winter breeding areas		Summer feeding areas		
	Lesser Antilles	CVI	Gulf of Maine	Iceland	Norway/Barents Sea
Greater Antilles	-.0023 (-.017 - 0.019)	0.017*** (0.0085 - 0.029)	-0.0004 (-0.0029 - 0.0024)	0.0019 (-0.0003 - 0.0048)	0.0042*** (0.002 - 0.0069)
Lesser Antilles		0.0098** (-0.0083 - 0.035)	0.0010 (-0.015 - 0.020)	-0.0039 (-0.017 - 0.016)	-.0031 (-0.016 - 0.017)
CVI			0.0144*** (0.0062 - 0.026)	0.0097*** (0.0028-0.019)	0.0084*** (0.0015 - 0.017)
Gulf of Maine				0.0016 (-0.0005 - 0.004)	0.0034** (0.001 - 0.006)
Iceland					0.0016** (0.0004- 0.0030)

Notes: No asterisk denotes P>0.05, ** 0.001<P<0.01, and *** P<0.001. Range in parenthesis is 95% confidence interval.

Table 8. Estimates of genetic divergence (θ_{WC}) estimated from ddRADSeq SNP genotype data

	Winter breeding area	Summer feeding areas		
	CVI	Gulf of Maine	Iceland	Norway/Barents Sea
Greater Antilles	0.0088 (-0.0007 - 0.021)	0.0036 (-0.0003 - 0.0098)	0.0038 (0.0004 - 0.0081)	0.0077*** (0.0048 - 0.012)
CVI		0.011 (0.0027 - 0.022)	0.0084 (0.0007 - 0.019)	0.0054 (-0.0029 - 0.017)
Gulf of Maine			0.0005 (-0.0025 - 0.0045)	0.0036 (-0.0001 - 0.0071)
Iceland				0.0014 (-0.0011 - 0.0041)

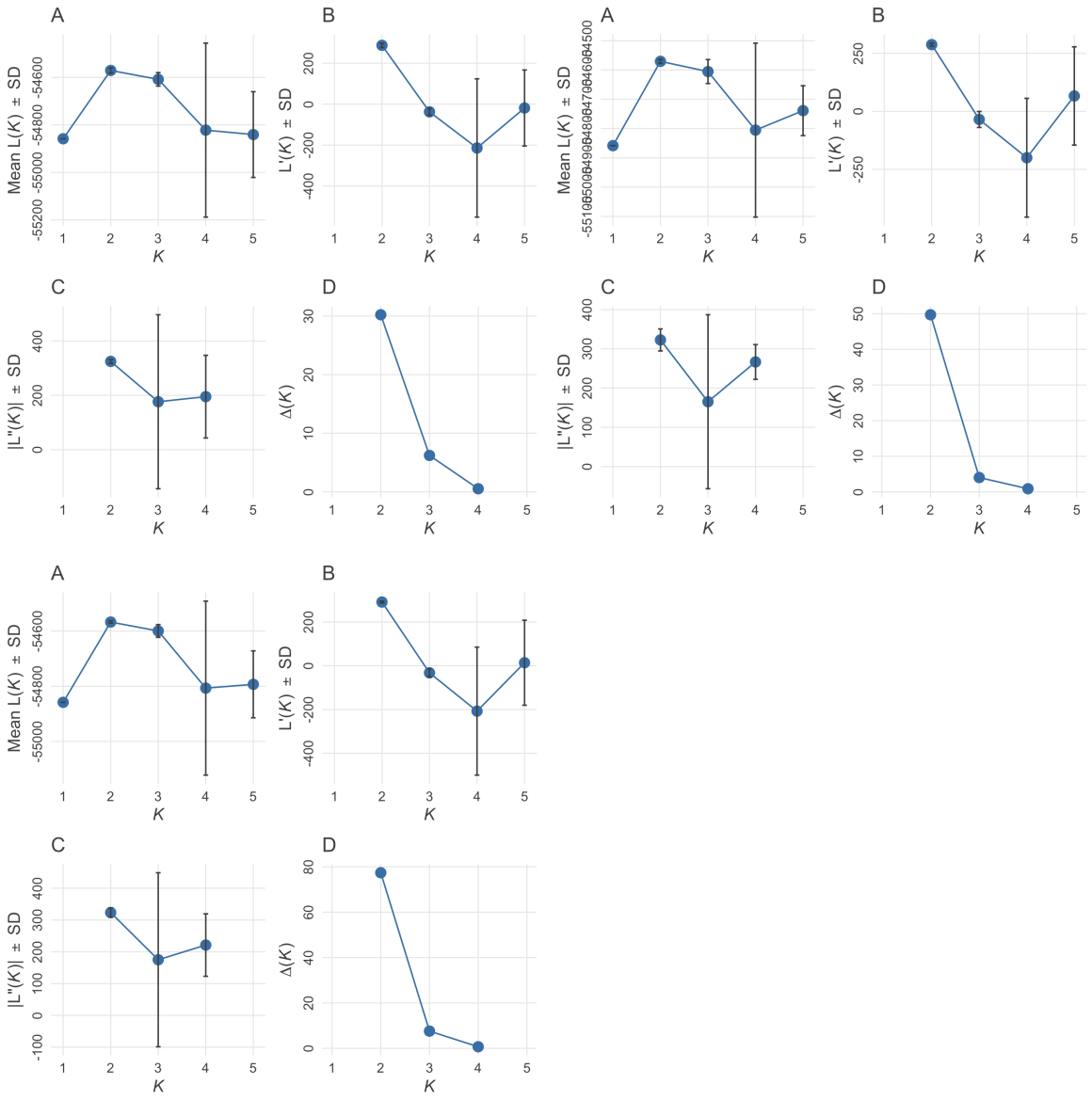
Notes: No asterisk denotes P>0.05, ** 0.001<P<0.01, and *** P<0.001.

STRUCTURE assessment

Convergence in the STRUCTURE analyses was achieved with 120,000 and 600,000 MCMC iterations for ddRADseq SNP and STR genotypes, respectively, of which the initial one third iterations were discarded. The three estimations from each data set were based on 15 replicates for K between one and five. λ was estimated at 0.977 and 0.487 for the ddRADSeq and STR genotype data, respectively.

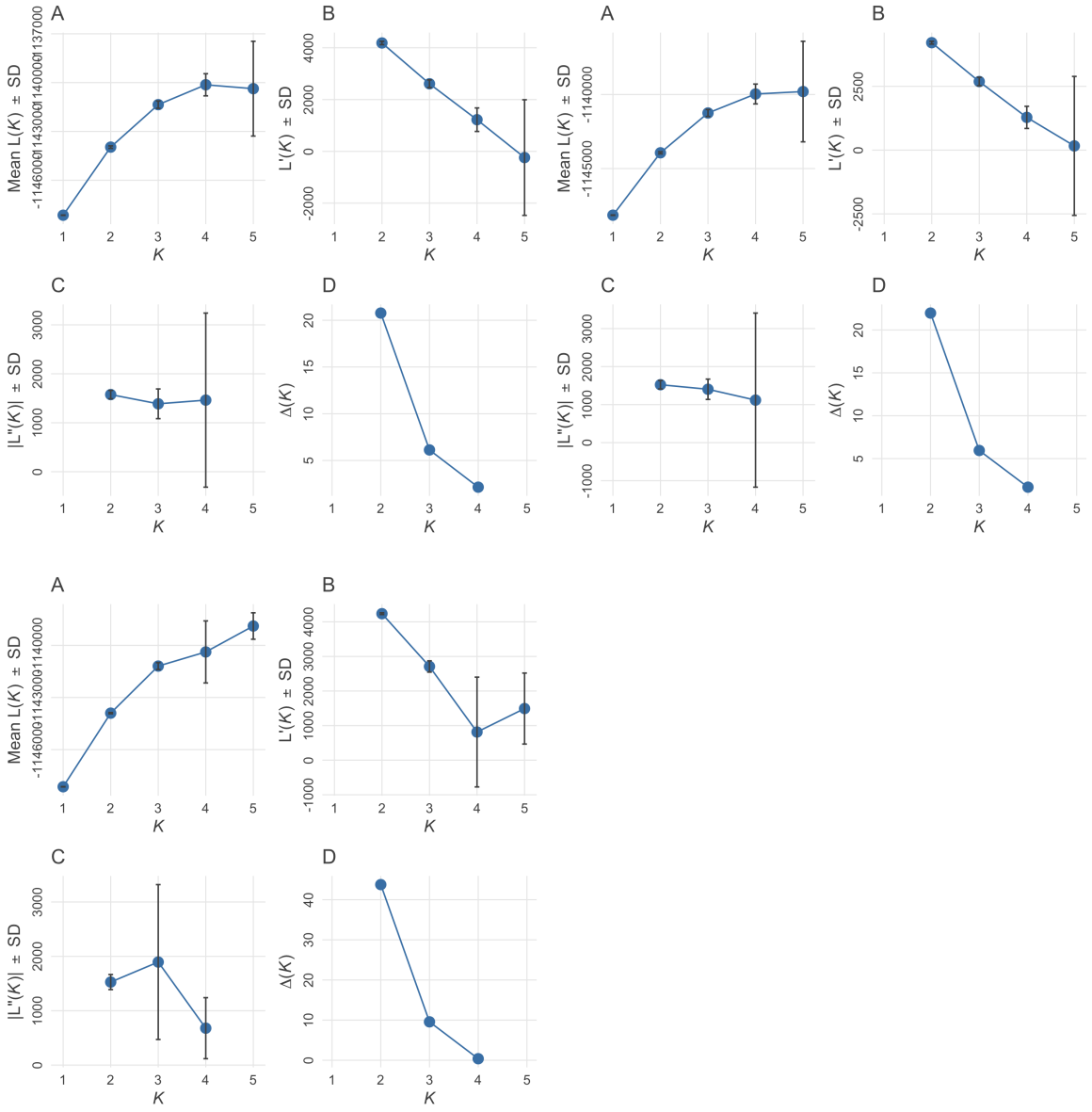
In case of the estimations based on STR genotypes, both Ln(K) and Evanno's ΔK supported a value of two (Fig. 2A). Evanno's ΔK also supported a value of K at two when the estimation was based on the SNP genotype data, whereas Ln(K) increased with K (Fig. 2B). The admixture proportions at $K = 2$ divided the samples into a "western" (Cluster 1, Fig. 2C-D) and "eastern" North Atlantic cluster (Cluster 2, Fig. 2C-D). Individuals from the Greater and Lesser Antilles as well as the Gulf of Maine were all assigned to the western North Atlantic cluster and did not show any indications of a substantial degree of admixture. Among the samples from CVI and the Norway/Barents Sea a few individuals were assigned fully (ddRADSeq SNPs) or almost fully (STR loci, same individuals as for ddRADSeq SNPs) to the eastern North Atlantic cluster. A larger proportion of individuals from CVI, Norway/Barents Sea and Iceland were admixed with a decline in degree and proportion of admixed individuals when moving from those sites respectively and further west (Fig. 2C-D).

Figure 2A. STR data plots of Ln(*K*) and statistics for Evanno's *ΔK*



Notes: Panels A depicts the mean Ln(*K*); B, the rate of change in mean Ln(*K*); C, the absolute value of the 2nd order of change in mean Ln(*K*); D, the estimated value of Evanno's *ΔK*

Figure 2B. ddRADSeq SNP data plots of Ln(*K*) and statistics for Evanno's *ΔK*



Notes: Panels A depicts the mean Ln(*K*); B, the rate of change in mean Ln(*K*); C, the absolute value of the 2nd order of change in mean Ln(*K*); D, the estimated value of Evanno's *ΔK*

Figure 2C. Admixture plots inferred from 793 20-locus STR genotypes. Based on 15 replicates for each value of K.

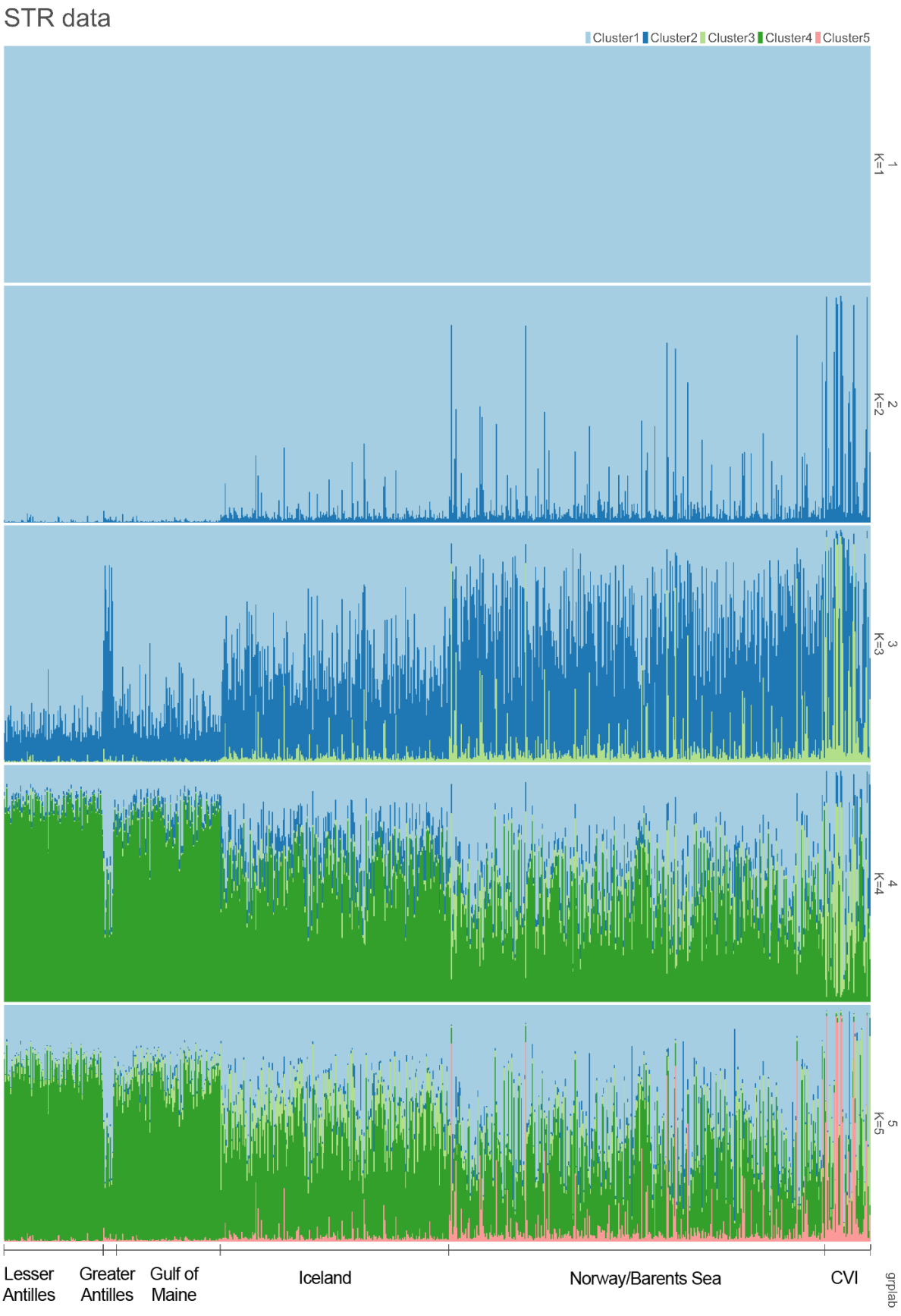


Figure 2D. Admixture plots inferred from 203 samples and 10,308 SNPs. Based on 15 replicates for each value of K.

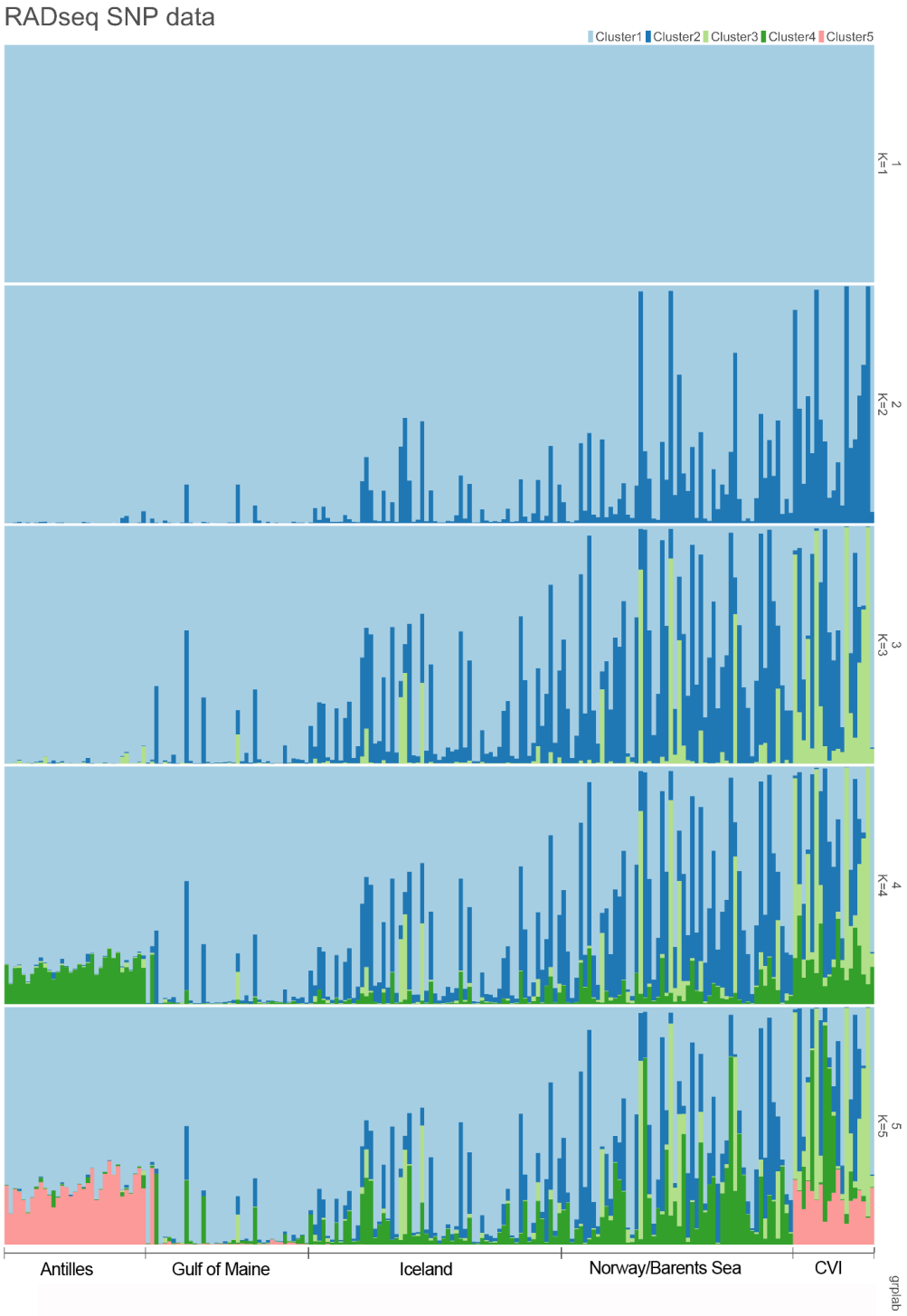
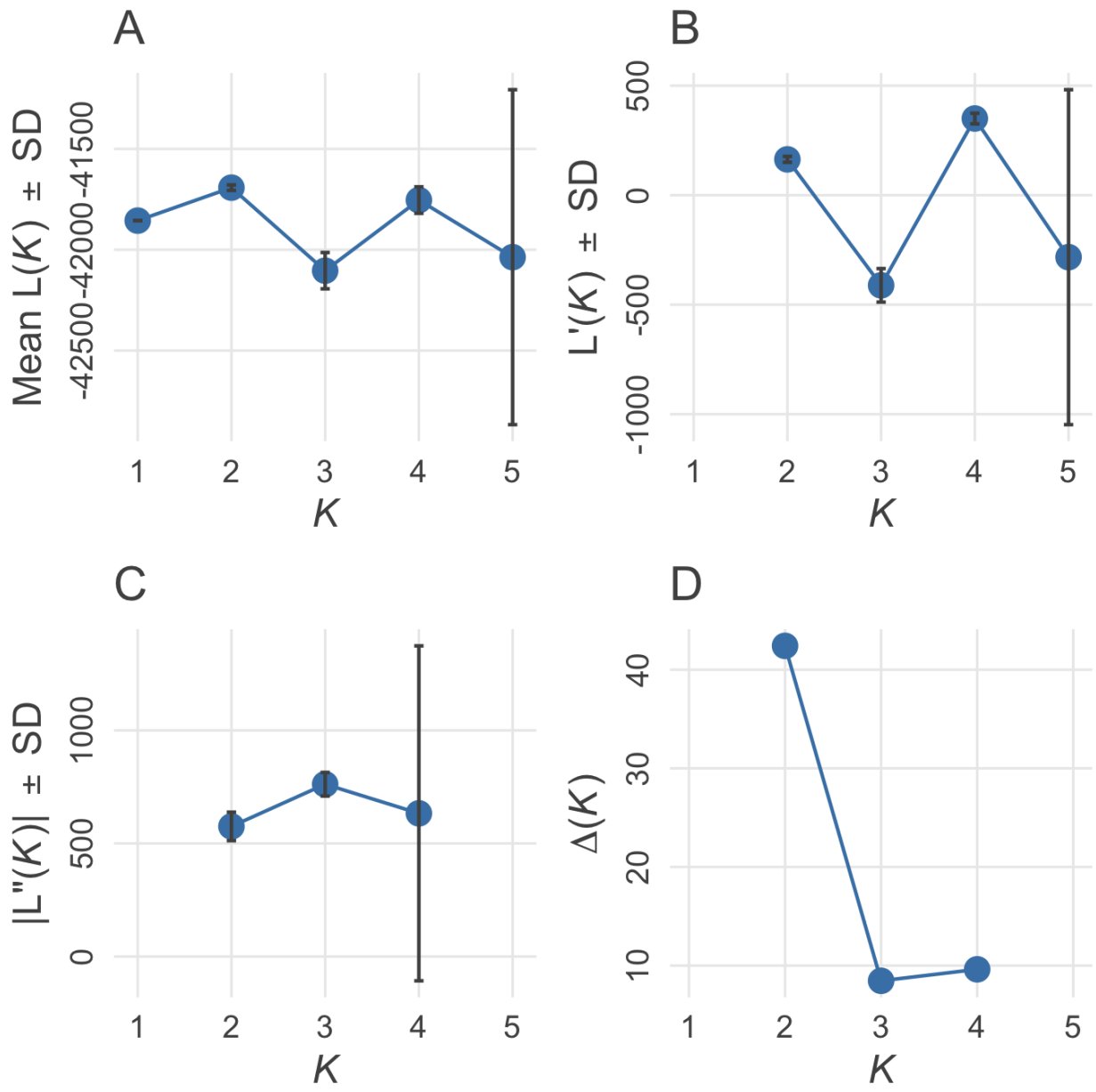


Figure 3A. STR data plots of Ln(*K*) and statistics for Evanno's *ΔK* (excluding samples from the Greater Antilles and the Gulf of Maine)



Notes: Panels A depicts the mean Ln(*K*); B, the rate of change in mean Ln(*K*); C, the absolute value of the 2nd order of change in mean Ln(*K*); D, the estimated value of Evanno's *ΔK*

Figure 3B. Admixture plots inferred from 607 20-locus STR genotypes. Based on 15 replicates for each value of K. The Greater Antilles and Gulf of Maine were excluded.

Contemporary migration rates

The BayesAss-based estimations of *m* from both data sets indicated a much higher fraction of Greater Antilles immigrants among the CVI samples than vice versa (Table 9). The difference was most pronounced in the estimates based on the STR data (.25 vs. .005) compared to the SNP data (.1 vs. .01). The final values for the Δ_M, Δ_A, Δ_F mixing priors were; 0.09, 0.25 and 0.15,

respectively for the STR-based estimates, and 0.13, 0.3, 0.02, respectively for the SNP-based estimates. The STR and ddRADSeq SNP genotype estimates were based on 10,000,000 and 100,000 MCMC iterations, respectively of which the first ⅓ were discarded.

Table 9. BayesAss acceptance rates and estimates

	Δ_M, Δ_A, Δ_F[1]	Proportion of immigrants (top) Inbreeding coefficients (bottom)	
		CVI	Greater Antilles
STR genotypes			
Estimation 1	.41, .42, .43	.252 (SE[2] .0283) .0297 (SE .0632)	.0052 (SE .005) .0157 (SE .0084)
Estimation 2	.41, .42, .43	.252 (SE .0278) .0281 (SE .0557)	.0052 (SE .005) .0157 (SE .0084)
Estimation 3	.41, .43, .43	.253 (SE .0282) .0314 (SE .0715)	.0052 (SE .005) .0157 (SE .0084)
ddRADSeq SNP genotypes			
Estimation 1	.43, .45, .39	.0929 (SE .0553) .0311 (SE .0093)	.01 (SE .0099) .0251 (SE .0072)
Estimation 2	.44, .45, .39	.0947 (SE .0554) .0295 (SE .0106)	.0095 (SE .0091) .0231 (SE .0073)
Estimation 3	.43, .45, .38	.0933 (SE .0533) .0249 (SE .0107)	.0102 (SE .0102) .0232 (SE .0072)

Notes: [1]Acceptance rates for migration, allele frequencies and inbreeding, respectively. [2]Standard error.

Discussion

Overall population structure and introgression.

The overall picture emerging from the data analyses agrees remarkably well among the different data sets and analyses. Overall the results suggest the presence of an eastern North Atlantic cluster of humpback whales; which is experiencing substantial introgression from western North Atlantic humpback whales, represented by the 2nd cluster identified in the STRUCTURE

assessment. The distribution of individuals with 100% or admixed eastern North Atlantic cluster genome was mostly confined to CVI and Norway/Barents Sea. The fraction of admixed individuals at Iceland was comparatively lower and the degree of admixture was generally lower. Humpback whales from the Lesser and Greater Antilles and the Gulf of Maine were generally not admixed (with the exception of a few slightly admixed individuals) and all were assigned to the "western" North Atlantic cluster. One possible interpretation of these results is the presence of a once distinct, historical, poorly recovering eastern North Atlantic breeding population that winters and breeds at CVI. The current, low estimates of abundance of humpback whales in this area at 272 (SE 10, Wenzel et al., 2020) is far below the estimated historical abundance at ~ 5,000 whales (Punt et al., 2006). In contrast, the recovery rate of humpback whales that winter and breed in the Greater Antilles is much less dire; from a historical abundance at ~25,000 (Punt et al., 2006) to a most recent estimate at ~11,000 (Stevick et al., 2003). Consequently, a relatively few immigrants from a recovering western North Atlantic population into a small, severely depleted eastern North Atlantic population during the last few generations could result in the spatial configuration among individuals observed in this study, i.e., leading to 10 - 25% immigration per generation (BayesAss estimates of m). If this level of immigration and introgression continues, the descendants of the putative historical "eastern" North Atlantic population will eventually disappear, and the eastern North Atlantic breeding population will be supplanted by individuals with western North Atlantic genomes.

The "eastern" North Atlantic cluster could potentially represent gene flow from South Atlantic humpback whales into the eastern North Atlantic. However, we did assess this possibility in collaboration with the Wildlife Conservation Society (Dr Rosenbaum) including samples from Gabon, off western Africa (Gabualdi, unpublished data). The Gabon samples were assigned to their own, third cluster.

There were some indications of additional "clusters" in the data, in particular the ddRADSeq SNP genotypes. However, some of the additional "clusters" seen in the admixture plots at K at four and five align with individual ddRADSeq libraries rather than geography and are thus likely so-called "batch effects", i.e., experimental issues. We attempted to remove some of this noise by removing SNP loci that yielded unusually high θ_{WC} estimates (>0.3) between ddRADSeq libraries. However, such pruning could perhaps have been conducted at an even lower threshold, since this batch effect was still evident at K of four and five. In addition, no corresponding clusters were evident at K of four and five in the admixture plots inferred from the STR genotypes. However, at $K = 3$, the admixture probabilities in the samples from the Lesser Antilles samples (Fig. 2C) mostly resemble the proportions of admixture observed in the samples from CVI and Norway/Barents Sea, which is consistent with recent photographic matches of individual humpback whales observed in both the Lesser Antilles and CVI (Stevick et al., 2016). In addition, the genetic divergence inferred from the mitochondrial DNA sequences was much lower between the Lesser Antilles and CVI than between the Lesser and Greater Antilles (Tables 4 & 5), whereas the opposite was the case for divergences estimated from the nuclear STR genotypes (Table 7, although significantly different). Humpback whales sighted in the

Northwest Atlantic (the Gulf of Maine, Atlantic Canada and west Greenland) have also been sighted (at much lower rates, however) in the Lesser Antilles (Stevick et al., 2018).

Tracking data from satellite-tracked humpback whales further confirm that the Greater and Lesser Antilles are both winter migratory destinations for whales that feed off Iceland and Norway (Kennedy et al., 2014). An additional STRUCTURE assessment excluding samples from the Greater Antilles and the Gulf of Maine supported the previous value of *K* (Fig. 3A), but some (subtle) support for the Lesser Antilles samples being similar to those from CVI and Norway/Barents Sea was only visible in the admixture plot at $K = 5$ in terms of the admixture proportion cluster number four (Fig. 3B). Thus, there may be some additional subtle affinity of whales that breed at the Lesser Antilles with the eastern North Atlantic. Unfortunately the ddRADSeq data from the Lesser Antilles samples were insufficient for inclusion in this analysis; additional ddRADSeq reads would need to be generated. The observed pattern of divergences and admixture in some STRUCTURE plots could, potentially, be due to maternally directed site-fidelity to a specific winter breeding area (as is the case for summer feeding areas) and occasional male-driven connectivity between winter breeding areas.

The immigration rates into the Greater Antilles from CVI, as inferred from BayesAss, were between 0.5% (STR genotypes) and 1% (ddRADSeq genotypes). Additional, STRUCTURE assessments based upon a larger Greater Antilles sample suggest even lower rates (Gabualdi, unpublished data; Palsbøll, unpublished data).

Further assessment of the above hypothesis of the recent introgression of western North Atlantic genomes into a historically more isolated eastern North Atlantic breeding population is currently being pursued by means of model selection using Approximate Bayesian Computation (work by MSc. student Yakamoz Kizildas and Dr. Jurjan van der Zee).

Comparing traditional and genome data

Comparing the results obtained from the more "traditional" data (STR genotypes and mitochondrial control region DNA sequences) with "genomic" data (i.e., ddRADseq SNP genotypes and mitochondrial genome DNA sequences) it seems that the choice of which kind of data to generate is not black and white. Both costs and the specific research objective need to be considered. Genome-level data remains more expensive on a per sample basis compared to traditional data. Hence, sample sizes are typically comparatively lower in analyses based upon genome data (e.g., here 203 versus 703 samples). The difference in sample sizes and the number of alleles per locus (two for SNPs) affects overall power resulting in fewer rejections of homogeneity in the SNP data, and more divergence estimates with a 95% confidence interval including zero, compared to the STR-based assessment. In addition, genome data, such as ddRADseq SNP genotypes are subject to considerable noise, e.g., among experiments. In the present study initial STRUCTURE assessments were dominated by clusters that aligned with libraries rather than geography (as was the case for the STR-based assessment). This effect was due to 66 inter-library outliers of 10,374 SNPs in total, or a mere 0.6%. We have similar

experiences with ddRADseq data from North Atlantic fin whales where 13 loci (of ~ 13,000) resulted in a highly distinct DAPC cluster that corresponded to a single ddRADseq library. Accordingly, it appears that (a) the outcome in ddRADseq data is highly dependent on the bioinformatic pipeline, and (b) the overall power in SNPs is considerably lower.

The analyses based on DNA sequences of the entire mitochondrial genome and mitochondrial control region were based on the same samples and hence the same sample size. In North Atlantic humpback whales, approximately 20% of all segregation sites in the mitochondrial genome are located in the region of the mitochondrial control region typically employed in this kind of population genetic assessment. The differences in divergence were mainly observed when the degree of divergence was in the higher end of the range (>0.25) but mostly the two assessments yielded similar results and a surprisingly similar picture of the spatial configuration of mitochondrial diversity across the sampled regions.

In conclusion, the assumed increase in power from genome data is not always realized, probably due to the fact that traditional genetic markers have targeted loci/regions with elevated levels of allelic/DNA sequence diversity. In contrast, SNPs are bi-allelic and the nucleotide diversity is often much lower in the regions outside the control region on the mitochondrial genome, thus, in both cases, resulting in a limited gain in precision and statistical power (at least in the present study). Accordingly, it may not always be the optimal strategy from a resource point of view to aim for genome data as a default but instead aim for larger sample sizes and traditional genetic markers.

Likely an ongoing, cryptic extinction of a historical eastern North Atlantic humpback whale breeding population

The results presented here suggest that uneven recovery rates of the western and eastern North Atlantic breeding populations is resulting in an introgression of western North Atlantic genomes into the eastern North Atlantic cluster identified in this study, which likely represents the remnants of what used to be a historically isolated eastern North Atlantic breeding population. The individuals of a predominantly eastern North Atlantic cluster origin appear confined to the winter and summer range in the eastern North Atlantic. The very high rate of immigration of western origin individuals into the eastern breeding population suggests the observed introgression is a recent phenomenon as a long term immigration rate at 10 - 25% would erase any signal of west-east genetic structure. The analyses presented here demonstrated that individuals with a 50 - 100 % eastern North Atlantic origin can easily be identified by genotyping 20 STR loci. Hence, it is possible to monitor if mortalities are from this much smaller population, and if certain areas or times are critical.

Acknowledgements

We are grateful to the many, many people who contributed to the field work as well as The Marine Institute (Norway), Northeast Fisheries Science Center, the Irish Research Council, the Norwegian Research Council (Ice Whales project) and Island Foundation for funding,

Cited literature

Amos, W., & Hoelzel, A. R. (1991). Long-term preservation of whale skin for DNA analysis. *Reports of the International Whaling Commission Special Issue*, *13*, 99–104.

Arnason, U., Gullberg, A., & Widegren, B. (1993). Cetacean mitochondrial DNA control region: Sequences of all extant baleen whales and two sperm whale species. *Molecular Biology and Evolution*, *10*(5), 960–970.

Bérubé, M., Aguilar, A., Dendanto, D., Larsen, F., Notarbartolo-di-Sciara, G., Sears, R., Sigurjonsson, J., Urban-R, J., & Palsbøll, P. J. (1998). Population genetic structure of North Atlantic, Mediterranean Sea and Sea of Cortez fin whales, *Balaenoptera physalus* (Linnaeus 1758): Analysis of mitochondrial and nuclear loci. *Molecular Ecology*, *7*(5), 585–599. https://doi.org/10.1046/j.1365-294x.1998.00359.x

Bérubé, M., Jørgensen, H., McEwing, R., & Palsbøll, P. J. (2000). Polymorphic di-nucleotide microsatellite loci isolated from the humpback whale, *Megaptera novaeangliae*. *Molecular Ecology*, *9*(12), 2181–2183. https://doi.org/10.1046/j.1365-294X.2000.105315.x

Bérubé, M., Rew, M. B., Skaug, H. J., Jørgensen, H., Robbins, J., Best, P. B., Sears, R., & Palsbøll, P. J. (2005). Polymorphic microsatellite loci isolated from humpback whale, *Megaptera novaeangliae*, and fin whale, *Balaenoptera physalus*. *Conservation Genetics*, *6*(4), 631–636. https://doi.org/10.1007/s10592-005-9017-5

Bukhman, Y. V., Morin, P. A., Meyer, S., Chu, L.-F. (Jack), Jacobsen, J. K.,

Antosiewicz-Bourget, J., Mamott, D., Gonzales, M., Argus, C., Bolin, J., Berres, M. E., Fedrigo, O., Steill, J., Swanson, S. A., Jiang, P., Rhie, A., Formenti, G., Phillippy, A. M., Harris, R. S., … Stewart, R. (2022). *A high-quality blue whale genome, segmental duplications, and historical demography* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-1910240/v1

Cabrera, A. A., Schall, E., Bérubé, M., Anderwald, P., Bachmann, L., Berrow, S., Best, P. B., Clapham, P. J., Cunha, H. A., Dalla Rosa, L., Dias, C., Findlay, K. P., Haug, T., Heide-Jørgensen, M. P., Hoelzel, A. R., Kovacs, K. M., Landry, S., Larsen, F., Lopes, X. M., … Palsbøll, P. J. (2022). Strong and lasting impacts of past global warming on baleen whales and their prey. *Global Change Biology*, *28*(8), 2657–2677. https://doi.org/10.1111/gcb.16085

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). *Stacks*: Building and Genotyping Loci *de novo* From Short-read Sequences. *G3 (Bethesda, Md.)*, *1*(3), 171–182. https://doi.org/10.1534/g3.111.000240

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics (Oxford, England)*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Molecular Ecology Notes*,

7(4), 574–578. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Franchini, P., Monné Parera, D., Kautt, A. F., & Meyer, A. (2017). QuaddRAD: a new high-multiplexing and PCR duplicate removal ddRAD protocol produces novel evolutionary insights in a nonradiating cichlid lineage. *Molecular Ecology*, *26*(10), 2783–2795. https://doi.org/10.1111/mec.14077

Francis, R. M. (2017). POPHELPER : an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27–32. https://doi.org/10.1111/1755-0998.12509

Hudson, R. R., Boos, D. D., & Kaplan, N. L. (1992). A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution*, *9*(1), 138–151. https://doi.org/10.1093/oxfordjournals.molbev.a040703

Kalinowski, S. T., Wagner, A. P., & Taper, M. L. (2006). ML-Relate: A computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes*, *6*(2), 576–579. https://doi.org/10.1111/j.1471-8286.2006.01256.x

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, *4*(8), 782–788. https://doi.org/10.1111/2041-210X.12067

Kennedy, A. S., Zerbini, A. N., Rone, B. K., & Clapham, P. J. (2014). Individual variation in movements of satellite-tracked humpback whales *Megaptera novaeangliae* in the eastern Aleutian Islands and Bering Sea. *Endangered Species Research*, *23*(2), 187–195. https://doi.org/doi: 10.3354/esr00570

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*

Methods, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Mackay, M., Bacon, C., Bouveret, L., Fossette, S., & Stevick, P. (2019). Humpback Whale (*Megaptera novaeangliae*) Intra/Inter- Seasonal Exchanges Between Puerto Rico and the Southeastern Caribbean. *Animal Behavior and Cognition*, *6*, 98–104. https://doi.org/10.26451/abc.06.02.02.2019

Mullis, K. B., & Faloona, F. A. (1987). [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. In *Methods in Enzymology* (Vol. 155, pp. 335–350). Academic Press. https://doi.org/10.1016/0076-6879(87)55023-6

Mussmann, S. M., Douglas, M. R., Chafin, T. K., & Douglas, M. E. (2019). BA3-SNPs: Contemporary migration reconfigured in BayesAss for next-generation sequence data. *Methods in Ecology and Evolution*, *10*(10), 1808–1813. https://doi.org/10.1111/2041-210X.13252

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, *76*(10), 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Paetkau, D., & Strobeck, C. (1994). Microsatellite analysis of genetic variation in black bear populations. *Molecular Ecology*, *3*(5), 489–495.

Palsbøll, P. J., Bérubé, M., Larsen, A. H., & Jørgensen, H. (1997). Primers for the amplification of tri- and tetramer microsatellite loci in baleen whales. *Molecular Ecology*, *6*(9), 893–895. https://doi.org/10.1046/j.1365-294X.1997.d01-214.x

Palsbøll, P. J., Larsen, F., & Sigurd Hansen, E. (1991). Sampling of Skin Biopsies from Free-Ranging Large Cetaceans in West Greenland: Development of new Biopsy Tips and Bolt Designs. *Report of the International Whaling Commission (Special Issue)*, *13*,

71–79.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: An Inexpensive Method for *de novo* SNP Discovery and Genotyping in Model and Non-model Species. *Plos One*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Punt, A. E., Friday, N., & Smith, T. D. (2006). Reconciling data on the trends and abundance of North Atlantic humpback whales within a population modeling framework. *Journal of Cetacean Research and Management*, *8*, 145–160.

Rew, M. B., Robbins, J., Mattila, D., Palsbøll, P. J., & Bérubé, M. (2011). How many genetic markers to tag an individual? An empirical assessment of false matching rates among close relatives. *Ecological Applications*, *21*(3), 877–887. https://doi.org/10.1890/10-0348.1

Rousset, F. (2008). genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution*, *34*(12), 3299–3302. https://doi.org/10.1093/molbev/msx248

Sambrook, J., & Russell, D. W. (2001). *Molecular Cloning. A laboratory manual* (Third edition, Vols. 1–3). Cold Spring Harbor Laboratory Press. www.molecularcloning.com

Smith, T. D., & Pike, D. G. (2009). The enigmatic whale: The North Atlantic humpback. *NAMMCO Scientific Publications*, *7*, 161–178.

Stevick, P. T., Allen, J., Clapham, P. J., Friday, N., Katona, S. K., Larsen, F., Lien, J., Mattila, D. K., Palsbøll, P. J., Sigurjónsson, J., Smith, T. D., Øien, N., & Hammond, P. S. (2003). North Atlantic humpback whale abundance and rate of increase four decades after protection from whaling. *Molecular Ecology Progress Series*, *258*, 263–273.

Stevick, P. T., Berrow, S. D., Bérubé, M., Bouveret, L., Broms, F., Jann, B., Kennedy, A., Suárez, P. L., Meunier, M., Ryan, C., & Wenzel, F. W. (2016). There and back again: Multiple and return exchange of humpback whales between breeding habitats separated by an ocean basin. *Journal of the Marine Biological Association of the United Kingdom*, 1–6. https://doi.org/10.1017/s0025315416000321

Stevick, P. T., Bouveret, L., Gandilhon, N., Rinaldi, C., Rinaldi, R., Broms, F., Carlson, C., Kennedy, A., Ward, N., & Wenzel, F. (2018). Migratory destinations and timing of humpback whales in the southeastern Caribbean differ from those off the Dominican Republic. *J. Cetacean Res. Manage.*, *18*(1), 127–133. https://doi.org/10.47536/jcrm.v18i1.442

Stevick, P. T., Carlson, C. A., & Balcomb, K. C. (1999). A note on migratory destinations of humpback whales from the eastern Caribbean. *Journal of Cetacean Research and Management*, *1*(3), 251–254.

Suárez-Menéndez, M., Bérubé, M., Furni, F., Rivera-León, V. E., Heide-Jørgensen, M.-P., Larsen, F., Sears, R., Ramp, C., Eriksson, B. K., Etienne, R. S., Robbins, J., & Palsbøll, P. J. (2022). *Direct estimation of genome mutation rates from pedigrees in free-ranging baleen whales* (p. 2022.10.06.510775). bioRxiv. https://doi.org/10.1101/2022.10.06.510775

Valsecchi, E., & Amos, W. (1996). Microsatellite markers for the study of cetacean populations.

Molecular Ecology, *5*(1), 151–156.

Waples, R. S., & Anderson, E. C. (2017). Purging putative siblings from population genetic data sets: A cautionary view. *Molecular Ecology*, *26*(5), 1211–1224. https://doi.org/10.1111/mec.14022

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.2307/2408641

Wenzel, F. W., Broms, F., López-Suárez, P., Lopes, K., Veiga, N., Yeoman, K., Rodrigues, M. S. D., Allen, J., Fernald, T. W., Stevick, P. T., Jones, L., Jann, B., Bouveret, L., Ryan, C., Berrow, S., & Corkeron, P. (2020). Humpback Whales (*Megaptera novaeangliae*) in the Cape Verde Islands: Migratory Patterns, Resightings, and Abundance. *Aquatic Mammals*, *46*(1), 21–31. https://doi.org/10.1578/AM.46.1.2020.21

Werle, E., Schneider, C., Renner, M., Volker, M., & Fiehn, W. (1994). Convenient Single-Step, One Tube Purification of Pcr Products for Direct Sequencing. *Nucleic Acids Research*, *22*(20), 4354–4355.

Wilson, G. a, & Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, *163*(3), 1177–1191.