# Genetic structure of the North Atlantic common minke whale (*Balaenoptera acutorostrata*) at feeding grounds: a combined microsatellite and mtDNA analysis

Ralph Tiedemann[1], Christophe Pampoulie[2], Anja R.R. Schneider[1], Katrin Kiemel[1], Lars Witting[3], Malene J. Simon[3], Maia E. Olsen[3], Hans J. Skaug[4], Nils Øien[5] and Gísli A. Víkingsson[2]

[1]Unit of Evolutionary Biology/Systematic Zoology, Institute of Biochemistry and Biology, University of Potsdam, D-14476 Potsdam, Germany

[2]Marine Research Institute, IS-101 Reykjavík, Iceland

[3]Greenland Institute of Natural Resources, 3900 Nuuk, Greenland

[4]University of Bergen, Bergen, Norway

[5]Institute of Marine Research, Bergen, Norway

## ABSTRACT

With the addition of 348 specimens typed at the conventional 16 microsatellites previously used on minke whales and 682 specimens sequenced at 369bp of the mitochondrial control region, we have compiled a data set of around 1,200 specimens of North Atlantic minke whales, representing – according to IWC stock assignment – the Western (West Greenland), the Central (East Greenland, Iceland), and the Eastern stock (Norway, Spitsbergen, Barent Sea, North Sea). Most of the genetic variation (over 99%) is assigned to the lowest level of geographic stratification in both microsatellites (i.e., the individual level) and mtDNA (i.e., the locality level). Nonetheless, there is a consistent tendency towards a subtle differentiation among the stocks. In all analysis, West Greenland and Eastern stock are slightly more differentiated. The Central stock is intermediate, with a closer affinity towards West Greenland. Locus-specific analysis reveals that (1) significance in the microsatellite data is due to divergence at a single locus, (2) levels of differentiation at mitochondrial DNA are similar to those revealed in a previous study, and (3) microsatellite $F_{ST}$ values – even if corrected for within population variability – are considerably lower than values derived from an earlier allozyme study. Possible reasons for these differences are discussed.

This study is generally compatible with the IWC-three stock hypothesis (W, C, E), but would not contradict a two stock hypothesis (W+C, E) either, as none of the analyses revealed any difference between W and C stock.

## INTRODUCTION

The common minke whale (*Balaenoptera acutorostrata*) is distributed throughout the North Atlantic. It has feeding grounds from the East of Canada and Western Greenland in the West over Iceland, East Greenland and Jan Mayen to the Spitsbergen, the Barent Sea, the Coast of Norway, and the North Sea in the East. It is currently poorly known (1) where North Atlantic minke whales breed (Víkingsson & Heide-Jørgensen 2014) and (2) whether they form a single breeding population.

Earlier studies on allozymes (Danielsdóttir et al. 1992) and mtDNA/microsatellites (Andersen et al. 2003) have reported some population structure. More recent mtDNA/microsatellite studies based on larger sample sizes suggest the existence of two breeding sites, but the absence of any spatial genetic structure on the feeding grounds (Pampoulie et al. 2008), though results of cluster analyses have been interpreted as indicative of cryptic population structure (Anderwald et al. 2011).

This paper provides an update of genetic resources for the North Atlantic minke whale based on the conventional genetic markers (16 microsatellite loci and the D-loop of the mtDNA) applied in Pampoulie et al. (2008) and puts these data into perspective with the findings of older studies. With regard to traditional IWC Schedule stock assignments, our analysis covers the Western (West Greenland), the Central (East Greenland/Iceland), and the Eastern (Norway, Spitsbergen, Barent Sea, North Sea) stocks (Donovan 1991).

## MATERIALS AND METHODS

For this paper, n=348 specimens were typed at 16 microsatellites (EV001, EV=037, EV094, EV096, GATA028, GATA053, GATA098, GATA417, GT011, GT023, GT195, GT211, GT310, GT509, GT575, and Sam25) and n=682 specimens were sequenced at 369bp of the mitochondrial Control Region according to the protocols provided in Pampoulie et al. (2008). These samples originated from Iceland and Greenland (figure 1, table 1). For further analysis, fetuses (n=18) and duplicate samples (n=2) were removed from the data set. The data were combined with the data from Pampoulie et al. 2008 as well as additional Norwegian data from 1991-2006, rendering a total of 1191 and 1215 specimens with known origin for the 16 microsatellites and the mtDNA, respectively (table 2). Twelve randomly chosen samples typed for microsatellites at the Marine Research Institute in Reykjavík were re-typed at the University of Potsdam. Out of 192 re-typed alleles, 190 were scored identical in both labs, translating into an error rate of 1%. The two mismatches could be resolved after additional re-typing.

First we looked for temporal differentiation with a hierarchical Analysis of Molecular Variance (AMOVA) using ARLEQUIN ver. 3.5 (Excoffier & Lischer 2010), having "stock" as the uppermost level and sampling period as second hierarchical level (stratified into before and after year 2000 for any location). After having detected a temporal pattern in the Icelandic microsatellite data (see results), we excluded samples from before 1990 from subsequent analyses (table 2), rendering a remaining total of 949 and 997 specimens typed for the 16 microsatellite loci and mtDNA, respectively.

These data were first analyzed without any a priori location/stock assignment (microsatellites only) using STRUCTURE ver. 2.3.2 (Pritchard et al. 2000) and subsequently with an assignment to locations and IWC stocks (table 2), performing again a hierarchical AMOVA, now with "stock" as the uppermost level and sampling location as second hierarchical level. With the same program, pairwise $F_{ST}$ values were calculated and exact tests of population differentiation performed. Significance was established by permutations.

We also calculated locus-wise $F_{ST}$ values which we corrected for within-population diversity by

$$F_{ST}' = \frac{F_{ST}}{(1-H)} \qquad \text{(Hedrick 2005)} \qquad (1)$$

where H is the locus-specific expected within-population heterozygosity (haplotype diversity in case of mtDNA), averaged over the compared populations. All H values were corrected for sample size n by multiplication with n/(n-1). For comparison, we also calculated locus-specific $F_{ST}$'-values for two published studies on North Atlantic minke whales which cover all three IWC stocks, i.e., for 9 allozyme loci from Danielsdóttir et al. (1992) and for the mitochondrial Control Region data from Andersen et al. (2003).

Finally, we performed a one-level AMOVA where we collapsed all sample location within a stock into a single population, in order to calculate $F_{ST}$ values among stocks.

Table 1: Sampling location and year of sampling of common minke whale individuals additionally scored at 16 microsatellites and the mtDNA Control Region. Brei: Breiðafjörður Bay, West Iceland, Fax: Faxaflói Bay, SW-Iceland, Horn: Hornafjörður, SE-Iceland; Hún: Húnaflói Bay, N-Iceland (See Fig. 1 for Icelandic localities).

| **Sampling-acronym** | **Microsatellites** | **mtDNA** |
|---|---|---|
| Brei2007 | 1 | 1 |
| Brei2008 | 35 | 35 |
| Brei2009 | 16 | 15 |
| Brei2010 | 1 | 1 |
| Brei2012 | 8 | 9 |
| Fax2009 | 60 | 60 |
| Fax2010 | 47 | 47 |
| Fax2011 | 60 | 61 |
| Fax2012 | 41 | 39 |
| Horn1992 | 1 | 1 |
| Horn2004 | 1 | 1 |
| Horn2007 | 1 | 1 |
| Hun2007 | 1 | 1 |
| Hun2008 | 2 | 2 |
| Hun2009 | 4 | 2 |
| Hun2011 | 3 | 3 |
| Iceland - others | 3 | 3 |
| EastGreen2000 | 2 | 2 |
| EastGreen2001 | 1 | 1 |
| EastGreen2002 | 3 | 3 |
| EastGreen2004 | | 8 |
| EastGreen2006 | 2 | 2 |
| EastGreen2007 | 1 | 1 |
| EastGreen2009 | 2 | 2 |
| EastGreen2010 | 1 | 1 |
| EastGreen2011 | 1 | 1 |
| EastGreen2012 | 3 | 3 |
| WestGreen2000 | | 20 |
| WestGreen2001 | | 3 |
| WestGreen2002 | | 6 |
| WestGreen2003 | | 11 |
| WestGreen2004 | | 41 |
| WestGreen2005 | | 6 |
| WestGreen2006 | | 42 |
| WestGreen2007 | 20 | 40 |
| WestGreen2008 | | 43 |
| WestGreen2009 | 27 | 40 |
| WestGreen2010 | | 43 |
| WestGreen2011 | | 40 |
| WestGreen2012 | | 41 |
| Sum Iceland | 285 | 282 |
| Sum Greenland | 63 | 400 |
| **Total** | **348** | **682** |

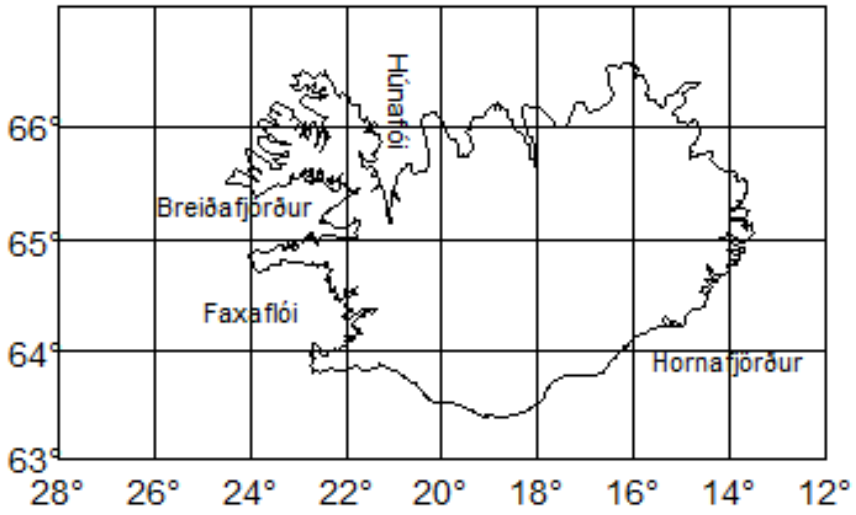


Figure 1: Sampling locations around Iceland.

**RESULTS**

*Microsatellite loci*

We first looked for a temporal pattern in the total data set of microsatellite data from 1191 common minke whale samples since 1980. In a hierarchical AMOVA with "stock" as the uppermost level and sampling period as second hierarchical level (stratified into before and after year 2000), there was a tendency towards a divergence between the two periods ($F_{CT}$=0.001, p=0.063). This was because of a differentiation among the Icelandic samples of the two periods ($F_{ST}$=0.002, p<0.001). This effect disappeared if samples from before 1990 were excluded. Hence, all further analyses were based on samples from 1990 onwards only.

All local samples were in Hardy-Weinberg-equilibrium (non-significant $F_{IS}$), except for the sample from Western Iceland (i.e., Faxaflói, Breiðafjörður; $F_{IS}$=0.020, p=0.019).

Table 2: Sampling location and assignment to IWC stocks (Donovan 1991) of minke whale individuals scored at 16 microsatellites and the mtDNA Control Region since 1990 (combined data from table 1, further Norwegian samples from 1991-2006, and data from Pampoulie et al. 2008). For Iceland and West Greenland, additional samples from 1981-1985 were used for the analysis of temporal patterns (additional sample size in parentheses).

| Sampling location | Stock | Microsatellites | mtDNA |
|---|---|---|---|
| Iceland | Central (C) | 448 (+123) | 443 (+157) |
| East Greenland | Central (C) | 16 | 24 |
| West Greenland | Western (W) | 47 (+19) | 370 (+61) |
| Norway | Eastern (E) | 318 | 40 |
| North Sea | Eastern (E) | 7 | 7 |
| Barent Sea | Eastern (E) | 50 | 50 |
| Spitsbergen | Eastern (E) | 63 | 63 |
| Sum Central | | 464 (+123) | 467 (+157) |
| Sum Western | | 47 (+19) | 370 (+61) |
| Sum Eastern | | 438 | 160 |
| Total | | 949 (+142) | 997 (+218) |

Table 3: Analysis of Molecular Variance on North Atlantic minke whales.

| Source of variation | d.f. | % variation | p-value |
|---|---|---|---|
| **Microsatellites** | | | |
| Among IWC stocks | 2 | 0.16 | 0.069 |
| Among localities within IWC stocks | 4 | 0.01 | 0.416 |
| Among individuals within localities | 942 | 0.00 | 0.711 |
| Within individuals | 949 | 99.83 | |
| Total | 1897 | | |
| **mtDNA** | | | |
| Among IWC stocks | 2 | 0.93 | 0.075 |
| Among localities within IWC stocks | 4 | 0.00 | 0.825 |
| Within localities | 990 | 99.07 | |
| Total | 996 | | |

Without any *a priori* assignment of samples to localities/stocks, the most likely number of groups according to STRUCTURE was k=1. Running STRUCTURE with k-values of 2 to 5 did not reveal any substructure in the microsatellite data (data not shown). If samples are assigned to localities and localities to IWC stocks (cf. table 2), the vast majority of the variation (over 99%) is due to variation within individuals (table 3). A small (0.16%), but marginally significant (p=0.069) part of the variation is due to assignment to IWC stocks. Neither pairwise $F_{ST}$-values nor exact tests among localities revealed any significance (data not shown). Pairwise $F_{ST}$-values among stocks reveal that the Eastern stock is slightly differentiated both from the Central ($F_{ST}$=0.002, p<0.001) and the Western one ($F_{ST}$=0.005, p<0.001), while there is no difference between Central and Western stock ($F_{ST}$=0.000, p=0.442). Locus-wise $F_{ST}$ values range from 0 to 0.027, translating into adjusted $F_{ST}'$ values ranging from 0 to 0.157 (table 4). One locus (sam25) exhibits by far the largest $F_{ST}$ value. If this locus is omitted from the AMOVA analysis, there is no significant population structure among stocks ($F_{ST}$=0.000).

*mtDNA*

We first looked for a temporal pattern in the total data set of mtDNA data from 1215 minke whale samples since 1980 in a hierarchical AMOVA as for the microsatellite data (see above). For mtDNA, there was no difference between samples before and after year 2000 ($F_{CT}$=0.000, p=0.603). Nonetheless, samples before 1990 were not further considered in order to base both the microsatellite and the mtDNA analysis on samples from the same time period.

Table 4: Locus-specific $F_{ST}$, $H_E$, and $F_{ST}'$.

| Locus | $F_{ST}$ | $H_E$ W stock | $H_E$ C stock | $H_E$ E stock | Mean $H_E$ | $F_{ST}'$ |
|---|---|---|---|---|---|---|
| **GT310** | 0.000 | 0.674 | 0.681 | 0.676 | 0.677 | 0.000 |
| **GT575** | 0.000 | 0.767 | 0.779 | 0.769 | 0.772 | 0.002 |
| **GT211** | 0.000 | 0.795 | 0.782 | 0.791 | 0.790 | 0.000 |
| **EV094** | 0.000 | 0.330 | 0.329 | 0.310 | 0.323 | 0.000 |
| **GT011** | 0.000 | 0.500 | 0.424 | 0.434 | 0.453 | 0.000 |
| **GT195** | 0.001 | 0.532 | 0.409 | 0.389 | 0.443 | 0.003 |
| **sam25** | 0.027 | 0.834 | 0.858 | 0.791 | 0.828 | 0.157 |
| **GATA028** | 0.002 | 0.843 | 0.835 | 0.819 | 0.832 | 0.011 |
| **GATA098** | 0.001 | 0.691 | 0.727 | 0.728 | 0.715 | 0.002 |
| **GATA417** | 0.000 | 0.837 | 0.858 | 0.852 | 0.849 | 0.000 |
| **EV037** | 0.000 | 0.725 | 0.667 | 0.656 | 0.683 | 0.000 |
| **EV096** | 0.000 | 0.661 | 0.716 | 0.708 | 0.695 | 0.000 |
| **GT023** | 0.002 | 0.821 | 0.760 | 0.773 | 0.785 | 0.010 |
| **EV001** | 0.000 | 0.823 | 0.835 | 0.819 | 0.826 | 0.000 |
| **GT509** | 0.000 | 0.837 | 0.820 | 0.821 | 0.826 | 0.000 |
| **GATA053** | 0.000 | 0.194 | 0.160 | 0.155 | 0.170 | 0.000 |

Among the remaining 997 specimens analyzed, there were 66 different mtDNA haplotypes found. If samples are assigned to sampling locations and locations to IWC stocks (cf. table 2), the vast majority of the variation (over 99%) is due to variation within localities (table 3). A small (0.93%), but marginally significant (p=0.075) part of the variation is due to assignment to IWC stocks. Pairwise $F_{ST}$-values among localities are all non-significant (data not shown), while exact tests among localities were all highly significant (p<0.001 in all pairwise comparisons except for North Sea vs. East Greenland with p=0.007).

Pairwise $F_{ST}$-values among stocks reveal that the Eastern stock is slightly differentiated from the Western one ($F_{ST}$=0.006, p=0.061), while there is no significant difference, neither between Eastern and Central stock ($F_{ST}$=0.003, p=0.142) nor between Central and Western stock ($F_{ST}$=0.000, p=0.820).

## DISCUSSION

Jointly analyzing samples from a period of over 30 years (1980 – 2012) reveals a subtle, but significant change in microsatellite allele frequencies in the largest local sample (i.e., Iceland). In order not to obscure potential spatial structure by temporal variation, the oldest samples (before 1990) were excluded from further analyses.

Based on the remaining almost 1,000 specimens sampled between 1990 and 2012, it became evident that the vast amount of genetic variation is assigned to the lowest level of analysis (individuals for microsatellites, sampling locations for mtDNA). Within the traditional IWC stocks, there is no differentiation among locations. Among these stocks there is a consistent tendency towards a subtle differentiation: West Greenland and North Eastern stock are differentiated in both marker systems, the Central stock (East Greenland/Iceland) stands intermediate, with a closer affinity to the West Greenland than to the Northeastern stock. It should be emphasized – however - that the observed differentiation among IWC stocks is only subtle. Moreover, locus-specific analysis revealed that the inferred differentiation at microsatellites is due to a single locus (sam25; Figure 2).

A deviation from this general pattern is observed in the exact tests on mtDNA haplotype distribution, which were significant among all pairs of locations. Mitochondrial DNA is clonally maternally inherited from mother to offspring. Hence, any matrilineal association of specimens in any area would translate into a higher-than-random occurrence of that particular haplotype. Consequently, significance in this measure may reflect some matrilineal social structure, rather than true differentiation among stocks/populations. Here, relatedness analysis (e.g., Skaug et al. 2010) will reveal whether there are dyads of related individuals at some locations.

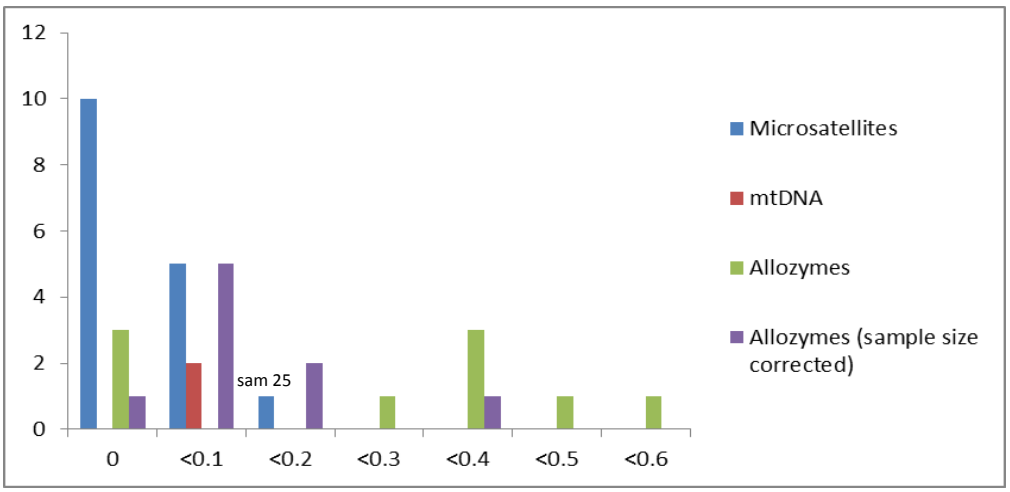


Figure 2: Absolute numbers of locus-specific $F_{ST}$'-values (i.e., $F_{ST}$ values corrected for within population diversity). Allozyme $F_{ST}$'-values were calculated from Danielsdóttir et al. (1993). mtDNA $F_{ST}$'-values were determined twice, i.e., once from data of this study and once from Andersen et al. (2003). sam 25 is the microsatellite locus with highest $F_{ST}$' value.

The level of mtDNA differentiation found in our study is similar to that found in a previous study on North Atlantic minke whales with a similar geographical coverage, i.e., $F_{ST}$' = 0.008 (this study) resp. 0.018 (recalculated from Andersen et al. 2003). Locus-specific microsatellite $F_{ST}$' values range from 0 to 0.157 (highest value for microsatellite locus sam25). Some locus specific $F_{ST}$' values calculated from an allozyme study of the same area (Danielsdóttir et al. 1993) reach a considerably higher level (up to 0.508). A possible explanation might be the uneven sampling scheme of that study, having 50, 114, but only 12 samples from W, C, and E stock, respectively. Here, $F_{ST}$ was re-calculated locus-wise

$$F_{ST} = \frac{H_{Etotal} - H_{Epop}}{H_{Etotal}} \qquad \text{(Hedrick 2005)} \qquad (2)$$

where $H_{Etotal}$ and $H_{Epop}$ are the expected heterozygosity of the entire data set and the mean expected heterozygosity of the populations, respectively. In this calculation, the impact of allele frequencies on $H_{Etotal}$ are much larger for a large rather than a small sample, even if all $H_E$ values are corrected for sample size. If we estimate $H_{Etotal}$ instead from the mean of the population-specific allele frequencies (thereby weighting populations equally regardless of sample size), the allozyme $F_{ST}$' values range from 0 to 0.330 and are much more similar to the microsatellite $F_{ST}$' values, i.e., most locus-specific values are below 0.1 (Figure 2).

In summary, the outcome of this study is generally compatible with the current IWC stock definition in North Atlantic minke whales, i.e., a Western (W), a Central (C), and an Eastern (E) stock. Within these stocks, no indication of subdivision was detected. The difference between W and E was found in both microsatellite and mtDNA data, while C and E where only significantly differentiated in the microsatellite data. Note here, that (1) all significances in the microsatellite data set were due to a single locus and (2) no analysis revealed any differentiation between W and C stock. Our data hence do not contradict a two stock hypothesis (W+C, E) either. Note – however – that all our W samples originate from Western Greenland (WG), whereas no samples are included from the Canadian East coast (WC). In addition, the WG sample for microsatellites is still rather small (47 samples since 1990), but the analysis of further 300 samples is underway. This will allow for more precise estimates of microsatellite differentiation among the W and the other stocks.

**Acknowledgements**

Lab assistance was provided by Ulrike Matthes. Financial Support is acknowledged from the University of Potsdam and the Greenland Institute of Natural Resources in Nuuk.

**References**

Andersen LW, Born EW, Dietz R, Haug T, Øien N, Bendixen C (2003) Genetic population structure of minke whales *Balaenoptera acutorostrata* from Greenland, the North East Atlantic and the North Sea probably reflects different ecological niches. Marine Ecology Progress Series 247, 263-280.

Anderwald P, Daníelsdóttir AK, Haug T, Larsen F, Lesage V, Reid RJ, Víkingsson GA, Hoelzel AR (2011) Possible cryptic stock structure for minke whales in the North Atlantic: Implications for conservation and management. Biological Conservation 144, 2479-2489.

Danielsdóttir AK, Duke EJ, Arnason A (1992) Genetic variation at enzyme loci in North Atlantic minke whales, *Balaenoptera acutorostrata*. Biochemical Genetics 30, 189-202.

Donovan GP (1991) A review of IWC stock boundaries. Report of the International Whaling Commission (special issue 13), 39-68.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10, 564–567.

Hedrick PW (2005) Genetics of populations, 3rd edition. Jones and Bartlett, Sudbury, MA.

Pampoulie C, Danielsdóttir AK, Víkingsson GA (2008) Genetic structure of the North Atlantic common minke whale (*Balaenoptera acutorostrata*) at feeding grounds: a microsatellite loci and mtDNA analysis. IWC SC/F13/SP17 SC/60/PFI10.

Pritchard JK, Stephens M, Donnely P (2000) Inference of population structure using multilocus genotype data. Genetics 155, 945-959.

Skaug HJ, Bérubé M, Palsbøll P (2010) Detecting dyads of related individuals in large collections of DNA-profiles by controlling the false discovery rate. Molecular Ecology Resources 10, 693-700.

Víkingsson GA, Heide-Jørgensen MP (2014) First indications of autumn migration routes and destination of common minke whales tracked by satellite in the North Atlantic during 2001 - 2011. Mar Mammal Sci In Press