SC/68D/SD & DNA/04

Sub-committees/working group name: SD & DNA

Population genetic structure in the minke whale, Balaenoptera acutorostrata acutorostrata in the North East Atlantic: Fifteen years of data

Maria Quintela Sanchez, Hans J. Skaug

Population genetic structure in the minke whale, *Balaenoptera acutorostrata acutorostrata* in the North East Atlantic: Fifteen years of data

Maria Quintela Sanchez
Institute of Marine Research, PO box 1870, Nordnes N-5817, Bergen, Norway

Hans J. Skaug

Department of Mathematics, University of Bergen, Bergen, Norway

INTRODUCTION

Norway conducts a commercial harvest of minke whale, *Balaenoptera acutorostrata acutorostrata* in the Northeast Atlantic and, each year, approximately 500 whales are captured across five IWC Management Areas (see Fig. 1). In order to enforce domestic regulation and compliance within this harvest, an individual-based DNA register (NMDR) has been maintained since 1996 (Glover *et al.* 2012). This register contains information on ten microsatellite loci and mtDNA, as well as biometric data and the geographic position of all the catches.

The population structure of the North Atlantic minke whale has been the object of debate (e.g. Andersen *et al.* 2003; Anderwald *et al.* 2011). In 2014, the study of 2990 whales harvested in the NE Atlantic IWC Management Areas during the period 2004 and 2007–2011 and genotyped at ten microsatellite loci and 331 bp of the mitochondrial D-loop concluded that no spatial or temporal genetic differentiation was observed at any of those markers. Likewise, the combination of the information provided by nuclear loci and *in silico* simulations of random mating showed clear evidence of a single panmictic population in the minke whale in the NE Atlantic (Quintela *et al.* 2014).

The objective of the present study is to extend the temporal scope of Quintela *et al.* (2014) until 2020 and assess if any spatial or temporal variation in the genetic structure of the Atlantic minke whale occurred in the last years.

MATERIAL AND METHODS

Sampling and microsatellite genotyping

A total of 7755 whales (5663 females and 2092 males) were harvested between April and September across fifteen years; *i.e.* 2004 and the period 2007-2020. DNA was extracted twice from muscle stored in ethanol using Qiagen DNeasy Blood & Tissue Kit following manufacturer´s instructions and DNA concentration was measured on a Nanodrop. Ten microsatellite loci: EV1*Pm*, EV037*Mn* (Valsecchi & Amos 1996); GATA028, GATA098, GATA417 (Palsbøll *et al.* 1997); GT023, GT211, GT310, GT509, GT575 (Bérubé *et al.* 2000) were amplified in three multiplex reactions based on a 2 minute hot start at 94°C, denaturizing for 20 seconds at 94°C, annealing for 45 seconds, elongation at 72°C for 1 minute and a final hold at 4°C. Multiplex specific conditions are detailed in Glover *et al*. (2012). Individuals were sexed using specific primers for the ZFY/ZFX gene (Bérubé & Palsbøll 1996).

Statistical analysis

The total number of alleles, number of private alleles, and allelic richness per harvest year was calculated using MSA 4.05 (Dieringer & Schlötterer 2003). The observed (H_o) and unbiased expected heterozygosity (uH_e) as well as the inbreeding coefficient (F_{IS}) were computed for each sample with GenAlEx v6.1 (Peakall & Smouse 2006). The genotype frequency of each locus and its direction (heterozygote deficit or excess) was compared with Hardy-Weinberg expectations (HWE) using the program Genepop (Rousset 2008) as was linkage disequilibrium (LD) between pairwise loci. Bonferroni correction was applied to p-values to control for Type I errors.

Genetic structure assessed using pairwise F_{ST} (Weir & Cockerham 1984) was computed with Arlequin v.3.5.1.2 (Excoffier *et al.* 2005) on the harvest year basis as well as on the Small Management Areas (SMA), the latter by combining different years as well as on the year basis. The relationship among samples were also examined using Principal Components Analysis (PCA) and the Discriminant Analysis of Principal Components (DAPC) (Jombart *et al.* 2010) implemented in *adegenet* (Jombart 2008) in which groups were defined *a priori* using a) harvest year and b) management areas. A number of principal components ranging from 40 to 120 was tested to determine the optimal number of PCs to avoid overfitting of the data and creating artificially large separation between groups (Jombart & Collins 2015; Miller *et al.* 2020). The trade-off between power of discrimination and overfitting was measured using the a-score, which is the difference between the proportion of successful reassignment of the analysis (observed discrimination) and values obtained using random groups (random discrimination); in other words, the proportion of successful reassignment corrected for the number of retained PCs.

RESULTS

The numbers of individuals harvested per year ranged between 424 (in 2019) and 718 (in 2014). The average ratio F:M was 2.7, spanning between 2 and 3.6, except in 2017, when the number of captured females was 6 fold-larger. The geographic distribution of the catches according sex was quite similar

although females seemed to venture slightly further north (Fig. 2); *i.e.* no male was captured at >81.18 °N in opposition to some 38 females. The number of individuals captured on the SMA basis was very uneven. Thus, the proportion of the catches per area between 2004 and 2020 was as follows: CM (0.6%), EN (5.6%), EB (13.3%), EW (27.9%) and ES (52.5%), see pie chart in Figure 1 and Table 1.

Genetic variation as computed through allelic richness, H_o or uH_e took very similar values across different years (Table 2) and no signs of inbreeding were detected as F_{IS} was centred in zero. None of the pairwise comparisons conducted on the year basis turned a F_{ST} value that was significantly different from zero (Table 3). Likewise, and in spite of the unevenness in sampling sizes, no genetic differentiation was registered between SMA (Table 4). The differentiation among SMA was also computed on the year basis showing no statistical differentiation either overall (Table 5) or pairwise.

In alignment with these results, the PCA plot revealed a major overlapping of the individuals harvested across these fifteen years (Fig. 3) as well as across SMA (Fig. 4). The same lack of differentiation was revealed through DAPC in spite of using groups defined *a priori*: none of the axis of variation managed to discriminate any of the year harvests (Fig. 5) nor the SMA (Fig. 6).

DISCUSSION

The suite of statistical approaches used to analyse the minke whales harvested in the period 2004 and 2007-2020 and genotyped at ten microsatellite loci revealed a total lack of temporal differentiation, in alignment with the results obtained for the period 2004 and 2006-2010 (Quintela *et al.* 2014). Likewise, no genetic differentiation was detected across Small Management Areas, thus supporting previous results (Quintela *et al.* 2014).

The temporal follow-up presented here confirms further the idea of minke whale in the NE Atlantic being a panmictic entity. However, given that the suite of microsatellites used here screens a very small proportion of the genome, it is worth noting that a different outcome can never be ruled out when using powerful genomic tools with higher discriminating capacity.

REFERENCES

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

Anderwald P, Daníelsdóttir AK, Haug T*, et al.* (2011) Possible cryptic stock structure for minke whales in the North Atlantic: Implications for conservation and management. *Biological Conservation* **144**, 2479-2489.

Bérubé M, Jørgensen H, McEwing R, Palsbøll PJ (2000) Polymorphic di-nucleotide microsatellite loci isolated from the humpback whale, *Megaptera novaeangliae*. *Molecular Ecology* **9**, 2181-2183.

Bérubé M, Palsbøll P (1996) Identification of sex in cetaceans by multiplexing with three ZFX and ZFY specific primers. *Molecular Ecology* **5**, 283-287.

Dieringer D, Schlötterer C (2003) MICROSATELLITE ANALYSER (MSA): A platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* **3**, 167-169.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.

Glover KA, Haug T, Øien N*, et al.* (2012) The Norwegian minke whale DNA register: a data base monitoring commercial harvest and trade of whale products. *Fish and Fisheries* **13**, 313-332.

Jombart T (2008) *adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403-1405.

Jombart T, Collins C (2015) A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0.0.

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

Miller JM, Cullingham CI, Peery RM (2020) The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. *Heredity* **125**, 269-280.

Palsbøll PJ, Bérubé M, Larsen AH, Jørgensen H (1997) Primers for the amplification of tri- and tetramer microsatellite loci in baleen whales. *Molecular Ecology* **6**, 893-895.

Peakall R, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Quintela M, Skaug HJ, Øien N*, et al.* (2014) Investigating population genetic structure in a highly mobile marine organism: The minke whale *Balaenoptera acutorostrata acutorostrata* in the North East Atlantic. *PLOS ONE* **9**, e108640.

Rousset F (2008) GENEPOP'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Valsecchi E, Amos W (1996) Microsatellite markers for the study of cetacean populations. *Molecular Ecology* **5**, 151-156.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

FIGURES

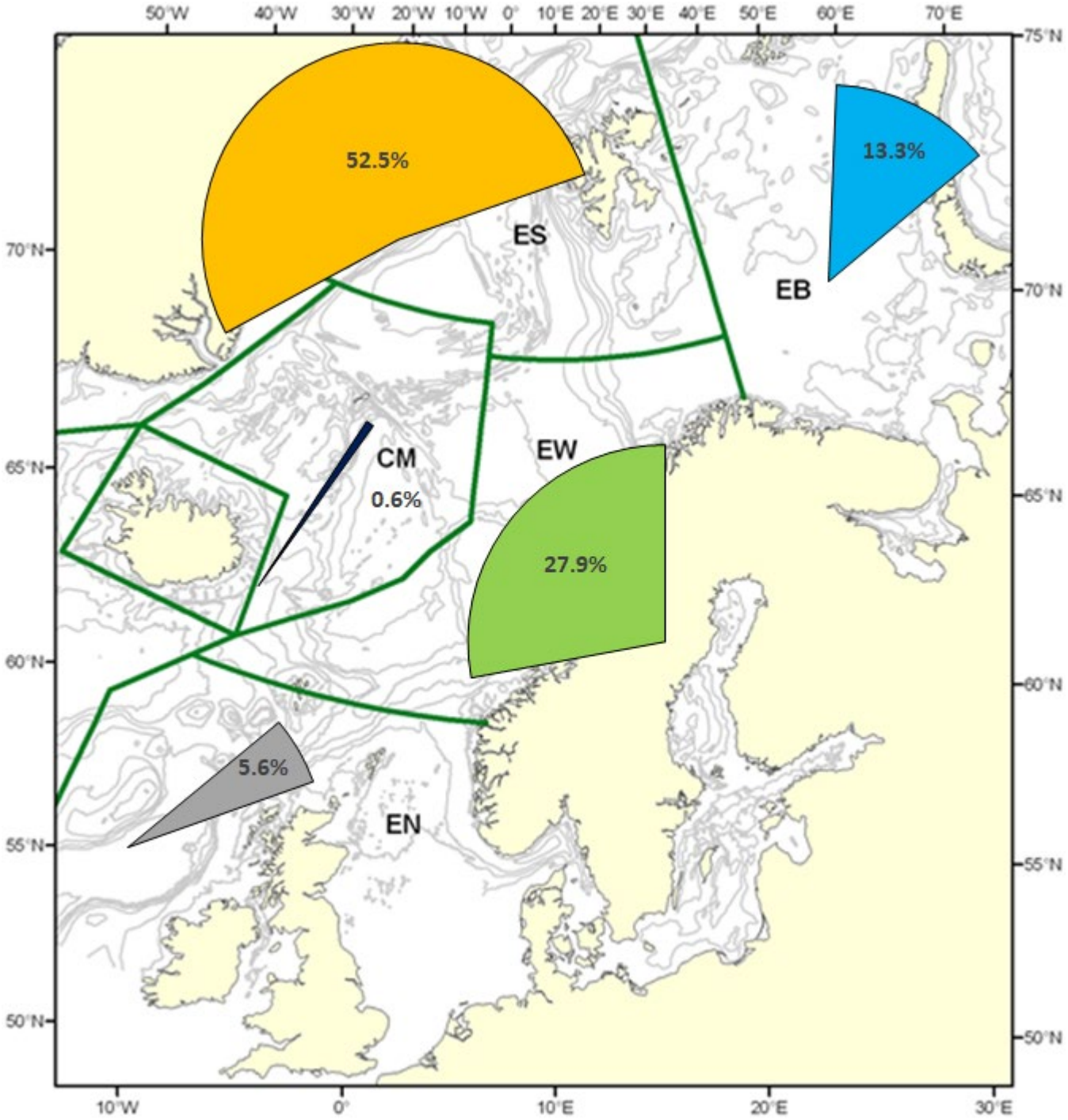

Fig 1. Geographic distribution of the five International Whaling Commission (IWC) Management Areas: ES (Svalbard-Bear Island area), EB (Eastern Barents Sea), EW (Norwegian Sea and coastal zones off North Norway, including the Lofoten area), EN (North Sea), and CM (Western Norwegian Sea-Jan Mayen area). In each Small Management Area, the section of a pie chart representing the proportion of individuals harvested in the period from 2004 and 2007-2020 is depicted.

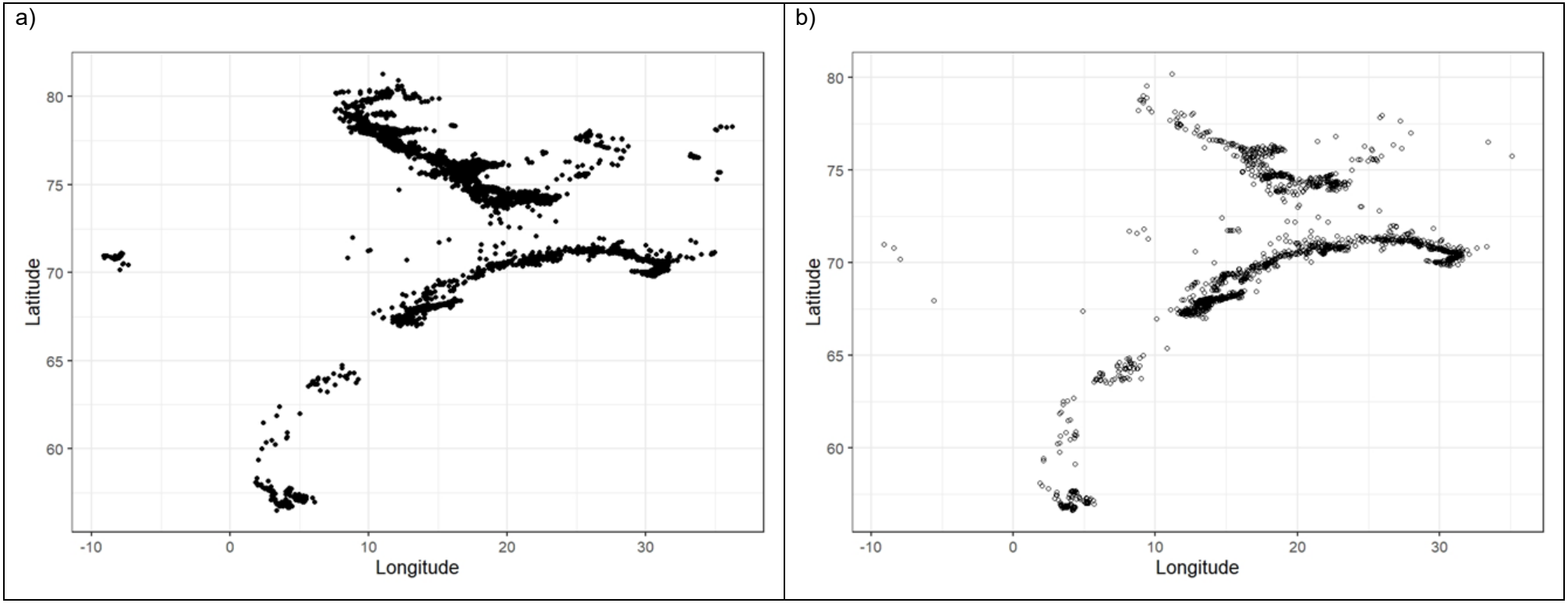

Fig 2. Geographic position of the catches for a) females (N=5663) and b) males (N=2092) harvested in the period 2004 and 2007-2020.

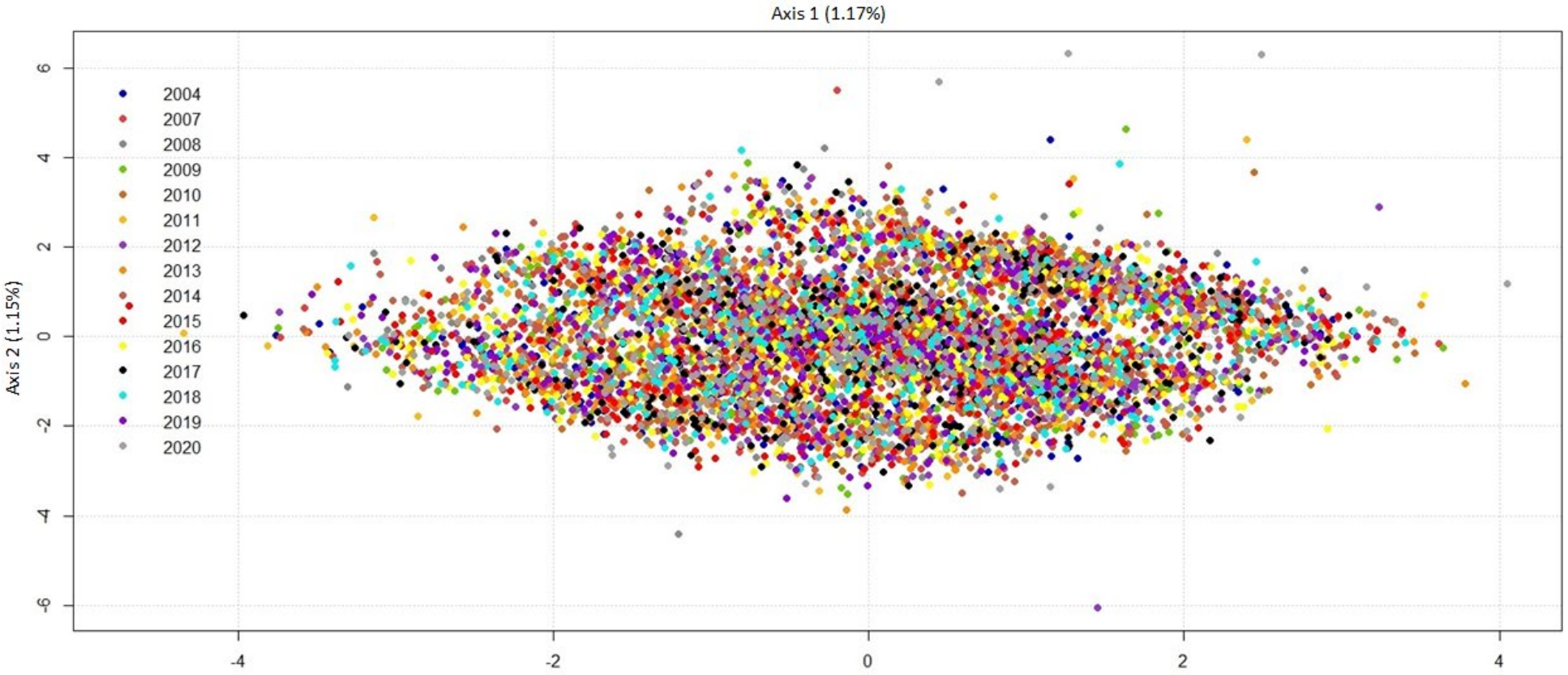

Fig. 3. Principal Component Analyses for the samples coloured according harvesting years.

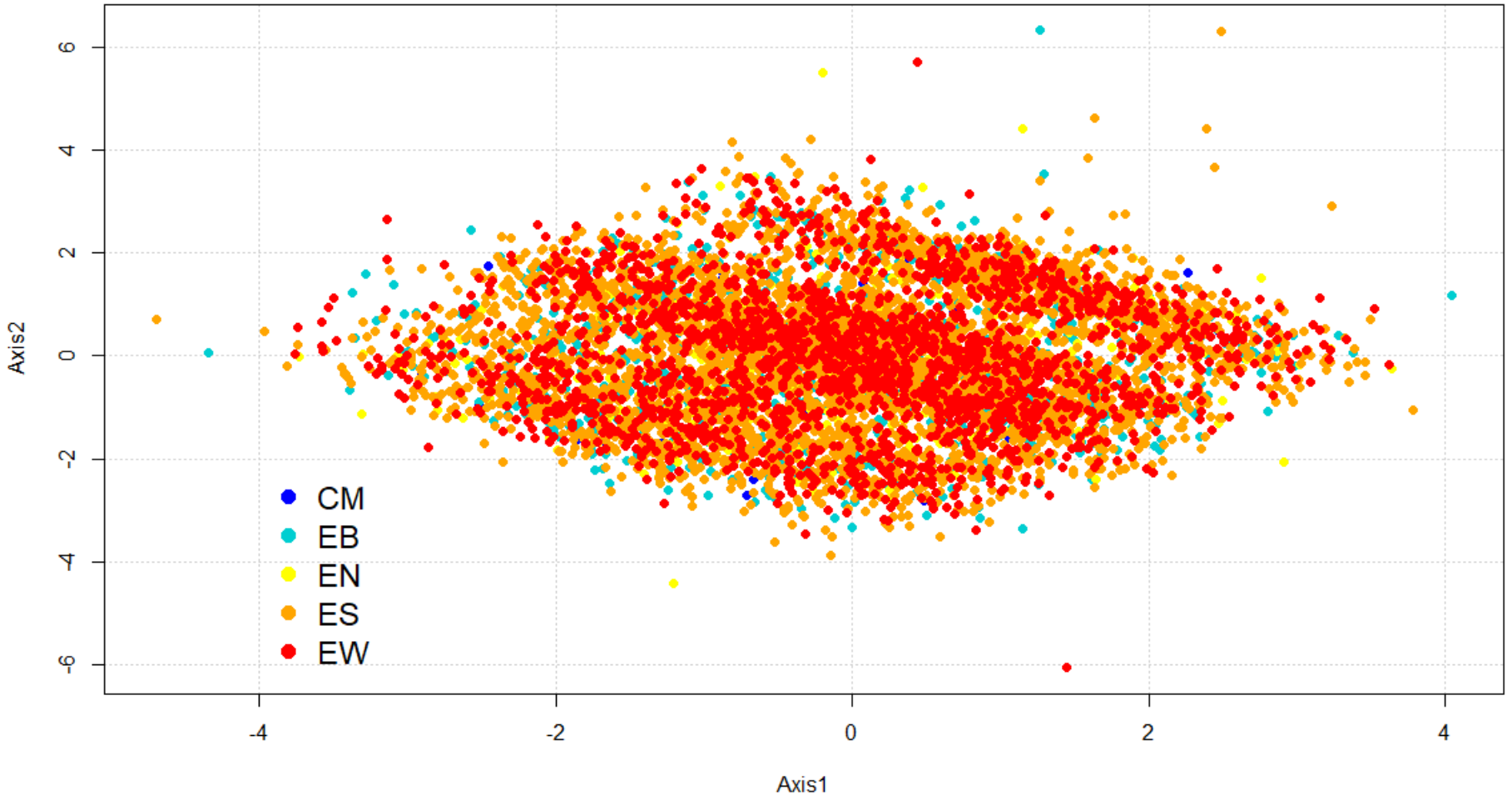

Fig. 4. Principal Component Analyses for the samples coloured according Small Management Areas between 2004 and 2020. Note that the proportion of individuals captured in each of them is extremely uneven (see Table 2). Axis 1 and 2 explain 1.17 and 1.15% of the variation, respectively.

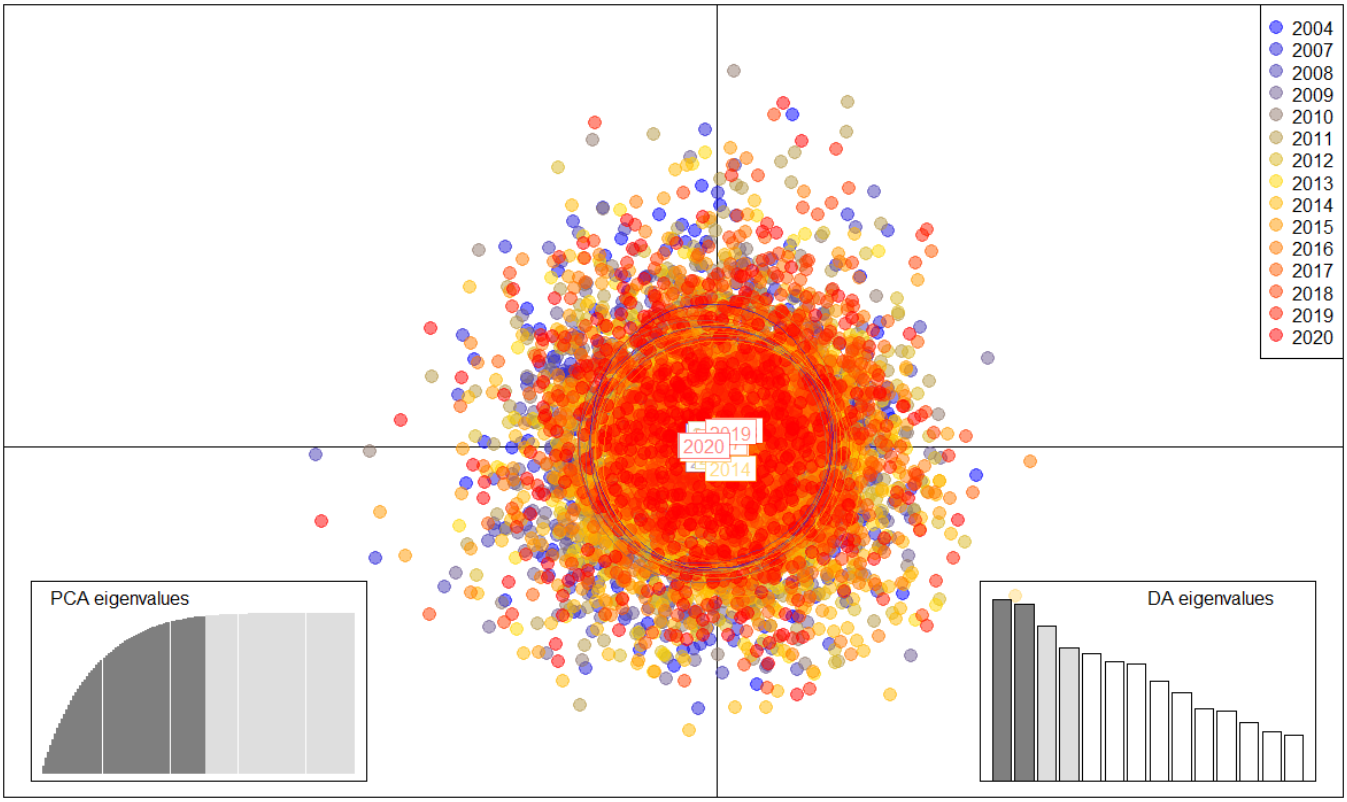

Fig. 5. Discriminant Analysis of Principal Components (DAPC) after retaining 70 principal components. Individuals from different harvesting years are represented by coloured dots, and inertia ellipses are centred on the mean of each geographically-explicit sample. The first axis explains 12.1% of the variation whereas the second one accounts for 11.9%.

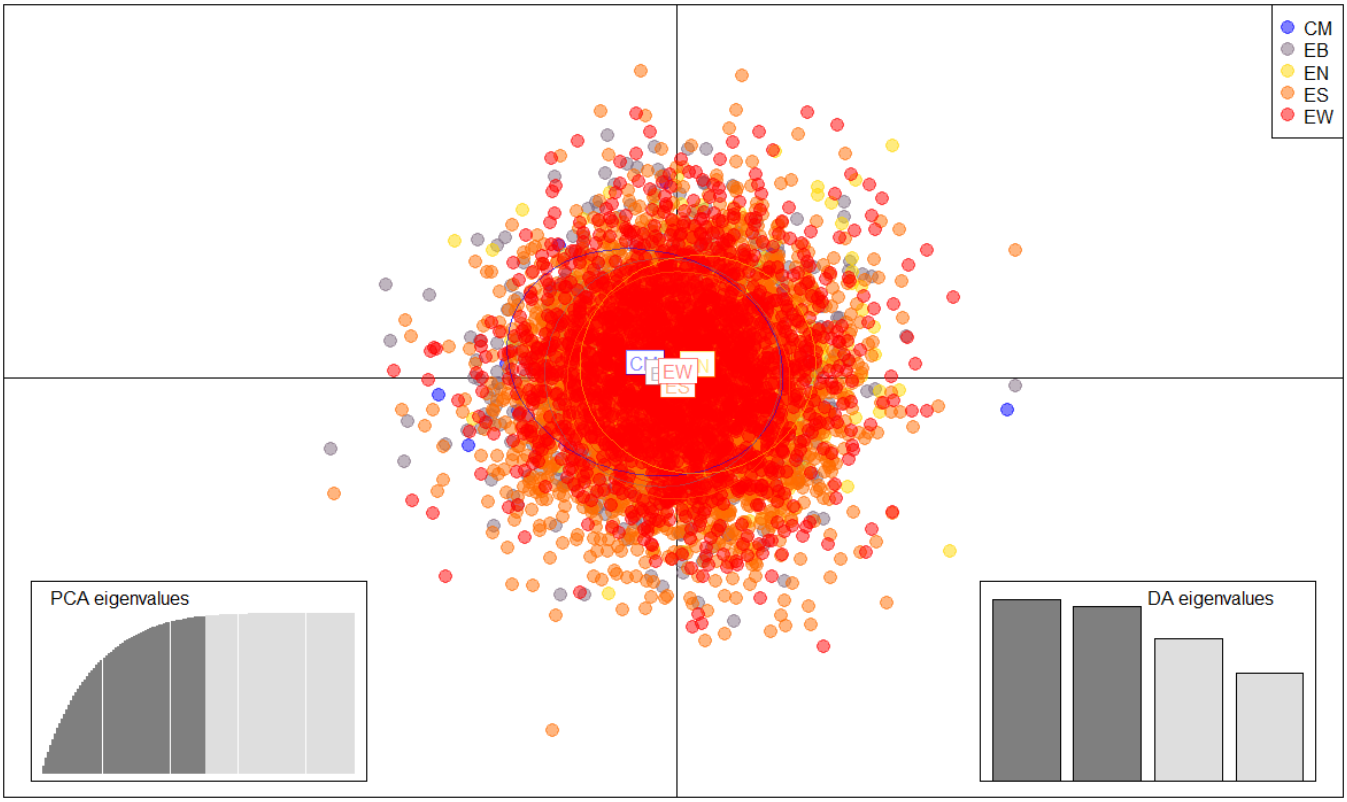

Fig. 6. Discriminant Analysis of Principal Components (DAPC) after retaining 70 principal components. Individuals from different Small Management Areas are represented by coloured dots, and inertia ellipses are centred on the mean of each geographically-explicit sample. The first axis explains 33.6% of the variation whereas the second one accounts for 24.5%.

TABLES

Table 1. Number of individuals harvested per Small Management Area on the year basis.

		Small Management Area				
Year	**No TOTAL**	**CM**	**EB**	**EN**	**ES**	**EW**
2004	515	17	123	81	109	185
2007	567	0	28	91	276	172
2008	498	30	20	86	220	142
2009	466	0	3	49	243	171
2010	449	1	17	27	264	140
2011	495	0	96	12	184	203
2012	443	0	5	8	235	195
2013	579	0	65	2	278	234
2014	718	0	106	17	362	233
2015	650	0	92	4	425	129
2016	578	0	59	15	421	83
2017	427	0	15	6	373	33
2018	446	0	124	0	209	113
2019	424	0	108	3	239	74
2020	500	0	170	34	237	59
SUM	**7755**	**48**	**1031**	**435**	**4075**	**2166**

Table 2. Sample summary statistics per harvest year: Number of individuals (N), total number of alleles, number of private alleles, allelic richness (Ar) based on a minimum sample of 424 diploid individuals; observed heterozygosity, H_o (mean ± SE); unbiased expected heterozygosity, uH_e (mean ± SE); inbreeding coefficient, F_{IS} (mean ± SE); number of deviations from Hardy-Weinberg equilibrium (HWE) at α=0.05; number of deviations from Linkage Disequilibrium (LD) at α=0.05 both before and after Bonferroni (B) correction.

Harvest year	No ind	No alleles	No private alleles	Ar	*Ho*	u*He*	F_{IS}	No dev HWE (B)	No dev LD (B)
2004	515	119	0	11.7	0.770 ± 0.022	0.775 ± 0.020	0.005 ± 0.008	1 (1)	2 (1)
2007	567	120	0	11.7	0.775 ± 0.021	0.772 ± 0.020	-0.005 ± 0.006	3 (2)	5 (0)
2008	498	123	1	12.0	0.786 ± 0.022	0.775 ± 0.022	-0.015 ± 0.006	1 (1)	5 (1)
2009	466	124	0	12.2	0.777 ± 0.022	0.774 ± 0.022	-0.006 ± 0.006	1 (1)	4 (0)
2010	449	116	0	11.5	0.765 ± 0.021	0.773 ± 0.022	0.009 ± 0.007	2 (1)	2 (0)
2011	495	116	1	11.4	0.777 ± 0.022	0.773 ± 0.021	-0.006 ± 0.005	1 (0)	2 (0)
2012	443	118	1	11.7	0.779 ± 0.021	0.774 ± 0.021	-0.007 ± 0.005	0 (0)	3 (1)
2013	579	116	0	11.2	0.776 ± 0.021	0.777 ± 0.021	0.001 ± 0.011	3 (2)	5 (0)
2014	718	124	0	11.8	0.776 ± 0.018	0.776 ± 0.020	-0.002 ± 0.005	0 (0)	1 (0)
2015	650	124	1	11.9	0.778 ± 0.023	0.773 ± 0.022	-0.007 ± 0.006	2 (1)	4 (0)
2016	578	118	0	11.5	0.776 ± 0.018	0.775 ± 0.020	-0.004 ± 0.007	1 (1)	2 (1)
2017	427	123	1	12.3	0.785 ± 0.022	0.776 ± 0.020	-0.012 ± 0.010	0 (0)	8 (2)
2018	446	120	0	11.9	0.777 ± 0.020	0.771 ± 0.021	-0.010 ± 0.010	2 (0)	4 (0)
2019	424	117	0	11.7	0.765 ± 0.023	0.773 ± 0.021	0.009 ± 0.005	0 (0)	3 (0)
2020	500	121	0	11.9	0.779 ± 0.022	0.777 ± 0.020	-0.003 ± 0.008	1 (1)	7 (0)

Table 3. Pairwise F_{ST} values across harvesting years. None of the comparisons was significantly different from zero.

	2004	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
2004	*														
2007	0.0000	*													
2008	0.0002	0.0004	*												
2009	0.0000	0.0000	0.0002	*											
2010	0.0000	0.0000	0.0002	0.0000	*										
2011	0.0001	0.0000	0.0000	0.0000	0.0002	*									
2012	0.0001	0.0000	0.0004	0.0000	0.0000	0.0000	*								
2013	0.0001	0.0001	0.0000	0.0000	0.0002	0.0000	0.0000	*							
2014	0.0001	0.0000	0.0004	0.0000	0.0004	0.0001	0.0000	0.0001	*						
2015	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	*					
2016	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	*				
2017	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*			
2018	0.0003	0.0002	0.0006	0.0003	0.0006	0.0001	0.0001	0.0003	0.0004	0.0005	0.0000	0.0000	*		
2019	0.0001	0.0001	0.0005	0.0001	0.0000	0.0000	0.0000	0.0002	0.0002	0.0001	0.0001	0.0000	0.0001	*	
2020	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0003	0.0000	*

Table 4. Pairwise F_{ST} values across Small Management Areas between 2004 and 2007-2020. Given the extremely uneven sampling sizes (see Table 1), a subset of 400 randomly chosen individuals was selected per area (with the exception of CM where N=48). None of the comparisons was significantly different from zero (*P* values ranging between 0.161 and 0.933). Likewise, the overall F_{ST} was 0.0001 (P=0.787).

	CM	**EB**	**EN**	**ES**	**EW**
CM	*				
EB	0.0000	*			
EN	0.0000	0.0000	*		
ES	0.0000	0.0000	0.0002	*	
EW	0.0000	0.0000	0.0003	0.0000	*

Table 5. Overall genetic differentiation among Small Management Areas per year class. It is worth noting that none of the pairwise comparisons between SMA conducted on the year basis was significantly different from zero.

Year	F_{ST}	*P*-value
2004	0.0000	0.6836
2007	0.0000	0.6964
2008	0.0000	0.9613
2009	0.0004	0.2406
2010	0.0000	0.9374
2011	0.0000	0.9867
2012	0.0009	0.1174
2013	0.0000	0.6982
2014	0.0000	0.9769
2015	0.0000	0.9464
2016	0.0003	0.3103
2017	0.0000	0.8521
2018	0.0000	0.7810
2019	0.0008	0.1664
2020	0.0000	0.8145