Int J Leg Med (1994) 106:254–257

International Journal of
Legal Medicine

DNA polymorphism in Greenland

Allele and profile frequencies in a Greenland population sample using the VNTR probes MS1, MS31, MS43a and YNH24

Birthe Eriksen, Ole Svensmark

Institute of Forensic Genetics, University of Copenhagen, 11 Frederik V's Vej, DK-2100 Copenhagen, Denmark

Received September 15, 1993 / Received in revised form December 16, 1993

Summary. Allele frequencies obtained by RFLP (Restriction Fragment Length Polymorphism) analysis using the VNTR (Variable Number Tandem Repeats) single locus probes MS1, MS31, MS43a and YNH24 on *Hin*fI-restricted DNA in a Greenland Eskimo (155 individuals) and a Danish Caucasian (616 individuals) population sample are reported. For MS1 the frequency distributions were almost identical whereas minor but significant differences were seen for the other 3 probes. The distribution of the frequencies of 139 Greenland complete DNA profiles was estimated using the Greenland and the Danish database. The average profile frequencies obtained with the Greenland database were approximately 10 times higher than the estimates obtained with the Danish database.

Key words: Allele frequencies – Profile frequencies – VNTR single locus probes – Greenland population

Zusammenfassung. Die Allel-Frequenzen von RFLP (Restriktions-Fragmentlängen-Polymorphismen), nachgewiesen mit den VNTR (Variable Number of Tandem Repeats) Single-Locus Proben MS1, MS31, MS43a und YNH24 an *Hin*fI-gespaltene DNA einer Populationsstichprobe von Grönland-Eskimos (n = 155 Individuen) und dänischen Kaukasiern (n = 616 Individuen) werden mitgeteilt. Für MS1 sind die Frequenz-Verteilungen nahezu identisch, während geringere, aber signifikante Unterschiede bei den anderen drei Sonden zu beobachten sind. Die Verteilung der Frequenzen der 139 kompletten Grönland-DNA-Profile wurde bestimmt mit Hilfe der grönländischen und dänischen Datenbank. Die durchschnittlichen Profil-Frequenzen, welche mit Hilfe der grönländischen Datenbank erhalten wurden, waren ungefähr 10-fach höher als jene unter Verwendung der dänischen Datenbank.

Schlüsselwörter: Allel-Frequenzen – Profil-Frequenzen – VNTR Single-Locus-Sonden – Grönländische Population

Correspondence to: B. Eriksen

Introduction

It was the object of the present study to compare a Greenland Eskimo and a Danish Caucasian population sample of restriction fragments (*Hin*fI) obtained with the VNTR probes MS1, MS31, MS43a and YNH24. For this purpose the distributions of the allele and profile frequencies were compared. To obtain normally distributed measurement errors independent of fragment length the fragment lengths were transformed into normalized migration distances (Eriksen et al. 1992). The cumulative frequency distributions (Eriksen and Svensmark 1993) were compared and the Kolmogorov-Smirnov test applied as a test of significance. The frequencies of 8-band profiles (n= 139) from the Greenland population sample were estimated with a Greenland population sample (300 bands for each probe) and with a Danish population sample (1200 bands for each probe) as reference samples. A preliminary account of part of this study has been published elsewhere (Thymann et al. 1993).

Material and methods

Blood samples from 616 unrelated Danish and 155 Greenland individuals involved in criminal cases or paternity cases were collected. The individuals involved in criminal cases were selected from Greenland cases by names so that only persons with characteristic Eskimo names were included. The individuals related to paternity cases were born in Greenland. RFLP analysis was performed on *Hin*fI (Boehringer) restricted DNA. The probes MS1, MS31, MS43a (Cellmark Diagnostics) and YNH24 (Promega Corporation) were used together with the SJ5000 (Amersham) or the Gibco BRL NICE DNA Analysis Ladder size markers. Preparation of DNA, RFLP analysis and calculation of fragment lengths were carried out as described previously (Eriksen et al. 1992). Duplicate determinations were performed on different gels. To obtain data with normally distributed measurement errors independent of the fragment length, all fragment lengths were transformed into normalized migration distances (Eriksen et al. 1992) according to the function:

$$m = 760/(3.7 + b^{1.5}) + 32.3$$

Table 1. Number of bandpairs, single band patterns and frequency of apparent homozygosity in a Greenland and a Danish database

Database Probe	Number of bandpairs	Apparent homozygotes	
		Number	Frequency
Greenland			
MS1	149	5	0.03
MS31	150	17	0.11
MS43a	140	14	0.10
YNH24	147	10	0.07
Total	586	46	0.08
Denmark			
MS1	603	24	0.04
MS31	600	30	0.05
MS43a	610	56	0.09
YNH24	602	27	0.04
Total	2415	137	0.06

where b is the fragment length in units of kilobasepairs (kb) and m the normalized migration distance in units of millimetres.

The frequency of a DNA profile was estimated from the allele frequencies by counting the number of bands within 3 mm intervals in the appropriate database and dividing by the total number of bands (N). In our laboratory a 3 mm interval corresponds to an interval of ±3 standard deviations (SD) which can be considered a conservative interval for frequency estimates. When the count was zero the frequency f was set to 1/(N+1) as a default allele frequency. If the target profile is in the database it is deleted from the database and replaced after completion of the calculation of the target profile frequency. Otherwise an overestimate of the frequency is obtained. Let f_1 be the allele frequency of band 1 and f_2 the frequency of band 2. Thus, assuming Hardy-Weinberg equilibrium, the frequency p of the band pair is $2f_1f_2$. Let p_1 be the frequency of the pattern obtained with MS1, p_2 with MS31, p_3 with MS43a and p_4 with YNH24, then the frequency of the DNA profile q is $p_1\,p_2\,p_3\,p_4$. In this study single band patterns were treated as close heterozygotes and the frequency calculated as $2f^2$. Cumulative frequency distributions were obtained for each probe by grouping the transformed fragment lengths into 1 mm bins, summing and plotting the cumulative relative frequencies against the fragment length transformed into migration distance according to the function given above.

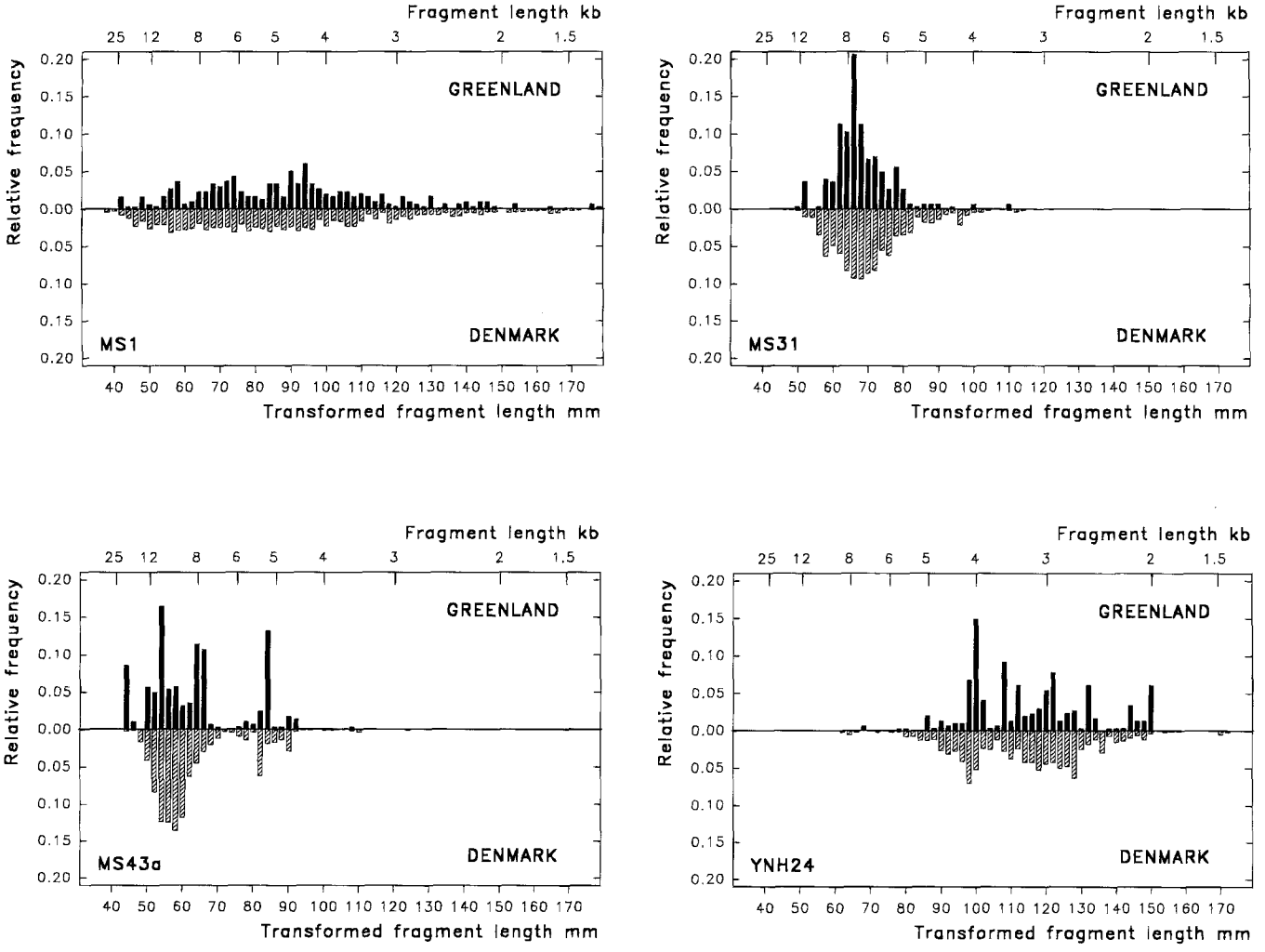

Fig. 1. The distribution of allele frequencies obtained with the probes MS1, MS31, MS43a and YNH24 in two population samples: 1. A Greenland sample consisting of 300 bands for each probe, and 2. A Danish sample consisting of 1200 bands for each probe

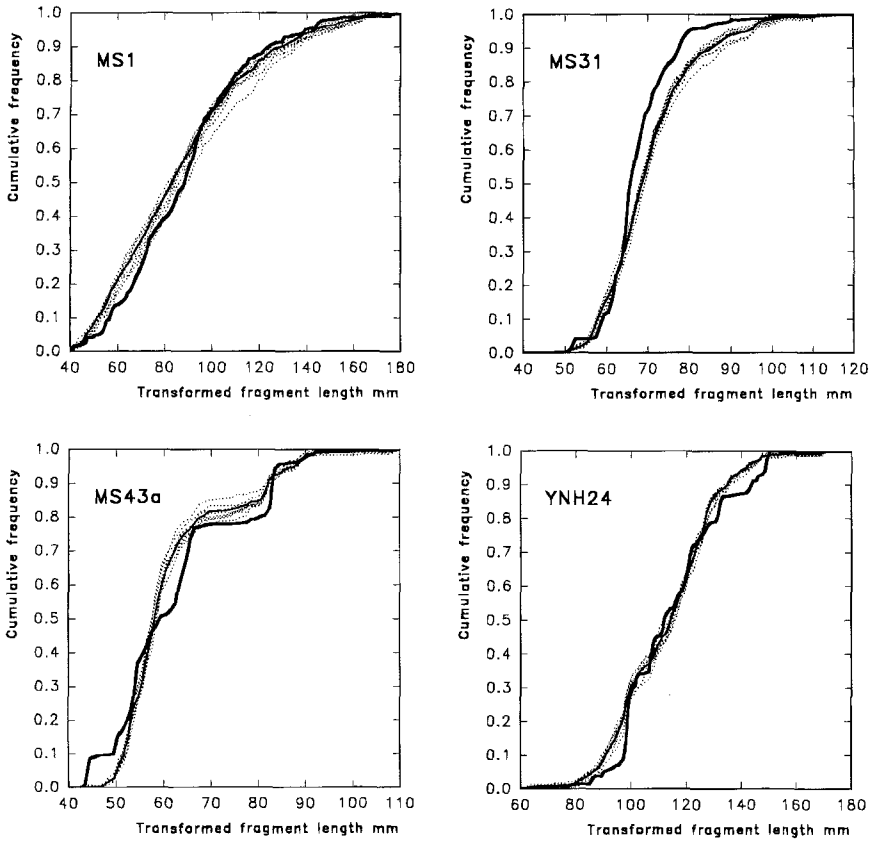

Fig. 2. The cumulative distribution of allele frequencies obtained with the probes MS1, MS31, MS43a and YNH24 in two population samples. *Thick line*: A Greenland sample consisting of 300 bands for each probe; *Thin line*: A Danish sample consisting of 1200 bands for each probe; *Dotted lines*: 10 random samples of 300 bands extracted from the Danish database using random numbers

Results

Distribution of allele frequencies

The number of bands, single band patterns and frequencies of apparent homozygotes for the Greenland and the Danish databases are given in Table 1. For MS31 the frequency of homozygotes, apparent or true, was significantly higher in the Greenland population sample than in the Danish sample. The distributions of the allele frequencies in the 2 population samples are given in Fig. 1. For MS1 the 2 distributions were almost identical although spike frequencies were slightly higher in the Greenland sample than in the Danish sample. This may be due to the small sample size. For MS31 the distributions differed considerably, especially around the peak at 7.5 kb. This peak may explain the high frequency of apparent or true homozygotes (Table 1). Out of 17 homozygotes 9 fell in the 7.5 kb bin and 16 out of 17 in the range 7–8.5 kb where the probability of close heterozygosity or true homozygosity is high. The Greenland distribution of the MS43a alleles differed markedly from the Danish distribution. The spikes at 8 and 11 kb were pronounced, and the spike at 20 kb fell in a region without observations in the Danish population sample. Also the Greenland distribution of YNH24 alleles was more spiky than the Danish distribution.

To investigate whether these differences are real or whether they are due to sampling variance, the cumulative distributions were studied (Fig. 2). For each probe the distributions were plotted for the Greenland population sample (thick lines) and for the Danish sample (thin lines). To demonstrate the effect of sampling errors, the distribution of 10 random samples consisting of 300 bands from the Danish population are shown (dotted lines). The two-sided Kolmogorov-Smirnov test was applied on the data in order to compare the distributions. The test statistic is the maximum vertical distance between the distributions. The critical value at the 0.05 significance level was 0.09

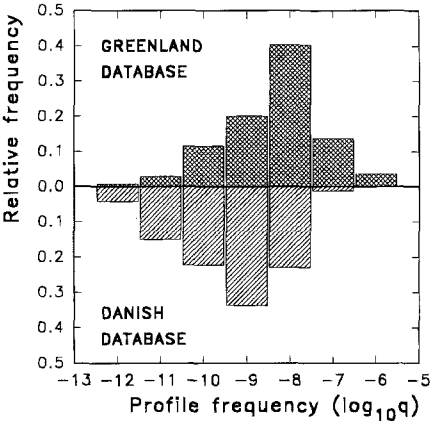

Fig. 3. The distribution of the frequencies of 139 DNA profiles from the Greenland population sample estimated with two different reference samples: 1. A Greenland sample consisting of 300 bands for each probe; 2. A Danish sample consisting of 1200 bands for each probe. The width of the window was 3 mm and the minimum default allele frequency was 1/(N+1). The logarithm of the frequencies ($\log_{10}q$) was used

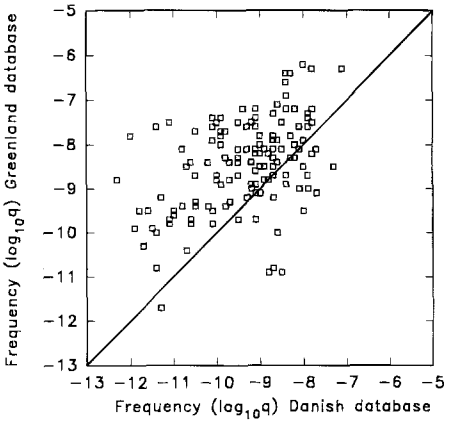

Fig. 4. Correlation of the frequencies of 139 DNA profiles from the Greenland population sample estimated with two different reference samples: 1. A Greenland sample consisting of 300 bands for each probe; 2. A Danish sample consisting of 1200 bands for each probe. The width of the window was 3 mm and the minimum default allele frequency was 1/(N+1). *Abscissa*: The logarithm of the frequency ($\log_{10}$q) as estimated with the Greenland reference sample. *Ordinate*: The logarithm of the frequency ($\log_{10}$q) as estimated with the Danish reference sample

on the ordinate axis. Significant differences between the 1200-band sample and the 10 random 300-band samples were not observed. The analysis confirms that the distributions of MS1 alleles were only slightly different whereas the differences for the other probes were significant and not due to sampling variance alone.

Frequencies of DNA profiles

The frequencies of 139 DNA profiles from the Greenland population sample were estimated with a Greenland reference sample and with a Danish database as reference. The distributions were approximately lognormal (Fig. 3). It appeared that the frequencies were on average approximately 10 times higher with the Greenland reference sample than with the Danish reference sample. The differences between the individual profiles may be as large as 10^4. However, the frequencies generally were less than 10^{-6} in either database (Fig. 4).

Discussion

The population in Greenland is of Eskimo origin but has been strongly intermixed with Scandinavians and other Caucasians, especially in the west coast. The present population sample derives mainly from the west coast, with only 4 individuals from the east coast. Thus, the present study will not be representative of a pure Eskimo population because an intermixture of Caucasian alleles has to be expected. The distributions of allele frequencies (Figs. 1, 2) exhibit some similarities with Caucasian distributions but distinct and significant dissimilarities were seen for MS31, MS43a and YNH24. For MS31 and MS43a high frequencies of high molecular weight bands were observed (12 and 20 kb). For YNH24 a high frequency of 2 kb bands was seen. For MS1 the difference may be insignificant. The low variability of MS1 between racial groups has been observed by others (Buffery et al. 1991; Flint et al. 1989). For MS31 and YNH24 the rates of apparent homozygosity in the Greenland population sample were higher than in the Danish population sample. The most frequent fragment lengths of the homozygotes coincide with spikes in the distributions, i.e. in regions where the probability of occurrence of close heterozygotes or true homozygotes was high. The distributions of the frequencies (Figs. 3, 4) derived from 139 profiles from the Greenland population sample differed significantly, dependent on whether the reference sample was Greenland or Danish. On average the frequencies were 10 times higher with the Greenland reference sample. This finding illustrates the incorrectness of using an inappropriate database for the estimation of Greenland profile frequencies. On the other hand, for practical purposes it is of minor importance whether the maximum frequency is 10^{-6} or 10^{-7}as far as an isolated population of moderate size (ca. 55000) is concerned.

Acknowledgements. We are obliged to Dr. Mariann Thymann for collecting blood samples from Greenland paternity cases, and to Mrs Ulla Sørensen and Mrs Susanne Billesbølle for excellent technical assistance.

References

Buffery C, Burridge F, Greenhalgh M, Jones S, Willot G (1991) Allele frequency distributions of four variable number tandem repeat (VNTR) loci in the London area. Forensic Sci Int 52: 53–64

Eriksen B, Svensmark O (1993) Analysis of a Danish Caucasian population sample of single locus DNA-profiles: allele frequencies, frequencies of DNA-profiles and heterozygosity. Forensic Sci Int 59:119–129

Eriksen B, Bertelsen A, Svensmark O (1992) Statistical analysis of the measurement errors in the determination of fragment length in DNA-RFLP analysis. Forensic Sci Int 52:181–191

Flint J, Boyce AJ, Martinson JJ, Clegg JB (1989) Population bottlenecks in Polynesia revealed by minisatellites. Hum Genet 83:257–263

Thymann M, Eriksen B, Svensmark O (1993) DNA polymorphism in Greenland. Comparison of allele frequencies in population samples from Greenland and Denmark using the VNTR systems D1S80 (MCT118), D1S7 (MS1), D7S21 (MS31), D12S11 (MS43a) and D2S44 (YNH24). Proceedings 2. International Symp on the forensic aspects of DNA analysis, FBI Academy, Quantico, VI, (in press)