International Congress Series 1288 (2006) 400–402

www.ics-elsevier.com

SE33 allele and genotype frequencies in the population of Schleswig-Holstein (Northern Germany)

M. Krause [a,*], K.-G. Heide [a], M. Krawczak [b]

[a] *Laboratory for Paternity Testing, Kiel, Germany*
[b] *Institute for Medical Informatics and Statistics, Christian-Albrechts-University Kiel, Germany*

Abstract. Allele and genotype frequencies for STR marker SE33 were determined in a sample of 1614 unrelated Germans, ascertained through routine paternity cases. Many rare alleles were observed. A significant deviation from Hardy–Weinberg equilibrium was noted in the underlying population which was mainly attributable to the over-representation of genotypes otherwise expected to be rare. © 2005 Published by Elsevier B.V.

Keywords: SE33; Allele frequencies; Hardy–Weinberg equilibrium; Population genetics; Northern Germany

1. Introduction

STR locus SE33, a complex tetranucleotide, is one of the most informative and polymorphic genetic markers used in forensics. Since the practical application of a polymorphic marker requires the availability of a suitable profile database from the reference population, we carried out a detailed analysis of the allele frequencies and other parameters of SE33 in our laboratory.

2. Materials and methods

EDTA–blood samples of unrelated individuals were selected from routine paternity cases. DNA was extracted by the salting out procedure [1]. STR locus SE33 was amplified

* Corresponding author. Tel.: +49 431 388252; fax: +49 431 388221.
E-mail address: krause@labor-krause.de (M. Krause).

0531-5131/ © 2005 Published by Elsevier B.V.
doi:10.1016/j.ics.2005.09.122

Table 1
SE33 allele frequencies

SE 33								
Allele	Quantity	Frequency	Allele	Quantity	Frequency	Allele	Quantity	Frequency
8	0	0.0000	17.2	1	0.0003	27.2	25.4	0.0787
9	0	0.0000	17.3	3	0.0009	27.3	0	0.0000
10	0	0.0000	18	214	0.0663	28	2	0.0006
10.2	0	0.0000	18.2	1	0.0003	28.2	28.2	0.0874
11	5	0.0015	18.3	3	0.0009	29	4	0.0012
11.2	2	0.0006	19	230	0.0713	29.2	217	0.0672
12	19	0.0059	19.2	16	0.0050	30	0	0.0000
12.2	1	0.0003	20	191	0.0592	30.2	190	0.0589
13	22	0.0068	20.2	23	0.0071	31	3	0.0009
13.2	6	0.0019	21	98	0.0304	31.2	82	0.0254
13.3	0	0.0000	21.2	61	0.0189	32	2	0.0006
14	87	0.0270	22	35	0.0108	32.2	32	0.0099
14.2	6	0.0019	22.2	82	0.0254	33	17	0.0053
14.3	0	0.0000	23	3	0.0009	33.1	0	0.0000
15	111	0.0344	23.2	127	0.0393	33.2	16	0.0050
15.2	1	0.0003	24	0	0.0000	34	11	0.0034
15.3	7	0.0022	24.2	90	0.0279	34.2	5	0.0015
16	148	0.0458	25	3	0.0009	35	2	0.0006
16.1	3	0.0009	25.2	135	0.0418	35.2	2	0.0006
16.2	3	0.0009	26	1	0.0003	36	3	0.0009
16.3	2	0.0006	26.2	177	0.0548	37	2	0.0006
17	182	0.0564	27	2	0.0006	38	1	0.0003

in a monoplex polymerase chain reaction (PCR). Some of the PCR product were analysed on an ABI prism ™ 310 Genetic Analyzer (Applied Biosystems) according to manufacturer's recommendations.

Table 2
Analysis of Hardy–Weinberg equilibrium

Number of different alleles observed	56
Number of different genotypes observed	343
Number of unrelated individuals genotyped	1614
Observed number of homozygotes	68
Expected number of homozygotes under Hardy–Weinberg condition	84.1
χ^2 for the comparison between observed and expected number of homozygotes	3.260
p-value (χ^2 distribution with one degree of freedom)	0.07101
χ^2 (CHIGEN) for the comparison between observed and expected genotype frequency spectrum, i.e. taking all observed genotypes into account	5293.707
PVLGEN	0.00062
PVLMCMC	0.00000

PVLGEN: Since CHIGEN includes many genotypes with small expected frequencies under Hardy–Weinberg conditions, the corresponding p-value could not be determined by large sample approximation. It had to be determined by simulation instead. To this end, pairs of alleles were drawn from the observed set of alleles, randomly and without replacement, thereby generating a new, equally sized genotype set under Hardy–Weinberg conditions. This procedure was repeated 1 million times for each marker. The empirical p-value (PVLGE) equals the relative number of times the χ^2 for a simulated genotype set was larger than the (actually observed) CHIGEN.

PVLMCMC: Instead of performing a χ^2 test, comparing observed and expected genotype frequencies, compatibility with Hardy–Weinberg conditions can also be assessed by taking into account that the exact genotype distribution would be a hypergeometric one. This is analogous to using Fisher's exact test, instead of a χ^2 test, to analyse ordinary contingency tables. However because of the large number of possible genotypes, the p-value could not be determined by exhaustive permutation but had to be approximated by simulation. An implementation of the Monte-Carlo Markov Chain (MCMC) method described by Guo and Thompson [2] was chosen for this purpose.

Table 3
Overexpression of genotypes

Genotype	Observed	Expected	p-value
12–35.2	2	0.01	0.00007
12–16.2	3	0.02	<0.000001
15.3–23.3	5	0.28	<0.000001
16.1–30.2	3	0.18	0.00014

Allele frequencies were estimated from the genotype numbers observed in the sample (gene counting method). A possible deviation from Hardy–Weinberg equilibrium (HWE) was tested for statistical significance in three ways, namely by (i) a comparison of the observed and expected homozygote frequency using a χ^2 test with one degree of freedom, (ii) a χ^2-based permutation test taking the complete genotype distribution into account, and (iii) a Monte-Carlo Markov Chain (MCMC) exact test [2].

The paternity index (PI) and power of exclusion (PE) were calculated by Huston [3].

3. Results and discussion

SE33 allele frequencies as estimated from our North German sample (Table 1) were found to be comparable to those previously obtained in the German population [4]; no significant differences were observed to publicly available data from other parts of the country.

The sample of 1614 individuals comprised 343 different genotypes for SE33. These were analysed for Hardy–Weinberg equilibrium as outlined in Table 2.

After Bonferroni correction, four genotypes were found to be significantly over-represented in the database (Table 3).

The power of exclusion (PE) of SE33 was estimated to be 0.867, and a typical paternity index (PI) would be 7.69.

References

[1] D. Dykes, The use of biotinylated DNA probes in parentage testing: non-isotopic l labelling and non-toxic extraction, Electrophoresis 9 (1988) 359–368.
[2] S.W Guo, E.A. Thompson, Performing the exact test of Hardy–Weinberg proportion for multiple alleles, Biometrics 48 (1992) 361–372.
[3] K.A. Huston, Statistical Analysis of STR Data Profiles in DNA Promega, 1998, pp. 14–15.
[4] Huckenbeck/Scheil, The Distribution of the Human DNA-PCR Polymorphisms Supplement 98/99/00/01/02/03/04, http://www.uni-duesseldorf.de/Med.Fak/Serology.