Genotyping and comparing the different alleles present at three microsatellite loci

Lynda Gabriela Arostegui

Abstract

This experiment analyzed and compared three different microsatellite loci. First genomic DNA was isolated from cheek cell samples. Then, multiplex PCR was used to amplify the following three loci: D14S306, D15S655, and D15S657. After, the samples were electrophoresed on a polyacrylamide gel. A relationship was established between the distance of the bands of the ladder and the base pair each band represented, so that the bands of the samples could be designated with a fragment size. Each band represented a different allele. The D14S306 locus was found to have 10 different alleles, allowing for 55 possible genotypes. The D15S655 locus had 11 alleles, producing 66 possible genotypes. Lastly, the D15S657 locus had 9 alleles, which could manifest in 45 different genotypes. The frequencies of each allele were calculated. Even more, at the D15S657 locus, the genotypic frequencies were calculated and tested to see if they were in accordance with Hardy-Weinberg Equilibrium. The chi-square value, 48.4, corresponded with a p-value of 0.065. Thus, the null hypothesis was accepted and the data was found to be in compliance with HWE. This suggests that no evolution is occurring at this locus.

Introduction

The purpose of this experiment was to identify the different alleles present at three different microsatellite loci and compare them amongst the subjects tested. By the extraction of DNA from cheek cell samples, amplification via multiplex polymerase chain reaction, and gel electrophoresis, the genotype of microsatellite loci could be determined. Microsatellites are short tandem repeats in DNA sequences that make up about 3% of the genome (Carlini, 2018). Though mutations hardly affect the phenotype because microsatellite loci usually lie within non-coding regions, they are highly polymorphic. This is a result of the variability in the number of

copies of the tandem units in different alleles, which arise during DNA replication. As mutations arise and are not repaired, the number of alleles increases (Vieira et all, 2016). As the number of alleles increases, the likelihood of an individual to be heterozygous at a given locus also increases. Consequently, with a sufficient number, about 10 to 15, of microsatellite loci genotyped, a unique DNA profile can be established for an individual, except identical twins. For these reasons, microsatellite genotyping is a common practice in genetics research, but also practical in criminal forensics and paternity suits (Carlini, 2018).

Microsatellite genotyping is often completed in union with polymerase chain reaction (PCR) and gel electrophoresis. PCR is a method of amplification that is able to multiply a target sequence with the assistance of primers. The method exploits the ability of DNA polymerase to synthesize new strands, based on a template strand (Polymerase Chain Reaction 2014). Typically, more than one microsatellite locus is in question, and multiplex PCR is used, which allows for the use of multiple primers at once (Carlini, 2018). Because of this technique, a drop of blood could provide sufficient DNA to determine a genotype. Once the target sequence has been amplified, it can be electrophoresed to expose the different base pair fragments. Different alleles appear as different band sizes, as gel electrophoresis is a method that can separate DNA according to molecular size. In this experiment, the following three microsatellite loci were amplified: D14S306, D15S655, and D15S657. Found on chromosome 14, D14S306 is a microsatellite locus with an allele size range of 190-210 base pairs. On chromosome 15, D15S655 has an allele size range of 234-252 base pairs and D15S657 has a range of 336-360 base pairs (Carlini, 2018).

Materials & Methods

The first part of this experiment required the extraction and isolation of genomic DNA. Multiple subjects drank, but did not swallow, 10mL of a saline solution (0.9% NaCl). The solution was spit into separate paper cups after the contents mixed in their mouths for thirty seconds. Using a pipette, 1.5mL of the mixture was placed into a 1.5mL microcentrifuge tube. All samples were centrifuged for five minutes at maximum speed. The majority of the supernatant was removed, leaving behind just enough to resuspend the cells. Next, these cells were transferred to a 0.5mL microcentrifuge tube which also contained 100µL of Chelex®. The tube was then placed on a vortex to thoroughly mix the contents. The samples were then heated using a thermal cycler; it was set at 100°C for ten minutes. The sample was then placed in a microcentrifuge again but only spun for one minute. Then, 30µL of the supernatant was pipetted into a new 1.5mL microcentrifuge tube.

For the multiplex PCR reaction setup, a 0.5x and a 0.25x dilutions of the genomic DNA were prepared. To make the 0.5x dilution, 5µL of the isolated genomic DNA was pipetted into a 0.5µL microcentrifuge tube, which was previously loaded with 5µL of ddH_2O. The contents were then mixed and the 0.25x dilution was prepared. From the 0.5x diluted genomic DNA, 5µL were extracted and placed in another 0.5µL microcentrifuge tube, which had also been supplemented with 5µl of ddH_2O. Next, two strip tubes were used; each was filled with 3.4µl of ddH2O, 0.6µl of 10X PCR buffer, 0.5µl 2.5mM of dNTPs, 0.2µl of microsatellite primer mix, 0.2µl of IRD-700 fluorescently labeled primer tag, and 0.1µl of Taq polymerase. Next, the strip tubes were placed in a thermal cycler and programmed for the following profile. The initial denature was 5 minutes at 95°C. The 5 cycles ensued 45 seconds at 95°C, 5 minutes at 68°C, and 1 minute at 72°C. After each cycle, the anneal temperature was decreased 2°C. The following five cycles were 45 seconds at 95°C, 2 minutes at 58°C, and 1 minute at 72°C . Again, after each

cycle the anneal temperature was lowered by 2°C. The next step occurred for 25 cycles of 45 seconds at 95°C, 2 minutes at 50°C, and 1 minute at 72°C. The final polymerization was 5 minutes at 72°C and then put on hold at 4°C.

Next, the samples were loaded onto an acrylamide gel for electrophoresis, along with a ladder. Once the image on the gel was produced, the genotypic data was acquired. Though done electronically, the band sizes were interpreted through the establishment of a relationship between the distance of each band on the ladder and the fragment length it represented on the gel. Each different fragment size represented a different allele at each locus. After all the data was collected, several different analyses were completed. First, the calculation of the allele frequencies was determined by dividing the number of times the given fragment size appeared in the sample by the total number of alleles. Next, the number of possible genotypes was defined by the equation, $\sum_{i=1}^{i=n} i$, where n is the number of alleles.

The alleles at the D15S657 locus were then further analyzed, to see if the data was in accordance with Hardy-Weinberg Equilibrium. This was done by completing a chi-square test. Because nine alleles were present at this locus, the following modified HWE equation was used: $p^2+q^2+r^2+s^2+t^2+u^2+v^2+w^2+x^2+2pq+2pr+2ps+2pt+2pu+2pv+2pw+2px+2qr+2qs+2qt+2qu+2qv+2qw+2qx+2rs+2rt+2ru+2rv+2rw+2rx+2st+2su+2sv+2sw+2sx+2tu+2tv+2tw+2tx+2uv+2uw+2ux+2vw+2vx+2wx=1$. The expected frequencies were calculated using the previously found allele frequencies, and then multiplied by the sample size to find the expected number of subjects. The following equation was used to determine the chi-square value: $X^2=\Sigma(o-e)^2/e$. The degrees of freedom was equal to number of genotypes minus one minus the number of independent allele frequencies estimated from the data.

Results

Locus: D14S306	
Allele	**Frequency**
198	0.0125
202	0.0375
206	0.0375
210	0.3125
222	0.1375
214	0.15
218	0.1875
226	0.1
234	0.0125
238	0.0125

Locus: D15S655	
Allele	**Frequency**
257	0.2875
248	0.025
254	0.2375
251	0.05
266	0.075
269	0.1375
260	0.0875
272	0.05
275	0.0125
278	0.025
263	0.0125

Locus: D15S657	
Allele	**Frequency**
344	0.024
352	0.2125
348	0.0375
356	0.225
368	0.1375
360	0.1625
364	0.15
372	0.0375
376	0.0125

Table 1. Frequencies for each allele.

Variable	**Genotype**	**Expected Frequency**	**Expected Number of Subjects**	**Observed Number of Subjects**	**(O-E)^2/E**
p^2	344/344	0.000625	0.025	0	0.025
q^2	348/348	0.00141	0.0564	0	0.0564
r^2	352/352	0.0454	1.816	1	0.366660793
s^2	356/356	0.0506	2.024	1	0.5180711462
t^2	360/360	0.0264	1.056	0	1.056
u^2	364/364	0.0225	0.9	1	0.0111111111
v^2	368/368	0.0189	0.756	0	0.756
w^2	372/372	0.0014	0.056	0	0.056
x^2	376/376	0.00015	0.006	0	0.006
2pq	344/348	0.00188	0.0752	0	0.0752
2pr	344/352	0.011	0.44	2	5.530909091
2ps	344/356	0.011	0.44	0	0.44
2pt	344/360	0.008	0.32	0	0.32
2pu	344/364	0.0075	0.3	0	0.3
2pv	344/368	0.0069	0.276	0	0.276
2pw	344/372	0.00187	0.0748	0	0.0748
2px	344/376	0.00125	0.05	0	0.05
2qr	348/352	0.015975	0.639	1	0.2039452269
2qs	348/356	0.016875	0.675	1	0.1564814815
2qt	348/360	0.0121875	0.4875	1	0.5387820513
2qu	348/364	0.01125	0.45	0	0.45
2qv	348/368	0.0103125	0.4125	0	0.4125
2qw	348/372	0.015975	0.639	0	0.639

2qx	348/376	0.0009375	0.0375	0	0.0375
2rs	352/356	0.09585	3.834	5	0.3546051122
2rt	352/360	0.069225	2.769	2	0.2135648248
2ru	352/364	0.0639	2.556	2	0.1209452269
2rv	352/368	0.058575	2.343	3	0.1842291933
2rw	352/372	0.015975	0.639	0	0.639
2rx	352/376	0.005325	0.213	0	0.213
2st	356/360	0.073125	2.925	2	0.2925213675
2su	356/364	0.0675	2.7	4	0.6259259259
2sv	356/368	0.061875	2.475	4	0.9396464646
2sw	356/372	0.016875	0.675	0	0.675
2sx	356/376	0.005625	0.225	0	0.225
2tu	360/364	0.04875	1.95	4	2.155128205
2tv	360/368	0.0446875	1.7875	3	0.822465035
2tw	360/372	0.0121875	0.4875	1	0.5387820513
2tx	360/376	0.0040625	0.1625	0	0.1625
2uv	364/368	0.04125	1.65	0	1.65
2uw	364/372	0.01125	0.45	0	0.45
2ux	364/376	0.00375	0.15	0	0.15
2vw	368/372	0.0103125	0.4125	1	0.8367424242
2vx	368/376	0.003	0.12	0	0.12
2wx	372/376	0.0009375	0.0375	1	24.70416667
				Chi Square Value	48.4295834
				Degrees of Freedom	35
				P-value	0.065

Table 2. Testing for Hardy-Weinberg Equilibrium at the D15S657 locus

Discussion

This experiment analyzed three different microsatellite loci. At D14S306, ten alleles were observed which correlates with a possibility of 55 genotypes. In the sample size, 18 genotypes were observed. The most common allele was 210, with a frequency of and 206, both of which had a frequency of 0.3125. The least common alleles were 234, 238, and 198, each with a frequency of 0.0125. The microsatellite locus D15S655 had the presence of eleven different alleles. Out of a possible 66 genotypes, 20 were observed. The most common allele was 257,

with a frequency of 0.2875, and the least common alleles were 275 and 263, tied with a frequency of 0.0125. Finally, at the D15S657 locus, nine alleles were detected and 18 out of 45 possible genotypes were observed. At this locus, the most common allele was 356, which had a frequency of 0.225 and the least common allele was 376, which had a frequency of 0.0125.

If considering all three loci, there are a possible 163,350 possible genotypes. The observed data speaks to the high polymorphic nature of microsatellite loci, and supports how a unique DNA profile could be created for an individual given a variety of 10-15 loci. Within the sample of this experiment, subjects shared the same genotype at one locus and less shared it at two, but no two students shared the same genotype at all three loci. Thus, if an unknown sample of DNA was obtained, it would be possible to unequivocally identify which subject it came from. However, this might not be the case if tested amongst a larger subject pool. This is why, in criminal forensics, more loci are used to increase the number of possible genotypes and eliminate the probability of two or more subjects matching, since the genotype is being compared to that of the seven billion people in this world.

Moreover, at the D15S657 locus, the genotypic frequencies observed were in agreement with the expectations of Hardy-Weinberg Equilibrium. The chi-square value, 48.4, calculated corresponded with a p-value of 0.065, which is above the accepted threshold of 0.05. Therefore the data was not significant and the null hypothesis was accepted. Because the data was found to be in accordance with HWE, the experiment indicates that no evolution is occurring. This is fitting because microsatellite loci are part of non-coding regions of DNA, which means certain alleles wouldn't be selected over others for reasons of "fitness."

References

Carlini, David. PCR Amplification of Microsatellite Loci. American University, 2018. Print.

Polymerase Chain Reaction (PCR). (2014, September 26). Retrieved April 1, 2018, from https://www.ncbi.nlm.nih.gov/probe/docs/techpcr/

Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. de F. (2016). Microsatellite markers: what they mean and why they are so useful. *Genetics and Molecular Biology*, *39*(3), 312–328. http://doi.org/10.1590/1678-4685-GMB-2016-0027