CHARACTERIZATION OF THE HUMAN STR LOCUS D18S535 IN CAUCASIAN AND ABORIGINAL POPULATIONS FROM NORTHERN ONTARIO FOR FORENSIC PURPOSES

by
Krista Ann Currie

A thesis submitted in partial fulfilment
of the requirements for the degree of
Master of Science in Biology

School of Graduate Studies
Laurentian University
Sudbury, Ontario

Your file Votre référence
ISBN: 0-612-85627-5
Our file Notre référence
ISBN: 0-612-85627-5

Abstract

Short tandem repeat (STR) loci display significant variation within human populations and are useful markers in the identification of individuals for forensic purposes. D18S535 is a simple hypervariable STR consisting of blocks of GATA repeats and was recently proposed for application into routine forensic casework. Genetic characterization of one Caucasian and three Aboriginal populations from Northern Ontario was undertaken in order to assess the utility of the D18S535 as a genetic marker for forensic DNA typing in Northern Ontario. The D18S535 was amplified by monoplex polymerase chain reaction (PCR), separated by denaturing polyacrylamide gel electrophoresis (PAGE), and visualized using the silver stain detection method. Population data generated showed that the D18S535 was highly polymorphic with a heterozygosity of ≥ 0.75. Based on the exact test, observed genotypic proportions at the D18S535 locus were not consistent with expected ones under Hardy-Weinberg Equilibrium (HWE) in two of the Northern Ontario Aboriginal populations studied. Statistical analyses were carried out to compare the four Northern Ontario populations with each other and with previously published D18S535 studies from Europe and South America, and to calculate the power of discrimination (PD) and chance of exclusion (CE). Forensic parameters indicated that the D18S535 locus in the four Northern Ontario populations was highly discriminating (PD ≥ 0.85, CE ≥ 0.51); however, the lack of HWE in some of the populations must be taken into account in the application of these results into Northern Ontario forensic casework.

Acknowledgements

I wish to extend my gratitude to Dr. Kabwe Nkongolo for supervising my research and for his knowledge and kindness that was extended toward me. I wish to thank Mr. Paul Michael for his patience, technical contributions, and guidance with this project. I would like to thank my committee members, Dr. Gilbert Coté, Director of the Genetics Department at the Sudbury Regional Hospital, and Dr. Annyck Ratiarson of Laurentian University for their valuable suggestions regarding this study.

I am grateful to Dr. George Carmody of Carleton University in Ottawa, Dr. Bruce Weir of North Carolina State University in North Carolina, Dr. Richard A. Nichols of Queen Mary and Westfield College in London, and Dr. Michael Persinger of Laurentian University for their many valuable discussions and statistical assistance. I would like to express my thanks to Dr. Garry Ferroni, Dr. Leo Leduc, and Dr. Eric Gauthier for use of their laboratory equipment.

Further thanks are due to Beatriz Quintáns Castro at the Institute of Legal Medicine, Santiago de Compostela, Galacia, Spain for providing the sequenced D18S535 allelic ladder used in this study. Finally, I would like to acknowledge the hospital laboratory staff at the Espanola General Hospital in Espanola, Sioux Lookout Zone Hospital in Sioux Lookout, and the Wikwemikong Health Centre in Wikwemikong for graciously donating blood samples.

Table of Contents

Page

List of Figures

List of Tables

Table **Page**

1. Introduction

1.1 Background Information on Genetic Markers for Forensic Purposes

The genetic uniqueness of individuals, with the exception of identical siblings, is defined by the combination of genetic markers that an individual inherits from his or her parents. These genetic markers can be analysed at the level of DNA sequence variation. Genetic markers are located at a particular physical position on the chromosome called a locus, but they exist in alternate forms, designated alleles. Alleles differ in their DNA sequence (Rudin and Inman, 2002).

The purpose of forensic genetic marker typing is to determine which alleles are present at genetically variable loci. If an evidentiary DNA profile does not match the reference sample of the individual in question, then the individual can be excluded with absolute certainty. If the DNA profiles match, it does not constitute an identification because some proportion of the general population may also share the same genetic marker combination (Committee on DNA Forensic Science, 1996).

Genetic markers are judged by their ability to distinguish between individuals. If the genotype (a pair of alleles) frequencies of the marker within a defined population are known, then the probability that two individuals chosen at random from this population would have the same genotypic frequencies could be calculated. The evidence of a matching profile is presented as a likelihood ratio (LR), defined as the probability of the evidence of a match between suspect and offender if those people are the same, divided by the probability if they are two unrelated people. Under the assumption that unrelated people have

independent genotypic frequencies, the LR reduces to the reciprocal of the profile frequency.

Hardy-Weinberg Equilibrium (HWE) is used to determine the stability of alleles at a particular locus in a population. If the population is in HWE for a particular marker, then the expected genotype frequencies can be calculated from the observed allele frequencies and allele frequencies can be multiplied across loci (Committee on DNA Forensic Science, 1996). Being able to multiply across loci is important because the use of more loci increases the ability to distinguish between individuals. Absence of HWE invalidates multiplication across and within loci (Devlin *et al.*, 1990).

The index value known as the power of discrimination (PD), is the probability that two random individuals will have different genetic types for the marker being tested. Higher PD values indicate a greater individualization potential because a larger portion of the population can be excluded. For a single marker system, the greater the number of alleles (degree of polymorphism) and the more even their distribution throughout a population (higher heterozygosity), the higher the power of discrimination (George Carmody 2002, personal communication; Rudin and Inman, 2002). The genetic diversity of tandem repeats provides this great individualization potential for DNA profiling used in forensics and comes from the fact that they are non-coding regions of DNA. Since these regions do not code for proteins, variations are functionally insignificant and hence well-tolerated during evolution (Committee on DNA Forensic Science, 1996).

1.2 Tandem Repeat Loci

As previously stated, tandem repeat loci display considerable variation within human

populations and are useful markers in the identification of individuals for forensic purposes (Gill *et al.*, 1985; Lygo *et al.*, 1994). Tandem repeats are tandemly reiterated sequences which are found in large numbers throughout the human genome and are usually found approximately every six to ten kilobases (Wong *et al.*, 1987; Weber and May, 1989; Beckman and Weber, 1992). Tandem repeat loci are polymorphic, with alleles differing in the number of repeat units (Craig *et al.*, 1988) and some differing in base pair sequence (Urquhart *et al.*, 1993; Möller *et al.*, 1994). A theoretical model which has been proposed for the production of different alleles is the stepwise mutation model (SMM). The SMM assumes that the entire sequence of allelic states can be expressed by integers (..., A_{-1}, A_0, A_1,...) and that if an allele changes by mutation, it moves either one step in the positive direction or one step in the negative direction with equal probability (Kimura and Ohta, 1975; Kimura and Ohta, 1978). Although the exact mechanism is currently unknown, the difference in repeats are thought to arise during meiosis from unequal recombination mechanisms between sister chromatids (Jeffreys *et al.*, 1985; Wolff *et al.*, 1989) or arise from replication slippage during meiosis when the DNA is replicated. The DNA polymerase enzyme can slip forward or backwards on the repeat units, deleting or adding repeat units to the daughter strand. This means that daughter strands may have slightly fewer or slightly more repeat units than the parent strand (Tautz and Renz, 1984; Tautz *et al.*, 1986; Schlötterer and Tautz, 1992).

There are two classes of tandem repeat loci distinguished by the number of repeat units of their core sequences: short tandem repeat (STR) loci and variable number of tandem repeat (VNTR) loci. STRs are also known as microsatellites and are composed of tandemly

repeated sequences of two to eight base pairs in length. VNTRs are also known as minisatellites and contain longer repeat units of nine to eighty base pairs (Litt and Luty, 1989). Based on their variable sequence, STR loci can be separated into three categories: simple repeats, compound repeats, and complex repeats. Simple repeats consist of three, four, or five base pairs repeated a variable number of times. Compound repeats consist of two or more adjacent simple repeats, and complex repeats consist of numerous repeat blocks of variable unit length along with a more or less variable intervening sequence (Urquhart *et al.*, 1994).

Tandem repeat loci were originally characterized by restriction enzyme digestion of genomic DNA. This was followed by agarose gel electrophoresis, Southern blotting, and hybridization with labelled minisatellite probes (Wyman and White, 1980). Polymerase chain reaction amplification (PCR) (Mullis and Faloona, 1987) of tandem repeat loci has greatly improved the sensitivity of the technique, enabling as little as 1 ng of genomic DNA to be typed (Jeffreys *et al.*, 1988). PCR is a process by which a specific region of DNA is targeted and replicated through a number of cyclic steps at different temperatures. Primers anneal to the unique conserved regions of the tandem repeat locus, which flank the variable regions, and specifically amplify a single tandem repeat locus. The PCR products can be analysed using a variety of different formats. Tandem repeats can be separated by native or denaturing polyacrylamide gel electrophoresis (PAGE) and visualized using both the silver stain detection method and a gel imaging system capable of detecting fluorescent tags if the PCR primers are fluorescently labelled (Frégeau and Fourney, 1993; Robertson, 1994; Lins *et al.*, 1996; Decorte and Cassiman, 1996; Micka *et al.*, 1999). Tandem repeats can also be

analysed by capillary electrophoresis (McCord *et al.*, 1993; Butler et al., 1994; Buel *et al.*, 1998; Wenz et al., 1998), time-of-flight matrix-assisted laser desorption/ionization (Tof MALDI) mass spectrometry (Braun *et al.*, 1997; Monforte and Becker, 1997), and microchip (Schmalzing *et al.*, 1997).

Restriction fragment length polymorphism (RFLP) analysis was the first DNA-based method applied to individual identification (Jeffreys *et al.*, 1985). This method is powerful in its ability to differentiate individuals but is limited by the quantity and quality of DNA required for clear and accurate results. It is also limited by the amount of time it takes to obtain a result. The PCR DNA amplification technology is ideally suited for the analysis of forensic DNA samples in that it is sensitive, rapid, and not as limited by the quality of DNA as the RFLP method.

1.3 Forensic Application of Short Tandem Repeat Loci

Given that individuals can be distinguished by their genetic marker profiles, the more markers compared, the more differences between individuals would be revealed. In other words, the probability of a match decreases as the number of loci increases or with more loci, the probability of discrimination increases (Jones, 1972). The Committee on DNA Forensic Science (1996) recommended research into the identification and validation of more and better marker systems for the analysis of forensic materials in order to make each profile unique; however, not all genetic markers are suitable for forensic applications. Genetic markers must be carefully selected with the intention of enhancing the discrimination power and broadening the assortment of evidence samples from which information can be acquired.

Currently, STRs are utilized as the marker of choice for forensic DNA testing. The criteria by which STR genetic markers are evaluated for forensic use are small allele sizes, a high heterozygosity, distinguishable alleles, low stutter, robust amplification, easy multiplexing, and a low mutation rate (Gill *et al.*, 1996; Lareu *et al.*, 1998; Carracedo and Lareu,1998; Wiegand *et al.*, 1999; Barral *et al.*, 2000).

Short tandem repeat loci with small allele sizes (< 300 base pairs) are important for forensic purposes because STRs with small allele sizes make old or degraded samples more amenable to analysis; therefore, less chance of allelic dropout. Allelic dropout is seen when the two alleles of a heterozygote vary by a large number of repeats. The smaller of the two alleles is preferentially amplified by PCR (Batanian *et al.*, 1990; Gécz, 1991; Lygo *et al.*, 1994) with an apparent allele loss in the heterozygote (two different alleles at a locus), having an individual appear as a homozygote (the same allele at a locus).

Heterozygosity is the proportion of a population that is heterozygous for a particular locus. Heterozygosity is a measure of the genetic variability of a population and a short tandem repeat locus should have > 70 % heterozygosity (Urquhart *et al.*, 1995 b; Gill *et al.*, 1996; Wiegand *et al.*, 1999) which would indicate a high degree of variability in the population. Having a high degree of variability in a population is important in forensics because it allows individuals to be discriminated from one another. Urquhart *et al.* (1995 b) indicated that the discrimination power of a locus should be > 80 %.

Tetranucleotide and pentanucleotide repeats are preferred for forensic casework since allele determination is easier as alleles often differ by a full repeat length and also for their low stutter characteristics (Holgersson *et al.*, 1994; Gill *et al.*, 1996, Pérez-Lezaun *et al.*,

1997). Stutter is observed on polyacrylamide gels as a fainter band which is one repeat unit less than the actual allelic band. Stutter is an artifact of PCR which is produced by strand slippage along the repeat region during PCR amplification (Hauge and Litt, 1993). Mixtures of fluids are often encountered in forensic analysis and stutter bands can make interpretation difficult.

Short tandem repeat loci are more useful in forensic analysis when they are robust (Gill *et al.*, 1996). This allows forensic samples with a varying degree of quality and quantity to be tested with the same STRs. Validation studies must be carried out on non-human samples, mixtures, and degraded samples in order to ensure that new STRs are robust.

Forensic STRs should have the ability to be incorporated into a multiplex PCR to increase the speed of the process of analysis and to limit consumption of a DNA sample. Multiplex PCR is when more than one primer set is added to a single amplification reaction. This allows more STR loci to be typed in a minimal amount of time and also allows more information to be obtained from a sample with a limited quantity (Edwards *et al.*, 1991; Edwards *et al.*, 1992; Kimpton *et al.*, 1994; Lins *et al.*, 1996).

Low mutation rate loci are important in forensic analysis because they tend to have large differences in allele frequency distributions from population to population and they may be thought of as having greater discriminatory power between individuals than fast mutating loci because populations and subpopulations will have frequency distributions which cannot be distinguished (Wall *et al.*, 1993; Wiegand *et al.* 2000).

1.4 The D18S535 Locus

D18S535 is a new STR locus which was first identified by Lareu *et al.* (1998) after a survey of the genome database, GenBank. The D18S535 locus was found in the human sequence tagged site (STS) (GenBank accession number G07985) by Murray *et al.* (1995), referred to as human STS CHLC.GATA13.p9341. The D18S535 is located on human chromosome 18 (q arm) and has been reported to be rather robust for forensic analyses. It also displays several of the highly desirable characteristics required for new STRs to be useful for forensic purposes (Lareu *et al.*, 1998). According to the guidelines published by Urquhart *et al.* (1994), the D18S535 is a simple hypervariable STR which consists of blocks of (GATA) repeats with a basic sequence structure $(GATA)_1$ $(GACA)_1$ $(GATA)_1$ $(GAT)_1$ $(GATA)_{9\text{-}16}$ (Lareu *et al.*, 1998).

Data from previously published D18S535 studies indicated that the D18S535 locus was a highly polymorphic STR with a heterozygosity > 0.75. This STR was also shown to have eight common alleles and one rare off ladder variant (see Appendix A for sequence structure and characteristics of the common alleles of the D18S535 locus). The D18S535 was also shown display very low stutter characteristics and all populations studied to date using D18S535 have been shown to be in HWE (Lareu *et al.*, 1998; Wiegand *et al.*, 1999; Nievas *et al.*, 1999; Samura *et al.*, 2000; Martínez-Jarreta *et al.*, 2000; Tomàs *et al.*, 2001; Grubić *et al.*, 2002; Albeza *et al.*, 2002).

1.5 Geography, Demography, and History of Northern Ontario

For the use of genetic markers in forensic analysis, it is necessary to collect

allele/genotype frequency data from relevant populations so that an estimate of the frequency of a profile can be provided. Current forensic practice is to estimate allele frequencies for broadly defined ethnic groups (e.g., Caucasian) but not for any sub-populations generated by, for example, geographic, religious, nor social divisions within these groups (Nichols and Balding, 1991); however, genotype frequencies can vary for a particular marker and can vary from population to population. For this reason, it is important to collect allele frequency data from relevant populations in various geographic locations (Lewontin and Hartl, 1991).

The D18S535 polymorphism has previously been studied in a few European and South American populations, but nothing is known about the genetic variation in the Northern Ontario region of Canada. Canada is a multi-ethnic and multi-cultural country which has a total land area of 9, 093, 507 square km with 31, 414, 000 people (Statistics Canada, 2001 b). The Province of Ontario is 907, 655.59 square km and consists of a population of 11, 697, 569 people (Statistics Canada, 2001 b). Ontario is centrally located in Canada and is bordered by the Great Lakes on the south, Hudson Bay on the north, Manitoba on the west, and Quebec on the east. Ontario is Canada's wealthiest, most industrialized, and most populated province (Brock, 1978). Ontario is consistent with the rest of Canada in regards to its population being ethnically and culturally diverse; however, Northern Ontario has been recognized as being geographically and culturally different from the rest of Ontario. Since Northern Ontario is geographically and culturally different from Southern Ontario, the people who inhabit Northern Ontario are also significantly different from people who inhabit Southern Ontario (Braithwaite, 1974).

The lines that divide Northern and Southern Ontario are not well defined; however,

the provincial government has defined Northern Ontario as the area of land from the Manitoba border to the Quebec border and extending south to the District of Parry Sound (Brock, 1978). Northern Ontario occupies approximately 75 % of Ontario's land mass but is populated by only 10 % of its people (Brock, 1978; White, 1985). The populations of Northern Ontario are more rural and isolated, and are characterized by low population density and population dispersion (McVey, 1979) (see Appendix B for a map indicating the locations of the eleven federal electoral districts of Northern Ontario; Appendix C for the demographic characteristics of the Northern Ontario federal electoral districts; Appendix D for the demographic characteristics of Northern Ontario, Southern Ontario, Ontario, and Canada according to Statistics Canada, 1996).

The genetic composition, which is the distribution of allele sizes, of human populations varies because of their differing evolutionary histories, patterns of dispersal, interbreeding, and the process of mutation, where alleles can change in size by one repeat unit in populations (Di Renzo *et al.*, 1994; Balding and Nichols, 1995). The history of Northern Ontario was greatly influenced by its geography, geology, and climate which influenced what peoples have inhabited it and where they have settled. Northern Ontario is heavily forested with countless lakes and rivers (Turnbull, 1966), and it is comprised of the immense Precambrian rock known as the Canadian Shield.

The ancestors of the North American Aboriginal peoples are thought to have entered Northern Ontario about 9, 000 years ago, settling along the shores of the Upper Great Lakes. By the 17th century the Aboriginal population had developed a temporary seasonal lifestyle which focussed on hunting and trading. They spoke the Algonkian language but were

divided into a number of dialect groups which included the Ojibwa (or Chippewa), the Cree, the Algonquin, the Ottawa, the Nipissing, and the Mississauga. The different Northern Ontario dialect groups were traditionally bands (aggregates of family hunting groups) related through marriage, clan ties, language, and way of life. They were also associated together through geographic location and common survival strategies (Dawson, 1977; Bray and Epp, 1984; White, 1985).

In the early part of the 17th century Northern Ontario encountered representatives of two European countries, England's Henry Hudson on Hudson Bay and France's Samuel de Champlain on Georgian Bay. It is estimated that during this time, the entire Aboriginal population of Canada was around 220,000 persons scattered in more than 500 tribes with 40,000 to 50,000 living in Ontario. It was not until the mid-1840's that Canadians started to settle in Northern Ontario after the Canadian government negotiated with the Ojibwa for title to their land following the awareness of the economic potential of the north from lumber and minerals. An increased presence in Northern Ontario was brought about by the construction of the railways in the 1850's and 1860's (Turnbull, 1966; Braithwaite, 1974; McVey, 1979; Bray and Epp, 1984).

The north attracted a culturally diverse population. In 1871 the population was almost entirely composed of the "founding peoples" of Canada: Aboriginals (50.3 %), British (32.4 %), and French (15.8 %). The population figures changed by 1911 so that there were Aboriginals (5.3 %), British (50.7 %), French (21.2 %), and 23 % from other countries with the foremost being Finns and Ukrainians. There were also Italians, Germans, Scandinavians, Poles, Dutch, Belgians, Swiss, Greeks, Chinese, and Japanese but they contributed only a

small portion of the total population. The different ethnic groups appeared to gravitate to particular communities and to particular types of work with the exception of British Canadians who were found in large numbers in all areas of work throughout the north (Schull, 1978; Bray and Epp, 1984).

Aboriginals were a greater visible presence in Northern Ontario than anywhere else in the province and this remains true today (White, 1985; Statistics Canada, 1996). In 2000, the total Aboriginal population in Canada was 675,499 with Ontario having the largest proportion of all regions with 153,946 (23 %) peoples. Of the 153,946 Aboriginals in Ontario, 78,346 (50.9 %) reside on a reserve which is a recognized Aboriginal settlement on Crown land. Of the Aboriginals which reside on a reserve, 36,376 (46.4 %) reside on a reserve classified as urban (one which is located within 50 km from the nearest service centre having year-round road access), 20,962 (26.8 %) reside on a reserve classified as rural (one which is located between 50 km and 350 km from the nearest service centre having year-round road access), 88 (0.1 %) reside on a reserve classified as remote (one which is located over 350 km from the nearest service centre having year round road access), and 20,915 (26.7 %) reside on a reserve classified as special access (one which has no year-round road access to the nearest service centre) (Department of Indian Affairs and Northern Development, 2002).

The Ojibwa, who are members of the Algonkian language family, make up most of the Aboriginal population of Ontario with their greatest concentration on the northern rims of Lake Huron and Lake Superior in Northern Ontario. Today, the term Ojibwa is a collective term which includes at least four tribal divisions which were originally territorially

based. The Ojibwa, Mississauga, Ottawa, and Potawatami are included under the term Ojibwa because of the linguistic and socio-cultural similarities among these people. Ojibwa bands merge with and are eventually replaced by Cree in the north (Szathmary *et al.*, 1978).

Northern Ontario has a very different urban history from that to the rest of Canada. Most cities and towns are not much more than a century old because the Canadian Shield was a huge physical barrier. Urbanization occurred at a later period when the railways were built. Northern Ontario today does not have a large metropolis, unlike Southern Ontario; therefore, small and medium-sized centres have a much more significant importance in Northern Ontario (Turnbull, 1966; Bray and Epp, 1984). Even in modern times there are still large parts of Northern Ontario which are sparsely populated and only slightly impacted by modern society (White, 1985).

1.6 Ethnicity vs. Race

When discussing populations, the idea of ethnicity and race needs to be addressed; however, the concepts are very complex issues. Ethnicity and race are far from being synonymous; nevertheless, the two terms are commonly used interchangeably (Caldwell and Popenoe, 1995). Ethnicity is a concept that refers to shared origins, geographical residence, social background, language, religious traditions, culture, and traditions that are distinctive and maintained between generations (Senior and Bhopal, 1994; Bagley, 1995). Race refers to a breeding population characterized by frequencies with which certain hereditary traits differ from those of other populations of the same species. Among these traits are features of external appearance which make it possible to visually identify members of different

populations (Brues, 1977; Goldsby, 1977).

Although the concept of ethnicity is preferred when referring to individuals and population groups, the word race is commonly used despite the fact that most biologists reject the traditional Western concept of race as identifiable biological groups. It is often thought of as a meaningless culturally generated identifier since most human genetic variation is clinal in nature. Clines are geographical gradients where the frequency of a gene or morphological character changes gradually throughout its distribution due to selective factors or to interbreeding of populations, which makes defining racial boundaries very difficult (Watts, 1997).

Despite the arguments against the concept of race, it remains a focus of forensic science research and application since individuals are often assigned to particular socially constructed racial categories based on features they display which may point to a particular ancestry (Sauer, 1992; Brace, 1995; Lewontin, 1997). It is recognized that most populations are mixed and that definitions can be somewhat arbitrary. Individuals often place themselves in categories which are based on geographical locations, physical appearance, and linguistics.

Among the Northern Ontario inhabitants, Aboriginals and Caucasians represent two of the most relevant population groups. The literature suggests that Aboriginal is a term which is inclusive of First Nations, Métis and/or Inuit peoples (Department of Indian Affairs and Northern Development, 2002) and Caucasian is a broad category of individuals traditionally of European, North African, and Middle Eastern descent (Goldsby, 1977; Lewontin, 1997).

1.7 Study Objectives

The aims of this investigation were firstly to obtain allele frequencies for D18S535 in Northern Ontario Caucasian and Aboriginal population samples using PCR. Secondly, to compare results between each of the Northern Ontario populations and with previously published populations from Europe and South America using multiple analysis of variance, pairwise comparisons, and cluster analysis. This was done in order to determine if populations differed from each other; therefore, determine if population data could be amalgamated. Thirdly, to test whether or not the genotype frequencies conformed to Hardy-Weinberg Equilibrium expectations in order to determine if the expected genotype frequencies could be calculated from the observed allele frequencies. Also, if genotype frequencies conformed to Hardy-Weinberg expectations, allele frequencies could be multiplied across loci which would allow for greater discrimination between individuals. Finally, to calculate the heterozygosity, the power of discrimination, and the chance of exclusion to ascertain the utility of the D18S535 as a genetic marker for Northern Ontario forensic casework.

2. Materials and Methods

2.1 Biological Samples

Discarded whole blood samples from anonymous individuals were donated from three hospital laboratories in the Northern Ontario region of Canada: Espanola General Hospital in Espanola, Sioux Lookout Zone Hospital in Sioux Lookout, and Wikwemikong Health Centre in Wikwemikong Unceded Indian Reservation located on Manitoulin Island. The geographical location of each of the sampling locations is displayed in Figure 1. The study was approved by the ethics committee at each facility.

Blood samples were collected in ethylene diamine tetraacetic acid (EDTA) *Vacutainer*® tubes by venipuncture from Caucasian individuals in Espanola and from Aboriginal individuals in Espanola, Sioux Lookout, and Wikwemikong. A total of one hundred samples from each population were collected over a three month period. All tubes were received labelled as either Caucasian or Aboriginal. Hospital laboratory personnel designated donors as Caucasian or Aboriginal visually, and on the basis of common first names and surnames. Determining ethnic origin by appearance and the use of surnames in sample selection has been proven to be an effective means of designating individuals into ethnic categories (Wijsman *et al.*, 1984; Piazza *et al.*, 1987; Guglielmino and Silvestri, 1995; Buroker *et al.*, 1997; Meng *et al.*, 1999). The samples were transported on ice to Laurentian University in Sudbury, Ontario. They were aliquoted into1.5 ml microcentrifuge tubes and held at 4 °C until further processing.

Figure 1. A map of Ontario indicating the locations of the Caucasian and Aboriginal populations analysed in this study (Caucasian samples from Espanola, and Aboriginal samples from Espanola, Sioux Lookout, and Wikwemikong)

2.2 Description of Sampling Sites

Espanola is located 70 km west of Sudbury and 455 km northwest of Toronto just south of the Trans-Canada highway. It is representative of most of the industrial towns of Northern Ontario because it is a single resource community. A pulp mill was established in 1905 and the town and the mill flourished until The Depression which turned Espanola into a ghost town. During World War II the mill was used to house German prisoners of war and was reopened as a mill following the war (Bray and Epp, 1984). Espanola today has a land area of 82.37 square km and contains a population of 5, 449 people (Statistics Canada, 2001 a). On the outskirts of Espanola are two Aboriginal reserves which Espanola General Hospital services: Whitefish River Aboriginal Reserve (population of 268 according to Statistics Canada, 2001 a) and Sagamok Aboriginal Reserve (population of 870 according to Statistics Canada, 2001 a).

Wikwemikong occupies 412.97 square km and has a population of 2, 427 (Statistics Canada, 2001 a). Wikwemikong is recognized as North America's only Unceded First Nation Reserve, which includes the Ojibwa peoples, meaning that it has not relinquished title to its land to the government (Montigny and Chambers, 2000). Wikwemikong is located on the easternmost section of land on Manitoulin Island, 554 km northwest of Toronto. Manitoulin Island is located in the northern part of Lake Huron and is the largest fresh-water island in the world being 129 km long and from 5 to 49 km wide (Turnbull, 1966). In the 1860's Manitoulin Island had been a Native Reserve for thirty years until it was opened up for non-Native use which lead to a large immigration of mostly Protestants of British descent from Southern Ontario communities. Manitoulin Island today is an agricultural settlement

whose main industry is tourism. Wikwemikong contains modern conveniences and is easily accessible by both land and water (Bray and Epp, 1984).

Sioux Lookout is a community which is located about 2000 km northwest of Toronto and is 378.64 square km with a population of 5, 336 people (Statistics Canada, 2001 a). Sioux Lookout is located in northwestern Ontario where Cree and Ojibwa peoples are found (Montigny and Chambers, 2000). Sioux Lookout developed as a service centre for the Aboriginal reserves of northwestern Ontario, including many like Sandy Lake (population of 1,704 people according to Statistics Canada, 2001 a) which is accessible only by air during most of the year and most members of the community still speak the traditional language of Ojibwa-Cree (Busch *et al.*, 1999). Sioux Lookout Zone Hospital serves as a medical hub for health care delivery to 26 Aboriginal communities in northwestern Ontario (see Figure 2 for the locations of the Aboriginal reserves) reaching as far north as Hudson Bay and comprising an area approximately 1/3 of the land mass of Ontario. The Aboriginal communities range in size from about 50 (Wawkapewin) to almost 2000 (Sandy Lake and Pikangikum) (Statistics Canada, 2001 a) with a total population of approximately 16, 000 Aboriginal peoples.

2.3 Laboratory Precautionary Measures

PCR is extremely sensitive and DNA contamination can produce a false positive or mistyping. Sample contamination was monitored by employing negative controls. Negative controls consisted of samples which contained all the necessary components of PCR, except for template DNA, during each PCR amplification. If the negative controls showed the

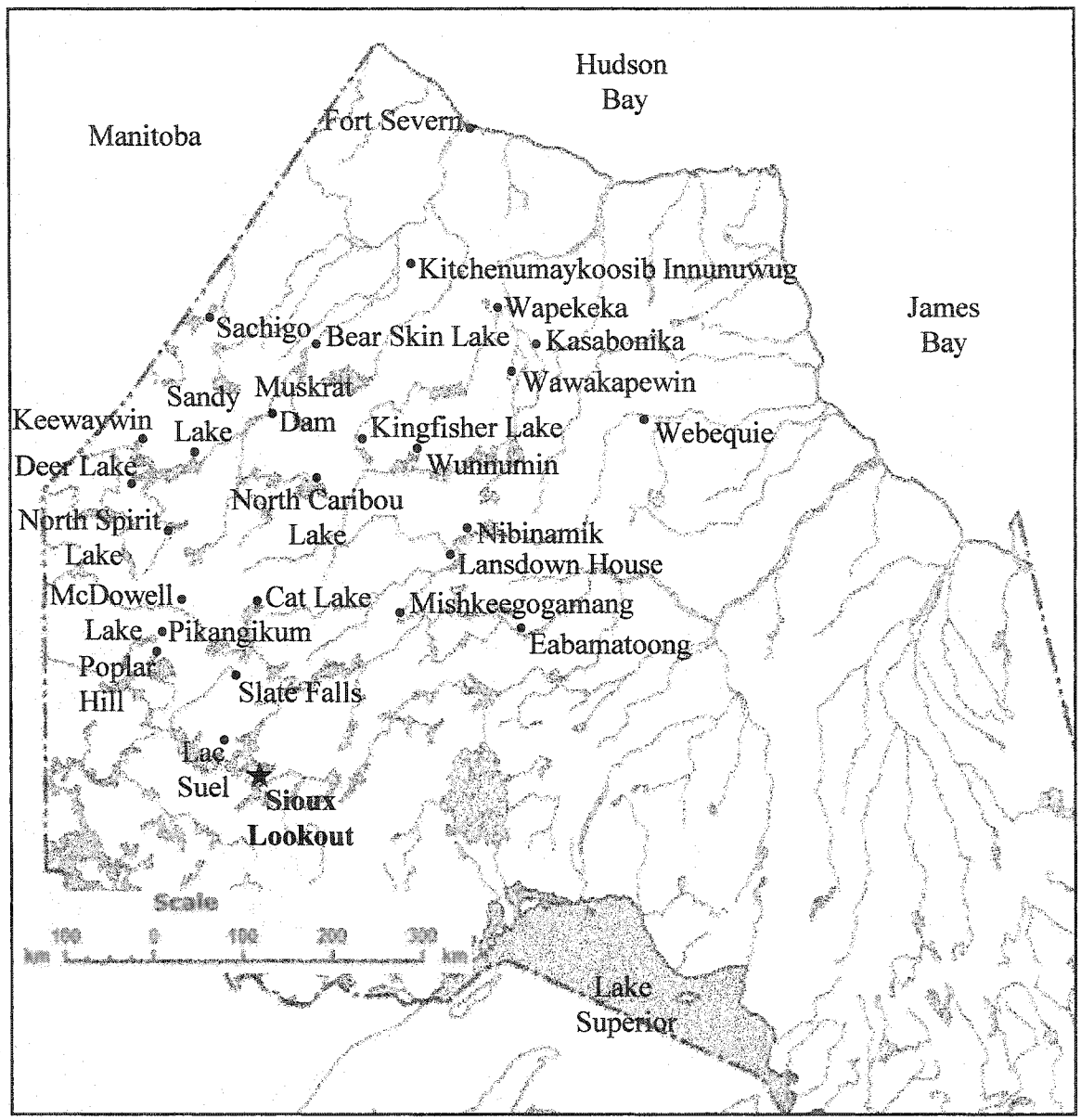


Figure 2. A map of Northwestern Ontario indicating the locations of the Aboriginal communities serviced by Sioux Lookout Zone Hospital in Sioux Lookout

presence of contaminating DNA after the PCR, the results would have been discarded. Contamination in the laboratory can occur with genomic DNA from the environment, between samples during preparation, and contamination of a sample with PCR products. Care was taken to avoid contamination of samples by designating working areas, routinely autoclaving solutions, aliquoting reagents, the use of disposable gloves, and aerosol resistant pipette tips (Kwok and Higuchi, 1989). Master mixes were used in order to reduce pipetting errors and provide a greater consistency between individual amplification reactions (Lygo *et al.*, 1994). In order to standardize the electrophoresis of samples and allow accurate comparison of alleles both within and between gels, the acrylamide solution for the polyacrylamide gel was degassed for 5 minutes in order to remove any air bubbles which would minimize band shifting due to gel irregularities. Universal precautions were observed throughout the study to prevent transmission of bloodborne pathogens.

2.4 DNA Extractions

Genomic DNA was extracted from leukocytes following a modified procedure described by the Promega Corporation, Technical Manual No. D004 (2800 Woods Hollow Road, Madison, WI, 53711-5399, USA), for cell lysis/Proteinase K DNA extraction. For each blood sample, 300 µl of whole blood was added into a sterile 1.5 ml microcentrifuge tube containing 0.5 ml of Lysis Buffer (see Appendix E). The samples were centrifuged at maximum speed (18,000 x g) for 1 minute at room temperature and the supernatant was removed and discarded. To resuspend each pellet, 1.0 ml of Lysis Buffer was added, the samples were vortexed for 30 seconds, and the supernatant was removed. This step was

repeated two more times or until no visible blood was present in the pellet. The clean pellets were resuspended in 0.5 ml of Digestion Buffer (see Appendix E) and 3 µl of Proteinase K (10 mg/ml). The samples were incubated at 60 °C for 1 hour in a water bath followed by incubation of the samples at 95 °C for 10 minutes to inactivate the Proteinase K. The samples were briefly centrifuged to bring the contents to the bottom of the tube and were then stored at -20 °C (Kawasaki, 1990).

2.5 Beta-Globin Screening

The polymerase chain reaction was used to assess the success of the DNA extractions. Primers that target the human Beta-Globin gene, which codes for the Beta-chain of the globin protein in red blood cells, were used. The Beta-Globin gene is a single-copy gene and possesses a conserved region in the human genome which guarantees its presence. The Beta-Globin primers GH20 (5'-CAACTTCATCCACGTTCACC-3') and PCO4 (5'-GAAGAGCCAAGGACAGGTAC-3') were synthesized by University Core DNA Services, The University of Calgary. All 400 samples were assessed using the Beta-Globin primers. The PCR reactions were carried out using 2.5 µl DNA extract, 0.5 µM of each primer GH20 and PCO4, 1.5 mM $MgCl_2$, 0.3 U of AmpliTaq™ DNA Polymerase, 200 µM of deoxynucleotide triphosphates (dNTP's), 10 mM Tris (pH 8.4), 50 mM KCL, and sterile double distilled water to a final reaction volume of 25 µl. Each reaction was overlaid with a single drop of mineral oil to avoid evaporation of the reaction mixture. Another drop of mineral oil was placed in each well of the DNA Thermocycler (Perkin Elmer Cetus, Emeryville, CA) to evenly distribute the heat. A total of 35 cycles were carried out for all

reactions. Each cycle consisted of denaturation at 95 °C for 1.5 minutes, primer annealing at 65 °C for 2 minutes, and primer extension at 72 °C for 1 minute. Prior to the first cycle, the reactions were heated at 95 °C for 5 minutes.

The products of the PCR amplification were visualized on a 1% agarose gel using a 5 µl sample from the PCR amplification mixed with 1 µl of Agarose Gel Loading Buffer (see Appendix E) as an indicator. The gels were run at 2.8 V/cm for 90 minutes in 1X TAE (Tris-acetate/EDTA) (see Appendix E). The gels were stained with 0.5 mg/ml ethidium bromide for 15 minutes followed by a 10 minute destaining in distilled water. The gels were then viewed and photographed under UV light in a UV Viewing System (FBTIV-88 Fisher Biotech, Fisher Scientific). A 1 Kb plus ladder (Invitrogen Life Technologies) was used as a molecular size marker. Samples negative for the expected 268 base pair fragment were reextracted and retested. Samples were chosen for further study if a 268 base pair fragment was visible.

2.6 D18S535 Locus Targeting

The polymerase chain reaction was performed on all 400 successful DNA extractions. Primers which target the D18S535 locus were used. The D18S535 primers 1 (5'-TCATGTGACAAAAGCCACAC-3') and 2 (5'-GACAGAAATATAGATGAGAATGCA-3') (Laureu *et al.*, 1998) were synthesized by Invitrogen Life Technologies. The PCR reactions were carried out in 25 µl volumes using the same reaction components as the Beta-Globin amplification with the following changes: 10 µl DNA extract, 0.25 µM of each primer, 1 U of AmpliTaq™ DNA Polymerase, and 0.01 % Gelatin. A total of 35 cycles were carried out

for all reactions. Each cycle consisted of denaturation at 94 °C for 1 minute, primer annealing at 60 °C for 2 minutes, and primer extension at 72 °C for 1.5 minutes. Prior to the first cycle, the reactions were heated at 95 °C for 5 minutes and a final extension of 72 °C for 7 minutes after the last cycle.

A sequenced allelic ladder containing the 8 common alleles for the D18S535 locus was constructed by adding samples together which contained the eight common alleles and amplified using PCR (Lareu *et al.*, 1998). The allelic ladder was provided by the Institute of Legal Medicine, Santiago de Compostela, Galacia, Spain for this study. The ladder was reamplified using 1 µl of a 1:2000 dilution following the conditions used for D18S535 locus targeting.

Since the use of polyacrylamide gels is a more costly and time consuming process, the PCR products were prescreened on a 2 % agarose gel to ensure the samples were successfully amplified. PCR products were prescreened on a 2 % agarose gel using a 5 µl sample from the PCR amplification mixed with 1 µl of Agarose Gel Loading Buffer (see Appendix E) as an indicator. The gels were run at 2.8 V/cm for 120 minutes in 1X TAE. The gels were stained with 0.5 mg/ml ethidium bromide for 15 minutes and destained for 10 minutes in distilled water. The gels were then viewed and photographed under UV light. Samples negative for the expected amplified 130 - 158 base pair fragment were reextracted or purified using a Phenol/Chloroform/Isoamyl alcohol procedure (see Appendix F). Only samples which were successfully amplified were used in the study.

The products of the PCR amplification were resolved on 0.4 mm denaturing polyacrylamide gels [6 % (19:1 acrylamide/bisacrylamide) containing 8 *M* Urea, degassed

for 5 minutes, 40 µl of 0.25 mg/µl Ammonium Persulfate and 40 µl TEMED were added after filtration] and run on a 20 x 40 cm Sequi-Gen® GT Nucleic Acid Electrophoresis Cell (BIO-RAD). The gel was allowed to polymerize for at least 1 hour. Four µl of sample from the PCR amplification was mixed with 4 µl of freshly prepared Polyacrylamide Gel Loading Buffer (see Appendix E) and 5 µl of the D18S535 allelic ladder was mixed with 5 µl of Polyacrylamide Gel Loading Buffer. Samples and D18S535 allelic ladders were denatured for 5 minutes in a boiling water bath, snap cooled for 2 minutes, briefly spun in a centrifuge to bring the contents to the bottom of the tube, and placed on ice before loading on denaturing gels. The gels were preheated at a constant power of 42 W,1800 V, and 22 mA until equilibrated to 50°C. Three sets of 21 samples and 3 D18S535 allelic ladders were loaded into a 24 well shark tooth comb, with 30 minutes between loadings, and electrophoresis was conducted for 2.5 hours total in 1X TBE (Tris-borate/EDTA) buffer (see Appendix E).

2.7 Silver Staining and Allelic Designation

After electrophoresis, the PCR products were visualized with Silver Sequence™ DNA Sequencing System by the Promega Corporation (2800 Woods Hollow Road, Madison, WI, 53711-5399, USA) according to manufacturers guidelines (Technical Manual No. D004). To summarize, the gel was placed in a tray, covered with 2 L of a fix/stop solution (see Appendix E), and agitated for 20 minutes followed by a 2 minute rinse in double distilled water using agitation. The rinse was repeated two more times. The gel was transferred to 2 L of a staining solution (see Appendix E) and was agitated for 30 minutes.

The gel was removed from the staining solution, briefly dipped into a tray containing double distilled water, drained, and immediately placed into a tray containing 1 L of chilled developing solution (see Appendix E). The gel was agitated by rocking the tray until the first bands were visible. The gel was then transferred to another tray containing 1 L of chilled developing solution and continued developing until all bands were visible. To terminate development and fix the gel, 1 L of the fix/stop solution was added directly to the gel in the developing solution and agitated for 2 minutes. The gel was then rinsed for 2 minutes using agitation and a photograph of the gel was taken. The gel was placed in a 10 % NaOH solution for 45 minutes, rinsed twice with distilled water for 2 minutes, and using filter paper was peeled off the glass plate. The gel was dried for 1 hour at 85 °C using a gel dryer (Bio-Rad Model 543) for a permanent record.

Allelic designation was accomplished by visually comparing allelic bands to the sequenced allelic ladder. Allelic ladders were utilised to ensure accurate designation of alleles and preliminary identification of rare alleles. The allelic bands were scored twice to ensure accurate designation and any ambiguous samples were eliminated from the study. Nomenclature of the STR alleles followed the recommendations of the DNA Commission of the International Society for Forensic Haemogenetics (DNA recommendations, 1992; DNA recommendations, 1994; Bär *et al.*, 1997) where alleles are named with digits according to the number of complete tandem repeats. If a partial repeat is present, it is designated by the number of bases prefixed by a decimal point.

2.8 Statistical Calculations

Allele frequencies for the D18S535 locus for all four Northern Ontario populations were obtained using the Promega Software PowerStats. Allele frequencies were determined using the gene counting method, where the frequency of each allele for the D18S535 locus was calculated from the numbers of each genotype in the sample set (Edwards *et al.*, 1992).

There is a large number of possible genotypes at STR loci; therefore, the sample size required to observe all possible genotypes is large. Consequently, estimation of genotypic probabilities must be based on allele frequencies (Chakraborty, 1992). This theory assumes that genotypic frequencies comply with Hardy-Weinberg Equilibrium expectations. Hardy-Weinberg proportions are used to predict genotypic proportions from allele frequencies at a Mendelian locus when certain conditions are satisfied. The necessary conditions are that the population is large, randomly mating, and free from mutation, migration and natural selection. If the conditions of the Hardy-Weinberg law are met, then the genotypic proportions at a Mendelian locus remain constant from generation to generation and they are given by the product of respective allele frequencies occurring in the proportions p^2, 2pq, and q^2 where p and q equal the frequencies of the alleles in the population (Ayala, 1982). These are idealistic conditions and most natural populations as well as genetic loci may violate one or more of these conditions. Since most Northern Ontario populations are small and rural, a certain level of inbreeding is expected to occur. The populations were tested using an exact test (Guo and Thompson, 1992) by the computer program GENEPOP version 1.2 (Raymond and Rousset, 1995) in order to determine if their genotypic proportions deviated from those predicted by the Hardy-Weinberg law.

The exact test was used to examine independence of frequencies for all genotypes. The exact test involved reconstruction of the genotype frequencies by random shuffling of all alleles among individuals in each population database, much like shuffling and dealing a pack of playing cards. GENEPOP created 2000 datasets from shuffling the specific observed alleles 2000 times. For each locus this gives a new set of individual genotypes of the same size as the original sample, but now with independence guaranteed by the randomness of the shuffling. Each of the 2000 sampled populations which were generated had identical allele frequencies as the sample database population of actual individuals. From every replication of shuffling, a new genotype frequency distribution was used to compute the respective test criteria. Levels of significance were determined from the number of times (out of 2000 replications of shuffling) the computed statistics exceeded the observed values of these statistics in the population samples. Interest was focussed on levels around the conventional .05 level, with 95 % confidence, by 2,000 permutations (Evett *et al.*, 1996). GENEPOP provided the probability of error when rejecting H_o (Raymond and Rousset, 1995).

The Statistical Program for Social Science (SPSS for Vax, Version 11) was used to perform a series of multivariate analysis of variance (MANOVA) tests in order to determine if there was an overall homogeneity of variance in the allelic distribution in the four Northern Ontario populations in conjunction with previously published D18S535 studies from Europe and South America, in the Espanola Caucasians in conjunction with previously published D18S535 studies from Europe, in the previously published D18S535 studies from Europe, in the Northern Ontario and South American Aboriginal populations, in the four Northern

Ontario populations, and in the three Northern Ontario Aboriginal populations.

The previously published studies from Europe and South America included a population from Croatia, Halla Area in Northeastern Germany, Aragon in Northeastern Spain, Galacia in Northwestern Spain, Asturias in Northern Spain, and Valencia in Eastern Spain. The previously studied populations also included the Spanish islands of Majorca, Minorca, and Ibiza, a Chuetas population (a small Jewish community from the island of Majorca), and a population from Puna in Northwestern Argentina (refer to Appendices G and H for maps indicating the locations of the previously published D18S535 studies from Europe and South America; Appendix I for tables displaying the relative allele frequencies for the previously published D18S535 studies from Europe and South America).

SPSS was also used to create contingency tables and Pearson chi-square test statistics, as *post hoc* analyses, to determine which D18S535 alleles differed in their distribution in the four Northern Ontario populations in conjunction with previously published D18S535 studies from Europe and South America, in the Espanola Caucasians in conjunction with previously published D18S535 studies from Europe, in the Northern Ontario and South American Aboriginal populations, in the four Northern Ontario populations, and in the three Northern Ontario Aboriginal populations.

The Northern Ontario populations and previously studied populations from Europe and South America were compared to each other using GENEPOP version 1.2 (Raymond and Rousset, 1995) to determine if the allelic distributions were identical. This was done by pairwise comparisons of the allele frequency distributions using the exact test. For the D18S535 locus, GENEPOP created a contingency table describing the allelic

composition in each population and tested the lack of differentiation for all pairs of populations. The estimate of the probability of error when rejecting H_o, which was that there was no allelic differentiation between the populations, and its standard error were provided (Raymond and Rousset, 1995). Also, the Espanola and Wikwemikong Aboriginals were pooled and compared to the Sioux Lookout Aboriginals using a pairwise comparison to determine if the more isolated Sioux Lookout Aboriginals were similar in their allelic distribution to the Aboriginals with more Caucasian contact.

In populations which indicated a significant difference in their allelic distributions using the pairwise comparisons, 2 x 2 contingency tables and Pearson's chi-square were generated using SPSS to determine which allele(s) was significantly different between the populations.

A statistical means of comparing two samples is to compare their allelic counts with a 2 x 2 contingency table and Pearson chi-square test statistic. A genetic approach is comparing allele frequency distributions between samples from different populations by means of Wright's *F* statistics (Wright, 1951; Weir and Cockerham, 1984; Evett and Weir, 1998). Two *F* statistics were estimated, using the GENEPOP statistical software for the Northern Ontario populations. F_{ST} was calculated as an indicator of the D18S535's differentiation level among populations (divergence among populations) and F_{IS} which is a measure of the amount of nonrandom mating within populations (intrapopulation inbreeding) and in this sense provided an estimate of the inbreeding coefficient, θ, in each population (Wright, 1951; Weir and Cockerham, 1984). *F* statistics have a theoretical range of 0 - 1 (Bowcock *et al.*, 1991) and if the *F* statistics values in the interval did not include zero, the

estimate was considered to be significantly different from zero.

To determine how the four Northern Ontario populations along with the previously studied populations from Europe and South America clustered together with respect to the D18S535 locus, a cluster analysis was performed using Euclidean distances measurement (the distance between two cases is the square root of the sum of the squared differences in values for each variable) to calculate distance and average linkage between groups to form the clusters. The cluster analysis was performed based on D18S535 allele frequency distributions. Contingency tables and Pearson chi-square analyses were used to determine if the clusters formed were affected by geography and ethnicity. A resultant Euclidean dissimilarity coefficient matrix based on D18S535 allelic distribution indicates the magnitude of difference between two populations. By inference, the smaller the statistic, the more similar were the groups compared. All analyses were carried out using SPSS.

The potential usefulness of the D18S535 in Northern Ontario forensic casework analysis was evaluated by means of the power of discrimination ($PD = 1 - \sum P^2_j$, where P_j is the frequency of each genotype) (Fisher, 1951; Jones, 1972) which is the probability that two individuals chosen at random from a given population have different genotypes and the heterozygosity value (h) (Nei and Roychoudhury, 1974) which was calculated from the observed numbers of heterozygotes and homozygotes within each sample set. Also, the "a priori" probability that a falsely accused father will be excluded, chance of exclusion [$CE = h^2(1-2hH^2)$, where h is the heterozygosity and H is the homozygosity] was calculated (Brenner and Morris, 1990). All forensic calculations were carried out using the Promega Software, PowerStats.

3. Results

Successful PCR amplification is dependent upon the presence of DNA in the sample. To confirm the presence of sufficient amplifiable human DNA and the absence of PCR inhibitors in each sample, the human Beta-Globin gene was amplified by PCR. A 1 Kb plus ladder was used as a molecular size marker. The results showed that the expected 268 base pair Beta-Globin fragment was present in most samples. The presence of this 268 bp band indicated that the sample contained enough DNA to be amplified and that the samples contained little or no PCR inhibitors. Sample concentrations produced variations in band intensities (see Figure 3). A strong band indicated a higher quantity and/or better quality of DNA in the sample than a sample which displayed a fainter band. The negative control showed no band which indicated that there was no contaminating DNA. The distribution of D18S535 alleles in the populations studied was determined by visually comparing allelic bands in denaturing polyacrylamide gels to the sequenced allelic ladder. Examples of the resolution of the D18S535 four base pair repeat unit in each of the four Northern Ontario populations are shown in Figures 4, 5, 6, and 7. The relative allele frequencies of the four Northern Ontario populations are shown in Table 1 and Figure 8.

The allelic designation was based on the number of GATA repeat units according to the recommendations of the DNA Commission of the International Society for Forensic Haemogenetics (1992; 1994). All the observed alleles appeared to migrate with the alleles present in the allelic ladder; therefore, no new alleles in relation to those previously found were observed. A total of 8 alleles were observed in the Espanola Caucasian population with

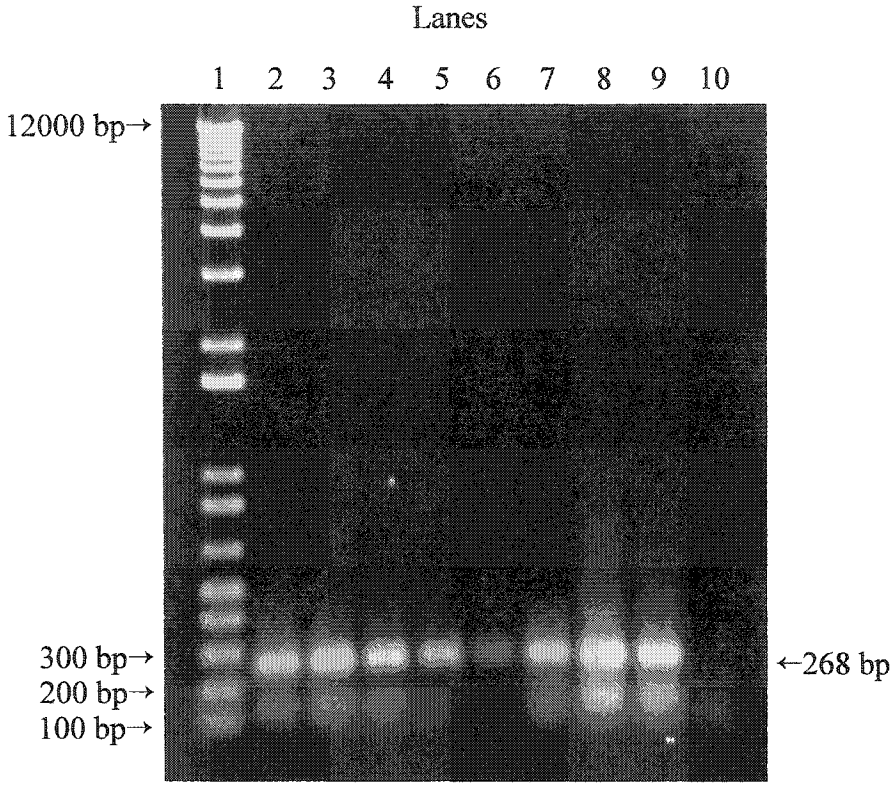


Figure 3. Agarose gel showing the 268 bp fragment of the Beta-Globin locus in the Wikwemikong Aboriginal population (lane 1 is the 1 Kb plus ladder containing 100 bp (base pairs) fragment at the bottom to 12000 bp at the top, lanes 2 to 9 contain the PCR amplified 268 bp Beta-Globin fragment, lane 10 contains the negative control which showed no amplification)

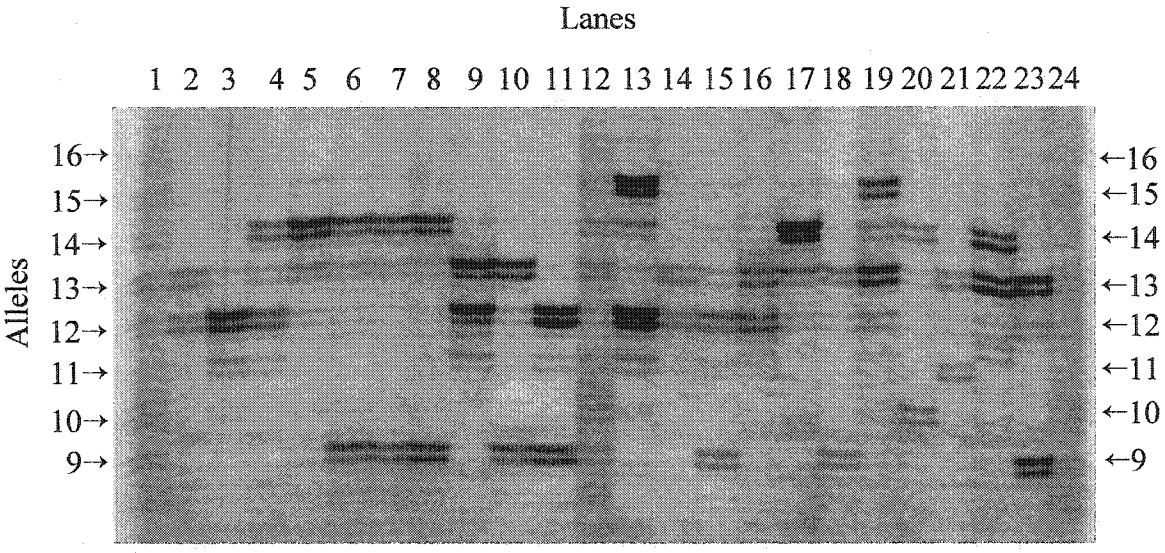


Figure 4. Silver-stained denaturing 6 % polyacrylamide gel displaying D18S535 profiles in the Espanola Caucasian population (the D18S535 genotypes from left to right are lane 2: 13-12, lane 3: 12-12, lane 4: 14-12, lane 5: 14-14, lane 6: 14-9, lane 7: 14-9, lane 8: 14-9, lane 9: 13-12, lane 10: 13-9, lane 11: 12-9, lane 13: 15-12, lane 14: 13-12, lane 15: 12-9, lane 16: 13-12, lane 17: 14-14, lane 18: 13-9, lane 19: 15-13, lane 20: 14-10, lane 21: 13-11, lane 22: 14-13, and lane 23: 13-9. The ladders are composed of alleles 9, 10, 11, 12, 13, 14, 15 and 16, and are shown in lanes 1, 12, and 24)

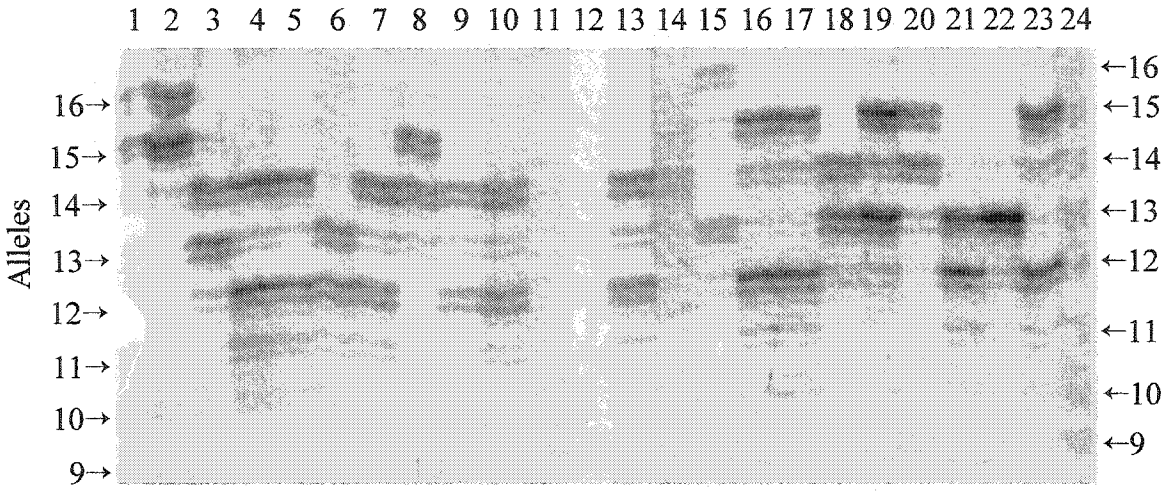


Figure 5. Silver-stained denaturing 6 % polyacrylamide gel displaying D18S535 profiles in the Espanola Aboriginal population (the D18S535 genotypes from left to right are lane 2: 16-15, lane 3: 14-13, lane 4: 14-12, lane 5: 14-12, lane 6: 13-12, lane 7: 14-12, lane 8: 15-14, lane 9: 14-12, lane 10: 14-12, lane 11: no amplification, lane 13: 14-12, lane 14: 14-14, lane 15: 16-13, lane 16: 15-12, lane 17: 15-12, lane 18: 14-13, lane 19: 15-13, lane 20: 15-14, lane 21: 13-12, lane 22: 13-13, and lane 23: 15-12. The ladders are composed of alleles 9, 10, 11, 12, 13, 14, 15 and 16, and are shown in lanes 1, 12, and 24)

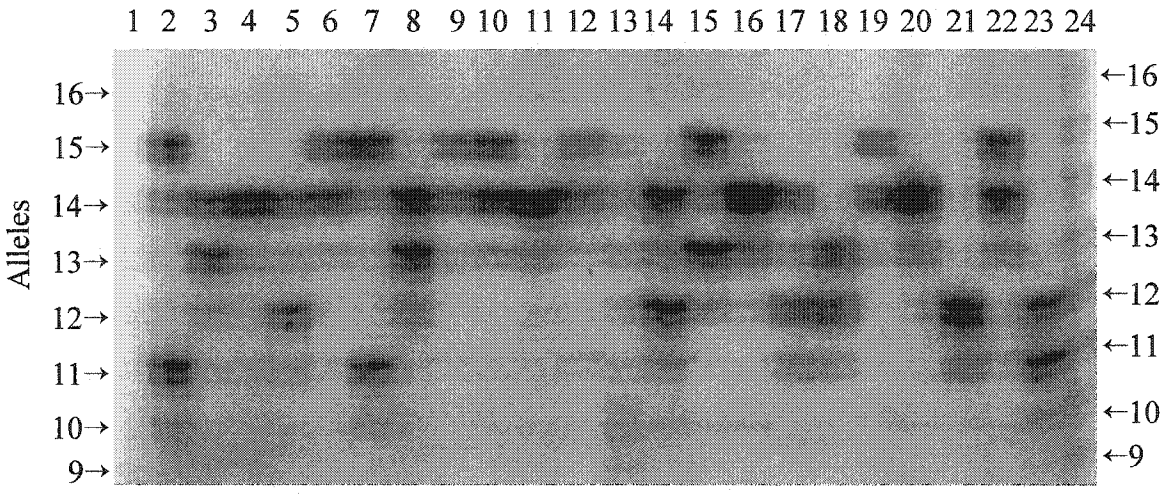


Figure 6. Silver-stained denaturing 6 % polyacrylamide gel displaying D18S535 profiles in the Sioux Lookout Aboriginal population (the D18S535 genotypes from left to right are lane 2: 15-11, lane 3: 14-13, lane 4: 14-14, lane 5: 14-12, lane 6: 15-14, lane 7: 15-11, lane 8: 14-13, lane 9: 15-14, lane 10: 15-14, lane 11: 14-14, lane 12: 15-14, lane 14: 14-12, lane 15: 15-13, lane 16: 14-14, lane 17: 14-12, lane 18: 13-12, lane 19: 15-14, lane 20: 14-14, lane 21: 12-12, lane 22: 15-14, and lane 23: 12-11. The ladders are composed of alleles 9, 10, 11, 12, 13, 14, 15 and 16, and are shown in lanes 1, 13, and 24)

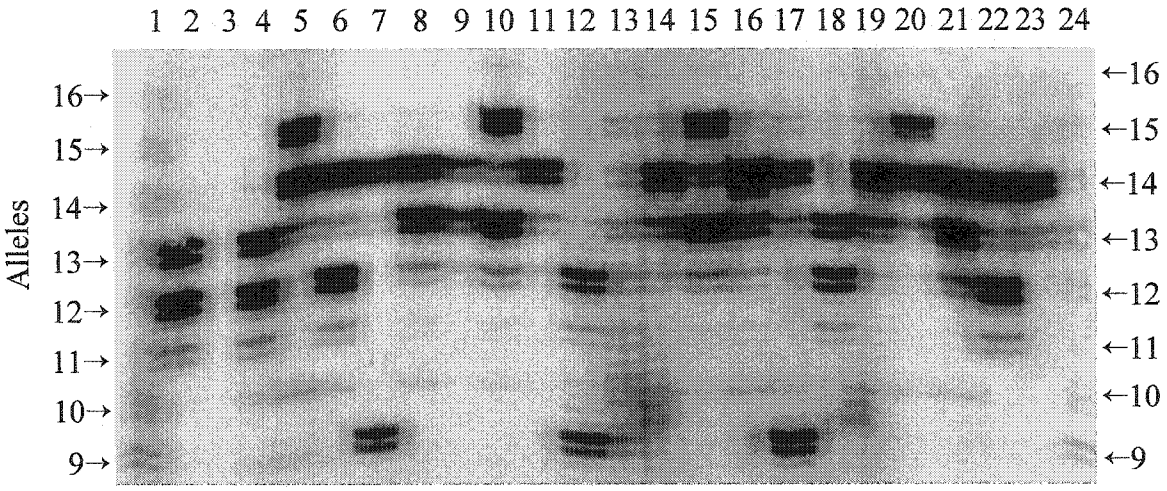


Figure 7. Silver-stained denaturing 6 % polyacrylamide gel displaying D18S535 profiles in the Wikwemikong Aboriginal population (the D18S535 genotypes from left to right are lane 2: 13-12, lane 3: no amplification, lane 4: 13-12, lane 5: 15-14, lane 6: 14-12, lane 7: 14-9, lane 8: 14-13, lane 9: 14-13, lane 10: 15-13, lane 11: 14-14, lane 12: 12-9, lane 14: 14-14, lane 15: 15-13, lane 16: 14-14, lane 17: 14-9, lane 18: 13-12, lane 19: 14-14, lane 20: 15-14, lane 21: 14-13, lane 22: 14-12, and lane 23: 14-14. The ladders are composed of alleles 9, 10, 11, 12, 13, 14, 15 and 16, and are shown in lanes 1, 13, and 24)

Table 1. Relative D18S535 allele frequency distribution in Caucasian and Aboriginal populations from Northern Ontario

Allele	Relative Frequency (f)			
	Espanola Caucasians (N = 99)	Espanola Aboriginals (N = 99)	Sioux Lookout Aboriginals (N = 100)	Wikwemikong Aboriginals (N = 99)
9	.136	.015	.005	.126
10	.020	.005	.000	.000
11	.010	.010	.080	.005
12	.303	.288	.195	.126
13	.212	.167	.170	.131
14	.207	.369	.395	.444
15	.086	.121	.155	.152
16	.025	.025	.000	.015

N = the number of individuals

a sample size of 99. Alleles 12 (relative frequency = 0.303) and 13 (relative frequency = 0.212) were found to be the most common alleles. A total of 8 alleles were observed in the 99 samples from the Espanola Aboriginal population. Alleles 12 (relative frequency = 0.288) and 14 (relative frequency = 0.369) were found to be the most common alleles. A total of 6 alleles were observed in the 100 samples from the Sioux Lookout Aboriginal population. Alleles 12 (relative frequency = 0.195) and 14 (relative frequency = 0.395) were determined to be the most common alleles. In the Wikwemikong population of 99 samples, a total of 7 alleles were observed. Alleles 14 (relative frequency = 0.444) and 15 (relative frequency = 0.152) were found to be the most common alleles.

The range of alleles varied greatly at the D18S535 locus across the four Northern Ontario populations. The range tended to be the narrowest for Sioux Lookout Aboriginals (6 alleles), while the Wikwemikong Aboriginals (7 alleles), Espanola Caucasians (8 alleles), and Espanola Aboriginals (8 alleles) displayed a wider range. Uneven allele distributions were observed in the populations shown by distinctive alleles with high frequencies, specifically the mid-sized alleles (refer to Table 1 and Figure 8).

A total of 36 possible genotypes at the D18S535 locus were expected to occur in each population based on the combinations of the eight common D18S535 alleles. There was a distinct difference between population groups in terms of the number of genotype combinations, which was much larger in the Espanola Caucasian population. Table 2 shows the observed and expected D18S535 locus genotypes in the four Northern Ontario populations. A total of 19 genotypes were represented in the Espanola Caucasian population, and as expected from the allele frequencies, the most common genotypes were 12-13

Table 2. Observed and expected D18S535 locus genotypes in Caucasian and Aboriginal populations from Northern Ontario

Genotype	Espanola Caucasians		Espanola Aboriginals		Sioux Lookout Aboriginals		Wikwemikong Aboriginals	
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
9,9	0	1.78	0	0.02	0	0.00	0	1.52
10,9	0	0.55	0	0.02	0	0.00	0	0.00
10,10	0	0.03	0	0.00	0	0.00	0	0.00
11,9	0	0.27	0	0.03	0	0.08	0	0.13
11,10	0	0.04	0	0.01	0	0.00	0	0.00
11,11	0	0.01	0	0.01	0	0.60	0	0.00
12,9	9	8.22	2	0.87	0	0.20	3	3.17
12,10	0	1.22	0	0.29	0	0.00	0	0.00
12,11	0	0.61	0	0.58	9	3.14	0	0.13
12,12	11	8.99	1	8.10	6	3.72	3	1.52
13,9	9	5.76	0	0.50	0	0.17	3	3.30
13,10	0	0.85	0	0.17	0	0.00	0	0.00
13,11	1	0.43	0	0.34	1	2.73	0	0.13
13,12	13	12.79	17	9.55	4	6.66	5	3.30
13,13	4	4.37	2	2.68	1	2.81	1	1.65
14,9	5	5.62	0	1.11	0	0.40	14	11.17
14,10	1	0.83	0	0.37	0	0.00	0	0.00
14,11	1	0.42	1	0.74	2	6.35	0	0.45
14,12	8	12.49	29	21.12	11	15.48	9	11.17
14,13	8	8.74	7	12.23	22	13.50	9	11.61
14,14	8	4.16	13	13.34	15	15.48	20	19.43
15,9	4	2.33	1	0.37	1	0.16	4	3.81
15,10	3	0.35	1	0.12	0	0.00	0	0.00

Table 2. (Continued)

Genotype	Espanola Caucasians		Espanola Aboriginals		Sioux Lookout Aboriginals		Wikwemikong Aboriginals	
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
15,11	0	0.17	1	0.24	4	2.49	0	0.15
15,12	4	5.18	7	6.94	3	6.08	2	3.81
15,13	3	3.62	3	4.02	5	5.30	7	3.96
15,14	2	3.54	9	8.89	14	12.31	15	13.40
15,15	0	0.69	0	1.40	2	2.34	1	2.21
16,9	0	0.69	0	0.08	0	0.00	1	0.38
16,10	0	0.10	0	0.03	0	0.00	0	0.00
16,11	0	0.05	0	0.05	0	0.00	1	0.02
16,12	4	1.52	0	1.45	0	0.00	0	0.38
16,13	0	1.07	2	0.84	0	0.00	0	0.40
16,14	0	1.04	1	1.85	0	0.00	1	1.34
16,15	1	0.43	2	0.61	0	0.00	0	0.46
16,16	0	0.05	0	0.05	0	0.00	0	0.02

Obs. = Observed
Exp. = Expected

(relative frequency = 0.131) and 12-12 (relative frequency = 0.112). There were 17 different genotypes observed in the Espanola Aboriginal population and the most common genotypes observed were 12-14 (relative frequency = 0.293) and 12-13 (relative frequency = 0.172). In the Sioux Lookout Aboriginal population a total of 15 genotypes were observed with genotypes 13-14 (relative frequency = 0.220) and 14-14 (relative frequency = 0.150) being the most common genotypes. There were 17 different genotypes observed in the Wikwemikong Aboriginal population and the most common genotypes observed were 14-14 (relative frequency = 0.202) and 14-15 (relative frequency = 0.152).

To test whether genotype frequencies conformed to Hardy-Weinberg Equilibrium expectations at the D18S535 locus, an exact test was performed using the null hypothesis which was that there were no detectable deviations from observed and expected genotypes. The Hardy-Weinberg Equilibrium results are summarized in Table 3 and indicated that the observed genotype values showed agreement with those expected under Hardy-Weinberg Equilibrium ($p > 0.05$) for the Espanola Caucasian and the Wikwemikong Aboriginal populations. The results showed detectable departures from HWE in the Espanola Aboriginal and Sioux Lookout Aboriginal populations ($p < 0.05$). If populations deviate from HWE at a particular locus, an excess of homozygotes could be evident in the population sample (Devlin *et al.*, 1990); therefore, possible causes of deviation from HWE was further tested by testing an alternate hypothesis of heterozygote deficiency using the GENEPOP statistical program. Tests for heterozygote deficiency were carried out using 2000 shuffled arrays to determine *p*-values and standard errors. These results suggested that the deviant populations did not demonstrate heterozygote deficiency (Espanola Aboriginals:

Table 3. Statistical values of interest for the Hardy-Weinberg equilibrium and statistical parameters of forensic interest for the D18S535 in Caucasian and Aboriginal populations from Northern Ontario

Population	Exact Test (*P*-value)	Power of Discrimination (PD)	Chance of Exclusion (CE)	Heterozygosity (h)
Espanola Caucasians	0.0530	0.92	0.54	0.77
Espanola Aboriginals	0.0012*	0.85	0.67	0.84
Sioux Lookout Aboriginals	0.0031*	0.88	0.53	0.76
Wikwemikong Aboriginals	0.2324	0.89	0.51	0.75

* indicates a significant value ($p < 0.05$)

p = 0.99, standard error = 0.001; Sioux Lookout Aboriginals: p = 0.68, standard error = 0.009) but did show a significant heterozygote excess since the p-values were not significant for this test ($p > 0.05$).

Inbreeding is the correlation between uniting gametes which produces a decrease in the proportion of heterozygotes and a corresponding increase in homozygotes in a population. The amount of inbreeding in a population is revealed by the inbreeding coefficient, F_{IS}, where $0 \leq F_{IS} \leq 1$ (Ward and Sing, 1970). The null hypothesis of no inbreeding in these populations was $F_{IS} = 0$. The F_{IS} values in the Northern Ontario populations were all negative except for the Espanola Caucasian population (see Table 4). The positive and negative F_{IS} values observed indicated that homozygote proportions were altered in all the Northern Ontario populations. The negative F_{IS} values observed in the Espanola Aboriginal and Sioux Lookout Aboriginal populations were in agreement with the deviations in HWE previously tested. The deviations from HWE were in the direction of excess heterozygotes in the Espanola Aboriginal and Sioux Lookout Aboriginal populations as indicated by the negative sign (George Carmody 2002, personal communication). These results corroborate the conclusion of heterozygote excess already indicated by the previously tested alternate hypothesis. The alteration of homozygotes proportions did not affect HWE in the Espanola Caucasian and Wikwemikong Aboriginal populations.

Distribution of all the D18S535 alleles are shown in Figures 8 to 11, and the distribution of each individual allele across the Northern Ontario populations and all other populations studied to date are shown in Figures 12 to 20. Based on the distributions observed in the figures, the populations displayed a uni- or bimodal pattern of allele

Table 4. Wright's F statistic, $F_{IS,}$ for the D18S535 locus in Caucasian and Aboriginal populations from Northern Ontario

Population	F_{IS}
Espanola Caucasians	+0.037
Espanola Aboriginals	-0.131
Sioux Lookout Aboriginals	-0.013
Wikwemikong Aboriginals	-0.019

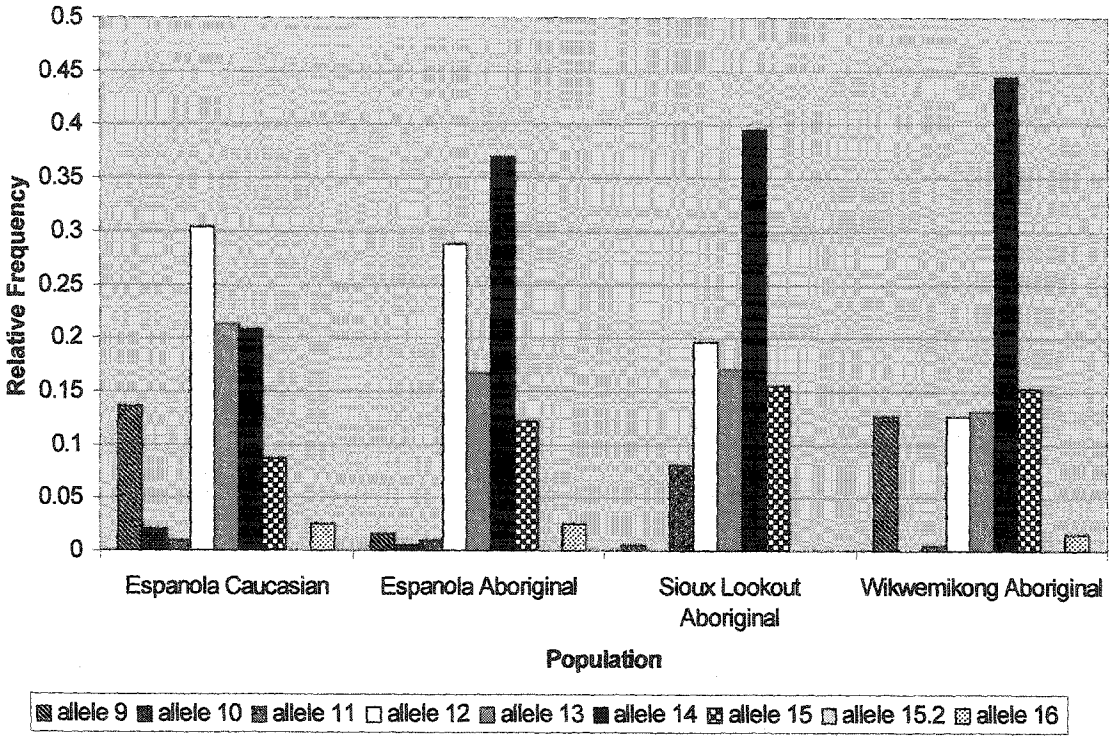


Figure 8. Distribution of the D18S535 alleles in Caucasian and Aboriginal populations from Northern Ontario

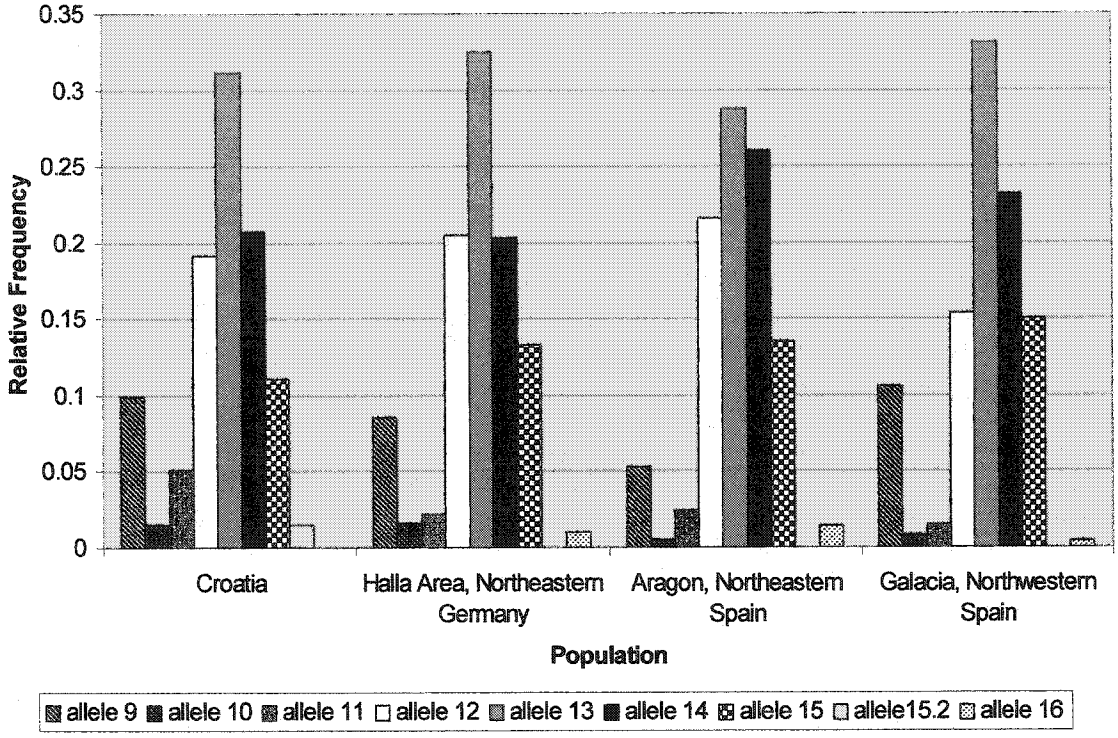


Figure 9. Distribution of the D18S535 alleles in populations from Croatia, Halla Area in Northeastern Germany, Aragon in Northeastern Spain, and Galacia in Northwestern Spain

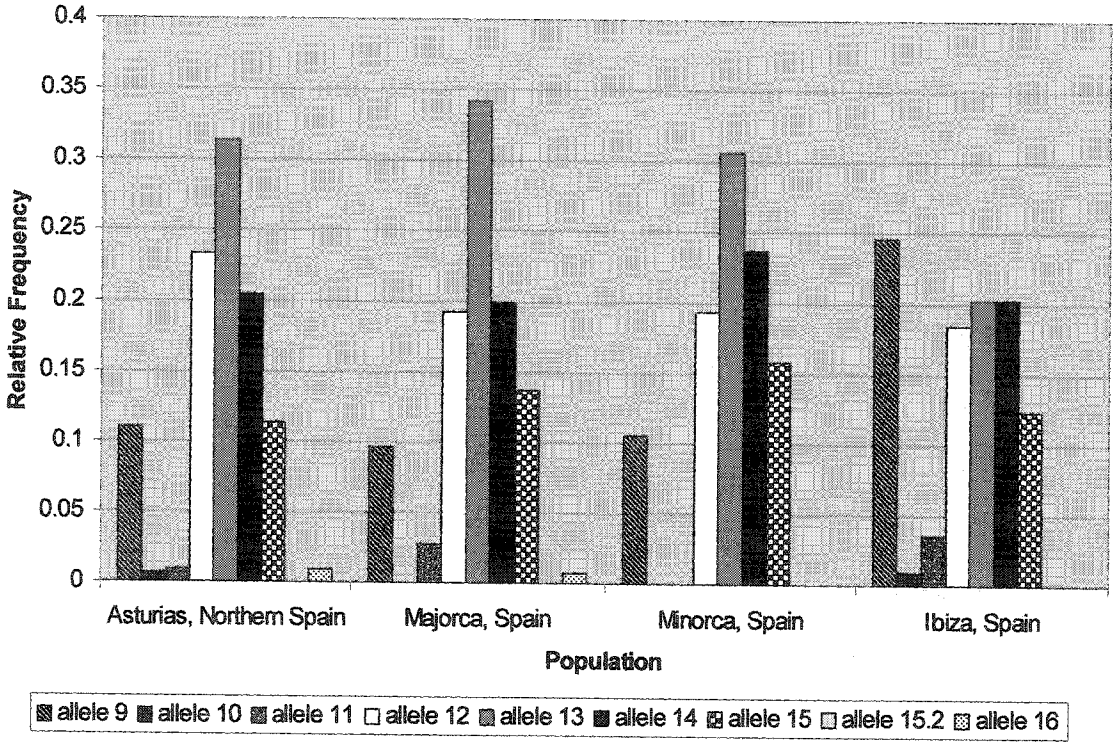


Figure 10. Distribution of the D18S535 alleles in populations from Asturias in Northern Spain, and the Spanish Islands of Majorca, Minorca, and Ibiza

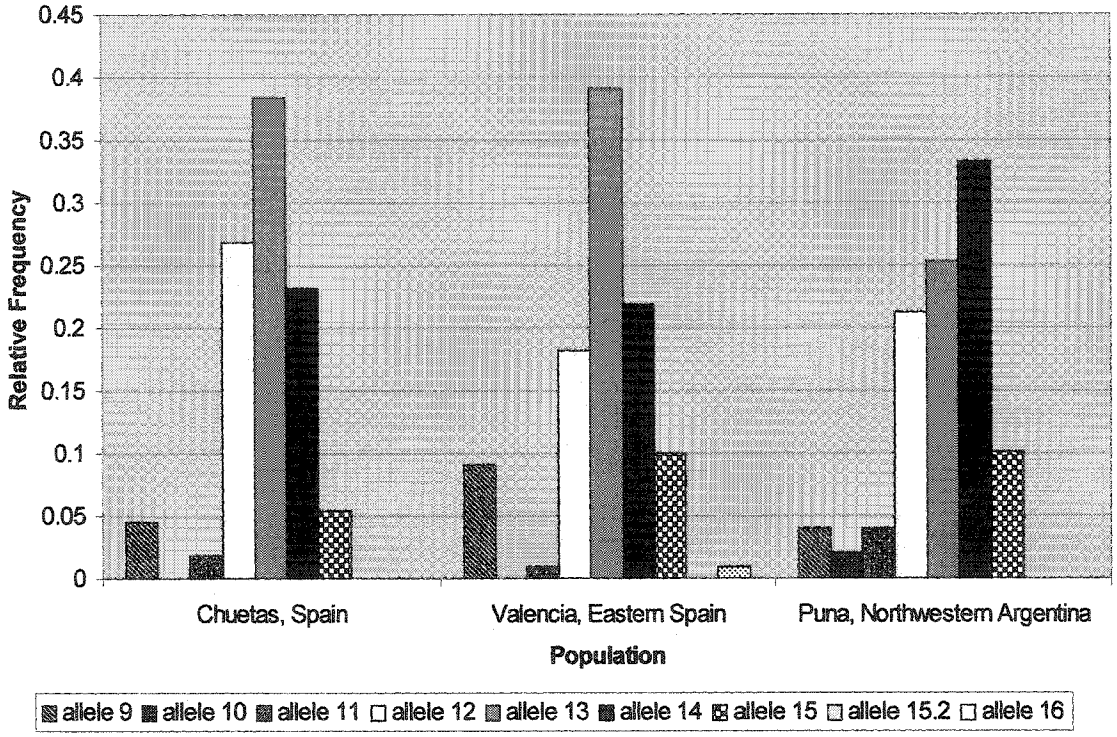


Figure 11. Distribution of the D18S535 alleles in the Spanish populations of Chuetas and Valencia in Eastern Spain, and Puna in Northwestern Argentina

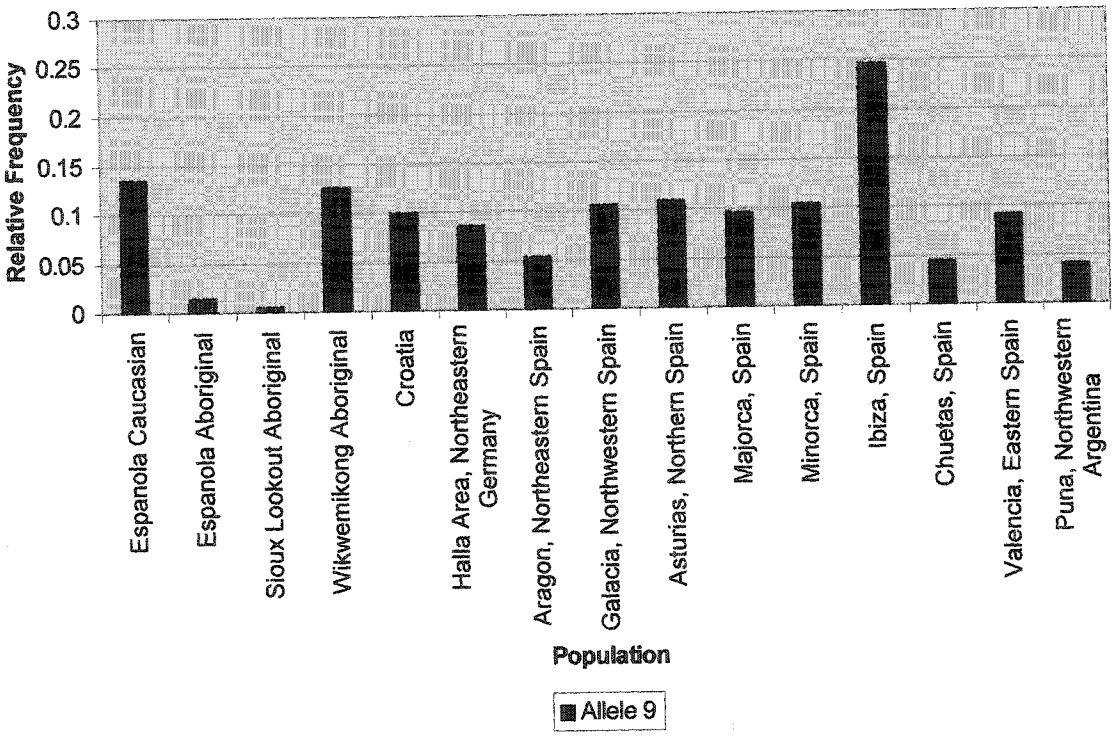


Figure 12. Distribution of the D18S535 allele 9 across Caucasian and Aboriginal populations from Northern Ontario and previously published D18S535 studies in populations from Europe and South America

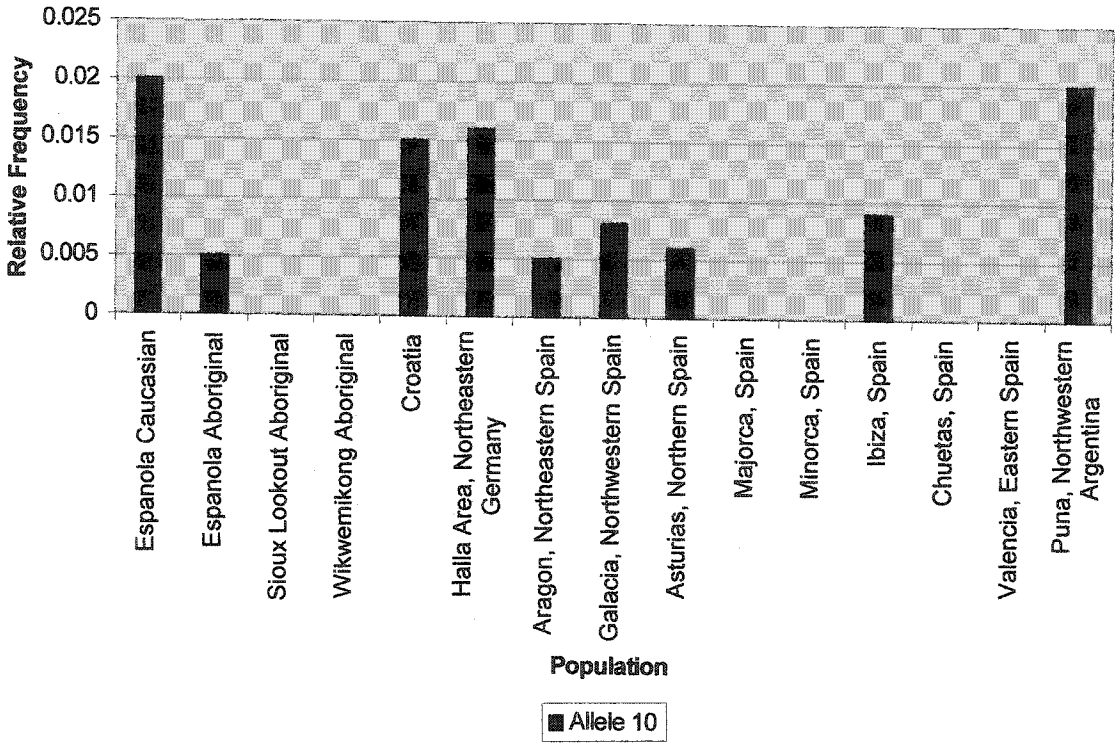


Figure 13. Distribution of the D18S535 allele 10 across Caucasian and Aboriginal populations from Northern Ontario and previously published D18S535 studies in populations from Europe and South America

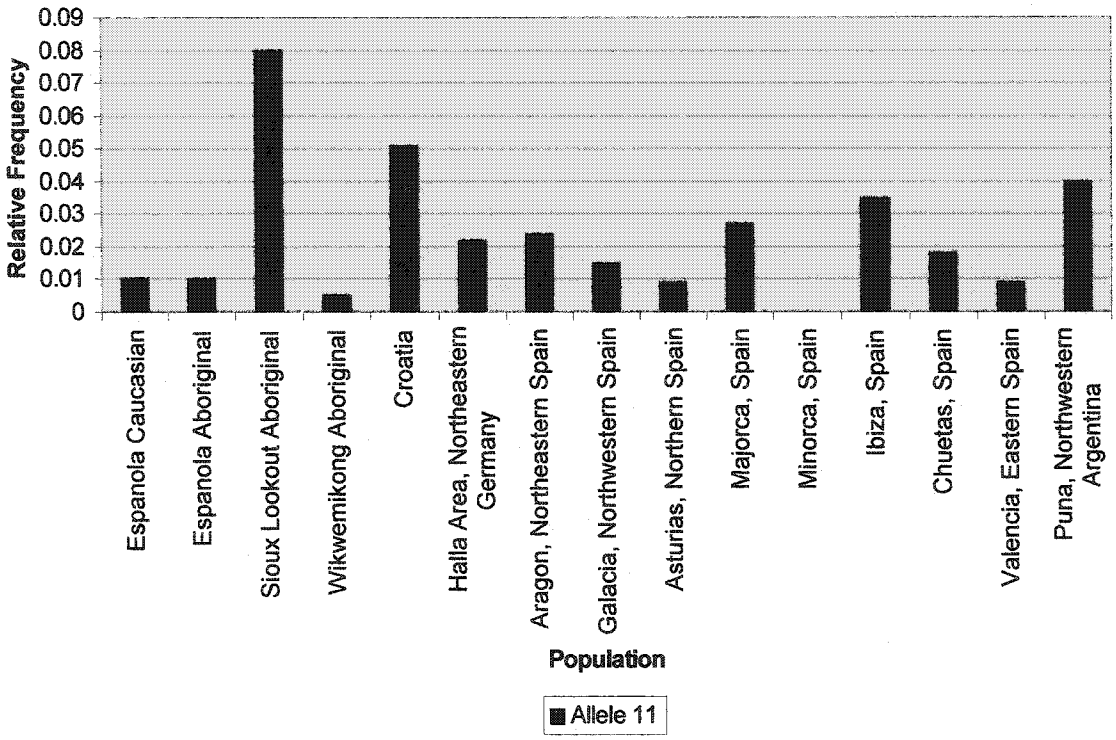


Figure 14. Distribution of the D18S535 allele 11 across Caucasian and Aboriginal populations from Northern Ontario and previously published D18S535 studies in populations from Europe and South America

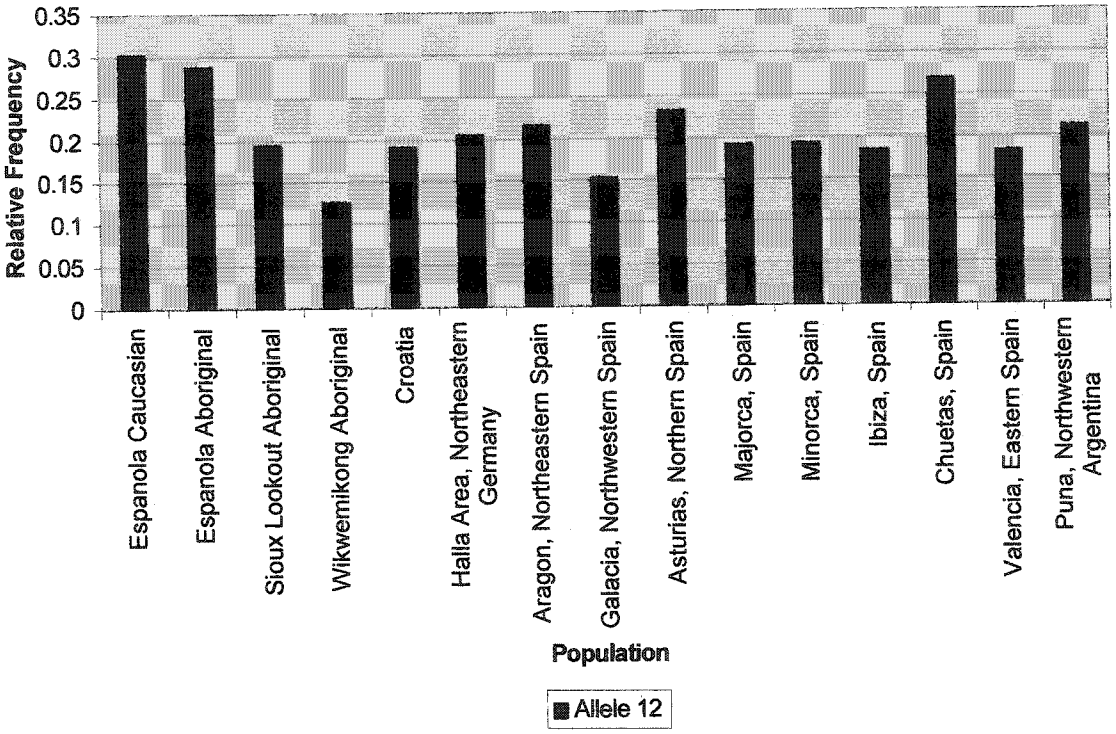


Figure 15. Distribution of the D18S535 allele 12 across Caucasian and Aboriginal populations from Northern Ontario and previously published D18S535 studies in populations from Europe and South America

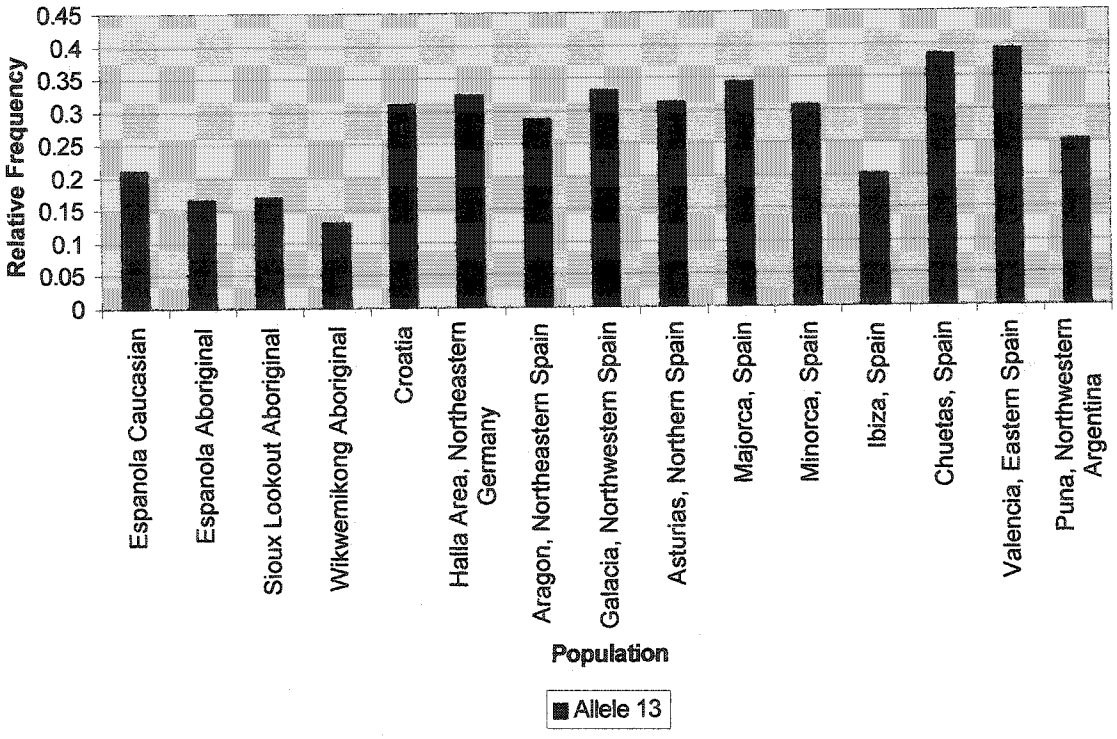


Figure 16. Distribution of the D18S535 allele 13 across Caucasian and Aboriginal populations from Northern Ontario and previously published D18S535 studies in populations from Europe and South America

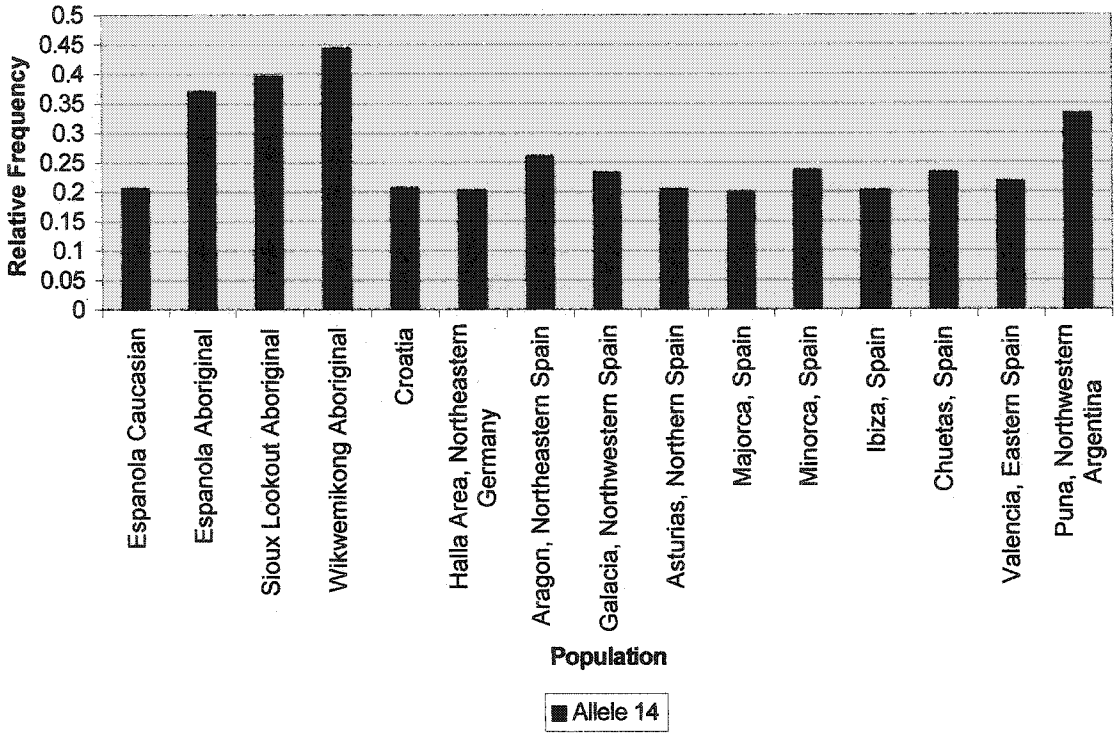


Figure 17. Distribution of the D18S535 allele 14 across Caucasian and Aboriginal populations from Northern Ontario and previously published D18S535 studies in populations from Europe and South America

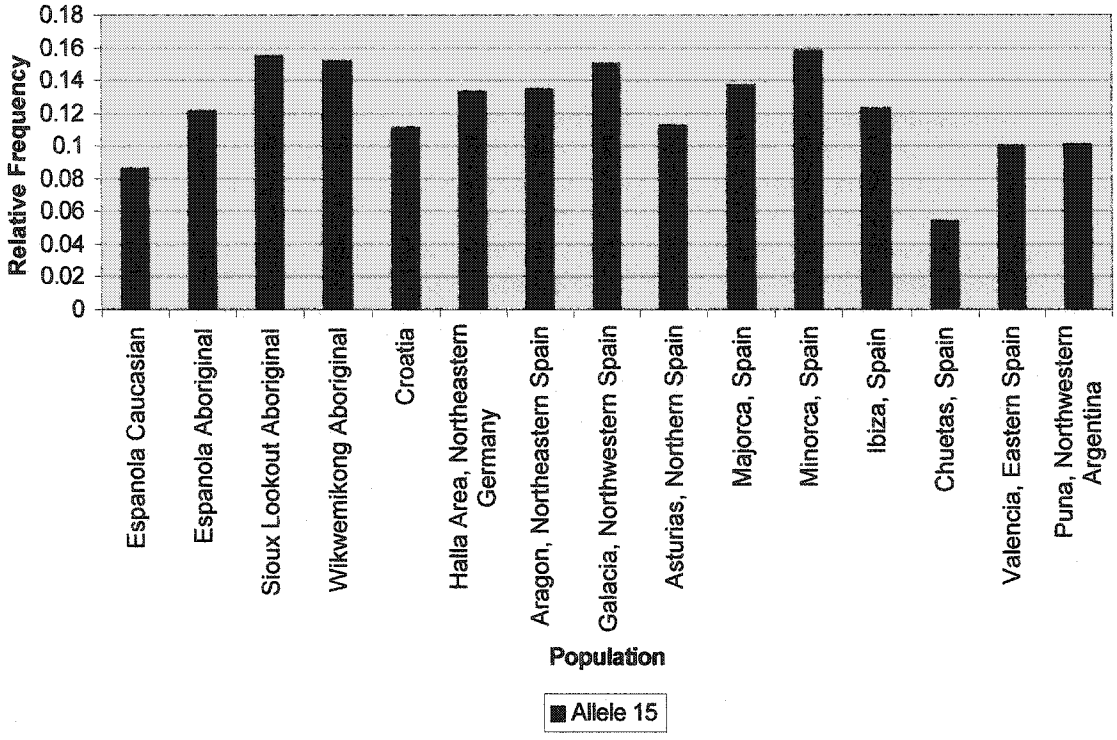


Figure 18. Distribution of the D18S535 allele 15 across Caucasian and Aboriginal populations from Northern Ontario and previously published D18S535 studies in populations from Europe and South America

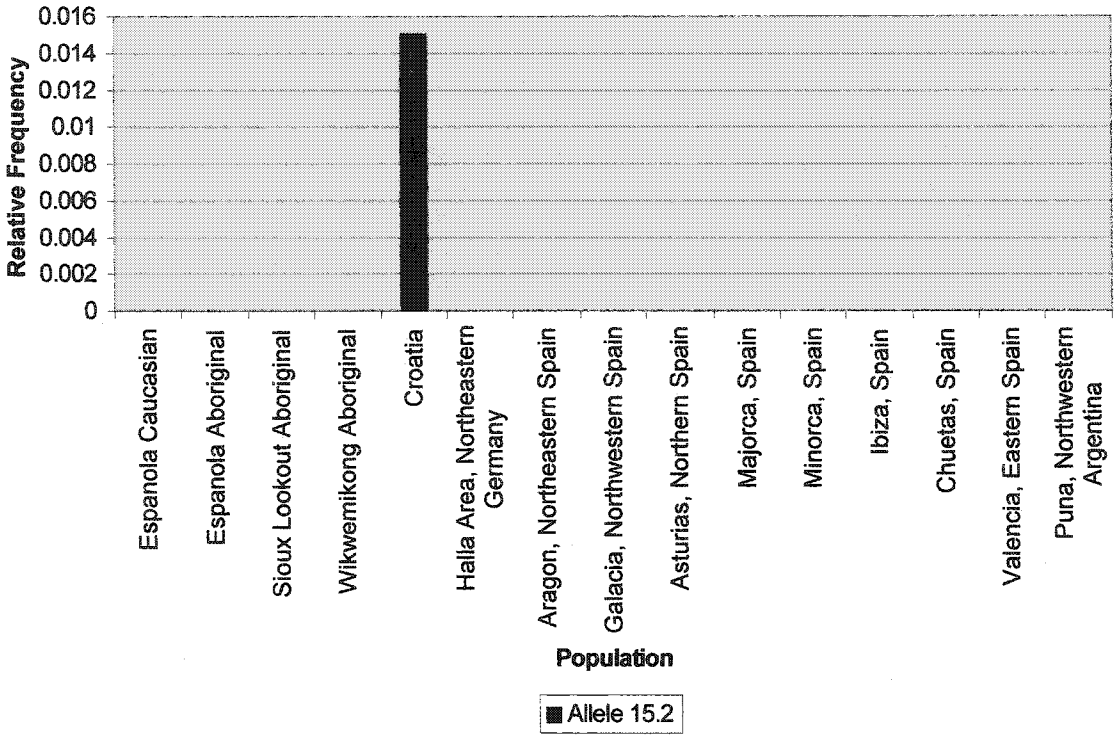


Figure 19. Distribution of the D18S535 allele 15.2 across Caucasian and Aboriginal populations from Northern Ontario and previously published D18S535 studies in populations from Europe and South America

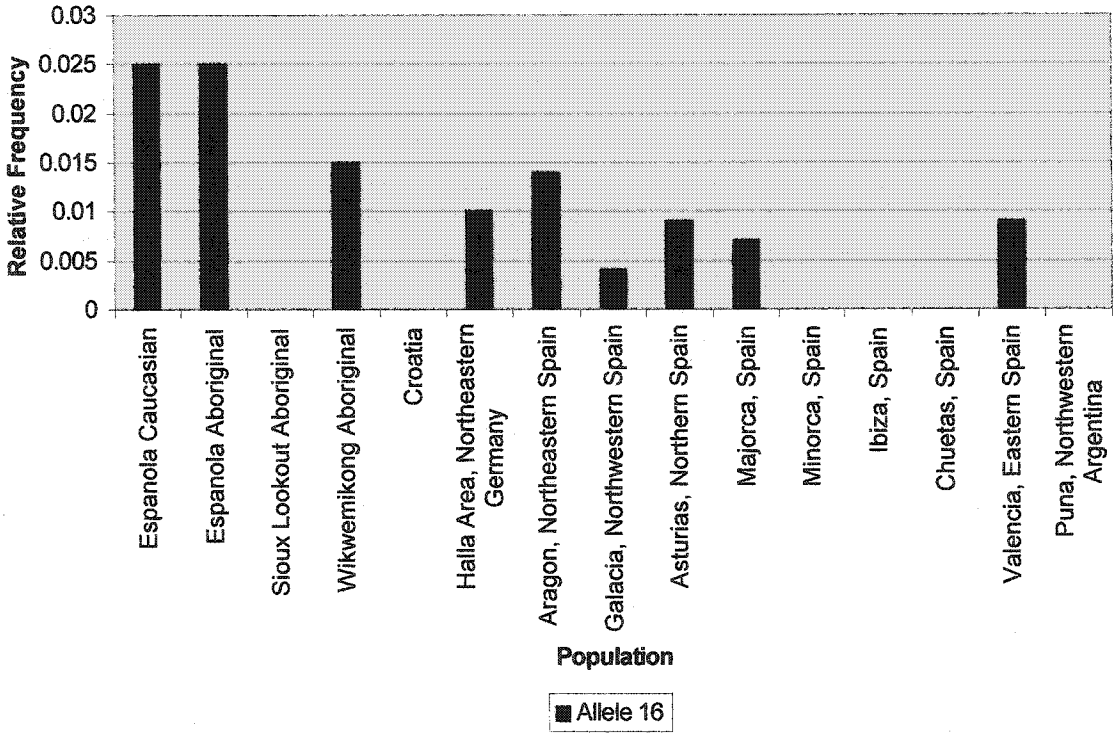


Figure 20. Distribution of the D18S535 allele 16 across Caucasian and Aboriginal populations from Northern Ontario and previously published D18S535 studies in populations from Europe and South America

distribution.

A MANOVA was employed in order to determine if there was an overall homogeneity of variance in the D18S535 allelic distribution in the Northern Ontario populations along with previously published D18S535 studies from Europe and South America. The results indicated that there was an overall significant main effect in the allelic distribution of the populations ($F = 296.51, p < 0.01$, $eta^2 = 9.0$ %). A two-way interaction between alleles by population was also statistically significant ($F = 4.75, p < 0.01$, $eta^2 = 2.2$ %).

A 2 x 15 contingency table and Pearson chi-square test were used as a *post hoc* analysis to determine which D18S535 alleles differed in their distribution across the Northern Ontario populations and all other populations studied to date. The results are shown in Table 5 and indicated that alleles 9, 11, 12, 13, 14, and 16 were significantly different ($p < 0.05$) in their distribution.

The contingency table and Pearson chi-square test was rerun to determine if there was an overall homogeneity of variance in the D18S535 allelic distribution in the Northern Ontario populations along with previously published D18S535 studies from Europe and South America after populations with extreme values at each allele were removed. The outliers were removed to discern if they were the cause of the significant and non-significant differences observed. The results are shown in Table 6 and indicated that after removing the Espanola and Sioux Lookout Aboriginals, allele 9 remained significantly different ($p < 0.05$) in its distribution. After removing Sioux Lookout Aboriginals, and the populations from Croatia and Puna, allele 11 was no longer significantly different ($p > 0.05$) in its distribution.

Table 5. Comparison of the relative frequency distribution of each D18S535 allele in Caucasian and Aboriginal populations from Northern Ontario and with previously published D18S535 studies from Europe and South America

Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
9	86.92	14	0.0001*
10	18.89	14	0.1693
11	53.42	14	0.0001*
12	37.92	14	0.0005*
13	80.59	14	0.0001*
14	97.42	14	0.0001*
15	16.09	14	0.3079
16	24.79	14	0.0367*

note: no data for allele 15.2 could be computed due to only one of the 15 populations displaying this allele

* indicates a significant value ($p < 0.05$)

Table 6. Comparison of the relative frequency distribution of each D18S535 allele in Caucasian and Aboriginal populations from Northern Ontario and with previously published D18S535 studies from Europe and South America after removing extreme values at each allele

Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
9	20.96	11	0.0338*
10	12.81	11	0.3062
12	17.16	12	0.1435
13	17.76	9	0.0381*
14	4.54	10	0.9199
15	10.44	13	0.6576

note: no data for allele 15.2 could be computed due to only one of the 15 populations displaying this allele

* indicates a significant value ($p < 0.05$)

Allele 12 also was no longer significantly different ($p > 0.05$) in its distribution after removing Espanola Caucasians and Espanola Aboriginals.

When all three Aboriginal populations from Northern Ontario, and the populations from Chuetas and Valencia were removed from the analysis, allele 13 remained significantly different ($p < 0.05$) in its distribution. Allele 14 was no longer significantly different ($p > 0.05$) in its distribution after removing the three Northern Ontario Aboriginal populations and the Puna population. After removing the Chuetas population, allele 15 remained non-significant ($p < 0.05$) in its distribution. Alleles 10 and 16 contained no outliers and were not rerun in the analysis.

A MANOVA was employed in order to determine if there was an overall homogeneity of variance in the D18S535 allelic distribution in Espanola Caucasians with previously published D18S535 studies from Europe. The results indicated that there was an overall significant main effect in the allelic distribution of the Caucasian populations ($F = 210.12$, $p < 0.01$, $eta^2 = 8.6\ \%$). A two-way interaction between alleles by population was also statistically significant ($F = 1.86$, $p < 0.01$, $eta^2 = 0.8\ \%$).

A 2 x 11 contingency table and Pearson chi-square test were used as a *post hoc* analysis to determine which D18S535 alleles differed in their distribution across the Caucasian populations. The results are shown in Table 7 and indicated that alleles 9,11,12, and 13 were significantly different ($p < 0.05$) in their distribution.

A MANOVA was employed in order to determine if there was an overall homogeneity of variance in the D18S535 allelic distribution in the ten European populations. The results indicated that there was an overall significant main effect in the allelic

Table 7. Comparison of the relative frequency distribution of each D18S535 allele in Espanola Caucasians with previously published D18S535 studies from Europe

Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
9	35.45	10	0.0001*
10	10.72	10	0.3795
11	24.43	10	0.0065*
12	21.90	10	0.0156*
13	22.81	10	0.0115*
14	4.54	10	0.9199
15	11.96	10	0.2877
16	15.05	10	0.1304

note: no data for allele 15.2 could be computed due to only one of the 11 populations displaying this allele

* indicates a significant value ($p < 0.05$)

distribution of the European populations ($F = 194.98$, $p < 0.01$, $eta^2 = 8.8$ %). A two-way interaction between alleles by population was also statistically significant ($F = 6.60$, $p < 0.02$, $eta^2 = 0.6$ %).

A 2 x 10 contingency table and Pearson chi-square test were used as a *post hoc* analysis to determine which D18S535 alleles differed in their distribution across the European populations. The results are shown in Table 8 and indicated that alleles 9 and 11 were significantly different ($p < 0.05$) in their distribution.

A MANOVA was employed in order to determine if there was an overall homogeneity of variance in the D18S535 allelic distribution in the Northern Ontario and South American Aboriginal populations. The results indicated that there was an overall significant main effect in the allelic distribution of the four Aboriginal populations ($F = 125.00$, $p < 0.01$, $eta^2 = 13.7$ %). A two-way interaction between alleles by population was also statistically significant ($F = 3.89$, $p < 0.01$, $eta^2 = 1.5$ %).

A 2 x 4 contingency table and Pearson chi-square test were used as a *post hoc* analysis to determine which D18S535 alleles differed in their distribution across the Aboriginal populations. The results are shown in Table 9 and indicated that alleles 9, 10, 11, 12, 13, and 16 were significantly different ($p < 0.05$) in their distribution.

A MANOVA was employed in order to determine if there was an overall homogeneity of variance in the D18S535 allelic distribution in the four Northern Ontario populations. The results indicated that there was an overall significant main effect in the allelic distribution of the four Northern Ontario populations ($F = 108.82$, $p < 0.01$, $eta^2 = 12.1$ %). A two-way interaction between alleles by population was also statistically significant

Table 8. Comparison of the relative frequency distribution of each D18S535 allele in previously published D18S535 studies from Europe

Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
9	33.98	9	0.0001*
10	9.29	9	0.4114
11	21.79	9	0.0096*
12	10.36	9	0.3224
13	13.13	9	0.1568
14	4.43	9	0.8810
15	9.37	9	0.4036
16	9.68	9	0.3766

note: no data for allele 15.2 could be computed due to only one of the 10 populations displaying this allele

* indicates a significant value ($p < 0.05$)

Table 9. Comparison of the relative frequency distribution of each D18S535 allele in Northern Ontario and South American Aboriginal populations

Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
9	40.53	3	0.0001*
10	8.79	3	0.0323*
11	21.60	3	0.0001*
12	16.05	3	0.0011*
13	11.02	3	0.0116*
14	6.09	3	0.1071
15	3.03	3	0.3472
16	9.09	3	0.0281*

* indicates a significant value ($p < 0.05$)

($F = 6.83, p < 0.01, eta^2 = 2.5\ \%$).

A 2 x 4 contingency table and Pearson chi-square test were used as a *post hoc* analysis to determine which D18S535 alleles differed in their distribution across the Northern Ontario populations. The results are shown in Table 10 and indicated that alleles 9, 10, 11, 12, and 14 were significantly different ($p < 0.05$) in their distribution.

A MANOVA was employed in order to determine if there was an overall homogeneity of variance in the D18S535 allelic distribution in the three Northern Ontario Aboriginal populations. The results indicated that there was an overall significant main effect in the allelic distribution of the three Northern Ontario Aboriginal populations ($F = 101.73, p < 0.01, eta^2 = 14.6\ \%$). A two-way interaction between alleles by population was also statistically significant ($F = 4.09, p < 0.01, eta^2 = 1.4\ \%$).

A 2 x 3 contingency table and Pearson chi-square test were used as a *post hoc* analysis to determine which D18S535 alleles differed in their distribution across the Northern Ontario Aboriginal populations. The results are shown in Table 11 and indicated that alleles 9, 11, and 12 were significantly different ($p < 0.05$) in their distribution.

To evaluate the differences between the pooled Espanola and Wikwemikong Aboriginals and the Sioux Lookout Aboriginals, the four Northern Ontario populations, and the four Northern Ontario populations with previously studied populations from Europe and South America, a series of pairwise comparisons were made. The results of the pairwise comparisons are displayed in Table 12. The allele frequency distribution for D18S535 did show statistical significant differences ($p < 0.05$) between the pooled Espanola and Wikwemikong Aboriginals and the Sioux Lookout Aboriginals, and all four Northern

Table 10. Comparison of the relative frequency distribution of each D18S535 allele in Caucasian and Aboriginal populations from Northern Ontario

Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
9	44.83	3	0.0001*
10	8.69	3	0.0337*
11	29.91	3	0.0001*
12	23.25	3	0.0001*
13	4.60	3	0.2031
14	27.43	3	0.0001*
15	5.50	3	0.1386
16	5.29	3	0.1520

* indicates a significant value ($p < 0.05$)

Table 11. Comparison of the relative frequency distribution of each D18S535 allele in the three Aboriginal populations from Northern Ontario

Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
9	38.79	2	0.0001*
10	2.01	2	0.3654
11	22.67	2	0.0001*
12	16.10	2	0.0003*
13	1.38	2	0.5003
14	2.44	2	0.2958
15	1.12	2	0.5710
16	4.86	2	0.0881

* indicates a significant value ($p < 0.05$)

Table 12. Pairwise comparisons of the relative D18S535 allele frequency distributions between Caucasian and Aboriginal populations from Northern Ontario and with previously published D18S535 studies from Europe and South America

Populations	Probability	Standard Error	Populations	Probability	Standard Error
EC & EA	0.0001*	0.0001	EA & CR	0.0001*	0.0001
EC & SLA	0.0001*	0.0001	EA & NEG	0.0001*	0.0001
EC & WA	0.0001*	0.0001	EA & AR	0.0029*	0.0006
EC & CR	0.0001*	0.0006	EA & GA	0.0001*	0.0001
EC & NEG	0.0124*	0.0012	EA & AS	0.0001*	0.0001
EC & AR	0.0055*	0.0007	EA & MA	0.0001*	0.0001
EC & GA	0.0002*	0.0001	EA & MI	0.0001*	0.0001
EC & AS	0.0393*	0.0024	EA & IB	0.0001*	0.0001
EC & MA	0.0140*	0.0016	EA & CH	0.0001*	0.0001
EC & MI	0.0165*	0.0016	EA & VA	0.0001*	0.0001
EC & IB	0.0212*	0.0015	EA & PU	0.0057*	0.0007
EC & CH	0.0043*	0.0005	SLA & WA	0.0001*	0.0001
EC & VA	0.0146*	0.0012	SLA & CR	0.0001*	0.0001
EC & PU	0.0002*	0.0001	SLA & NEG	0.0001*	0.0001
EA & SLA	0.0008*	0.0003	SLA & AR	0.0001*	0.0001
EA & WA	0.0001*	0.0001	SLA & GA	0.0001*	0.0001

Table 12. (Continued)

Populations	Probability	Standard Error	Populations	Probability	Standard Error
SLA & AS	0.0001*	0.0001	WA & GA	0.0001*	0.0001
SLA & MA	0.0001*	0.0001	WA & AS	0.0001*	0.0001
SLA & MI	0.0001*	0.0001	WA & MA	0.0001*	0.0001
SLA & IB	0.0001*	0.0001	WA & MI	0.0001*	0.0001
SLA & CH	0.0001*	0.0001	WA & IB	0.0001*	0.0001
SLA & VA	0.0001*	0.0001	WA & CH	0.0001*	0.0001
SLA & PU	0.0037*	0.0008	WA & VA	0.0001*	0.0001
WA & CR	0.0001*	0.0001	WA & PU	0.0001*	0.0001
WA & NEG	0.0001*	0.0001	EA + WA & SLA	0.0001*	0.0001
WA & AR	0.0001*	0.0001			

EC = Espanola Caucasians; EA = Espanola Aboriginals; SLA = Sioux Lookout Aboriginals; WA = Wikwemikong Aboriginals; CR = Croatia; NEG = Halla Area , Northeastern Germany; AR = Aragon, Northeastern Spain; GA = Galacia, Northwestern Spain; AS = Asturias, Northern Spain; MA = Majorca, Spain; MI = Minorca, Spain; IB = Ibiza, Spain; CH = Chuetas, Spain; VA = Valencia, Eastern Spain; PU = Puna, Northwestern Argentina; EA + WA= pooled Espanola and Wikwemikong Aboriginals

* indicates a significant value ($p < 0.05$)

Ontario populations. The four Northern Ontario populations also showed significant differences ($p < 0.05$) between all previously studied populations.

To evaluate the differences between the previously studied populations from Europe and South America, a series of pairwise comparisons were made. The results of the pairwise comparisons are displayed in Table 13. All the populations were significantly different ($p < 0.05$) from the Ibiza population and all but the population from Aragon were significantly different ($p < 0.05$) from the Puna population. The Ibiza and the Puna population were significantly different ($p < 0.05$) from each other. The Croatia population was also significantly different ($p < 0.05$) from the population from the Halla Area in Northeastern Germany, the Galacia population, the Asturias population, and the Chuetas population. The Galacia and Minorca populations were also significantly different ($p < 0.05$) from the Chuetas population.

Two by two contingency tables and Pearson chi-square test were used to identify the alleles which were the source of the pairwise population differences between the pooled Espanola and Wikwemikong Aboriginals and the Sioux Lookout Aboriginals, the Caucasian and Aboriginal populations from Northern Ontario, and the Caucasian and Aboriginal populations from Northern Ontario with previously published D18S535 studies from Europe and South America. The results are shown in Table 14. Figures 21 to 27 display the allelic distributions between the pooled Espanola and Wikwemikong Aboriginals and the Sioux Lookout Aboriginals, and between the Northern Ontario populations which differed from each other. Figures 28 to 71 display the allelic distributions between the previously published D18S535 studies from Europe and South America populations which differed

Table 13. Pairwise comparisons of the relative D18S535 allele frequency distributions between previously published D18S535 studies from Europe and South America

Populations	Probability	Standard Error	Populations	Probability	Standard Error
CR & NEG	0.0414*	0.0035	NEG & CH	0.1323	0.0047
CR & AR	0.0668	0.0034	NEG & VA	0.8905	0.0030
CR & GA	0.0273*	0.0028	NEG & PU	0.0074*	0.0010
CR & AS	0.0010*	0.0002	AR & GA	0.3151	0.0066
CR & MA	0.1974	0.0061	AR & AS	0.1006	0.0046
CR & MI	0.0677	0.0031	AR & MA	0.6732	0.0055
CR & IB	0.0152*	0.0012	AR & MI	0.3666	0.0064
CR & CH	0.0317*	0.0020	AR & IB	0.0006*	0.0002
CR & VA	0.1288	0.0045	AR & CH	0.16110	0.0050
CR & PU	0.0086*	0.0012	AR & VA	0.5021	0.0065
NEG & AR	0.7103	0.0064	AR & PU	0.5274	0.0066
NEG & GA	0.7160	0.0054	GA & AS	0.3196	0.0076
NEG & AS	0.5320	0.0071	GA & MA	0.8793	0.0033
NEG & MA	0.9573	0.0016	GA & MI	0.8914	0.0032
NEG & MI	0.5938	0.0067	GA & IB	0.0120*	0.0013
NEG & IB	0.0024*	0.0006	GA & CH	0.0185*	0.0015

Table 13. (Continued)

Populations	Probability	Standard Error	Populations	Probability	Standard Error
GA & VA	0.8040	0.0044	MA & PU	0.0118*	0.0012
GA & PU	0.0021*	0.0005	MI & IB	0.0141*	0.0012
AS & MA	0.4888	0.0077	MI & CH	0.0320*	0.0017
AS & MI	0.8103	0.0048	MI & VA	0.4344	0.0054
AS & IB	0.0039*	0.0007	MI & PU	0.0108*	0.0010
AS & CH	0.1313	0.0047	IB & CH	0.0001*	0.0001
AS & VA	0.4904	0.0069	IB & VA	0.0039*	0.0005
AS & PU	0.0001*	0.0001	IB & PU	0.0001*	0.0001
MA & MI	0.5667	0.0054	CH & VA	0.3278	0.0057
MA & IB	0.0133*	0.0015	CH & PU	0.0492*	0.0025
MA & CH	0.1302	0.0041	VA & PU	0.0128*	0.0012
MA & VA	0.9719	0.0009			

CR = Croatia; NEG = Halla Area , Northeastern Germany; AR = Aragon, Northeastern Spain; GA = Galacia, Northwestern Spain; AS = Asturias, Northern Spain; MA = Majorca, Spain; MI = Minorca, Spain; IB = Ibiza, Spain; CH = Chuetas, Spain; VA = Valencia, Eastern Spain; PU = Puna, Northwestern Argentina

* indicates a significant value ($p < 0.05$)

Table 14. Allelic differences observed in pairwise comparisons between Caucasian and Aboriginal populations from Northern Ontario and with previously published D18S535 studies from Europe and South America which demonstrated significant differences

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
EC & EA	9	20.77	1	0.0001*
	10	1.82	1	0.1770
	11	0.00	1	1.0000
	12	0.11	1	0.7411
	13	1.33	1	0.2484
	14	12.61	1	0.0004*
	15	1.33	1	0.2483
	16	0.00	1	1.0000
EC & SLA	9	26.25	1	0.0001*
	10	4.08	1	0.0434*
	11	11.25	1	0.0008*
	12	6.21	1	0.0127*
	13	1.14	1	0.2851
	14	16.69	1	0.0001*
	15	4.48	1	0.0342*
	16	5.11	1	0.0237*
EC & WA	9	0.09	1	0.7660
	10	4.04	1	0.0444*
	11	0.34	1	0.5622
	12	18.35	1	0.0001*
	13	4.55	1	0.0330*
	14	25.40	1	0.0001*
	15	4.08	1	0.0434*
	16	0.51	1	0.4750

Table 14. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
EC & CR	9	1.64	1	0.1995
	10	0.04	1	0.8324
	11	6.41	1	0.0113*
	12	9.30	1	0.0023*
	13	5.56	1	0.0183*
	14	0.01	1	0.9738
	15	0.92	1	0.3388
	16	8.67	1	0.0032*
EC & NEG	9	3.10	1	0.0784
	10	0.36	1	0.5507
	11	0.74	1	0.3898
	12	5.99	1	0.0144*
	13	7.74	1	0.0054*
	14	0.37	1	0.8477
	15	2.65	1	0.1034
	16	0.95	1	0.3285
EC & AR	9	7.23	1	0.0072*
	10	0.78	1	0.3769
	11	1.85	1	0.1743
	12	4.46	1	0.0348*
	13	3.14	1	0.0762
	14	1.56	1	0.2113
	15	2.45	1	0.1178
	16	1.46	1	0.2263

Table 14. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
EC & GA	9	0.87	1	0.3497
	10	0.14	1	0.2435
	11	0.24	1	0.6223
	12	14.66	1	0.0001*
	13	7.86	1	0.0051*
	14	0.37	1	0.5445
	15	3.87	1	0.0492*
	16	2.30	1	0.1291
EC & AS	9	0.57	1	0.4503
	10	2.03	1	0.1542
	11	0.22	1	0.6358
	12	2.99	1	0.0836
	13	6.63	1	0.0100*
	14	0.02	1	0.9012
	15	1.04	1	0.3086
	16	3.25	1	0.0716
EC & MA	9	1.41	1	0.2343
	10	3.02	1	0.0820
	11	1.42	1	0.2328
	12	5.79	1	0.0161*
	13	6.86	1	0.0088*
	14	0.01	1	0.9207
	15	2.15	1	0.1423
	16	0.59	1	0.4429

Table 14. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
EC & MI	9	0.73	1	0.3934
	10	2.37	1	0.1234
	11	1.18	1	0.2775
	12	4.87	1	0.0273*
	13	3.78	1	0.0519
	14	0.50	1	0.4786
	15	3.55	1	0.0596
	16	2.98	1	0.0844
EC & IB	9	5.58	1	0.0181*
	10	0.03	1	0.8533
	11	2.32	1	0.1277
	12	6.57	1	0.0104*
	13	0.01	1	0.9127
	14	0.01	1	0.9971
	15	1.00	1	0.3180
	16	2.98	1	0.0845
EC & CH	9	5.36	1	0.0206*
	10	2.33	1	0.1267
	11	0.32	1	0.5736
	12	0.56	1	0.4541
	13	10.95	1	0.0009*
	14	0.19	1	0.6636
	15	1.17	1	0.2795
	16	2.93	1	0.0872

Table 14. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
EC & VA	9	1.64	1	0.2006
	10	2.33	1	0.1267
	11	0.32	1	0.5736
	12	6.18	1	0.0129*
	13	10.95	1	0.0009*
	14	0.01	1	0.9423
	15	0.32	1	0.5698
	16	0.20	1	0.6580
EC & PU	9	11.11	1	0.0009*
	10	0.01	1	0.9884
	11	3.75	1	0.0526
	12	14.04	1	0.0444*
	13	1.02	1	0.3133
	14	7.19	1	0.0073*
	15	.30	1	0.5819
	16	5.01	1	0.0252*
EA & SLA	9	1.03	1	0.3100
	10	1.01	1	0.3143
	11	11.26	1	0.0008*
	12	4.69	1	0.0304*
	13	0.01	1	0.9292
	14	0.29	1	0.5890
	15	0.95	1	0.3288
	16	5.11	1	0.0237

Table 14. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
EA & WA	9	18.60	1	0.0001*
	10	1.00	1	0.3167
	11	0.34	1	0.5622
	12	15.75	1	0.0001*
	13	0.98	1	0.3232
	14	2.32	1	0.1249
	15	0.77	1	0.3796
	16	0.51	1	0.4750
EA & CR	9	14.07	1	0.0002*
	10	1.55	1	0.2137
	11	6.42	1	0.0113*
	12	7.13	1	0.0076*
	13	12.78	1	0.0004*
	14	17.00	1	0.0001*
	15	0.11	1	0.7407
	16	8.67	1	0.0032*
EA & NEG	9	11.12	1	0.0009*
	10	0.82	1	0.3642
	11	0.74	1	0.3898
	12	4.33	1	0.0375*
	13	15.75	1	0.0001*
	14	17.34	1	0.0001*
	15	0.16	1	0.6924
	16	0.95	1	0.3285

Table 14. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
EA & AR	9	5.16	1	0.0231*
	10	0.29	1	0.5914
	11	1.85	1	0.1743
	12	3.16	1	0.0753
	13	8.52	1	0.0035*
	14	5.61	1	0.0178*
	15	0.16	1	0.6862
	16	1.46	1	0.2263
EA & GA	9	15.25	1	0.0001*
	10	0.12	1	0.7284
	11	0.24	1	0.6223
	12	12.10	1	0.0005*
	13	15.74	1	0.0001*
	14	10.38	1	0.0013*
	15	0.60	1	0.4396
	16	2.30	1	0.1291
EA & AS	9	16.94	1	0.0001*
	10	0.03	1	0.8531
	11	0.22	1	0.6358
	12	1.84	1	0.1740
	13	14.24	1	0.0002*
	14	17.19	1	0.0001*
	15	0.06	1	0.8023
	16	3.24	1	0.0716

Table 14. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
EA & MA	9	11.44	1	0.0007*
	10	0.75	1	0.3866
	11	1.42	1	0.2328
	12	4.45	1	0.0349*
	13	13.61	1	0.0002*
	14	11.16	1	0.0008*
	15	0.15	1	0.7005
	16	0.59	1	0.4429
EA & MI	9	12.54	1	0.0004*
	10	0.59	1	0.4433
	11	1.18	1	0.2775
	12	3.75	1	0.0529
	13	8.81	1	0.0030*
	14	5.43	1	0.0198*
	15	0.73	1	0.3935
	16	2.98	1	0.0845
EA & IB	9	42.07	1	0.0001*
	10	1.15	1	0.2838
	11	2.32	1	0.1277
	12	5.27	1	0.0217*
	13	0.80	1	0.3721
	14	8.97	1	0.0028*
	15	0.01	1	0.9891
	16	2.98	1	0.0844

Table 14. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
EA & CH	9	3.63	1	0.0568
	10	0.58	1	0.4473
	11	0.32	1	0.5736
	12	0.22	1	0.6391
	13	18.72	1	0.0001*
	14	6.60	1	0.0102*
	15	3.92	1	0.0479*
	16	2.93	1	0.0872
EA & VA	9	9.54	1	0.0020*
	10	0.58	1	0.4472
	11	0.32	1	0.5736
	12	4.92	1	0.0265
	13	18.72	1	0.0001*
	14	8.45	1	0.0037*
	15	0.18	1	0.6711
	16	0.20	1	0.6580
EA & PU	9	2.39	1	0.1220
	10	1.85	1	0.1733
	11	3.76	1	0.0526
	12	2.84	1	0.0922
	13	4.63	1	0.0314*
	14	0.77	1	0.3797
	15	0.36	1	0.5458
	16	5.01	1	0.0252*

Table 14. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
SLA & WA	9	23.96	1	0.0001*
	10	--	--	--
	11	13.67	1	0.0002*
	12	3.48	1	0.0620
	13	1.16	1	0.2808
	14	1.00	1	0.3176
	15	0.01	1	0.9231
	16	3.05	1	0.0806
SLA & CR	9	18.75	1	0.0001*
	10	3.57	1	0.0589
	11	1.56	1	0.2113
	12	0.04	1	0.8468
	13	12.22	1	0.0005*
	14	22.54	1	0.0001*
	15	2.11	1	0.1461
	16	--	--	--
SLA & NEG	9	15.67	1	0.0001*
	10	2.69	1	0.1011
	11	10.27	1	0.0014*
	12	0.10	1	0.7502
	13	15.16	1	0.0001*
	14	22.73	1	0.0001*
	15	0.46	1	0.4965
	16	2.69	1	0.1011

Table 14. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
SLA & AR	9	9.18	1	0.0025*
	10	1.93	1	0.1645
	11	5.23	1	0.0222*
	12	0.17	1	0.6783
	13	8.07	1	0.0045*
	14	8.51	1	0.0035*
	15	0.34	1	0.5584
	16	1.93	1	0.1645
SLA & GA	9	20.18	1	0.0001*
	10	1.55	1	0.2139
	11	11.35	1	0.0008*
	12	1.35	1	0.2460
	13	15.15	1	0.0075*
	14	14.46	1	0.0001*
	15	0.07	1	0.7922
	16	1.55	1	0.2139
SLA & AS	9	21.83	1	0.0001*
	10	1.27	1	0.2596
	11	19.74	1	0.0001*
	12	1.10	1	0.2945
	13	13.66	1	0.0002*
	14	22.65	1	0.0001*
	15	1.83	1	0.1762
	16	1.27	1	0.2596

Table 14. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
SLA & MA	9	16.55	1	0.0001*
	10	--	--	--
	11	4.41	1	0.0358
	12	0.02	1	0.8919
	13	13.08	1	0.0003*
	14	14.62	1	0.0001*
	15	0.27	1	0.6044
	16	2.71	1	0.0992
SLA & MI	9	18.04	1	0.0001*
	10	--	--	--
	11	9.77	1	0.0018*
	12	0.01	1	0.9076
	13	8.39	1	0.0038*
	14	7.74	1	0.0054*
	15	0.01	1	0.9967
	16	--	--	--
SLA & IB	9	49.22	1	0.0001*
	10	3.47	1	0.0625
	11	2.56	1	0.1092
	12	0.25	1	0.6195
	13	0.67	1	0.4141
	14	11.82	1	0.0006*
	15	0.71	1	0.4002
	16	--	--	--

Table 14. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
SLA & CH	9	7.56	1	0.0060*
	10	--	--	--
	11	5.24	1	0.0220*
	12	1.97	1	0.1607
	13	18.14	1	0.0001*
	14	9.09	1	0.0026*
	15	7.32	1	0.0068*
	16	--	--	--
SLA & VA	9	14.70	1	0.0001*
	10	--	--	--
	11	5.24	1	0.0220*
	12	0.18	1	0.6696
	13	18.14	1	0.0001*
	14	11.21	1	0.0008*
	15	1.52	1	0.2177
	16	3.53	1	0.0602
SLA & PU	9	5.71	1	0.0168*
	10	4.12	1	0.0423*
	11	2.67	1	0.1023
	12	0.23	1	0.6343
	13	4.29	1	0.0383
	14	2.01	1	0.1561
	15	2.47	1	0.1157
	16	--	--	--

Table 14. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
WA & CR	9	0.88	1	0.3471
	10	3.53	1	0.0601
	11	8.42	1	0.0037*
	12	3.48	1	0.0621
	13	20.81	1	0.0001*
	14	34.33	1	0.0001*
	15	1.79	1	0.1809
	16	5.18	1	0.0228*
WA & NEG	9	2.03	1	0.1538
	10	2.66	1	0.1028
	11	1.92	1	0.1655
	12	5.35	1	0.0207*
	13	24.34	1	0.0001*
	14	34.12	1	0.0001*
	15	0.33	1	0.5678
	16	0.03	1	0.8661
WA & AR	9	5.76	1	0.0164*
	10	1.91	1	0.1666
	11	3.39	1	0.0656
	12	5.23	1	0.0222*
	13	15.00	1	0.0001*
	14	15.23	1	0.0001*
	15	0.23	1	0.6267
	16	0..26	1	0.6132

Table 14. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
WA & GA	9	0.37	1	0.5383
	10	1.53	1	0.2162
	11	1.11	1	0.2917
	12	0.70	1	0.4020
	13	24.20	1	0.0001*
	14	23.46	1	0.0001*
	15	0.03	1	0.8730
	16	0.58	1	0.4466
WA & AS	9	0.17	1	0.6738
	10	1.26	1	0.2620
	11	0.03	1	0.8531
	12	9.12	1	0.0025*
	13	22.55	1	0.0001*
	14	34.20	1	0.0001*
	15	1.54	1	0.2151
	16	0.98	1	0.3213
WA & MA	9	0.85	1	0.3568
	10	--	--	--
	11	2.87	1	0.0901
	12	2.59	1	0.1079
	13	21.08	1	0.0001*
	14	22.02	1	0.0001*
	15	0.18	1	0.6681
	16	0.02	1	0.8995

Table 14. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
WA & MI	9	0.37	1	0.5450
	10	--	--	--
	11	0.59	1	0.4433
	12	2.31	1	0.1286
	13	14.80	1	0.0001*
	14	12.95	1	0.0003*
	15	0.01	1	0.9307
	16	1.77	1	0.1828
WA & IB	9	6.91	1	0.0086*
	10	3.43	1	0.0638
	11	4.04	1	0.0443*
	12	1.27	1	0.2600
	13	3.12	1	0.0773
	14	17.99	1	0.0001*
	15	0.58	1	0.4476
	16	1.77	1	0.1828
WA & CH	9	4.38	1	0.0363*
	10	--	--	--
	11	1.19	1	0.2762
	12	9.34	1	0.0023*
	13	26.96	1	0.0001*
	14	14.61	1	0.0001*
	15	6.93	1	0.0085*
	16	1.74	1	0.1866

Table 14. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
WA & VA	9	1.08	1	0.2989
	10	--	--	--
	11	1.19	1	0.2762
	12	1.42	1	0.2338
	13	26.96	1	0.0001*
	14	17.20	1	0.0001*
	15	1.33	1	0.2491
	16	0.03	1	0.8713
WA & PU	9	9.37	1	0.0022*
	10	4.08	1	0.0433*
	11	5.64	1	0.0175*
	12	5.41	1	0.0201*
	13	9.69	1	0.0019*
	14	5.78	1	0.0162*
	15	2.18	1	0.1402
	16	2.99	1	0.0837
EA + WA & SLA	9	12.39	1	0.0004*
	10	0.51	1	0.4769
	11	22.58	1	0.0001*
	12	0.12	1	0.7294
	13	0.45	1	0.5045
	14	0.07	1	0.7857
	15	0.38	1	0.5390
	16	4.10	1	0.0430*

Table 14. (Continued)
EC = Espanola Caucasians; EA = Espanola Aboriginals; SLA = Sioux Lookout Aboriginals; WA = Wikwemikong Aboriginals; CR = Croatia; NEG = Halla Area, Northeastern Germany; AR = Aragon, Northeastern Spain; GA = Galacia, Northwestern Spain; AS = Asturias, Northern Spain; MA = Majorca, Spain; MI = Minorca, Spain; IB = Ibiza, Spain; CH = Chuetas, Spain; VA = Valencia, Eastern Spain; PU = Puna, Northwestern Argentina; EA + WA= pooled Espanola and Wikwemikong Aboriginals
* indicates a significant value ($p < 0.05$)
– statistics could not be computed since the number of non-empty rows or columns was one

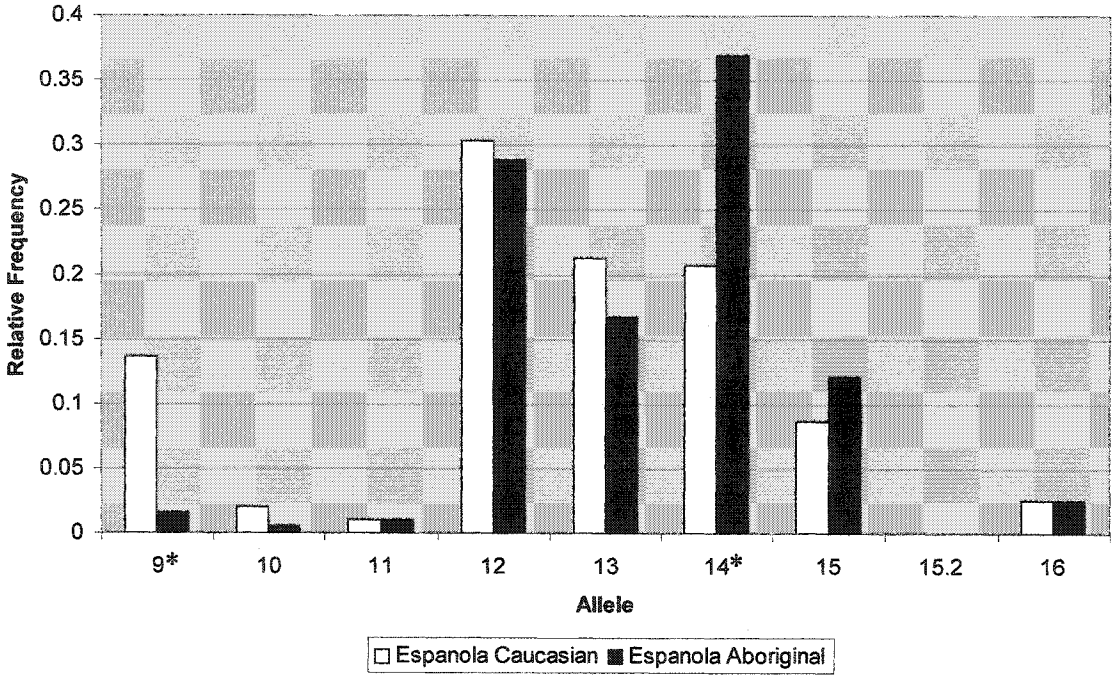


Figure 21. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Caucasian and Espanola Aboriginal populations [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

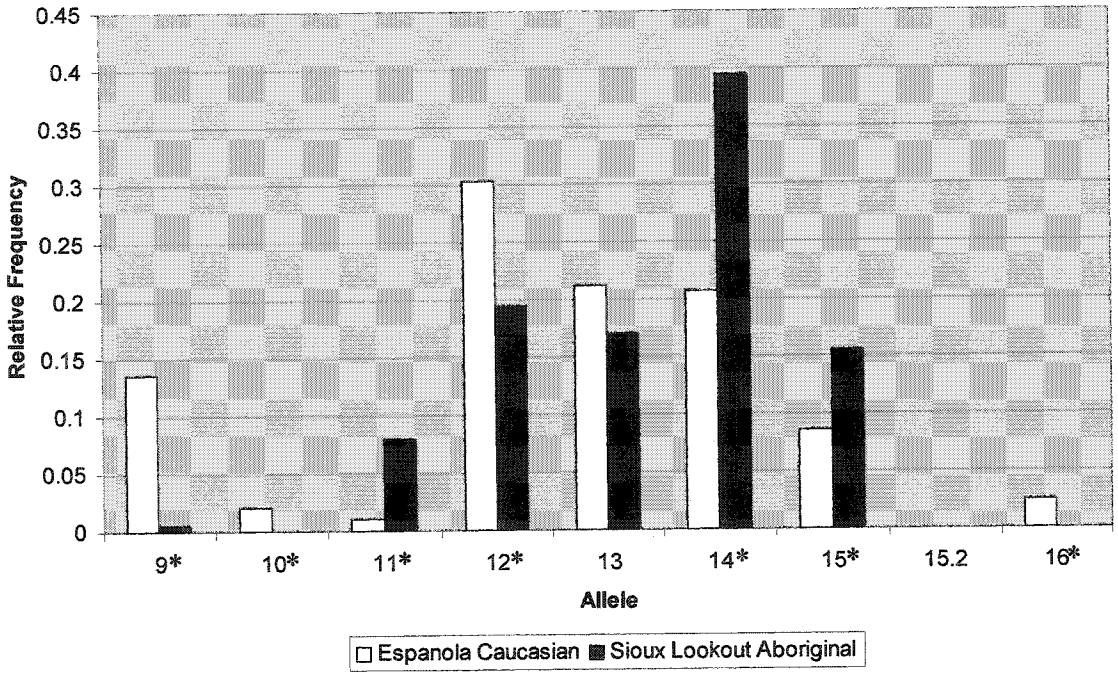


Figure 22. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Caucasian and Sioux Lookout Aboriginal populations [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

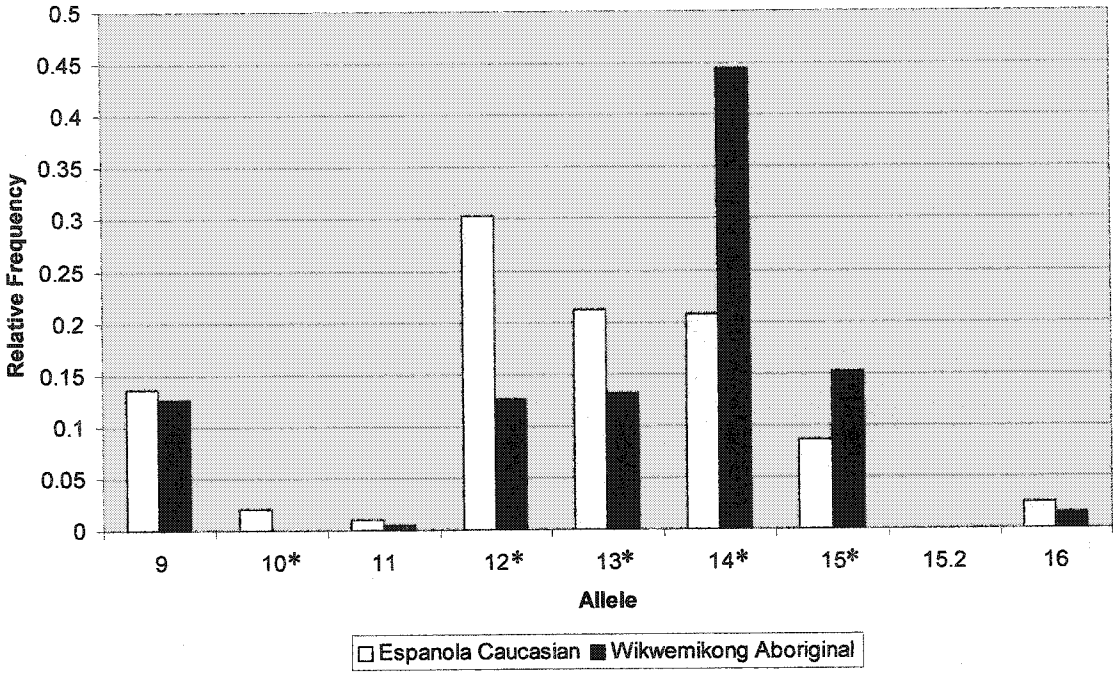


Figure 23. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Caucasian and Wikwemikong Aboriginal populations [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

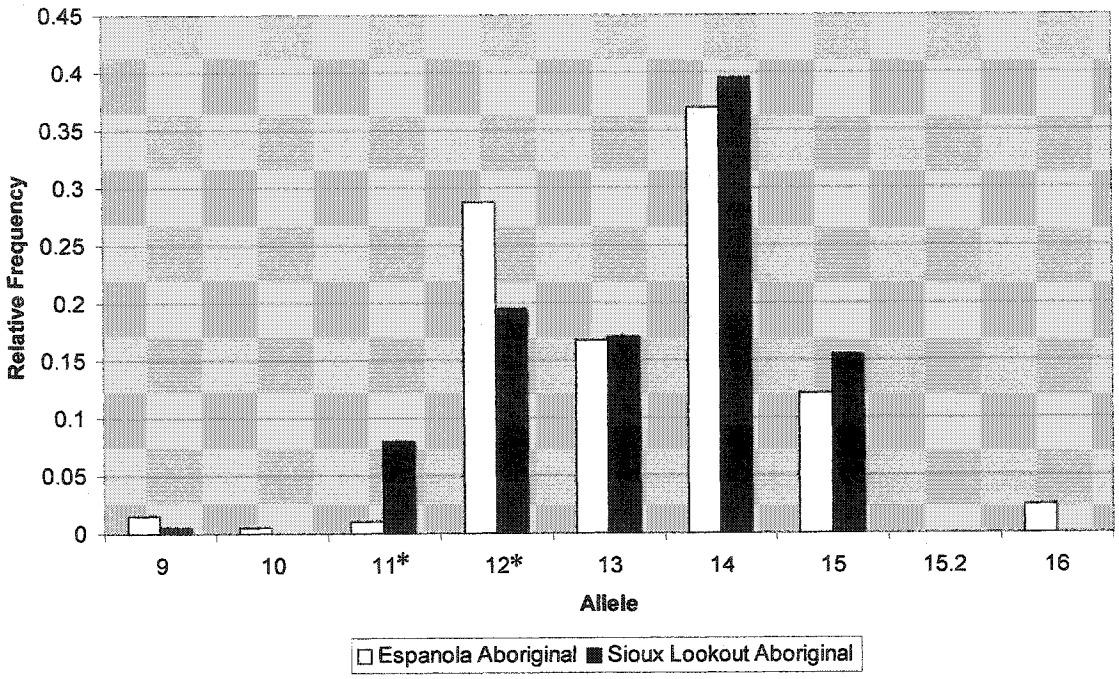


Figure 24. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Aboriginal and Sioux Lookout Aboriginal populations [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

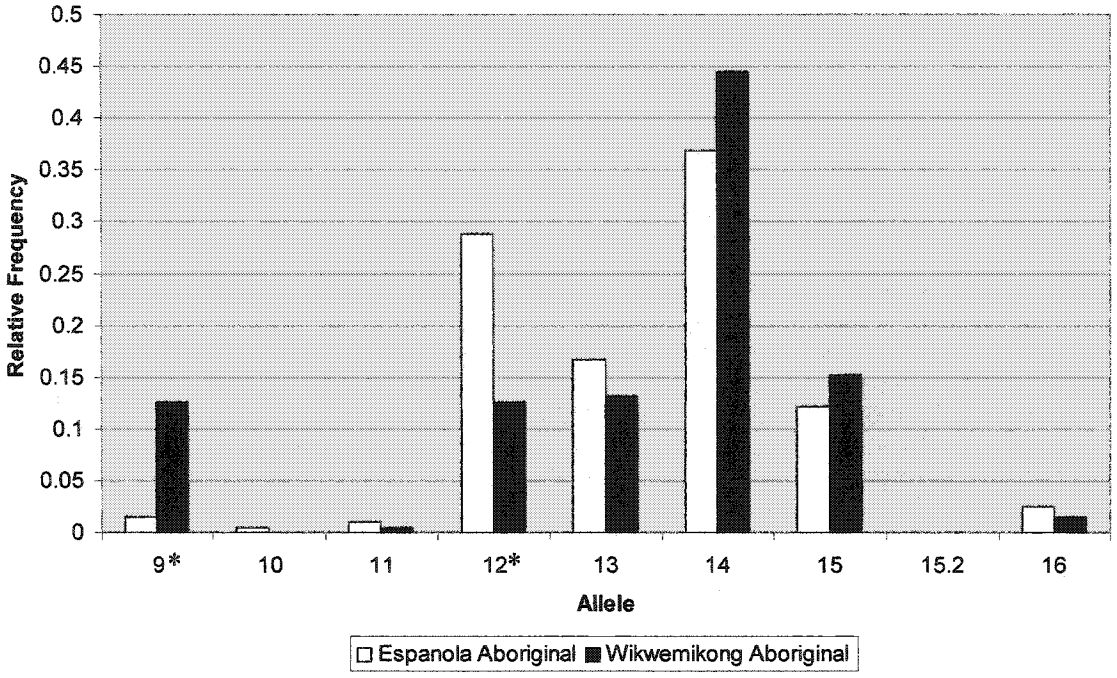


Figure 25. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Aboriginal and Wikwemikong Aboriginal populations [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

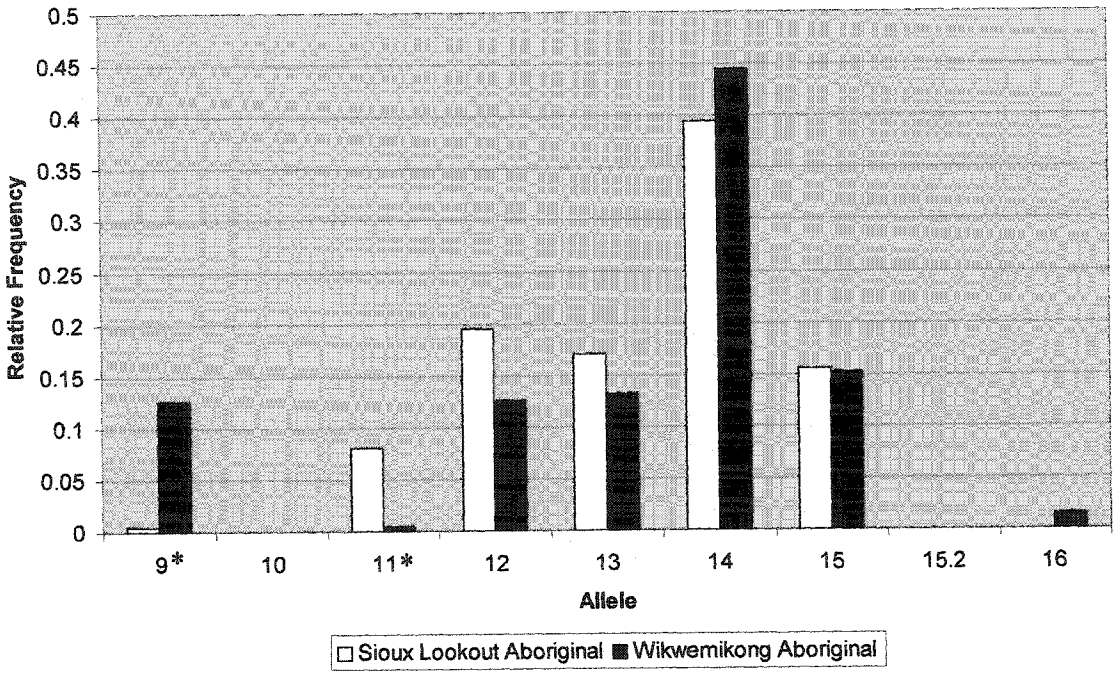


Figure 26. Pairwise comparison of the relative D18S535 allele frequency distributions between Sioux Lookout Aboriginal and Wikwemikong Aboriginal populations [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

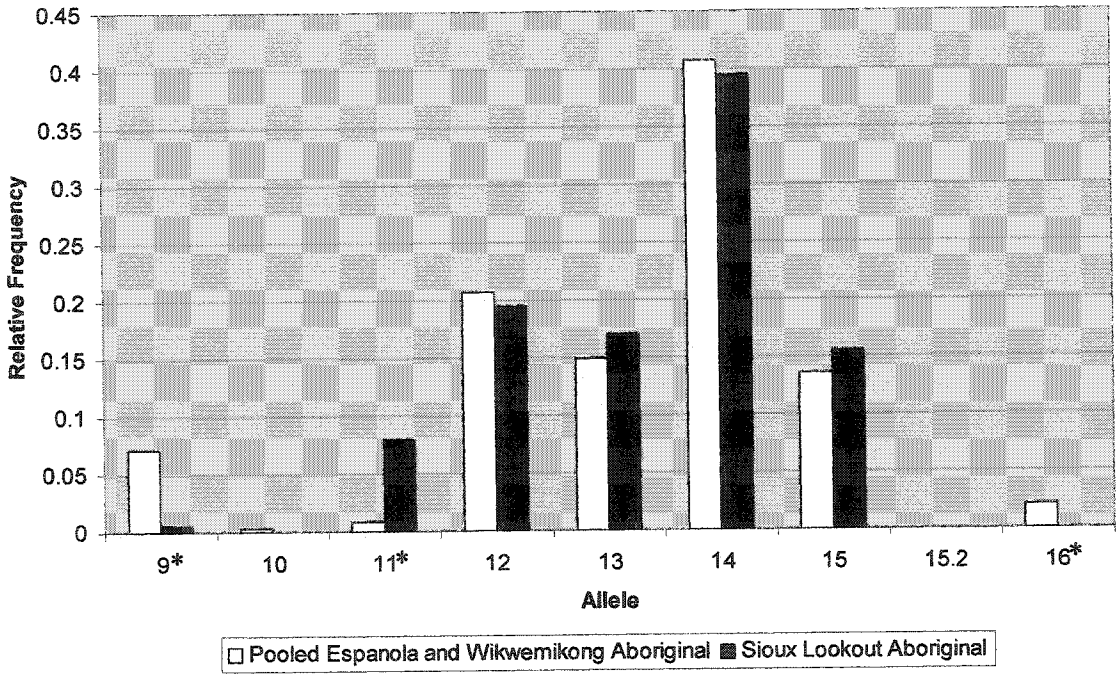


Figure 27. Pairwise comparison of the relative D18S535 allele frequency distributions between a pooled Espanola and Wikwemikong Aboriginal and Sioux Lookout Aboriginals [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

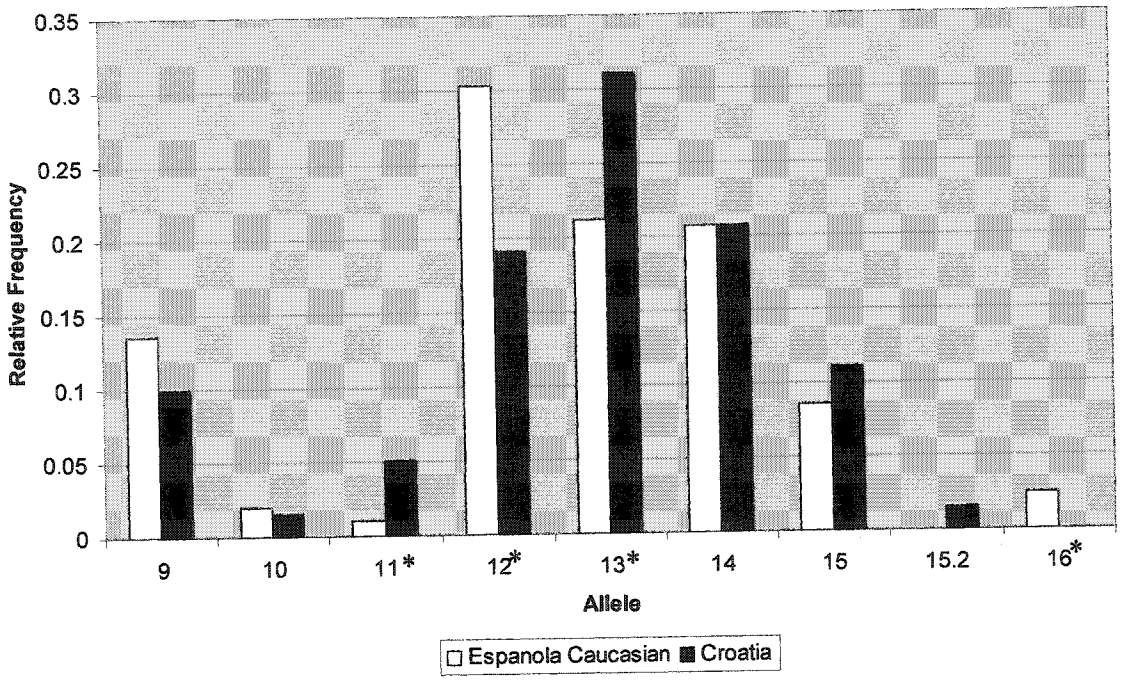


Figure 28. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Caucasians and a population from Croatia [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

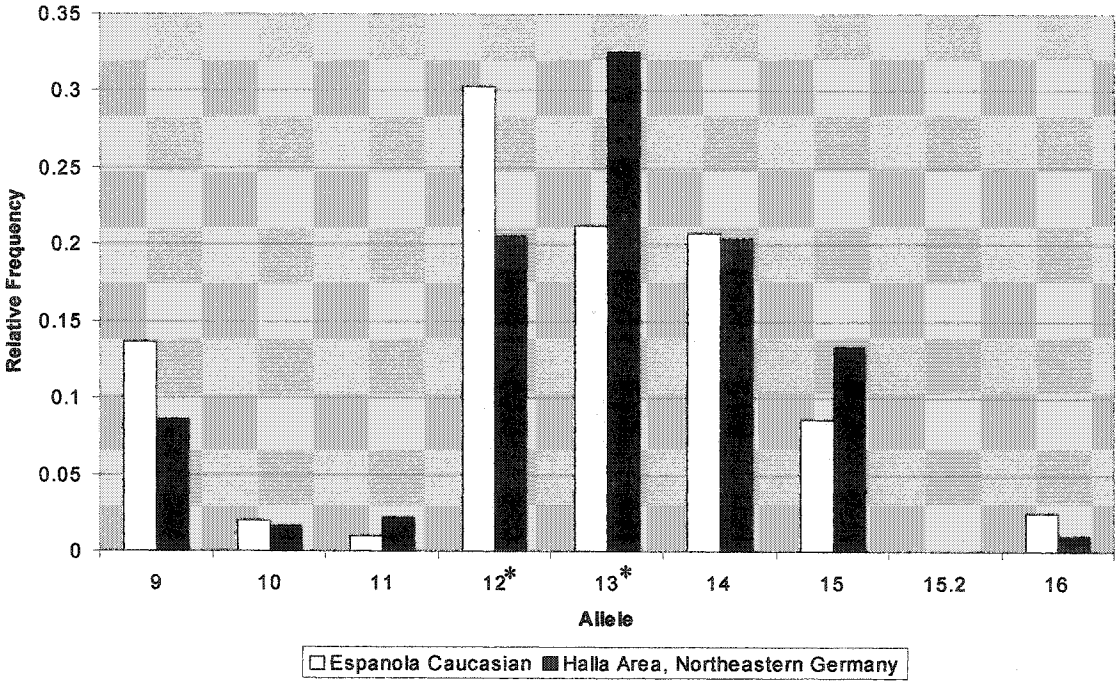


Figure 29. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Caucasians and a population from Halla Area in Northeastern Germany [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

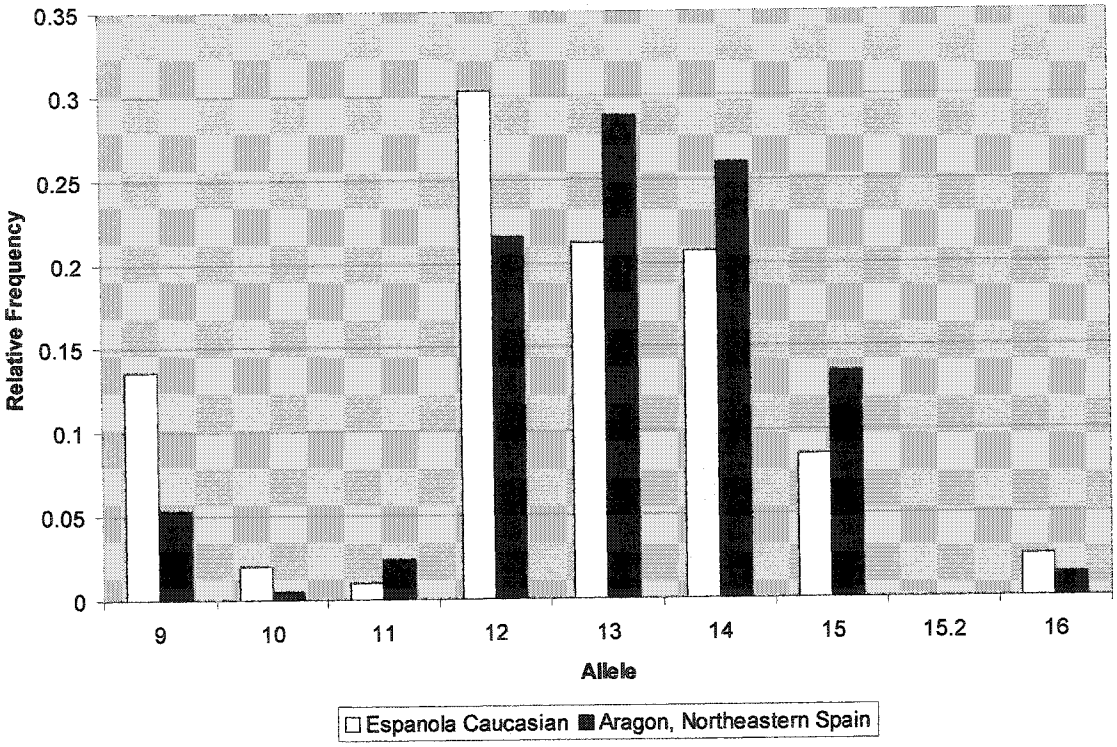


Figure 30. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Caucasians and a population from Aragon in Northeastern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

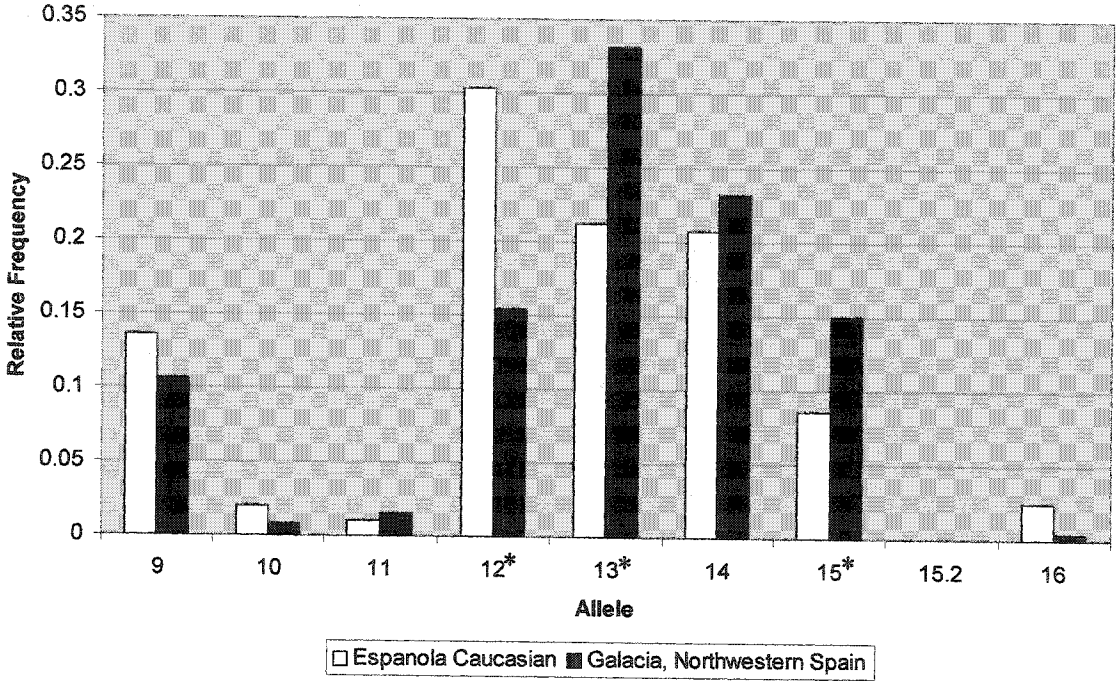


Figure 31. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Caucasians and a population from Galacia in Northwestern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

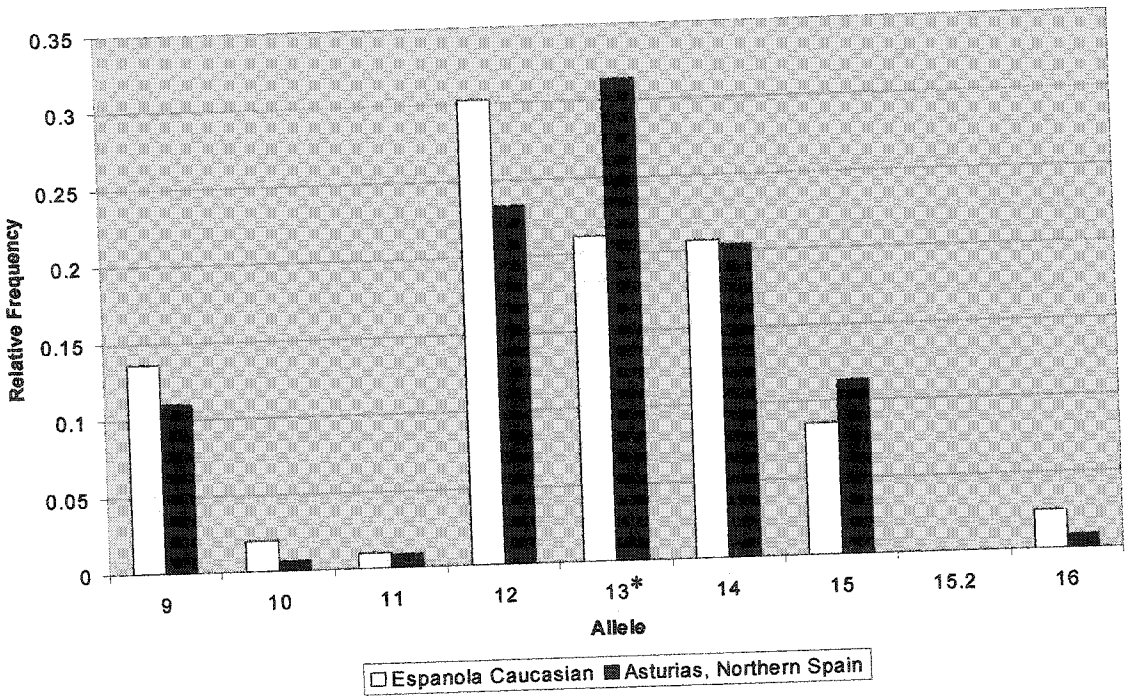


Figure 32. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Caucasians and a population from Asturias in Northern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

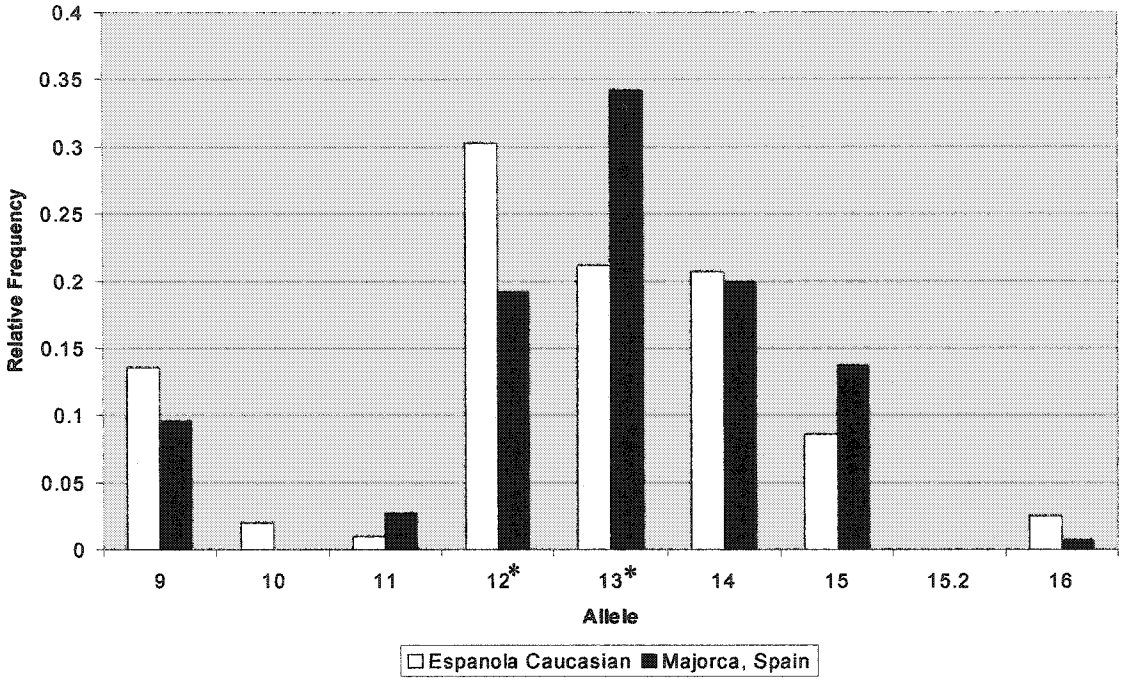


Figure 33. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Caucasians and a population from the island of Majorca in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

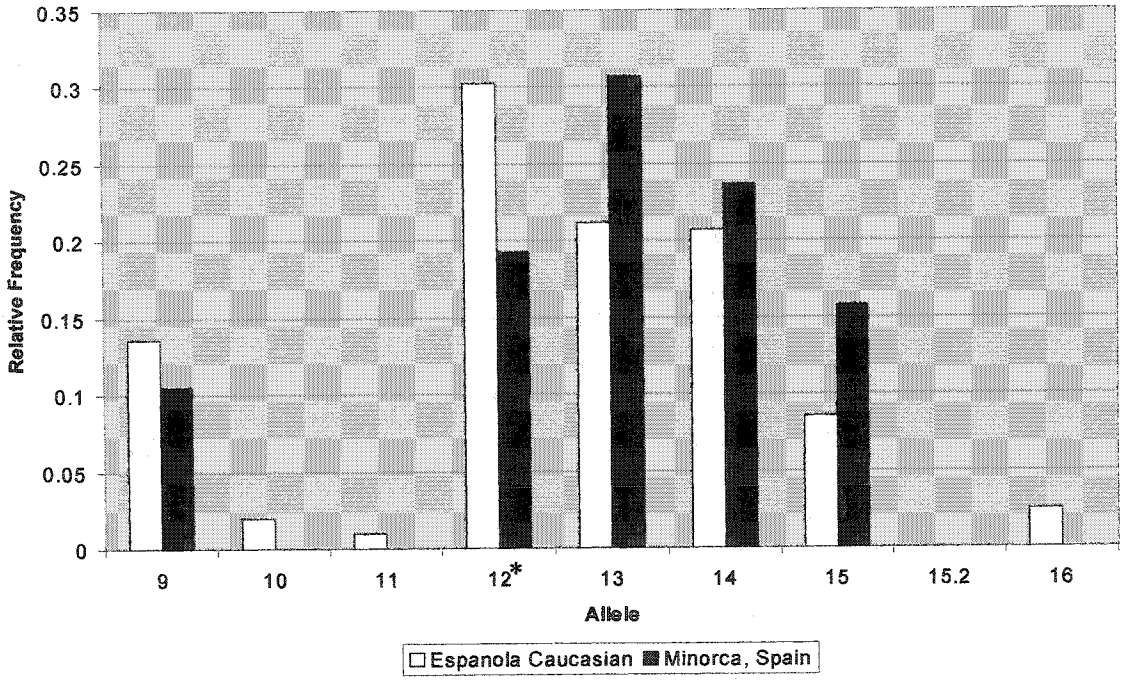


Figure 34. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Caucasians and a population from the island of Minorca in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

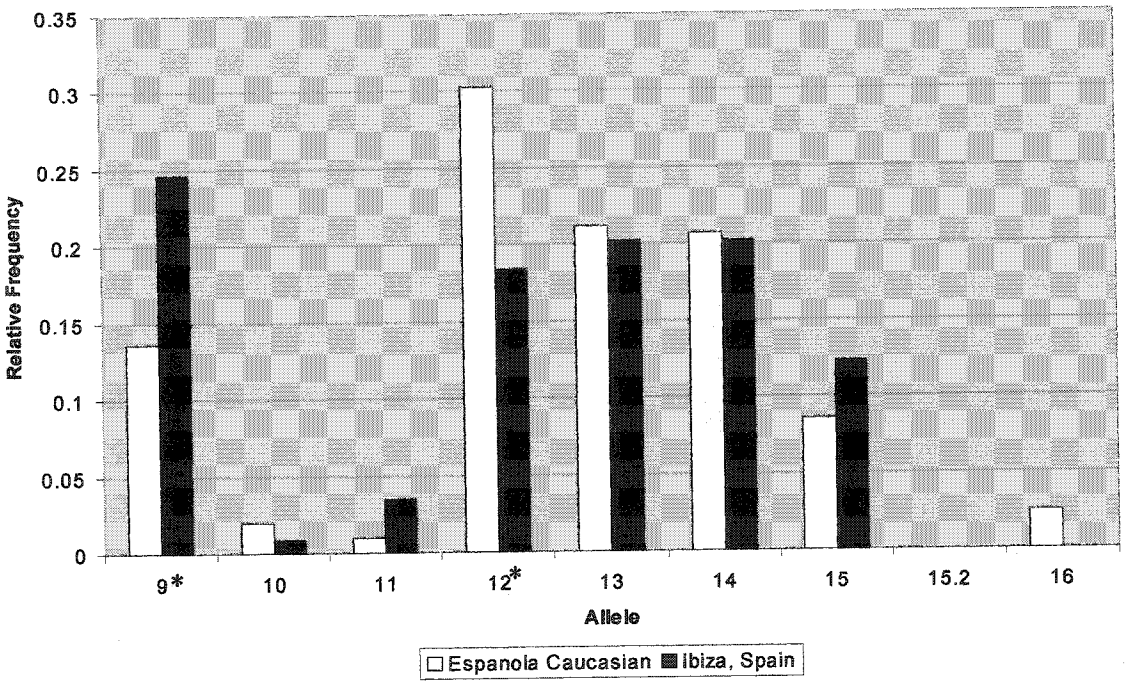


Figure 35. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Caucasians and a population from the island of Ibiza in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

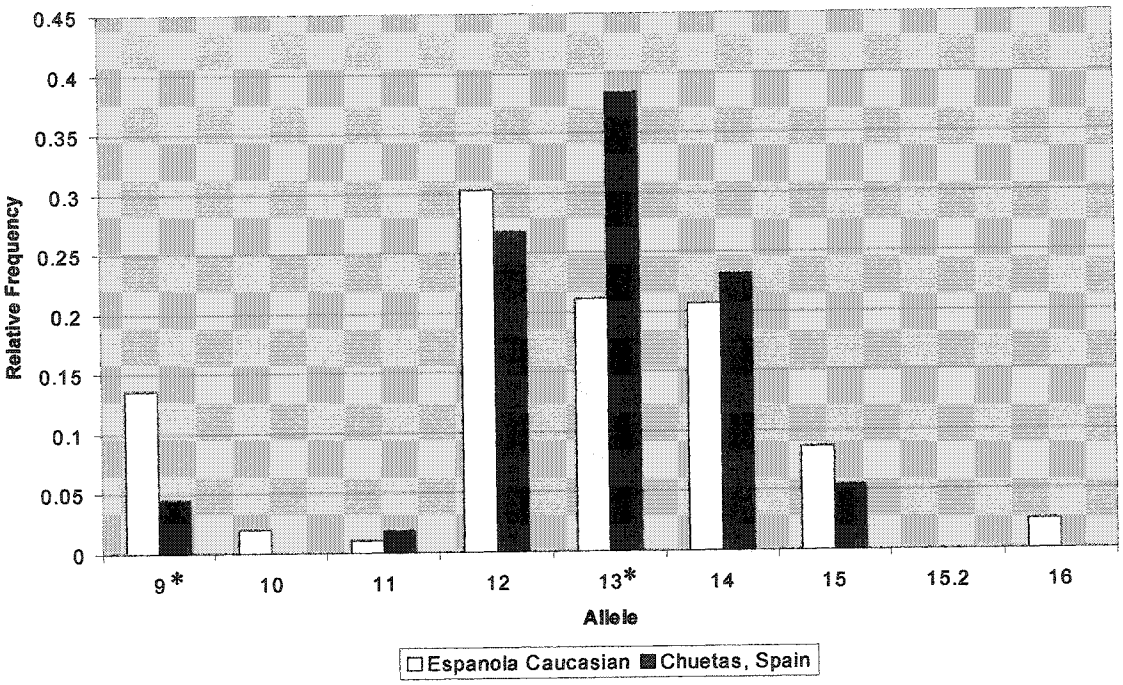


Figure 36. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Caucasians and a Chuetas population from Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

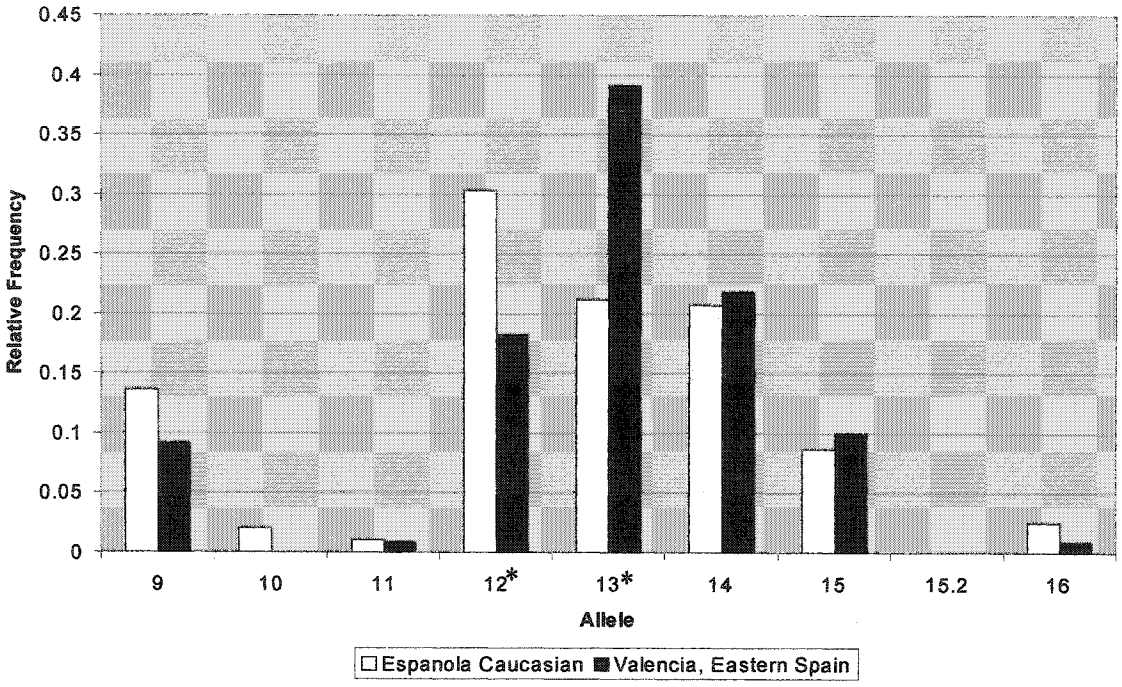


Figure 37. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Caucasians and a population from Valencia in Eastern [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

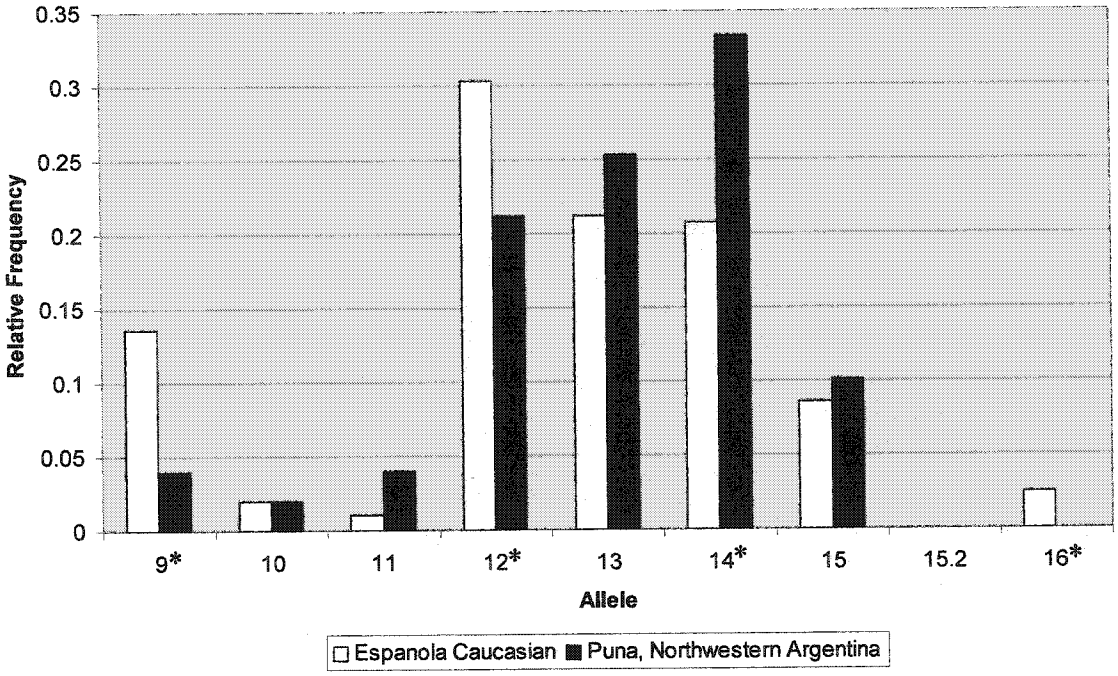


Figure 38. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Caucasians and a population from Puna in Northwestern Argentina [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

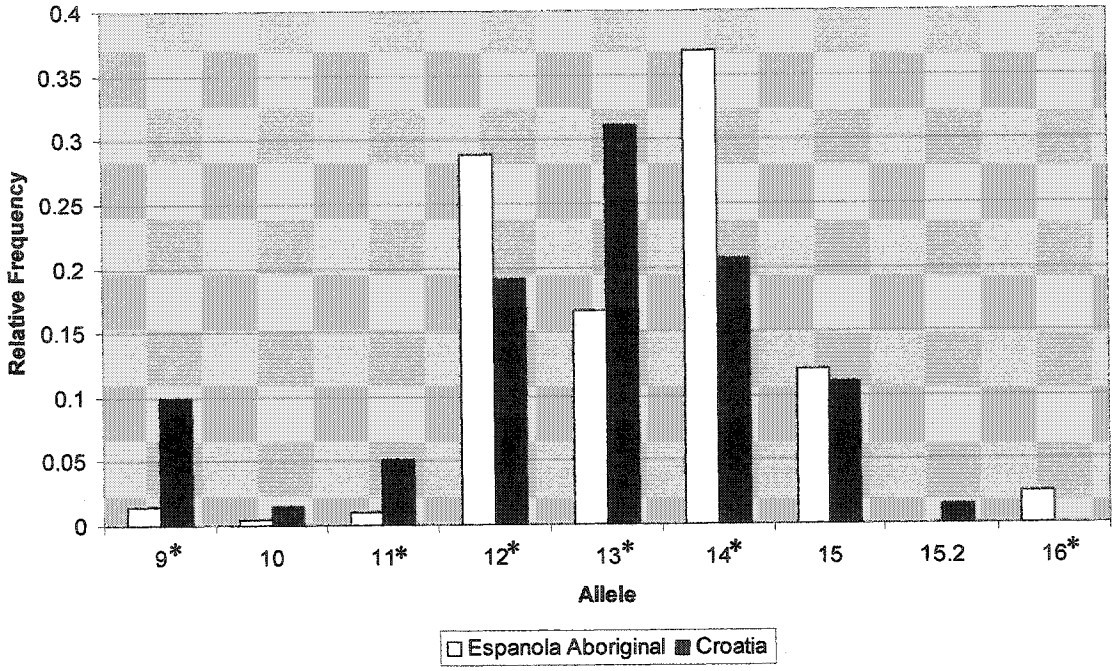


Figure 39. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Aboriginals and a population from Croatia [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

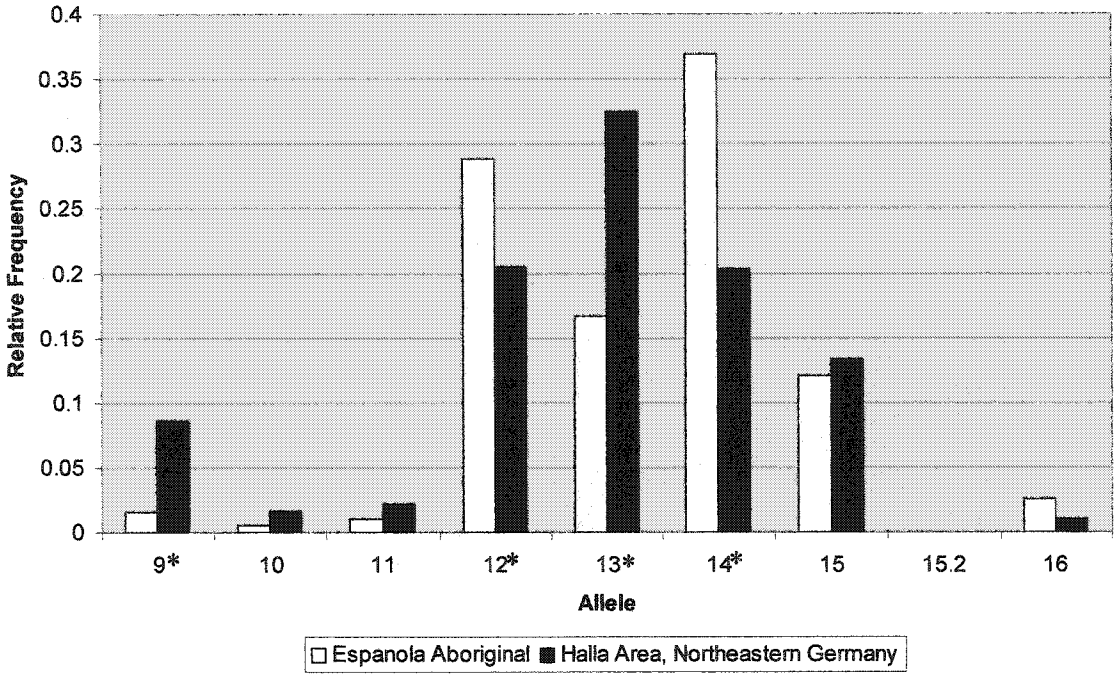


Figure 40. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Aboriginals and a population from Halla Area in Northeastern Germany [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

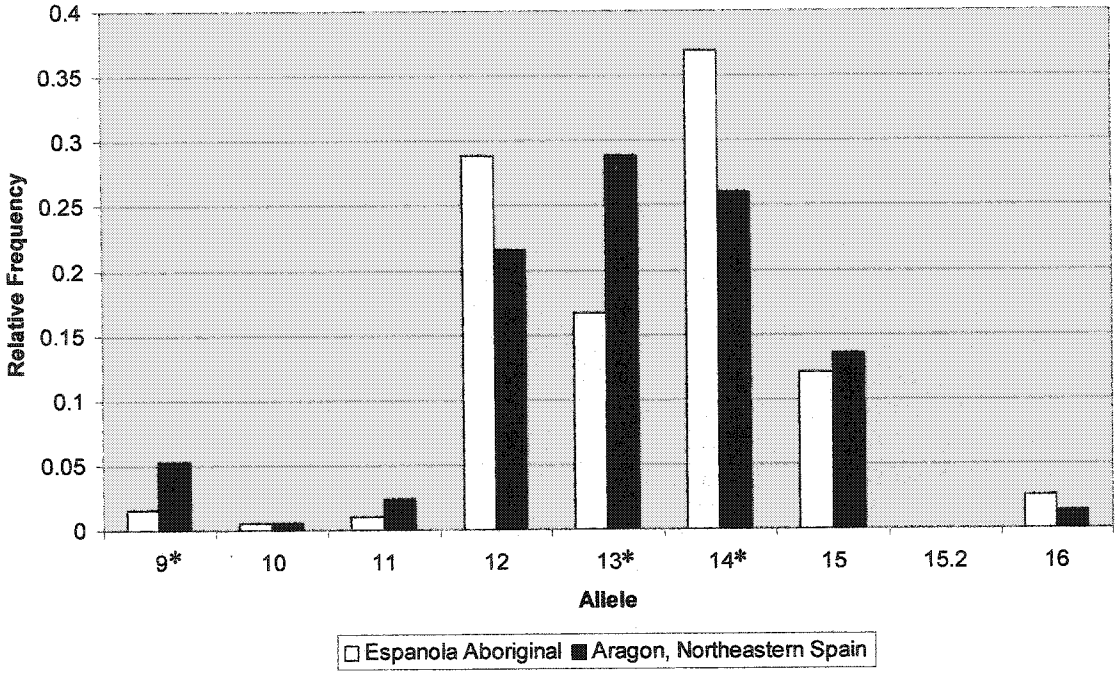


Figure 41. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Aboriginals and a population from Aragon in Northeastern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

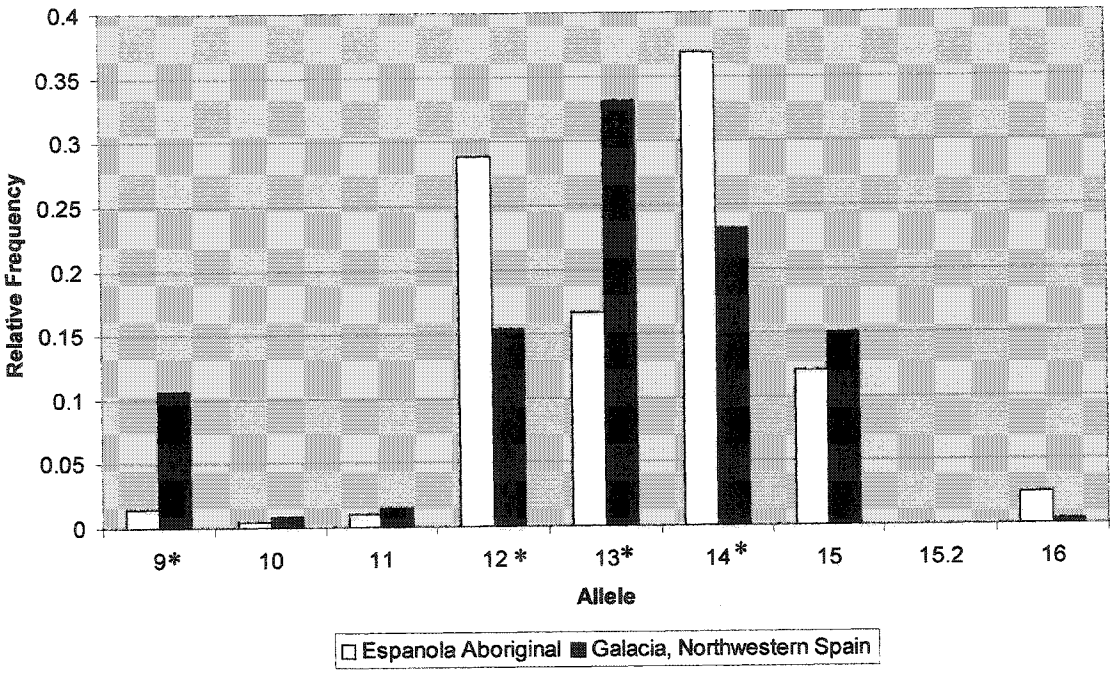


Figure 42. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Aboriginals and a population from Galacia in Northwestern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

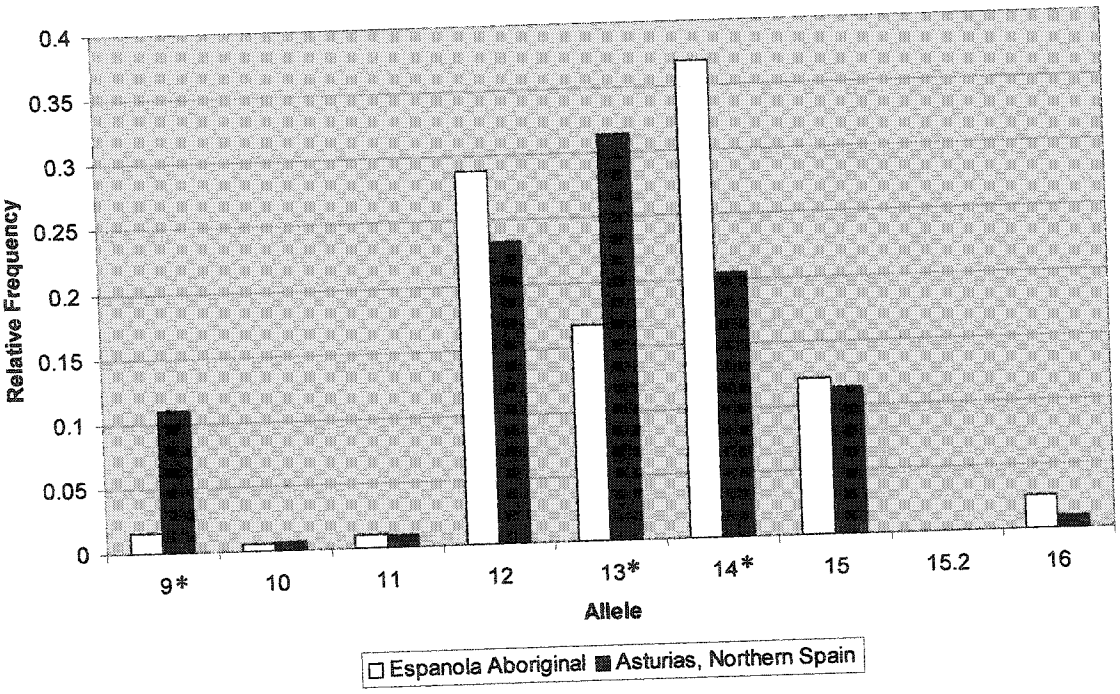


Figure 43. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Aboriginals and a population from Asturias in Northern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

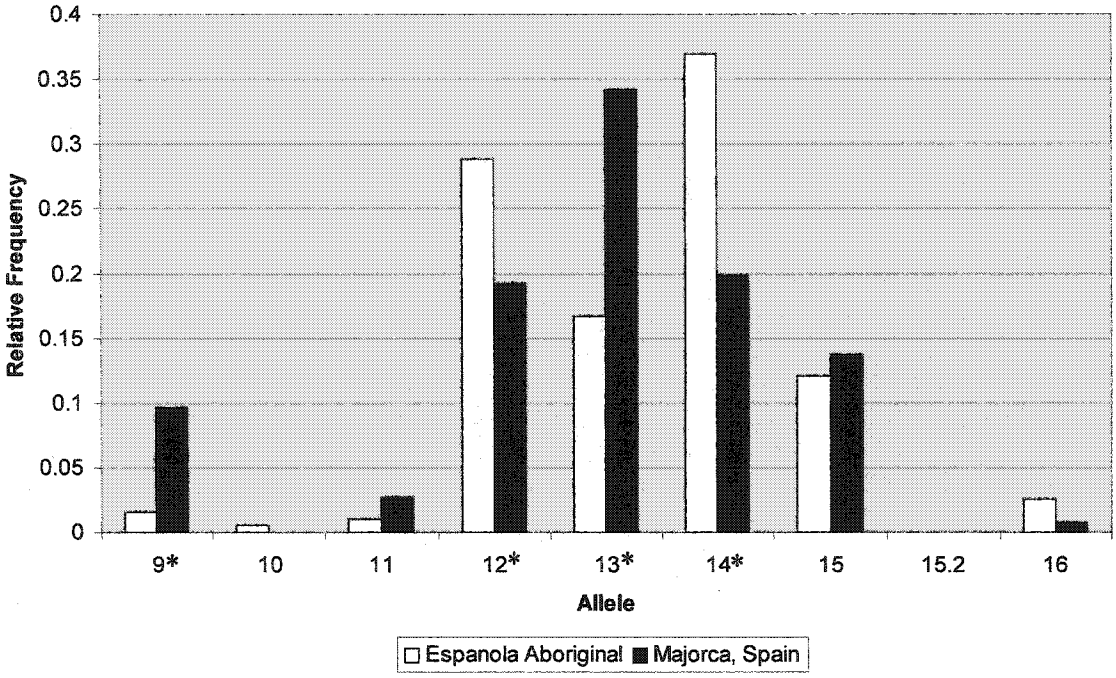


Figure 44. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Aboriginals and a population from the island of Majorca in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

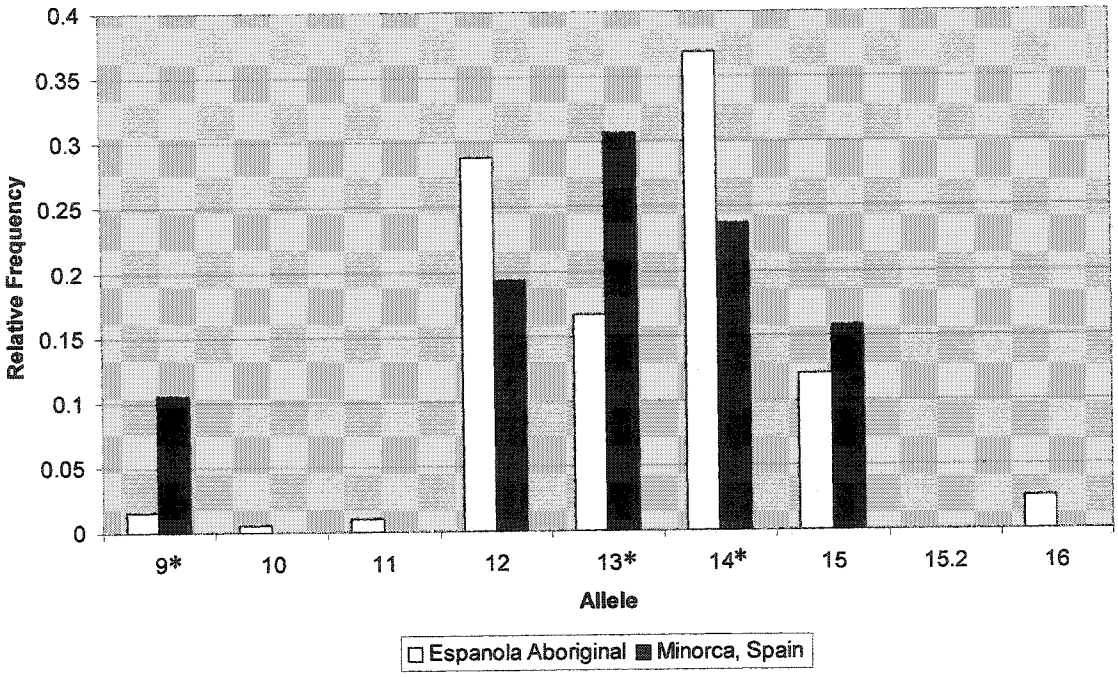


Figure 45. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Aboriginals and a population from the island of Minorca in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

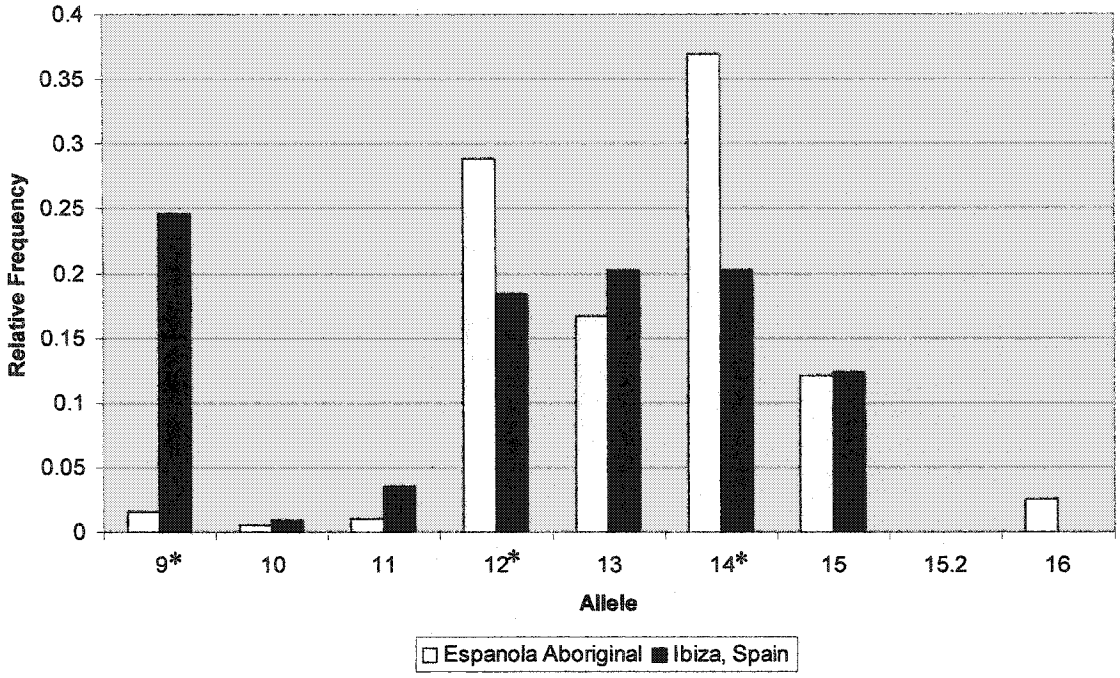


Figure 46. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Aboriginals and a population from the island of Ibiza in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

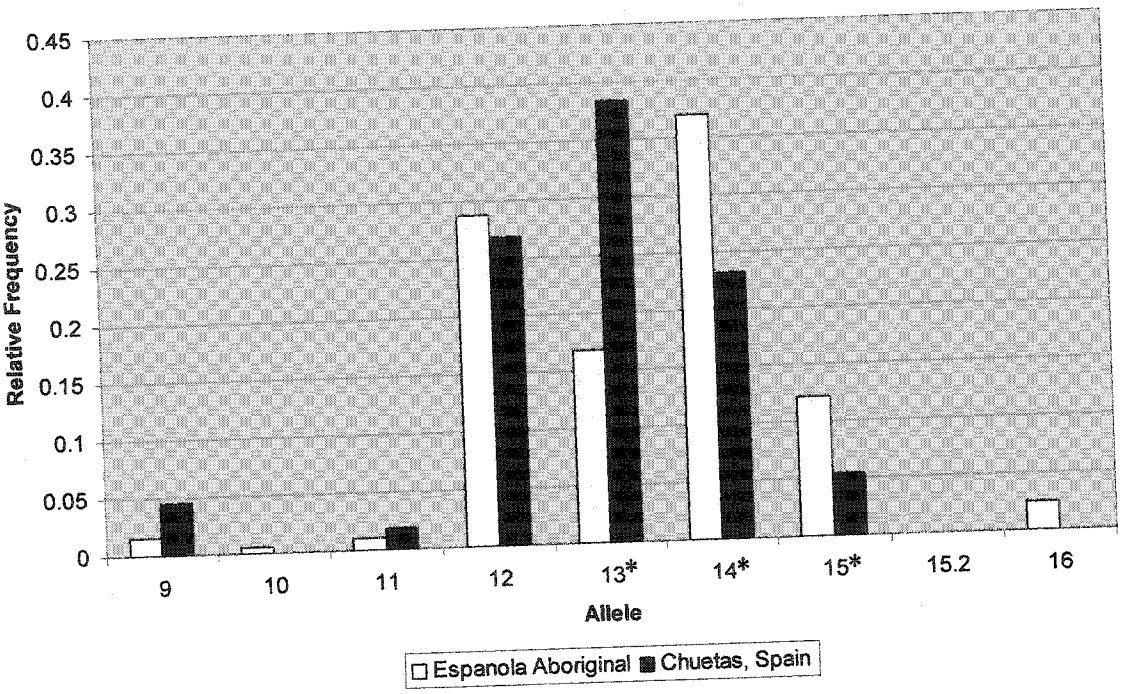


Figure 47. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Aboriginals and a Chuetas population from Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

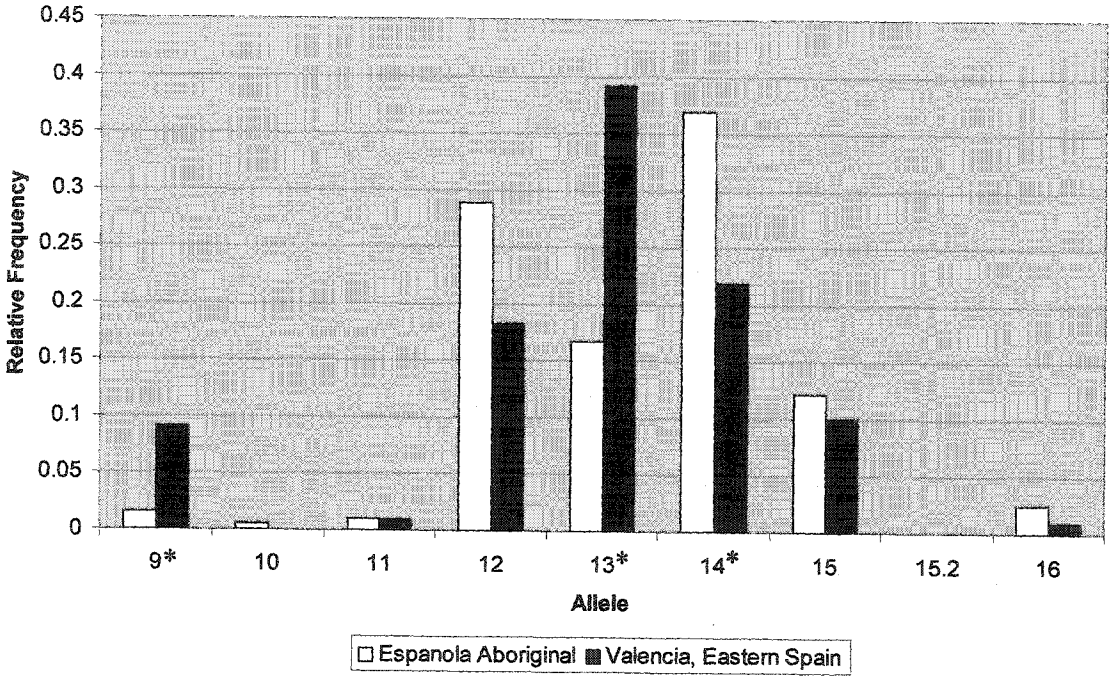


Figure 48. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Aboriginals and a population from Valencia in Eastern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

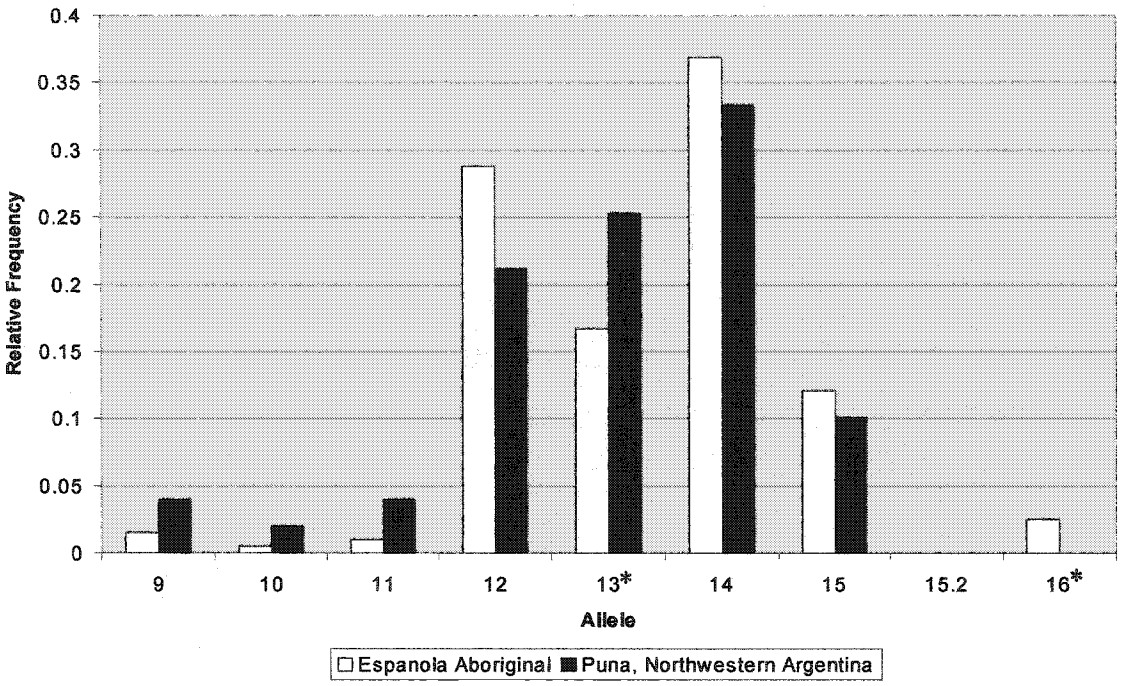


Figure 49. Pairwise comparison of the relative D18S535 allele frequency distributions between Espanola Aboriginals and a population from Puna in Northwestern Argentina [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

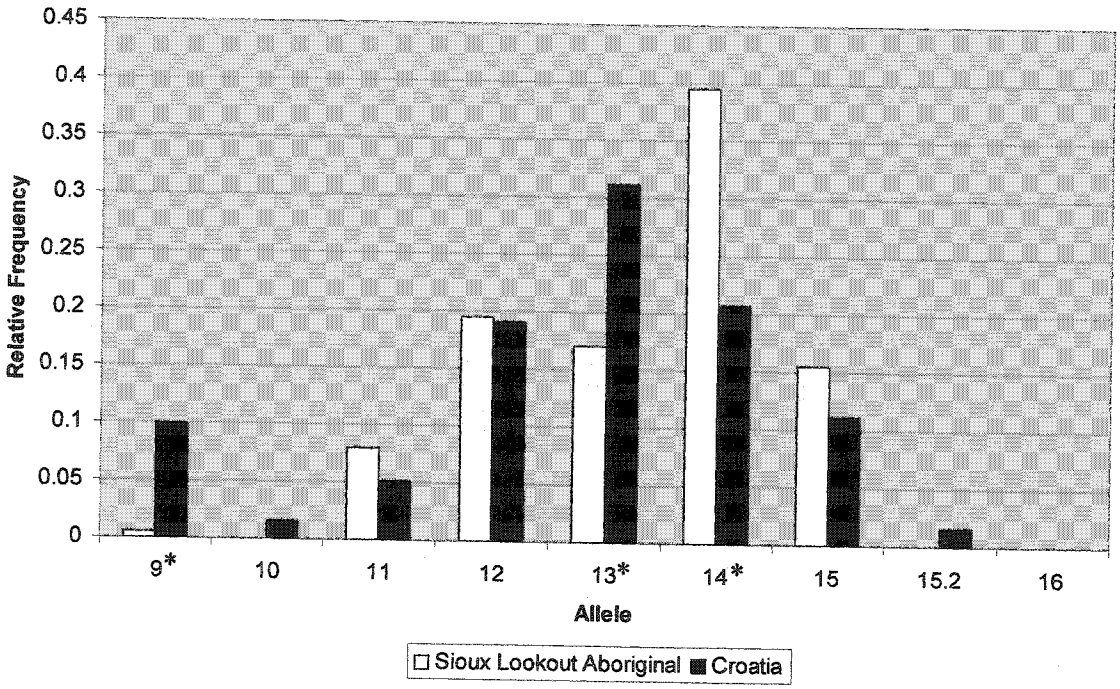


Figure 50. Pairwise comparison of the relative D18S535 allele frequency distributions between Sioux Lookout Aboriginals and a population from Croatia [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

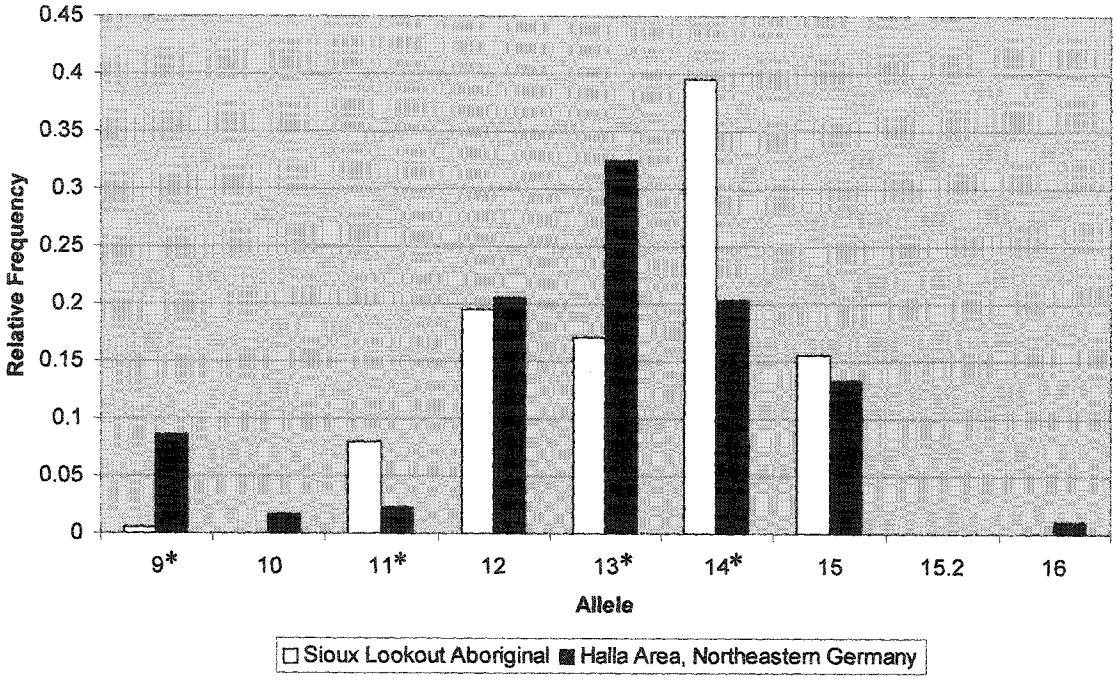


Figure 51. Pairwise comparison of the relative D18S535 allele frequency distributions between Sioux Lookout Aboriginals and a population from Halla Area in Northeastern Germany [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

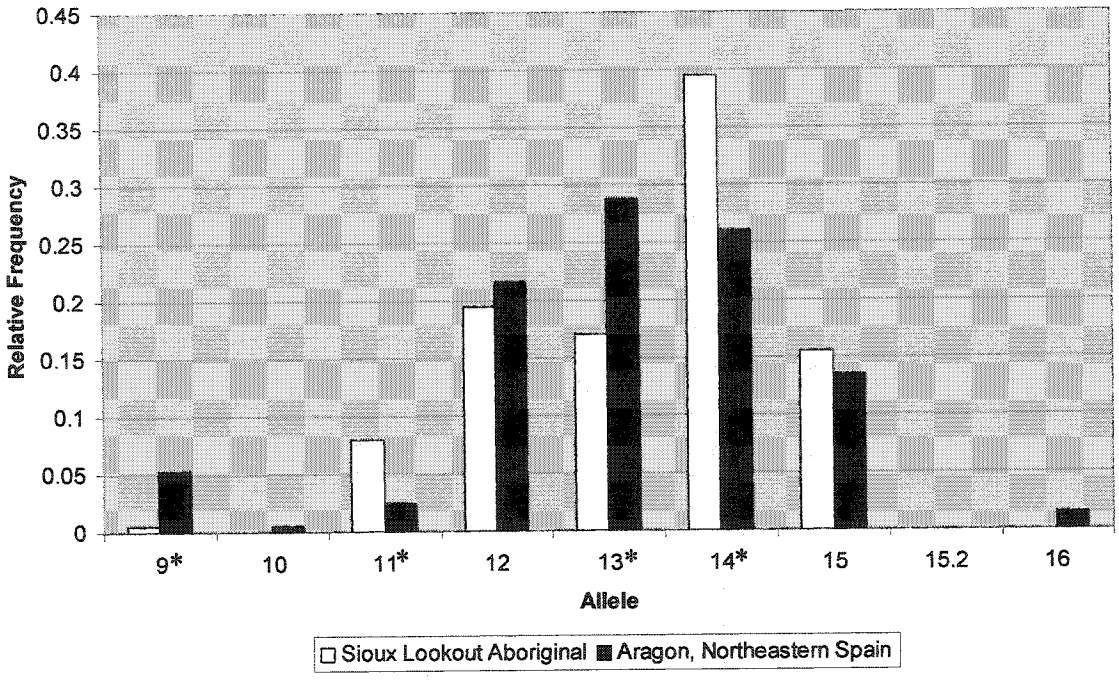


Figure 52. Pairwise comparison of the relative D18S535 allele frequency distributions between Sioux Lookout Aboriginals and a population from Aragon in Northeastern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

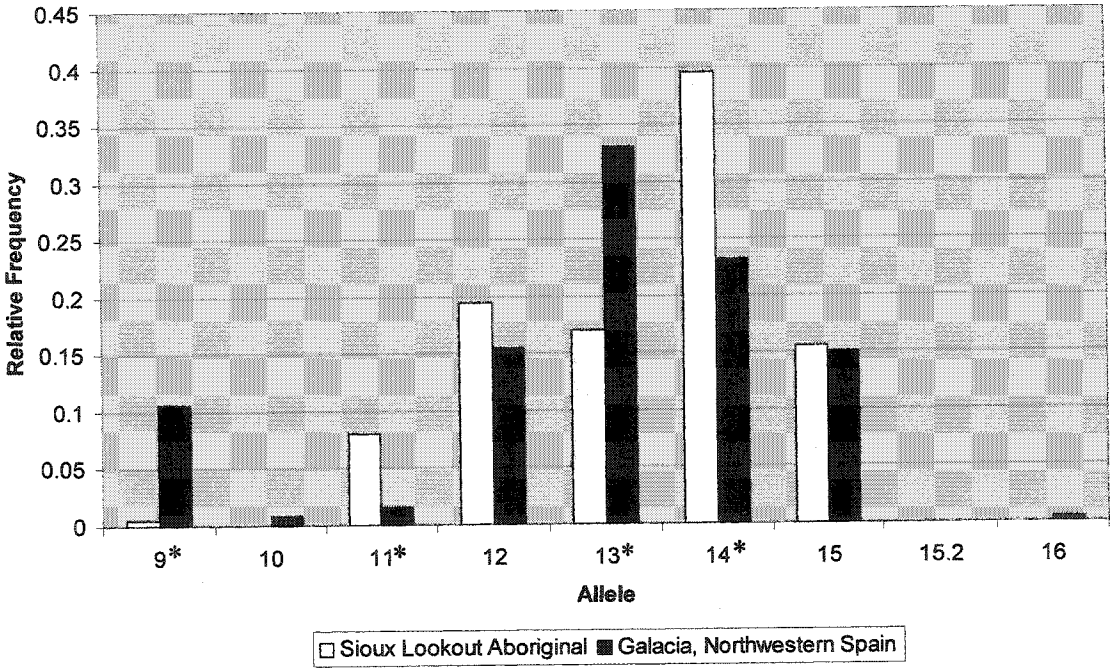


Figure 53. Pairwise comparison of the relative D18S535 allele frequency distributions between Sioux Lookout Aboriginals and a population from Galacia in Northwestern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

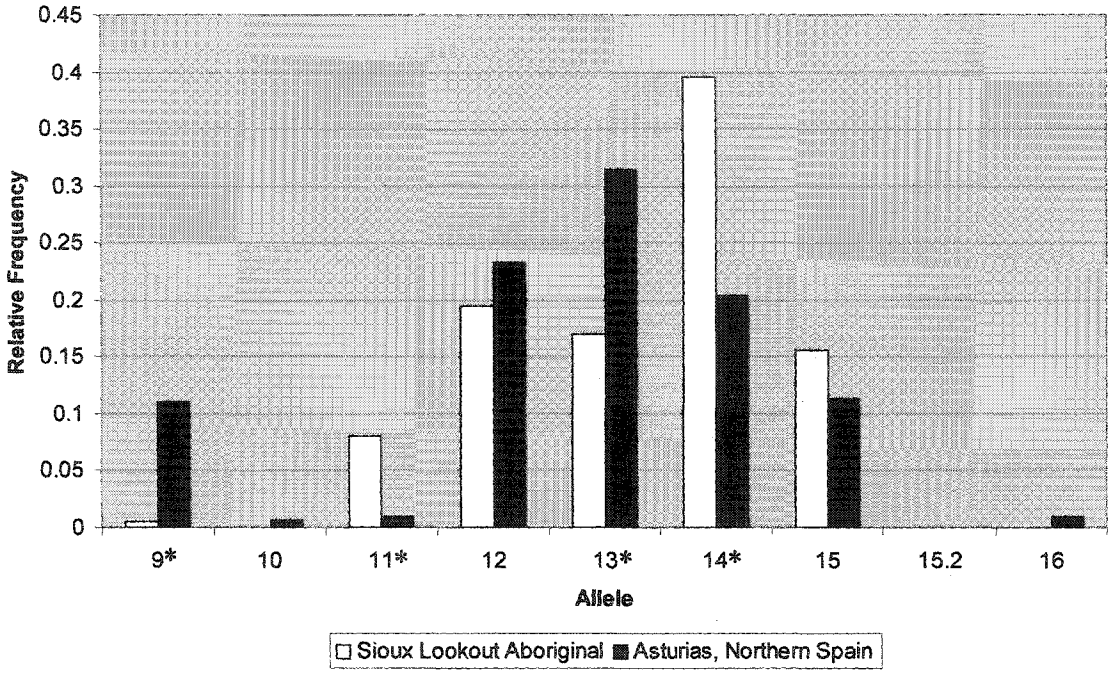


Figure 54. Pairwise comparison of the relative D18S535 allele frequency distributions between Sioux Lookout Aboriginals and a population from Asturias in Northern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

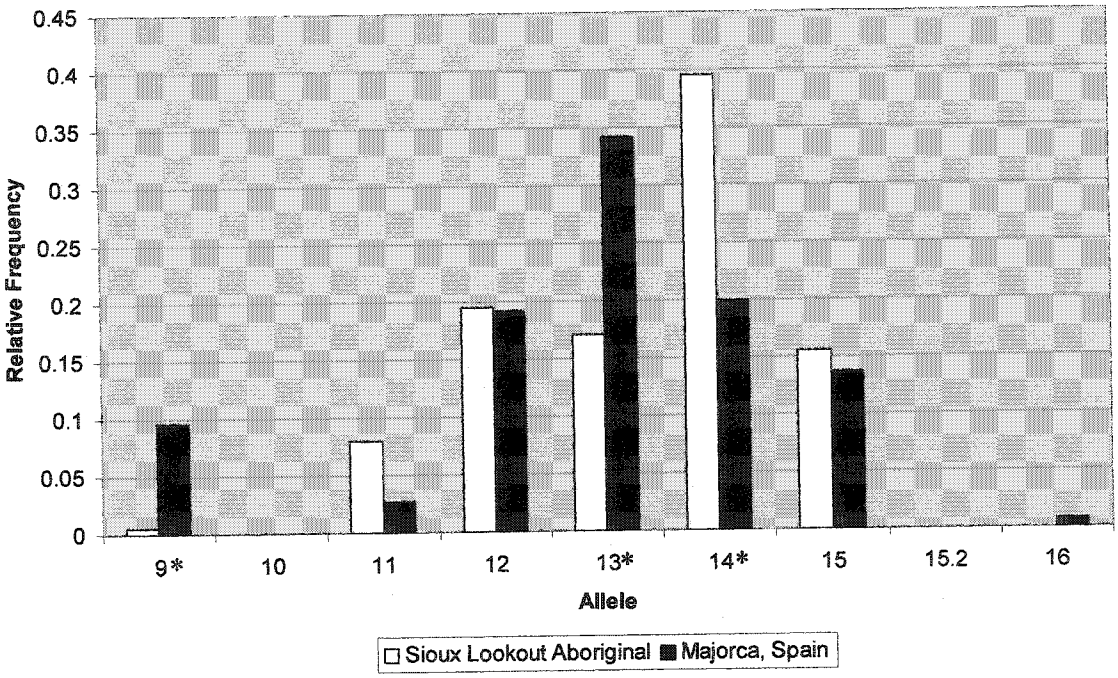


Figure 55. Pairwise comparison of the relative D18S535 allele frequency distributions between Sioux Lookout Aboriginals and a population from the island of Majorca in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

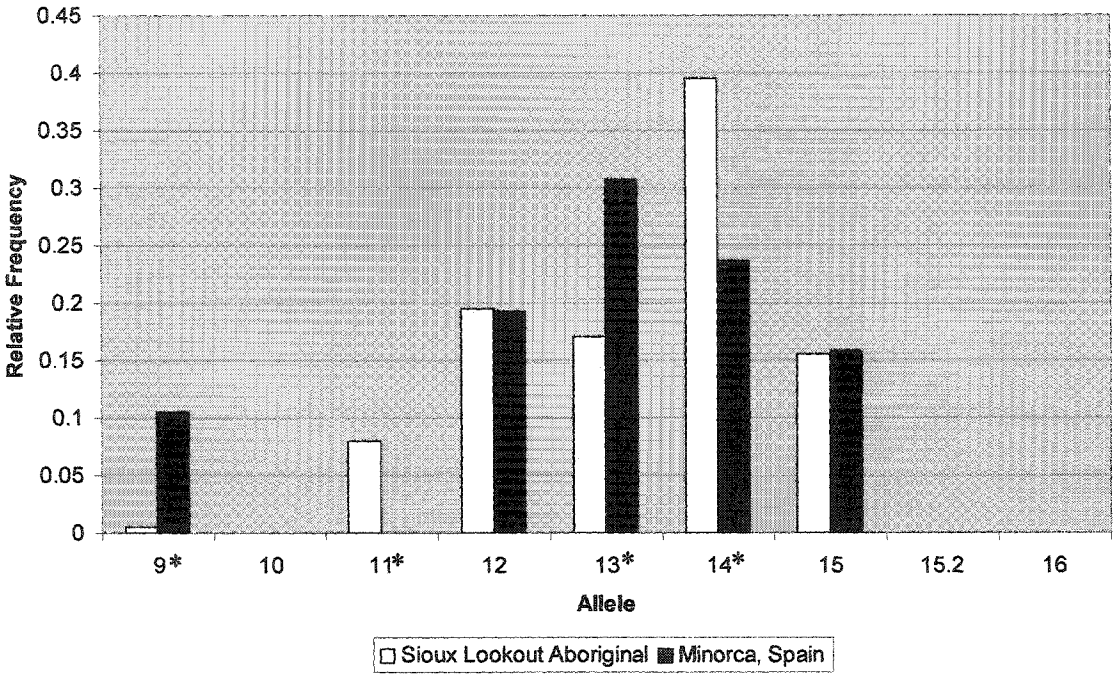


Figure 56. Pairwise comparison of the relative D18S535 allele frequency distributions between Sioux Lookout Aboriginals and a population from the island of Minorca in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

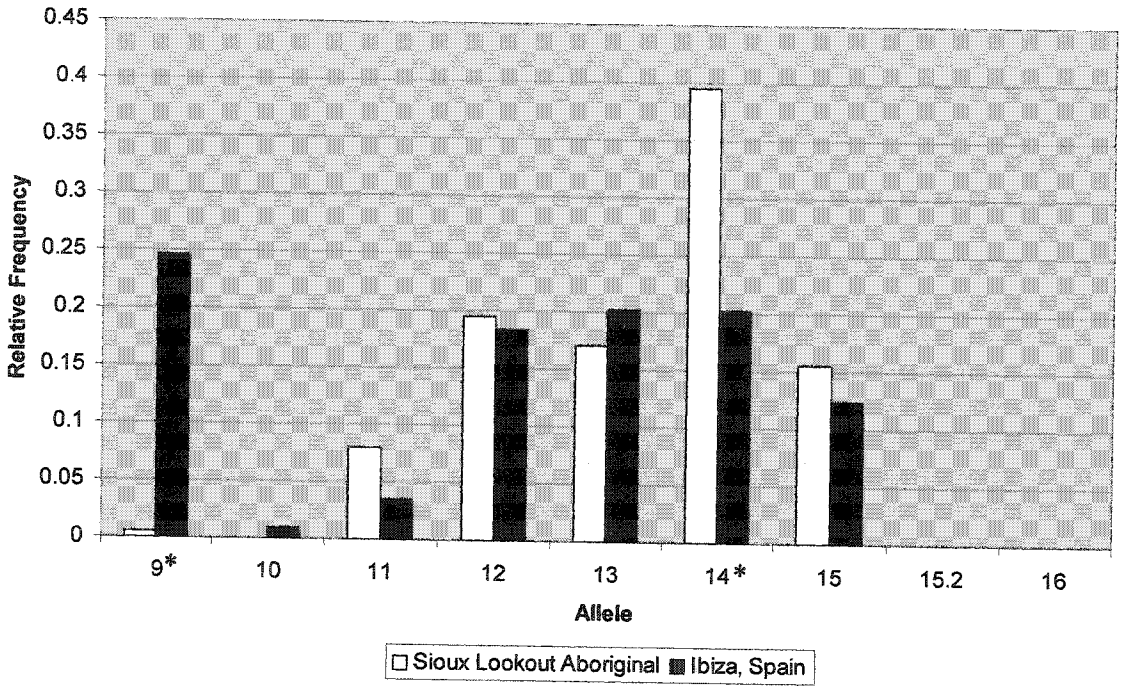


Figure 57. Pairwise comparison of the relative D18S535 allele frequency distributions between Sioux Lookout Aboriginals and a population from the island of Ibiza in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

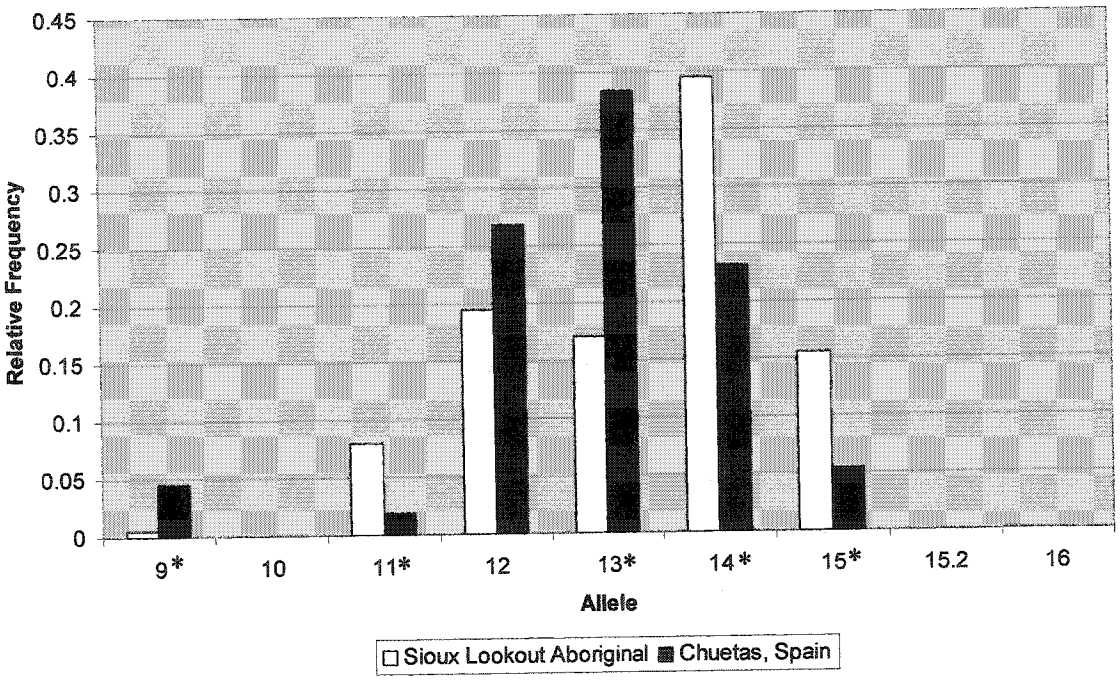


Figure 58. Pairwise comparison of the relative D18S535 allele frequency distributions between Sioux Lookout Aboriginals and a Chuetas population from Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

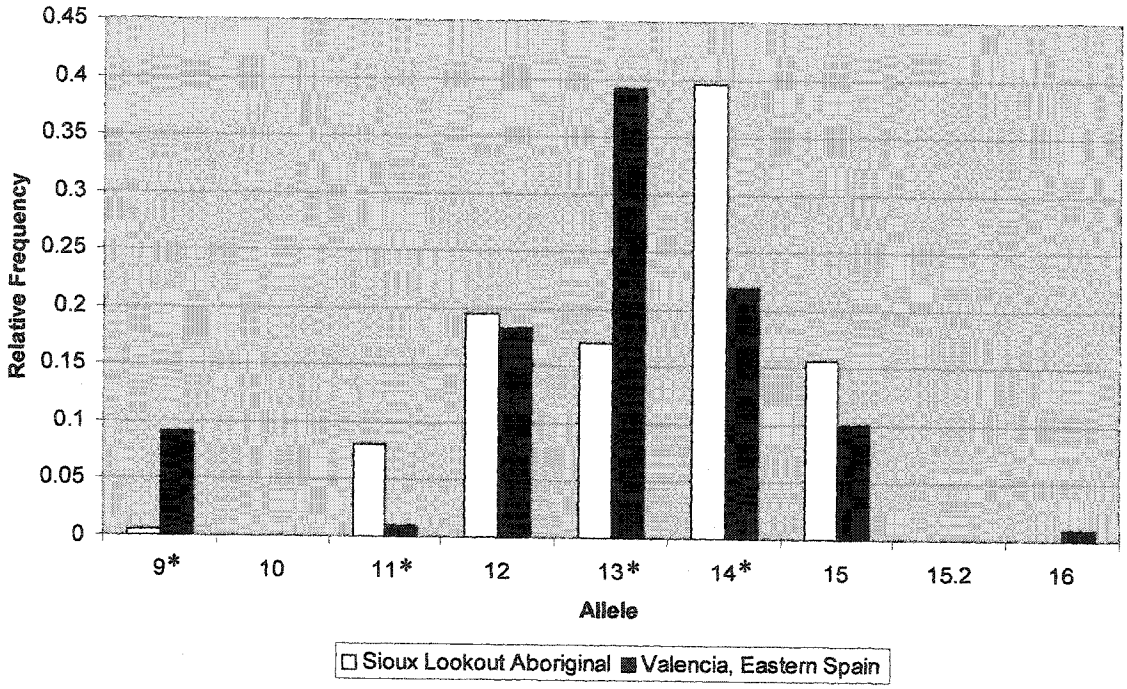


Figure 59. Pairwise comparison of the relative D18S535 allele frequency distributions between Sioux Lookout Aboriginals and a population from Valencia in Eastern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

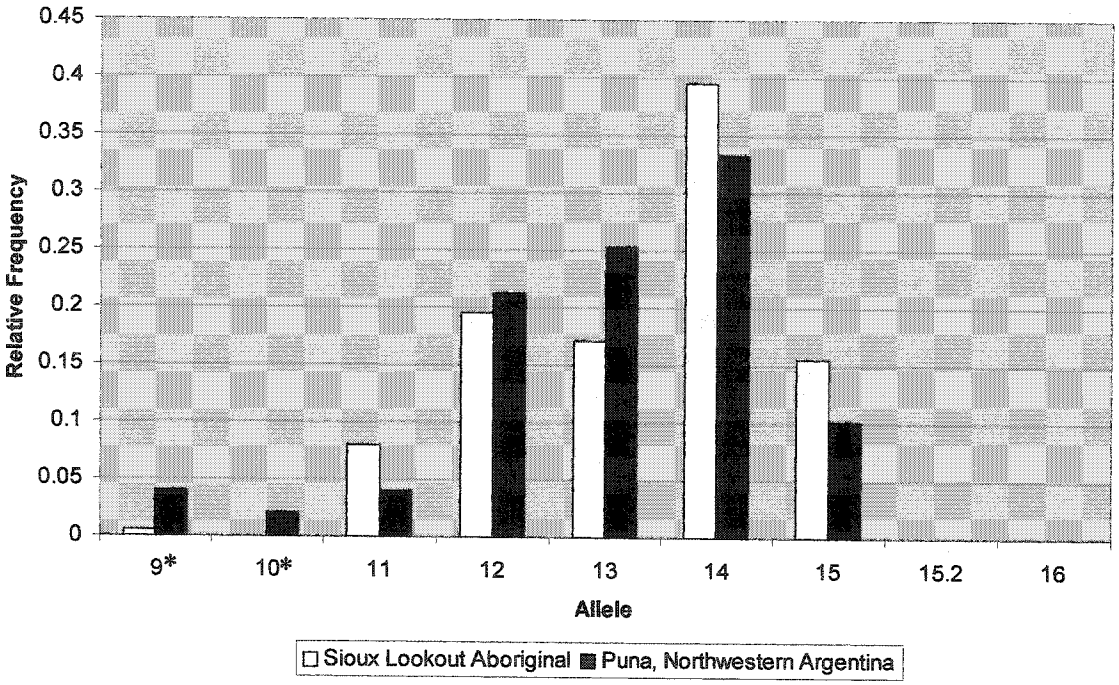


Figure 60. Pairwise comparison of the relative D18S535 allele frequency distributions between Sioux Lookout Aboriginals and a population from Puna in Northwestern Argentina [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

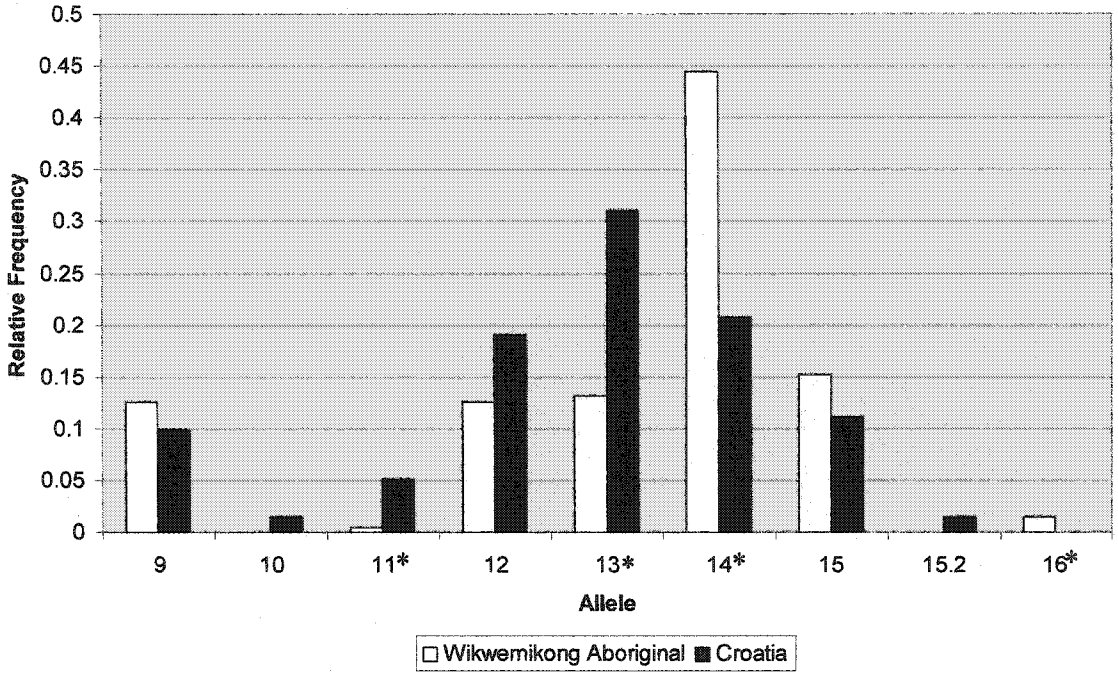


Figure 61. Pairwise comparison of the relative D18S535 allele frequency distributions between Wikwemikong Aboriginals and a population from Croatia [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

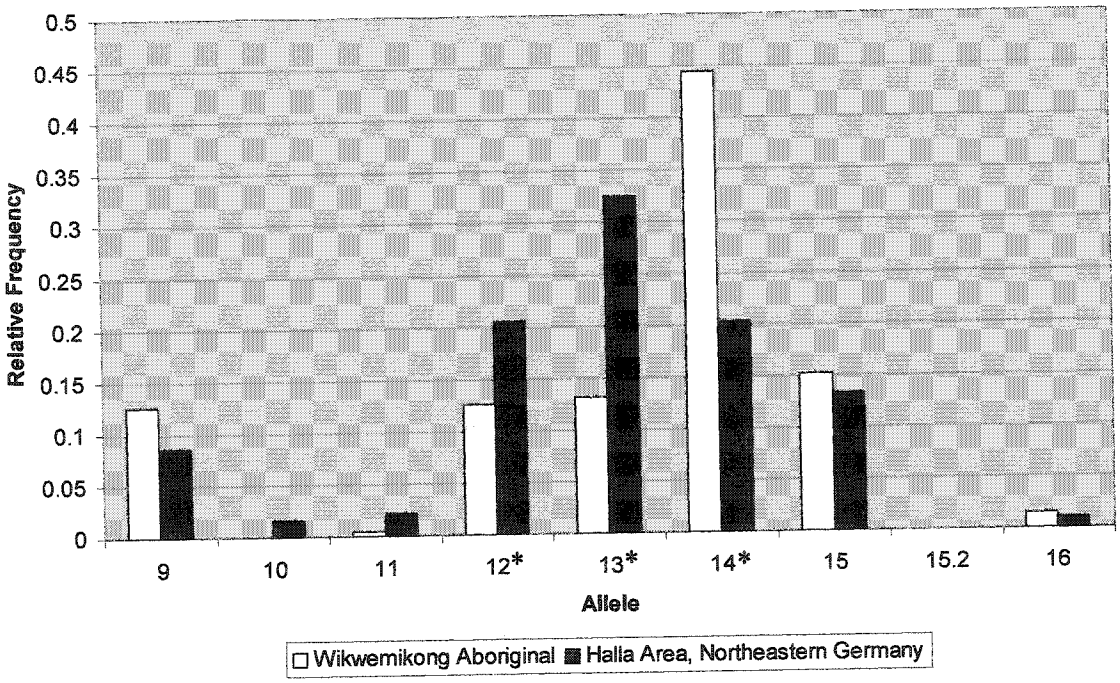


Figure 62. Pairwise comparison of the relative D18S535 allele frequency distributions between Wikwemikong Aboriginals and a population from Halla Area in Northeastern Germany [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

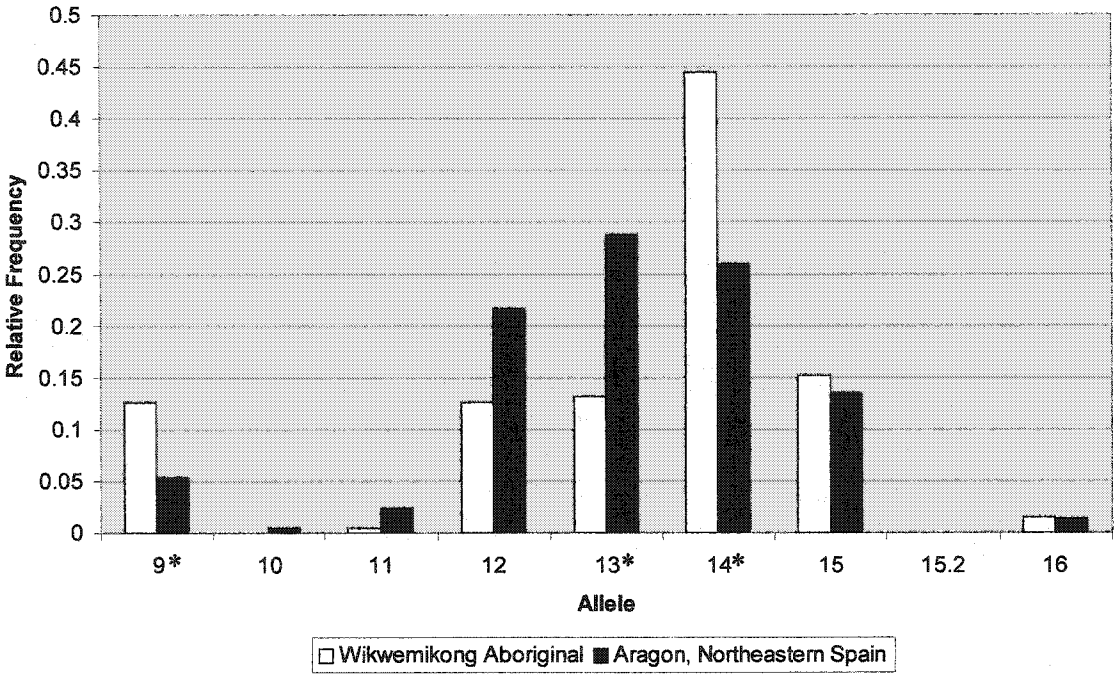


Figure 63. Pairwise comparison of the relative D18S535 allele frequency distributions between Wikwemikong Aboriginals and a population from Aragon in Northeastern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

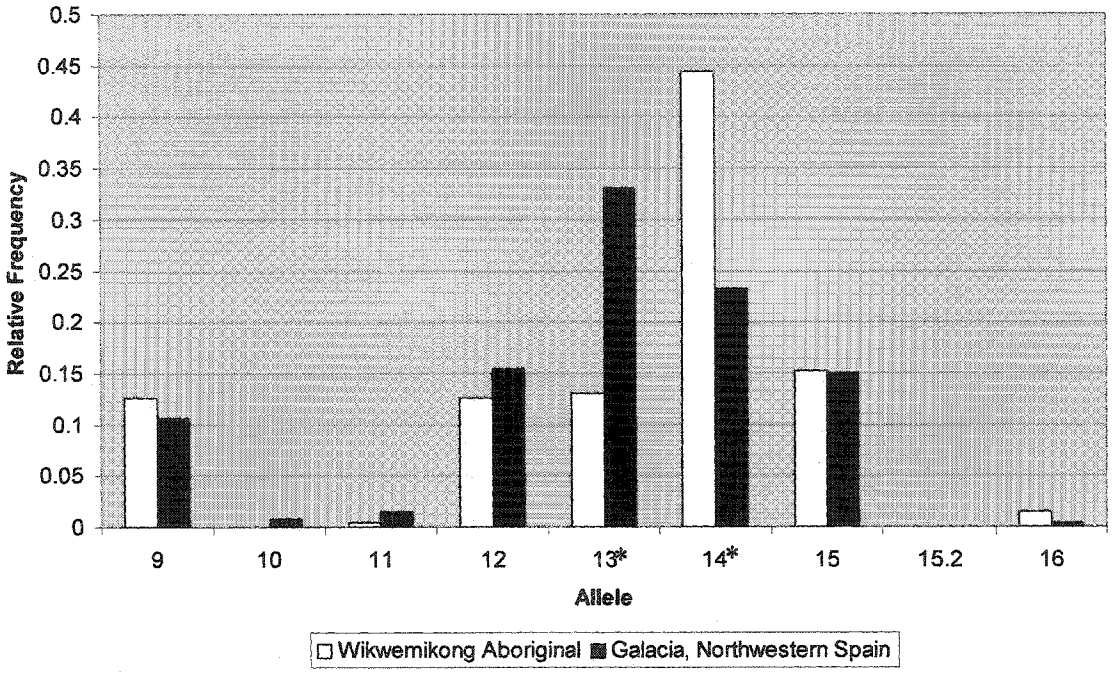


Figure 64. Pairwise comparison of the relative D18S535 allele frequency distributions between Wikwemikong Aboriginals and a population from Galacia in Northwestern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

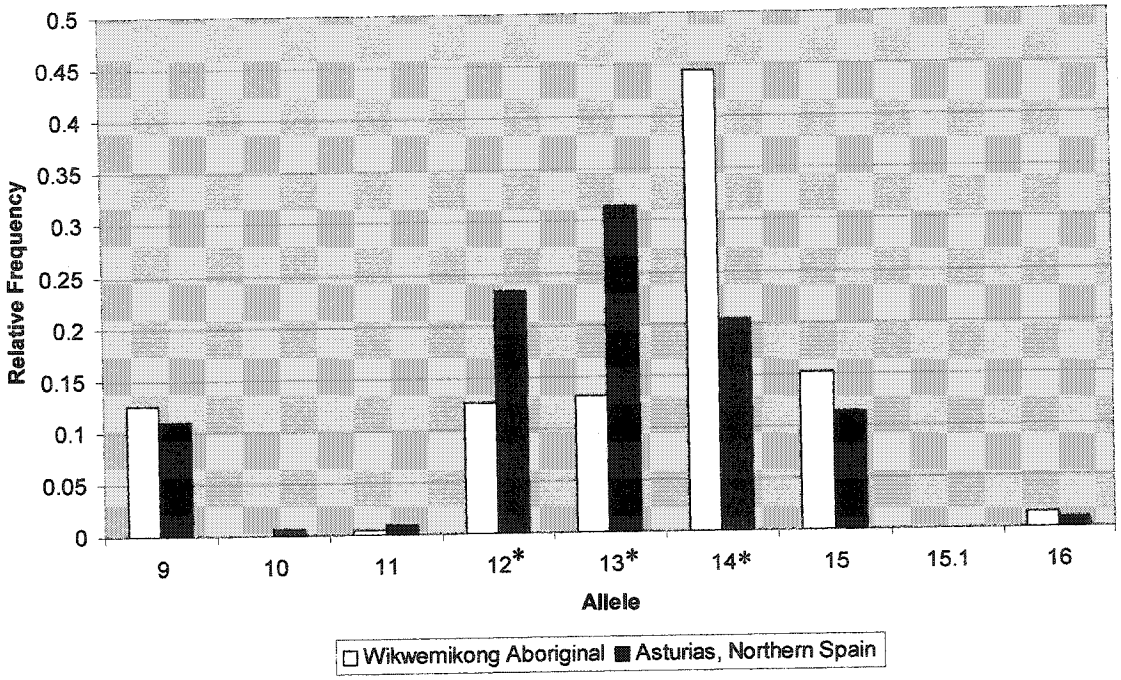


Figure 65. Pairwise comparison of the relative D18S535 allele frequency distributions between Wikwemikong Aboriginals and a population from Asturias in Northern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

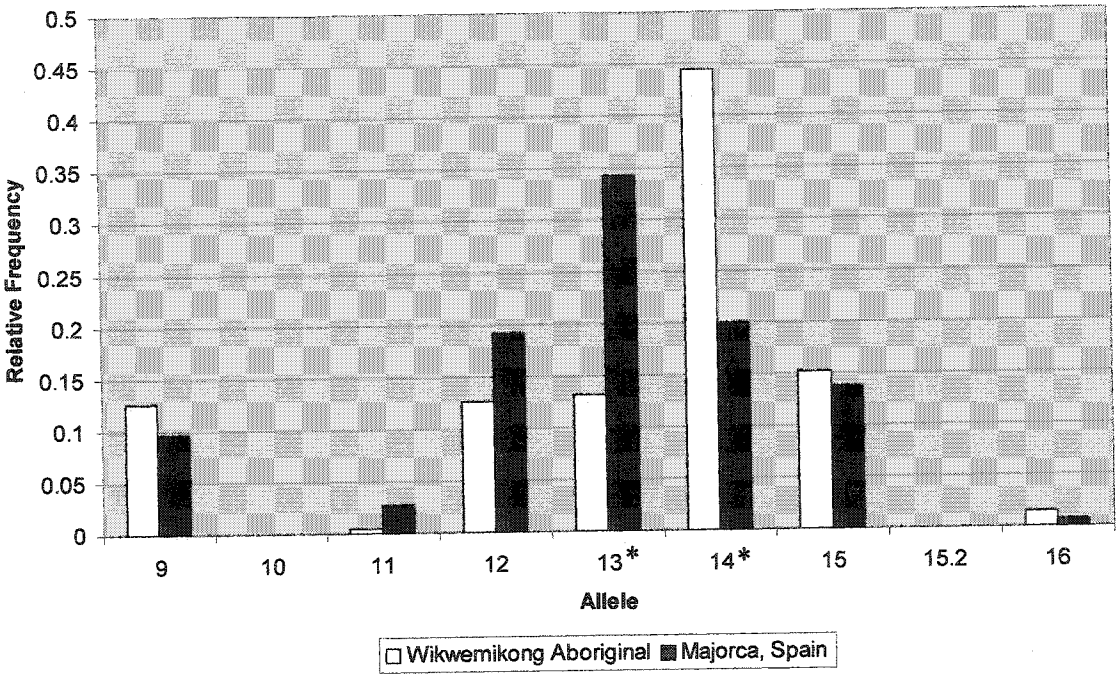


Figure 66. Pairwise comparison of the relative D18S535 allele frequency distributions between Wikwemikong Aboriginals and a population from the island of Majorca in Spain[* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

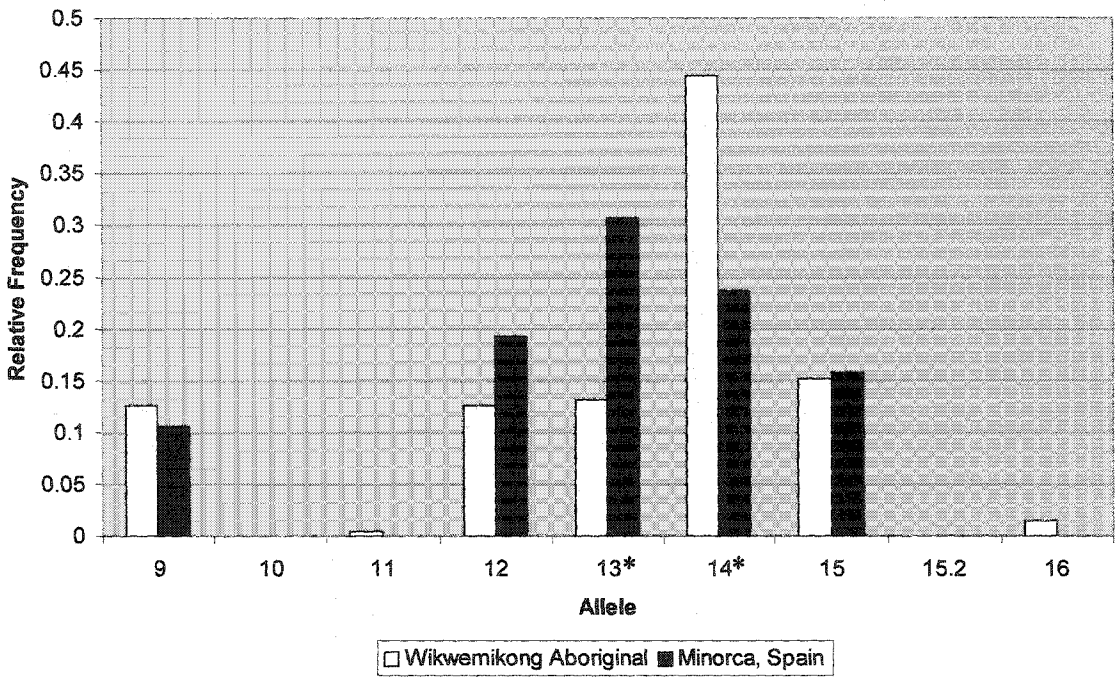


Figure 67. Pairwise comparison of the relative D18S535 allele frequency distributions between Wikwemikong Aboriginals and a population from the island of Minorca in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

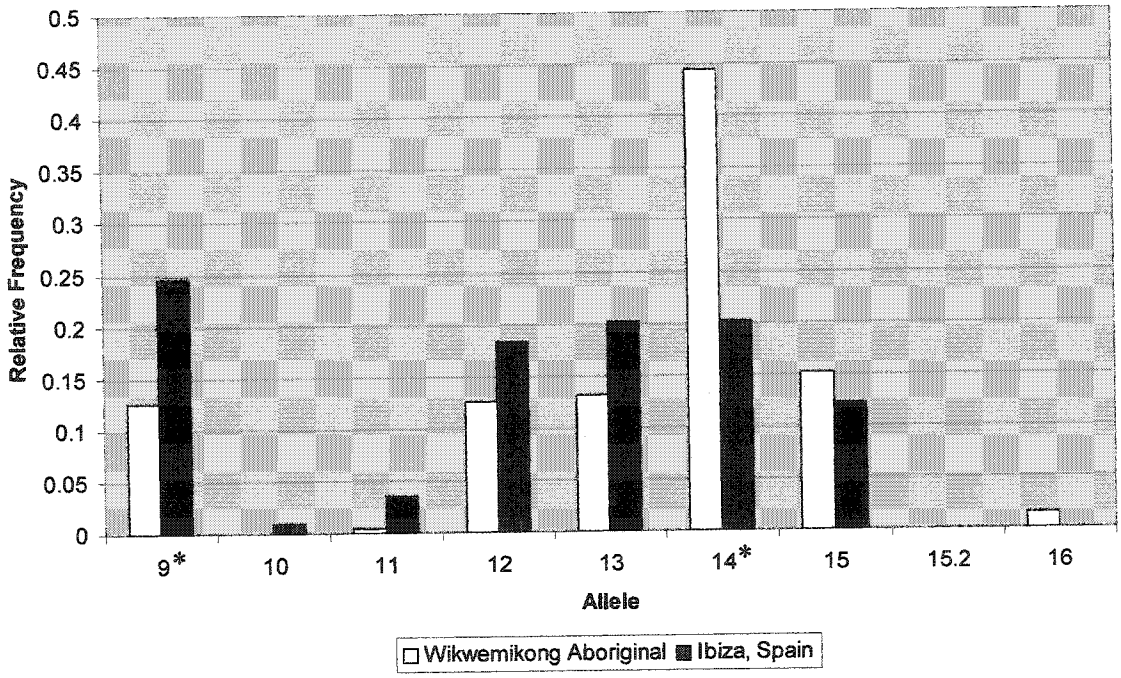


Figure 68. Pairwise comparison of the relative D18S535 allele frequency distributions between Wikwemikong Aboriginals and a population from the island of Ibiza in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

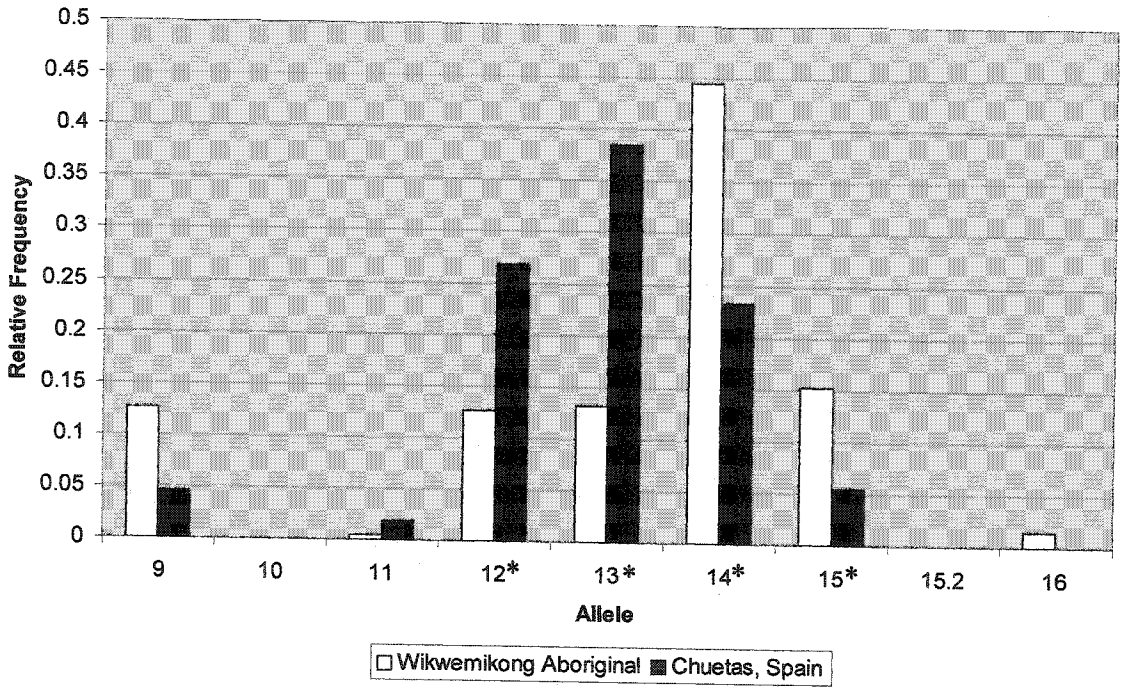


Figure 69. Pairwise comparison of the relative D18S535 allele frequency distributions between Wikwemikong Aboriginals and a Chuetas population from Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

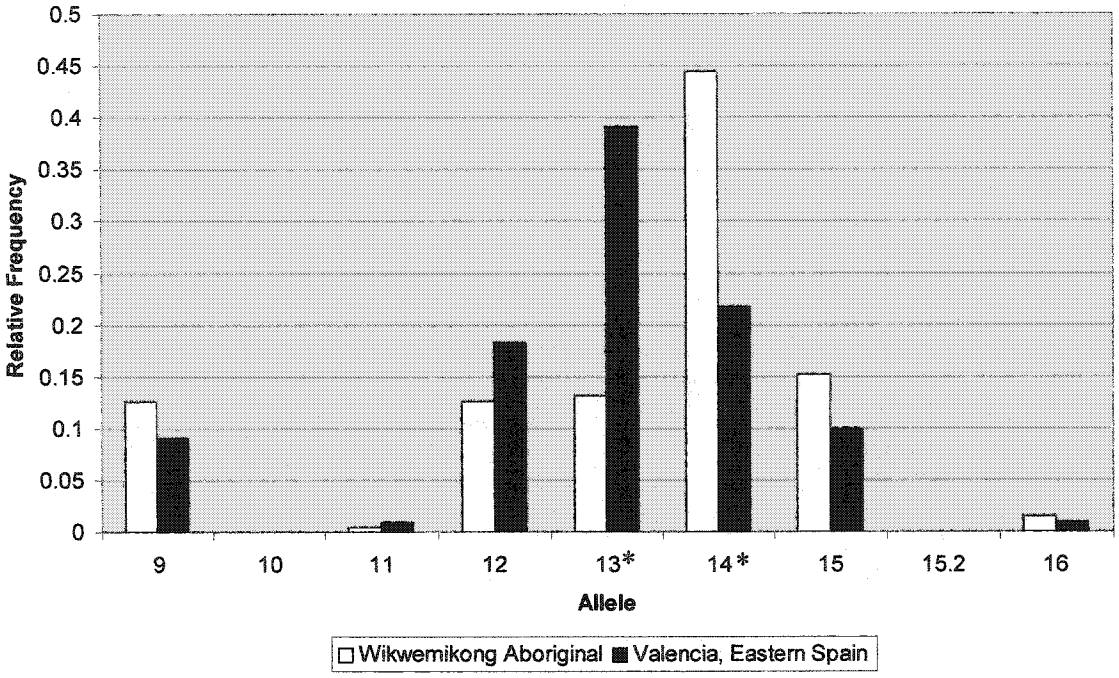


Figure 70. Pairwise comparison of the relative D18S535 allele frequency distributions between Wikwemikong Aboriginals and a population from Valencia in Eastern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

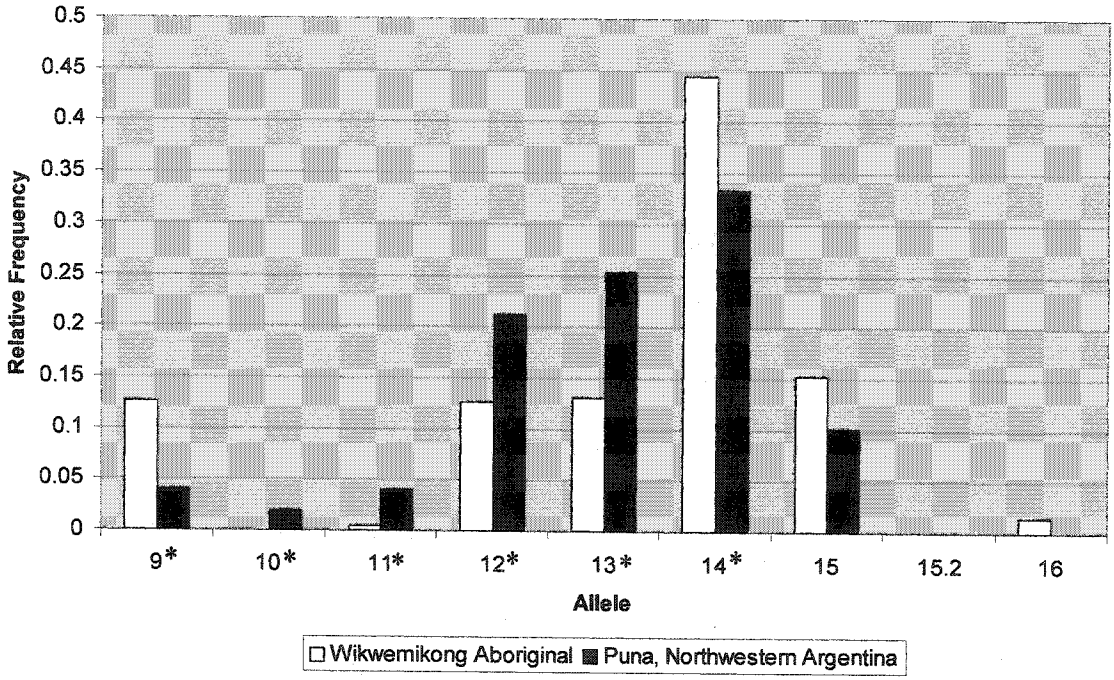


Figure 71. Pairwise comparison of the relative D18S535 allele frequency distributions between Wikwemikong Aboriginals and a population from Puna in Northwestern Argentina [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

from the Northern Ontario populations. All figures indicate where the allelic differences occurred.

A similar comparison was carried out between the previously studied populations from Europe and South America to identify the alleles which were the source of the pairwise population differences. The results of the pairwise comparisons are shown in Table 15 and Figures 72 to 95 display the allelic distributions between the populations which differed from the each other and indicates where the allelic differences occurred.

Pairwise comparisons of the Northern Ontario population samples were also carried out by means of F_{ST} values. F_{ST} estimates correspond to the amount divergence between populations. Populations are least divergent when F_{ST} is close to zero, most divergent the farther away from zero, and no divergence if F_{ST} is zero. Table 16 summarizes the information about F_{ST} estimates. F_{ST} values ranged from 0.0061-0.0562 which means that the amount divergence between the Northern Ontario populations was minimal. Pairwise F_{ST} comparisons showed that the Espanola Aboriginals and the Sioux Lookout Aboriginals were the least divergent of the four groups (F_{ST} = 0.0061). The Espanola Caucasians and Wikwemikong Aboriginals were the most divergent (F_{ST} = 0.0562), with the other populations showing an intermediate level of divergence (F_{ST} = 0.0145 - 0.0426).

A cluster analysis using euclidean distances measurement to calculate distance and average linkage between groups to form the clusters was performed based on D18S535 allele frequency distribution in order to determine how the four Northern Ontario populations along with the previously studied populations from Europe and South America clustered at this locus. Table 17 displays the euclidean dissimilarity coefficient matrix based on D18S535

Table 15. Allelic differences observed in pairwise comparisons between previously published D18S535 studies from Europe which demonstrated significant differences

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
CR & NEG	9	0.33	1	0.5636
	10	0.19	1	0.6606
	11	4.79	1	0.0286*
	12	0.34	1	0.5584
	13	0.32	1	0.5724
	14	0.03	1	0.8536
	15	0.69	1	0.4052
	16	4.56	1	0.0327*
CR & GA	9	0.09	1	0.7590
	10	1.11	1	0.2921
	11	5.88	1	0.0153*
	12	1.22	1	0.2702
	13	0.42	1	0.5161
	14	0.54	1	0.4634
	15	1.58	1	0.2095
	16	2.62	1	0.1053
CR & AS	9	0.33	1	0.5638
	10	1.74	1	0.1869
	11	12.04	1	0.0005*
	12	2.08	1	0.1491
	13	0.09	1	0.7700
	14	0.01	1	0.9153
	15	0.01	1	0.9304
	16	2.16	1	0.1418

Table 15. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
CR & IB	9	14.72	1	0.0001*
	10	0.01	1	0.9771
	11	0.64	1	0.4231
	12	0.14	1	0.7043
	13	4.20	1	0.0405*
	14	0.01	1	0.9814
	15	0.07	1	0.7940
	16	--	--	--
CR & CH	9	2.38	1	0.1226
	10	2.04	1	0.1533
	11	2.54	1	0.1110
	12	2.91	1	0.0875
	13	2.49	1	0.1144
	14	0.25	1	0.6156
	15	4.08	1	0.0434*
	16	--	--	--
CR & PU	9	6.03	1	0.0141*
	10	0.05	1	0.8200
	11	0.40	1	0.5292
	12	0.53	1	0.4659
	13	1.57	1	0.2108
	14	9.65	1	0.0019*
	15	0.12	1	0.7271
	16	--	--	--

Table 15. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
NEG & IB	9	17.73	1	0.0001*
	10	0.09	1	0.7643
	11	0.75	1	0.3871
	12	0.62	1	0.4310
	13	5.78	1	0.0161*
	14	0.02	1	0.8751
	15	0.12	1	0.7308
	16	1.56	1	0.2114
NEG & PU	9	3.90	1	0.0482*
	10	0.37	1	0.5409
	11	1.87	1	0.1712
	12	0.04	1	0.8385
	13	2.90	1	0.0886
	14	10.12	1	0.0015*
	15	1.09	1	0.2961
	16	2.63	1	0.1046
AR & IB	9	23.22	1	0.0001*
	10	0.36	1	0.5512
	11	0.08	1	0.7785
	12	0.71	1	0.3964
	13	2.58	1	0.1082
	14	1.13	1	0.2873
	15	0.13	1	0.7205
	16	1.12	1	0.2894

Table 15. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
GA & IB	9	11.31	1	0.0008*
	10	0.69	1	0.4045
	11	1.40	1	0.2359
	12	0.21	1	0.6498
	13	5.95	1	0.0147
	14	0.26	1	0.6077
	15	0.44	1	0.5088
	16	0.90	1	0.3436
GA & CH	9	2.91	1	0.0882
	10	0.88	1	0.3478
	11	0.02	1	0.8784
	12	6.22	1	0.0126*
	13	1.06	1	0.3022
	14	0.01	1	0.9545
	15	6.68	1	0.0098*
	16	0.88	1	0.3478
GA & PU	9	6.87	1	0.0088*
	10	1.39	1	0.2281
	11	2.82	1	0.0931
	12	2.77	1	0.0962
	13	3.06	1	0.0804
	14	5.18	1	0.0229*
	15	1.96	1	0.1616
	16	1.51	1	0.2185

Table 15. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
AS & IB	9	10.92	1	0.0010*
	10	1.01	1	0.2940
	11	4.91	1	0.0267*
	12	1.90	1	0.1679
	13	4.98	1	0.0257*
	14	0.01	1	0.9205
	15	0.04	1	0.8455
	16	0.74	1	0.3904
AS & PU	9	8.23	1	0.0041*
	10	2.07	1	0.1502
	11	7.15	1	0.0061*
	12	0.27	1	0.6012
	13	2.20	1	0.1382
	14	9.92	1	0.0016*
	15	0.18	1	0.6754
	16	1.25	1	0.2644
MA & IB	9	10.47	1	0.0012*
	10	2.57	1	0.1088
	11	0.12	1	0.7258
	12	0.12	1	0.7258
	13	5.53	1	0.0187*
	14	0.01	1	0.9332
	15	0.12	1	0.7280
	16	1.58	1	0.2088

Table 15. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
MA & PU	9	4.07	1	0.0436*
	10	3.06	1	0.0804
	11	0.48	1	0.4901
	12	0.33	1	0.5670
	13	2.80	1	0.0943
	14	6.51	1	0.0107*
	15	0.90	1	0.3431
	16	2.66	1	0.1026
MI & IB	9	7.73	1	0.0054*
	10	2.02	1	0.1555
	11	4.07	1	0.0437*
	12	0.12	1	0.7331
	13	3.24	1	0.0720
	14	0.40	1	0.5289
	15	0.58	1	0.4463
	16	--	--	--
MI & CH	9	2.06	1	0.1514
	10	--	--	--
	11	2.05	1	0.1519
	12	1.78	1	0.1827
	13	1.45	1	0.2286
	14	0.06	1	0.8118
	15	6.47	1	0.0110*
	16	--	--	--

Table 15. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
MI & PU	9	4.76	1	0.0291*
	10	2.40	1	0.1215
	11	4.86	1	0.0275*
	12	0.27	1	0.6026
	13	1.11	1	0.2913
	14	2.54	1	0.1109
	15	1.92	1	0.1655
	16	--	--	--
IB & CH	9	16.26	1	0.0001*
	10	1.98	1	0.1591
	11	0.65	1	0.4204
	12	2.78	1	0.0952
	13	8.85	1	0.0029*
	14	0.15	1	0.6971
	15	3.35	1	0.0670
	16	--	--	--
IB & VA	9	9.84	1	0.0017*
	10	1.98	1	0.1591
	11	0.65	1	0.4204
	12	0.01	1	0.9518
	13	8.85	1	0.0029*
	14	0.01	1	0.9460
	15	0.14	1	0.7118
	16	2.05	1	0.1519

Table 15. (Continued)

Populations	Allele	Pearson Chi-Square Value	Degrees of Freedom	*P*-value
IB & PU	9	28.72	1	0.0001*
	10	0.04	1	0.8440
	11	0.08	1	0.7786
	12	0.80	1	0.3704
	13	0.94	1	0.3333
	14	5.15	1	0.0232*
	15	0.26	1	0.6094
	16	--	--	--
CH & PU	9	0.23	1	0.6291
	10	2.36	1	0.1247
	11	1.25	1	0.2635
	12	0.97	1	0.3258
	13	5.84	1	0.0157*
	14	3.39	1	0.0655
	15	2.29	1	0.1302
	16	--	--	--
VA & PU	9	2.90	1	0.0886
	10	2.36	1	0.1247
	11	1.25	1	0.2635
	12	0.68	1	0.4100
	13	5.84	1	0.0157*
	14	4.77	1	0.0289*
	15	0.01	1	0.9284
	16	3.46	1	0.0628

Table 15. (Continued)
CR = Croatia; NEG = Halla Area, Northeastern Germany; AR = Aragon, Northeastern Spain; GA = Galacia, Northwestern Spain; AS = Asturias, Northern Spain; MA = Majorca, Spain; MI = Minorca, Spain; IB = Ibiza, Spain; CH = Chuetas, Spain; VA = Valencia, Eastern Spain; PU = Puna, Northwestern Argentina
* indicates a significant value ($p < 0.05$)
– statistics could not be computed since the number of non-empty rows or columns was one

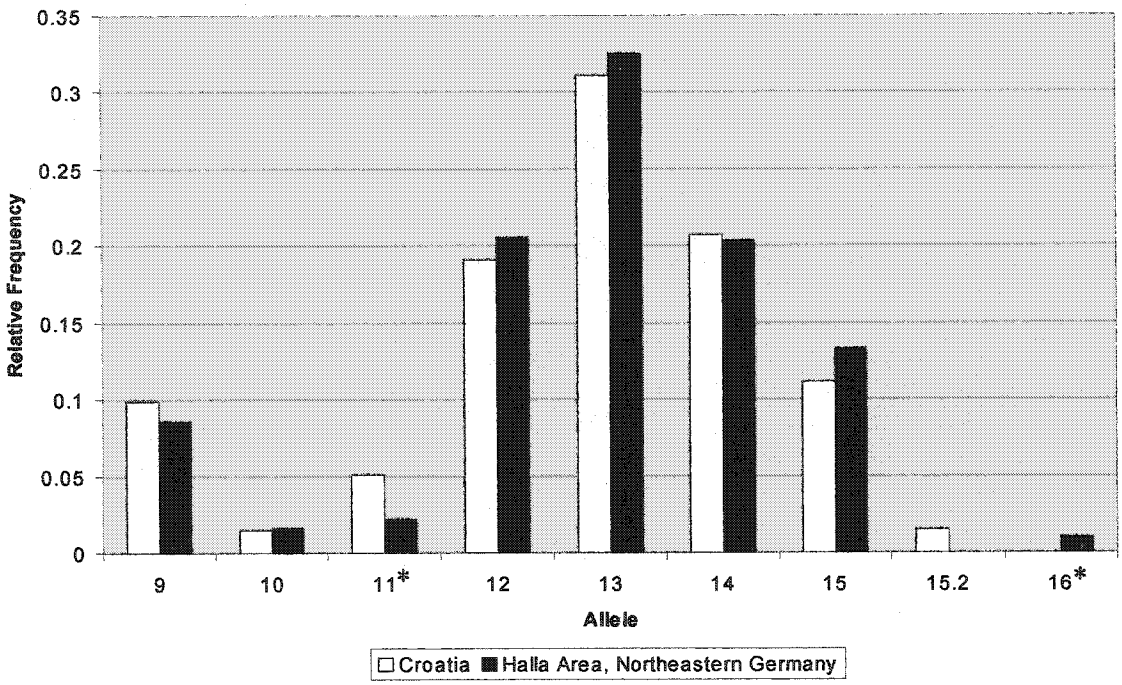


Figure 72. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from Croatia and a population from Halla Area in Northeastern Germany [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

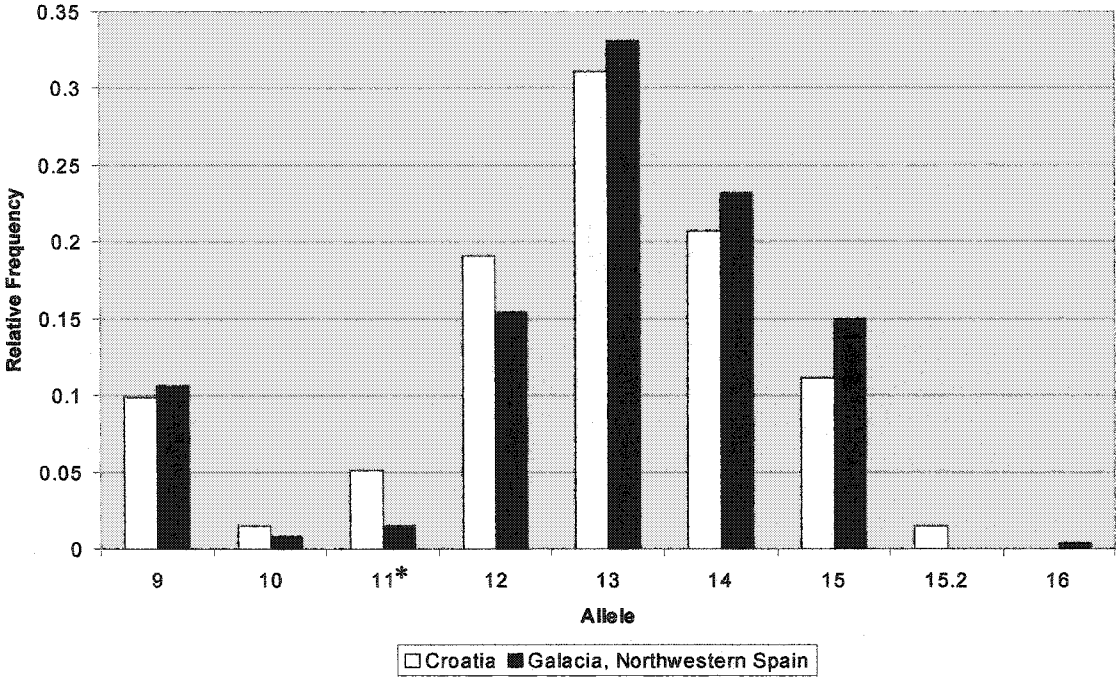


Figure 73. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from Croatia and a population from Galacia in Northwestern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

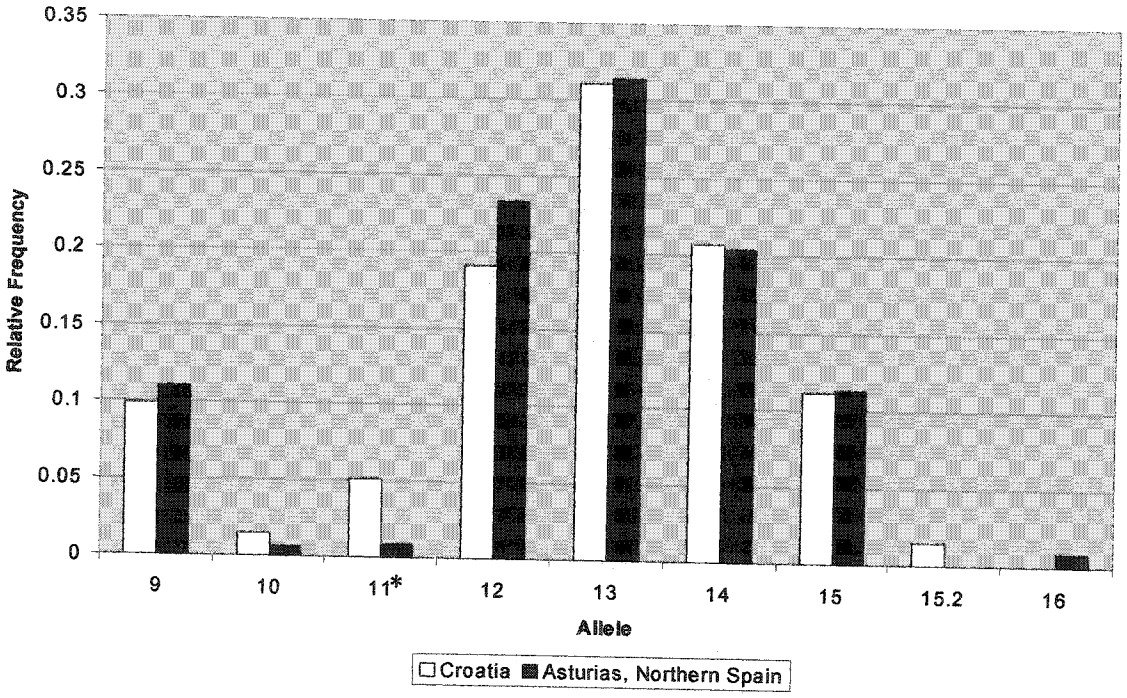


Figure 74. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from Croatia and a population from Asturias in Northern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

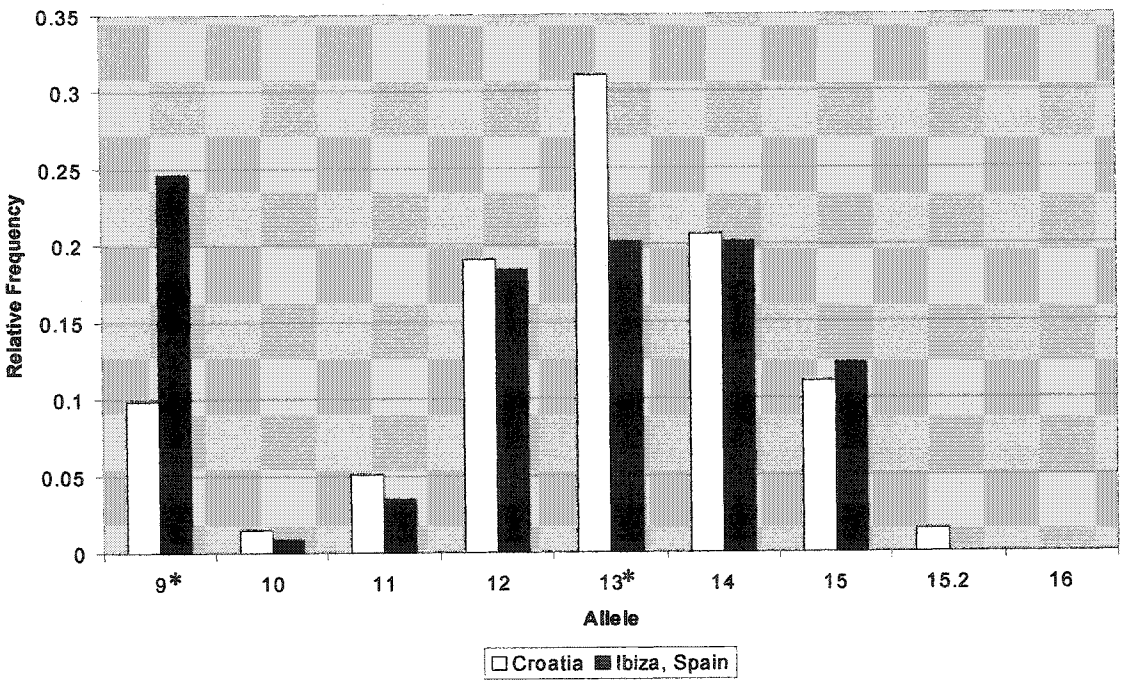


Figure 75. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from Croatia and a population from the island of Ibiza in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

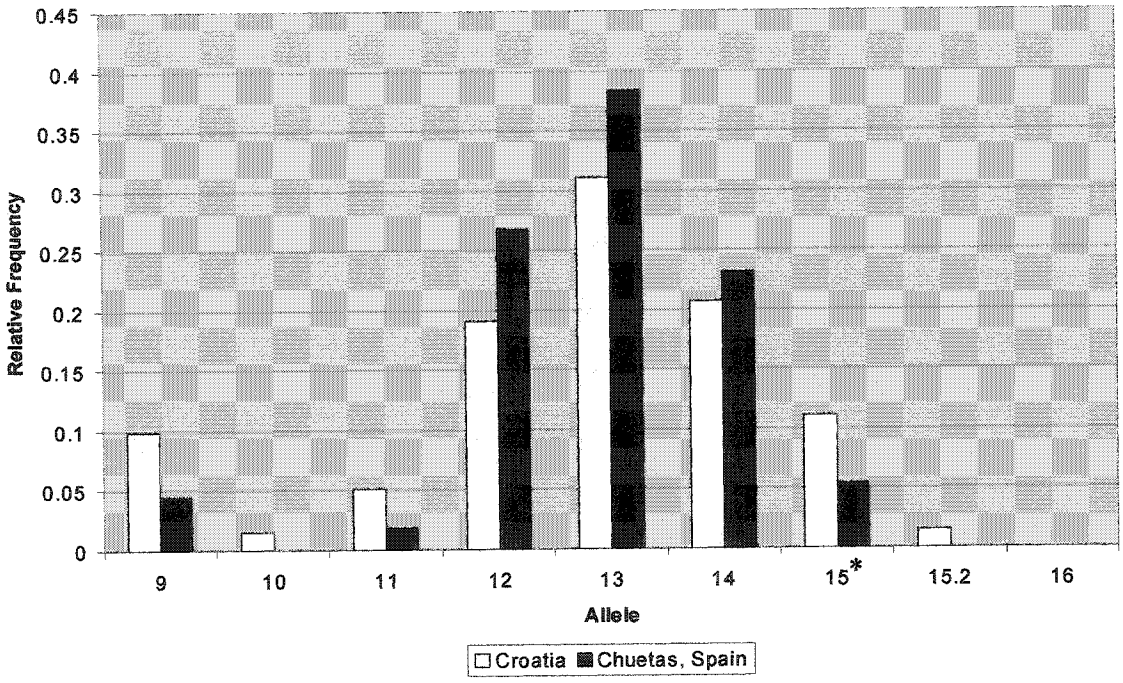


Figure 76. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from Croatia and a Chuetas population from Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

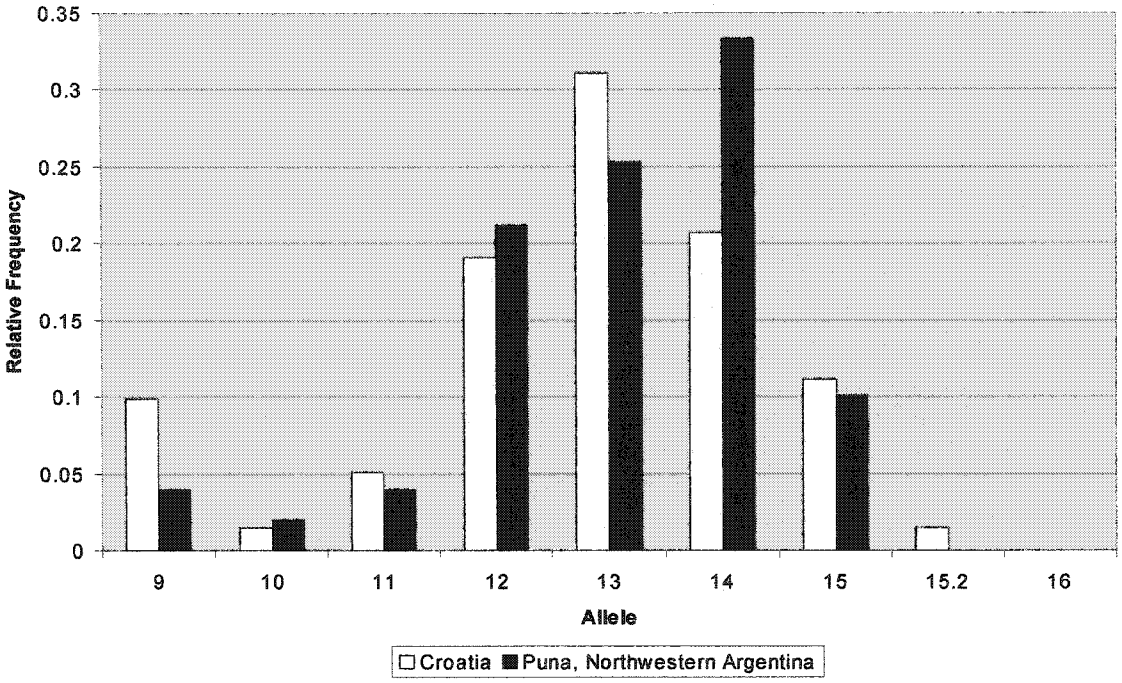


Figure 77. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from Croatia and a population from Puna in Northwestern Argentina [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

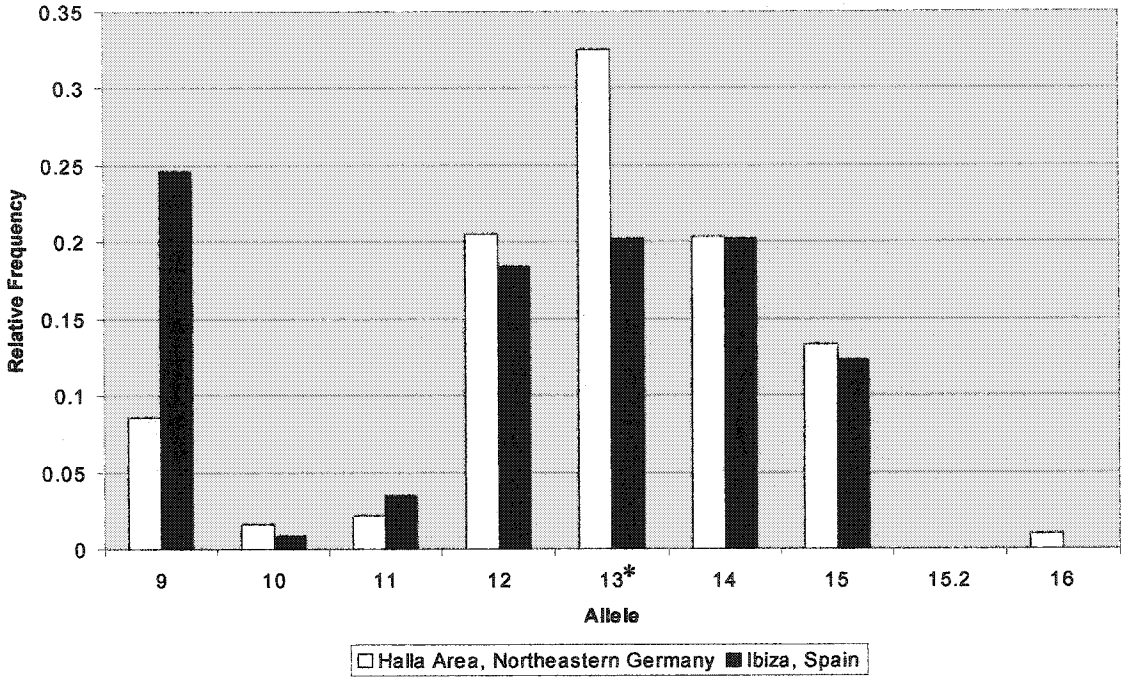


Figure 78. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from Halla Area in Northeastern Germany and a population from the island of Ibiza in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

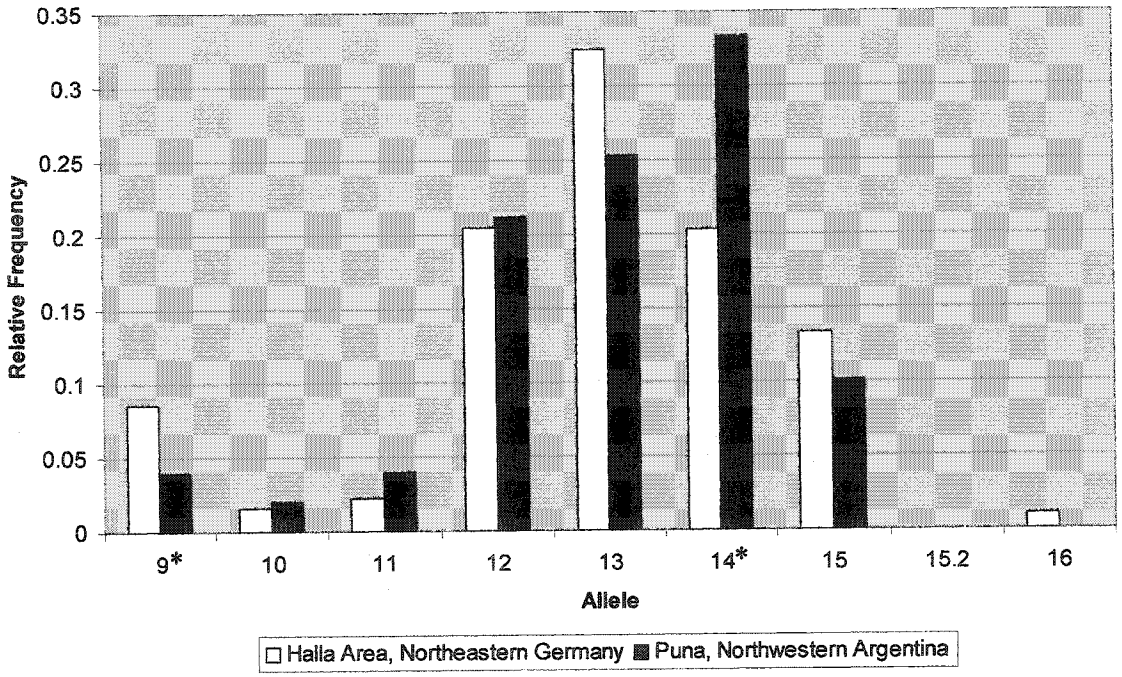


Figure 79. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from Halla Area in Northeastern Germany and a population from Puna in Northwestern Argentina [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

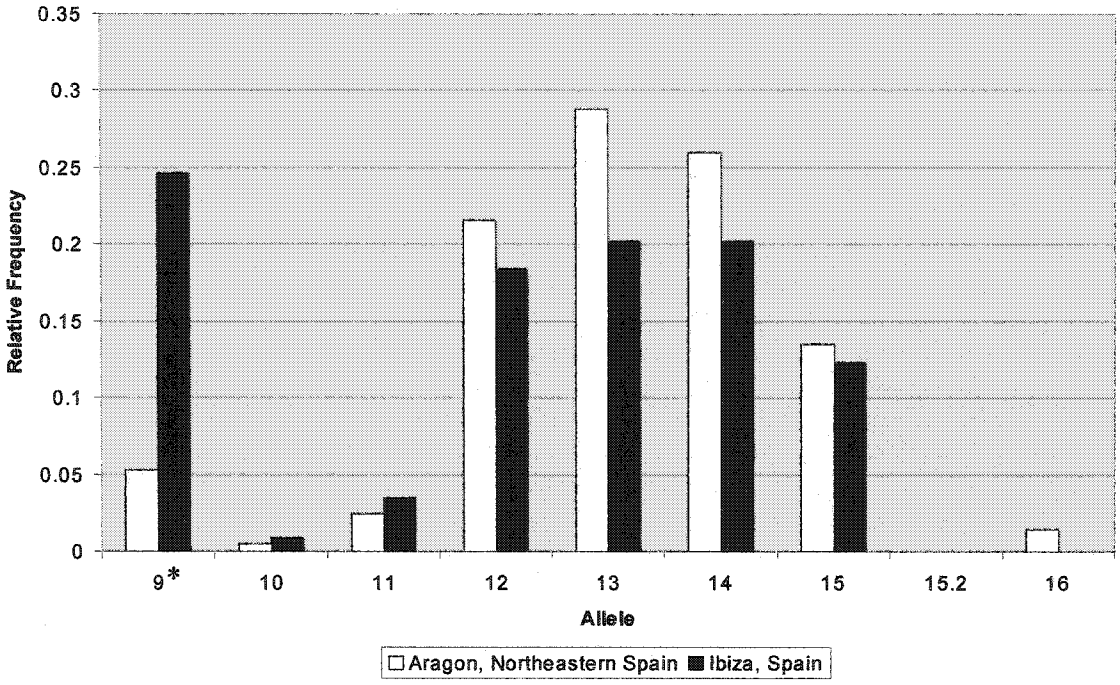


Figure 80. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from Aragon in Northeastern Spain and a population from the island of Ibiza in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

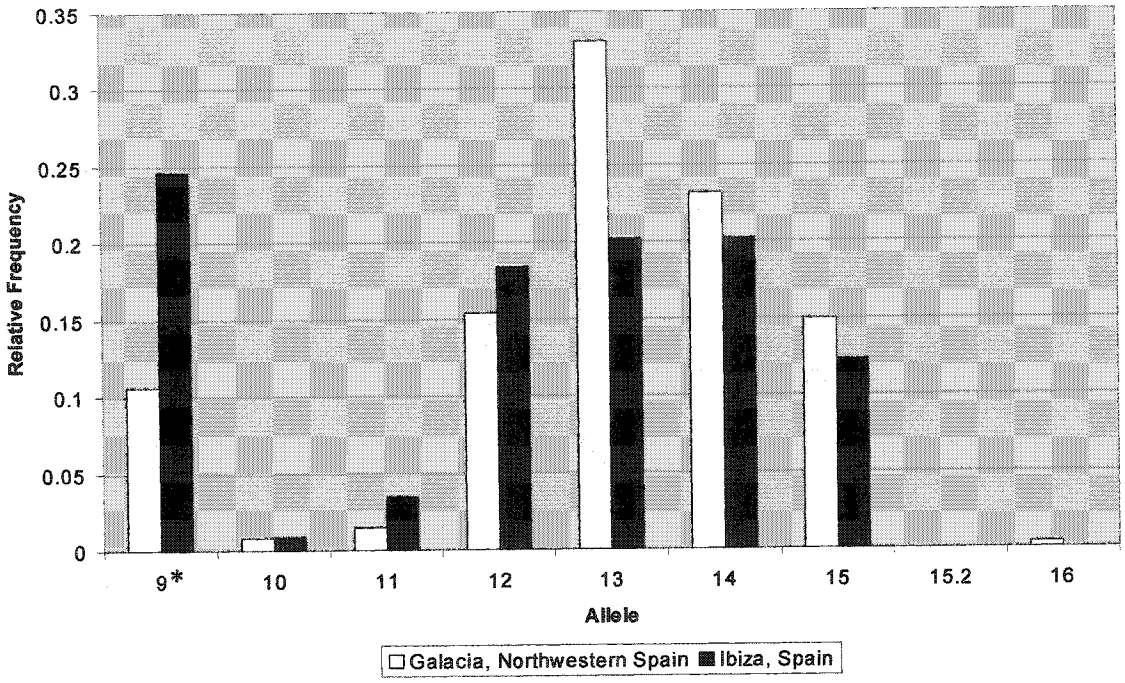


Figure 81. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from Galacia in Northwestern Spain and a population from the island of Ibiza in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

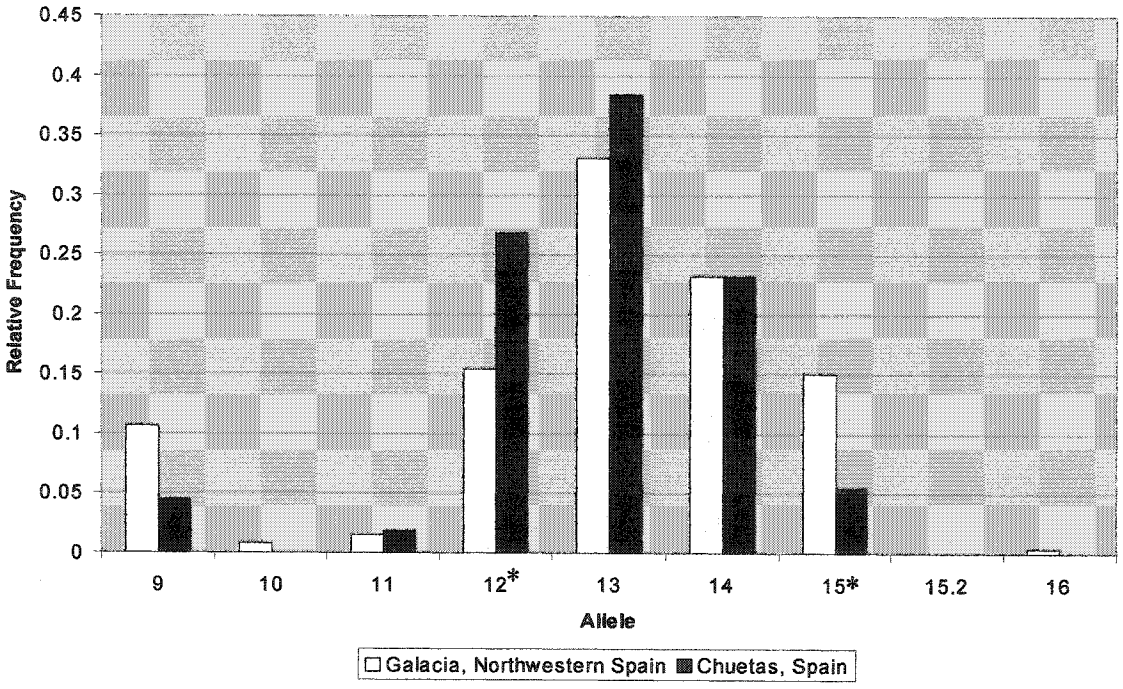


Figure 82. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from Galacia in Northwestern Spain and a Chuetas population from Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

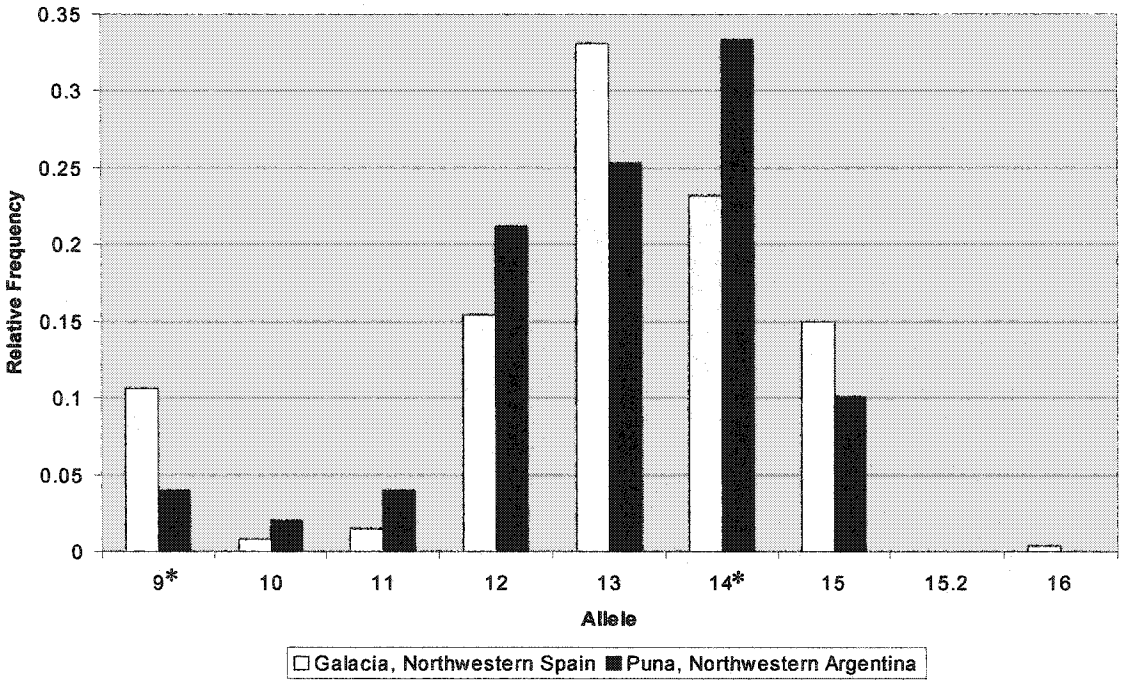


Figure 83. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from Galacia in Northwestern Spain and a population from Puna in Northwestern Argentina [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

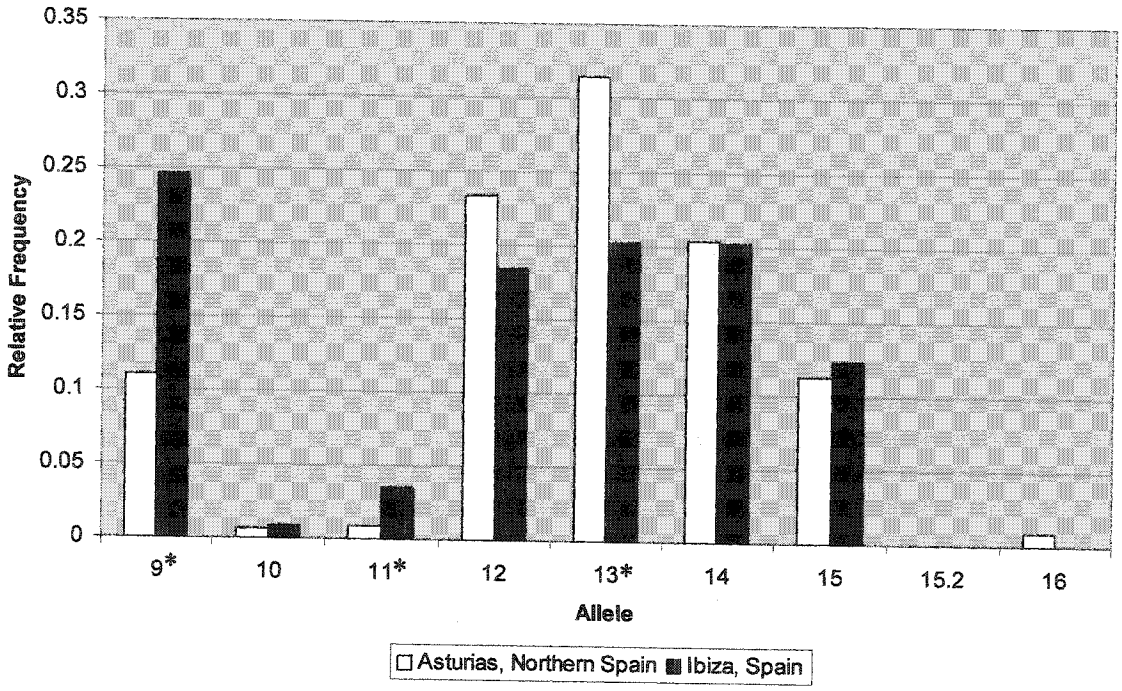


Figure 84. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from Asturias in Northern Spain and a population from the island of Ibiza in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

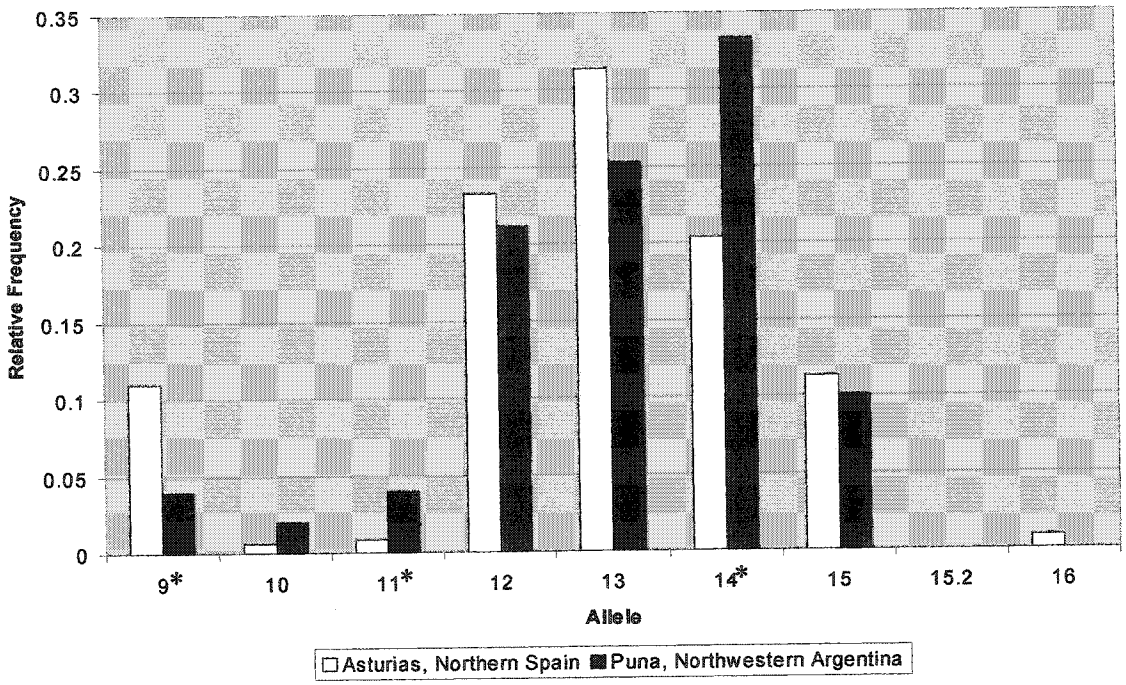


Figure 85. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from Asturias in Northern Spain and a population from Puna in Northwestern Argentina [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

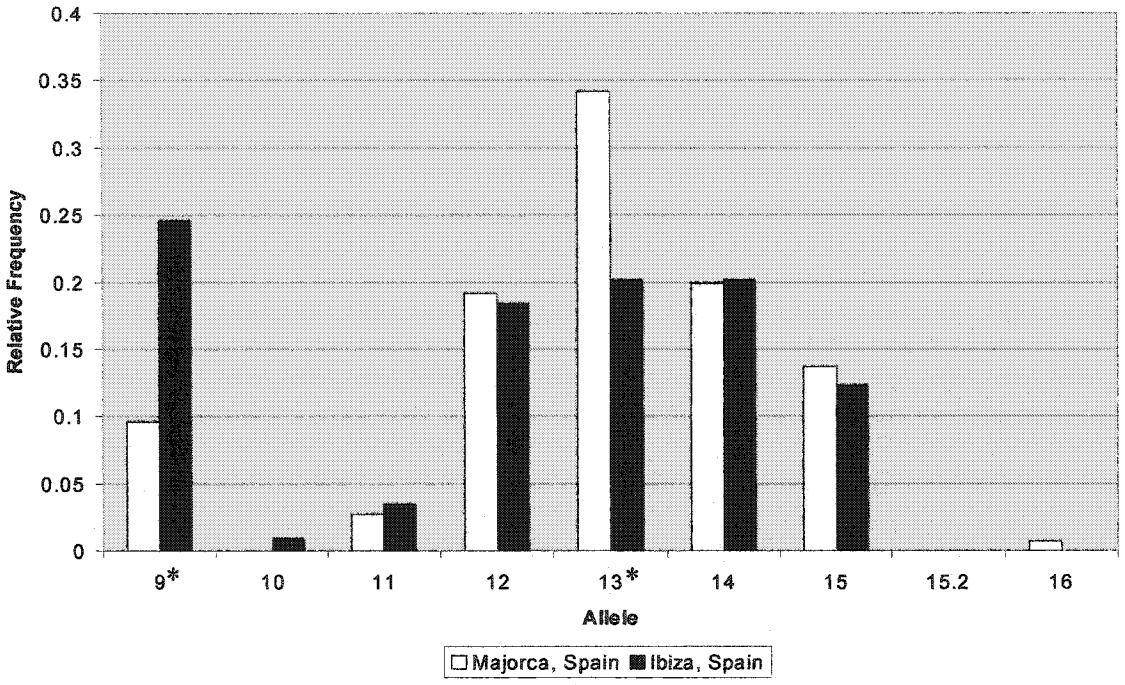


Figure 86. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from the island of Majorca in Spain and a population from the island of Ibiza in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

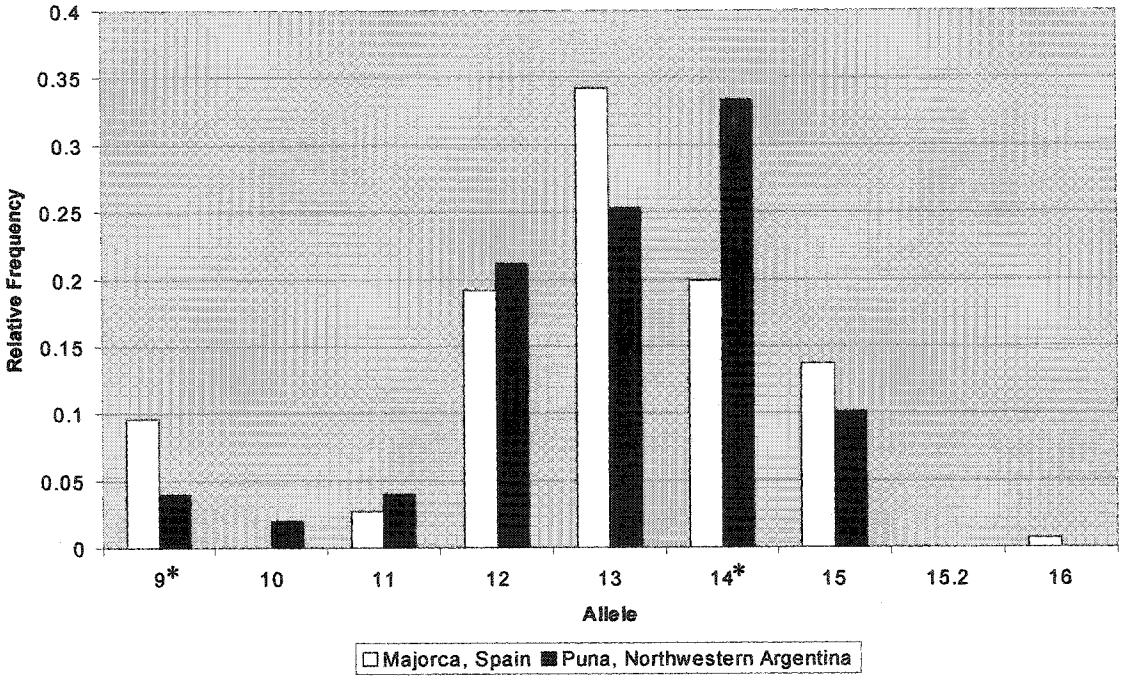


Figure 87. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from the island of Majorca in Spain and a population from Puna in Northwestern Argentina [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

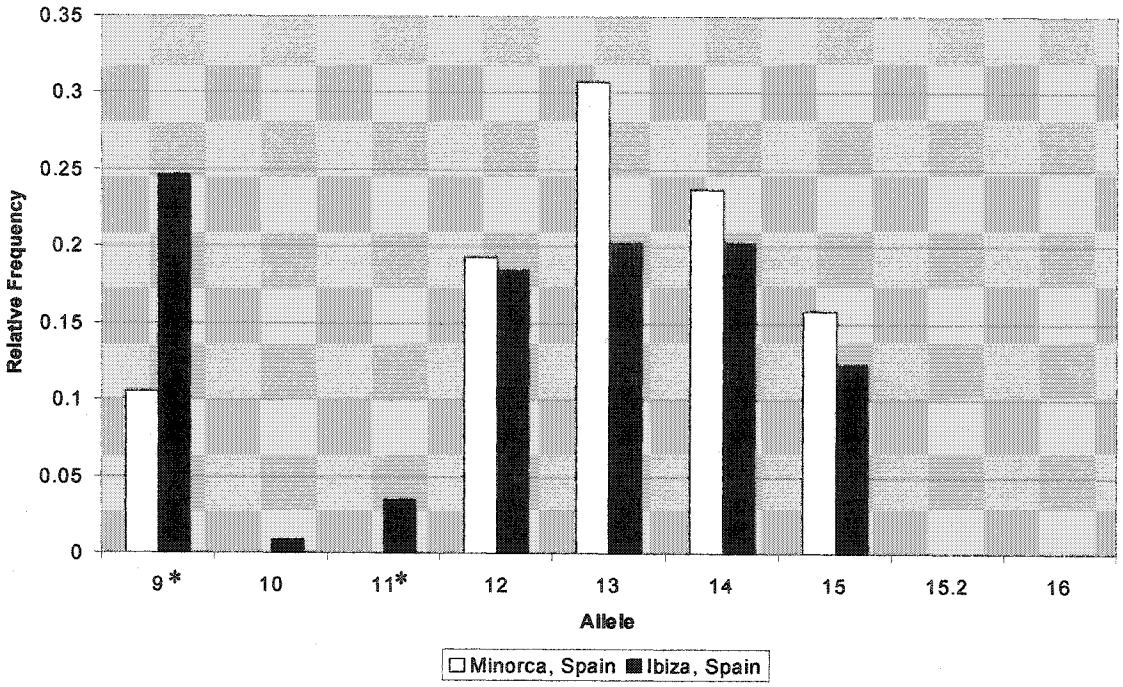


Figure 88. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from the island of Minorca in Spain and a population from the island of Ibiza in Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

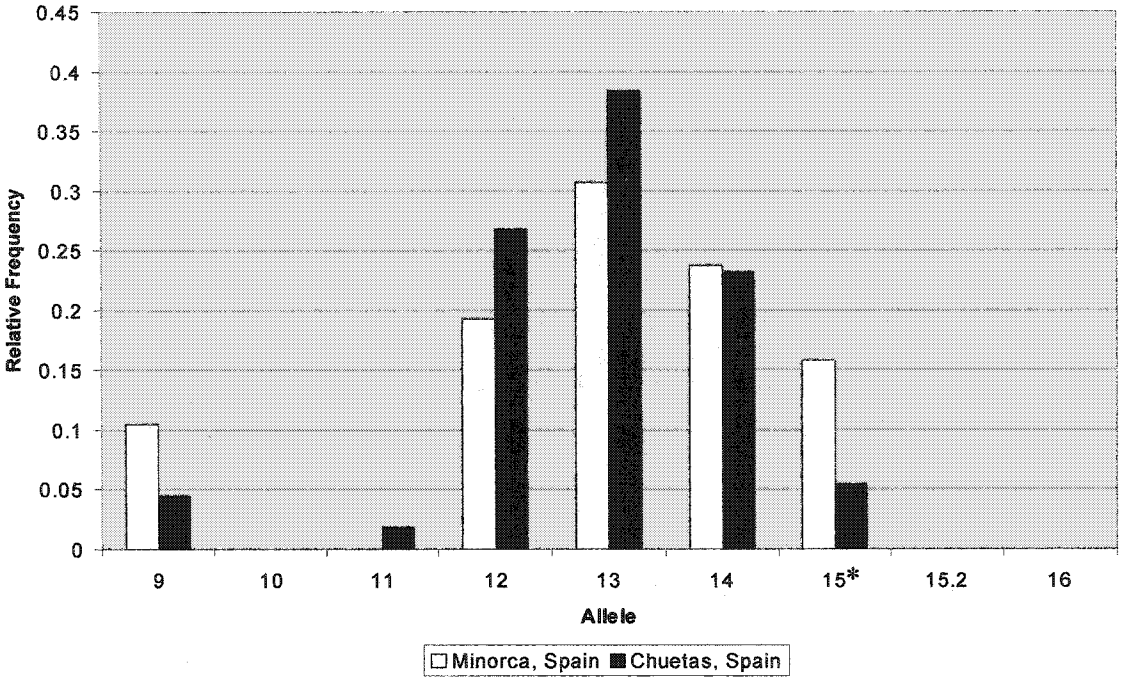


Figure 89. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from the island of Minorca in Spain and a Chuetas population from Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

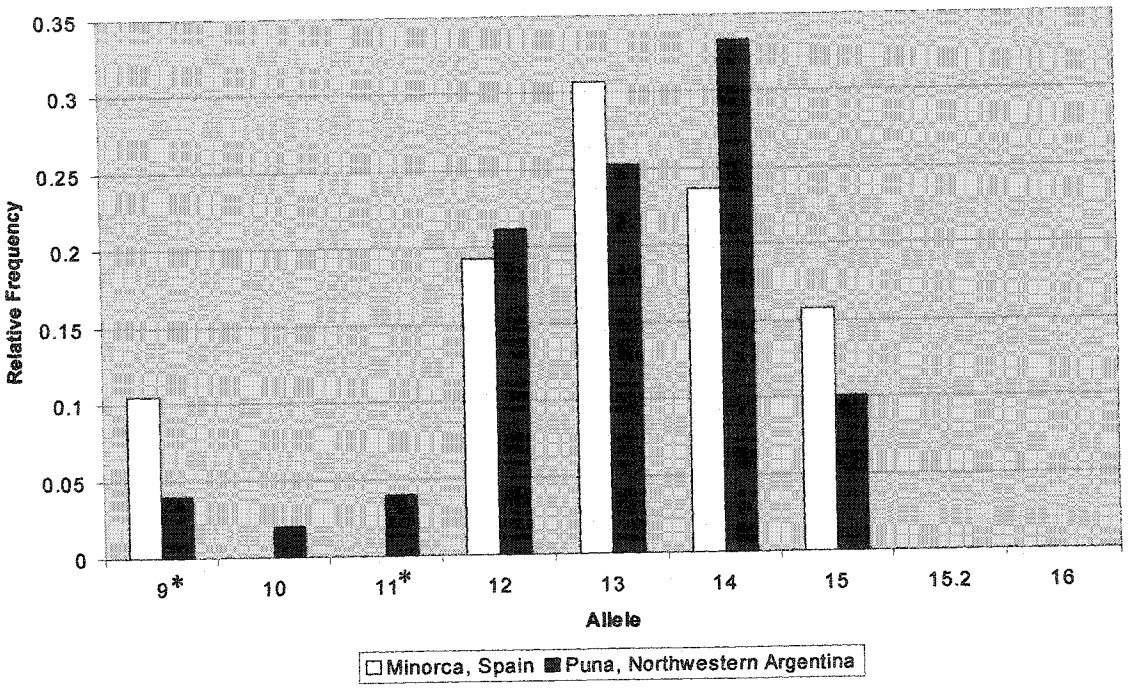


Figure 90. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from the island of Minorca in Spain and a population from Puna in Northwestern Argentina [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

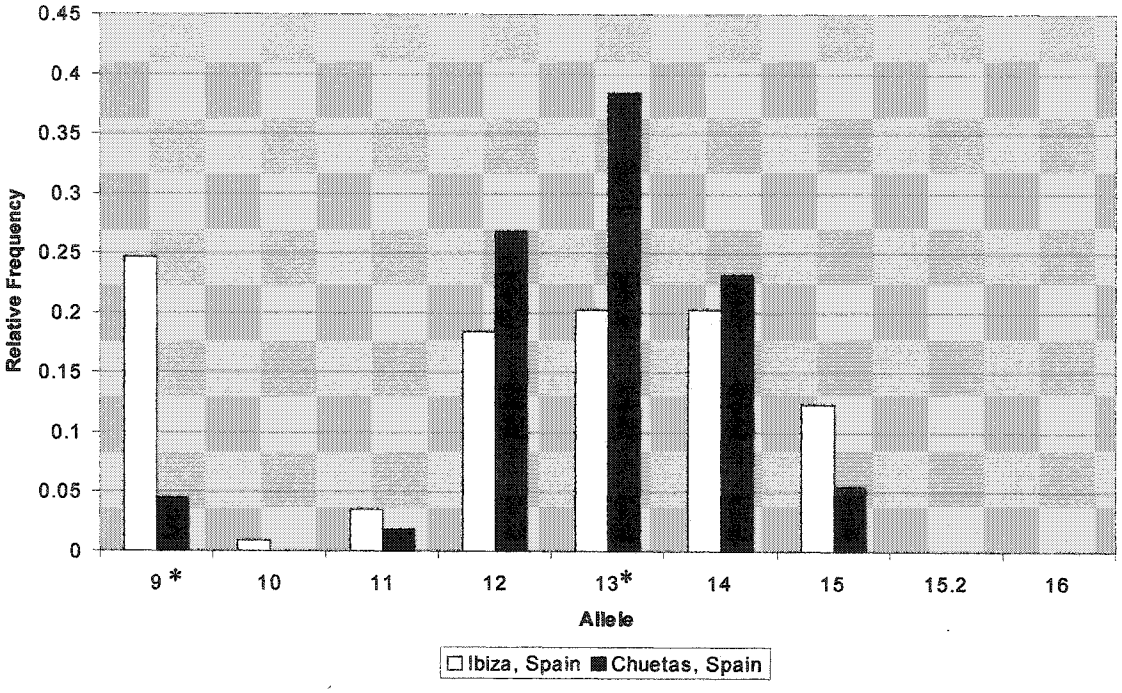


Figure 91. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from the island of Ibiza in Spain and a Chuetas population from Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

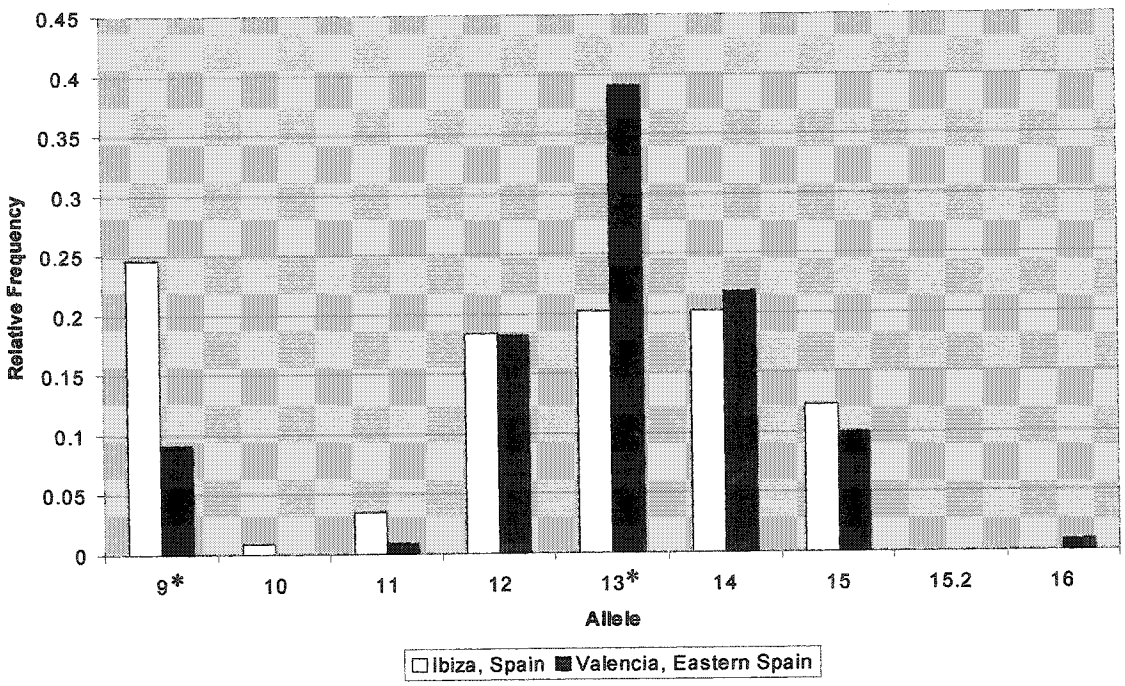


Figure 92. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from the island of Ibiza in Spain and a population from Valencia in Eastern Spain [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

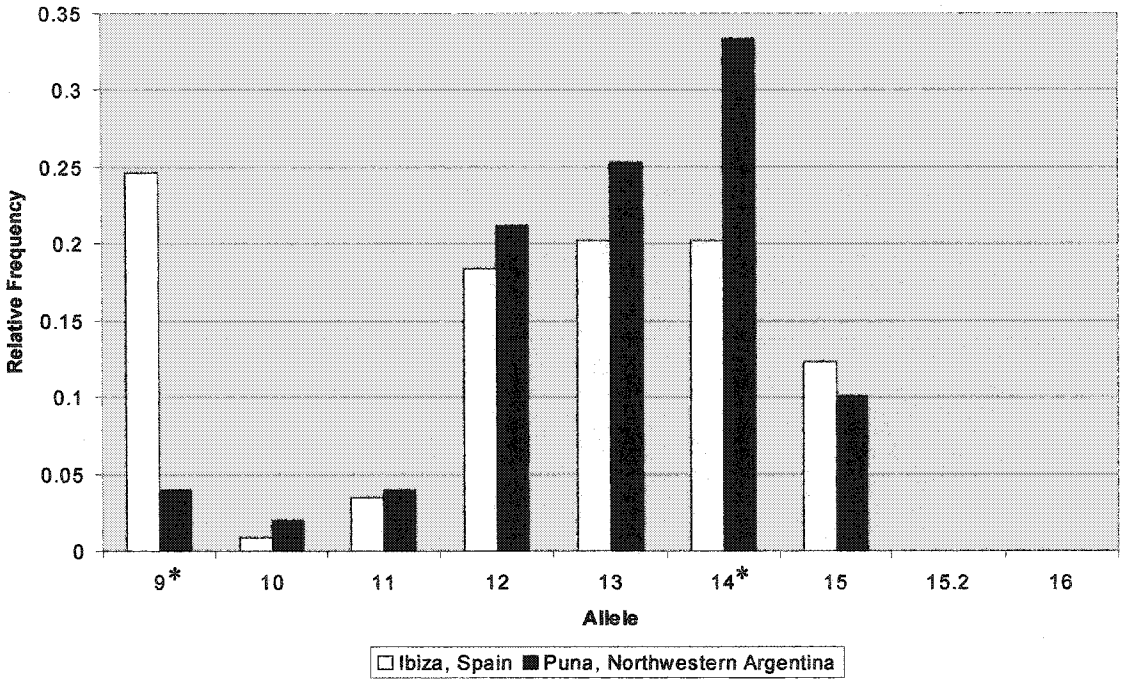


Figure 93. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from the island of Ibiza in Spain and a population from Puna in Northwestern Argentina [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

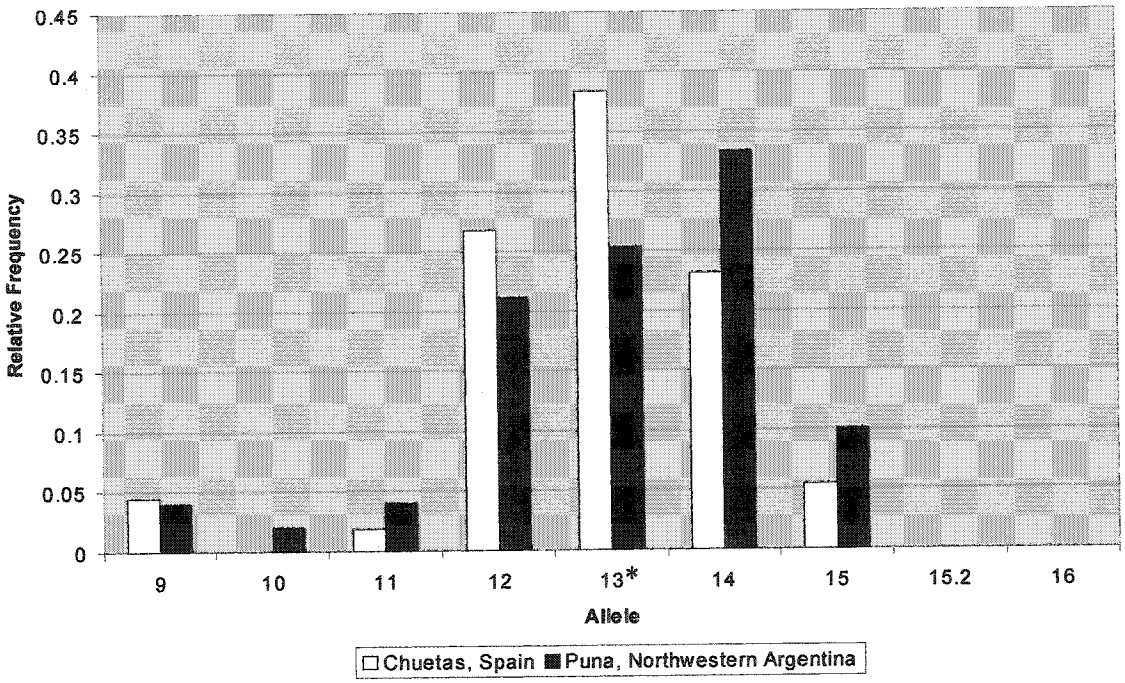


Figure 94. Pairwise comparison of the relative D18S535 allele frequency distributions between a Chuetas population from Spain and a population from Puna in Northwestern Argentina [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

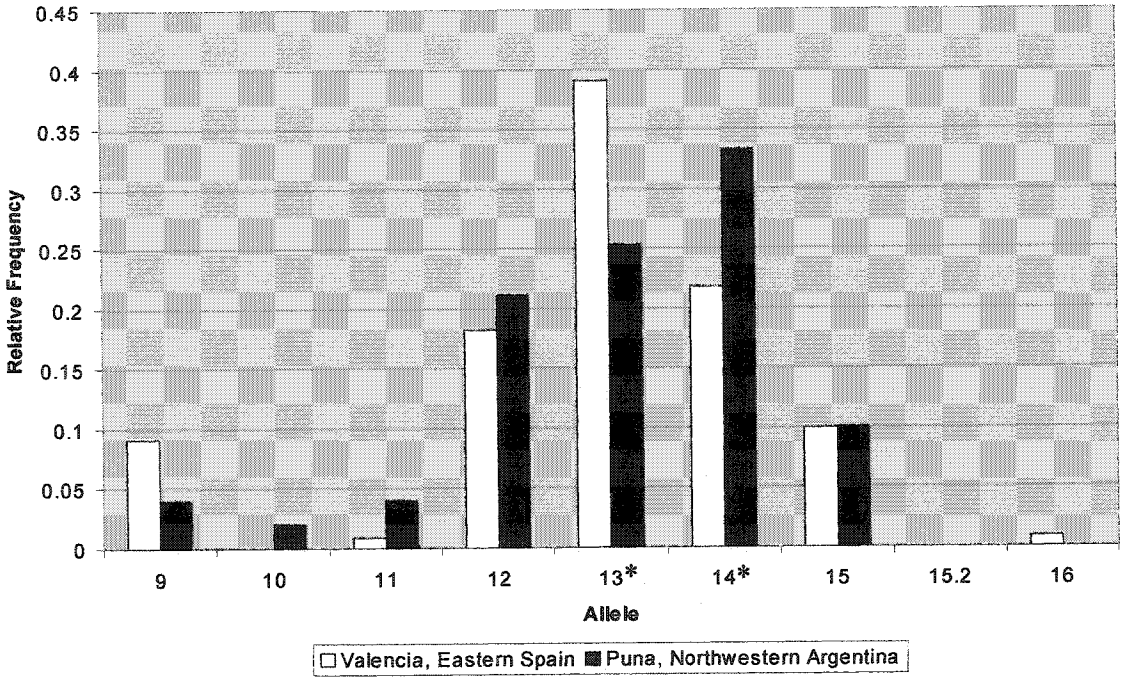


Figure 95. Pairwise comparison of the relative D18S535 allele frequency distributions between a population from Valencia in Spain and a population from Puna in Northwestern Argentina [* indicates a significant difference ($p < 0.05$) using a 2 x 2 contingency table and Pearson's chi-square value]

Table 16. Wright's F statistic, F_{ST}, at the D18S535 locus for Caucasian and Aboriginal populations from Northern Ontario

Population	Espanola Caucasians	Espanola Aboriginals	Sioux Lookout Aboriginals
Espanola Aboriginals	0.0236		
Sioux Lookout Aboriginals	0.0426	0.0061	
Wikwemikong Aboriginals	0.0562	0.0262	0.0145

Table 17. Euclidean dissimilarity coefficient matrix based on relative D18S535 allele frequency distributions in Caucasian and Aboriginal populations from Northern Ontario and with previously published D18S535 studies from Europe and South America

	EC	EA	SLA	WA	CR	NEG	AR	GA	AS	MA	MI	IB	CH	VA
EA	.211													
SLA	.277	.127												
WA	.315	.216	.171											
CR	.163	.257	.257	.310										
NEG	.166	.258	.272	.327	.052									
AR	.161	.187	.196	.267	.083	.078								
GA	.204	.271	.264	.295	.072	.067	.094							
AS	.132	.251	.281	.325	.069	.047	.093	.093						
MA	.184	.271	.277	.327	.053	.027	.088	.051	.062					
MI	.175	.237	.244	.280	.079	.061	.069	.048	.073	.060				
IB	.175	.306	.311	.282	.174	.200	.212	.187	.181	.199	.184			
CH	.203	.274	.309	.382	.141	.124	.143	.164	.118	.131	.157	.285		
VA	.223	.302	.307	.354	.093	.075	.124	.078	.098	.060	.101	.238	.111	
PU	.187	.133	.134	.219	.147	.158	.085	.161	.162	.166	.139	.243	.180	.188

EC = Espanola Caucasians; EA = Espanola Aboriginals; SLA = Sioux Lookout Aboriginals; WA = Wikwemikong Aboriginals; CR = Croatia; NEG = Halla Area , Northeastern Germany; AR = Aragon, Northeastern Spain; GA = Galacia, Northwestern Spain; AS = Asturias, Northern Spain; MA = Majorca, Spain; MI = Minorca, Spain; IB = Ibiza, Spain; CH = Chuetas, Spain; VA = Valencia, Eastern Spain; PU = Puna, Northwestern Argentina

Note: the larger the coefficient, the greater the distance

allelic distribution. The Euclidean dissimilarity coefficient matrix indicated that there were relatively large distances among the four Northern Ontario populations and that the shortest distances were typically among the European populations. Table 18 indicates the formation of the clusters based on the euclidean distances. The results indicated that two major clusters were formed based on D18S535 allele frequency distribution and that there was no significant clustering beyond cluster 2. One cluster consisted of the Espanola Aboriginals, Sioux Lookout Aboriginals, and Wikwemikong Aboriginals. The other cluster consisted of Espanola Caucasians, Croatia, Halla Area of Northeastern Germany, Aragon, Galacia, Asturias, Majorca, Minorca, Ibiza, Chuetas, Valencia, and the Puna population.

Contingency tables and chi-square analyses were used to determine if the clusters formed in the cluster analysis were affected by geography. The geographic locations tested were Northern Ontario, South America, and Europe. The Northern Ontario populations consisted of Espanola Caucasians, Espanola Aboriginals, Sioux Lookout Aboriginals, and Wikwemikong Aboriginals, and the South American population consisted of the Puna population in Northwestern Argentina. The European populations consisted of Croatia, the Halla Area in Northeastern Germany, Aragon in Northeastern Spain, Galacia in Northwestern Spain, Asturias in Northern Spain, Valencia in Eastern Spain, Majorca, Minorca, Ibiza, Chuetas, and Puna in Northwestern Argentina. The results demonstrated that there was a significant effect of geography on clustering with respect to (Pearson chi-square = 10.31, d.f. = 2, $p = 0.006$). Cluster 2 showed that the three Northern Ontario Aboriginal populations clustered by themselves while the Espanola Caucasians, and previously published D18S535 studies from Europe and South America formed another cluster.

Table 18. Cluster analysis formation of Caucasian and Aboriginal populations from Northern Ontario along with previously published D18S535 studies from Europe and South America based on relative D18S535 allele frequency distributions

Population	Geography	Ethnicity	Cluster 2*	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
EC	N. Ontario	C	1	1	1	1	1	1	1
EA	N. Ontario	A	2	2	2	2	2	2	2
SLA	N. Ontario	A	2	2	2	2	2	3	3
WA	N. Ontario	A	2	2	3	3	3	4	4
CR	Europe	C	1	3	4	4	4	5	5
NEG	Europe	C	1	3	4	4	4	5	5
AR	Europe	C	1	3	4	4	4	5	6
GA	Europe	C	1	3	4	4	4	5	5
AS	Europe	C	1	3	4	4	4	5	5
MA	Europe	C	1	3	4	4	4	5	5
MI	Europe	C	1	3	4	4	4	5	5
IB	Europe	C	1	1	1	5	5	6	7
CH	Europe	C	1	3	4	4	6	7	8
VA	Europe	C	1	3	4	4	4	5	5
PU	S. America	A	1	3	4	4	4	5	6

Table 18. (Continued)

Population	Geography	Ethnicity	Cluster 9	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14
EC	N. Ontario	C	1	1	1	1	1	1
EA	N. Ontario	A	2	2	2	2	2	2
SLA	N. Ontario	A	3	3	3	3	3	3
WA	N. Ontario	A	4	4	4	4	4	4
CR	Europe	C	5	5	5	5	5	5
NEG	Europe	C	5	5	5	5	6	6
AR	Europe	C	6	6	6	6	7	7
GA	Europe	C	5	5	7	7	8	8
AS	Europe	C	5	5	5	8	9	9
MA	Europe	C	5	5	5	5	6	6
MI	Europe	C	5	5	7	7	8	10
IB	Europe	C	7	7	8	9	10	11
CH	Europe	C	8	8	9	10	11	12
VA	Europe	C	5	9	10	11	12	13
PU	S. America	A	9	10	11	12	13	14

EC = Espanola Caucasians; EA = Espanola Aboriginals; SLA = Sioux Lookout Aboriginals; WA = Wikwemikong Aboriginals; CR = Croatia; NEG = Halla Area , Northeastern Germany; AR = Aragon, Northeastern Spain; GA = Galacia, Northwestern Spain; AS = Asturias, Northern Spain; MA = Majorca, Spain; MI = Minorca, Spain; IB = Ibiza, Spain; CH = Chuetas, Spain; VA = Valencia, Eastern Spain; PU = Puna, Northwestern Argentina; * indicates a significant value ($p < 0.05$); C = Caucasian; A = Aboriginal

Contingency tables and chi-square analyses were also used to determine if the clusters formed in the cluster analysis were affected by ethnicity. The ethnic categories were Caucasian and Aboriginals. The Caucasians included the Espanola Caucasians and all the previously published D18S535 studies from Europe, while the Aboriginals included all three Northern Ontario Aboriginal populations along with the Puna population from Northwestern Argentina (for details about the Aboriginal background of the Puna population see Albeza *et al.*, 2002). The results demonstrated that there was a significant effect of ethnicity on clustering (Pearson chi-square = 10.31, d.f. = 1, $p = 0.013$). Cluster 2 showed that the three Northern Ontario Aboriginal populations clustered by themselves while the Espanola Caucasians and previously published populations from Europe and South America formed another cluster.

The power of discrimination and the chance of exclusion are statistical parameters of forensic interest which reveal the usefulness of the D18S535 for personal identification and the results are shown in Table 3. The power of discrimination was high overall, ranging from 0.85 in the Espanola Aboriginal population to 0.92 in the Espanola Caucasian population. The chance of exclusion was also relatively high overall, ranging from 0.51 in the Wikwemikong Aboriginal population to 0.67 in the Espanola Aboriginal population.

The maximum number of bands at the D18S535 locus for an individual is two, one from each parent. One band observed indicated a homozygote and two bands indicated a heterozygote. Examples of the different bands are shown in Figures 4 to 8. The observed heterozygosity values (h) for all the populations are shown in Table 3. Heterozygosities also provide an indication of the usefulness of a locus for personal identification and were

consistently high in all populations, ranging from 0.75 in the Wikwemikong Aboriginal population to 0.84 in the Espanola Aboriginal population.

4. Discussion

4.1 Samples

Aboriginal and Caucasian populations were chosen as populations to study the D18S535 locus in because they were two distinct, representative, and forensically relevant populations of Northern Ontario. Canada has an abundant Aboriginal population which form well-characterized populations based on linguistics, anthropology, and geography (Frégeau *et al.*, 1998). The Aboriginal populations used in this study were representative of the Aboriginal populations in Northern Ontario based on geographical locale, culture, and linguistic affiliations. The choice of the Aboriginal communities also depended on several factors such as contrasting environments and different amounts of Caucasian contact, both genetically and culturally. Espanola, Wikwemikong, and Sioux Lookout Aboriginals represent dissimilar environments and varying degrees of Caucasian contact. It was important to choose contrasting Aboriginal subpopulations in order to determine the extent of genetic differentiation; therefore, determine whether the Aboriginal subpopulations could be amalgamated into a single database. Although no data was collected other than the ethnicity of the samples obtained, Espanola and Wikwemikong Aboriginal samples are most likely members of the Ojibwa subgroup while the Aboriginal samples from Sioux Lookout most likely represent members of both Ojibwa and Cree subgroups. The Ojibwa and Cree are both derived from the Algonkian linguistic family.

A previous genetic study by Frégeau *et al.* (1998) demonstrated that Canadian Caucasian samples from different provinces were not distinguishable from one another and

that one Caucasian population can be representative of all Caucasian populations across Canada. For this reason, only one Caucasian population was chosen to study.

Although the populations sampled in this study were selected based on geographical and cultural isolation, they were also chosen based on the availability of sufficient samples to compile DNA databases. For statistical purposes, unrelated samples from scientifically relevant populations are desirable when studying allele frequency data. This ideal can rarely be met when collecting samples from small populations because small populations unavoidably consist of interrelated persons.

Also, it is important to obtain random samples when studying allele frequency data. Blood samples for this study were considered convenience samples since they were obtained from hospital laboratories and might be considered nonrandom; however, convenience samples have been shown to be appropriate for forensic uses (Devlin and Risch, 1992; Committee on DNA Forensic Science, 1996; Foreman *et al.*, 1998; Chakraborty *et al.*, 1999). This is due to the fact that most loci used are unlikely to have any functional significance; therefore, are unlikely to be correlated with any physical or behavioural traits.

The samples used in this study were selected on the basis of visual designation and on the basis of common first names and surnames by hospital laboratory personnel. While some individuals may not be accurately classified by this method, the number of these individuals should be a small percentage of the population groups since both visual and name criteria were employed, the relatively large sample size of the populations in this study, and the large Caucasian and Aboriginal populations in the geographic areas where the samples were obtained.

Sample size in this study is defined as the number of individuals sampled from the population (n); hence, the number of alleles sampled is $2n$ because the human genome is diploid. A population sample must include all common alleles with a certain amount of confidence when determining allele frequency data. The estimation of non observed rare alleles becomes more accurate with larger databases. Chakraborty (1992) stated that 100-150 individuals per population is sufficient for estimates of allele and genotype frequencies at tandem repeat loci. The population sample sizes in this study, $n \geq 99$, are large enough to not be subject to sampling effects using the D18S535 locus (George Carmody 2002, personal communication).

4.2 Technical Considerations

Technical aspects of this study were reviewed in order to ensure that there were no technical errors involved in allelic scoring in the four Northern Ontario populations. Errors made in allelic designation have been shown to be greatly reduced with the use of an allelic ladder (Committee on DNA Forensic Science, 1996). A D18S535 sequenced allelic ladder was employed during this study in order to reduce any errors in allelic scoring and all samples were rescored by different individuals to ensure accurate designation of the D18S535 alleles.

The D18S535 locus displayed a faint band one repeat length under each allele band (shown in Figures 4 to 8). These artifacts are not uncommon among PCR amplification products and their origin is thought to be caused by strand slippage (Hauge and Litt, 1993; Litt *et al.*, 1993) along the tetrameric repeat region resulting in amplification products which

are four bases shorter. These PCR artifacts have been shown to have an intensity which are < 10 % of the main band (Urquhart *et al.*, 1995 a) and that the proportion of stutter band generally increases with allele length due to longer stretches of core repeat sequence (Walsh *et al.*, 1996). The common methods used to interpret results when these PCR artifacts are observed is to read the band which shows the strongest intensity and to read the band which has the slowest electrophoretic mobility (Tautz, 1989; Ingvarsson *et al.*, 2000). The stutter observed in the polyacrylamide gels in this study did not prevent accurate typing of the D18S535 alleles even in heterozygotes with alleles separated by one repeat unit. This was due to the high resolution power of the polyacrylamide electrophoresis system and the use of the common methods for interpreting results when PCR artifacts are observed. The faint stutter bands can actually assist in genotype determination relative to the allelic ladder in non-mixed samples (Edwards *et al.*, 1992). Stutter bands can be minimized with extensive optimization of the PCR protocol, increasing the denaturing conditions during gel electrophoresis with the use of formamide, and maintaining an electrophoresis gel temperature above 43 °C (Litt *et al.*, 1993; Meldgaard and Morling, 1997). All the precautions to minimize stutter bands were taken during this study.

Extracted DNA was not quantified since a set volume of DNA extract repeatedly gave a good yield in almost all samples when the Beta-Globin gene was targeted using PCR (see Figure 3). This method was also used in a study by Phillips *et al.* (1998). The stutter observed at this locus could be attributed to the fact that the DNA was not quantitated before the amplification reaction. A previous study by Lygo *et al.* (1994) showed that amplification of larger than 3 ng of template DNA produced an increase in stutter due to enzyme slippage.

Terminal nucleotide addition (Walsh *et al.*, 1996; Smith *et al.*, 1995) occurs when *Taq* DNA Polymerase catalyses nontemplated addition of an adenine nucleotide to the 3'-termini of amplified DNA fragments. A band which is one base shorter than the expected allele may result from the incomplete addition of adenine to a majority of PCR product strands. An artifact band is generated when this terminal addition does not occur with 100 % efficiency and may be visualized as an extra band. The addition of a final extension step and extended time at 4°C, or room temperature, increases the amount of product that contains the added terminal nucleotide; thus, minimizing the shorter artifact band (Magnuson *et al.*, 1996). With silver stain detection, both strands of DNA are detected. In some loci, the difference in opposing strands causes them to migrate at different rates. This results in doublets for each allele (see Figure 3). Artifact bands are also detected with this system. The doublets which were produced for each D18S535 allele and artifact bands did not appear to interfere with accurate scoring of the alleles.

PCR products differ in size and can be analysed directly by gel electrophoresis. Denaturing polyacrylamide gels were used because while opposing DNA strands of a single allele are the same length, they may migrate differentially due to sequence differences. The electrophoretic mobility of a single-stranded DNA on denaturing polyacrylamide gels was shown in a previous study (Saitoh *et al.*, 1998) to depend on the proportion of AC/GT content with AC-rich strands migrating faster than GT-rich strands. To avoid mistyping of alleles in this study, the allelic ladder together with the D18S535 PCR products were denatured so that scoring of the allelic bands were based on length and not base pair composition. The use of polyacrylamide gels enables resolution of PCR products differing

in size by just one base pair (Gill *et al.*, 1990; Lygo *et al.*, 1994) which assists in the precise designation of alleles.

Variation in signal intensity of different allelic bands within heterozygotes can occur on polyacrylamide gels. This has been shown to be a function of the electrophoresis system and also of the amplification reaction directly related to specific primer concentrations (Lygo *et al.*, 1994). Variation in signal intensity did not appear to have occurred in this study; therefore, should not have affected accurate allele scoring.

Every effort was undertaken to ensure that no technical problems would interfere with the accurate scoring of the D18S535 alleles in the Northern Ontario populations. After reviewing all possible sources of error, there was no indication that error occurred in accurate allele scoring in this study.

4.3 D18S535 Allele Frequencies Among Northern Ontario Populations

This initial D18S535 population study in four Northern Ontario populations showed that the Espanola Caucasian and Espanola Aboriginal populations had a more uniform pattern of allele frequency distribution when compared to the Wikwemikong and Sioux Lookout Aboriginal populations at the D18S535 locus (see Table 1). All eight alleles were observed in the Espanola Caucasian and Espanola Aboriginal populations while Sioux Lookout Aboriginals and Wikwemikong Aboriginals had 6 and 7 observed alleles, respectively. This difference in allele pattern and the amount of observed alleles is probably not due to the sample size of the populations studied, rather it is probably due to genetic drift (Zago *et al.*, 1996; Deka *et al.*, 1995). This is a likely explanation for the observed difference

since the populations in this study were relatively small and both Szathmary *et al.* (1975) and Wall *et al.*(1993) stated that allele frequencies in small founding populations were significantly affected by genetic drift, while allele frequencies in large populations were affected far more by mutation events.

Genetic drift is the random change of allele frequencies at a locus. Genetic drift results from continually small population sizes and fluctuations in the number of individuals which leads to a fixation of alleles and loss of genetic variation. There are two types of genetic drift: the founder effect and the bottleneck effect. In the founder effect alleles may be over- or under-represented in existing individuals in the founding population. This is caused by a large amount of random drift in the first generation of a new subpopulation when the founding members are limited in numbers. If the new subpopulation then expands, it produces a substantial difference from the main population (Scriver, 2001). The bottleneck effect results in a reduction of larger populations by events such as disease, change of environment, and disasters. A bottleneck causes the loss of many low frequency alleles and the loss in the average number of alleles per locus which is greatly affected by the size of the bottleneck. Both the founder effect and the bottleneck effect change the allele composition in a random fashion and allele frequencies in small surviving groups will be altered compared with their ancestors (Gaedigk *et al.*, 2001).

The presence of all eight D18S535 alleles and the more uniform pattern of allele frequency distribution in the Espanola Aboriginal population and not in the Sioux Lookout and Wikwemikong Aboriginal populations is also likely due to the amount of Caucasian gene flow each population has been exposed to in conjunction with genetic drift.

Ojibwa comprise a wide geographic distribution and dispersion into small communities (Jantz and Meadows, 1995). Interracial marriages are common among Aboriginal people, consequently exposing these populations to admixture which is created by interracial gene flow (Buroker *et al.*, 1997). Gene flow (admixture) is a microevolutionary force which opposes genetic differentiation. Caucasian admixture varies widely on the continent and is highest in regions where European contact has been the most intensive and of longest duration (Szathmary, 1993). Studies on blood group markers in the Ojibwa people who inhabit the regions around the Great Lakes indicated that some groups contain detectable European admixture where the likely source of this gene flow were French or British settlers who entered the region during the European migration into North America (Szathmary and Reed, 1972; Szathmary, 1984). The rate of gene admixture from Caucasians has been estimated to be 3 % in northwestern Ontario Ojibwa, in the location of Sioux Lookout, while some Ojibwa groups in northeastern Ontario have been shown to have almost 30 % Caucasian admixture, in the location of Espanola and Wikwemikong Aboriginals, based on the presence Caucasian alleles (Szathmary *et al.*, 1978). Admixture has most likely continued generating Aboriginal populations with a certain degree of Caucasian admixture.

The Aboriginal population samples used in this study have not been evaluated for their degree of admixture. Alleles which were relatively frequent in the Caucasian population sample may have been introduced to the Aboriginal populations during the centuries of European settlement. The Ojibwa from the northern part of Northern Ontario appear to have been the least influenced by Europeans unlike the Ojibwa from the southern part of Northern Ontario, near Lake Huron, who appear to have a great European

influence (Scozzari *et al.*, 1997). This is a possible explanation for the increased number of observed alleles in the Espanola and Wikwemikong Aboriginals when compared to the more geographically isolated Aboriginals from Sioux Lookout.

Although the populations displayed frequency differences of individual alleles across the populations in this study, the common alleles were present in almost every population and the alleles which were not shared by all populations did not have large frequencies. The more frequent alleles in the populations may be founder alleles of the modern populations. At the time that populations diverge, there is a founder effect and it is at this time that alleles are lost (Wall *et al.*, 1993). The founder effect may be the factor responsible for the high frequencies of allele 14 (see Table 1) in the three Northern Ontario Aboriginal populations in comparison to the high frequency of allele 12 in the Espanola Caucasian population.

The reduction in the number of alleles, primarily due to the absence of rare alleles, in the Sioux Lookout and Wikwemikong Aboriginals, relative to the Espanola Caucasian and Espanola Aboriginal populations, did not affect the number of heterozygotes observed. The Sioux Lookout and Wikwemikong Aboriginals had $\geq$ 75 % heterozygosity (see Table 3). This high observed heterozygosity indicated that in spite of the lower genetic diversity in the Sioux Lookout and Wikwemikong Aboriginals, the D18S535 locus displayed a high degree of polymorphism. The Sioux Lookout and Wikwemikong populations were relatively small, and Szathmary (1993) indicated that population size and the size of the area inhabited influences the heterozygosity and gene diversity. Population size has also been shown to be important in the maintenance of genetic variation by Leberg (1992) because the longer a small population remains small, the more genetic variation it will lose.

Populations with small sizes are susceptible to inbreeding which reduces heterozygosity. The level of heterozygosity is associated with fitness and reduction in genetic variation affects the populations' ability to adapt to changing environmental conditions (Luijten *et al.*, 2000). Also a decrease in heterozygosity and a loss of rare alleles is correlated with a decrease in the breeding population size after a population bottleneck (Wright, 1931). A bottleneck allele diversity was predicted to decrease faster than heterozygosity and bottlenecks have been shown to have a much larger effect on allelic diversity than on heterozygosity by Nei *et al.* (1975). The average heterozygosity decreases but then starts to increase as the population size becomes large (Nei *et al.*, 1975).

Genetic drift has been shown to cause the loss of specific alleles in small populations, especially rare alleles (Cavalli-Sforza *et al.*, 1994), and may cause substantial variation in their frequency in some subgroups (Committee on DNA Forensic Science, 1996). This was observed in the three Northern Ontario Aboriginal populations; therefore, genetic drift may be the factor affecting the low incidence of rare alleles in these populations. The low incidence of rare alleles may also be the result of sampling or from the isolation of these groups from other populations (Jaffe, 1992; Cavalli-Sforza *et al.*, 1994; Bonatto and Salzano, 1997). Similarly, the isolation of the Northern Ontario Aboriginals, particularly the Sioux Lookout Aboriginals, may explain their bias for the mid-sized alleles. The low incidence of some alleles (0.5%) is more likely due to a sampling effect, as opposed to a founder effect, resulting in the isolation of these alleles in a specific group (Frégeau *et al.*, 1998). The presence of similar alleles in different populations indicated that the origin of these alleles may predate divergence (Deka *et al.*, 1994; Zago *et al.*,1996).

4.4 Genetic Differences Among Northern Ontario Populations

As previously stated, more Caucasian contact is likely with the Espanola and the Wikwemikong Aboriginal populations than with the Sioux Lookout Aboriginal population due to access to service centres. This would allow for mixture (gene flow) with distinct populations; therefore, differentiation. Based on this fact, the Espanola and Wikwemikong Aboriginal populations were pooled and compared to the Sioux Lookout Aboriginals. The results indicated a significant difference in their allelic distributions. Heterogeneity was also observed at the D18S535 locus in the three Northern Ontario Aboriginal populations as shown by the significance ($p < 0.01$) of the MANOVA used to determine if there was an overall homogeneity of variance in the D18S535 allelic distribution in the three Northern Ontario Aboriginal populations. This result is not unusual among members of a single language family (Szathmary *et al.*, 1975). The genetic differences are likely due to the greater presence of foreign genes that had entered some of the populations and not the others. Szathmary and Reed (1972) stated that the entry of Caucasian alleles into Aboriginal gene pools was one of the factors generating genetic variability in modern Aboriginal populations.

Szathmary *et al.* (1975) performed a study on a population sample from Wikwemikong and from Pikangikum, which is located in Northwestern Ontario, using several blood group, serum protein, and red cell enzyme genes. Their study found extensive differences in their blood group genetic characteristics. Szathmary *et al.* (1975) stated that genetic drift was considered principally responsible for the genetic heterogeneity between the two populations since there was undeniable unity of origin for the groups and the environments within which the groups were found were identical. They suggested that

genetic divergence among the Ojibwa, due to genetic drift, probably occurred after nomadic hunting bands were settled on reservations. The bands were sometimes at considerable distances from each other, with very little contact with groups at the extremes of the Ojibwa range. Szathmary *et al.* (1975) also indicated that natural selection was probably the least important agent of diversification and that dissimilar amounts of European admixture only aided divergence of the Aboriginal populations. Szathmary (1993) stated that there was considerable genetic variation at the local level in Aboriginals and genetic differentiation within North America could explain existing Aboriginal genetic diversity.

Based on the results of the study by Szathmary *et al.* (1975), genetic drift and Caucasian gene flow are likely the cause of the observed difference between the Sioux Lookout Aboriginals in Northwestern Ontario, and the Wikwemikong and Espanola Aboriginals in Northeastern Ontario. These populations are almost 1300 km apart, have had no known contact since the initial Ojibwa separation, and have been affected by differing amounts of Caucasian gene flow. Genetic drift is a possible force considering that during the first European contact, the Aboriginals lived in small groups dispersed over a large area, and were not a single homogenous group. The differing geographical and climatological pressures of the three Northern Ontario populations may have also produced a recognizable effect (Szathmary *et al.*, 1975). Also, drift is considered to have had an effect if there are alleles in the populations which have noticeably abnormal frequencies and if populations exhibit a high potential for genetic drift indicated by a low breeding population size (Cavalli-Sforza, 1973; Szathmary *et al.*, 1975). The combination of high drift potential plus anomalous allele frequencies (Table 1) relative to other populations suggest that genetic drift

has affected the incidence of allele 9 (relative frequency = 0.126) in the Wikwemikong Aboriginals, allele 11 (relative frequency = 0.080) in the Sioux Lookout Aboriginals, and allele 16 (relative frequency = 0.025) in the Espanola Aboriginals.

Another indication that drift might be the force producing a difference in the allele frequencies was the difference in the major alleles in the Espanola and Wikwemikong Aboriginal populations, and the different patterns of allele distribution. The general trend is that the Espanola and Wikwemikong Aboriginal populations were bimodal, compared not only to the Espanola Caucasians but also to the Sioux Lookout Aboriginals. The unimodal and bimodal patterns of the four Northern Ontario populations can be observed in Figure 8.

A previous study by Nei and Roychoudhury (1997) found that in both North and South America, the Aboriginal tribes had undergone extensive genetic differentiation from one another. The high degree of genetic differentiation of Aboriginal tribes was suggested to have occurred principally due to the bottleneck effect. Although the causes of significant allelic heterogeneity between the three Northern Ontario Aboriginal populations at the D18S535 locus cannot be determined, genetic differentiation of populations is principally caused by isolation either by cultural factors or geographic distance; therefore, the interpopulation variance at the D18S535 locus was most likely influenced by genetic drift and evolutionary separation of these semi-isolated populations as opposed to mutation (Nei and Roychoudhury, 1997; Chakraborty *et al.*, 1999; Da Silva *et al.*, 1999).

The findings of the studies by Szathmary *et al.* (1975), Nei and Roychoudhury (1997), and the findings of this present study indicated that the study of a small number of tribes may not be representative of the variability of Aboriginals, even if a large number of

individuals were studied. To depict the whole range of genetic variability of Aboriginals, it would be necessary to study a large number of populations. The lack of homogeneity in the allele frequencies of the three Northern Ontario Aboriginal populations indicated that the genetic forces which have acted on these populations have a statistically recognizable effect. Forces such as inbreeding, population mixture, selection, and non random mating can all affect allelic proportions. The three Northern Ontario Aboriginal populations cannot be amalgamated into a single database since they contain differences in their allelic frequencies (Rahman *et al.*, 2001), particularly alleles 9, 11, and 12 (see Table 11).

The D18S535 locus shows not only considerable variability within each population, possibly caused by sampling which may have included closely related individuals, but also some significant differences in allele frequencies between the four Northern Ontario populations in this study, as shown by the significance ($p < 0.01$) of the MANOVA used to determine if there was an overall homogeneity of variance in the D18S535 allelic distribution in the four Northern Ontario populations. The difference in allele distribution was caused by alleles 9, 10, 11, 12, and 14 (see Table 10). This difference might be an indication that gene flow is unidirectional, from the Caucasian to the Aboriginals, and that the Caucasians contain little, if any, Aboriginal alleles. The heterogeneity of the four Northern Ontario populations corroborates the idea that databases are needed for different ethnic populations despite being from the same geographic locale.

A previous study by Picornell *et al.* (1996) indicated that genetic distances were significantly correlated with linguistic distances, whereas the genetic distances were not significantly correlated with geographic distances. The F_{ST} statistics (see Table 11) showed

a moderately low level of differentiation among populations, implying a high level of gene flow. The F_{ST} values, which were a genetic approach to comparing the allelic distribution between populations, since it measures the degree of association between the uniting gametes (Li and Horvitz, 1953), were not consistent with the statistical tests used to compare the allelic distribution between populations. The genetic distances showed closest relationships between Espanola Aboriginals and Sioux Lookout Aboriginals, and may be explained by the linguistic affiliations shared by these populations. Also, the Espanola Caucasians and the Espanola Aboriginals showed a close relationship and may be explained by the geographical proximity of these populations to each other. In spite of being in different geographic regions, the Aboriginal populations who belong to the same linguistic family were all similar to each other; therefore, both geographical proximity and linguistic affiliations seem to correlate with genetic relationships between populations. This interpretation is somewhat contradictory to a study by Rolf *et al.* (1998) which showed that geographical proximity seemed to correlate with genetic relationships between populations better than linguistic affiliations.

4.5 Comparison of Northern Ontario Populations with Populations From Europe and South America

A series of MANOVA were employed to determine if there was an overall homogeneity of variance in the D18S535 allelic distribution in the four Northern Ontario populations along with previously published D18S535 studies from Europe and South America, in the Espanola Caucasians in conjunction with previously published D18S535

studies from Europe, in the previously published D18S535 studies from Europe, and in the Northern Ontario and South American Aboriginal populations. The results demonstrated a heterogeneity of variance in all the comparisons (see Tables 5, 6, 7, 8, and 9). This demonstrates the variability of the D18S535 locus.

To evaluate the differences between the four Northern Ontario populations and the South American and European populations, a series of pairwise comparisons were performed. The results (see Table 12 and Figures 28 to 71) showed that the D18S535 allele frequencies for the Espanola Caucasian population and the Caucasian-based populations from previously published European countries were significantly different from each other. This was unexpected because based on its history, Espanola is an admixture of European ancestry; therefore, this population should exhibit frequencies similar to its parental populations. A similar result was observed for the three Northern Ontario Aboriginal populations and the Puna population from South America which showed that their D18S535 allele frequencies were significantly different from each other (see Table 12). Szathmary (1993) stated that time since the initial separation of the populations is the factor which determines the amount of genetic differences between populations.

A series of pairwise comparisons between the European populations indicated that most European populations were similar to each other with respect to their D18S535 allelic distributions (see Table 13). This was expected because a study by Nei and Roychoudhury (1997) found that European populations were genetically closely related to one another.

Szathmary (1993) indicated that no matter what is the source of a populations' genes, all of the evolutionary forces (mutation, genetic drift, gene flow, and natural selection) will

ultimately act on the gene pool and generate changes in gene frequency. The dissimilarities in allele frequencies between the compared populations indicated that a significant difference in profile frequency estimates is likely if an alternate subgroup sample population were used as a reference database for the D18S535 locus (Budowle *et al.*, 1999).

When compared with previously published data from Europe and South America, the predominant alleles, the mid-sized alleles, did not appear to be the same in the various populations as shown in Figures 8 to 11. The Espanola Caucasian population was centred around allele 12 and previously published populations from Europe tended to center around allele 13. The Ibiza population was different from the other previously published populations from Europe and all other populations in that the most common allele was allele 9. Espanola Aboriginals, Sioux Lookout Aboriginals, Wikwemikong Aboriginals, and the Puna population from Northwestern Argentina all appeared to center around allele 14. This shift observed in the mid-sized alleles appeared to be the cause of the allelic differences observed in the pairwise comparisons between Caucasian and Aboriginal population samples from Northern Ontario and with previously published populations from Europe and South America (See Table 14). As previously stated, these differences were most likely due to genetic drift.

When populations go through a bottleneck, they show a rather large genetic distance even with their neighbouring populations (Nei and Roychoudhury, 1997). The Euclidean dissimilarity coefficient matrix based on D18S535 allele frequency distributions (Table 17) indicated that there was a relatively large distance between the four Northern Ontario populations and that the shortest distances were typically between the European populations.

This was another indication that a bottleneck effect may be the factor contributing to the differences observed between the Northern Ontario populations. Mutation, selection, and genetic drift, are all evolutionary changes which affect populations. The effects of these forces are highly locus-dependent which makes comparisons of allele frequencies for one or two loci not as reliable due to large variances which are produced. Comparing one locus produces only the history of that locus; therefore, in order for genetic relationships among populations to become clear, a large number of loci would have to be examined. Examining multiple loci allows the effects of genetic drift and natural selection, which vary between the loci, to be averaged out (Szathmary and Ossenberg, 1978; Pamilo and Nei, 1988; Nei and Roychoudhury, 1997).

Cluster analysis produces details of population relationships. The results of the cluster analysis, using euclidean distances measurement to calculate distance and average linkage between groups to form the clusters, indicated that based on D18S535 allele frequency distribution, two major clusters were formed. One cluster consisted of the three Northern Ontario Aboriginal populations while the other cluster consisted of the Espanola Caucasians and the previously published D18S535 populations from Europe and South America. This was expected based on the linguistics and geography of the three Northern Ontario Aboriginal populations. A study by Nei and Roychoudhury (1997), using multiple loci, found that North American and South American Aboriginal populations were clustered separately, with the exception of the Northern Ontario Ojibwa, who were closer to the South American Aboriginals than to the Aboriginals of Alaska. The results of this study were different from the results found by Nei and Roychoudhury (1997) since the Puna population

from Northwestern Argentina did not cluster with the three Northern Ontario Aboriginal populations. This is probably a reflection of the fact that, as previously stated, the effects of microevolutionary forces are highly locus-dependent, possibly reflecting the genetic effects of isolation of these populations, and reiterates the need to compile databases for new STR loci in many different populations.

Contingency tables and chi-square analyses also demonstrated that the ethnicity and geography of the populations had a significant effect on the formation of the clusters. This was expected and demonstrated that populations tend to mate within their own subpopulation as stated elsewhere (Committee on DNA Forensic Science, 1996).

4.6 Analysis of Northern Ontario Populations with Reference to Hardy-Weinberg Equilibrium

Northern Ontario is a rich mixture of different ethnic subpopulations. In more rural areas, such as the populations used in this study, people may perhaps breed within their own subpopulation. This is due to a small degree of relatedness because of the possibility of common ancestors for members of the population; therefore, invalidating the use of the product rule (Weir, 1994). Genotypic frequencies are distorted by inbreeding from the frequencies expected from Hardy-Weinberg Equilibrium (HWE), with more homozygotes than expected (Weir, 1994; Rahman *et al.*, 2001). In order to detect possible divergence from HWE expectations in the Northern Ontario populations, an exact test was used. An exact test is preferred for population genetic analyses of hypervariable markers, such as short tandem repeats, because it has a high statistical power to detect Hardy-Weinberg deviations

(Guo and Thompson, 1992; Chakraborty and Zhong, 1994; Morin *et al.*, 1994; Hou *et al.*, 1994). Other traditionally used tests such as the chi-square goodness-of-fit test (Nei and Roychoudhury, 1974; Nei, 1978; Drmić *et al.*, 1998) and the likelihood ratio test (Zar, 1996) were not attempted because they are less stringent than the exact test (Guo and Thompson, 1992).

HWE is an equilibrium of the distribution of alleles in a population among the different genotypes. HWE indicates that the alleles associate randomly with each other at a particular locus and that there is no detectable population heterogeneity (Budowle *et al.*, 1991). A significant deviation from HWE should be observed if only a portion of the possible genotypes are present in a population sample. The advantage of exact tests is that they are not negatively affected by small expected values and may be the only test which is valid if sample sizes are small and some alleles are rare (Lessios, 1992). A genotype represents a chance encounter between two alleles; therefore, the exact test can be applied because it generates random genotypic datasets obtained through simulation of allelic assortment (Holt *et al.*, 2000). Goodness-of-fit chi-square tests are severely affected by small expected frequencies which makes it necessary to pool genotype classes. One way to avoid genotype pooling is to test homozygosities/heterozygosities rather than genotype frequencies by pooling all homozygotes and pooling all heterozygotes at a locus and test them against those predicted from the Hardy-Weinberg law. This method is suitable when deviations from HWE are thought to be due to population mating structure which affects overall heterozygosity regardless of allele (Lessios, 1992). This test however has less statistical power than tests which include information about individual genotypes and is not able to

detect deviations from HWE due to selection and mutation which act on the frequency of individual alleles or genotypes (Pamilo and Vario-Aho, 1984). Pooled data which shows non-significant results may be due to pooling rather than a good fit between expected and observed genotype distributions (Lessios, 1992). Likelihood-ratio tests are used to test the hypothesis that deviations of observed from expected genotype frequencies are equal to zero. Likelihood-ratio tests are like chi-square tests in that it has a disadvantage if the population being tested has small expected frequencies (Lessios, 1992).

The Hardy-Weinberg test is a useful means of evaluating the quality of a data set. If a population is in Hardy-Weinberg equilibrium, then the proportions of the alleles remain constant from one generation to another which is important in population genetics; therefore, to forensic DNA testing. It is a test to determine the stability of the alleles in a population by testing for deviations from the genotypic ratios expected when the population is large, random mating, and in the absence of disturbing forces such as selection, mutation and migration. Any deviations observed might indicate hidden ethnic stratification of a population sample or typing errors (Drobnič *et al.*, 2000).

With as many as 8 alleles, a total of 36 possible genotypes at the D18S535 locus were expected to occur in each population. The most common genotypes (see Table 2) contained the mid-sized alleles which was expected based on their high allelic frequencies. Since some of the alleles were very rare (relative allele frequencies = 0.000-0.005), the minimum sample size required in order to observe all genotypes in a sample would be very large (Chakraborty, 1992). With the data obtained from this study, problems in testing the Hardy-Weinberg equilibrium are expected to occur. The exact test was conducted to see if the D18S535

genotypes in this study conformed to Hardy-Weinberg expectations. The results (see Table 3) showed that the genotype distributions were in accordance with their Hardy-Weinberg expectations in the Espanola Caucasian and Wikwemikong Aboriginal populations. This result is unexpected since the Wikwemikong Aboriginals are located on the Manitoulin Island in Lake Huron and island populations are typically small in size; therefore, violating the Hardy-Weinberg assumptions of a large sized population. Geographically isolated island populations can be further subdivided by ethnicity, religion, and socioeconomic factors, which further reduces their effective sizes, and assist genetic changes due to evolutionary processes such as inbreeding (Martin *et al.*, 2000).

The Espanola Aboriginal and Sioux Lookout Aboriginal populations showed genotypic distributions which were not in agreement with their Hardy-Weinberg Equilibrium expectations. The observed allele frequencies at D18S535 do not supply appropriate estimations of forensic casework sample genotype frequencies in the Espanola Aboriginal and Sioux Lookout Aboriginal populations due to HWE deviation.

Nonrandom association of alleles (deviation from HWE) at loci is often interpreted as proof of either intrapopulation heterogeneity as a result of recent admixture of genetically distinguishable subpopulations (population mixing or non-random mating) (Prout, 1973; Ohta, 1982), selection, inbreeding, presence of a silent or unidentified allele, an error in the classification of the alleles (usually with RFLPs), or population subdivision (Devlin *et al.*, 1990; Rousset and Raymond, 1995). Population subdivision, one which contains two or more groups within the population whose individuals mate between groups, has been shown to be the most important variable for tandem repeat loci in human populations (Lander, 1989;

Cohen, 1990).

Further examination of the populations which did not conform to Hardy-Weinberg Equilibrium, as shown by the exact test, revealed that they contained rare alleles. These rare alleles can be a cause of significant departures from expectation even if the population from which the samples are drawn is in Hardy-Weinberg proportions (Committee on DNA Forensic Science, 1996). Many Aboriginal groups live in remote northern areas but there are significant differences in the degree of Euro-Canadian influence on their communities (Young *et al.*, 1990). This may be a factor which has affected the genotypic distributions of the populations.

The Aboriginal samples from Sioux Lookout are comprised of Cree and Ojibwa tribes; however, the samples could not be identified as coming from one or the other. The Espanola Aboriginal and Sioux Lookout Aboriginal populations both consist of individuals derived from different local subpopulations which utilize the Espanola General Hospital and the Sioux Lookout Zone Hospital; however precise details of the origin of these samples cannot be ascertained because of lack of information obtained from donated samples. This substructuring is a possible explanation for the HWE deviation observed in these population samples.

Since the null hypothesis was rejected for three of the four Northern Ontario populations, a new hypothesis was developed to explain the data. Through testing this alternate hypothesis it was demonstrated that, even in the rural populations of the Espanola Aboriginals, Sioux Lookout Aboriginals, and the Wikwemikong Aboriginals, these populations showed an excess number of heterozygotes. This was an unusual observation

for the Espanola Aboriginal, Sioux Lookout Aboriginal, and Wikwemikong populations since a study performed by Kidd *et al.* (1991) revealed that homozygosity is much greater in Aboriginal populations when compared to Caucasian populations due to genetic isolation and lack of migration. Also, Ojibwa marriages have historically been noted as having a large number of cousin marriages and parents had considerable power in controlling the choice of their children's marriage partners (Szathmary *et al.*, 1975). These two marriage patterns are thought to produce more homozygotes in a population.

This result was also corroborated by negative F_{IS} values which indicated a deviation from HWE in the direction of excess heterozygotes; however, the Wikwemikong Aboriginal population, which showed no deviation from HWE, showed a slightly more negative F_{IS} value than the Sioux Lookout Aboriginal population, which did show deviations from HWE.

Deviations from HWE are usually in the direction of excess homozygotes (Chakraborty *et al.*, 1994; Buroker *et al.*, 1997; George Carmody 2002, personal communication). Technical aspects of this study were previously reviewed in order to ensure that there was accurate allelic designation in the four Northern Ontario populations. The results of suggested that technical errors did not prevent accurate allele designation; therefore, other factors needed to be examined in order to determine the cause of the excess heterozygosity observed in the Espanola Aboriginal and Sioux Lookout Aboriginal populations.

The real underlying cause for the excess heterozygotes might be the result of a sampling effect. There are other possibilities for the excess heterozygotes which include the fact that small populations, such as those used in this study, generally show heterozygote

excess, which decrease with increasing population size. This is thought to occur because the heterozygous individuals in small populations may be survivors of a formerly larger population with relatively high fitness (Luijten *et al.*, 2000).

Heterozygote excesses which were observed in the Espanola Aboriginal and Sioux Lookout Aboriginal populations may possibly be accredited to substructuring of the populations into a large number of families which would cause an over representation of certain alleles showing up as heterozygotes (Launey *et al.*, 2001; George Carmody 2002, personal communication). Random drift can also be an important factor in small populations in producing heterozygote excesses (Sbordoni *et al.*, 1986; Hedgecock and Sly, 1990). Sampling bias of alleles due to a small amount of founders and non-random mating can have an affect on the amount of heterozygotes in a population (Launey *et al.*, 2001). Other studies (Hedgecock and Sly, 1990; Leberg, 1992) have shown that the heterozygosity levels can increase temporarily in the generation immediately following a bottleneck, attributed to allelic evening out. Another possibility is that the parents of the sample populations may be from a generation where partners tended to be drawn from two different subpopulations (Richard A. Nichols 2003, personal communication).

Inbreeding seems to be avoided (inbreeding avoidance) by the Espanola Aboriginal and Sioux Lookout Aboriginal populations and this would be a plausible explanation for the excess of heterozygotes observed in these populations (George Carmody 2002, personal communication). Inbreeding, mating between individuals related by a common ancestry, causes individuals to have noticeably lower survival. Animals avoid inbreeding by dispersal to avoid contact with close relatives, by recognizing and avoiding kin as mates, by extra-

pair/extra-group copulations, by delayed maturation or reproductive suppression, and avoiding familiar individuals regardless of kinship (Boulin and Boulin, 1988; Pusey and Wolf, 1996). Human populations tend to avoid close inbreeding and this inbreeding avoidance might be the phenomenon which is occurring in the Northern Ontario populations causing observed homozygote deficiencies in these populations. Selective pressures may favour inbreeding avoidance and the selection of highly dissimilar mates (Amos *et al.*, 2001). Heterozygote advantage may be an important demographic-genetic mechanism to conserve genetic variation during population bottlenecks (Luijten *et al.*, 2000). Selection may have favoured the survival of heterozygotes in the small populations used in this study; therefore, it is possible that the action of selection might be affecting the allelic distribution.

Although the D18S535 is a noncoding repetitive element which does not direct functional RNA or protein products, a possible selection factor producing the excess heterozygotes might be Fat-free mass (FFM). Chagnon *et al.* (2000) performed a genome-wide search for genes which may be involved in the regulation of the lean component of body mass in humans. FFM consists mostly of skeletal muscle and bone tissues. Significant statistical linkage was observed for D18S535. There may be some sort of natural selection due to fitness differences among individuals with different genotypes in the four populations in this study; however, Aboriginal peoples in Canada have observed a rise in the occurrence of obesity due to an adopted diet which is high in energy, saturated fat and simple sugars, and a sedentary lifestyle with reduced physical activity (Thouez *et al.*, 1989). This occurrence makes it unlikely that individuals were selectively choosing their mates based on FFM.

Another possible selection factor producing the excess heterozygotes observed in the

Northern Ontario Aboriginal populations might be Early Onset Diabetes (EOD). A recent study to evaluate the relative contribution of known and unknown susceptibility genes in families with EOD revealed that the D18S535 locus showed a significant statistical linkage with EOD (Lindgren *et al.*, 2002). Early Onset Diabetes, whose trigger appears to be impaired insulin secretion, is a familial diabetes which shows the onset of diabetes ≤ 45 years.

Diabetes Mellitus is a endocrine disease due to a deficiency of insulin. Two major classifications for diabetes are type 1 and type 2. Type 1 diabetes is caused by a destruction of pancreatic beta cells which leads to an absolute deficiency of insulin. Type 2 diabetes is the result of insulin resistance in the body's tissues and/or a defect in the pancreas to secrete insulin and is the most common with 90 % of people with diabetes having type 2. Type 2 diabetes has been shown to have a genetic basis but that environmental factors are also involved in disease onset (Canadian Diabetes Association, Canadian Medical Association, 1998).

Diabetes Mellitus, especially Early Onset Diabetes, is emerging as a prevalent risk factor in certain Aboriginal subgroups (Young *et al.*, 1990; Health Canada, 2000) and has been shown to affect ~ 6 % of Aboriginal adults compared to ~ 2 % of all Canadian adults (MacMillan *et al.*, 1996). Studies on the Oji-Cree of Northwestern Ontario, in the Sioux Lookout Zone region, have indicated that they have a frequency of non-insulin-dependent (type 2) diabetes of ~ 40 % which is among the world's highest (Harris *et al.*, 1997). On reserve Aboriginals appear to have a higher rate of diabetes than off reserve Aboriginals (Health Canada, 2000) and most Aboriginals have type 2 diabetes, type 1 is rare (Canadian

Pediatric Society, 1994). Type 2 diabetes usually occurs later in life but there is a high occurrence of it among Aboriginal children in Canada (Dean *et al.*, 1992).

There are some instances of STRs being associated with genetic diseases; however, these STRs all have three base pair repeats and the STR used in this study, D18S535, was a four base pair repeat. Humans choose their mates preferentially but independently of the genetic system being tested; nevertheless, the statistical linkage to diabetes may cause a distinct bias through selective advantage or behavioural or physical traits which may influence mate selection. If this scenario is the case then the relevance of forensic genotype frequency estimates using population databases derived from the D18S535 locus could be affected; however, it seems unlikely that selection at one locus would be associated with a different, unlinked locus (Evett *et al.*, 1996). Also, selection seems an unlikely occurrence at the D18S535 locus since previously studied populations from Europe and South America did not show any deviations from HWE; therefore, did not indicate selection.

Information needed in order to detect selection operating on genetic polymorphisms were not collected for the populations used in this present study. If demographic information such as sex or age were available, then detection of significant differences in D18S535 genotypic proportions between the sexes, or between the age groups might have been possible (Szathmary *et al.*, 1975).

In order to determine the true cause of the excess heterozygotes observed in the Espanola Aboriginal and Sioux Lookout Aboriginal populations, further studies should be performed. Since destabilizing forces, such as selection and mutation, act on a single locus, while others, such as population subdivision and inbreeding, act on all loci (Bowcock *et al.*,

1991; Lessios, 1992), further studies could include testing these four Northern Ontario populations with additional STR loci since population phenomenons such as inbreeding avoidance and genetic drift are detectable at other loci (George Carmody 2002, personal communication), retest the population samples using D18S535 to see if the same results are obtained, and retest another population sample from the same populations to compare the results.

4.7 Forensic Application of D18S535 in the Four Northern Ontario Populations

The data presented from this study indicated the effects of population substructuring to be great in the forensic context. Genotype frequency estimation for multiallelic loci relies on component allele frequencies since population data represent only a small portion of possible DNA profiles. Deviations from Hardy-Weinberg proportions has a practical consequence in genotype frequency estimation since a genotype frequency from observed allele frequencies is completely tied to the behaviour of alleles in a population (Holt *et al.*, 2000). Also, without the assumption of Hardy-Weinberg Equilibrium, the product rule cannot be employed to calculate the probability of occurrence of a multi-locus DNA profile. Even though some populations showed deviations from Hardy-Weinberg expectations, the data generated in this study still has practical significance in forensic casework if the actual observed genotype frequency data are used when estimating profile frequencies for the D18S535 locus, for those populations which showed departures from Hardy-Weinberg proportions. This avoids assumptions of independence since genotypes are independent of HWE (Martínez-Jarreta *et al.*, 1998; George Carmody 2002, personal communication). Not

all genotypes were observed in this study; therefore, a minimum genotype frequency of 5/N, where N is the size of the sampled population, should be used for genotypes which were not observed. This minimum genotype frequency provides a conservative frequency estimate of rare genotypes (George Carmody 2002, personal communication).

When interpreting and assessing the weight of DNA profiles, focus should be on conditional profile probabilities (Balding and Nichols, 1994; Committee on DNA Forensic Science, 1996; Ayres and Overall, 1999; Bruce Weir 2003, personal communication). Conditional profile probabilities determine the probability that an unknown person has a profile when it is known that one person already has the profile. Human populations are stratified so that individuals who are related tend to associate together (Nichols and Balding, 1991). Subpopulations, population substructure involving various ethnic populations within a large ethnic population, which are indistinguishable in terms of anthropology, geography, ethnicity or culture, may still exhibit major genetic differentiation (Zhivotovsky *et al.*, 2001). This stratification may result in a serious underestimation of the probability of two DNA profiles matching (Cohen, 1990). The Committee on DNA Forensic Science (1996) provided guidelines for the evaluation of forensic DNA evidence and have suggested equations for conditional probabilities which involve population allele frequencies and θ (F_{ST} values), which represents the level of coancestry within a population to account for the effect of population structure. Conditional probabilities are modifications to correct bias in homozygote estimates. The formula $p^2 + p(1-p)\theta$ should be used for the homozygote genotype frequency at a locus instead of p^2 while the product rule formula, $2p_ip_j$, should still be used for heterozygote genotype frequency calculations. The Committee on DNA Forensic

Science (1996) suggested a conservative θ value of 0.03 for all population groups for PCR-based systems which would provide a conservative approximation (erring in favour of the suspect). The use of conditional profile probabilities eliminates the need to have data from all subpopulations.

The results of the pairwise population comparisons and the F_{ST} indices in this study suggest that the possibility of population differentiation should be taken into account in the calculation of match probabilities in Northern Ontario population forensic cases. However, conditional profile equations do not take into account heterozygotes excess and the hope would be that a sufficiently large θ value would be sufficient in cases where excess heterozygotes were observed (Bruce Weir 2003, personal communication). Also, if the excess heterozygosity is due to within-subpopulation processes, such as disassortative mating and differences in mortality, then incorporating the inbreeding coefficient F_{IS}, in addition to F_{ST}, according to Ayres and Overall (1999), might help deal with the excess heterozygotes when calculating forensic match probabilities (Richard A. Nichols 2003, personal communication). If the cause of the excess heterozygotes is outbreeding, then the conditional profile probabilities suggested by Balding and Nichols (1994) using just the F_{ST} indices, should be adequate to account for the excess heterozygotes (Richard A. Nichols 2003, personal communication).

The high informativeness of the D18S535 was demonstrated by the relatively high number of detected alleles and > 75 % heterozygosity. The power of discrimination of the locus was > 0.85 for all of the Northern Ontario populations tested and the calculated chance of exclusion varied from 0.51 to 0.67 which were similar to previously published

D18S535 studies from Europe and South America. Theoretical values of forensic interest, such as the heterozygosity, power of discrimination, and chance of exclusion value were calculated from the genotype frequencies obtained in this study; therefore, lack of Hardy-Weinberg proportions in some of the populations did not affect the interpretation of these values.

The power of discrimination is indicative of how frequently two randomly drawn people in the population will have the same typing. Although the values found in this study were high, they are not sufficient to exclude the majority of suspects from matching a crime scene sample. This value would increase if this locus was added to a battery of loci which are currently used; therefore, increase the significance of any match, especially in circumstances involving partial profiles in Northern Ontario forensic casework. Before incorporation of the D18S535 into an already established STR package it would be necessary to check for linkage equilibrium.

5. Conclusion

STR analysis is an integral part of forensic DNA analysis. When a match is attained, a calculation is made to provide an estimate of the scarcity of the DNA profile (Committee on DNA Forensic Science,1996). The frequencies for a DNA profile frequency estimate are derived from sample population databases which contain the major population groups of the geographic area. Establishment of databases such as these are crucial in the successful application of any new human identification markers for forensic DNA typing in order to calculate population specific allele frequencies. Allele and genotype frequencies can vary across subpopulations which would have implications for profile frequency estimates in forensic casework. Northern Ontario consists of abundant Aboriginal and Caucasian populations; therefore, the aim of this research was to generate allele frequency databases for a new STR locus D18S535, for these forensically relevant populations, and determine its usefulness in forensic casework in Northern Ontario. The populations chosen for study, Espanola Caucasians, Espanola Aboriginals, Sioux Lookout Aboriginals, and Wikwemikong Aboriginals, were representative of the populations of Northern Ontario in terms of geography, linguistics, and culture.

The level of polymorphism at the D18S535 locus was high in the Northern Ontario populations. Even in isolated populations, such as the Aboriginals, the extent of heterozygosity was the same as that of the less isolated populations of the Caucasians. The observed allele frequency differences among the populations in this study were typical of this

locus, in the sense that the common alleles (e.g., alleles 12, 13, and14) were present in all populations.

Comparison of allele frequencies of the D18S535 locus between the four Northern Ontario populations, and between the four Northern Ontario populations and previously published D18S535 studies from Europe and South America showed significant differences in their distribution. The dissimilarity in allele frequencies indicated that a significant difference in profile frequency estimates would be likely if an alternate subgroup sample population were used as a reference database for the D18S535 locus. In addition, comparison of the allelic patterns of the three Aboriginal populations in this study resulted in significant differences demonstrating a heterogeneity of the Northern Ontario Aboriginal populations. Thus, amalgamation of the Aboriginal populations was impossible for this locus.

The genotype data at the D18S535 locus was in conformance with the predictions of Hardy-Weinberg Equilibrium in the Espanola Caucasian and Wikwemikong Aboriginal populations, and deviated from the predictions in the Espanola Aboriginal and Sioux Lookout Aboriginal populations. These findings indicated that the observed allele frequencies may be used to estimate genotype frequencies and the product rule used to calculate combined genotype frequencies across loci in the Espanola Caucasian and Wikwemikong Aboriginal populations only. The actual observed genotype frequencies could be used with the Espanola Aboriginal and Sioux Lookout Aboriginal populations when estimating profile frequencies.

The deviations from HWE, in the direction of excess heterozygotes in some of the populations, indicated that genetic pressures may have had an influence within the populations. Additional studies are needed to verify the true cause of the HWE deviations at this locus in these populations. Further studies could include examination of several different loci in these populations, and retesting the D18S535 locus in the same samples and in new samples from the same populations.

Since typing the D18S535 locus proved to be easy, and the high heterozygosity, power of discrimination, and chance of exclusion observed at the D18S535 in the Northern Ontario populations, the D18S535 locus would be a very useful marker for forensic analysis. The D18S535 was a highly informative marker and the data presented can be used to calculate matching probabilities in forensic casework if individuals from these Northern Ontario populations were considered as a source of DNA evidence. The capacity to discriminate among individuals can be strengthened when multiple loci are analysed. Future studies might focus on incorporating the D18S535 locus into a multiplex PCR package and multicolour fluorescence detection which would offer high throughput and automated genetic analysis. With a D18S535 allelic ladder available from the Institute of Legal Medicine, Santiago de Compostela, Galacia, Spain, compilation of population databases could be rapidly accomplished.

References

Albeza, M. V., A. Picornell, N. Acreche, C. Tomas, J. A. Castro, and M. M. Ramon. (2002). Genetic variability at 14 STR loci in the Puna population of northwestern Argentina. International Journal of Legal Medicine, 116:126-132.

Amos, W., J. W. Wilmer, K. Fullard, T. M. Burg, J. P. Croxall, D. Bloch, and T. Coulson. (2001). The influence of parental relatedness on reproductive success. Proceedings of the Royal Society of London, 268(1480):2021-2027.

Ayala, F. J. (1982). *Populations and evolutionary genetics: a primer.* The Benjamin/Cummings Publishing Company, Inc., California.

Ayres, K. L., and A. D. L. Overall. (1999). Allowing for within-subpopulation inbreeding in forensic match probabilities. Forensic Science International, 103:207-216.

Bagley, C. (1995). A plea for ignoring race and including insured status in American research reports on social science and medicine. Social Science and Medicine, 40 (8):1017-1019.

Balding, D. J., and R. A. Nichols. (1994). DNA profile match probability calculations: how to allow for population stratification, relatedness, database selection and single bands. Forensic Science International, 64:125-140.

Balding, D. J., and R. A. Nichols. (1995). A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. Genetica, 96:3-12.

Bär, W., B. Brinkmann, B. Budowle, A. Carracedo, P. Gill, P. Lincoln, W. Mayr, and B. Olaisen. (1997). DNA recommendations - Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. International Journal of Legal Medicine, 110:175-176.

Barral, S., M. V. Lareu, A. Salas, and A. Carracedo. (2000). Sequence variation of two hypervariable short tandem repeats at the D22S683 and D6S477 loci. International Journal of Legal Medicine, 113:146-149.

Batanian, J. R., S. A. Ledbetter, R. K. Wolff, Y. Nakamura, R. White, W. B. Dobbins, and D. H. Ledbetter. (1990). Rapid Diagnosis of Miller-Dicker syndrome and isolated lissencephaly sequence by polymerase chain reaction. Human Genetics, 85:555-559.

Beckman, J. S., and J. L. Weber. (1992). Survey of Human and Rat Microsatellites. Genomics, 12:627-631.

Bonatto, S. L., and F. M. Salzano. (1997). A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. Proceedings of the National Academy of Sciences USA, 94:1866-1871.

Boulin, S. F., and M. Boulin. (1988). Inbreeding avoidance behaviours. Trends in Ecology and Evolution, 3(9):230-233.

Bowcock, A .M., J. R. Kidd, J. L. Mountain, J. M. Hebert, L. Carotenuto, K. K. Kidd, and L. L. Cavalli-Sforza. (1991). Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. Proceedings of the National Academy of Sciences USA, 88:839-843.

Brace, C. L. (1995). Region does not mean "race" - reality versus convention in forensic anthropology. Journal of Forensic Sciences, 40:171-175.

Braithwaite, M. (1974). *Max Braithwaite's Ontario*. J. J. Douglas Limited, West Vancouver, British Columbia, Canada.

Braun, A., D. Reuter, B. Darnhofer-Demar, K. Tang, D. Little, and H. Köster. (1997). Improved analysis of microsatellites using mass spectrometry. Genomics, 46(1):18-23.

Bray, M., and E. Epp. (1984). *A vast and magnificent land.* The Ontario Ministry of Northern Affairs, Toronto, Ontario.

Brenner, C., and J. W. Morris. (1990). Paternity index calculations in single locus hypervariable DNA probes: validation and other studies, p.21-53. In Proceedings for the International Symposium on Human Identification 1989. Promega Corporation, Madison, WI.

Brock, G. (1978). *The province of Northern Ontario*. Highway Book Shop, Cobalt, Ontario.

Brues, A. M. (1977). *People and races*. Macmillan Publishing Company, Inc., New York, New York.

Budowle, B., R. Chakraborty, A. M. Giusti, A. J. Eisenberg, and R. C. Allen. (1991). Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. American Journal of Human Genetics, 48:137-144.

Budowle, B., T. R. Moretti, A. L. Baumstark, D. A. Defenbaugh, and K. M. Keys. (1999). Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. Journal of Forensic Sciences, 44(6):1277-1286.

Buel, E., M. Schwartz, and M. LaFountain. (1998). Capillary electrophoresis STR analysis: comparison to gel-based systems. Journal of Forensic Sciences, 43(1):164-170.

Buroker, N., J. Day, D. Maclaren, K. Sweeny, C. R. Scott, and S-H. Chen. (1997). Population genetics of the D4S139, D10S28, D17S74 and D17S70 VNTR loci among Asian, Black, Caucasian, Hispanic and Native American populations from Seattle. Gene Geography, 11:1-14.

Busch, C. P., S. B. Harris, A. J. G. Hanley, B. Zinman, and R. A. Hegele. (1999). The ADD1 G460W polymorphism is not associated with variation in blood pressure in Canadian Oji-Cree. Journal of Human Genetics, 44(4):225-229.

Butler, J. M., B. R. McCord, J. M. Jung, and R. O. Allen. (1994). Rapid analysis of the short tandem repeat HUMTHO1 by capillary electrophoresis. Biotechniques, 17(6):1062-1070.

Caldwell, S. H., and R. Popenoe. (1995). Perceptions and misconceptions of skin color. Annals of Internal Medicine, 122(8):614-617.

Canadian Diabetes Association, Canadian Medical Association. (1998). 1998 clinical practice guidelines for the management of diabetes in Canada. Canadian Medical Association Journal, 158 (8 Suppl):S1-S29.

Canadian Pediatric Society. (1994). Diabetes and the First Nations. Canadian Journal of Pediatrics, 1:222-224.

Carracedo, A., and M. V. Lareu. (1998). Development of new STRs for forensic casework: Criteria for selection, sequencing & population data and forensic validation. In Proceedings for the International Symposium on Human Identification. Promega Corporation, Madison, WI.

Cavalli-Sforza, L. L. (1973). Analytic review: some current problems of human population genetics. American Journal of Human Genetics, 25:82-104.

Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza. (1994). *The history and Geography of human genes*. Princeton University Press, Princeton, NJ.

Chagnon, Y. C., I. B. Borecki, L. Pérusse, S. Roy, M. Lacaille, M. Chagnon, M. A. Ho-Kim, T. Rice, M. A. Province, D. C. Rao, and C. Bouchard. (2000). Genome-wide search for genes related to the fat-free body mass in the Québec family study. Metabolism, 49(2):203-207.

Chakraborty, R. (1992). Sample size requirements for addressing the population genetic issues of forensic use of DNA typing. Human Biology, 64:141-159.

Chakraborty, R., Y. Zhong, L. Jin, and B. Budowle. (1994). Nondetectability of restriction fragments and independence of DNA fragment sizes within and between loci in RFLP typing of DNA. American Journal of Human Genetics, 55:391-401.

Chakraborty, R., and Y. Zhong. (1994). Statistical power of an exact test of Hardy-Weinberg proportions of genotypic data at a multi-allelic locus. Human Heredity, 44:1-9.

Chakraborty, R., D. N. Slivers, B. Su, Y. Zhong, and B. Budowle. (1999). Nucleic acids-the utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems (review). Electrophoresis, 20(8):1682-1696.

Cohen, J. E. (1990). DNA fingerprinting for forensic identification: Potential effects on data interpretation of subpopulation heterogeneity and band number variability. American Journal of Human Genetics, 46(2):358-368.

Committee on DNA Forensic Science. (1996). *The evaluation of forensic DNA evidence*. Committee on DNA Forensic Science: an update. National Research Council, National Academy Press, Washington, D. C.

Craig, J., S. Fowler, L. A. Burgoyne, A. C. Scott, and H. W. Harding. (1988). Repetitive deoxyribonucleic acid (DNA) and human genome variation. A concise review relevant to forensic biology. Journal of Forensic Sciences, 33:1111-1126.

Da Silva, W. A., M. C. Bortolini, D. Meyer, F. M. Salzano, J. Elion, R. Krishnamoorthy, M. P. Schneider, D. C. De Guerra, Z. Layrisse, H. M. Castellano, T. D. Weimer, and M. A. Zago. (1999). Genetic diversity of two African and sixteen South American populations determined on the basis of six hypervariable loci. American Journal of Physical Anthropology, 109(4):425-437.

Dawson, K .C. A. (1977). An application of the direct historical approach to the Algonkians of Northern Ontario. Canadian Journal of Archaeology, 1:151-181.

Dean, H., R. Mundy, and M. Moffatt. (1992). Non-insulin dependent diabetes mellitus in Indian children in Manitoba. Canadian Medical Association Journal, 147:52-57.

Decorte, R., and J. J. Cassiman. (1996). Evaluation of the ALF DNA sequencer for high-speed sizing of short tandem repeat alleles. Electrophoresis, 17:1542-1549.

Deka, R., L. DeCroo, L. Jin, S. T. McGarvey, F. Rothhammer, R. E. Ferrell, and R. Chakraborty. (1994). Population genetic characteristics of the D1S80 locus in seven human populations. Human Genetics, 94:252-258.

Deka, R., M. D. Shriver, L. M. Yu, R. E. Ferrell, and R. Chakraborty. (1995). Diversity at short tandem repeat loci in diverse populations of the world: intra- and inter-population . Electrophoresis, 16(9):1659-1664.

Department of Indian Affairs and Northern Development. (2002). *Basic Departmental Data 2001*. Minister of Indian Affairs and Northern Development, Ottawa, Canada.

Devlin, B., N. Risch, and K. Roeder. (1990). No excess of homozygosity at loci used for DNA fingerprinting. Science, 249:1416-1420.

Devlin, B., and N. Risch. (1992). Ethnic differentiation at VNTR loci, with special reference to forensic applications. American Journal of Human Genetics, 51(3):534-548.

Di Renzo, A., C. Peterson, J. C. Garza, A. M. Valdes, M. Slatkin, and N. B. Freimer. (1994). Mutational processes of simple-sequence repeat loci in human populations. Proceedings of the National Academy of Sciences USA, 91:3166-3170.

DNA Recommendations. (1992). 1992 report concerning recommendations of the DNA Commission of the International Society of Forensic Haemogenetics relating to the use of PCR-based polymorphisms. International Journal of Legal Medicine, 105:63-64.

DNA Recommendations. (1994). 1994 report concerning further recommendations of the DNA Commission of the ISFH regarding PCR-based polymorphisms in STR (short tandem repeat) systems. International Journal of Legal Medicine, 107:159-160.

Drmić, I., M. S. Schanfield, S. Andjelinović, R. Galavotti, M. D. Gojanović, E. Trabetti, D. Marasović, D. Primorac, and P. F. Pignatti. (1998). Allele frequencies of six highly polymorphic DNA loci in the Croatian population. Human Biology, 70(5):949-957.

Drobnič, K., A. Regent, and B. Budowle. (2000). The Slovenian population data on the PCR based loci HLA-DQA1, LDLR, GYPA, HBGG, D7S8, GC, and D1S80. Journal of Forensic Sciences, 45(3):689-691.

Edwards, A., A. Civitello, H. A. Hammond, and C. T. Caskey. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. American Journal of Human Genetics, 49:746-756.

Edwards, A., H. A. Hammond, L. Jin, C. T. Caskey, and R. Chakraborty. (1992). Genetic Variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics, 12:241-253.

Evett, I. W., P. D. Gill, J. K. Scranage, and B. S. Weir. (1996). Establishing the robustness of short-tandem-repeat statistics for forensic applications. American journal of Human Genetics, 58:398-407.

Evett, I. W., and B. S. Weir. (1998). *Interpreting DNA evidence: statistical genetics for forensic scientists*. Sunderland, MA: Sinauer.

Fisher, R. A. (1951). Standard calculations for evaluating a blood group system. Heredity, 5:95-102.

Foreman, L. A., J. A. Lambert, and I. W. Evett. (1998). Regional genetic variation in Caucasians. Forensic Science International, 95:27-37.

Frégeau, C., and R. Fourney. (1993). DNA typing with flourescently tagged short tandem repeats: a sensitive and accurate approach to human identification. Biotechniques, 15:100-119.

Frégeau, C. J., W. F. Tan-Siew, K. H. Yap, G. R. Carmody, S. T. Chow, and R. M. Fourney. (1998). Population genetic characteristics of the STR loci D21S11 and FGA in eight diverse human populations. Human Biology, 70(5):813-844.

Gaedigk, A.,W .L. Casley, R. F. Tyndale, E. M. Sellers, M. Jurima-Romet, and J.S. Leeder. (2001). Cytochrome P4502C9 (CYP2C9) allele frequencies in Canadian Native Indian and Inuit populations. Canadian Journal of Physiology and Pharmacology, 79(10):841-847.

Gécz, J. (1991). PCR amplification of large VNTR alleles of D17S5 (YNZ22) locus. Nucleic Acids Research, 19(20):5806.

Gill, P., A. J. Jeffreys, and D. J. Werrett. (1985). Forensic application of DNA 'fingerprints'. Nature, 318:577-579.

Gill, P., K. M. Sullivan, and D. J. Werrett. (1990). The analysis of hypervariable DNA profiles: problems associated with the objective determination of the probability of a match. Human Genetics, 85:75-79.

Gill, P., A. Urquhart, E. Millican, N. Oldroyd, S. Watson, R. Sparkes, and C. P. Kimpton. (1996). A new method of STR interpretation using inferential logic-Development of a criminal intelligence database. International Journal of Legal Medicine, 109:14-22.

Goldsby, R. A. (1977). *Race and races*. Second Edition. Macmillan Publishing Company, Inc., New York, New York.

Grubić, Z., K. Štingl, and A. Kaštelan. (2002). Determination of polymorphism at eight STR loci in the Croatian population using automated detection. Forensic Science International, 127 (1):147-149.

Guglielmino, C. R., and A. Silvestri. (1995). Surname sampling for the study of the genetic structure of an Italian province. Human Biology, 67:613-628.

Guo, S. W., and E. A. Thompson. (1992). Performing the exact Hardy-Weinberg proportion for multiple alleles. Biometrics, 48:361-372.

Harris, S. B., J. Gittelsohn, A. J. G. Hanley, A. Barnie, T. Wolever, J. Gao, A. Logan, and B. Zinman. (1997). The prevalence of NIDDM and associated risk factors in native Canadians. Diabetes Care, 20:185-97.

Hauge, X. Y., and M. Litt. (1993). A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR. Human Molecular Genetics, 2(4):411-415.

Health Canada. (2000). *Diabetes among Aboriginal (First Nations, Inuit and Métis) people in Canada: the evidence*. Government of Canada.

Hedgecock, D., and F. Sly. (1990). Genetic drift and effective population sizes of hatchery-propagated stocks of the Pacific oyster, Crassostrea gigas. Aquaculture, 88:21-38.

Holgersson, S., J-A. Karlsson, A. Kihlgren, B. Rosen, P. Savolainen, and U. Gyllensten. (1994). Fluorescent-based typing of the two short tandem repeat loci HUMTH01 and HUMACTBP2: Reproducibility of size measurements and genetic variation in the Swedish population. Electrophoresis, 15(17):890-895.

Holt, C. L., C. Stauffer, J. M. Wallin, K. D. Lazaruk, T. Nguyen, B. Budowle, and P. S. Walsh. (2000). Practical applications of genotypic surveys for forensic STR testing. Forensic Science International, 112:91-109.

Hou, Y., M. Prinz, and M. Staak. (1994). Comparison of different tests for deviation from Hardy-Weinberg equilibrium of AMPFLP population data. Advances in Forensic Haemogenetics, 5:468-470.

Ingvarsson, S., V. Finnsdottir, A. Sigurdsson, and G. Geirsson. (2000). Population studies and validation of paternity determinations by six microsatellite loci. Journal of Forensic Sciences, 45(3):692-695).

Jaffe, A. J. 1992. *The first immigrants from Asia: A population history of the North American Indians*. Plenum Press, New York.

Jantz, R. L., and L. Meadows. (1995). Population structure of Algonquin speakers. Human Biology, 67(3):375-386.

Jeffreys, A. J., V. Wilson, and S. L. Thein. (1985). Hypervariable 'minisatellite' regions in human DNA. Nature, 314(6006):67-73.

Jeffreys, A. J., V. Wilson, R. Neumann, and J. Keyte. (1988). Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. Nucleic Acids Research, 16:10953-10971.

Jones, D. A. (1972). Blood samples: probability of discrimination. Journal of Forensic Science Society, 12:355-359.

Kaye, D. H. (1993). DNA Evidence: Probability, population genetics, and the courts. Harvard Journal of Law and Technology, 7:101-172.

Kawasaki, E. (1990). *Sample preparation from blood, cells, and other fluids.* In: PCR Protocols: A Guide to Methods and Applications. Academic Press, 146-152.

Kidd, J. R., F. L. Black, K. M. Weiss, I. Balazs, and K. K. Kidd. (1991). Studies of three Amerindian populations using nuclear DNA polymorphisms. Human Biology, 63(6):775-794.

Kimpton, C. P., D. Fisher, S. Watson, M. Adam, A. Urquhart, and P. Gill. (1994). Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. International Journal of Legal Medicine, 106:302-311.

Kimura, M., and T. Ohta. (1975). Distribution of allelic frequencies in a finite population under stepwise production of neutral alleles. Proceedings of the National Academy of Sciences USA, 72(7):2761-2764.

Kimura, M., and T. Ohta. (1978). Stepwise mutation model and distribution of allelic frequencies in a finite population. Proceedings of the National Academy of Sciences USA, 75(6):2868-2872.

Kwok, S., and R. Higuchi. (1989). Avoiding false positives with PCR. Nature, 339:237-238.

Lander, E. S. (1989). DNA fingerprinting on trial. Nature, 339(6225):501-505.

Lareu, M. V., S. Barral, A. Salal, and A. Carracedo. (1998). Sequence variation of a variable short tandem repeat at the D18S535 locus. International Journal of Legal Medicine, 111:337-339.

Launey, S., M. Barre, A. Gerard, and Y. Naciri-Graven. (2001). Population bottleneck and effective size in Bonamia ostreae-resistant populations of Ostrea edulis as inferred by microsatellite markers. Genetical Research, 78(3):259-270.

Leberg, P. (1992). Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. Evolution, 46:477-494.

Lessios, H. A. (1992). Testing electrophoretic data for agreement with Hardy-Weinberg expectations. Marine Biology, 112:517-523.

Lewontin, R. C., and D. L. Hartl. (1991). Population genetics in forensic DNA typing. Science, 254(5039):1745-1750.

Lewontin, R. C. (1997). The apportionment of Human Diversity. In: *The concept of "race" in natural and social science*. Garland Publishing, Inc., New York, New York..

Li, C. C., and D. G. Horvtiz. (1953). Some methods of estimating the inbreeding coefficient. American Journal of Human Genetics, 5:107-117.

Lindgren, C. M., E. Widén, T. Tuomi, H. Li, P. Almgren, T. Kanninen, O. Melander, J. Weng, M. Lehto, and L. C. Groop. (2002). Contribution of known and unknown susceptibility genes to early-onset diabetes in Scandinavia: Evidence for heterogeneity. Diabetes, 51:1609-1617.

Lins, A. M., C. J. Sprecher, C. Puers, and J. W. Schumm. (1996). Multiplex sets for the amplification of polymorphic short tandem repeat loci - silver stain and fluorescence detection. Biotechniques, 20(5):882-889.

Litt, M., and J. A. Luty. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. American Journal of Human Genetics, 44:397-401.

Litt, M., X. Hauge and V. Sharma. (1993). Shadow bands seen when typing polymorphic dinucleotide repeats: some causes and cures. Biotechniques, 15(2):280-284.

Luijten, S. H., A. Dierick, J. Gerard, B. Oostermeijer, L. E. L. Raijmann, and H. C. M. Den Nijs. (2000). Population size, genetic variation, and reproductive success in a rapidly declining, self-incompatible perennial (Arnica montana) in the Netherlands. Conservation Biology, 14(6):1776-1787.

Lygo, J. E., P. E. Johnson, D. J. Holdaway, S. Woodroffe, J. P. Whitaker, T. M. Clayton, C. P. Kimpton, and P. Gill. (1994). The validation of short tandem repeat (STR) loci for use in forensic casework. International Journal of Legal Medicine, 107:77-89.

MacMillan, H., A. MacMillan, D. Offord, and J. Dingle. (1996). Aboriginal health. Canadian Medical Association Journal, 155:1569-1578.

Magnuson, V. L., D. S. Ally, S. J. Nylund, Z. E. Karanjawala, J. B. Rayman, J. I. Knapp, A. L. Lowe, S. Ghosh, and F. S. Collins. (1996). Substrate nucleotide-determined non-templated addition of adenine by Taq DNA polymerase: implications for PCR-based genotyping and cloning. Biotechniques, 21(4):700-709.

Martin, L. J., M. H. Crawford, T. Koertvelyessy, D. Keeping, M. Collins, and R. Huntsman. (2000). The population structure of ten Newfoundland outports. Human Biology, 72(6):997-1016.

Martínez-Jarreta, B., B. Budowle, E. Abecia, B. Bell, Y. Casalod, and M. Castellano. (1998). PM and D1S80 loci gene frequencies in the Zaragoza population of Northern Spain. Journal of Forensic Sciences, 43(5):1094-1096.

Martínez-Jarreta, B., P. N. Marco, E. A. Martínez, M. V. Lareu Huidobro, and B. Budowle. (2000). Population genetics of the D1S1656, D12S391, and D18S535 in Asturias (North Spain). Journal of Forensic Sciences, 45(2):442-444.

McCord, B. R., J. M. Jung, and E. A. Holleran. (1993). High resolution capillary electrophoresis of forensic DNA using a non-gel sieving buffer. Journal of Liquid Chromatography, 16(9&10):1963-1981.

McVey, K. (1979). *The demographic bases of Canadian society*. M^{c}Graw-Hill Ryerson Limited, Canada.

Meldgaard, M., and N. Morling. (1997). Detection and quantitative characterization of artificial extra peaks following polymerase chain reaction amplification of 14 short tandem repeat systems used in forensic investigation. Electrophoresis, 18:1928-1935.

Meng, H. Y., Y. P. Hou, G. D. Chen, Y. B. Li, J. Wu, H. Walter, and M. Prinz. (1999). Frequencies of D8S384 alleles and genotypes in European, African-American, Chinese, and Japanese populations. Journal of Forensic Sciences, 44(6):1273-1276.

Micka, K., E. A. Amiott, T. L. Hockenberry, C. J. Sprecher, A. M. Lins, D. R. Rabbach, J. A. Taylor, J. W. Bacher, D. E. Glidewell, S. D. Gibson, C. A. Crouse, and J. W. Schumm. (1999). TWGDAM validation of a nine-locus and a four-locus fluorescent STR multiplex system. Journal of Forensic Sciences, 44(6):1243-1257.

Möller, A., E. Meyer, and B. Brinkmann. (1994). Different types of structural variation in STRs: HumFES/FPS, HumVWA and HumD21S11. International Journal of Legal Medicine, 106:319-323.

Monforte, J. A., and C. H. Becker. (1997). High-throughput DNA analysis by time-of-flight mass spectrometry. Nature Medicine, 3(3):360-362.

Montigny, E. A., and L. Chambers. (2000). *Ontario since confederation: a reader*. University of Toronto Press, Canada.

Morin, P. A., J. J. Moore, R. Chakraborty, L. Jin, J. Goodall, and D. S. Woodruff. (1994). Kin selection, social structure, gene flow, and the evolution of chimpanzees. Science, 265:1193-1201.

Mullis, K. B., and F. A. Faloona. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods in Enzymology, 155:335-350.

Murray, J., V. Sheffield, J. L. Weber, G. Duyuk, and K. H. Buetow. (1995). Human STS CHLC.GATA13.p9341 (GenBank accession number G07985). Cooperative Human Linkage Center, University of Iowa, unpublished.

Nei, M., and A. K. Roychoudhury. (1974). Sampling variances of heterozygosity and genetic distance. Genetics, 76:379-390.

Nei, M., T. Maruyama, and R. Chakraborty. (1975). The bottleneck effect and genetic variability in populations. Evolution, 29:1-10.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics, 89:583-590.

Nei, M., and A. K. Roychoudhury. (1997). Genetic relationship and evolution of human races. In: *The concept of "race" in natural and social science*. Garland Publishing, Inc., New York, New York.

Nichols, R. A., and D. J. Balding. (1991). Effects of population structure on DNA fingerprint analysis in forensic science. Heredity, 66:297-302.

Nievas, P., B. Martinez-Jarreta, E. Abecia, and M. V. Lareu. (1999). Fluorescence-based amplification of the STR loci D18S535, D1S1656 and D12S391 in a population sample from Aragon (North Spain). International Journal of Legal Medicine, 113:58-59.

Ohta, T. (1982). Linkage disequilibrium due to random genetic drift in finite subdivided populations. Proceedings of the National Academy of Sciences USA, 79:1940-1944.

Pamilo, P., and S. Vario-Aho. (1984). Testing genotype frequencies and heterozygosities. Marine Biology, 79(1):99-100.

Pamilo, P., and M. Nei. (1988). Relationship between gene trees and species trees. Molecular Biology and Evolution, 5:568-583.

Pérez-Lezaun, A., F. Calafell, E. Mateu, D. Comas, R. Ruiz-Pacheco, and J. Bertranpetit. (1997). Microsatellite variation and the differentiation of modern humans. Human Genetics, 99:1-7.

Phillips, C. P., C. D. Syndercombe, M. V. Lareu, J. Hasskamp, and A. Carracedo. (1998). Band shift analysis of three base-pair repeat alleles in the short tandem repeat locus *D12S391*. Forensic Science International, 93(2):79-88.

Piazza, A., G. Rendine, G. Zei, A. Moroni, and L. L. Cavalli-Sforza. (1987). Migration rates of human populations from surname distribution. Nature, 329:714-716.

Picornell, A., A. Miguel, J. A. Castro, M. Misericordia Ramon, R. Arya, and M. H. Crawford. (1996). Genetic variation in the population of Ibiza (Spain): genetic structure, geography, and language. Human Biology, 68(6):899-913.

Prout, T. (1973). Appendix to the paper by J. B. Mitton and R. K. Koehn. Genetics, 73:493-496.

Pusey, A., and M. Wolf. (1996). Inbreeding avoidance in animals. Trends in Ecology and Evolution, 11(5):201-206.

Rahman, Z., T. Afroze, and B. S. Weir. (2001). DNA typing results from two urban subpopulations of Pakistan. Journal of Forensic Sciences, 46(1):111-115.

Raymond, M., and F. Rousset. (1995). GENEPOP (Version 1.2): A population genetics software for exact tests and ecumenicism. Journal of Heredity, 86:248-249.

Robertson, J. M. (1994). Evaluation of native and denaturing polyacrylamide gel electrophoresis for short tandem repeat analysis. Advances in Forensic Haemogenetics, 5:320-322.

Rolf, B., B. Horst, A. Eigel, T. Sagansermsri, B. Brinkmann, and J. Horst. (1998). Microsatellite profiles reveal an unexpected genetic relationship between Asian populations. Human Genetics, 102(6):647-652.

Rousset, F., and M. Raymond. (1995). Testing heterozygote excess and deficiency. Genetics, 140:1413-1421.

Rudin, N., and K. Inman. (2002). *An introduction to forensic DNA analysis*. Second Edition, CRC Press LLC, Boca Raton, Florida.

Saitoh, H., S. Ueda, K. Kurosaki, and M. Kiuchi. (1998). The different mobility of complementary strands depends on the proportion of AC/GT. Forensic Science International, 91(2):81-90.

Samura, O., B. Pertl, S. Sohda, K. L. Johnson, A. Sekizawa, V. M. Falco, R. S. Elmes, and D. W. Bianchi. (2000). Female fetal cells in maternal blood: use of DNA polymorphisms to prove origin. Human Genetics, 107:28-32.

Sauer, N. J. (1992). Forensic anthropology and the concept of race: if races don't exist, why are forensic anthropologists so good at identifying them? Social Science and Medicine, 34(2):107-111.

Sbordoni, V., E. De Matthaeis, M. Cobolli Sbordini, G. La Rosa, and M. Mattoccia. (1986). Bottleneck effects and the depression of genetic variability in hatchery stocks of Penaeus japonicus (Crustacea, Decapoda). Aquaculture, 57:239-251.

Schlötterer, C., and D. Tautz. (1992). Slippage synthesis of simple sequence DNA. Nucleic Acids Research, 20(2):211-215.

Schmalzing, D., L. Koutny, A. Adourian, P. Belgrader, P. Matsudaira, and D. Ehrlich. (1997). DNA typing in thirty seconds with a microfabricated device. Proceedings of the National Academy of Sciences USA, 94:10273-10278.

Schull, J. (1978). *Ontario since 1867*. The Canadian Publishers M°Clelland and Stewart Limited, Toronto, Ontario, Canada.

Scozzari, R., F. Cruciani, P. Santolmazza, D. Sellitto, D. E. C. Cole, L. A. Rubin, D. Labuda, E. Marini, V. Succa, G. Vona, and A. Torroni. (1997). mtDNA and Y chromosome-specific polymorphisms in modern Ojibwa: implications about the origin of their gene pool. American Journal of Human Genetics, 60:241-244.

Scriver, C. R. (2001). Human genetics: lessons form Quebec populations. Annual Review of Genomics and Human Genetics, 2:69-101.

Senior, P. A., and R. Bhopal. (1994). Ethnicity as a variable in epidemiological research. British Medical Journal, 309:327-330.

Smith, J. R., J. D. Carpten, M. J. Brownstein, S. Ghosh, V. L. Magnuson, D. A. Gilbert, J. M. Trent, and F. S. Collins. (1995). Approach to genotyping errors caused by nontemplated nucleotide addition by Taq DNA polymerase. Genome research, 5(3):312-317.

SPSS Inc. SPSS for Vax, Version 11. Chicago, U. S. A.

Statistics Canada. (1996). Federal electoral district profile. Ottawa, Canada.

Statistics Canada. (2001 a). 2001 community profile. Ottawa, Canada.

Statistics Canada. (2001 b). Canadian Statistics. Ottawa, Canada.

Szathmary, E. J. E., and T. E. Reed. (1972). Caucasian admixture in two Ojibwa Indian communities in Ontario. Human Biology, 44(4):655-671.

Szathmary, E. J. E., J. F. Mohn, H. Gershowitz, R. M. Lambert, and T. E. Reed. (1975). The northern and southeastern Ojibwa: blood group systems and the causes of genetic divergence. Human Biology, 47(3):351-368.

Szathmary, E. J. E., D. W. Cox, H. Gershowitz, D. L. Rucknagel, and M. S. Schanfield. (1978). The northern and southeastern Ojibwa: serum proteins and red cell enzyme systems. American Journal of Physical Anthropology, 40:49-66.

Szathmary, E. J. E., and N. S. Ossenberg. (1978). Are the biological differences between North American Indians and Eskimos truly profound? Current Anthropology, 4:673-701.

Szathmary, E. J. E. (1984). Peopling of northern North America: clues from genetic studies. Acta Anthropogenetica, 8:79-109.

Szathmary, E. J. E. (1993). Genetics of Aboriginal North Americans. Evolutionary Anthropology, 1:202-220.

Tautz, D., and M. Renz. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Research, 12(10):4127-4138.

Tautz, D., M. Trick, and G. A. Dover. (1986). Cryptic simplicity in DNA is a major source of genetic variation. Nature, 322:652-656.

Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research, 17(16);6464-6471.

Thouez, J. P., A. Rannou, and P. Foggin. (1989). The other face of development: native population, health status and indicators of malnutrition-the case of the Cree and Inuit in northern Quebec. Social Science and Medicine, 29:965-974.

Tomàs, C., A. Picornell, J. A. Castro, and M. Misericòrdia Ramon. (2001). Genetic analyses in five Western Mediterranean populations: Variation at five tetrameric short tandem repeat loci. Human Biology, 73(3):349-363.

Turnbull, R. (1966). *Canada: the story of Ontario*. McGraw-Hill Company of Canada Limited.

Urquhart, A., C. P. Kimpton, and P. Gill. (1993). Sequence variability of the tetranucleotide repeat of the human beta-actin related pseudo-gene H-beta-Ac-psi-2 (ACTBP2) locus. Human Genetics, 92:637-638.

Urquhart, A., C. P. Kimpton, T. J. Downes, and P. Gill. (1994). Variation in short tandem repeat sequences - a survey of twelve microsatellite loci for use as forensic identification markers. International Journal of Legal Medicine, 107:13-20.

Urquhart, A., N. J. Oldroyd, C. P. Kimpton, and P. Gill. (1995 a). Highly discriminating heptaplex short tandem PCR system for forensic identification. Biotechiques, 18(1):116-121.

Urquhart, A. J., N. J. Oldroyd, T. Downes, M. Barber, R. Alliston-Greiner, C. P. Kimpton, and P. D. Gill. (1995 b). Selection of STR loci for forensic identification systems. 16th Congress of the International Society for Forensic Haemogenetics, Santiago De Compostela, pp115-117.

Wall, J. W., R. Williamson, M. Petrou, D. Papaloannou, and B. H. Parkin. (1993). Variation of short tandem repeats within and between populations. Human Molecular Genetics, 2(7):1023-1029.

Walsh, P. S., N. J. Fildes, and R. Reynolds. (1996). Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. Nucleic Acids Research, 24(14):2807-2812.

Ward, R. H., and C. F. Sing. (1970). A consideration of the power of the χ^2 test to detect inbreeding effects in natural populations. The American Naturalist, 104(938):355-365.

Watts, E. S. (1997). The biological race concept and diseases of modern man. In: *The concept of "race" in natural and social science*. Garland Publishing, Inc., New York, New York.

Weber, J. L., and P. E. May. (1989). *Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction.* American Journal of Human Genetics, 44:338-396.

Weir, B. S., and C. C. Cockerham. (1984). Estimating F-statistics for the analysis of population structure. Evolution, 38:1358-1370.

Weir, B. S. (1994). The effects of inbreeding on forensic calculations. Annual Review of Genetics, 28:597-621.

Wenz, H. M., J. M. Robertson, S. Menchen, F. Oaks, D. M. Demorest, D. Scheibler, B. B. Rosenblum, C. Wike, D. A. Gilbert, and J. W. Efcavitch. (1998). High-precision genotyping by denaturing capillary electrophoresis. Genome Research, 8(1):69-80.

White, R. (1985). *Ontario 1610-1985: a political and economic history*. Dundurn Press Limited, Toronto, Ontario, Canada.

Wiegand, P., M. V. Lareu, M. Schürenkamp, M. Kleiber, and B. Brinkman. (1999). D18S535, D1S1656 and D10S2325: three efficient short tandem repeats for forensic genetics. International Journal of Legal Medicine, 112:360-363.

Wiegand, P., E. Meyer, and B. Brinkmann. (2000). Microsatellite structures in the context of human evolution (review). Electrophoresis, 21(5):889-895.

Wijsman, E., G. Zei, A. Moroni, and L. L. Cavalli-Sforza. (1984). Surnames in Sardinia: Computation of migration matrices from surname distributions in different periods. Annals of Human Genetics, 48:65-78.

Wolff, R. K., P. Rosemaire, A. J. Jeffreys, and R. White. (1989). Unequal crossing over between homologous chromosomes is not the major mechanism involved in the generation of new alleles at VNTR loci. Genomics, 5: 382-384.

Wong, Z., V. Wilson, I. Patel, S. Povey, and A. J. Jeffreys. (1987). Characterization of a panel of highly variable minisatellites cloned from human DNA. Annals of Human Genetics, 51:269-288.

Wright, S. (1931). Evolution in Mendelian populations. Genetics, 16:97-159.

Wright, S. (1951). The genetical structure of populations. Annals of Eugenetics, 15:323-354.

Wyman A. R. and R. White. (1980). A highly polymorphic locus in human DNA. Proceedings of the National Academy of Sciences USA, 77(11):6754-6758.

Young, T. K., E. J. E. Szathmary, S. Evers, and B. Wheatly. (1990). Geographical distribution of diabetes among the native population of Canada: a national survey. Social Science and Medicine, 31:129-39.

Zago, M. A., W. A. Silva Jr., M. H. Tavella, S. E. B. Santos, J. F. Guerreiro, and M. S. Figueiredo. (1996). Interpopulation and intrapopulational genetic diversity of Amerindians as revealed by six variable number of tandem repeats. Human Heredity, 46:274-289.

Zar, J. H. (1996). *Biostatistical analysis*. Third Edition, Prentice-Hall, Inc., New Jersey.

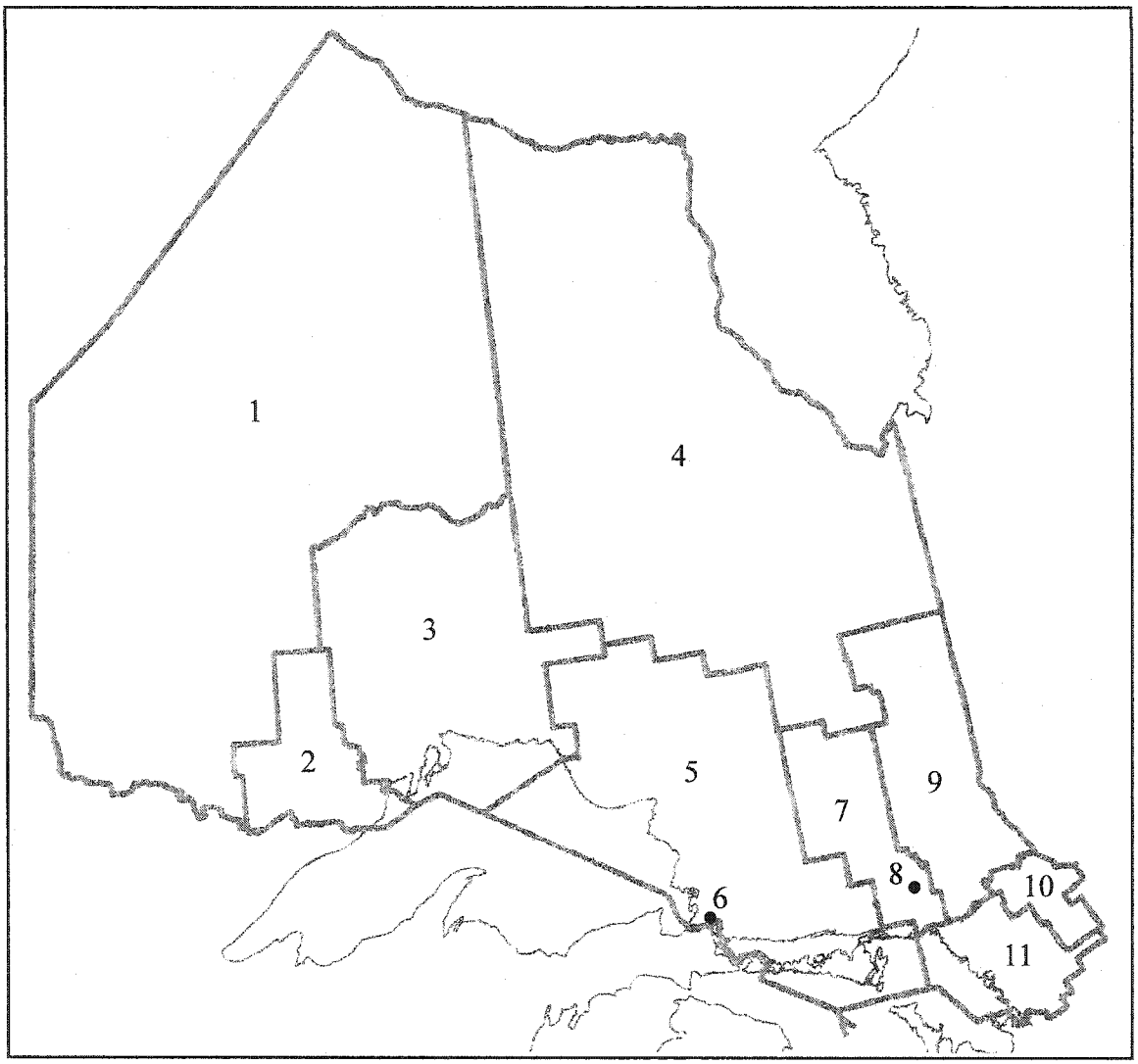
Zhivotovsky, L. K., S. Ahmed, W. Wang, and A. H. Bittles. (2001). The forensic DNA implications of genetic differentiation between endogamous communities. Forensic Science International, 119:269-272.

Appendix A: Sequence structure and length of alleles 9-16 of the D18S535 locus (Lareu *et al.*, 1998)

Allele	Sequence	Length (base pairs)
9	$(GATA)_9$	129
10	$(GATA)_{10}$	133
11	$(GATA)_{11}$	137
12	$(GATA)_{12}$	141
13	$(GATA)_{13}$	145
14	$(GATA)_{14}$	149
15	$(GATA)_{15}$	153
16	$(GATA)_{16}$	157

Complete sequence: TCA TGT GAC AAA AGC CAC ACC CAT AAC TTT TTG CCT CTA GAT AGA CAG ATA GAT **$(GATA)_{9-16}$** TAG ATT CTC TTT CTC TGC ATT CTC ATC TAT ATT TCT GTC T

Appendix B: Map indicating the locations of the eleven federal electoral districts of Northern Ontario



Northern Ontario Federal Electoral Districts

1. Kenora-Rainy River
2. Thunder Bay-Atikokan
3. Thunder Bay-Superior North
4. Timmins-James Bay
5. Algoma-Manitoulin
6. Sault Ste. Marie
7. Nickel Belt
8. Sudbury
9. Timiskaming-Cochrane
10. Nipissing
11. Parry Sound-Muskoka

Appendix C: Demographic characteristics of the Northern Ontario federal electoral districts (Statistics Canada, 1996)

	Kenora-Rainy River	Thunder Bay-Atikokan	Thunder Bay-Superior North	Timmins-James Bay	Algoma-Manitoulin	Sault Ste. Marie
Total Population	79550	78360	79680	79625	75120	80055
Total Population by Visible Minority Population	78970	77625	79055	79025	74610	79055
Total Visible Minority Population	570	1415	1595	885	465	995
Black	70	265	210	205	75	150
South Asian	120	230	265	85	65	240
Chinese	115	225	360	330	130	270
Korean	0	10	10	10	10	25
Japanese	135	295	195	70	0	25
Southeast Asian	30	95	145	15	0	20
Filipino	90	105	175	50	45	75
Arab/West Asian	0	20	105	45	70	80
Latin America	10	145	85	40	30	70
Others	0	25	0	20	15	25

Appendix C: (Continued) (Statistics Canada, 1996)

	Kenora-Rainy River	Thunder Bay-Atikokan	Thunder Bay-Superior North	Timmins-James Bay	Algoma-Manitoulin	Sault Ste. Marie
Multiple Visible Minority	0	0	45	15	25	15
Total Population by Immigrant Status	78985	77625	79055	79020	74610	79065
Non-Immigrant Population	73760	69255	69660	75830	69740	69255
Immigrants	5010	8305	9290	3115	4805	9745
Non-Permanent Residents	215	65	105	75	65	65
Total Population by Mother Tongue	77600	76725	77840	76875	73545	78165
English	61385	65265	61375	33235	58360	65910
French	1805	1475	4935	36290	10585	3050
Non-Official Languages	14410	9985	11530	7350	4600	9205
Total Aboriginal Population	20215	4515	6950	9420	7560	3345
Total Non-Aboriginal Population	58770	73115	72105	69595	67055	75715

Appendix C: (Continued) (Statistics Canada, 1996)

	Nickel Belt	Sudbury	Timiskaming-Cochrane	Nipissing	Parry Sound-Muskoka
Total Population	82575	86245	75770	76050	82855
Total Population by Visible Minority Population	82460	84885	74875	74880	81805
Total Visible Minority Population	700	2195	560	920	530
Black	295	685	135	205	115
South Asian	150	425	100	145	70
Chinese	80	455	210	270	160
Korean	0	0	15	10	20
Japanese	0	0	25	50	20
Southeast Asian	10	20	20	10	10
Filipino	55	130	10	90	25
Arab/West Asian	10	220	20	55	20
Latin America	25	175	15	25	20
Others	55	40	0	20	30
Multiple Visible Minority	20	45	10	40	40

Appendix C: (Continued) (Statistics Canada, 1996)

	Nickel Belt	Sudbury	Timiskaming-Cochrane	Nipissing	Parry Sound-Muskoka
Total Population by Immigrant Status	82450	84890	74880	74870	81820
Non-Immigrant Population	79035	75795	71965	70690	74835
Immigrants	3400	8980	2900	4100	6910
Non-Permanent Residents	15	115	15	80	75
Total Population by Mother Tongue	80760	82810	73540	73900	81475
English	49345	53115	42020	59045	76105
French	27750	19770	29080	11935	1185
Non-Official Languages	3665	9925	2440	2920	4185
Total Aboriginal Population	1935	3015	2510	3380	2235
Total Non-Aboriginal Population	80510	81875	72360	71495	79580

Appendix D: Demographic characteristics of Northern Ontario, Southern Ontario, Ontario, and Canada (Statistics Canada, 1996)

	Northern Ontario	Southern Ontario	Ontario	Canada
Total Population	875885	9877685	10753570	28846765
Total Population by Visible Minority Population	867215	9775565	10642780	28528100
Total Visible Minority Population	10830	1671205	1682035	3197460
Black	2410	353805	356215	573860
South Asian	1895	388160	390055	670585
Chinese	2605	388485	391090	860150
Korean	110	35290	35400	64835
Japanese	815	23460	24275	68130
Southeast Asian	375	75530	75905	172760
Filipino	850	116515	117365	234195
Arab/West Asian	645	118010	118655	244660
Latin America	640	85105	85745	176970
Others	230	51940	52170	69745
Multiple Visible Minority	255	34905	35160	61570

Appendix D: (Continued) (Statistics Canada, 1996)

	Northern Ontario	Southern Ontario	Ontario	Canada
Total Population by Immigrant Status	867270	9775525	10642795	28528125
Non-Immigrant Population	799820	7044555	7844375	23390340
Immigrants	66560	2657930	2724490	4971070
Non-Permanent Residents	890	73040	73930	166715
Total Population by Mother Tongue	784135	9686355	10470490	28125560
English	625160	7069470	7694630	16890615
French	147860	331430	479290	6636655
Non-Official Languages	80215	2216355	2296570	4598290
Total Aboriginal Population	65080	76445	141525	799010
Total Non-Aboriginal Population	802175	9699090	10501265	27729115

Appendix E: Protocols and compositions of solutions

Table 1. Lysis Buffer

Composition	Concentration
Sucrose	0.32 M
Tris-HCl	10 mM
$MgCl_2$	5 mM
TritonX-100	1 %

Autoclave the solution and store at 4°C.

Table 2. Digestion Buffer

Composition	Concentration
Tris-HCl	10 mM
KCl	50 mM
$MgCl_2$	2.5 mM
TritonX-100	0.45 %
Tween® 20	0.45 %
Gelatin	0.1 mg/ml

Autoclave the solution and store at 4°C.

Table 3. Agarose Gel Loading Buffer (6x)

Composition	Concentration
Sucrose	40 % (w/v)
Xylene Cyanol	0.25 % (w/v)
Bromophenol Blue	0.25 % (w/v)

Store at 4°C and dilute the solution by a factor of six before use.

Table 4. 25 X TAE Buffer (1 Litre Stock Solution)

Composition	Mass
Tris Base	121.0 g
Glacial Acetic Acid	28.6 ml
EDTA (disodium, dihydrate)	18.6 g

Adjust pH of the solution to 8.5 and dilute it 25 times before use in electrophoresis.

Table 5. 5 X TBE Buffer (1 Litre Stock Solution)

Composition	Concentration
Tris Base	0.45 M
Boric Acid	0.45 M
EDTA (free acid)	10 mM

Not necessary to adjust pH of this buffer; pH should be about 8.3.

Table 6. Polyacrylamide Gel Loading Buffer

Composition	Concentration
Xylene Cyanol	50 mg/µl
Bromophenol Blue	50 mg/µl
NaOH	1 N
Deionized Formamide	50 % v/v

Fill up to volume with double distilled water. Store at 4 °C.

Table 7. Fix/Stop Solution

Composition	Concentration
Glacial Acteic Acid	10 % v/v

Fill up to volume with double distilled water.

Table 8. Silver Staining Solution

Composition	Concentration
Silver Nitrate ($AgNO_3$)	1 g/L
37% Formaldehyde (H_2CO)	0.056 % v/v

Fill up to volume with double distilled water. Add formaldehyde after silver nitrate has been dissolved.

Table 9. Developing Solution

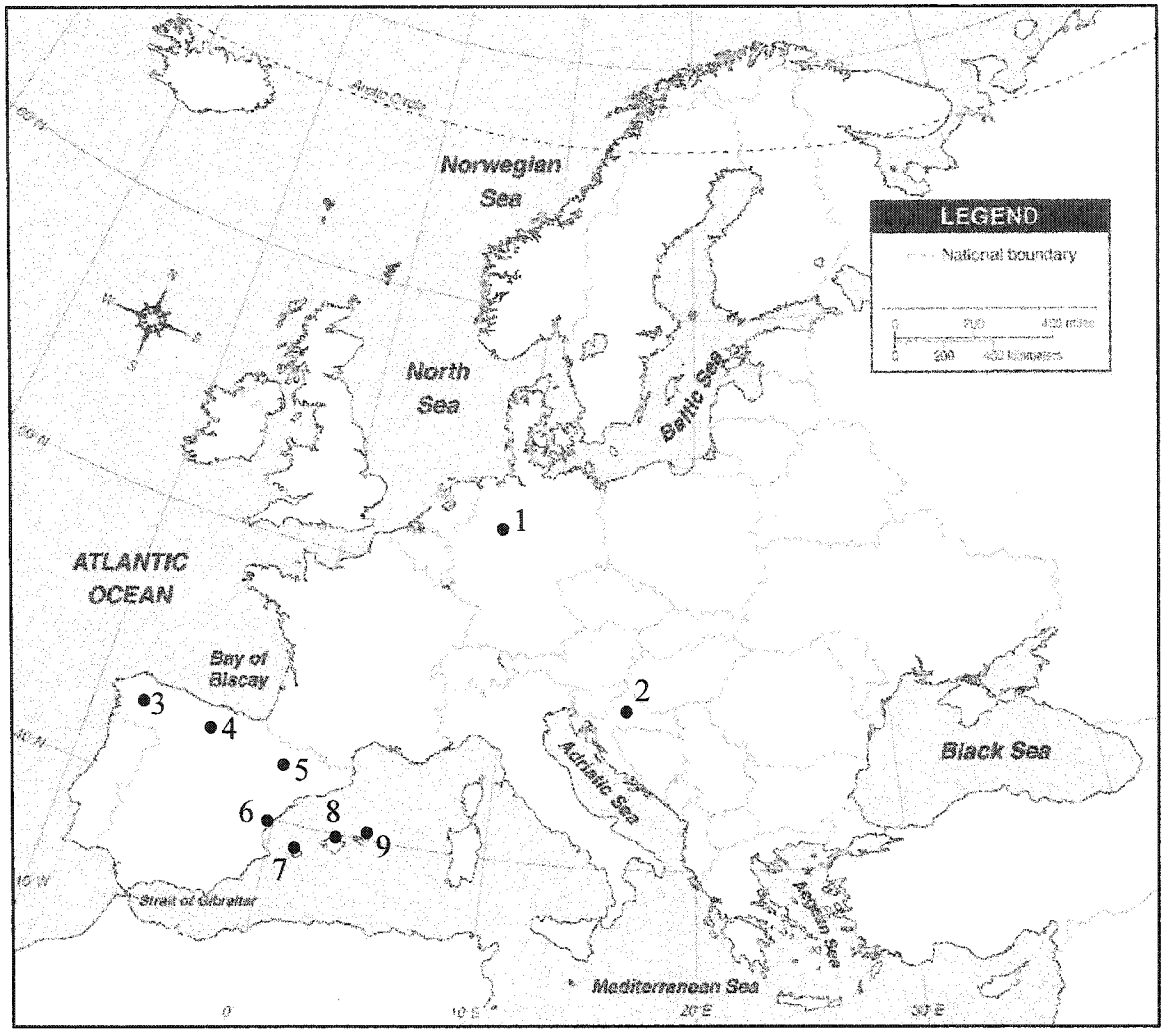
Composition	Concentration
37% Formaldehyde (H_2CO)	0.056 % v/v
Sodium Carbonate (Na_2CO_3)	30 g/ml
10 mg/ml Sodium Thiosulfate ($Na_2S_2O_3 \bullet 5H_20$)	2 mg/L

Fill up to volume with double distilled water. Prepare fresh and chill to 10 °C before each use. Add sodium thiosulfate a few minutes before use.

Appendix F: DNA purification using a Phenol/Chloroform/Isoamyl alcohol procedure

The Phenol/Chloroform/Isoamyl alcohol procedure involved adding 50 µl of Phenol:Chloroform:Isoamyl alcohol (25:24:1) to a microcentrifuge tube containing 50 µl of DNA extract in TE [10 mM Tris-HCL (pH 8.0), 1 mM EDTA]. The mixture was then inverted several times to mix and centrifuged at 18000 x g for 5 minutes to separate the phases. The aqueous phase was transferred to a clean tube. This procedure was repeated again followed by a Chloroform:Isoamyl alcohol (24:1) purification. The DNA was precipitated by adding ammonium acetate to a final concentration of 2 *M* and mixing by inversion several times. Twice the volume of 95 % ethanol (-20 °C) was then added to the microcentrifuge tube, mixed by inverting several times and left over night at -20 °C. The next day, the tubes were centrifuged for 30 minutes at 18000 x g to pellet the DNA, the 95 % ethanol was removed, and 200 µl of 70 % ethanol was added and allowed to sit for 5 minutes. The final pellet was collected by centrifugation (18000 x g, 20 minutes), was drained, remaining droplets removed under vacuum, air dried for 10 minutes, and resuspended in 25 µl of TE buffer. The tubes were then placed at 4 °C overnight.

Appendix G: Map indicating the locations of the previously published D18S535 studies from Europe



Previously Published D18S535 Populations From Europe

1. Halle Area, Northeastern Germany
2. Croatia
3. Galacia, Northwestern Spain
4. Asturias, Northern Spain
5. Aragon, Northeastern Spain
6. Valencia, Eastern Spain
7. Ibiza, Spain
8. Majorca, Spain
9. Minorca, Spain

Appendix H: Map indicating the location of the previously published D18S535 study from South America

Appendix I: Relative allele frequencies for the previously published D18S535 studies from Europe and South America

	Relative Frequency (f)				
Allele	Croatia (N = 167)	Halla Area, Northeastern Germany (N = 150)	Aragon, Northeastern Spain (N = 104)	Galacia, Northwestern Spain (N = 129)	Asturias, Northern Spain (N = 159)
9	0.099	0.086	0.053	0.106	0.110
10	0.015	0.016	0.005	0.008	0.006
11	0.051	0.022	0.024	0.015	0.009
12	0.191	0.205	0.216	0.153	0.233
13	0.311	0.325	0.288	0.331	0.314
14	0.207	0.203	0.260	0.232	0.204
15	0.111	0.133	0.135	0.150	0.113
15.2	0.015	0.000	0.000	0.000	0.000
16	0.000	0.010	0.014	0.004	0.009

N = the number of individuals

Appendix I: (Continued)

Allele	Relative Frequency (f)					
	Majorca, Spain (N = 73)	Minorca, Spain (N = 57)	Ibiza, Spain (N = 57)	Chuetas, Spain (N = 56)	Valencia, Eastern Spain (N = 55)	Puna, Northwestern Argentina (N = 99)
9	0.096	0.105	0.246	0.045	0.091	0.040
10	0.000	0.000	0.009	0.000	0.000	0.020
11	0.027	0.000	0.035	0.018	0.009	0.040
12	0.192	0.193	0.184	0.268	0.182	0.212
13	0.342	0.307	0.202	0.384	0.391	0.253
14	0.199	0.237	0.202	0.232	0.218	0.333
15	0.137	0.158	0.123	0.054	0.100	0.101
15.2	0.000	0.000	0.000	0.000	0.000	0.000
16	0.007	0.000	0.000	0.000	0.009	0.000

N = the number of individuals