

VOLUME VII

C A N A D A

IN THE COURT OF QUEEN'S BENCH OF NEW BRUNSWICK

TRIAL DIVISION

JUDICIAL DISTRICT OF FREDERICTON

B E T W E E N:

HER MAJESTY THE QUEEN

- and -

ALLAN JOSEPH LEGERE

VOIR DIRE PROCEEDINGS held before Mr. Justice
David M. Dickson at the Burton Courthouse, Burton,
New Brunswick, on the 6th day of May, 1991.

APPEARANCES:

Graham Sleeth, Esq.,)
Anthony Allman, Esq.,) for the Crown.
John Walsh, Esq.,)

Weldon Furlotte, Esq.,)
Michael Ryan, Esq.,) for the Defence.

Copyright 1991
Department of Justice
Province of New Brunswick

Nancy Patterson
Court Reporter

May 6/91

I N D E X

Exhibit	Description	Page
VD-57	Curriculum Vitae of Dr. Carmody	3
VD-58	Two-page document	12
VD-59	Xerox copy of slide	14
VD-60	Xerox copy of slide	15
VD-61	Xerox copy of slide	16
VD-62	Xerox copy of slide	18
VD-63	Xerox copy of slide	18
VD-64	Total rebin Caucasian Database dated December 3/90	48
VD-65	Compilation of calculations	53

* * * * *

Dr. George Carmody - Direct Examination - Page 2
- Cross Examination - Page 73

(COURT RECONVENED MAY 6, 1991 at 9:00 a.m.)

THE COURT: Now, let me see. The accused is present. Same counsel as on Friday. Mr. Walsh you were going to call a witness. I had a message yesterday from the Clerk saying there has been some change in the plan.

MR. WALSH: Yes. Yesterday I received a call from Dr. Kidd. He became ill over the Saturday night. He had hoped to make the flight and he just couldn't. He told me he will do everything he can -- he hopes he will be feeling better by the end of this week and we are to check with his office this week and hopefully be able to get him to testify next week.

THE COURT: He is ill is he?

MR. WALSH: Yes. I understand Dr. Kidd is ill. I don't know anything about Dr. Kidd's general health. I just know that he is ill and according to my people that does happen on occasion with Dr. Kidd because he does have some health problems.

THE COURT: Where is he?

MR. WALSH: Yale University.

THE COURT: Yale is in --

MR. WALSH: New Haven, Connecticut. He hoped to make the flight yesterday and he said he was still dizzy. He was still having problems and he didn't think he could take the flight and then everything that follows, so he is trying to free up next week and hopefully he will be feeling better at the end of the week and he was very apologetic obviously to --

THE COURT: Well, he does appreciate does he the --

MR. WALSH: Oh, very much so, My Lord. This is something -- apparently Dr. Kidd has spent a lot of time in his work in other areas and different countries and things of that nature and as a result his health has suffered.

This is something that is understood and he is very apologetic to the court. He is not that type of person I can assure you, My Lord. He is a very conscientious man.

THE COURT: Well, what does this do to your schedule?

MR. WALSH: Well, I have Dr. Carmody here. As we indicated, Dr. Kidd was in fact going to be put in slightly out of place so it doesn't affect my schedule in that great a regard. I have Dr. Carmody available to testify and then to be followed by Dr. Bowen. At that point we could see how far into the week we are at that point, but I expect I would have witnesses available as each finishes.

THE COURT: But I mean you see us putting the whole week to good advantage?

MR. WALSH: Yes. I have no intention of putting the court in a position where we don't use the time that is available to us. It is just a matter of the order in which I called my witnesses. As you can appreciate it is difficult sometimes, human --

THE COURT: Oh, I realize we have to make allowances.

MR. WALSH: Thank you, My Lord.

THE COURT: You are calling Dr. Carmody.

MR. WALSH: Yes.

THE COURT: We are in a voir dire, of course, again.

DR. GEORGE CARMODY, called as a witness, having been duly sworn on the voir dire, testified as follows:

DIRECT EXAMINATION BY MR. WALSH:

Q. Give the court your name please?

A. My name is George Carmody.

Q. Your present occupation?

A. I am an Associate Professor of biology at Carleton University.

Q. I show you, Doctor, this particular document and ask you if you could identify it for me please?

A. This is my curriculum vitae or more commonly known perhaps as resume.

MR. WALSH: I would have this entered, My Lord, please.

THE COURT: That would be VD-57.

MR. WALSH: With Your Lordship's permission I wish to take Dr. Carmody through his curriculum vitae.

THE COURT: All right.

Q. Doctor, you have a Bachelor of Science with major in chemistry from Columbia University, New York.

A. That's correct.

Q. You received that in 1960.

A. 1960.

Q. You have a Ph.D. in zoology from Columbia University in New York and you received that in 1967.

A. Yes.

Q. You were a Post Doctoral Fellow in population biology at the University of Chicago between 1967 and 1969.

A. Yes. That's correct.

Q. At that particular time, Doctor, did you collaborate with anyone in the particular fields of population biology?

A. Yes, there were a few people but notably I was working in the laboratory of Dr. Richard Lewontin.

Q. And he is presently at Harvard?

A. He is presently at Harvard University, yes.

- Q. You were a Senior Fellow in genetics at the University of Nottingham in England between 1976 and 1977.
- A. That's correct.
- Q. What field would that be in? What part of the field of genetics would you have been working in at that time?
- A. That is in population genetics.
- Q. You were a visiting researcher in genetics at the National Institute of Environmental Health Scientists. I understand that is in North Carolina.
- A. That's right. It is a Federal U.S. Government lab in the Research Triangle in North Carolina.
- Q. And the area of genetics that you were working in?
- A. In population genetics.
- Q. You have been a visiting professor in genetics at the University of Hawaii.
- A. That's correct. This past year.
- Q. That is -- what field were you actually working in?
- A. In population genetics.
- Q. Doctor, perhaps if you could just tell the court briefly, what is population genetics and what application would it have to the matters here. Just briefly.
- A. Population genetics is the study of the behaviour of genes in populations and how changes occur in the course of time in those populations, and particularly how different populations can be different genetically from others.
- Q. What application would it have in forensics?
- A. Well, in fact to calculate what the probabilities are or what the frequency of a particular genotype is for these VNTR loci that we are using for forensic identification, one has to know what the frequencies of

these various types are in real human populations.

- Q. Are there different -- under the umbrella of population genetics, Doctor, are there various -- studies of various life forms or are there specialties or sub-specialties?
- A. Yes, very much so. You can break down population genetics first by the type of organisms that a person would study and basically you have people studying animal populations, plant populations. You have people being more theoretically applied in their interests and in other cases you have them more interested in experimental details.
- Q. Could you give some examples of what you mean by theoretical application, experimental application and the various kinds of studies that go on under population genetics?
- A. Well, to give an example of theoretical studies there are people who are very highly mathematically oriented in their studies who basically try to construct models of what is happening in the course of evolution and what is happening in the course of geographic distribution of species where they derive equations that attempt to predict the observations that one would expect to make in real populations. Then there is another group of people amongst which I would tend to classify myself that are more interested not so much in the mathematical models per se but are interested in corroborating or testing those models with some real data from actual populations. Some of these studies involved -- in my own case I work in a particular insect called drosophila. Some of these studies involve

trying to understand what happens during the process of producing different species. The process of speciation.

- Q. Drosophila. What importance does drosophila have to the population geneticists?
- A. It is the prime organism that all the population genetics has been based on since about 1905. It is an organism that is very easy to work with experimentally, that has a very interesting natural history, that is very easy to approach in terms of understanding the genetic differences that are present in different individuals and in different populations. I would say that population genetics and experimental population genetics most of the original work has been done using this organism at one time or another.
- Q. Do the different fields of population genetics, do they have common theoretical basis? Common working principles?
- A. Yes. Very much. The theory applies in a very broad way to all life forms.
- Q. Would that include humans?
- A. That includes humans.
- Q. Are there people actually working or would you consider to be in the specialty of human population genetics?
- A. Yes, there are. Amongst whom I would classify Dr. Kidd, for example, who is not able to get here today. There are some unique problems or questions that come up in studying human population genetics. That is not the case in studying drosophila genetics or any other organism.
- Q. Does statistics form part of population genetics generally?

- A. Yes. Very much. As I mentioned earlier, the use of mathematical models means that in order to test those models and the predictions of those models, you need to collect some data and then test the numbers you derive statistically from that data against the models and the predictions that the models make, so statistics is an essential component.
- Q. The theory of population genetics generally and the principles in theory and working formulas that you use, can you use those and apply those to human populations?
- A. Yes, you can.
- Q. Have you actually done work applying -- in the area you actually deal with, have you attempted to look at these models and apply them to human populations?
- A. Yes, I have. In the last eight months or so, since last August, I have been working with the population databases that the R.C.M.P. in Ottawa have been amassing over the last couple of years.
- Q. Doctor, you -- perhaps if we could go on. You have been -- you are presently an Associate Professor of biology at Carleton University. You were an Associate Dean of science at Carleton University and I understand also, Doctor, you are chairman of Integrated Science Studies at Carleton University.
- A. Yes.
- Q. Tell the court what kind of teaching duties you would have associated as Associate Professor of biology. What kind of things would you be teaching?
- A. Well, I teach an introductory course in genetics. I teach an introductory course in molecular genetics.

Some years I teach a fourth year population genetics course and on alternate years I teach a graduate course for graduate students in evolutionary genetics jointly with a colleague at the University of Ottawa.

- Q. Evolutionary genetics, would that have application to population genetics?
- A. Yes, it does. I would say that population genetics is actually the groundwork and the basis for trying to understand changes that occur during the processes of evolution.
- Q. And as chairman of Integrated Science Studies at Carleton what duties do you have there?
- A. We have about forty students involved in rather interesting imaginative programs. I am very proud of this program, I must say, to the court. It is unique in Canada in that it allows students to combine a study of an area in science with an area not in science. So we have students, for example, who are combining studies in anthropology and archeology with biology. We have students combining an area of science with an area of business or law and the combinations are almost as great as you could imagine. It is a very interesting program.
- Q. Professional memberships, Doctor, in relation -- as a population geneticist, the Genetic Society of America and the Genetic Society of Canada and the Society for the Study of Evolution, would they have application to population genetics?
- A. Yes. All those societies have population geneticists like myself as members although there are also other people in those societies who I would not call population geneticists too.

Q. You have listed in your CV research interests and you have indicated among them molecular evolution of DNA sequences and genetics of population differentiation as speciation.

A. Yes.

Q. They have application to population genetics I take it.

A. All of those interests stem out of my fundamental interest in the studies of the differences between different individuals, different populations, different races and different species.

Q. That is an area of research to your interest?

A. That's correct.

Q. Doctor, what is your relationship with the R.C.M.P.? Are you a member of the R.C.M.P.?

A. I am not a member of the R.C.M.P.. I strictly do this on a consulting basis. I am not paid by the R.C.M.P. I am interested in this as a fundamental science.

Q. Doctor, if we could sum up what would your field of science be?

A. I would classify myself as/^apopulation geneticist. If I wanted to make a further subdivision I would say I was an experimental population geneticist.

MR. WALSH: My Lord, at this time I am going to ask that Dr. Carmody be declared an expert in the field of population genetics.

THE COURT: Do you have any objections?

MR. FURLOTTE: I have no questions. No objections.

THE COURT: I would -- I am satisfied -- Just one thing. You are a doctor of -- let me see. You have a Ph.D. in zoology. I always wondered what zoology was. I conjure up visions of giraffes and all sorts of animals. What is it? Because it starts with 'zoo'.

A. Yes. It means the study of animals as opposed to the study of plants. It really is an administrative division in many universities where you divide up biology. One of the easiest ways to divide it up is to divide it up into botany and zoology. Many universities now are reversing that process and combining these into departments of biology or biological sciences. At the time that I was a student one had to be either a Ph.D. candidate either in botany or in zoology. I was more interested in the animal population genetics and the professor I chose to work under was in the zoology department so even though I took my degree in population genetics I was officially in the department of zoology. But you had people in that department who were working at the molecular level or that in fact were doing the kinds of studies that that name conjures up, of studying bird behaviour or going to Africa and studying the behaviour of elephants and whatever. It also comes under that umbrella so that is where the name pertains.

Q. Does the human species come within the umbrella of zoology?

A. Yes, but typically in universities that tends to be relegated to medical studies and you tend to find pathology and anatomy and so forth in medical school, and people interested strictly in the human end of it would tend to be there.

THE COURT: You made a request to the court.

MR. WALSH: Yes. Sorry.

THE COURT: I do declare the witness for the purpose of this trial and expert in the field of population genetics.

MR. WALSH: Thank you, My Lord.

THE COURT: That doesn't make you an expert. You can't charge more, you know, for your consultation. Just that we are going to allow you to give opinion evidence.

A. Thank you.

Q. Dr. Carmody, are you familiar with the databases that presently the R.C.M.P. have?

A. Yes, I would say I am quite familiar with them.

Q. And that they presently use for DNA typing?

A. Yes.

Q. Do you know the various types of databases -- do they head databases by race?

A. Yes. They basically have at the present time two main databases. One that is made up of various samples from Caucasian populations in Canada, and another that is made up of two separate samples from native aboriginal populations.

Q. Which would be the larger database?

A. The largest database would be the Caucasian database.

Q. Could you describe the composition of that database? First of all, is the composition of that database -- do you have that summarized anywhere?

A. Yes, I do. I provided to you a two-sheet document that has a summary of how those two samples or three samples were obtained.

Q. Of the Caucasian database.

A. The Caucasian database.

Q. Is this the document here, Doctor?

A. Yes, it is.

MR. WALSH: My Lord, with permission I would have this entered.

THE COURT: VD-58.

Q. I am handing you a document entered through His Lordship. I would give you a copy of that and I would ask you if you would please just summarize, if you can, the Caucasian database and what it comprises.

A. Well, there were basically three components to it and three samples taken. The first one that I will mention is a sample of 356 from the Vancouver area that was provided by Dr. Lorne Kirby at the University of British Columbia, and that was collected from the Grace Maternity Hospital there during the five month period of January to May 1989, and that those samples were sent to the forensic labs in Ottawa and have been subsequently analyzed. There are two samples from Ontario. There is a smaller sample that is continuing to grow from the Ottawa area that ^{presently} consists of 97 samples that were obtained from the Ottawa branch of the Canadian Red Cross. These samples were, as summarized in the submission, obtained during the period from March to July 1988. These are samples that are randomly taken from people who have donated blood to the Canadian Red Cross and there was every effort made to make sure that no two people had donated twice and that there were no identical twins in the sample.

The third sample and the largest of them was obtained at a blood donor clinic again run by the Canadian Red Cross on the Canadian Forces Base in Kingston, Ontario. That sample was obtained during the two month period of August to September 1988 and consists of 526 samples.

- Q. Again, Doctor, what was attempted to be obtained by those particular databases and in the fashion they were --
- A. Well, a very important consideration in designing these studies is to try and get as representative a sample of the Canadian Caucasian population as one could contrive.
- Q. Obtaining databases in that fashion would that be a reasonably reliable way of going about that?
- A. These are a reasonably reliable way. One can imagine still theoretically better ways and one would still want to have samples from other areas of Canada to completely flush out these three samples, but these three samples as we have analyzed so far are an excellent cross section and representation of Canada.
- Q. The areas that are being studying with forensic RFLP typing, does that come into the consideration of how representative your database must be?
- A. Yes, it does, because if one is studying something that shows a great amount of local differences, then taking a sample becomes more tricky because you have to be sure to get represented in your sample every single possible area where there might be some difference. For the parts of the human genome that have been looked at forensically in the studies I have been involved with in these databases there seems to be very little difference geographically throughout Canada.
- Q. If you could give the court some indication of how representative these databases are of the Canadian population, the Maritimes and New Brunswick..
- A. Yes. I have some slides to show to exhibit how in fact

these samples particularly -- and in fact I guess I am going to talk almost exclusively about the Canadian Forces sample from the Kingston area and how representative they are particularly of the Maritime area of Canada.

Q. Before you start, Doctor, I am going to -- Doctor, that schematic you have on the stand, I have a paper representation I believe of that schematic. Would you look at it and tell me if that represents the schematic on the screen?

A. Yes, it is a xerox of the slide that was made.

MR. WALSH: The only difference, My Lord, would be the color. This is black and white and unfortunately we don't have one in color yet. It is coming down but I don't think we have received it yet.

THE COURT: Want to mark that?

MR. WALSH: Please, My Lord.

THE COURT: VD-59.

Q. Doctor, would you tell us please what you are attempting to depict there?

A. I am attempting to show in this slide, in actual census numbers, Canada in terms of representation of different provinces by number of population and I think what I want to show particularly in this slide is that -- this pie chart -- that in fact the Maritime provinces -- the first here in green being New Brunswick -- that total part of the population constitutes somewhat less than 10% of the total Canadian population. I have an equivalent chart to this where I have expressed these numbers now in percentages.

Q. These numbers that are on -- do they represent individuals?

A. They represent individuals censused in 1988 I guess it was in Canada.

Q. '86 I believe the screen says, Doctor.

A. '86. I can't see it. If it says '86 it is '86. This was taken from Statistics Canada -- a Statistics Canada publication.

Q. I will show you another -- do you wish to go to the next slide?

A. Yes.

Q. I show you this paper representation. Does this accurately depict what is on the screen?

A. Yes, it does.

Q. Again, My Lord, it does except for color. I do not have a color reproduction.

MR. WALSH: I would ask to have that entered.

THE COURT: This slide would be VD-60.

Q. Continue, Doctor, please.

A. This slide is essentially the same pie chart as the first slide but here we have the numbers depicting the percentage --that is the fraction out of 100 -- that is present in each of the provinces and the point here is that New Brunswick constitutes 2.8% as of the 1986 census and the Maritimes overall if you would add those up comes out to be a bit less than 10%.

Q. Of the total Canadian population.

A. Of the total Canadian population.

THE COURT: I think earlier you said New Brunswick was about 10% or less. You meant the Maritimes.

A. I meant the Maritimes. I meant the total Maritimes -- New Brunswick in fact is as this indicates. 2.8%.

Q. Do you have another, Doctor?

A. Yes, I do.

Q. I will show you this document here. Is this an accurate depiction of what is on the screen now?

A. Yes, it is. Yes.

MR. WALSH: I would ask to have this particular document entered.

THE COURT: VD-61.

A. This table shows the percentage composition by birth place at the Canadian Forces Base in Kingston of both military personnel and dependants. This was provided to us by the base at Kingston and represents the profile by birth place of people who were contributing to the data bank that we have gotten from samples from there.

Q. What does that indicate to you, Doctor?

A. Well, these numbers are percentages and I think you can see there -- let's just point out New Brunswick first right here -- that in fact there were 5% of the personnel there who were from -- by birth place -- from New Brunswick. And if you recall on the previous slide New Brunswick constitutes 2.8% or somewhat less than 3% of the Canadian population. Just running through the Maritime Provinces there in general you will see -- if you add up Newfoundland and Prince Edward Island, Nova Scotia and New Brunswick, that you come up with a number that is over 20% so this sample that we have of the Canadian Forces represents -- in fact, if anything, over represents -- the Maritime Provinces and has approximately twice as many people from New Brunswick as per Canadian population.

THE COURT: It also means that unemployment is double here.

A. People joining the Forces because of that. Perhaps.

It is interesting that -- I found it interesting that

in fact Newfoundland contributes 7% when in fact in terms of population it is much smaller.

Q. When you say over represented, is that -- for/ ^{the} purposes that we are working with today in forensics, is that a good indicator or bad indicator for the work you are trying to do?

A. That is a good indicator. If, for example, we were using this to ascertain what the frequencies were in British Columbia for example, it would not be a good sample to use. However, I can -- just to anticipate a bit here -- one finds that if you look at the genetic profile in this sample and compare it to the profile of Vancouver they are statistically indistinguishable.

Q. What does that indicate to you in terms of how representative your database has to be?

A. That indicates to me that in fact there is almost no or at least no statistically detectable genetic differences in Caucasians from various parts of Canada. I have by the way I think another slide here --

Q. Yes.

A. -- that indicates this by percentage again in a pie chart and I think graphically here you can see that the Maritime Provinces, if you recall in terms of the census data, constituted an area roughly this size. Now, in terms of our sample constitute almost a quarter of the sample from the Canadian Forces Base.

Q. I have a copy of what is on the screen. Would you look at that please? Is that an accurate depiction of what is on the screen?

A. Yes, it is.

MR. WALSH: Again, My Lord, it is not in color. I would move to have it entered.

THE COURT: VD-62.

Q. Continue, Doctor, please.

A. Well, the point of this slide is just to show graphically that in fact the sample is quite enriched, if you will, in a component drawing from the Maritime Provinces. I can go on to another slide where in fact I make this comparison directly where in this histogram the population of Canada is indicated by the black vertical bars and the population at the Canadian Forces Base in Kingston by the red vertical bars. You will see the Maritimes down here on the left all have red bars that are higher -- in some cases considerably higher. Almost double -- compared to the Canadian population. Indeed in New Brunswick here you will see that we have more than an adequate sample it seems to me.

MR. WALSH: I have a copy of what appears to be on the screen. Is that an accurate reflection of what is on the screen?

A. Yes, it is.

MR. WALSH: Have this entered, My Lord.

THE COURT: VD-63.

MR. WALSH: Again unfortunately it is not in color.

Q. Doctor, the type of data that we have been looking at or you have been showing the court, is this the kind of data that a population geneticist would refer to to look at -- to see how representative the sample population is?

A. Yes, it certainly is in any sampling program, and when you are trying to study animal or plant populations it is almost always the case that you have to rely on samples taken from those actual populations. A major

question is whether the sample you have represents accurately and precisely what was really out there that you are trying to study, and the purpose of doing this kind of analysis, of comparing your samples and the origin of your samples to the actual population you are trying to make statistical inferences about, is exactly why we did this.

- Q. Doctor, what, if anything, was done to combine the separate databases, Vancouver, Ottawa and CFB Kingston? Before I ask you that. Can you tell me -- going out in a blood donor -- obtaining samples from the Red Cross in the fashion that you have described, is that an accepted manner of actually obtaining samples for purposes like these?
- A. Yes, it is. One would imagine that the people that were donating blood were samples from the population irrespective of the particular genetic types that they were for the loci that these samples were going to be used to study.
- Q. Again, the question I wanted to ask was: With respect to Vancouver, Ottawa and CFB Kingston, what, if anything, have you done in terms of determining whether you can in fact amalgamate these and whether in fact they were amalgamated?
- A. I looked at the genetic profile for each of these three populations and did statistical tests and was able to show that there were no significant statistical differences between these three samples, so it was a perfectly legitimate process to amalgamate those into one combined sample that we call 'Canadian Total Caucasian Population'.

- Q. These statistical tests, are they tests that are used in population genetics generally?
- A. Yes, they are. It is a test called the chi-square test and I did another test that is also known as a likelihood ratio test. Both of these are standard tests that I used to compare distributions of genotypes in populations.
- Q. And a genotype, for the record, is what?
- A. Is the genetic blueprint at the particular locus that you are interested in studying of an individual.
- Q. For the record, chi-square would be spelled c-h-i?
- A. Right. It comes from a Greek letter.
- Q. You did these statistical tests and you say that there is or isn't statistical difference?
- A. There is no statistical difference in these three samples, one combined to the other. I have done all the pair-wise comparisons. I have looked at them all three at a time and four. Any of the five loci that I have studied in these populations that are part of the R.C.M.P. database, there are no statistically significant differences in the genetic profiles of these three samples.
- Q. What conclusions can you draw from that, Doctor?
- A. I can draw the conclusion with great assurance that as we get greater samples and as we increase sample sizes from other areas of Canada, that it is very unlikely that they will show differences from the existing samples that we have. Nevertheless it is important to continue to get further samples because we can't say with 100% certainty that there won't be some differences.

Q. These differences, would they have an effect forensically?

A. It turns out that if you calculate through the effect of some of the differences that I have found by comparing these samples to some Caucasian samples from the United States and from France that there is no significant forensic implication to the differences that I found between some Canadian samples and some American samples.

Q. Doctor, what would -- what is the size now of the Canadian Caucasian database for the R.C.M.P.?

A. Well, it varies slightly from one genetic locus to another genetic locus but it is on the order of 1,500.

Q. Would you tell the court please your opinion as to the adequacy of that size for the purposes here?

A. This is a very good sized sample and a measure of the adequacy of a statistically analyzed sample is how much you would expect that sample estimate to vary if you were to take another sample of the same size, and you will see that -- I will show some data later -- that in fact if you were to take further samples of the same size, the amount of difference you would expect to see is quite, quite small.

Q. Doctor, if you would now -- you have touched on the database. If you would I would like to get into the frequency calculations and how you go about frequency calculations for the purposes of forensics. Could you tell me please what the is the method for generating frequency statistics from this database for use in forensic DNA RFLP typing?

A. Right. Basically for each of the samples they have the

DNA extracted and the DNA is run by the procedure that I understand Dr. Wayne described in earlier testimony, but basically each individual sample is run and where the two bands appears for each of the five probes that are looked at, estimates are made of the molecular weight size of each of the two bands, and that molecular weight size then is an indicator of the genetic profile or the genotype at that locus for the individual that contributed that sample. That is done for each of the five different loci that are looked at.

Q. Each of the five different probes.

A. Five different probes. Then after you have done that for all the specimens in your sample that data is put together locus by locus. Each probes data is kept separately like that, and you have a whole bunch of numbers. Those numbers can be used to create a histogram or a spectrum of distribution of the size fragments in that population.

Q. That were seen by each probe.

A. By each probe. So you have then a genetic profile of that sample. It is basically a distribution that indicates what the size variation is in that sample.

Q. All these bands that the probe has seen in the sample population are the alleles or the fragment sizes, what do you do with that to prepare your histogram? Where do you put these numbers that you have obtained?

A. These numbers are put into categories so that you can draw an equivalent histogram spectrum of the distribution like that and so you group together sizes of very close size and you call them one bin. The terminology is you put them into a bin. It is basically

just a category. It is an area between two size standards. Anything falling between those two size standards are called to be in that bin. So that if you do this across this spectrum you create a number of bins then, each of which has a number of bands that you have looked at in there and you create then a frequency of that particular bin, that particular slice, through this profile.

- Q. And if I understand you correctly, Doctor, there is a bin or a series of bins kept for each probe?
- A. That's correct.
- Q. The number of bins you would have for each probe, is that the same or does it vary?
- A. It varies from probe to probe because different probes have slightly different spectrum of profile of molecular weight sizes and some probes have a very narrow spectrum of sizes that are lit up by that probe. Other have a very broad spectrum so for some loci we have as few as perhaps eight bins. For other loci, for other probes, we have as many as 30 bins.
- Q. Doctor, if I understand you correctly, the more bands in a particular bin the more frequent the frequency is?
- A. That's correct.
- Q. What happens if you happen to have a few bands in a bin? I take it that would indicate a very rare phenomenon.
- A. That is a very rare phenomenon.
- Q. What do the R.C.M.P. do when they have something like that?
- A. When they have something like that and there are fewer than five in a sample in any particular bin, that bin

is amalgamated or consolidated with the adjacent bin to give a number that is always greater than five. One of the reasons for doing that is that if you have a number less than five in a bin, or as you can appreciate I think, the smaller the number in a bin the more unreliable that number is. If you took a sample of 1,500 and you found only one specimen or one band in your data in that bin, it would be very unreliable in the sense that if you took another sample you might anticipate there could have been five in that bin or there could have been zero in that bin. The number there will be very unstable so by consolidating that with an adjacent bin you are able to bring the number up so that it has a greater precision if you were to resample. You can have greater confidence in the value where you have a greater number.

- Q. The frequency of the bins that have less than five, if they are collapsed into another bin, what does that do with respect to the frequencies?
- A. It increases the frequencies and particularly in the usage of this data one wants to always try to be as conservative as possible. One wants to never underestimate the size of a bin so this would tend to increase the size of that category.
- Q. Is there anything else -- in actually binning these frequencies, is there anything else that is done by the R.C.M.P. to follow this conservative approach?
- A. Well, the other main thing that is done by the R.C.M.P. to be conservative in any inferences using this histogram of bins, is that if you have a specimen that falls near the bin boundary you look at the adjacent bin

and you take the higher of the two frequencies, either in the bin that the band actually formally falls in or if the adjacent bin is higher in frequency you take the higher of those two frequencies. Even though the band properly and nominally does not really fall in that bin, if it is higher and it is adjacent you take the higher of those two as the frequencies.

Q. What about the width of the bins, Doctor? How are the bin boundaries determined and what can you tell me about the width of these bins?

A. These bin boundaries are always wider than the precision of the estimates that one can derive by this technique. That is, I think -- maybe it is not appreciated, but the techniques that are used in this DNA analysis, though they give you an estimate for the size of the band that will come out of the computer that will be down to the base pair -- that it will be 1,749 -- that number has an imprecision in it and it is deceptive in the sense that we really don't know it is exactly 1,749. It could as well be 1,735 to 1,765. That is that there is an imprecision in that measurement. That means that if you were to look at a particular band, and though you estimate it to be of a certain size precisely to the base pair, if you were to run that again you might get a slight difference in the estimate. So the size of these bins is always larger than that window of imprecision in the estimates that are derived from the actual gels.

Q. And for the R.C.M.P. purposes could you give the court some indication of how the bin sizes range?

A. The bin sizes range from close to 6% of the molecular weight to over 15% of the molecular weight.

- Q. And the R.C.M.P. match window is what?
- A. Is 5.2%.
- Q. So the lowest is --
- A. -- is 5.7 and the highest is 15. Most of them, if you were to look at them, are on the order of 10%. Most of the bin widths are approximately twice the size of the window of imprecision.
- Q. And what effect does that have on this attempt to be conservative?
- A. It is again a conservative procedure in that it would be statistically correct to use bins that were exactly as wide as your window of imprecision and because you are using bins that are in fact wider than that they are necessarily going to have a greater frequency than a narrower bin. You are being conservative by taking a wider bin.
- Q. What is the purpose of this binning again, Doctor? What are you attempting to extract from the binning process? When you have finished your binning what do you want to know?
- A. One wants to know the frequency of that genetic variant in a population. What is the frequency of a band of a certain size in the population.
- Q. The frequency of one individual band.
- A. One individual band.
- Q. I am showing you a document that has been entered on this voir dire as number 49. Would you look at it please and tell me if you can identify that document?
- A. Yes. This is a copy of -- it is a preprint of a publication that has appeared in the American Journal of Human Genetics written by -- collaborated on by people at the R.C.M.P. and by people in the F.B.I.

- Q. What is this? The title is "The Fixed Bin Analysis for Statistical Evaluation of Continuous Distributions of Allelic Data from VNTR Loci for Use in Forensic Comparisons". What is contained in that document?
- A. That document describes the procedure of doing this fixed binning on DNA measurements and shows the statistical validity of this process. It is a peer reviewed paper that has appeared in a publication I think two months ago.
- Q. Okay. Doctor, at this point in time, if I understand you correctly, with the binning you have a way then of looking at individual allele or individual band frequencies. Now, what would you do next to determine, for example, a band pattern or -- excuse me -- a pattern of two bands, for example, or a one band pattern? What would you actually do to -- where would you go from there?
- A. Well, if I were to try and predict the frequency of a certain genotype, that is of an individual, for each of these loci, there would be two copies -- there would be two bands present. In some individuals if the two bands were identical they would appear on the gel to overlap one another and you would see one band. That kind of a genotype is called a homozygote. Most people -- and in fact homozygotes constitute typically for these probes something on the order of 10% or less than 10% of the population. Most individuals have a pattern where for each of these probes you get two bands.
- Q. Like shown on this diagram, VD-45.

- A. That's correct. These would indicate particular double-banded patterns which is the overwhelming majority of people.
- Q. Okay. Doctor, if I wanted to find out, for example, after finding out the individual allele frequency, the individual band frequency, by binning, if I wanted to find out what is the frequency of two bands appearing in the population what would I do?
- A. I could use those separate estimates and in fact the procedure is to take the estimate of the frequency of this band and to multiply that by the frequency -- estimate of the frequency of this band and to calculate the frequency of that particular double-banded pattern.
- Q. What is the scientific or mathematical expression of that particular calculation?
- A. That particular calculation and the expression that is used is to say that -- when you use that procedure you are using the Hardy-Weinberg Equilibrium law.
- Q. I have heard the term Hardy-Weinberg Equation. Is that synonymous?
- A. Hardy-Weinberg Equation, yes.
- Q. If I had a one-band pattern how would the equation apply?
- A. A one-band pattern would be -- let's say if this were a single band in this lane you would take the frequency of that band in your bin, the bin that it falls in, and multiply it by itself because there would be two copies of that present. You are not able to see both copies singly because in fact they would overlap.
- Q. So if I understand -- and correct me if I am wrong, Doctor -- I hope I am not taking too many liberties, My Lord, but just to clarify -- using the Hardy-Weinberg

Equation, if I understand you that would mean that by that equation, if I look at the frequency of one band that you have obtained by binning, and look at the frequency of a second band that you have obtained by binning, you can project what the frequency of having both bands together in this location?

- A. That's correct. I can predict the expected frequency of that particular pattern, that particular genotype at that probe position, in the population from the sample that I have.
- Q. This equation, can you tell us something about this equation? Is it something that has just been developed for forensic purposes or how --
- A. No. It was an equation that in fact has a long historical basis to it. It was actually first proposed in roughly 1904 by two people. A mathematician in England, Sir Jeffrey Hardy, and by a German physician, Eli Weinberg, They propose it is really an algebraic conclusion that one comes to if mating and if genes are associating in populations at random. Because you can then say if these processes are occurring at random then in fact the probability of these two coming together is simply the product of the two separate probabilities.
- Q. Doctor, now that -- if I was trying to do this and I now know the frequency -- I know the individual frequency by using Hardy-Weinberg. I know the frequency of these two bands together. For example, Doctor, hypothetically if I were to do another probe and find these two bands in another location or two bands in another location and I wanted to know -- to calculate the frequency of these new two bands I would use the same process you described.

- A. Use the same process. You would do that for each probe. Each of the four or five, whatever number of probes you have, and you can calculate a separate frequency for each of the genotypes at each of the five probes.
- Q So if I wanted to -- using the Hardy-Weinberg Equation I can determine each probe frequency and depending on the number of probes if I wanted to get the total genotype frequency I would multiply them across.
- A. You then multiply them across the probes. That's right.
- Q. That is called what in mathematics?
- A. Well, it is an extension of Hardy-Weinberg Equilibrium to more than one locus but we talk about the process of -- we are making an assumption of linkage equilibrium.
- Q. Okay. But the equation of actually multiplying one probe by one probe by one probe is what?
- A. It is called the product rule of being able to take two separate occurrences and taking their probabilities and multiplying them one times the other.
- Q. Again, Doctor, the product rule, is that something that has simply been developed for forensics or what kind of application does that have in science and mathematics?
- A. That is one of the fundamental rules of probability theory and that probability theory and the development of the mathematics of that goes back a couple of centuries actually. It is a fundamental axiom, if you will, of probabilities of events that are independent of one another. You multiply the probabilities.
- Q. Doctor, now that we have made such a calculation, is that what the number that is generated in a forensic case -- the conclusion that is generated? Is that how you go about it?

A. Yes, it is.

Q. So the number that you would actually get recorded in a forensic case, one in so many, is the end result of the individual binning, the Hardy-Weinberg Equation and then the product rule.

A. That's right.

Q. Doctor, what do population geneticists want to know about the database that is used in order to assess the reliability of the frequencies as you have described how they are calculated? What do you want to know?

A. Well, first of all, as I have testified earlier, it is very important that in fact the sample that these calculations are based upon represents the actual population that you are making your inferences about so that you want what is called a random sample -- certainly a representative sample -- of the population you are studying.

Q. Now, just to clarify, Doctor, does that mean if a crime was committed in Burton, New Brunswick, that we actually have to have a sample population from Burton, New Brunswick?

A. Well, it potentially could except that our studies on the Caucasian database drawn from these three samples in Canada would indicate that in fact there is no local geographic genetic differentiation that is present in our Caucasian population or at least none that is statistically significant enough to be seen in our samples, and that would mean that the calculations that I did using the data in the R.C.M.P. database would hold whether we were making the inference about British Columbia, Ontario or the Maritimes.

- Q. Continue, Doctor. What would you want to know about the database?
- A. In addition to it being, first of all, representative ^{actual} of the/population, one would want to know that in fact the appearance of these various bands in that database for each of the probes in fact were occurring at random with respect to one another. That is one would want to be assured that wherever the individuals that carried let's say this larger band here -- and the larger ones being towards the top of the gel -- carrying this one, did not have any strong correlation with the occurrence of a particular other band down here or other particular size fragments -- that is, that the appearance of a particular size band was occurring in individuals independent of the size of the second band.
- Q. Randomly associating.
- A. Randomly associating. So one needs to look at that original database to convince yourself that indeed there is no non-randomness present in the association of the various bands. If there were then you could not properly use the Hardy-Weinberg Equation.
- Q. Have you looked at this question?
- A. Yes, I have. I have done this using a test that is a non-parametric test that is a test that could pick up strong correlations between size fragments at each of these loci. These tests show that there is no evidence of a strong correlation being present, and the conclusion is that the spectrum and distribution of each of the band sizes is independent of one another.
- Q. This test, is this a test that you have developed simply for forensics or is it a test that is used for these kinds of purposes generally in science?

- A. It is a test that is used for these types of purposes in general. It is not one that is specifically used here. I have to also say, however, that because of the very, very large number of possible combinations -- say if you have 20 bins, for example, there are over 200 different combinations of one band with any of the others. That means that one really has to look at over 200 different categories to see whether all of those categories are occurring in their statistically expected frequency. That means that in a database of a size of even 1,500, if you are putting those all into 200 categories, some of those categories are likely not to be represented in your sample, and so to really do a highly refined statistical test one needs a massively sized sample. Unfortunately we, at the present time, don't have that or in fact in the foreseeable future we will not have that, because when I say massively sized I mean 100,000 or a million individuals to be able to really test if there is complete randomness present.
- Q. What if there was strong correlation?
- A. Yes.
- Q. I take it by -- when you say 'strong correlation' there may be -- a test can't pick up slight correlation. If I am using the correct term. How would that affect what you are doing? You have indicated that you need an immense database to actually determine I take it slight correlations.
- A. Right.
- Q. But since you can't test for slight correlation what effect does that have on what we are doing here in forensics?

- A. Very slight correlations are not going to have any substantive impact on any of the numbers that we generate. If there were strong correlations, however, what that would mean is that for example this particular pattern would, if there were strong correlations between this sized band and that sized band, it would mean that those two were tending to occur together and that using the Hardy-Weinberg Equation of multiplying the probabilities of each of those separately, would not be a correct estimate, but if there was a high correlation they would occur more frequently in the actual population than you would have predicted from your separate calculations.
- Q. You tested for that.
- A. I have tested for that and there are no strong correlations at any of the loci when you look at one band compared to another band.
- Q. Assume for a moment, Doctor, there is correlation in the sense it is non-random association between each individual band you see, but that non-random correlation is less than what this test will pick up, what effect would that have on the figures that are being generated by the R.C.M.P. for forensic purposes or more importantly the use of the Hardy-Weinberg Equation?
- A. It would have a very slight effect. We are talking about effects in the third or fourth decimal place in the estimates that we are ascertaining from these bin frequencies.
- Q. Would ^{it} affect the reliability of the figures that are being generated?
- A. It would have no substantive effect on the figures that were generated.

- Q. What are the other questions that you have to determine, Doctor?
- A. Well, there is a question -- and it comes up in this technique and has been used to criticize this technique, because when you analyze the data you find that in these databases there are more single band patterns than you would predict from the Hardy-Weinberg Equation. That is, that there is an over representation in your sample of individuals it seems that have just a single band. That can occur for at least three different reasons, two of which are strictly a limitation of the technique. For example, to talk about the limitations of a technique and why we think there could be over representation of single band patterns because of the technique. If two bands, though they are different, are below the resolution limit of the technique they would not be seen as two bands so there certainly is a difference that one cannot resolve using this technique. It is below the window of resolution.
- Q. This test for excess homozygosity, this was I understand originally tried when they were looking at this forensic application?
- A. That's right.
- Q. And what are the conclusions that scientists have drawn about the use of that test for trying to determine whether you can use the Hardy-Weinberg Equation?
- A. The conclusions that have been drawn is that because of the limitations in the technique that that is not a correct test to use because you are getting artificial so-called homozygotes that are not real homozygotes.
- Q. I show you a document that has been marked on this hearing 53. Would you look at that for me please and if

you could just read the title for the court and what application that has to what you have just talked about?

- A. Yes. The title is "No Excess of Homozygosity at Loci used for DNA Fingerprinting". It is a publication in the Journal of Science and it appeared last September.
- Q. Is that the Devlin and Risch article?
- A. The authors are Devlin, Risch and Roeder, yes.
- Q. What is the bottom line for this particular article?
- A. That article develops an alternative statistical technique that would allow you to analyze the data at each of these five probes. That does not use the criteria of per cent homozygosity versus per cent heterozygosity and is immune to the artificial effects of the limitations in this technique in scoring real homozygotes. The bottom line is that in their tests using this refined technique they have been able to demonstrate that there is no true actual biological excess homozygosity in the samples that they have analyzed.
- Q. What other things do you want to know, Doctor, in order to determine whether these frequency calculations using binning and Hardy-Weinberg and the product rule, what do you want to know about your database?
- A. The next thing one wants to know about the database pertains to the next step in this procedure of calculations, namely, the product rule. Multiplying the probabilities that you obtain from each locus which you are now confident are good reliable estimates because they are fitting Hardy-Weinberg Equilibrium. You now want to be sure that there are no correlations between what genotype occurs for probe one and probe two, probe

two and probe three and so on through all the combinations

- Q. You want to ensure the difference between each probe, there is no non-random association from probe to probe.
- A. That's right. In the same way that one wanted to be sure there was no correlation between the pattern of a band -- the two bands at each locus, one wants to be able to be sure that there is no correlation of the band pattern at one locus and the band pattern at another locus.
- Q. Okay. What do scientists or population geneticists do in that regard?
- A. Again it is a test to show that the frequency of appearance of a particular genotype at one locus is independent and uncorrelated with the frequency of a band pattern or genotype at a second locus, and one does statistical tests where what you do is to look at all the genotype frequencies at locus one and compare those to see how frequently each of those occurs with all of the potential genotypes at locus two or probe two to see if there is any deviation from what you would expect. If there is any correlation of one particular genotype at locus one with those at locus two.
- Q. Okay. Tell us please from a scientific point of view what the name of the test you would use and what are you actually in science testing for?
- A. To do that test I used a non-parametric test. That is, a variant of what I used looking at each individual locus. It is a non-parametric median test that I used. It has the ability to pick up strong correlations. It again has a limitation and the limitations are even greater in the case of comparing correlations between

probes because the number of categories that you expect is still higher than the number for each individual probe itself. That is, if for each of these probes there were 200 potential genotypes then the number of categories that you are looking at are 200 times 200 or roughly 4,000 different categories, and in a data set sample where you have 1,500 individuals it necessarily is the case that all of those categories are not going to be represented. So the test that you have to use is not able to pick up in a refined way all of the slight correlations that might exist but it would pick up any correlations that were of statistical significance in the use of this data for forensic purposes.

Q. That is what I wanted to ask you, Doctor. Using that test are you able to satisfy yourself with respect to the association from probe to probe for the purposes that we are dealing with here?

A. I have been able to satisfy myself that there are no strong correlations of the genotype frequencies from one probe to another probe.

Q. If there was, what would that term be called?

A. That would be called linkage disequilibrium or gametic phase disequilibrium.

Q. And if in fact there is no association -- random association, I take it that would be linkage equilibrium.

A. Yes.

Q. What, if anything, else would you want to know about the database you are working with and/or what, if any,

other tests would you do to ensure yourself with respect to the reliability of the frequency figures given?

- A. Well, one could do calculations on subpopulations. That is, we have the three populations I have mentioned doing a statistical test and we showed there was no difference in the bin frequencies. One could in fact do the calculations in a particular forensic case on each of the populations separately to just assure yourself that in fact it didn't make any difference which population you used. You could compare that to the estimates that you derived from the population -- from the composite population when you had put all those together. Furthermore you could in fact do these same calculations using other databases. I mentioned earlier that we have two native aboriginal samples. One from the Winnipeg area and one from the west coast of Canada, the coast of British Columbia, and one could do the calculations using those genotype frequencies and those bin frequencies. One could compare it to other frequencies from other populations and samples of Caucasians from other parts of the world and from other parts of North America.
- Q. Would you tell us please what are you testing for when you are doing this? When you are looking at you say sub-groups?
- A. Right. I am -- basically when I am doing that I am looking at the robustness of the numbers of the estimates that we come up with.
- Q. Is robustness -- does it have a particular meaning in the field you work in?

A. It does. I am using it in a statistical sense and it means that the number would not significantly change if you based that calculation on other information; on taking data from other areas, taking data from other populations, taking data from other geographic regions. If that number does not change substantially then you say and you have confidence that the estimate that you derived on the sample that you had is a reliable indication of what it would be under all further samples.

Q. Doctor, have you looked at Caucasian databases in other places?

A. Yes, I have. I have had access to the Caucasian databases that the F.B.I. in the United States have been putting together. I have looked at databases derived from separate geographic regions in the United States, one from Dade County, Florida, one from the State of Minnesota, one from the Fort Worth, Texas region, and I have some preliminary data on some Caucasian samples from France.

Q. And what were the conclusions that you drew with respect to the robustness of the frequency calculations that are used by the R.C.M.P. or generated by the R.C.M.P.?

A. My conclusions were that the R.C.M.P. data and my analysis of it was a true reflection of the occurrence of these variants in virtually any Caucasian population in North America. There were some slight differences for France for the two probes that I looked at there. It is difficult to say what the net effect of the differences between France and North

America would be in terms of doing all the calculations because I don't have data on all five probes from France.

Q. What about referring to United States? What conclusions did you draw there?

A. The conclusions that I could draw there were that it didn't significantly make any difference as to which Caucasian database you used from the United States or whether you used the Canadian database that we had. The net forensic implications were statistically trivially different.

Q. Here is where I want to hone in here, Doctor. You indicated when you compared Vancouver -- correct me if I am wrong -- you indicated when you compared Vancouver with Ottawa, the Canadian Forces Base Kingston, I believe the words you used, there was no statistical difference in the bin frequencies.

A. That's right.

Q. When you compared these databases in Canada with the databases of the F.B.I., Dade County, Florida, Fort Worth, Texas, and Minnesota, did you notice any difference in bin frequencies?

A. Yes. For some loci for some probes, particularly D2, D10 and in some cases D17, there were statistically significant differences between the bin frequencies in Florida and in Texas. Minnesota it turns out -- perhaps not surprisingly-- is more like the profile of Canada than either Texas or Florida is. However, conceding that there are differences and statistically significant differences in bin frequencies still had virtually had no effect on doing

the forensic calculations as we have gone through for each locus and for the product rule of ultimately getting the forensic probability. Using any of these databases they were within what we call the confidence interval that you would have in that estimate that I had derived based on the Canadian population.

- Q. Did you notice -- I don't know if you mentioned the F.B.I. Did you notice any significant statistical difference in the bin frequencies between the F.B.I. and the R.C.M.P.?
- A. There were I believe -- and I have the data that I can refer back to -- I believe for D2 and D10 there were some statistically significant differences.
- Q. Explain to His Lordship and the court, when you say there is a statistically significant difference in the bin frequencies, yet for forensic purposes there is no difference. Is that right?
- A. That's correct.
- Q. Would you explain what you mean by that and how that is?
- A. Well, it is -- if you have two histograms they can be a slight bit different, but when you are doing these calculations you are using frequencies coming from a number of different bins or at least two different bins for each locus. When you have differences between two populations that are statistically significant some of those bins are going to be higher in frequency. Some of those bins are going to be lower in frequency. The net effect often when you are multiplying between two bins that

bins that are statistically different in two populations is such that they can cancel one another out. That one might be bigger. The other one might be smaller. -- comparing two populations. When you multiply them together the net product of that calculation is often remarkably close so the net effect when you do the calculation for the probability for locus and then when you do the probability calculation using the product rule between loci is often a number that is insignificantly different from the forensic number you originally calculated on your original database.

- Q. So if you get a number, even though you might have individual bin frequencies from the database say the F.B.I. and the R.C.M.P. when you do the calculations because they balance out over the multiplication, you will get a close figure.
- A. That's right.
- Q. Now, when you say close -- here is something else I would like you to clarify for the court -- when you are dealing with, like, in my cheque book I am dealing with very low numbers. I would like to know how -- if the same type of mental processes go into the application of very high numbers. Now, that is something I think is important.
- A. This is important because we are talking about numbers here that are trying to look and predict frequencies of very, very rare occurrences, of very infrequent occurrences of matches and so forth, and these numbers are typically on the order of one in a million or one in ten million, one in a hundre million. That is a

number that is very low. In fact that number is so low that I think most of us -- I know myself -- often have difficulty in thinking of some kind of metaphorical example of a number that is that low.

Q. You say low. I think of it as high.

A. What I mean by low, I mean infrequent. I mean that when we express these it is one divided by a very large number, so one over ten million is a very tiny number, is a very small number, in terms of what that rarity of the event is that it is projecting. So when I say low I mean infrequent. It is a very infrequent number. The number that I am talking about, obviously ten million or hundred million, is a large number but because that is in the denominator of the equation it makes the total number, that estimate of the frequency, very, very low. It means, for example, that when we are basing these calculations on numbers even though our sample is as large as 1,500 or perhaps if we had a sample of --

Q. You say 1,500. You mentioned that before. Is that individuals or bands or --

A. That is individual bands. That means actually exactly half of that number of individuals because we get two bands -- an estimate of two bands from every individual that was run, so this sample that I am talking about of 1,500 -- rough number of 1,500 -- really represents a sample of about 750 individuals.

Q. Let's continue, Doctor, please. Would you explain to us at that high number what differences are we talking about?

A. Yes. That high number -- and it can be deceptive because

that high number is really a number that is trying to indicate the rarity of a certain occurrence. That number statistically can be seemingly very different but not really indicating a significant difference in the rarity of that event. What I mean is if I said that a number was one in a hundred thousand as compared to another estimate that was one in two hundred thousand, you would say quite properly that one hundred thousand is indeed very different from two hundred thousand, but when I do the calculation of one over a hundred thousand and one over two hundred thousand the rarity of those events are insignificantly different. That is what I mean by talking about the numbers at that very low infrequent level not being significantly different so I am saying that if a number is one in a hundred thousand, one in a million, one in two million, they are insignificantly different from one another. The precision of our estimates is not so great that we can say that it is exactly and precisely one in 1.1 million. We would have to give some kind of interval of that estimate to really reliably indicate where we thought that estimate actually was.

- Q. Okay. Hypothetically, Doctor, if we were dealing with a number -- say the difference between -- if you had a report at one in five million and you had another report of one in nine or one in ten million, what would be your opinion with respect to the statistical significance of the difference in those numbers?
- A. Statistically there would not be a test based on the sample sizes that are used in forensic work that could

discriminate and that would say that one in five million is statistically different from one in ten million. In fact I have done calculations to try and give a feeling for how accurate our estimates are in this particular case, and I can describe that in terms of an interval or a span of probabilities that we can be 99% absolutely certain span the range that really is the frequency in an actual population, and that number spans a greater distance than one in five million to one in ten million. I have written down -- and I can refer to my notes on that if we want to get to that later.

- Q. Okay. Perhaps we will when we get into the case specific evidence, but I wanted to clarify that. The difference at those high powers, there is not a great statistical significance?
- A. There is not a great significance, and in fact in these intervals often when you have an estimate of one in five million you could not exclude in fact all the way up to one in ten billion, and on the low side that one in five billion could be as small as one in two million.
- Q. Would you expect -- I ask you this question, Doctor -- if the R.C.M.P. resampled -- went out and got another Caucasian database would you expect differences in the bin frequencies?
- A. Absolutely. It would be very extremely unlikely -- I would say the probability would be unimaginatively small that you would get the identical bin frequency if you took another sample from the same populations in Canada. It would not be identical.

- Q. Would you expect, for example, a difference in the calculations made by an F.B.I. Caucasian database and an R.C.M.P. database?
- A. Very much, yes.
- Q. But would you expect any forensic difference or any statistical difference in the figures that are actually generated?
- A. I have seen in the studies I have done, I see, and I would expect no significant forensic difference in the implications and from the numbers that you would derive from any of those calculations.
- Q. At this time I am going to ask you a series of questions. Perhaps before I do -- you have been talking about the binning method of determining individual band frequencies. Would there ever be an occasion to rebin data? Would it ever be necessary to actually rebin data that has already been binned?
- A. The times that that are done goes back earlier to finding bins where the numbers are very, very low and typically the rule of thumb that is used by both the F.B.I. And the R.C.M.P. is if you have a bin where the frequency is less than 5 you rebin the data so that you put and coalesce adjacent bins so that you get a larger bin now that has a frequency that is never less than 5. That is called a rebinning process.
- Q. And if I wanted to use a new probe would I need -- I would have to bin again would I not?
- A. A new probe? You would have to start and establish what the bins were for the new probe, yes.
- Q. I am going to show you this document here, Doctor. Would you tell me whether you can recognize that?

A. This is the frequency distribution of the bins for the loci that we used in this specific case. They are in fact the database dated December 3, 1990, the total Caucasian rebin database that the calculations in this particular case were based upon.

Q. And you are familiar with this data?

A. Yes, I am.

MR. WALSH: My Lord, I would move to have this entered at the hearing.

THE COURT: VD-64.

MR. WALSH: My Lord, I am going to ask the Doctor a series of questions and then I would suggest a break would be appropriate after I do that. Thank you.

Q. Again I will ask you a series of questions. Is the methodology -- I want your opinion as to the methodology for selecting a database for VNTR forensic purposes as you have described? What is your opinion as to the reasonable reliability of such methodology?

A. I think from the tests and from the readings that I have done in this area that this is a very reliable technique. I anticipate that it will become more refined in time as we get more data and that the techniques themselves might be slightly modified in the future using different approaches, but that fundamentally the reliability of this technique is very great.

Q. And your opinion as to the general acceptance in the scientific community of the methodology for selecting databases in the fashion you have described for VNTR forensic purposes? Your opinion.

- A. In my opinion there is an emerging consensus in the forensic community that this is a reliable way to go about doing it.
- Q. And your opinion as to the reliability of the binning method for forensic purposes.
- A. I think it is very reliable and I know that it is a very conservative way of going about the procedure. That there are built into it a number of places where one makes a very conservative decision of always going towards overestimating the bin frequency rather than underestimating the bin frequency.
- Q. Your opinion as to the general acceptance in the scientific community of the binning method for the calculation of individual allele frequencies for forensic purposes?
- A. In my opinion it is the accepted current standard that people are using.
- Q. Your opinion, Doctor, as to the reasonable reliability of the Caucasian database employed by the R.C.M.P. and the method of pattern frequency calculations made by the R.C.M.P. from that database for forensic purposes?
- A. In my opinion it is a very reliable technique.
- Q. And your opinion as to its scientific acceptability in the general scientific community?
- A. In my opinion it is generally accepted.
- Q. Your opinion, Doctor, as to whether or not the allele frequencies generated from the R.C.M.P. Caucasian database as you have described, whether they reflect the Canadian Caucasian population as a whole including New Brunswick.

- A. Yes. In my opinion, and I think I am able to demonstrate quite conclusively that they represent very well and very accurately and very precisely the Canadian Caucasian population.
- Q. Including New Brunswick?
- A. Including New Brunswick.
- Q. And, Doctor, finally, in this section of our direct, your opinion as to what, if any, bias would be found in the probability figures generated in forensic DNA cases by the R.C.M.P. lab?
- A. In my opinion the techniques and procedures used in the R.C.M.P. lab are going to generate conservative estimates of the frequency of a match.
- Q. In whose favour would that be?
- A. That would be in favour of a defendant or of an accused.

MR. WALSH: At this time, My Lord, if I could suggest a break would be appropriate.

THE COURT: Fifteen minutes then.

(RECESS: 10:50 - 11:15)

THE COURT: Mr. Walsh?

- Q. Dr. Carmody, at this point in time I would like to turn to the --

THE COURT: That is an awful expression. 'At this point in time'. They used that on the Watergate hearings. Either 'at this point' or 'at this time'. You never say 'at this point in time'.

MR. WALSH: Sorry, My Lord.

THE COURT: Am I right, Dr. Carmody?

- A. I think that is correct, Your Honour. I would probably make the same mistake myself.

THE COURT: Mr. Furlotte, what do you say? Am I right?

MR. FURLOTTE: You don't want another argument from me do you?

THE COURT: It is not very serious mind you but --

MR. WALSH: My Lord?

THE COURT: Yes.

MR. WALSH: May I proceed now?

THE COURT: Yes. (laughter)

Q. Dr. Carmody, at this point I would like to turn to the case specific evidence.

A. Yes.

Q. I am showing you a document marked on this hearing 54. Look at that for me please and tell me whether or not you can identify it.

A. Yes. My understanding is that the report that Dr. John Bowen filed on the forensic specimens he was given to analyze from the forensic labs.

Q. With respect to this case.

A. With respect to this case.

Q. Have you had occasion to look at the frequency calculations that Dr. Bowen generated in this report?

A. Yes, I have. I had access to Dr. Bowen's original notes and I went through and verified the arithmetic in all of his calculations.

Q. Did you arrive at some conclusions with respect to these particular statistics?

A. The conclusions I arrived at were first that his calculations indeed were correct. I could find no flaw or evidence of mistake in them. In addition I was able to calculate what I referred to earlier as a kind of confidence band around the particular estimates

that derive from the frequencies, the bin frequencies, and so forth.

Q. Okay. To clarify, Doctor, these calculations that you are referring to -- correct me if I am wrong -- are as follows: "For the DNA typing profile obtained from exhibit 1I (D4S139 matches that of exhibit 56A-69A) the estimated frequency of occurrence in the Caucasian population is less than 1 in 68 male Caucasians." Is that one of the figures you have looked at?

A. That is one of the figures that I verified, yes.

Q. "For DNA typing profile obtained from exhibit 1J (D1S7, D4S139, D10S28 and D17S79 matches that of exhibit 56A-69A) the estimated frequency of occurrence in the Caucasian population is less than 1 in 5.2 million male Caucasians." Is that one of the figures you have reviewed?

A. Yes, it is.

THE COURT: Mr. Walsh, you don't have a spare copy of VD-54 do you?

MR. WALSH: Yes, I do, My Lord.

THE COURT: This is VD-54 is it or a copy?

MR. WALSH: It is a copy of VD-54, My Lord.

Q. "For the DNA typing profile obtained from exhibit 1I0 (D4S139, D10S28 match that of exhibit 56A-69A) the estimated frequency of occurrence in the Caucasian population is less than 1 in 7,400 male Caucasians." Have you looked at that figure?

A. Yes, I have and I verified that.

Q. "For the DNA/typing profile obtained from exhibit 135 (D1S7, D2S44, D4S139, D10S28 and D17S79 match that of exhibit 56A-69A) the estimated frequency of occurrence in the

Caucasian population is less than 1 in 310 million male Caucasians." Have you looked at that figure?

A. Yes. I recalculated that and I verified it as correct.

Q. And the calculations you looked at, the frequencies were obtained in the manner you described earlier this morning?

A. Yes, they were. I was using the bin frequencies in the R.C.M.P. database in that document that you submitted earlier.

Q. Have you had occasion to run these particular frequencies or bin frequencies and the calculations through other databases other than the R.C.M.P. Caucasian database?

A. Yes, I did. I ran through and I believe you have a sheet that summarizes the calculations I made there.

Q. I show you this document.

A. Yes. This is a compilation of the calculations that I made.

Q. These were prepared by you.

A. Yes, they were.

MR. WALSH: I move to have this entered at this hearing.

THE COURT: VD-65.

Q. I will show you VD-65, Dr. Carmody. I will show you what appears to be a duplicate. Is that an accurate duplicate?

A. Yes, it is.

MR. WALSH: I would give the duplicate to His Lordship.

Q. Dr. Carmody, would you please explain what you have done here and explain the figures you have generated in relation to this particular case, the case of Allan Joseph Legere?

A. Well, if I start with looking at the individual probes, they are indicated in the left-hand column as D1S7, D2S44, D4, D10, D17, etc. I have indicated in the column next to that in the middle of that column I have what is called the point estimate or the estimate that one would calculate using the bin frequencies and so for D1, using the Canadian Caucasian database, one gets an estimate of 1 in 78 as the frequency of the appearance of that particular genotype of the defendant in the Caucasian database. Flanking that on either side, the 1 in 56 and the 1 in 129, are the boundaries on what we referred to statistically as the 99% confidence interval.

Q. Meaning?

A. That is that that is the span of a frequency. 1 in 56 to 1 in 129. That if we were to resample and take additional samples from the identical Canadian Caucasian population we would expect the numbers that you calculated for that locus to fall within, that range greater than 99% of the time. That is, that if one took another sample you might get 1 in 78 again, but it could be 1 in 81, it could be 1 in 65. It could deviate as far as from 1 in 56 to 1 in 29 (sic). That interval is an attempt to give a feeling for the range that that estimate would move about upon resampling.

Q. So it is a method of allowing for sampling error. Would I be correct?

A. That's correct.

Q. Now, Doctor --

A. And since we are making these inferences based on a

particular finite sample, though it be 750 odd individuals, it nevertheless will change if we were to take another sample.

Q. Because of its size.

A. Because of its size and just by the fact that you are randomly sampling. You are going to get maybe a few things in higher frequency, a few things in lower frequency.

Q. The application of confidence intervals, is that an accepted method of determining these kinds of probability figures?

A. That is a standard statistical technique to give a feeling for the range within a particular estimate.

Q. And in the frequencies that Dr. Bowen generated what would the individual bin frequency be for D1S7 that Dr. Bowen would have used?

A. I don't have that data before me. I have it in my notes, but there would be perhaps in one bin it might be 1 in 11. In another bin it might be 1 in 5. Whatever. And you multiply those through to get the 1 in 78.

Q. Okay.

A. I have that in my notes if you would like --

Q. No, that is fine.

A. If you would like me to continue through that table --

Q. Please.

A. I have done the equivalent for each of the five loci that are pertinent to this particular case, and you will see, just running down that, that for the D2 locus the estimate is 1 in 59 and the confidence interval on that runs from 1 in 44 to 1 in 88. The D4 locus,

the estimate is 1 in 68 and the confidence interval is 1 in 50 to 1 in 107. For D10 it is 108 ranging from 1 in 77 to 1 in 184. For the D17 locus it is one-eighth and it ranges from 1 in 7 to 1 in 12.

- Q. These would be 1 in 8 --
- A. -- individuals would expect to have this same genotype at that locus.
- Q. In the Caucasian population.
- A. In the Caucasian population.
- Q. Now, the numbers that would be demonstrated in the report that is filed, the numbers that would actually be used would be 1 in 78, 1 in 59, 1 in 68, 1 in 108 and 1 in 8.
- A. That's right. And for the one calculation which is given in the very last row under five loci you multiply together those middle figures of 1 in 78, 1 in 59, 1 in 68, 1 in 108 and 1 in 8. Actually they carry to more decimal places than that, but you multiply those together and you get the estimate that is indicated in the lowest row there of 1 in 310 million.
- Q. For the four loci?
- A. And for the four loci, multiplying together D1, D4, D10 and D17, you get the figure of 1 in 5.2 million.
- Q. What are those numbers on either side of that?
- A. Those numbers again are the equivalent 99% confidence limits that one would place on that estimate. That is, that one would expect upon resampling. That in the case of the four loci where the estimate is 1 in 5.2 million that upon resampling from the same population the number that you generated could fall anywhere between 1 in 3.1 million and 1 in 17 million.

It gives a sense to the fact that 1 in 5 million, 1 in 10 million, 1 in 3 million are not statistically significantly different from each other.

Q. What you had discussed earlier this morning.

A. That's right. And that even though we quote a figure in this, and our very best estimate given the data that we have is that middle value, that is our best estimate, and indeed if we took another sample and made another estimate it would likely fall reasonably close to that, but we can be 99% certain it would not fall outside the range that I have indicated with that confidence interval.

Q. Okay. If I understand you correctly, Doctor, between what Dr. Bowen declared a match between exhibit 1J and 56A-69A using those four loci in which he generates 1 in 5.2 million male Caucasians being the probability--

A. Right.

Q. -- you are saying that that on a resampling could go as low as 1 in 3.1 million male Caucasians or as high, as 1 in 17 million male Caucasians.

A. That's right.

Q. The probability of finding someone else with that same pattern over those four loci.

A. That's right. And given the fact that we have looked at merely 750 individuals we have to take into account the fact that additional samples might give slightly different answers.

Q. And that is the application of the confidence interval.

A. That's right.

Q. All right. Would you go to the five loci.

A. For the five loci, in an equivalent way, there is just an additional multiplication and a few more calculations to get the confidence interval but again there the figure in the centre, the 1 in 310 million, is the estimated frequency of the particular banding patterns that were found at all five loci that were calculated by Dr. Bowen and that give the 1 in 310 million with a span in that 99% confidence region being 1 in 175 million to 1 in 1.3 billion.

Q. Doctor, meaning that Dr. Bowen's report where he says there was a match between exhibit 135 and 56A-69A at five loci, he has generated a figure of 1 in 310 million male Caucasians, the probability of seeing this pattern. When you apply 99% confidence intervals to that, if I am correct, that could be a pattern as frequent as 1 in 175 million?

A. Yes.

Q. Or as infrequent as 1 in 1.3 billion male Caucasians before you would see that pattern again.

A. Yes. Now, in this document I have also done calculations to indicate how, if you use other Caucasian databases, those numbers that were calculated by Dr. Bowen might vary. So just on the top most area where I am looking at individual loci, for D1 where the calculation based on the R.C.M.P. total Caucasian database is 1 in 78, if you use the F.B.I. Caucasian database you would get and derive a figure of 1 in 96. If you use the database I had to work with from Dade County, Florida it would be 1 in 80. If I use the database from the State of Minnesota it would be 1 in 76. Alas I did not have any information from France on that locus so I can't

indicate a figure for that, but I think in running across you will see that 1 in 78 with the confidence interval of 1 in 56 to 1 in 129 encompasses in fact even the geographic differences that are picked up in the American databases. One from a composite of Caucasian populations that are put together by the F.B.I., one deriving solely from Dade County, Florida and one coming from the State of Minnesota. Equivalently going along for each of the other four loci that were used in this study where Dr. Bowen got 1 in 59; using the F.B.I. database is 1 in 70, the Florida database is 1 in 73; the Minnesota database is 1 in 48; and the database from France is 1 in 34. The database from France falls below that confidence interval for that locus and is more frequent -- a more frequent variant genotype found in the French population. Looking at D4 where the calculation for the Canadian database is 1 in 68; the F.B.I. database gives 1 in 98; the Florida database 1 in 100; the Minnesota database 1 in 73. D10 where Canadian calculation gives us 1 in 108; F.B.I. is 1 in 92; Minnesota is 1 in 143. I did not have information on that locus from Florida to work with so I couldn't do the calculation. And I had information from France. In that case it is 1 in 54. Again the French variants are -- that French -- frequency for that genotype is more frequent than is found in any North American Caucasian populations. Lastly for the D17 locus where the calculation rounded off is 1 in 8; using the F.B.I. database it is 1 in 11; Florida database it is 1 in 12;

Minnesota database is 1 in 10. I have taken those individual calculations for both the Minnesota populations and the F.B.I. populations where I had all of the same loci calculated and ran through the chain or product rule calculation to derive the equivalent number that we have for four loci where in the Canadian case we get 1 in 5.2 million; using the database for Minnesota that calculation would be 1 in 8.4 million; using the F.B.I. database it would be 1 in 9.9 million.

Q. Now, Doctor, let's stop there. The difference, 1 in 5.2 million, as done by the R.C.M.P., and Minnesota 1 in 8.4 and the F.B.I. is 1 in 9.9 million. Again would you tell the court what statistical difference that has for you?

A. Given that these are based on samples of the size roughly of 750 individuals and a total therefore of double that number of bands that were analyzed, these numbers are not statistically significantly different from one another even though I would have to concede in any normal use of a word of 5.2 million is certainly different from 9.9 in the way we normally think of millions. When you are down at this very infrequent level these numbers cannot be statistically separated from one another.

Q. And they span the confidence intervals between the Canadian, the Minnesota and the F.B.I., they span the confidence intervals -- 99% confidence intervals.

A. Even if you were to resample the same population, yes, so that forensically the conclusion that I am presenting here says that though there are differences, those differences forensically are not significant.

Q. And with respect to the comparison between Minnesota and the F.B.I. and the five loci?

A. Again, in that case where the estimate that Dr. Bowen calculated and that I corroborated, is 1 in 310 million. If you were to use the Minnesota database it would come out to be 1 in 402 million or in the case of the F.B.I. composite Caucasian database it is 1 in 698 million. Those numbers fall well within the 99% confidence interval that I have calculated on resampling the Canadian database.

Q. And your conclusions would be, Doctor?

A. That in fact the differences in those numbers are not significant. Forensically significant. Statistically significant.

Q. What does that tell you about Caucasian populations in North America generally?

A. It says that using this technique as it applies to this case gives reliable -- as we would say in statistics and the term I used earlier -- robust estimates of the actual frequency in Caucasian populations for these genetic variants.

Q. Now -- but when you compare -- even though there is statistical differences between Canada and United States Caucasian populations, when you did the comparison between the Caucasian databases developed in Canada I understand that you didn't find any statistical difference even there.

A. That when I compared the bin frequencies --

Q. The bin frequencies, yes.

A. The bin frequencies, there were indeed differences for some loci between the Canadian database and Florida and Fort Worth, Texas.

Q. Right.

A. For the bin frequencies. But as you will see here these differences that you generate for individual loci for genotype frequencies and for the ultimate genotype frequency based on four loci or five loci are not statistically significantly different from one another.

Q. And when you did your bin frequency comparison between Vancouver, Ottawa and Kingston, your findings there?

A. That even for the bin frequencies there there were no statistical differences and it gave the appearance of a statistical homogeneous population.

Q. What about France, Doctor? You have that added in there. It is not part of North America but --

A. Right.

Q. What significance would that have? What meaning or interpretation can you give to that?

A. I think it is trying to give a sense of how Caucasian populations might vary when we get larger samples taken worldwide. These are taken from two loci that are known to vary quite a bit from one Caucasian population to the next though they don't in Canada, but these two loci, D2 and D10 in particular, show quite a bit of variation in bin frequencies. Statistically significant differences in bin frequencies from one Caucasian to the next and they are indicating that in fact these variants are much more common in France than they are in present-day Caucasian populations in North America.

Q. But do they in any way invalidate any of the opinions or cause you any concern with respect to the opinions that you have actually given?

- A. They caused me no concern because I think from looking at the database and representation, the geographic representation in our database that this is based on for the Canadian calculations, that there is a good representation from the Province of Quebec, for example, and from the Maritime Provinces, where we know the majority of decendants from France that have settled in North America reside.
- Q. With respect to the bin frequencies that were generated in this particular case, can you tell the court something about how common or -- the bin frequencies, were they in bins that had a lot of bands in them or were they in very rare bins?
- A. In all cases, I would refer to my notes again to get the specific details, but in all cases I can state categorically all of the particular genetic variants that were present in these forensic samples are for each probe very common bands. They are not rare ones. They are not ones that were ever in bins that had to be rebinned and put together because there were too few variants in the sample. They are in some of the most common bins at each locus.
- Q. Does that tell you something about the sample here in relation to the Caucasian database?
- A. Well, it says that these numbers are more reliable. That is, if you have a number that is higher in frequency you in general have a greater precision and a greater reliability of that estimate than if you have a very rare variant. The rare variants, by virtue of the fact that you have a finite sample, are going to occur only very rarely in that sample, and so the numbers from one sample to another are likely to move

around much more than when you have a much more common variant where it will be more stable in your samples.

- Q. So that the bands that were seen in the samples that were tested here, the matches, those were bands that were very commonly found in the samples from Vancouver, Kingston, and Ottawa?
- A. Yes, they were. They were -- in no case were they rare variants or rare frequency bands.
- Q. Doctor, what is the figures that Dr. Bowen generated? The ones -- particularly 1 in 5.2 million male Caucasians, 1 in 310 million male Caucasians. What is the meaning that can be taken by the court from those particular -- from a population geneticist's point of view, what meaning can be taken from the existence of those matches?
- A. The inference is that the occurrence of that particular genetic type has the estimated frequency of that value of -- in the case of the four loci -- 1 in 5.2 million in Canadian Caucasian population. And in the case of looking at five loci it is 1 in 310 million is the expected occurrence of that same genotype in the Canadian Caucasian population.
- Q. From a qualitative statement point of view what does that mean to you as a population geneticist?
- A. These are very, very rare genotypes. Virtually each of us are unique in terms of our genotype if you can look at enough of it. At the present state of development of the technique of DNA fingerprinting we can only look at a few sections of our genotypes and we can only see part of the differences between individuals. If we

could look at enough of the DNA in us we could show every individual is unique. This is saying by looking at these five snapshots of the genotype of the DNA in an individual that the estimated frequency of that is 1 in 5.2 million using four loci or 1 in 310 million. In other words, these are very rare variants. Each of us is going to turn out ultimately when we get more information to be shown to be genetically unique except for identical twins, but we will be genetically unique. The reasons that we have to go through these calculations and we can't show that this would never occur in any other individual is that we are limited at the present time by only being able to sample these particular sites in the DNA.

THE COURT: Just to enlarge on that a little. What you are in effect saying, if I understand you correctly, is that in Canada in the Caucasian population 1 out of 5.2 million people would have their bands showing, say under column "B", in two precise locations using one probe and in certain other combinations of two using three other probes.

A. That's correct, My Lord. Exactly. But that is the probability that you would have the various combinations for all five probes occurring in that individual and that it is -- it is a measure of the rarity of that particular combination.

THE COURT: But when Dr. Bowen says, where the DNA typing profile obtained from exhibit 1J say, the estimated frequency of occurrence in the Caucasian population is less than 1 in 5.2 -- does that mean the

person being matched and the matchee, does that mean
2 out of 5.2 million or --

- A. No. It doesn't. The inference is there that the
chance of that -- having a match coming from two
different individual is 1 in 5.2 million.

THE COURT: You see what I mean.

- A. Yes, but it doesn't -- I guess you could take the
inference too, and you could put it in terms you used.
That the chance that there would be two people that
matched like that would be one in 5.2 million, yes,
for those four probes. Yes, that would be correct.

THE COURT: So it is really two. Not one. Well --

- A. Yes.

THE COURT: -- we are talking the same language.

- A. Yes. It is. But you wouldn't -- the way this is
stated is that if you were -- how large a sample would
you typically need in order to find one of these
would be 5.2 and it wouldn't mean that if I took one
more in my sample that I would get of them. The
chance of getting another one would be 1 in 5.2 million
so in fact the chance of getting two of them from
two separate samples would in fact be the product of
those. 1 in 5.2 million times 1 in 5.2 million.

- Q. Dr. Carmody, at this point in time do you know a person
by the name of Dr. William Shields?

- A. Yes, I do. I met him at a previous case.

- Q. Are you aware of any work or any recent work Dr.
Shields has done in relation to comparisons between
the R.C.M.P. database and the F.B.I. database?

- A. Yes. I understand he has made some comparisons and
he has had the F.B.I. Caucasian database and the R.C.M.P.

database and he has made some comparisons of the bin frequencies.

Q. Do you have any opinions with respect to the work that he has done?

A. I think the work that he has done is correct. I think he has found statistically significant differences for some loci in the bin frequencies between the Canadian database and the R.C.M.P. database, as I have, and I have corroborated his findings.

Q. But how does that interpret in terms -- for forensic use?

A. In terms of forensic use, as I have indicated in this document that I have just been discussing, one can see the results in this particular case. That in fact there is no forensically significant difference even though the bin frequencies are slightly and in fact statistically significantly different in the F.B.I. database than in the database that we used in the Canadian calculations.

Q. And you are referring to the document that is marked VD-65.

A. Yes. That is the document. To just go back to that. Making the calculations as I did for the four loci it is 1 in 5.2 million. In fact if you use the F.B.I. database it becomes 1 in 9.9 million, and as I pointed out, that is statistically not significantly different one from the other even though in most usages of arithmetic we think 5.2 million is considerably different from 9.9, but when you are dealing with these extremely rare events those cannot be discriminated statistically.

- Q. It is the power of those high numbers.
- A. That's correct. Even though one is almost double the other you are dealing with such low frequencies here that in fact one cannot discriminate between those.
- Q. I see. Does -- Doctor, do you have any other opinions with respect to the conclusions Dr. Shields has drawn with respect to his comparison?
- A. One of the conclusions he drew/I disagree with was that he made a statement in a report or an affidavit that he had filed that I had seen that -- he said that if you use a different database you will invariably get lower estimates and be more prejudicial against the defendant if you use the wrong database. In general I would argue that that is incorrect because if you have two different databases some bins -- and if they are statistically different -- some bins are going to be larger, some bins are going to be smaller. As I have shown here, by making comparisons with a number of U.S. databases against the Canadian one you will see sometimes they go up and sometimes they go down but they are not statistically different from one another. In most cases, in this particular instance, they will lower when you use the U.S. Caucasian database than if you use the Canadian one. That is they are more rare and they are more prejudicial against the defendant, but in general that statistically is not what you would conclude when you use two different distributions of bin frequencies. Sometimes for a particular bin it will be higher. Sometimes it will be lower. In

general and on average they may well cancel out.

- Q. Doctor, there is terminology in the case law dealing with substructuring. We have touched on this this morning but I would just like to -- perhaps if we could just recap the application of the term 'substructuring' to what you have testified to this morning.
- A. Substructuring is a term that is used in population genetics to indicate that if you are looking at trying to understand the genetic dynamic evolutionary processes of a population that you could not treat that population as one homogeneous unit where there was complete random mixing of all the genetic variants. It means that in fact by substructuring that one really has to be sensitive to the fact that within that geographic or demographic unit that one is studying that there are smaller components within which there might be some differences, and so one of the uses of substructuring is to indicate and try and convey the idea that the population that you are studying in toto is not a homogeneous unit and should not be treated statistically or mathematically as a homogeneous unit. The consequences of having smaller and substructuring in populations are that you can get deviations from the predictions of the Hardy-Weinberg equation, from the predictions of the product rule and so forth. That you would get perhaps an excess of homozygosity. That you would get gene frequency and bin frequency differences geographically. All of those could be a consequence of having substructuring in a population and they are very necessary to be aware of that possibility when studying

human populations because, as has been well documented, human populations are not one homogeneous inter-breeding genetically uniform mixture like that. They indeed are made up of separate ethnic, geographic, socio-economic geographic units within which sometimes there is not complete random mating.

- Q. You accept that that does in fact go on in human populations.
- A. It certainly does and it has been well documented.
- Q. Both in Canada and United States you accept that.
- A. Both in Canada and United States and certainly in Europe and Asia.
- Q. Does that in any way affect the opinions you have given to this court with respect to the validity of the numbers that are generated by the R.C.M.P. database and in particular as it applies to this case?
- A. It does not invalidate them because I have been able to show, for the Canadian database that we use for these calculations, that there was absolutely no evidence of what we call substructuring. That is, that we could treat the Canadian Caucasian population as a homogeneous genetic unit.
- Q. And even the comparisons between -- what would be the statistically significant understanding you would have or the result from seeing something that is statistically significantly different? Like, for example, between the F.B.I. and the R.C.M.P. database? What does that tell you in relation to substructuring?
- A. It tells me that there is some substructuring present. That is, that you do find these differences in bin frequencies, and that -- but that that substructuring,

when you look at it in terms of its effect on any calculations that we are going to make in terms of forensic inferences, has no consequence.

- Q. If you were doing something other than forensics, would substructuring, even of a minor degree, have an important effect?
- A. Depending on the study it might well have an effect and indeed some of the more interesting studies in this area are to look at some of our aboriginal populations to see how different they are, because one would expect just from some of the anthropological information we have that they could show much higher degrees of substructuring, for example, and amongst perhaps some ethnic groups there might be some evidence in future, as we get larger samples, of some substructuring occurring within those groups.
- Q. But the effect of substructuring as you have seen it or between the Canadian and the United States population, does that give you any cause for concern in terms of the figures used here?
- A. It gives me no cause for concern when I do the calculations and try to look at any effect of this substructuring. It has really an effect that is inconsequential for the forensic inferences that we are using this data for.
- Q. Doctor, the gentleman that is sitting over against the wall between the two police officers, with the white shirt, could you tell me please what racial group you would say he belongs to?
- A. I would say he belongs to a Caucasian group.
- Q. Is there anything, Doctor, that perhaps you would like to add that I haven't covered?

A. No, there isn't. I guess I would say that I have confidence in the techniques that are used and certainly in the statistical techniques that are used to make the inferences that are being supported in the court.

MR. WALSH: Thank you, Doctor. I have no further questions, My Lord.

THE COURT: Thank you very much. Mr. Furlotte?

MR. FURLOTTE: My Lord, I would prefer to wait until after lunch break to begin my cross examination on this witness. If we take an early lunch break we could come back early.

THE COURT: That is fair enough.

(RECESS: 12:00 - 1:30)

THE COURT: Cross examination, Mr. Furlotte?

MR. RYAN: Prior to cross examination, My Lord, if I may and with the court's indulgence, I would advise the court that during the morning recess and the noon recess Mr. Legere and I have been discussing a matter that he wanted brought to the court's attention with respect to his own personal view and in putting on the record at this point his dissatisfaction with the situation regarding the request for an adjournment to allow Mr. Furlotte to prepare further for cross examination and the court's decision on that. Understanding the protocol that the court has decided with respect to commentaries between the court itself and Mr. Legere, I advised it probably would be best if I put forth those remarks to the court on the record. Thank you, My Lord.

THE COURT: Mr. Furlotte?

CROSS EXAMINATION BY MR. FURLOTTE:

- Q. Dr. Carmody, I believe you said you studied or you were working in a lab with Dr. Richard Lewontin.
- A. That's correct.
- Q. How do you spell his last name?
- A. L-e-w-o-n-t-i-n.
- Q. Okay. I was always pronouncing it myself as Lewontin.
- A. I know. There is a difference in pronunciation but that is the way he pronounces it so I take that to be the correct way.
- Q. He deserves the dignity of having his name pronounced right. How would Dr. Lewontin stand in the scientific community as to his expertise?
- A. He is one of the premier population geneticists in the world.
- Q. Probably rated as the best?
- A. Possibly by some people, yes.
- Q. How would Dr. Eric Lander stand?
- A. He would have a very high standing as well although he is not strictly a population geneticist. He is more qualified as a molecular biologist.
- Q. How would Dr. Kidd compare to Dr. Lewontin and Dr. Landers?
- A. I would say he would have equal status particularly in the area of human population genetics. I think in both the cases of Dr. Lander and Dr. Lewontin, they are strictly -- their expertise is more general and is not directed specifically towards human populations.

- Q. I believe you stated that for yourself you are more interested rather in formulating these models or whatever? As I understood, you are more interested in going out and testing to see whether these models or theories are correct.
- A. That's correct. That is my interest.
- Q. In other words, would you consider yourself more like a field worker? To go out and collect the data to prove or disprove the theories?
- A. Well, I am more interested in getting the data. Some of the data in fact is generated in the laboratory and then in testing that data against the predictions of theories. Although I do do field work.
- Q. The data that you are using to test those theories, that is basically data that other people collect and you use it.
- A. In the case of the forensic DNA material that I am testifying on today, yes, it is data that was collected and produced in the laboratories at the R.C.M.P. or equivalent laboratories at other places when I referred to the American or F.B.I. data.
- Q. Have you yourself ever run autorads and gels?
- A. Yes, I have. On drosophila.
- Q. On human DNA?
- A. I have not run any on human DNA myself.
- Q. I assume you have no experience in forensic analysis whatsoever.
- A. Not prior to eight months -- well, that is not strictly true. I was involved in a case for the Ministry of Natural Resources about five years ago. It involved some testimony about whether some frogs

legs came from bull frogs or from some other species of frogs and my expert testimony -- although the case never went to trial, I produced the results that they used -- were going to use as evidence, so I have that experience forensically.

Q. How many times have you come to court to testify as an expert witness?

A. This is my second appearance.

Q. And your prior appearance was in which trial?

A. It was in the voir dire of Bourguignon in Ottawa.

Q. At which Dr. William Shields also testified.

A. That's correct, and that is where I met him for the first time.

Q. I understand that you and Dr. Shields were in basic agreement in that trial were you?

A. Well, to be honest I wasn't present at his testimony. He had some disagreements about the conservative nature of the R.C.M.P. system that I would disagree with him about. I would say we disagreed more than we agreed. I think we both agreed that this evidence is very useful and has great potential to be used in forensic work. If I were to give my impression of his feelings about it, is that he is more concerned that the actual databases are not stable enough to give reliable estimates. That would be my interpretation I think of what his feelings are.

Q. This problem was -- I won't say a problem but the technique we have -- that you have testified today about comparing the samples in a specific case with different databases across Canada and throughout United States, North America, and you even went to Europe I see, why did this come about?

- A. Well, there is always the question as to whether in fact there is population substructuring, and what the effect of that would be on these inferences that we are trying to draw. As in any new area of science, people are always trying to discover what the problems are and what the solutions to those problems are so you naturally examine databases in this fashion to see whether in fact -- as I did with the Canadian populations -- whether in fact it was legitimate statistically to amalgamate those into one unit.
- Q. Did somebody specifically ask you to make this comparison?
- A. In the case -- it was brought to my attention specifically, the information that Dr. Shields had, in an affidavit that was given to me to look at, yes.
- Q. You were made aware that since Dr. Shields was going to I assume testify in this proceeding here, that he had used the information I gave him to make the comparison in a court hearing in the States. Would that be basically --
- A. I was given this affidavit that he had used at a U.S. court to get my opinion about and I redid some of those calculations just to convince myself that they were correct, and then in fact got some other data from U.S. populations that were not part of that affidavit to test my ideas about that further. To see whether in fact the substructuring that he claimed was there made any significant difference forensically.
- Q. Who provided you with that affidavit? The R.C.M.P. lab or the F.B.I.?

- A. It was provided for me by the Crown here in New Brunswick.
- Q. So you knew that basically Dr. Shields would be coming to this court proceeding and providing the same information as he gave in the hearing in the States.
- A. That's correct. Well, I wasn't sure what he was going to be providing but at least I was asked my opinion about that affidavit.
- Q. So a lot of the evidence that you gave and the comparisons, this is basically rebuttal evidence. It is what you expect Dr. Shields to come in and testify to.
- A. I was just asked to give information of what I thought of this evidence in the affidavit.
- Q. It never came to you mind to make these comparisons in the last trial you testified in. The Bourguignon case?
- A. We didn't have that data at that time actually. I did not have the F.B.I. data. In fact as a research program we intend to continue this kind of analysis particularly extending it further geographically to European populations and further populations in Canada indeed, so at the time that I testified in the Bourguignon trial in early January this data was not available to me and I hadn't yet started on a fuller analysis.
- Q. Would you admit, Dr. Carmody, that the idea that Dr. Shields come forward with is a good idea to either prove or disprove the ability to use the Hardy-Weinberg formula?
- A. Yes. I think it is important to make these comparisons.

Indeed I had made the comparisons on the data that I had prior to that. I did not have access to any of the American data prior to getting that affidavit. I agree that it is a good idea to do that, and indeed I have been doing that with the Canadian data that has been provided to me.

Q. Now, whether or not the differences in the number of different databases throughout North America and Europe, whether or not the ones in North America are substantial enough to prove substructure or linkage equilibrium, would it be safe to say that the data that you were using and which you formed your opinion on, that that in itself has not went to the general scientific community yet to establish whether or not your opinion is correct or whether Dr. Shields' opinion is correct?

A. That is true. It has not been published in peer reviewed -- in the peer reviewed literature at the present time.

Q. So you would admit that Dr. Shields' opinion may in the end be accepted in the general scientific community rather than your own?

A. That is possible.

Q. Population genetic studies within a population -- I mean like within an ethnic group or within a race or within geographic locations -- I understand that Dr. Kidd has done a study and he has a paper in press. Are you aware of that?

A. I am not aware of that. I know he has done many studies. I am not aware of any particular one that you are referring to.

Q. Where he has studied what, American Indians or South American Indians?

A. He has been involved in South American Indians in the past and a number of African populations and a number of populations, but I am not aware of the particular paper you are referring to.

Q. You are not aware of the impress of the Amerindian data?

A. No, I am not. I am not aware of that data.

Q. Are you aware of the studies done by Dr. Fournery with Canadian aboriginal Indians?

A. I have been involved in some of the comparisons that have been made there with the aboriginal Indians. I am not aware of those that have appeared in the literature at the present time, Dr. Fournery's data on that.

Q. Are you aware as to whether or not the study that he did shows that there is substructure within the Canadian Indians that would throw the linkage -- that would show linkage disequilibrium amongst the Indian groups?

A. Well, I am aware of the data and in fact I have made some comparisons myself of the two Indian populations that the R.C.M.P. has and indeed they are different in their bin frequencies. I am not aware of any implications of that in terms of linkage disequilibrium. I think it is important, and I know this is a technical point, but those studies on bin frequencies do not allow you to say anything about potential linkage disequilibrium in and of itself, because in order to do that you have to have the combinations of what alleles are present at these various probe sites

in the same individual and the bin frequencies basically look in isolation just at one probe site at a time and so you don't have the information in those kinds of studies to show the linkage disequilibrium.

Q. But the fact that there is a difference, and a substantial difference, in the frequencies or the bin frequencies for each probe, doesn't that really put you on guard that there is a strong likelihood that there is linkage disequilibrium? I will get my tongue around that word some time or other.

A. Yes. Well, my answer to that -- the short answer is no. What it does make me worry about, let's say in a forensic case, if it involved a Canadian aboriginal, I would worry about which database would be relevant to that and indeed I was consulted about a case in British Columbia where we had some data on the coastal Indian tribes and we had the bin frequencies for those, and we had data from Northern Ontario and around Winnipeg and they were so different that when I was asked / ^{and} said, well, what would you do in the case of -- this was around Penticton -- What would you say would be the relevant database there and what would be the frequencies to use? And I said at the time, and I would agree, that I don't know what the frequencies are likely to be because they were so different, the one to the other, and each of them were so different from Caucasian frequencies that I have most experience with, that indeed in the Penticton region, you could almost take your pick as to what frequencies you would use. So I was worried there not so much that within each of these groups there might be

some deviation from Hardy-Weinberg equilibrium or linkage disequilibrium or whatever. I think in those cases I was -- in fact I know in those cases I was more concerned about the fact of what the relevant population was. Furthermore when I was told that indeed the individuals in that area and the particular suspect in that case may not have had his birthplace in that area, though he was aboriginal, it threw up in the air entirely what frequencies to use. You could calculate frequencies on anything you had and then say, well, take your choice, but I couldn't choose which ones to use. So I was more concerned of that which we call more sort of population subdivision which is sort of an aspect of the substructuring. We call it population subdivision where in fact there is much more likely to be a coherent sort of homogeneous population just locally where there is very little migration in and out. There you do have to worry about that. When you have the evidence confronting you from samples that we do have, knowing that that is the case in these aboriginal areas, I was more worried with that. Then concerned about there would be linkage disequilibrium or non-Hardy-Weinberg equilibrium within each of them. I am not too concerned about that actually. I would think that you are likely to still^{see}/the equilibrium within those areas.

Q. Yes. Within those areas. Like in forensics when you have an accused person you can track down that accused person as to whether he may have come from and who his parents were and where he was born and all that, but you can't track the evidence that was found at the

scene because it is unknown. You haven't got a clue where he come from.

- A. That's right. And there may be some forensic background where you can say, well, this was an isolated community and the alternative person that committed this crime is very, very likely to have come from that general geographic area. You can never rule out somebody having flown in from Europe or something like that, but the more reasonable assumption is that it was likely to be a suspect from that area. It makes more sense to use a database from that particular local area where you have that kind of subdivision.
- Q. Yes. That is fine if we are comparing here where you compared the database from France and from Canada and there is enough significant difference there that you would not be able to use the Hardy-Weinberg formula. Is that -- do I understand that correct?
- A. Yes. I wouldn't say that. I would say in the French data unfortunately I just had information on two probes and I think that is all that is available at the present time. I would still, if I had data on all five of the probes and comparable information, I would still like to do the calculation because I am not convinced if you went through and did the total calculation you would come up with a number that is necessarily significantly different from the number that we have in the Canadian database. I would have to see more full data. I provided that because it had been given to me in some data that we obtained from the F.B.I. They had somebody from France working in

their lab generating the data on the same system that the F.B.I. used so it was nicely controlled that way. And it was of interest to see how different it would be. For the two loci that I did have the information on indeed they were somewhat different in terms of their individual locus frequencies, but then I only had two out of the five loci and if I had the other information and carried it through, it could well be insignificantly different from the Canadian sample even based on the French data.

- Q. If you were given a grant to prove linkage disequilibrium in the Canadian Caucasian database how would you go about it?
- A. First of all, you would talk about getting very large samples. We would have to get samples on the order I would estimate at this time of 50,000, 100,000 individuals and what you would have to do is you would not only have to get the information on the individual but you would have to get information on both their parents, because when you have information on an individual you basically, at each locus, have information on the two bands. What you want to know to satisfy yourself about linkage disequilibrium is you have to know, for each of those two -- let's say if we have two loci like that and suppose they were two different probes on that one gel, which is not typically the case, but you would want to know which of those in that first lane, which of those two bands came from the mother, which came from the father. In the second lane you would want to know which of those two bands came from the mother, which came from

the father, and then when you had that information on a large enough sample size, knowing which band was inherited with which other band, you could then use that information to survey the population to say what the frequency of that combination is, because linkage disequilibrium is a notion that a particular band at one locus shows a nonrandom association with a particular band at another locus. In order to show that in a rigorous way, in a rigorous incontrovertible way, you would need first of all a very large database and you would want to get information not only on people that had contributed specimens to your sample but on both of their biological parents to show that in fact what was being inherited was in fact a nonrandom combination or a random combination, whichever it turned out to be the case. Strictly speaking that is what linkage disequilibrium means. It means that the band that is present at one locus is not completely free to have an association with any other band at another locus. That is what we mean by disequilibrium, tend to be yoked together in some way.

- Q. That is why your database has to come from unrelated individuals, because in related individuals your band is not free because it is tied in with your parents.
- A. Well, if you get related individuals it is a problem there because you are more likely to get combinations that -- because of their biological relatedness -- are going to have the same combinations. And what you want to know is to look at a random sample so that you don't get an over-representation of people that are showing this combination strictly because of their recent common ancestry.

- Q. I assume it would take the same method to prove that there was linkage equilibrium in a population.
- A. That's right. One goes hand in hand with the other. I mean these are hypotheses that you test as either/or. You have either linkage disequilibrium or if you can show there is no linkage disequilibrium then you can conclude there is linkage equilibrium. The simpler hypothesis to test is whether there is equilibrium and the reason for that being more simple is that if you say that there is equilibrium it makes a prediction. It makes a prediction that in fact what band is present at one locus is not going to allow you to conclude what band is going to be present at another locus. It has got to be free to vary. So it makes that extremely accurate prediction of what you would expect to see. It is a neat mathematical model that way. Whereas the model -- and so you can test it. If your results don't fit that then you can say, well, we have been able to rule out equilibrium. It must be disequilibrium then. On the other hand, when you are setting yourself up to test for disequilibrium there can be all different magnitudes of disequilibrium and it doesn't make a simple prediction. It makes a prediction that there will be some correlations, but depending on the amount of disequilibrium the correlations could be stronger or weaker, so it is a less easily tested mathematical model because the predictions that it makes are much more fuzzy.
- Q. And therefore less reliable?
- A. And much more difficult to test because you have to --

any test of a scientific hypothesis is made more easy if you have very definite predictions that it makes that you can test against. Linkage disequilibrium is a whole range of possible correlations that you could have because some loci and not other loci and so forth. That is very difficult to test because the precision of the prediction it makes is not that great so I am just setting these up as kind of two sides of the same coin in a way, that if you are able to disprove equilibrium then you say, well, I have disproved equilibrium, therefore it must be disequilibrium

- Q. Okay. But in your studies and in the forensic field here in the databases equilibrium has never been proven.
- A. It hasn't been proven but there has been no evidence to disprove it. Let me put it that way. That in fact --
- Q. I agree a hundred per cent.
- A. Okay.
- Q. But there is evidence to suspect that disequilibrium might be there.
- A. There is evidence that these bin frequencies are different. It is not an automatic or even natural corollary of that that in fact there would be some kind of correlation between what is happening at one locus and another. It doesn't necessarily follow.
- Q. But within the Canadian Indians has it -- again disequilibrium has not been proven or it has?
- A. There has not been any demonstration that there is disequilibrium there.

- Q. But there is much stronger doubts within the Canadian Indians than there is in the Canadian Caucasians is there?
- A. Well, in the Canadian Indian you are showing differences between populations. That is not the equivalent of a disequilibrium. I would make the point again that there are two quite separate ideas in the sense that if you have this population subdivision there can well be -- and it cannot be taken as evidence of disequilibrium.
- Q. It just doesn't go that where you prove one, you can't assume the other. But it makes the other more probable.
- A. I would say even that -- I don't think there is any connection between the two ideas. Between the fact that you can have different bin frequencies in two areas has nothing necessarily to do with disequilibrium being present in each of those.
- Q. Am I to understand then basically what you are saying is, okay, Canadian Indians, yes, we have two different sub-populations. We can't throw them out into one pile.
- A. Right.
- Q. But within the two sub-populations there is random selection. Is that what you are saying?
- A. There is a random mixture of alleles, yes. Now, as you pointed out and where you would find linkage disequilibrium you would expect is if these really were separate populations and behaving the way we have just been describing, but suppose you were blind to that and you just assumed it was one big homogeneous unit

and it really wasn't, and you did the calculations based on that, there would then come out to be an apparent disequilibrium. Because of the combinations present in one place would be randomly behaving but they would be the particular combinations at that place, and in the other place that had different frequencies they could have their own combinations. If you had put those together they would appear as though there were perhaps some linkage disequilibrium even though within each of them separately there wasn't. I know it gets into somewhat --

- Q. -- a mind-boggling situation.
- A. Well, it is -- abstract idea, and unless you have thought about this and seen it in some actual cases it is sometimes difficult to convey. But that is one of the problems in that if there really is this substructure that is present and you are not aware of it or you treat the data as though there weren't, you can see these effects of the substructure being reflected in the deviations from Hardy-Weinberg equilibrium, the appearance of linkage disequilibrium and the fact that allele frequencies can vary from one sub-population to the next.
- Q. Okay. If you were doing your studies in a population to find out if there was linkage disequilibrium -- if I take my time with it I am okay -- what might be the first indication that you would come across for you to suspect that, gee, maybe there is?
- A. Well, the very first thing and the simplest thing to test would be these bin frequencies. If the bin frequencies were the same in the two places and there

was no evidence of heterogeneity that way or population substructuring that way, I would say, well, we could still look for disequilibrium but we are less likely to find it. If you found some differences in gene frequencies and bin frequencies in the two places then you would say, hey, maybe we should pursue this further and maybe there will be some disequilibrium. Maybe there will be some deviations from Hardy-Weinberg equilibrium.

Q. Like you did with the Canadian Indians. That situation.

A. That's right. In the case of the Caucasian populations we found no evidence of that when we compared the bin frequencies so we therefore said, well, -- and we are still going to pursue and look for linkage disequilibrium in those because we need bigger samples if we are going to continue to look for that. In the preliminary test that I did, which I admitted were not terribly strong, there was no evidence of strong disequilibrium being present. There might be some weak linkage disequilibrium, but it was below our power to resolve and see it.

Q. Fine. In a bin frequency/^{as you say}your first indication might be in different populations of the bin frequency -- the difference in that. How much difference in bin frequency would be necessary to clue you in that maybe I should study further? That it is significant. How much difference is needed to be significant?

A. The answer to that unfortunately depends on the size of your sample. The larger the sample the less the difference has to be before you can say it is

significant. I could give you an analogy in terms of tossing coins, if you will. It may not be appropriate in a court, Your Lordship.

Q. That's fine. Try it.

A. But, for example, if you had a coin and you weren't sure that it had a 50/50 chance of being heads and tails and I tossed it ten times and I got seven heads and three tails, well, I think just our experience would be that, well, sometimes a real balanced coin is going to give you that type of a result. But if I flipped it a thousand times and I got 700 heads and 300 tails, that would be statistically significant. That would say that that coin does not have an equal probability of coming up heads and tails, so it is an example of how the sample size enters into trying to make this decision as to whether the two things are different or not. Whether you try to say the probability of 50/50 or is it 70/30 or 60/40 depends how big a sample. The larger the sample size you have the more resolution you have to pick up smaller differences. If I had a large enough sample I could say, well, this coin has a probability of .551 or something of turning up a head and really not .50, but I would need a very large sample to do that. In the size samples that we have typically in the R.C.M.P. database -- I haven't actually done these calculations so I am going to wing this off my head a little bit -- but I would say you would need a difference in a bin of on the order of four or five per cent in order for that to be statistically significantly different. The other complication in

calling these bins different is the fact that you are not looking at single bins at a time. You basically have this spectrum of bins. Perhaps 20, 25, whatever, bins like that, and you have another histogram of 25 bins. Some of them are going to be up and some are going to be down. So you don't go about it just comparing individual bins like that. You have to compare the whole distributions like that. Individual ones can vary quite a bit and still the overall profile not be significantly different from one another. They can be just random like that because you have so many categories that can go up and down like this that you can't just look at them one at a time. They are all kind of yoked together because if one goes up then another one has to have gone down someplace. You can't look at these in isolation and the statistical tests that I have done, the likelihood ratios test, the contingency table chi-square test, or whatever, look at the whole spectrum concurrently in the two distributions like that, and allow you then to make decisions about whether these two distributions could have come from the same sample. That is could they -- what you test is whether those two distributions that you observe really could have been taken from the same population, and that is the test that you make.

- Q. That is why you said, like, in your histogram it would probably balance out in the end anyway once you go over your four or five or --
- A. Some of them go up. Some of them go down. That is what you expect. There can be extreme situations where you can imagine two histograms and one having a

big peak over here and another one having a big peak over there. That is unlikely. We have never seen anything comparable to that in any Caucasian populations, but if you were comparing some of the Canadian Indian populations to Caucasian populations, you do find that. You find that some parts of that profile would be quite high where in the Caucasian population it is quite low. If you were to calculate the frequencies of one based on the frequencies of the other you would get very different probability values.

- Q. Okay. I am a little confused or at least I want you to educate me as to if for each bin the frequency -- with the size of the R.C.M.P. database -- if you would need a variation of four or five per cent in the bin frequency, why is it that after you use the product rule for each bin, should you not be within the four or five per cent if you are calculating different databases?
- A. I am not sure I understand the question.
- Q. Well, you said that -- as I understood you to say that the bin frequencies, if there is a variation of more than four or five per cent, then they are statistically significant. A statistical significant difference. That is for one bin. Okay. And you are going to multiply all these bins to get your one in millions. Okay. Should that four or five per cent statistical difference hold true for your end product after you use the product rule?
- A. That is that the two net results should be within four or five per cent in order for them to --

Q. Yes. Otherwise there would be --

A. -- be significantly different.

Q. -- significant difference.

A. Well, what happens -- even in the one calculation -- let's say I did the calculation based on the Canadian database and there could be some differences, when you multiply these through you get a number that in fact essentially sort of amplifies any difference at any one locus. That is, if there were differences at each of these loci, if something is four per cent different here and the second one is four per cent different there, that four per cent gets multiplied by a four per cent, by another four per cent, by another four per cent, and so it is as though the estimate that comes out at the end has a wider kind of margin of potential error, and that is why in the data that I presented if you look at the confidence interval at the end of those calculations, they are wider than the confidence interval for any of the individual loci. It is true. What happens -- it is in fact much greater than four or five per cent because that gets sort of amplified out as you concatenate one locus to another. Keep multiplying them through. It is as though, if you were to multiply four times five times five times five times five and you get a number, and instead you multiply six by six by six by six by six. You would see at the end that you are going to be much greater than just one integer different even though each of those -- you have gone from a five to a six. When you multiply five times five times five and you compare that to

six times six times six, each of those individuals are only different by one integer. By one-fifth. By twenty per cent. But at the end you will see that the net outcome of that multiplication is much greater than the individual differences of each of those loci. You get this amplifying effect and that is why that net end result has a wider error margin on it, has a wider area that we can't be sure of.

Q. So four per cent turns into a hundred per cent.

A. Yes.

Q. But does that rule out the theory that they are going to balance out in the end? Once you go over four or five probes or bands or locus that they don't balance out?

A. Well, in fact in many cases the actual estimate that comes out is very close to the original one. In many cases. But in fact the confidence limits on that are very much wider. The estimate itself comes out to be the same or very much the same. You sometimes get this cancelling out though not always and I wouldn't say that in every case you are going to get that cancelling out. In some cases you are not going to get that cancelling out and it would be quite a bit different, but again it is very likely to fall within that relatively wide band of the 95% confidence interval.

Q. 95 or 99?

A. 99 I used. Three standard errors. 99. Sorry.

Q. I believe you were mentioning how important it was that New Brunswick be representative of the general population database and since New Brunswick was

represented from the Kingston base blood bank -- 10% I understand of the -- no. It was 5% of the service personnel at Kingston were from New Brunswick and New Brunswick only represented 2.8% of the population?

- A. I think those were the figures if I recall.
- Q. Were you able to check as to how many people were from New Brunswick in the -- of the samples taken from Kingston?
- A. I haven't seen that information and I am not sure that that is available. I haven't seen that information so I don't know.
- Q. I am just concerned that if they just went to the blood bank in Kingston for the base and they just looked at the names, how would they know if any of them came from New Brunswick?
- A. They had origins of birth place. Because these were military personnel I believe they could go back to the records if they had to, and I am not sure whether they did or not. There will be a person testifying later this week or early next week, Dr. Fourney from the R.C.M.P., who is intimately more familiar with that aspect of it. I cannot say certainly to my own knowledge that we know that in that sample from the Canadian Forces Base that those numbers that I used for the Canadian Forces personnel stationed at that base really are the numbers that were present in the sample.
- Q. What would be the significance if none of them were from New Brunswick? That there is not one person from New Brunswick in the R.C.M.P. database.

- A. I would say from my studies and in comparing Caucasian populations that it wouldn't make any difference.
- Q. Because I believe Dr. Wayne testified that he could use -- the tests in this case he could run down and just use the database from the F.B.I. or from anybody and it would give you a true --
- A. It would give you a close enough forensic number I think for forensic purposes that wouldn't be statistically different from whether I use the Ottawa sample or I use the Vancouver sample or I use the Kingston sample or whether I use the Minnesota or Florida or Texas or F.B.I. composite sample.
- Q. Dr. Wayne I believe also testified that it wouldn't have mattered even if you use the databases contained in England or in Europe, but now that you have received this information about France you would have your doubts about that?
- A. I would want to look at the data, yes. I would want to look at the data and I wouldn't feel safe in saying that it wouldn't make any difference. I would want to look at the data.
- MR. WALSH: My Lord, if I may interject with an objection. I am not quite sure for the record whether Dr. Wayne actually said that you could use the database from each country. I do remember Dr. Wayne testifying with respect to allele frequencies or frequencies he has seen with respect to Caucasian populations in other countries. I am not so sure Dr. Wayne actually made a statement that you could actually use the database from any country and it wouldn't matter.

MR. FURLOTTE: I will give Mr. Walsh that. I am not sure he stated as such, but where Dr. Waye stated that there was no significant difference in the allele frequencies between Caucasians in North America or Caucasians in England or Europe.

Q. From what you have found from France you might tend to disagree with that would you?

A. I would tend to disagree with that. I think that -- that is what I had heard up until about a month ago when I was given some data and when I was given Shields' information, that I actually did the tests, and my understanding up to that point was that there was no difference. I believe now that if you looked at these five loci some of them would show differences in different national populations.

Q. I am at a disadvantage because I don't even have Dr. Shields' affidavit.

MR. WALSH: I will provide him a copy. I assumed he had it.

MR. FURLOTTE: No, I don't have it.

MR. WALSH: I will provide him one.

THE COURT: May I just be clarified here as to what this affidavit is? What reference was made by Dr. Carmody to that this morning?

MR. WALSH: Apparently what happened, My Lord, for the record, Mr. Furlotte, as is his right, he sent the information that we provided in the disclosure to Dr. Shields.

THE COURT: That was in the form of what? A brief of what your witnesses were going to --

MR. WALSH: No. Some of the data from the R.C.M.P. The

rebin data, distribution tables and things of that nature. -- for Dr. Shields to assess and apparently Dr. Shields was testifying in a murder case in New Hampshire or one of the northern states, New York, I am not quite sure, and in surrebuttal there was -- he apparently testified and then there was rebuttal evidence from a Dr. Budowle, a chief scientist with the F.B.I., and in order to try and get surrebuttal I understand Dr. Shields filed an affidavit in that murder case in the States to show why he should be allowed to get back on the stand again, and part of that affidavit included reference to studies that he did from data that he received from the R.C.M.P. and comparative data that he made to the F.B.I. database. As a result I became aware of it through the circulation in the scientific community that in fact Dr. Shields had made the data comparison between the R.C.M.P. and F.B.I. database and as a result Dr. Carmody was asked to comment on that particular data.

THE COURT: Have you got a copy of the affidavit?

MR. WALSH: I have a copy of the affidavit.

THE COURT: Can you give a copy to Dr. Carmody and -- Yes, well, I just wondered what this affidavit was all about. Now you are going to either -- have you got a copy for Mr. Furlotte?

MR. WALSH: Yes. Certainly I will get a copy.

MR. FURLOTTE: I will have a look at it during the break.

MR. WALSH: I just assumed he had it.

MR. FURLOTTE: What Dr. Shields used, for the record, My Lord, was exhibit Vd-64 for comparison purpose with the F.B.I. database.

THE COURT: Did he get his surrebuttal?

MR. WALSH: I don't know, My Lord.

MR. FURLOTTE: I don't mind the Crown being aware of this information beforehand because they would be allowed to call Dr. Carmody back in rebuttal after so this way we are saving time. That is why I couldn't care less. As long as the court is aware that this was our idea and not the Crown's comparisons.

Q. Dr. Carmody, if I was able to go out and do a study of the Newcastle area of New Brunswick where these crimes were committed to get a population database for that area -- Mr. Walsh probably knows the population much better than I do. What would it be, Jack? Five, six thousand?

MR. WALSH: I can't comment on that.

Q. Under ten thousand people anyway. How big a database would one need?

THE COURT: More than that. More than ten thousand. However, it is not material here. Twenty thousand.

MR. FURLOTTE: Well, if they only need 750 for Canada hopefully you need a much smaller one for a smaller area.

A. Well, indeed you could get by with a smaller one probably but it would depend on how precise you wanted your estimates of bin frequencies to be actually. That is strictly a question of sample size. The larger the sample size -- irrespective of the population that you are trying to sample you would need on the order of 100 or 200 to get precise enough estimates of the band frequencies, the bin frequencies, to make these calculations so that the error margins

would be small enough that you would feel confident that you weren't having too small a sample.

Q. If I was going to go out and do that, which there is no possible way, how much of a difference would have to be shown say for the Newcastle area so that you would not be able to use the database of the R.C.M.P. in calculating the frequencies? Could you give me any kind of an idea on that?

A. Umm hmm. Well, I would say that if you calculated the forensic probability of having a particular genotype at all of the four or five loci, depending on the specimen, if you came up with a number that was outside that -- and I would even use a smaller confidence level for that. I would use a 95% confidence level -- which means that the error margin would be a slight bit narrower, but I would say that if in your calculations you came up with a number that was outside that 95% region, I would say that that population was significantly different and you should use frequencies derived only from that population in that case.

Q. Are you talking about the end product or are you talking about the bin frequency?

A. I am talking about the end product. The forensic end product. Comparing the bin frequencies I can't give you a number for particular frequencies. I can say and couch it in terms of the statistical test that is done. I can tell you in terms of how large the test statistic, the chi-square, would have to be for the number of bins that you had in order for that to be significantly different.

- Q. Okay. The only thing, we have a problem here is because I want to go on the assumption that we don't have an accused person, we don't have a crime even committed. There is no evidence. There is no accused person. Just doing a study on whether it is Newcastle or whether it is some small community in Ontario or in Nova Scotia, say 5,000, 6,000 people, what kind of a difference would you need in there to cause you concern? To say, gee, we better not use the database of the R.C.M.P. to make these calculations.
- A. I would again say that you would have to do the calculations and come up with an average forensic implication drawn from that data to be able to say that you shouldn't use it, and to have that number come out to be outside the range that you would predict it should fall within from the R.C.M.P. data.
- Q. What if each bin for the different probes, if -- just say for example that the R.C.M.P. bin for each probe, that there are say ten bands. Everyone was 1 in 20.
- A. All right.
- Q. Okay. And if you did one for -- take for example the Newcastle area, and every one came 1 in 15. Would that be significant enough to say that we cannot use the R.C.M.P. database to draw any frequencies on any crimes committed in this community?
- A. I would say again that you would have to take the calculations further, and I couldn't say categorically and simply that just a difference of 1 in 20 and 1 in 15 for each of the band -- for each of the bins would be sufficient to raise a difference that would be

great enough in the net end calculation that you would not be able to use it. I think it comes down to the end calculation as to what your decision would be as to whether you would have to throw that data out or get some new data or calculate it on a different database rather than using the R.C.M.P. database.

- Q. Basically what you are telling me is there is no standards that we can set beforehand to make all these calculations? Just have to do it ad hoc?
- A. No. I am saying that in terms of the acceptance forensically one has to ask the pertinent questions -- forensically pertinent questions. If you are asking questions about the population in some theoretical way and you want to know whether the bin frequencies are significantly different in one place or another place, I can do that. In fact we have shown that they are different for some Canadian populations compared to the U.S. populations, and for the native aboriginal populations between two of them. That does not necessarily allow you to conclude that in fact doing the forensic calculations is going to mislead you if you use one set of bin frequencies or the other set of bin frequencies even though they could be different.
- Q. I think I am beginning to understand you now. There is a certain set of rules for the general population and populations genetics in principle but then if you are going to apply it to a forensic setting then they can be watered down so to speak.
- A. Well, I wouldn't say watered down. I would say they

have to be looked at, and in any statistical test you have to look at the particular test that you are interested in applying. Whether it is a question of linkage disequilibrium or meeting Hardy-Weinberg or bin frequencies or the forensic frequencies that get generated from the original database. All those have to be looked at in the light of the question that you are asking. If the question that you are asking on the bin frequency is different then we can do a statistical test. If we want to know whether the forensic frequency that you are going to generate from these two different databases are going to be different statistically I can do that too. But it means that you do a different calculation and your statistical test is different.

- Q. The only thing I am concerned with, Doctor, is if there is such a difference between two populations, is using the Hardy-Weinberg formula and product rule, is it now valid to use it on the other population and ignore it? That is my concern.
- A. Well, I guess I would answer your concern with the fact that at least within Canada there is no evidence in Caucasian populations that there is difference in bin frequencies, so just there we have no evidence to indicate that there would be any problem in using this data to do any further forensic calculations at any other place in Canada. So I think there we are getting into somewhat academic questions about, well, suppose these frequencies were like the frequencies in Florida, should we use them in that case? Well, I have shown that if you apply it in a particular case

it doesn't make a great deal of difference and there is no statistical difference if you use frequencies from Florida or from Minnesota or the F.B.I. composite frequencies in this particular case. I am not going to say that in some other case it wouldn't make a difference. It possibly could but it would have to be looked at in those instances on a case by case basis. However, in Canada there is no indication of this population subdivision. There is no indication for Canadian Caucasian populations that there is population substructuring. So I feel quite confident in using the database -- this relatively large database. In fact at the present time in terms of databases on this information it is the largest available in the world. There is no other database -- and I have looked at a lot of them now -- that is larger than the Canadian Caucasian database in terms of numbers of samples in that database.

Q. In saying that just brings back to what you stated on direct. There seems to be very little difference in data among populations throughout Canada.

A. That's correct.

Q. And you have compared the database from Vancouver was it?

A. Vancouver, Ottawa, and the Canadian Forces Base Kingston.

Q. Now, we know that Canadian Forces Bases, their service personnel come from all over Canada so you have a nice wide spread of the general population of Canada there. We also know that in Vancouver it is a relatively new city and you have a lot of people

from all over Canada who have moved to Vancouver. Is it safe to assume that?

A. I think it is. That is my general impression anyway.

Q. And we know that Ottawa again is another big city and there is a good chance that people in Ottawa come from all over the country so you got a nice wide spread population randomly selected, but when you get into a small community, you know, three, four, five, ten thousand people, people are generally not moving into these small communities. They are moving out. So you don't have such a nice random selection to do a database with. Is that a safe assumption?

A. Well, I would concede it is a concern among human population geneticists. On the other hand, I would also -- can speak about some theoretical studies that have been done to look at just this problem in terms of isolated populations, as to how much interchange between them it would take to keep them basically together genetically and you can show theoretically that it takes very little immigration-emigration in and out of these communities, and that is basically one immigrant and emigrant per generation, should be enough to keep those essentially genetically linked tight enough that they would not be drifting apart. That is the theoretical studies.

Q. I am concerned and probably wrongfully so, but when I see this database being gathered and statistics coming from it, that -- and the degree of efficiency or whatever -- I believe with Dr. Waye we used an example like public opinion polls. I am concerned, like, public opinion polls saying who is going to win

the next election, Liberals or Conservatives, and they do a nice poll all throughout Canada. Randomly selected throughout the whole country. And it would give you a general idea as to who is going to win the next election, is that right, whether the Liberals or Conservatives, but it sure as heck can't tell you who is going to win the seat in --

A. -- a particular riding.

Q. -- a small community in New Brunswick, some large city in Ontario or some small community in British Columbia. In other words, you say, well, the Conservatives are going to win 70% of the seats. That doesn't mean there is a 70% chance that the Conservatives are going to win the seat in, say, Newcastle, New Brunswick.

A. That's right.

Q. Can't draw the conclusion on that whatsoever can you?

A. In that case, no.

Q. Could you distinguish the difference there in using the R.C.M.P. database for forensic purposes?

A. Well, first of all, one of the most important differences in the type of data that we are using here is that unlike an opinion poll where at most you are expressing your opinion about three categories. Whether you are Liberal, Conservative, NDP or perhaps undecided. Maybe four categories. We have in these cases typically 20, 25 categories. That means that in fact you have what is called in statistics -- I hate to introduce these technical terms -- but a lot more degrees of freedom. A lot more places where you can have options in there. That is

very useful because it means that when you have that many more options there is going to be -- it is going to be more difficult for different areas to be significantly different from one another, because they could have some low frequencies that were changing but for all of these frequencies to change, it is very much more difficult for them to all change in a particular way than it is to have shifts between the four categories of Liberal, NDP, Conservative, don't know. Into those categories. So that is one of the things that are different. Furthermore we are dealing here with traits that are inherited. Traits that are not influenced by the environment. Traits that are not easily changed from one category to another by any process that we are aware of. So these traits are very stable in terms of their fixedness. You are going to pass on to your offspring and they to their offspring the genetic variants that you have and that your spouse has and so forth, so it is unlike inherited patterns or people changing their mind in the course of a week or so and depending on the news, on the economic news, on the general feeling in the country and so forth, and so you have categories first of all that there are very many of them, and they are much more stable. They are a physical thing that are not, you know, somebody changing their mind or hearing a particular newscast or war breaks out in a certain place and we reacted to that can influence their opinion about it. Not going to change that. So what is present is something

that is biologically determined, that has been determined historically, evolutionarily, that you don't have, you know, voluntary control over, as these opinion polls data you are talking about. I think those are the two main things that I see being different in those cases. But I agree that in the case of opinion polls, as we know, making a prediction nationally is very different from making predictions for individual ridings.

- Q. But when you scientists in the field of population genetics are looking for a substructure it is almost on comparison with a public opinion poll in relation to say elections, how they are going to go. You think there might be some small community out there that is off the norm.
- A. There is the same concern taken in terms of aggregating samples. That one has to be sure to try and pick up all of the local heterogeneity or substructure that might potentially be there and so you want to construct your sampling design -- when you sample any community that you try to draw statistical inferences about in such a way that you feel it is as representative of the entity you are trying to describe as you possibly can get it. That means in the case of Canada, for example in this case, you want to get them as widespread geographically as you can and you want to be able to convince yourself statistically that there is no significant difference from one region of the country to the next.
- Q. I will be getting on this later on but I believe there are some scientists who believe that population

genetic study for the purpose of forensics that they should be doing population genetic studies for each area of the country.

A. That's correct. There are people -- and I would not say that that is incorrect. Certainly there is evidence in the aboriginal population in Canada, there is evidence in black populations, Hispanic populations in the United States, that there are very many differences from one local area to another local area.

Q. Within blacks and within Hispanics.

A. Within blacks and within Hispanics. In the case of Hispanics they are Caucasian. They are classified as Caucasian biologically. It is strictly a linguistic category to put them into that category, Hispanic, obviously. There there are significant differences and so it is natural, and I would support any proposal, to do further studies on a local scale, just to make absolutely and nail-down tight, the fact of whether there is local enough variation that we have to worry about the forensic implications or whether there isn't. In being a scientist I want to see the evidence. I don't like just going by what people's feelings are.

Q. Just one curious question before I go on to the rest. We know now, at least we think, that somehow the people in France are quite different than the people in Canada.

A. Well, the two loci that I have information on their probabilities of the particular genotypes that I was looking at in this case were more frequent than they were in the Canadian Caucasian population.

- Q. Where would you put a person who was of mixed parentage? Say his father from French, from France, and his mother was Spanish or something like that or Indian? How would you deal with these people?
- A. Well, generally forensically what you are dealing with in a situation is that you are trying to look at a certain community. You would want to know in a certain community. You are not looking specifically at an individual as such but you have a forensic specimen. The question forensically is: what is the probability that somebody in this community could have contributed that particular forensic specimen? Typically in those cases you don't know anything outside of the fact that the person contributing that specimen is likely to be residing in that local community or perhaps in Canada, if you want to look at it that way. So in those cases the proper statistical reference point is to use -- if you think that they were Caucasians, is to use the Caucasian database. If you had some prior knowledge that perhaps they were aboriginal, perhaps the use of aboriginal databases or to try and obtain a local aboriginal database if that was the case, or to look at the frequencies of the calculations based on all of these different databases and see how they compare. In almost every case, because of the way we are able to apply this technique and because of the amount of variation and difference between individuals, in almost every case, no matter what database you use, you are going to come up with very infrequent numbers. You are going to come up with very small numbers. You

are going to come up with rare events. Rare possibilities. And whether you come up with a number that says it is one in a million, one in 25 million, I think that that is telling me that it is a very rare occurrence that you are going to find somebody with that particular virtually unique genotype for those things we are able to get a snapshot of genetically.

Q. That is why you say there is no significant difference whenever -- depending on which database you use even though there is a significant difference in the bin frequencies, the end product there is no significant difference. So therefor it is valid.

A. In the cases that I have run that is true.

Q. The thing I have a problem about that with, Dr. Carmody, is are you sure you are not using the numbers to support the theory rather than using a valid scientific theory in order to obtain the numbers? Are you putting the cart before the horse?

A. Well, I would say that I am testing the theory in the sense that the theory says that there -- it makes a prediction. It says there won't be any differences. I say I am going to test that. I see if there are differences. If there are differences then it allows me to throw out the theory. But I find that there are no differences, so at the present time I accept the theory.

Q. But I am concerned that you are ignoring the principles upon which the theory was found and valid.

A. I am not sure that / ^{there} is a theory in the sense that --

Q. Well, the scientific theory or the validity of the

Hardy-Weinberg formula and the validity of the product rule. There has got to be a theory behind it.

A. Yes.

Q. And it has got to be based on factual empirical data.

A. But finding that these -- the result of these calculations based on a different database gives you the same number is not really testing Hardy-Weinberg or the product rule. It is testing to see whether in fact the bin frequencies that you did observe and that you were able to show were different in some cases don't result in any difference when you do the further calculations. It is not really testing the theories about product rules and linkage disequilibrium. It is conceivable that in the Canadian database that there may be some weak linkage disequilibrium or some slight deviations of Hardy-Weinberg equilibrium that would be below our limits of statistical testing right now based on the size samples that we have. We don't know that and so we say that with the tests that we have we don't see any significant statistical difference. Let's accept that fact, do the calculations, do the calculations on other databases, see whether in fact they give statistically similar results, statistically indistinguishable results, and assume that it is okay.

Q. I realize that, but you see my problem is I get hung up in this stupid case law that I read and I believe everything I read. What I have been reading is we have experts, especially for the police department, that is coming to town and coming to court and they

are justifying their use of the Hardy-Weinberg formula and the product rule in that the data that they are going out and collecting, it is within the Hardy-Weinberg equilibrium and it is also within linkage equilibrium, and if either one fails then you cannot use the product rule. It is scientifically invalid to use the product rule.

- A. Even if you found that there were some deviations there are ways for correcting if you know what those deviations are. In the case of the Hardy-Weinberg, the exhibit that was put in earlier -- the publication of Devlin, Risch and Roeder -- showed statistically that there was no evidence for excess homozygosity which would mean that you had deviation from Hardy-Weinberg equilibrium. Furthermore, the test that I did on the data, I tested specifically to see if there was any strong indication of deviations from Hardy-Weinberg equilibrium in the Canadian Caucasian database. I did not find any. I did some, again admittedly weak tests, but tests that would pick up any strong deviations from linkage disequilibrium and I was not able to find them in any of the tests I did on the Canadian Caucasian database. If it turns out with larger samples, looking at more populations, doing more refined statistical tests, that indeed there are some slight deviations from the ideal of linkage equilibrium and some small deviations -- real deviations -- from Hardy-Weinberg equilibrium, we will know what they are. We are able to correct for those, and they -- if they are that small -- are going to be numbers that are going to be down in

the third and fourth decimal place and are not significantly going to impact on the calculations that we derive from those numbers.

Q. You got to understand though why I might be a little skeptical as a defence lawyer or for any other scientist out there. When I first read case law and these cases were coming to court, the experts were coming and saying 'there is no such thing as band shifting'. Once band shifting was proven the defendants were trying to prove band shifting to show it wasn't reliable. The experts are saying there is no such thing as band shifting. Now they admit band shifting but we can correct it, and now there is no such thing as deviation from Hardy-Weinberg or linkage equilibrium, but now we -- if we do bring in evidence then you say, well, yes, there is but we can correct it.

MR. WALSH: My Lord ---

Q. Is that a proper --

MR. WALSH: My Lord -- objection. He is making a statement --

Q. Is that a proper procedure?

THE COURT: Well, just a minute.

MR. WALSH: He is making a statement and in fact the statement even if not relevant to -- he is straying away from the field of expertise, but he is making a statement with respect to experts saying there is band shifting and now there isn't band shifting; there is deviation and there isn't. I object on that particular basis. He hasn't asked a question but additionally it is somewhat misleading because if he

is reading the same case law I am reading the whole issue of deviation from, and being able to correct from, certain things like Hardy-Weinberg equilibrium, one of the very first cases Mr. Furlotte would have read the judge actually did talk about deviations from and correction factors from Hardy-Weinberg, and I refer you to New York and Wesley so even though he is making a statement I object to, he is also making a patently incorrect statement.

THE COURT: Okay. Well, that has been noted. Anyway I will allow Mr. Furlotte to make his premise --

MR. FURLOTTE: I believe New York and Wesley is being retried.

THE COURT: Well, you are asking for this witness's opinion on --

MR. FURLOTTE: Well, I just told him my concerns and, as defence lawyers, concerns, and that some other scientists in his own field, what their concerns are.

THE COURT: Well, for Mr. Walsh's --

MR. FURLOTTE: And is that a legitimate concern.

THE COURT: Yes. But for Mr. Walsh's benefit first I want to point out that you will have a chance to re-examine Mr. Walsh, and if you want to bring out discrepancies of this nature, you can in your re-examination.

MR. WALSH: No, My Lord, I just --

THE COURT: I get your point.

MR. WALSH: My initial objection was he wasn't asking a question and my objection went a little farther than that. Even the statement he was making to the court was in my humble opinion was incorrect.

THE COURT: All right. Go ahead.

Q. Now, if I can remember what I asked you.

THE COURT: The witness remembers.

A. I was going to say, My Lord, would you rephrase the question?

Q. Do you understand the point that I was trying to make? You know, for us to be concerned about ad hoc corrections I suppose once the theory is proven wrong.

A. Well, this is a developing technology and as in all of science we are learning things as we go along. I think in the case of band shifting that is easily detected and I don't think that is a big problem. I think even in this case of population subdivision I don't see it as a big problem particularly in relation to the R.C.M.P. database. I would have to say for the court that I have not had access to databases other than Caucasian databases in North America and this one in France, and as my colleagues could testify in the forensic labs at the R.C.M.P., I am chomping at the bit to get at any other databases that could be made available to me. We won't know the final answers until we know the final answers and that is the very nature and process of science, is all I can say. If you see that as a process of going along and making ad hoc corrections and then retracking and changing your mind about things, I would defend that as science as usual.

Q. Intriguing problem, is it not, Dr. Carmody?

A. I mean there are always new problems and to every answer there are generated more problems.

THE COURT: In other words, science never has final answers.

- A. That's correct. They are always approximate. They are always tentative. They are always given the data that is available and they are going to be refined further.
- Q. Every time we poke a hole in the dike you stick your finger in it.

THE COURT: Why not break off there?

MR. FURLOTTE: Sounds like a good idea.

(RECESS: 2:50 - 3:15)

- Q. Dr. Carmody, I recall when you were mentioning about New Brunswick being over-represented at the Kingston base and you mentioned that that was good here in this particular case because of the over-representation, and you said I believe British Columbia was under-represented.
- A. In that particular -- Kingston base.
- Q. And that was bad for B.C.
- A. Right.
- Q. But yet when I asked you a few minutes before we recessed, I said, well, what significance would it be if New Brunswick had no representation whatsoever in the database, and you said, none whatsoever. So I wonder how it can be advantageous on one hand but not disadvantageous on the other hand?
- A. Well, I am saying that after having looked at the databases, I was making the judgment about the representation in the Kingston database somewhat artificially if I had not analyzed that data yet. That is, just saying where that sample came from sort of prior to doing any analysis on it. Now that I have

done the analysis on all three databases and all three samples, I see no problem even if there weren't representation from New Brunswick. So I am saying it from a different perspective so to speak when I am making that call.

- Q. Okay. I would look at it ignorantly and I would probably say, well, it is the numbers that are influencing your decision rather than the principles upon which you use to justify the numbers.
- A. Well, that is partly true but that is I think always the case. One has to make some decision based on empirical evidence, and I am using the empirical evidence to inform my decision rather than coming upon it in some abstract way from prior principles.
- Q. As I noted in your direct examination when you were talking about the representative samples, you stated that if the sample represents accurately the actual population you are trying to make references about, so again you found at that time that you should have references about the actual population of which the person fits into.
- A. Yes, but I am saying that after we look at the data it wouldn't make any difference now. After we have looked at the data.
- Q. Yes. I understand that is what you were saying. You mentioned you used the -- to show that there was no statistical difference in the three areas which you checked, Vancouver, Kingston base, and Ottawa, you use the chi-square?
- A. It is called chi-square and there is also a likelihood ratio test that is basically using a chi-square in a slightly different way but I have done both tests.

- Q. Now, when you use that test, say, with the difference in the databases, the R.C.M.P. and the F.B.I. and the one in Minnesota is it?
- A. Minnesota, Florida, Fort Worth.
- Q. How did that chi-square test perform there?
- A. For some of the loci, particularly for the D2 locus and the D10 locus, there were differences for the R.C.M.P. versus Florida versus Texas. Not versus Minnesota so there were, in the bin frequencies, differences, statistically significant differences, in the bin frequencies when you compared the Caucasian R.C.M.P. data against those for two of the loci that I was using in this case. Correct that. Three. The D17 as well. D2, D10 and D17.
- Q. So it would appear that our political boundaries do make a difference.
- A. Well, they seem to expand to Minnesota anyway. There is potentially some difference there. I don't know how much in the case of Fort Worth and Dade County, Florida, how much of that might be some mixture with Hispanic populations and so forth in those areas. It is difficult to say without further sampling.
- Q. Would you admit that if we did, say, check out -- and in this particular case -- the Newcastle area or some other small community in Canada, that you might see as much difference between the databases from Vancouver, Ottawa and Kingston? Just as much difference say between those and a small community down this end of the country as you did see between the R.C.M.P. database and the database from Florida or the F.B.I.?

- A. I wouldn't expect as much difference just based on the Canadian three samples that I have looked at, but I couldn't say absolutely. I mean I wouldn't be as certain. I would say, you know, statisticians can express this in confidence levels and I could say something like I would be 90% sure that it wouldn't make a difference, but I wouldn't have 99% confidence in saying that.
- Q. There is room for error in that.
- A. Yes.
- Q. You have voiced your concern about -- you say when there is fewer than five events in a database -- in a bin then that is not sufficient enough to draw any reliable conclusions so you amalgamate it with the next highest bin.
- A. The adjacent bin, yes.
- Q. Up or down as to which one has the highest amount of events or --
- A. Well, you just amalgamate the two so that if there are less than five in one of them, then the net joint combined bin will have more than five. And you keep amalgamating them. For example, if you had a bin with one in it and the adjacent one had two and the adjacent to that had three, you would coalesce all three of those so you had a total of six in that coalesced bin.
- Q. Okay. Say if bin seven had four events, bin six had 20 and bin eight had 25, where would you put the four? In the bin 25?
- A. Right. In those cases, you put it in the smaller of the two if you are going to coalesce them like that, yes.
- Q. In that case you are not being conservative with the -- that is not a conservative measure.

- A. No, it is. It is because you are increasing the frequency.
- Q. Yes. You are increasing it to -- you are putting it in with the smaller bin where I have been told that the larger the bin the more conservative it is.
- A. The larger the bin the more conservative in that the frequency is greater so we consider a greater frequency to be more conservative, more in favour of having that -- seeing that in the population.
- Q. So there is the distinct difference. You know darned well it doesn't fit in either bin but you are going to be conservative and put it into a larger bin.
- A. That's right.
- Q. Because such a small sample is not reliable --
- A. That's right.
- Q. -- for drawing --
- A. For drawing these types of conclusions, yes.
- Q. How does that compare when -- do you know how many hyper-variable regions there are in human DNA?
- A. We don't know fully.
- Q. It is in the hundreds? Three hundred? Maybe more?
- A. I would say -- I don't really know. I know that there are at least dozens and there may turn out to be hundreds or several hundred but I couldn't say beyond that. They haven't all been uncovered yet is the problem and you are trying to make an inference from what you do know to parts of the genome that we don't know at all. If I can give you a feeling for what a small part of the genome these probes are actually able to see and we are actually sampling by looking at these sites, I once did a calculation -- and I often

use this in classes to impress my students with how much DNA is actually present in each one of our cells. DNA, as you have been told, is a very thin structure, but if you represented that DNA by a quarter inch rope -- quarter of an inch. Expanded it up -- it would reach, and I did the calculation, from Ottawa to Calgary and back. What we are looking at with these probes are regions that are about two meters in length each and we are looking at five two-meter sections. We are looking at a total of ten meters on a piece of DNA that goes from Ottawa to Calgary and back. That is the small size that we are actually able to sample in all of that DNA so it is really a very, very small sample. You consider that amount in the total distance from Ottawa to Calgary and back.

Q. I am just wondering if sampling such a small size of the whole, does that really give you a good picture of frequencies?

A. Well, indeed we know --

Q. No, but since five events -- something under five events is not sufficient to draw your basis on out of 200, why are three events, probes, or six bands sufficient enough when all that out there is available?

A. Because it is a different question here. We are taking the exact same sample from every individual. Every individual that we look at we know has exactly this same region that will be, as we call it, hyper-variable. We know that every individual has it. There may be the rare individual that has a deletion or something that doesn't have it but they haven't turned

up yet in the samples that we have. So we are sampling the identical thing in every individual. In fact two copies of identical thing in every individual, That is different from saying what variant is present at those sites. I mean it is a different sort of statistical question that the variants that we are looking at could be there zero times or they could be there ten thousand times, and if we only see a few of them when we have taken a sample of a thousand, that is a different statistical situation than when we are looking at the same thing in every known sample that we have.

- Q. And now we are dealing roughly what, one out of a hundred variables in a bin size or --
- A. In a size that we are talking about for each locus, each probe, we are looking at something on the order of twenty different size categories in there. So I am saying if you find only one of those particular 20 present in a sample of 1,000 less than 5% of the time, that is a very rare sample that you are looking at, and that is different from saying that these are sites that we know everybody has, that we are looking at specifically that we know show this great variation in them, and we are not making an inference about the rest of the DNA that we are not able to sample from these five or ten sites that we are looking at, and try to extrapolate that to all of the rest of that length of rope so to speak.
- Q. But you are drawing the conclusion and the assumption that if you check out for five regions and they match in five regions or loci, that they are definitely going to

match throughout the remaining. They will be identical throughout the remaining 300 or whatever.

- A. Not necessarily. Each one of these behaves quite independently of another. There in fact are some of these regions, these tandem repetitive regions -- one of which the R.C.M.P. uses -- that is the same, is identical, in every single human being. It is called a monomorphic locus and that locus in fact is not only constant in the same and every human, it is present and the same in all higher primates. So there are going to be some of these regions that are going to have that property. That are going to be identical in everybody. Other will be hyper-variables -- show a lot of variation in the lengths. As these have been chosen to be, there will be others that will be intermediate. There will be others that exist in four different size categories. There will be others that exist in two categories, eight categories, whatever. Some of which may be even more variable than this may exist in thousands of categories.
- Q. But to get my understanding straight, that if we have a known sample and an unknown sample and it matches say in four or five probes, there is a high degree of probability that if you continued matching them in every polymorphic locus, site, that you are going to get an exact match from here to eternity.
- A. That's correct. That's correct.
- Q. You say that most bins are about twice the size of the match window of imprecision that is being used.
- A. That's correct.
- Q. So that everybody in the same bin doesn't mean they all have the same size fragments.

- A. Not really if you got down to the molecular level, that's right.
- Q. But they are treated as if they all have the same size fragments.
- A. That's right. And that is a conservative decision actually because in fact by putting them together you are going to get a greater number than if you were actually able to look at a smaller size and be able to discriminate between things that you are lumping together.
- Q. Would it be safe to say if the system was so discreet that it could detect band sizes within a dozen base pairs? That there is a good chance that no bin would ever have more than five events?
- A. No. Not in a sample that was in the order of a thousand or 1,500 specimens. Indeed there is a technology coming on line where people are doing this and it is called a technique using polarized chain reaction where you can amplify up and you can get enough DNA from each of these regions to actual see the discreet and actual size of the fragment. In those cases for some of those loci you find that in fact there are let's say on the order of 18 different discreet actual sizes over the total span. Some of those in a sample of a thousand might be less than five, but most of them are going to actually occur with a frequency greater than five. You would have a frequency spectrum not too unlike what you get with this more imprecise technology that is available now.
- Q. For the R.C.M.P. or even the F.B.I., when they establish their database and they set up their bins and

they put them in, they basically what? They will take a DNA sample from a blood sample and they run that on the one gel and they measure it?

A. And they measure it.

Q. And they put it in.

A. They put it into a bin. Typically this is done not singly at a time but you have a gel where you run on the order of 20 or two dozen samples together. You have some standards on there so you know the actual calibration of the sizes of the pieces that you have run.

Q. If you were to run a single person's DNA on a single gel and you run the same DNA in two different lanes and -- same as like if you run it over the second time you might get a slight variation in the band sizings?

A. That's right. You do get a slight variation.

Q. But when you run them on a separate gel like they do for -- they only run one test for the purpose of forming their database -- say, your DNA is one little sample. They only run you once, but if they run you twice on two different gels the band sizing could be significant could it not?

A. Well, they could be different but that is taken into account by this match window of 5.2 per cent. That you never find a difference greater than that even using forensic specimens, and typically the difference even from gel to gel is more on the order of one or two per cent difference in the estimates.

Q. What would happen or what would be the cause of -- if you got the same person's DNA exceeding the match window in two different gels? What would that tell you or could that be possible?

A. It would be theoretically possible. I imagine it has happened at times. There are a number of mistakes one can make in running two gels. It could be that when you made up the second gel the wrong buffer was used or the resuspension of the DNA was done improperly, the concentration of agarose that was used to make up the gel. There are a number of technical points where you could account for a difference that would lead to a size difference in an identical sample being greater than 5.2 per cent.

Q. My concern is that is nice. It could exceed even five or six per cent could it not? Could be up as high as ten per cent?

A. It could exceed it by any amount you wanted to --

Q. -- imagine. Couldn't -- when the R.C.M.P. or F.B.I. is running their samples to establish their database, they only run the sample once so they never ever know what kind of errors they are committing and how broad they are.

A. Well, I think that -- and I don't want to put words in the mouth of a future witness, but Dr. Fourney from the R.C.M.P. is the person who could address that directly because as part of this program there is a quality assurance met where in fact the identical samples have been run again and again on different gels to establish exactly what the variation is, and so --

Q. That is how they establish their window.

A. That's right. That is to establish their window and to show that in fact when you run the same samples he can talk to you about -- and explicitly say how often you get this much deviation and how often you get that much deviation running the same samples, and from that

quality assurance -- that is a program that is in place -- it is very unlikely that even though you run new samples that are taken into the database just once, that there is likely to be that kind of error occurring.

- Q. If you run a sample once -- take here exhibit VD-45 for an example. Take lane A. You run that sample once and if your matching window is 5.2 per cent -- they know they have made mistakes up to that and they have averaged it out or that is their highest -- however they have established their window -- then you never know whether you should be up here by 5.2 per cent or you never know if this band should be down here by another 5.2 per cent; is that right?
- A. That would be the error but you would expect it to be the same for both bands. If they were going to shift they should shift in the same direction.
- Q. Both bands would shift in the same direction.
- A. Same direction.
- Q. But you never know which direction. If there is going to be an error on one gel, one test, you never know if they slowed down by 5.2 per cent or they speeded up by 5.2 per cent.
- A. Well, let me just make a minor correction. That you are talking about 2.6 per cent in either direction so is a total of 5.2, but to get to the main point of your question I think you don't know whether it is shifted down by 2.6 or up by 2.6. As much as that.
- Q. Right. As much as that.
- A. Yes.
- Q. So each band could be out by you say --

- A. -- 2.6 per cent, and they would have to be going in the same direction.
- Q. Yes. But when you just look at the band -- you have the band here. You realize that band could be up here by 2.6 per cent or it could be down here by 2.6 per cent.
- A. Yes.
- Q. So for each band there is a 5.2 per cent variation where that you don't know/it should lie.
- A. Well, you know --
- Q. -- or could lie.
- A. -- again, 99 per cent confidence that it lies within that width. That band width. That window.
- Q. My question is: if this band did not run at -- say, this one runs slow at 2.6 per cent and this one runs fast at 2.6 per cent, you don't know if that come from the same person though.
- A. That would lead to a false exclusion.
- Q. That would lead to a false exclusion. Right.
Now, if you can calculate the false exclusions on that basis, can you not also calculate false inclusions? In other words, if two bands are straight across like these two, if this one should be up here -- could be up here -- we don't know -- it could be up there or it could be down there. We haven't got a clue do we?
- A. Well, we know it falls in there but we also know that that should hold true of the other band in that same lane. That is, that the shifting, if there is some kind of biophysical phenomenon causing that shifting in that way, that it should apply to all of the bands in that lane and it should not only apply to the bands in the

lane when you probe it with this probe, but then when you probe that same gel with another probe, the band should also have been shifted in that other probe.

Q. In the same direction.

A. In the same direction. That is, if -- we are imagining in this model of this system that in fact this biophysical phenomenon that is happening is going to affect all the DNA in that lane.

Q. Okay. If for some reason -- if this is a pristine sample in 'B' and this is an evidentiary sample in lane 'C' which may shift/^{due}to either degradation or contamination, we really don't know if the shift is going to be a positive one or a negative one.

A. That's correct. In general though I must say my understanding of what happens in degradation is that the band would not shift because all of the DNA molecules that make up that band, if they get degraded, some of them would be chewed down very small and you would get sort of a smear in the gel in that case. You wouldn't get a sort of nice neat band, all of which shifting precisely and exactly by the same amount. Although there is some evidence that that can happen and the actual details at a molecular level aren't understood as to how you can get that band shifting, but it should affect all the DNA and to call it as a band shift, you should -- and you have to be able to show that at more than the single probe locus. You would have to be able to show that for all the probes. In general if you have two samples that really are biologically the same and one shifts, you are more likely to have a result that you would get an

exclusion of that individual and say it was not a match. Then by having ^{an} individual that was really not a match become a match because there would not only have to become a match at that particular probe site but it would have to become a match at all of the other probe sites as well. That is, the shifting and the combinations would be such that they would all, if they were really different, have to be shifted the identical amount and make them a match, and the probability of that happening for probe one, probe two, probe three, probe four and probe five, becomes banishingly small in terms of probability. That is, the technique I am saying is very much more likely to generate a false exclusion than to generate a false inclusion.

- Q. But it is not uncommon for individuals to have common band sharing in a probe so there could be a legitimate band matching say in two or three probes.
- A. Yes.
- Q. And then maybe for the fourth probe and the fifth probe band shifting could cause them to look like a match when they are actually an exclusion or inconclusive.
- A. No. Because you would expect then on the sites where you are saying they are a match, they should have been shifted as well.
- Q. They should have been shifted.
- A. And they should have been shifted out of where they were.
- Q. That is if you get band shifting in the same direction all the time.

A. Well -- and you would expect to get that. I mean band shifting has got to go either to make it faster or slower. That you don't expect band shifting to make the top band move faster and the lower band to move slower. I mean there is no biophysical explanation that would explain how you can get one band shifting faster and one band shifting slower.

MR. WALSH: My Lord, at this point I have an objection.

Q. You basically have the same --

THE COURT: Excuse me, Mr. Furlotte.

MR. WALSH: My Lord, at this time -- I have allowed him to proceed up this avenue without objecting. Dr. Carmody has not hesitated to provide answers. However, I think Dr. Carmody -- my objection is this. Dr. Carmody has been declared an expert in the field of population genetics. Mr. Furlotte now is delving into the areas of the molecular biology associated with the RFLP technique, and although Dr. Carmody certainly is familiar with that particular technique, we have not chosen to get Dr. Carmody into that field. It would seem to me that Mr. Furlotte is off the track for which Dr. Carmody has been declared an expert, although he has -- the witness has certainly --

MR. FURLOTTE: He has the qualifications to be an expert.

MR. WALSH: The point is, My Lord, that we have people who are going to be declared expert for the purpose of describing the technique and the matchings, etc. For Dr. Carmody's purpose, although he has been kind enough to answer the questions of Mr. Furlotte, it is not within where he has actually been declared an expert today.

THE COURT: Well, let's let Mr. Furlotte keep on going here until Dr. Carmody says, 'Look, I don't know anything about that field'. He has been willing so far. Go ahead, Mr. Furlotte. I forget just where we were.

Q. We were mentioning about band shifting possibly in different directions.

A. Right.

Q. You were saying that the band shifting for the two bands would be consistent or prorated, whatever --

A. That's right. It would be proportional but in the same direction.

Q. Right. One would not move faster than the other.

A. That's right. It would be in proportion. They don't run down the gel with an even speed because the lowermost one is a smaller molecular weight and the speed with which they move is a more complex function and it is actually moving more quickly as it goes down the gel.

Q. The lower one should shift more than the top one.

A. That's right. It should be in proportion to its size. In fact it moves as the logarithm of the size and so forth.

Q. Now, you mentioned the purpose of setting up your database is to see what the frequency is of one particular band size within the general population.

A. Yes.

Q. And like if you took 'B' and you calculate what the frequency of that is in the general population, you can take the second band in 'B' and calculate what the frequency of that is, and then you multiply the frequencies of the two to get --

- A. Yes. In fact you have to add -- multiply that by two again.
- Q. By two again.
- A. By two again.
- Q. Right, which is what Dr. Wayne testified to but you didn't mention that in your direct.
- A. Well, I thought it would be a complication. I can easily explain that.
- Q. Oh, no complication. Just that you forgot to mention it. That is fine. Just for my own purposes and maybe for the court, we would like to know exactly which way it is. The end product would not change, if you didn't multiply it by two, very much.
- A. No. Well, because if I didn't do that --
- Q. You had a sufficient number of probes.
- A. If I do that at each locus it is going to be 2^5 actually and that is not an insignificant change. Something like 32.
- Q. If I took my DNA and I wanted to analyze it and go to the population base and see how many people would share, say -- we will call these different probes rather than --
- A. Okay.
- Q. We will say this is one probe and one probe and this is all my DNA.
- A. All right.
- Q. And I wanted to see how many people out there shared this band with me, that band with me and maybe this band with me and then that band with me. (indicating)
- A. Right.
- Q. Would that be possible?
- A. Yes.

Q. What would I do then?

A. Well, you could look --

Q. For lane 'A' or probe 'A' we have the --

A. -- the frequency of the top one.

Q. The top one.

A. Okay. And the frequency of the bottom one.

Q. And then multiply.

A. Multiply by two to get the $2pq$, because you see when you have two bands in a lane the reason you have to multiply it by two is that the top one could come from your mother and the bottom one from your father or the reverse, and each of those possibilities has a probability of the 'p' of one -- the probability of one times the probability of the other, and it could happen either way because we can't tell which of those you got from your mother or father, so we have to multiply by two. If you have just a single band in the lane and there is another one, but if you are just asking questions about that you just use the probability of that in your database. So if we had two in a lane we have to do $2pq$ -- I use $2pq$. That is a population genetics term. It is really probability of one times the probability of another times two. If you have another band in another lane that you want to ask, well --

Q. Maybe this one here.

A. That's right. You take the probability of that and multiply that times the already calculated two times p_1 times p_2 .

Q. And then if there is another one down here on this probe --

A. In a second lane you use that alone. That's right.

Q. Then that would give me the frequency of how many people out there are going to share these four bands with me.

A. That's right.

THE COURT: Perhaps the record should show that this discussion has been in respect of exhibit VD-45.

MR. FURLOTTE: VD-45, and rather than lanes A, B and C, we can name them for the purpose, probe A, B, and C.

Q. Would that be correct, Doctor?

A. That would be correct.

Q. If I shared bands at this frequency -- say, four bands out of eight. There is eight there.

A. Right.

Q. 50 per cent. One the four probes that we run, could we also assume that throughout the other hypervariable regions that I might -- the person who I shared that with, I might also -- if there is a person out there that would share that many bands with me and these particular ones -- would it be safe to assume that if they run my profile and their profile with other probes that we might share 50 per cent bands in the rest of them that have not been conducted?

A. No, there wouldn't. It would be whatever the probabilities are in the database. Because if you match two the only thing you can say about it is on the basis of those two. Or in this case if it is four, you can only say it on the basis of those four. The other -- if there is real independence/^{which}is what we have tested for, then in fact the probability of having matches on others is not changed by that knowledge.

Q. It is not changed.

A. No.

Q. But if you were to generate four other probes -- which maybe we don't have at this time but if you would assess four other loci and have those probes and run -- say, with myself and found somebody else out there who shared these four bands --

A. Right.

Q. -- but not the other four, there is a chance that we are going to share bands on the other four probes also; is that what is called equilibrium?

A. There is a chance, but if there is linkage equilibrium and Hardy-Weinberg equilibrium as well, then the probability of your sharing other bands is independent of your sharing these four that we are talking about, and will be strictly the product of the probabilities of those other bands occurring in the population.

Q. Now, if -- I would understand that if I was going to share four bands out of eight, there would be a good chance that maybe one of my brothers would be a better candidate to go looking.

A. Right. Exactly.

Q. And if that was the case, because we are related, there is a good chance if we went on to four other ones we would also share some of those too.

A. There is a chance that is higher than in the population but let me just say, in terms of two brothers matching at any particular site it is roughly one in four.

Q. 25 per cent.

A. That's right. So across each locus then it is one-quarter raised to the fifth power. One-quarter times itself five times.

- Q. Any way you can tell from the database that -- how many bands I might share in the five probes that the R.C.M.P. use, how many bands I might share with any individual out there?
- A. If I could look and if that were your --
- Q. In unrelated individuals.
- A. Yes. I could look through the database and say how frequently each of those bands occur in the database. I could say that and I could give you that figure.
- Q. The database doesn't -- can you get out of the database how many people, say, in any one of the probes, how many people share the same fragment lengths in each bin? Say there is a hundred people in a bin and seven have probed whatever -- how many of those hundred people share both bands?
- A. I can get that out of the database, yes, because the database has the two bands for every individual in that database and I can go back for each individual and say that this individual has two bands that would match these two, and it would say yes or no, and I could go along for each individual in the database and I could tell you how many individuals in that database have that identical pattern let's say in that 'A' lane, and it might be -- let's say in a sample of 700, perhaps there is 30 people that might share that pattern. I could tell you -- and my estimate would be, if you said, well, how frequently does that occur in the Canadian Caucasian population, I would say, just divide 30 by 750 or whatever the total in the --
- Q. Or to doublecheck to see if your product rule worked you could also take the individual frequencies of each

one, multiply them, see if it compares with your actual empirical observation.

A. -- number -- that's right. That's right.

Q. Is that ever done?

A. That is a statistical test and the problem with doing that -- that is done and I have done that -- basically locus -- taking two loci at a time. The problem with that is that even in lane 'A', let's say if there are 20 bins, that there are over 200 possible combinations of two bands. That is, you could have bin 1 with bin 2, bin 1 with bin 3, bin 2 with bin 3, bin 2 with bin 4 and so on. There, if you have 20 bins you do the algebra on that and it comes out to be that there are over 200 different combinations just at one site. So if you take a sample that is of 750 people there are going to be some of those bins that probably are not going to be represented in a sample of 750 since they are not all equally frequent. So the problem in doing that test in a rigorous way as you are suggesting, taking each individual bin and seeing how many there are in your database, is that you know already without having a much, much larger sample than is available now, that there are going to be some bins that are poorly represented and in fact most bins in that case are going to have less than five individuals in. If you have 200 bins and there are 700, even if they were equally distributed, you are going to have less than five. So all those estimates are going to be very unreliable and to do a statistical test like that you don't have enough data available to really rigorously test that. So in fact the test that I did

was a relatively crude one. I said, well, suppose we can't do it with all those 200 possibilities like that. Let's divide up all the bands into four categories and say whether they are below the 25 per cent point quartile, between 25 and 50 per cent, between 50 and 75, 75 and the rest. I then asked, into those four bins, how frequently do the two bands that occur in individuals in that database fall in those various bins, because I can predict what the expected should be in that case. That is, there should be predictable numbers that should fall because you are using a non-parametric kind of test. I don't mean to sound too -

Q. No, no. I understand.

A. Okay. So the test that I used to look for non-randomness was basically to put them into four bins, four equally sized bins, based on the database, and I tested to see whether in fact the percentage of time that the two bands fell within various bins met the expectation of randomness and they did. There was no deviation. So that is only going to pick up -- I was quick to concede that that is only going to pick up coarse correlations, but there have to be very strong correlations to not be able to meet that test. There have been some criticisms in the literature, and the reason I thought of that test was in fact people had suggested it -- statisticians had suggested in the literature that in fact it should be done, so I did it and it showed that in fact there was no deviation from what you expected.

Q. Did I understand you to say that the frequency of homozygotes in a database, Caucasians, would be roughly ten per cent?

- A. That is the numbers for some loci and some loci it is even less than that, but it is on that order. There have been a few cases where you get higher than that. I think perhaps D17 might be higher than that. I don't remember exactly but it is on that order.
- Q. How high would you have to go before it would cause concern?
- A. It is not a simple answer to that because you have to know what the expecteds are and the expected proportion will depend on the number of bins and the relative disproportion of the distribution in the bins. That is, if everything were evenly distributed across all bins you would expect to get more homozygotes -- sorry -- fewer homozygotes than if they were skewed in their distribution across those bins, but as you can imagine it, there are some bins that are very common. than just randomly -- you would expect you would get more homozygotes because you would get the same bin coming up in two bands --
- Q. Two bands.
- A. -- more frequently if that were the nature of the distribution. On the other hand, if they were evenly distributed that would be more rare in its occurrence. So the answer of how much homozygosity that you have to observe before you would see that it deviated from randomness would depend on the distribution, and I can't give an absolute value to that. The value that it would have to be would be the per cent you observed compared to the per cent you expected. The per cent expected is going to be generated by the particular profile of distribution through those bins.

- Q. I have been reading that the validity of the product rule has to show that the database, your samples, are randomly taken and of unrelated people. How does that -- I see people being -- for the database they have to be unrelated. Now, when you attempt to use the database to draw the frequencies for an accused person, say like Mr. Legere here, don't you also have to assume that there is nobody out there related to Mr. Legere to calculate the frequencies?
- A. That there is nobody in the database related to MR. Legere?
- Q. No. That that is nobody out there who is related to Mr. Legere who may have committed these offences. If you are going to use unrelated people for your database don't you have to use unrelated people for the known and unknown samples?
- A. I am not sure I can fully grasp the question but I don't see a problem. If indeed this is a true representation of the actual frequency of these in the population, it should be a true representation of Mr. Legere and his relatives and anything else. I don't see a problem that way. If this is really a snapshot that we have of Canadian population genetically at these loci that it should apply to Mr. Legere or his relatives.
- Q. I am wondering, you mentioned that there is a 1 / 25 per cent chance of siblings sharing the same bands, fragment lengths, and the theory is that it is the same in everybody except for identical twins. Is there a frequency for just say a twin rather than identical twin?
- A. Well, if you know that they are fraternal twins, if you

know that they are not identical, then the frequency of them sharing the same bands is the same as any two siblings because fraternal twins are no more genetically related than siblings are. Born from separate gestations. So would be one-quarter for fraternal twins, the same as it is for siblings from different pregnancies. The same two biological parents.

- Q. I remember the judge asking you when you were doing your direct examination about one chance in five million. What did that mean? That only one chance in five million that there is somebody else out there with that band frequency, and he mentioned that the chance for there to be two people out there in five million then you would have to say the chance of two people coming together with that band frequency say in the population would be five million times five million and you would also multiply it by two again?
- A. No. In that case you don't multiply it by two. You are looking at the genotype frequencies and that is a frequency where you don't have any ambiguity about one being mother or father and so forth, and you just asking for a match on all five of those loci in that case and it would be one in five million times one in five million too.
- Q. So the chance of two people coming together at the same time and same place would be five million times five million.
- A. Yes.
- Q. One chance in 25 million.
- A. It is even higher than that. It is five million times

five million. It is getting into trillions actually. Doesn't just go up by that --

Q. Sorry about that.

A. It is ten to the twelfth power in fact. 22.5×10^{12} . I don't know if that is trillion. I guess it is trillion.

Q. I want to use, for an example, and see if I am correct. There was evidence last week or two weeks ago that there was a hair sample found and the expert witness gave that there was literature as to the statistical possibility of another hair being out there was one in 4,500.

A. Right.

Q. He thought that was being optimistic because his own test, he only went up to 200 samples. But he said there is literature on it in his field that says it is one in 4,500.

A. Right.

Q. So if you were going to calculate the possibility of two people coming together in a community with hair samples exactly the same, that would be again 4,500 times 4,500? Same as the DNA?

A. Well, this is a subtle point. In probability theory you have to distinguish the questions you are asking. If you have a hair there is a probability of one that you have that hair. There is not a probability of one in 4,500 that you have that hair. Okay? So you have what is called a prior probability of one. You want to say 'I have this hair'. Now, if I looked for another hair -- I don't have it yet, but I get this other hair, what is the chance that that hair is going

to match this hair that has a probability of one because I have it --

Q. Right.

A. -- is going to be one in 4,500.

Q. Okay.

A. If I said 'I don't have any hairs', what is the chance that I took two of them at random like that, and in fact they would be identical to some prior hair that I had, it would be one in 4,500 times one in 4,500, but both of these were like a previously known standard hair, so there is some subtleties in that -- asking those questions. I am not trying to belittle your knowledge in any way. Please don't misinform me.

Q. No. No. I am here to learn too, Doctor. I am here to learn mostly. Not too. So what would the chances be of two individuals -- we are not talking about the same person here -- one hair going and picking out an individual who might have that hair. What is the chances of two individuals being in the same room at the same time and both having hair samples that you cannot distinguish if the probability is one in 4,500?

A. The answer to that is not as simple as you would think. I can given you an analogy.

Q. Okay.

A. If you have a room full of people you can ask the question 'What is the probability that two people have the same birthday?' You say, well, the chance of your birthday falling on any particular day is one in 365. Probability of anybody else's birthday falling on any day of the year is one in 365. But I ask you, what

is the probability that these two people have the same birthday, it is not one in 365 over one in 365 because there are 365 possible birthdays that they could match on. If you see that. I don't know. It is a subtle point but you see, it is different from the question saying, well, what is the probability that each of these people had the birthday of May 6th? Then it is one in 365 times one in 365 because each of them had to have a birthday on May 6th, but to ask the question that they both have the same birthday, well, they could have a birthday on January 1st each, which is one in 365 times one in 365. They could have a birthday on the 2nd of January each, one in 365 times one in 365. You could go through the year like that. So it is one over 365 times one over 365 but there is 365 ways that they could match. So in fact it turns out to be one over 365 because the 365 that you multiplied cancels out one of the denominators. So the chance that two people had the same birthday, in the same way the chance that two people taken at random like that matched in their DNA, is going to be a more complex probability than just multiplying those two together, because they could match in any number of ways. You see, you are not asking for a specific match. You are not saying that they matched exactly in that pattern there. If you said, were they to match in that pattern exactly as indicated there, then I could give you the probabilities, but to say that they could match on any pattern, you see, you would have to actually calculate the probabilities of a match for each of the possible patterns of which there are a

very large number of patterns, and then multiply the total probabilities and the denominators that you multiply together by the total probabilities of all the possible matches. I am just trying to get to the point that there are many ways that two hairs could match and it is not just that they could match in only one over 4,500 ways.

Q. No, I don't think it is one in 4,500 ways that they could match. You would have to say -- if they did their database the same way that the R.C.M.P. is doing their database -- they are checking the hair samples and they are going out and trying to find out if they can get a match out of 4,500 people and they can't get a match out of 4,500 people at random selection --

A. Right.

Q. -- how do they come up with the figure of one in 4,500, I am not aware of how they collected their data. Okay?

A. I can tell you how they did it in fact. I was involved probably with that study.

Q. Okay.

A. They actually took 200 people, randomly chosen people, and took ten hairs from each of those people and had forensic hair and fibre experts look at each of the possible combinations. I mean it was a heroic experiment. You had to look at each of those hairs, note them -- and there is something like 20 characteristics that they score in hair and fibre analysis -- I mean I am not an expert in that end of it but I know statistically how they did it -- and because there is some variation that not all the hair

even from the same individual will necessarily match another hair from that individual -- so there is that complication, but they had to do a very complex study which involved a lot of matches of hairs and then from that estimate what the chance is that two hairs taken at random would match. The equivalent to that is to break these down in terms of the VNTR loci, to break that down into the frequencies of the bands at each locus and to use this model of equilibrium at a locus and between loci to calculate what the potential different combinations would be and what their relative probabilities would be, and then to use that data to generate the probability of having a specific specimen turn up again in the population at random.

Q. You were involved in that study yourself you say?

A. I was involved and it is in fact about eight years ago, with Dr. Barry Gaudet at the R.C.M.P. who has written a seminal paper on this topic, yes. It was published under Gaudet and Keeping, a mathematics professor at the University of Ottawa, and has become the reference standard for all forensic hair identification, and the number of one in 4,500 has been in the literature now and Dr. Gaudet has testified on that in numerous courts.

THE COURT: You contributed some of the hair?

A. Alas, it might seem that way but there are genetic reasons that explain my phenotype. (laughter)

Q. So if I was caught at the scene of a crime or anywhere and they found a hair and they matched it to me, then they could say, well, there is only one chance in 4,500 that it came from somebody else.

- A. Yes.
- Q. If they were able to prove that that hair did come from somebody else what would be the probability that both myself and that somebody else was there at the same time? How would you calculate that probability?
- A. I don't think that that can be rigorously calculated just from knowledge of the hair. I don't know a way of calculating that probability just on the basis of having a specimen of hair that you were both there.
- Q. I was just considering that it was two identical events happening at the same time and same place. Just like your two bands. If both bands matched. Can't use that analogy or can you? Is it possible?
- A. Well, the probability that -- I can talk about the probability of the bands matching in this case because I know how the probabilities of all the observed similarities derive, which I don't know for the hair comparison. In this case the chance that you had a match and it came from two separate individuals is the number that you would calculate by looking at the frequency of those types in the population, the $2pq$ for that particular locus. Two times p_1 times p_2 for that locus.
- Q. Although you don't know how to calculate that frequency like I mentioned, you know, two people with the same hair sample being in the same place at the same time, it would be much greater than one in 4,500 anyway. Even though we don't know how much.
- A. If I had other information about those two individuals -- you see, the question is not a probability one. It is a question about what kind of knowledge I have about the situation. When you say -- I could ask the

-- what is the probability of two people in Canada being in this courtroom today? I wouldn't know how to calculate that because it is not a random probability. In the same way, at a scene of a crime or whatever it is not a random probability and is not an equal probability for any person in Canada having been at that scene of the crime. So I don't know how to calculate it just based on the hair information. If you asked me, what is the probability of this person in particular and this person in particular had been there, I would say, well, -- I would want to know further information about what the possibilities were of this person having been there. I mean that is that if a person was from Brazil and I knew had been in Brazil at that time of the commission of the crime, I could say that for certain he was not at the crime. But I couldn't say just on the basis of a hair sample that they were both present concurrently just on the basis of the one hair sample, is what I am saying. I could give you a probability that this one contributed the hair sample or that one, but that is not the same question as to what is the probability that they were both at the scene at the same time.

- Q. I guess it would be basically that maybe this individual contributed a hair sample, that individual contributed a hair sample --
- A. Okay.
- Q. Let's say that both hair samples, your normal forensic laboratory test says that they look the same. There is only one chance in 4,500 they come from somebody else.
- A. Umm hnn.

Q. And it is proven that they do come from somebody else so that puts both individuals there at the same time. What were the probabilities that both people were there at the same time?

A. And both left a separate hair sample.

Q. And both left --

A. And there were two hair samples found.

Q. Two distinct -- distinctly two different people. No question about it.

A. Well, in that case the random match for one is one in 4,500. The random match for the other is one in 4,500.

Q. Could you multiply that?

A. For the joint concurrence of those two, yes, I would think so.

MR. FURLOTTE: My Lord, there is no way I can finish this witness today. I just wondered how late you want to go.

THE COURT: Well, we started at 9:00, very shortly after, this morning. I think this might be a good time to stop. You have been on your feet quite a bit of the day.

MR. FURLOTTE: I drove here from Moncton this morning so I have had a full day.

THE COURT: You have had a hard day. The witness has been on all day.

MR. FURLOTTE: He needs a break too.

THE COURT: Why don't we adjourn until 9:30 tomorrow.

(4:30 p.m. Court is adjourned to May 7th, 1991 at 9:30)

IN THE COURT OF QUEEN'S BENCH OF NEW BRUNSWICK
TRIAL DIVISION
JUDICIAL DISTRICT OF FREDERICTON
B E T W E E N:

HER MAJESTY THE QUEEN

- and -


ALLAN JOSEPH LEGERE

AFFIDAVIT

I, Nancy Patterson, make oath and say as follows:

1. THAT I am a stenographer duly appointed under the Recording of Evidence by Sound Recording Machine Act.
2. THAT this transcript is a true and correct transcription of the record of these proceedings made under Section 2 and certified pursuant to Section 3 of the Act.
3. THAT a true copy of the certificate made pursuant to Section 3(1) of the Act and accompanying the record at the time of its transcription is appended hereto as Schedule "A" to this Affidavit.

SWORN TO at the City of)
Fredericton, Province of)
New Brunswick, this 22nd)
day of May, A.D., 1991.)
BEFORE ME:)



Verna Peterson
A Commissioner of Oaths

MY COMMISSION EXPIRES
DECEMBER 31, 1994



SCHEDULE "A"

RECORDING OF EVIDENCE BY SOUND RECORDING MACHINE ACT

CERTIFICATE

I, Nancy Patterson, of Harvey Station, New Brunswick
certify that the sound recording tapes labelled

R -vs- Legere

initialled by me and enclosed in this envelope are the
record of the evidence (or a portion thereof) recorded
on a sound recording machine pursuant to Section 2 of
the Recording of Evidence by Sound Recording Machine Act
at the trial held in the above
proceeding on the 6th day of May 1991 at
Fredericton, New Brunswick, and that I was the person in
charge of the sound recording machine at the time the
evidence and proceedings were recorded.

DATED at Fredericton, N.B. the 6th May 1991