DNA FORENSICS: PAST, PRESENT, AND FUTURE

James F. Crow

Genetics Department, University of Wisconsin, Madison, WI 53706

My assignment, as chairman of the Research and Development Working Group of the National Commission on the Future of DNA Evidence, is to look ahead 10 years into the new technologies that are expected to emerge during this period and try to assess the impact that these will have on DNA forensics. I would like to tell you something of the group's thinking. Let me add that we welcome, indeed we solicit your suggestions and comments.

SOME HISTORY, MOSTLY ANCIENT

But before we look ahead into the 21st century, I invite you to glance back at the 20th, if only as caution against a too confident prediction of the future. I'm no expert on the next century, but I do regard myself as an authority on the present one, for the good reason of having lived through most of it. Consider the year 1916. That was the year that a new journal, *Genetics,* made its appearance. The first article was an epic, C. B. Bridges' use of nondisjunctional aneuploids to prove the chromosomal basis of inheritance: every exception in chromosome behavior was accompanied by a corresponding exception in inheritance. By 1916 most geneticists, but not all, were already convinced, and this work provided the clinching evidence. It was a momentous year for the science of genetics. It was also a momentous year for me. The new journal and I were both scheduled to appear in January 1916. I arrived on time, but the first issue of *Genetics was* two months late.

Let's go back still earlier, to the year 1900, a great year for genetics. As everyone knows, this is the year that Mendel's laws were rediscovered. It was also the year in which Karl Pearson invented the Chi-squared test. And, of special interest to this audience, 1900 produced the first reliable genetic marker, the ABO blood group system. It is paradoxical that, although geneticists immediately realized that blood groups were inherited, the mode of inheritance was not demonstrated for a quarter century. Felix Bernstein, using the voluminous blood group data in the records of World War I, was able to prove the three-allele theory, a beautiful demonstration of the analytic power of gene frequency analysis.

It was not until the 1920s that the second polymorphic marker was discovered, the MN blood groups, and by the same man, Landsteiner. And it was Landsteiner once again who discovered the Rh factor around 1940. By the 1960s there were 17 blood group systems, although not all were useful for identification. By the 1980s the HLA system was added and there were some 150 protein polymorphisms. Now there are more than 30,000 microsatellites, many available for exploitation. I don't know how many SNPs there are, but surely the number is in the millions. So we have come a long way. And note that the interval between successive new discoveries has gotten shorter and shorter.

Let me take a halfway look, 50 years back. In 1950 I taught my classes that the human chromosome number was 48. The correct number, 46, was not demonstrated until late in the 1950s. In the present era of chromosome painting, when anyone with color vision can distinguish each individual chromosome, it seems incredible that even the number was once uncertain. I have the dubious distinction of having studied with the man, T. S. Painter, who published the wrong number. He also made another error. Around 1930, C. B. Davenport suspected that mongolism (as it was then called) was caused by an extra chromosome, and he sent some chromosomal material to Painter. Such was the state of the art of the time that Painter was unable to find anything unusual. Had he seen the extra chromosome 21, human cytology might have been decades ahead. Let me quickly add, however, that the fault is the primitive methods of the time, not

Painter's eyesight. In fact, he more than redeemed himself by the discovery of giant salivary gland chromosomes in Drosophila larvae.

Chemical mutagenesis had been discovered during the war, but because of military secrecy was not generally known until the late 1940s. In 1950, recombination in bacteria and viruses was quite new. The main emphasis was on formal genetics, but the power of fine-structure analysis would soon point the way to homing in on the gene. The gene was still thought to be protein by most geneticists, although evidence for DNA was already strong. Then came the Watson-Crick model in 1953, and soon afterward, the central dogma, DNA --> RNA --> protein, and the cracking of the genetic code. It was some time afterward before geneticists realized that the DNA of a chromosome was a single molecule.

I find it interesting, looking back, that geneticists could not conceive of the gene as being linear. It did not seem possible for a linear molecule to carry the amount of information that the gene's complexity demanded. But no one had any inkling as to how exceedingly long a DNA molecule really is. I remember Francis Crick's remarking that we must seem amazingly short and fat to a DNA molecule.

RECENT PAST AND PRESENT

The great practical applications of this scientific revolution were not immediate. It was another third of a century in 1985 when Alex Jeffreys first used RFLPs for forensic work, thereby ushering in a new area in human identification. The FBI began using DNA in 1998.

It is convenient to divide recent history into two periods. From 1985 to 1995 VNTRs dominated the scene, although the convenience of DQα and polymarker led to their wide use. One of the drawbacks of VNTRs was that they required relatively large amounts of DNA. Then PCR changed everything. VNTRs were generally too large to be amplified by PCR so that in the second period, from 1995 on, the shorter STRs began to predominate. Last year, the FBI settled on 13 STR loci as Core loci and these are increasingly being used by testing laboratories. Six of the 13 loci are used by the British Forensic Science Service, so international comparisons are feasible.

In its early days, DNA was highly controversial, and not always accepted by the courts. The arguments and uncertainties led to a study by the National Academy of Sciences (NRC 1992). Publication of this report started with a bang. A badly mangled report by the usually reliable New York Times, implied that there should be a moratorium on DNA use. This was followed the next day by a front page retraction. The report had made no such recommendation, although it raised a number of questions about the techniques and their interpretation. In particular, the report was extremely conservative about population calculations. Fearing the effects of undetected population substructure, the report recommended the "interim ceiling principle", which opened the door to creative misinterpretation. Needless to say, the principle was roundly criticized and largely for this reason, a second study (NRC 1996), which I chaired, was authorized. The ceiling principle was quickly abandoned by most forensic scientists.

Despite the criticisms, the ceiling principle had two worthy objectives that I don't want us to overlook. One was to be conservative in the sense of favoring the defendant. Second, was to obviate the need to classify subjects by ethnic or racial group. I shall return to these desiderata later.

I am happy to say that the 1996 report was much more favorably received. The strongest criticisms came from the statistics community. This was not unexpected, for in soliciting advice before the study began we had substantial input from statisticians, but with a considerable disagreement among them.

The biggest mistake of the 1996 report, I think, was too much emphasis on VNTRs and not enough on STRs. Part of this is excusable, for our report was essentially finished a long time before it finally appeared in print. Nevertheless, I think we should have been more current. One consequence is that we were, I now think, excessively conservative in suggesting, for STRs, using 0.03 for the population structure measure, θ

(same as Wright's F_{ST}), pending more data. The data show that there was no reason to alter the recommendation of 0.01, which we had suggested for VNTRs.

We quoted data, especially from TWGDAM, showing substantial agreement with Hardy-Weinberg proportions for a single locus and with linkage equilibrium for two or three loci, and therefore presumably for any number. We gave formulae for taking substructure into account and established rough, empirical margins of error, regarding these as more realistic than more formal statistical tests based on binomial assumptions.

Now 50 states have databases. Most have changed from VNTRs to STRs or are in the process. The numbers are still small: about 190,000 in the convicted felon database and 9,000 in the forensic profile database, by the most recent count that I am familiar with. The limiting factor in extending these databases is money. But within a few years, the numbers should increase enormously.

TECHNICAL IMPROVEMENTS IN THE FUTURE

My discussion of the past 50 years, with all its surprises, should caution us against any attempts to predict the next 50. Hardly any of the innovations that we now take for granted were known, or guessed, 50 years ago. Yet, for a shorter period, 10 years, we have a better chance of predicting accurately. The working group is charged with assessing technological improvements in the next 5 and 10 years and their impacts on forensic practice. Even for this period, prediction is risky; but it is an interesting challenge and reasonable guesses are credible.

There are two general kinds of improvements that can be expected.

The first of these is better techniques applied to existing systems, STRs in particular. We can expect more foolproof systems that can provide unambiguous tests for more loci at once. Even now, some systems have added additional loci to the Core 13. This surely will continue.

We can expect more automation; it is already happening. There will be more integration of computerized analysis with the laboratory tests. Capillary electrophoresis will require less material and produce faster results. We can also expect miniaturization with attendant portability. I recently read of a hand-held chip that will analyze 8 STRs in a few minutes. We can foresee the time when analysis can take place at the crime scene. If immediate results are produced this can provide prompt clearance of erroneously identified suspects, avoiding much needless apprehension. I would emphasize, however, that what can be done in pilot experiments will usually not be good enough for forensic use, for which a system must be thoroughly tested and validated.

As we go to more loci, I expect more pentanucleotides to be used more often. These often have a larger number of alleles and a more even distribution of allele frequencies. They are also said to have less stutter.

The second kind of improvement is new systems. Some of these are already in effect. The Y chromosome now has several STRs and hundreds of SNPs. It is particularly useful in tracing male lineages. It is also useful for separating sperm samples in, for example, multiple rapes.

Mitochondrial DNA has two useful properties. First, it follows the female lineage; this can be a disadvantage when it fails to distinguish among maternally related relatives. Its greatest advantage, however, is that, since there are thousands of mitochondria in a cell, very small amounts of material are required. For example, individual hair shafts can be analyzed. There is a problem, however. Sometimes the mitochondria are heteroplasmic; that is, there are more than one kind of mitochondria within a person or within a cell. This requires special care, but can often be turned to advantage. Both Y chromosome and mitochondrial DNA require large databases. Since there is no recombination, there is no way to estimate the frequency of composite genotypes from their component frequencies. Hence the database size is limiting.

Single nucleotide probes (SNPs) are already being used. In fact they have been used in the DQα system for many years. But in recent times, thanks in part to the genome project, the number of SNPs becomes almost unlimited. The disadvantage is that these are usually diallelic, but this can be compensated by using a much larger number of loci. SNPs are certain to be widely used in the future; but how soon?

ALU sequences have sometimes been used and may be used more extensively in the future. And we can expect purely physical methods, such as mass spectroscopy, to be developed.

There is one clear conclusion for the near future, say 5 to 10 years. The 13 STR Core loci work very well. They represent a large investment for many labs. They are not likely to be replaced, even if better systems are developed. I believe that these will still be used for the 10 years that our group is trying to foresee, although probably with increasing supplementation from other methods.

Future DNA forensics are likely to involve plants and animals. Already there have been court cases, and there is a cat database. We can expect similar studies of other domestic animals, and perhaps various plants and wild animals.

Future analysis need not be confined to human DNA. Each of us carries a unique constellation of viruses, bacteria, and other parasites. It may well be that in the future these will play an important role in individualization, supplementing chromosomal DNA.

STATISTICAL AND POPULATION ISSUES

At present, several laboratories are recording population data on more and more refined groups. Further studies of the extent of departure from Hardy-Weinberg and linkage equilibrium proportions are being carried out, and more are expected. As studies become more extensive, minor departures from simplistic expectations are certain to be found. So calculations of the future can build in corrections for such departures. American Indians, with their tribal structure, have a great deal of variability between groups, and this greatly complicates the analysis. We can expect much better information in the near future.

It is important that such data be made readily available for forensic labs and witnesses. There is a difficulty; some journals are reluctant to accept such data collections on the grounds that they are not novel. Something needs to be done. The FBI has taken a step. With the distribution of a new CODIS version, to begin in a few weeks, population statistics are included in the files.

We can expect, I think, that conditional probabilities using empirical estimates of F_{ST} (or θ) will become routine. In a structured population it is likely that, with larger samples and a larger number of loci, departures from linkage equilibrium will be found. For example, if two individuals show identity at several of the first 10 loci analyzed, it is likely that they are relatives. In this case they will be correlated at the remaining 3 of the core 13. How important this is remains to be seen, but it may not be as negligible as is currently assumed.

There is always the possibility that two persons, contributor of the evidence sample and a suspect, are relatives or are from a small population subset (which is saying much the same thing). Can we find a principle that is robust with respect to these possibilities? I think there is promise in what I have called the Sib Rule.

The Sib Rule. If we have two subjects, we can calculate the match probability if they were sibs. This provides an upper limit on the true match probability, since we know that no other relatives have a higher probability of matching. So, we can use the Sib Rule as an upper limit.

This brings a bonus. Sibs always have a probability component of 1/4 of being identical with respect to any locus. This comes from simple Mendelian segregation regardless of inbreeding or any other complications. Furthermore, if the loci are independently inherited, linkage disequilibrium doesn't matter either. There are other terms, dependent on allele frequencies and inbreeding that add to the 1/4, but these are relatively minor, especially if we chose loci with a large number of alleles with evenly distributed frequencies, such as some pentanucleotides.

Here are the conditional sib probabilities for a single locus:

Homozygotes: A_iA_i: $(1 + 2p_i + p_i^2 + 4\theta)/4$
Heterozygotes: A_iA_j: $(1 + p_i + p_j + 2p_ip_j + 2\theta)/4$

A reasonable value of θ is 0.01 and p_i and p_j are the allele frequencies, usually much smaller than one. These single locus probabilities are multiplied to give a composite probability for multiple loci. Obviously, as I said earlier, the factor 1/4 dominates the expression.

By using the Sib Rule, we can largely avoid the complications of population structure and uncertainties about population equilibria. Not entirely, mind you, but to a considerable degree, especially with judicious choice of additional loci. With 13 loci the general match probability for sibs is about one in 300 thousand. With 21 loci, the value drops to one in 300 million. In two years, this number of loci will probably be readily available.

I think it is likely that the courts will prefer a higher match probability with fewer assumptions to a smaller one with greater uncertainty as to the validity of the assumptions. We'll see if the Sib Rule catches on.

Individualization (uniqueness). The FBI has recently argued that if the frequency of a profile is considerably less than the reciprocal of the population size, the profile can be declared unique. Several conservative factors were built in to the recommendation. First, they ask that the probability of not finding the profile be less than 0.01. Second, they chose the largest calculated profile frequency among the four common databases. Third, they multiply the frequency by 10, to allow for uncertainties in the assumptions following the empirical data from NRC 1996. This procedure has been criticized as simplistic, for ignoring population structure; but perhaps the safety factors sufficiently compensate for this. Balding (1999) has suggested a somewhat similar procedure, but one that places greater emphasis on the presence of undetected sibs in the population. There is, of course, no scientific definition of uniqueness, but there may be a political or legal definition within the next few years.

Suspect Identified by a Database Search. How are decisions based on match probabilities and likelihood ratios affected by the manner in which the suspect was found? If the suspect was identified by a search through a database that is representative of the general population, the larger the database, the greater is the chance of its including an innocent person whose DNA matches the evidence sample. Both NRC reports addressed this problem.

The first report (NRC 1992) recommended that different loci be used for courtroom evidence than were used to identify the suspect. Clearly, this is an unbiased procedure, but it may not be feasible if the number of loci is small. The second report (NRC 1996) agreed with this, but suggested that with limited loci the match probability or likelihood ratio be adjusted by the size of the database. When the match probability is much less than the reciprocal of the database, the approximation correction is to multiply the calculated probability by the size of the database. There may be the possibility of a solution by which a limited number of loci are used for identification, thus leaving enough loci for later courtroom use.

Evett and Weir (1998) argue that the likelihood ratio itself is not appreciably affected by the manner of ascertainment of the suspect. This is correct, but although the database search does not much affect the likelihood ratio, it does affect the prior likelihood that the evidence and suspect DNA came from the same

person. Such a prior is needed to place a probability interpretation on the likelihood ratio. The prior probability is likely to be low if the database comes from a population sample. On the other hand, if a database of convicted felons is used, the recidivism rate may be such as to substantially increase the prior probability. So computing prior odds is not obvious. Some have advocated what is already practiced in paternity testing, that the court be presented with a range of prior probabilities and the corresponding posterior probabilities. At present, the extent to which the courts will accept Bayes' Theorem is not clear. Presumably there will be some clarification in the next decade.

SOME OTHER POSSIBILITIES FOR THE FUTURE

Inferring Geographical Ancestry. It is already possible, with the 13 Core STR loci, that some profiles are appreciably more frequent in one geographical group than in another. If a profile were 100 times as likely in the Caucasian population as in Hispanics, this could help narrow the range of suspects searched. If this is done, it would be well to consider the background population frequency, used as a prior probability. I should emphasize that this is only probabilistic and is of value only in preliminary identification, not in establishing evidence of a common source of evidence and suspect DNAs.

Stronger evidence of this kind comes from other loci. For example, the Duffy-null blood group allele is quite common in parts of tropical Africa, because of its role in malaria protection, and rare elsewhere. But we have two cautions. One is to note that knowing that a person has some African ancestry may not be very indicative of actual phenotype. Second, such analysis depends on different kinds of loci than are used for ordinary forensic analysis. The Research and Development working group hopes for more emphasis on traits and less on geographical ancestry, in order to avoid any confusion with racial profiling.

Individual Traits. As a result of the genome project and other genetic research, more and more gene-determined individual traits are being identified. In the near future we can expect that genes for eye pigment, hair color, baldness, skin pigment, color-blindness, albinism, and others will have been identified. This is an area that is being intensively researched at present. It would be rash to predict, but it seems likely that a profile that could be useful in the search for suspects will be found in the near future.

Again, I emphasize:

(1) These depend on loci different from those used for forensic work.
(2) Phenotypic predictions will usually be probabilistic rather than certain.
(3) They may be able to narrow the range in a suspect search, but are not suitable for courtroom evidence.

Differences in Gene Expression Level. In recent years the development of micro-arrays has been rapid and spectacular. Of potential forensic interest are expression arrays. These permit not simply the identification of what alleles are present or absent, but also their degree of expression. Thus we can foresee the time when degree of gene expression can be added to the battery of tools available for identification. It is possible, since gene expression changes with age for many loci, that expression arrays can provide evidence of the age of the contributor.

The Possibility of Distinguishing between Identical Twins. Right now, analysis of sufficient DNA can distinguish among individuals, even close relatives. The exception is identical twins. It would be of value in forensic work if individual twins could be identified. There are several ways in which this might happen. One is to use loci that are highly mutable in somatic tissue. The immunoglobulins spring to mind, but there are perhaps others such as some trinucleotide repeats. A quite different approach is to study the possibly different parasites carried by twins. Each member of the pair might have a different array of viruses and bacteria. Or if they have acquired the same latent virus, it may have integrated at different sites in the two twins. I think it is a reasonable guess that within the 10 years we are trying to foresee, we shall be able to distinguish between identical twins.

DNA as a Standard. The forensic use of DNA has had a level of scrutiny that is unique in the annals of crime evidence. The techniques have been thoroughly tested for robustness and reproducibility before being validated. Laboratories have been subject to strict controls including regular audits and proficiency tests. And the population assumptions have been examined in great depth in and out of the courts. I believe that this has been good. DNA forensics is now on very solid grounds.

This raises the question: Are other sources of courtroom evidence similarly reliable? How rigorous and reproducible are ballistics, lie detectors, microscopic hair analysis, handwriting, and fibers, to say nothing of eyewitnesses. It is possible, and I think it would be desirable if these and other forms of conventional evidence were to receive more scrutiny. Perhaps DNA can serve as a model for the upgrading of the scientific basis of other types of evidence.

Universal Databases. In 10 years there is likely to be strong pressure for uniform DNA databases, perhaps from all newborns. There is also likely to be strong objection on privacy and libertarian grounds. One possibility is to keep a close watch on Iceland, which is undertaking such a program. Will it produce a great deal of useful medical information, facilitating early diagnosis and treatment? Will it continue to have the high level of social acceptance that it now enjoys in Iceland? This will give us an opportunity to see what happens on a smaller scale before we either run headlong into a large, possibly risky program, or bury a potentially useful idea before it ever has a chance to be tried.

Envoi. I needn't say that looking into the future is uncertain. It is easier to predict technical advances than their public acceptance or rejection, for example trait identification. I think it is safe to predict all sorts of technical advances -- many of them things we can't now imagine. I think it is also safe to predict that there will be controversy. Technical possibilities run head-on against concerns for privacy. But there is one certain prediction: there will be intellectual excitement. I hope to be around long enough to experience it.

I am indebted to the members of the Working Group on Research and Development for many of the ideas expressed here. I have, however, intruded some opinions of my own, for which the Working Group should not be held responsible.

REFERENCES

Balding, D. J. 1999. When can a DNA profile be regarded as unique? Science and Justice (in press).

Evett, I. W., and B. S. Weir. 1998. Interpreting DNA Evidence. Statistical Genetics for Forensic Scientists. Sinauer Associates, Sunderland, MA.

NRC 1992. DNA Technology in Forensic Science. National Academy Press, Washington, D. C.

NRC 1996. The Evaluation of Forensic DNA Evidence. National Academy Press, Washington, D. C.