

Genetic Identification of Missing Persons: DNA Analysis of Human Remains and Compromised Samples

M.J. Alvarez-Cubero^a M. Saiz^a L.J. Martinez-Gonzalez^b J.C. Alvarez^a
A.J. Eisenberg^c B. Budowle^c J.A. Lorente^{a, b}

^aLaboratory of Genetic Identification, Department of Legal Medicine and Toxicology, Medical Faculty, University of Granada, Granada, and ^bPfizer-University of Granada-Junta de Andalucia, Center for Genomics and Oncological, and Biomedical Research Center, Armilla, Spain; ^cInstitute of Applied Genetics, Department of Forensic and Investigative Genetics, University of North Texas Health Science Center, Fort Worth, Tex., USA

Key Words

Bioancestry · Databases · DNA typing · Forensic genetics · Genetic identification · Missing persons

Abstract

Human identification has made great strides over the past 2 decades due to the advent of DNA typing. Forensic DNA typing provides genetic data from a variety of materials and individuals, and is applied to many important issues that confront society. Part of the success of DNA typing is the generation of DNA databases to help identify missing persons and to develop investigative leads to assist law enforcement. DNA databases house DNA profiles from convicted felons (and in some jurisdictions arrestees), forensic evidence, human remains, and direct and family reference samples of missing persons. These databases are essential tools, which are becoming quite large (for example the US Database contains 10 million profiles). The scientific, governmental and private communities continue to work together to standardize genetic markers for more effective worldwide data sharing, to develop and validate robust DNA typing kits that contain the reagents necessary to type core identity genetic markers, to develop technologies that facilitate a number of

analytical processes and to develop policies to make human identity testing more effective. Indeed, DNA typing is integral to resolving a number of serious criminal and civil concerns, such as solving missing person cases and identifying victims of mass disasters and children who may have been victims of human trafficking, and provides information for historical studies. As more refined capabilities are still required, novel approaches are being sought, such as genetic testing by next-generation sequencing, mass spectrometry, chip arrays and pyrosequencing. Single nucleotide polymorphisms offer the potential to analyze severely compromised biological samples, to determine the facial phenotype of decomposed human remains and to predict the bioancestry of individuals, a new focus in analyzing this type of markers.

Copyright © 2012 S. Karger AG, Basel

Introduction: Forensic Genetics and Challenging Samples

Since the advent of forensic DNA analysis there have been two main objectives: (1) the identification of those who could be the source of biological evidence, which includes associations of individuals due to some alleged

kinship; and (2) to exclude individuals wrongly associated with evidence. The generation of reliable genetic profiles from unknown and reference samples, systematic and objective interpretation practices, and providing a statistical evaluation of the results are tantamount to a robust forensic DNA identification program. These criteria, seemingly obvious today, were envisioned 25 years ago by the forensic geneticists that developed this field. Moreover, the standards of practice used for forensic DNA typing are dramatically impacting in a positive way on the standards of criminalistics. Concepts and topics related to quality control and assurance, validation, proficiency tests, documentation and statistical evaluation, for example, are being reconsidered by forensic scientists in other disciplines to improve quality and reliability of their areas of forensic science. A further discussion on the direction of the forensic sciences can be found in the National Academy of Science Report (*Strengthening Forensic Science in the United States: A Path Forward* [1]).

Because of the success of forensic genetics in the identification of sources of biological evidence, developmental and innovative progress continues with more expectations of assisting investigators. Future endeavors could include predicting the phenotype of an individual from a bloodstain or human remains, determining factors related to cause of death, better capabilities to type severely damaged biological evidence, typing fetal DNA from maternal circulating blood, and novel genetic markers for identity testing and bioancestry, for example. These amazing realistic potential applications were not too long ago the substance of science fiction. However, today efforts on the generation of large databases to develop investigative leads when there is no known suspect(s) and improving capabilities to extract genetic information from challenged samples are driving developments in human DNA identity testing. These two areas are of forensic, criminal relevance and directly impact on society, as well as historic interests in a number of situations.

Large databases house DNA profiles from convicted felons (and in some jurisdictions arrestees), from forensic evidence, human remains, and direct and family reference samples of missing persons. There is a demand for more typed samples to be placed in these databases to help develop more investigative leads for crime solving. This need has motivated the community (government, academia and industry) to work collaboratively to develop and validate standard DNA typing kits that contain the reagents necessary to type core identity genetic mark-

ers and the concomitant automation of a number of processes (extraction, quantitation, amplification and to some degree interpretation of the results) to facilitate typing. The ability, for example, to type a reference blood or buccal sample without extracting the DNA away from interfering compounds would not have been imagined possible a few years ago and yet today such technology is readily available [2, 3].

The standard genetic markers used in essentially every forensic DNA typing laboratory in the world are autosomal short tandem repeat (STR) loci [4, 5]. The standard operating procedures employ a set of 10–17 STR loci, which provide a high level of diversity and resolution for identity testing [6–8]. Commercially available kits, such as the AmpF ℓ STR $^{\text{®}}$ Identifiler $^{\text{®}}$ PCR amplification kit (Applied Biosystems, Foster City, Calif., USA) or the PowerPlex $^{\text{®}}$ 16 system (Promega, Madison, Wisc., USA), enable analysis with high-quality materials and forensic samples. These kits, and STR loci, have been used extensively for the identification of human remains as well as in kinship cases, such as paternity testing and family reconstructions. While these STRs apply to the majority of biological evidence analyses, there are situations where autosomal STRs cannot yield sufficient information. Evidence, such as mixtures with a large amount of female DNA and a small amount of male DNA, and kinship cases where the reference sample derives from a relative separated several generations from the individual of interest require other markers, e.g. Y-STRs, X-chromosome-linked markers or mitochondrial DNA (mtDNA). These lineage-based systems provide additional power because of their unique biological qualities compared with autosomal markers. Additionally, when performing familial searching, i.e. searching for relatives of the true source of the sample, these lineage markers are extremely useful for reducing the number of adventitious associations in candidate lists. mtDNA sequencing is used to strengthen the genetic evidence when there are maternal relatives available to serve as references [9]. mtDNA, inherited through the maternal line, has a special property that makes it particularly useful for samples that are severely degraded and/or of limited quality, such as human remains in missing persons and mass disaster cases. There are hundreds to thousands of mtDNA molecules in a cell compared to only two copies of nuclear autosomal genetic markers [10]. Thus, when STR typing does not yield a result, there still is a good chance to obtain a result via mtDNA typing. In the following sections, the application of DNA analysis will be described in different areas where the common issue is the association of individuals via kinship analyses

either for identifying human remains, combating human trafficking or historical studies. These approaches generally share the process of indirect genetic comparisons to determine the identity of an individual and often the samples are compromised. The examples used are based on our own experience in developing and establishing missing persons' databases [in Spain and in the US, the so-called FENIX and the Missing Persons Program at the University of North Texas (UNT) Center for Human Identification (CHI), respectively], missing children identification, which is embodied by the DNA-PROKIDS (Program for Kids Identification with DNA Systems) program, and attempts to identify persons of historical interest.

Identification of Missing Persons

Strategies for the Identification of Missing Persons

The identification of human remains belonging to missing persons is one of the main challenges for forensic genetics. These identifications are a universal problem. When other human identification forensic techniques (e.g. dactyloscopy, anthropology, odontology and medico-legal examinations) provide limited information, or to support further or refute potential associations, DNA typing can be extremely valuable. DNA, in theory, can be recovered essentially from any tissue (e.g. soft, degraded tissues, bones, teeth and hairs). In addition, because close relatives on average share more genetic variants than unrelated individuals, relatives can provide reference samples to effect identifications of unknown individuals or their remains or, at least, to develop potential investigative leads for law enforcement. Lastly, DNA analysis can be a powerful exculpatory tool eliminating wrongly associated individuals, reducing candidates and redirecting resources to more viable avenues of investigation.

The Spanish Phoenix Program (Programa FENIX)

In November 1998, the Spanish Ministry of the Interior decided to support an initiative from the University of Granada that was presented to the Guardia Civil to implement a National Program to attempt to identify cadavers and bones from missing persons. The program was named 'Phoenix Program' (Programa Fénix, in Spanish) based on its purpose with the name derived from classic Greek mythology.

The Phoenix program contains two independent databases (the *questioned database*, containing STR profiles

and mtDNA sequences from bones, and the *reference database*, containing STR profiles and mtDNA sequences from relatives). The databases house genetic data that can be compared automatically to associate matching or related profiles, such as those from unknown remains and reference samples from relatives [11, 12].

The general procedure is shown in figure 1. Only persons signing a valid informed consent protocol are allowed to participate in the program and provide reference samples. People that have reported missing relatives are requested to contact the Phoenix Program by calling a toll-free telephone number. For those who voluntarily provide reference samples, two buccal swabs are obtained. A minimum of 2 and a maximum of 4 relatives (when available) can be sampled. All samples are bar coded and subsequent genetic data are coded to maintain confidentiality and reduce the misuse of genetic data. For *questioned database* samples, typically 2–4 fragments of at least 25 g of compact bone and/or teeth from non-identified cadavers and human remains are provided by trained specialists of the Guardia Civil.

Autosomal STRs and mtDNA analyses are routinely performed. In cases requiring additional information, such as when a maternal reference sample is not available, Y-chromosome STRs can be included. STR typing is facilitated due to the advent of multiplex commercial kits that provide the necessary reagents to type up to 15 STRs and a gender determination maker known as amelogenin. These kits are the PowerPlex16 kit (Promega) and the Identifiler kit (Life Technologies-Applied Biosystems). Nowadays, there is an improvement in the Identifiler kit (Identifiler Plus kit) and PowerPlex kit (PowerPlex 16 HS) which allows for better data recovery from degraded samples and a greater sensitivity of detection. Since these kits or variations of these kits are used worldwide, there is a high likelihood that any forensic or reference sample that is typed by a forensic laboratory will contain common genetic markers. Therefore, genetic results can be readily compared regardless of the country or laboratory that generates or maintains the data.

For mtDNA analysis, DNA from 1 buccal swab from 2 maternally related individuals per case is extracted and sequenced for the 2 most informative regions of the mtDNA genome. The regions are the hypervariable region 1 (HV1) and hypervariable region 2 (HV2) of the control region or d-loop of the mtDNA genome.

To date, >3,700 families have contacted Phoenix, 862 have enrolled in the program and at least 319 unidentified remains have so far been identified and returned to their relatives. When mtDNA and/or STR associations are

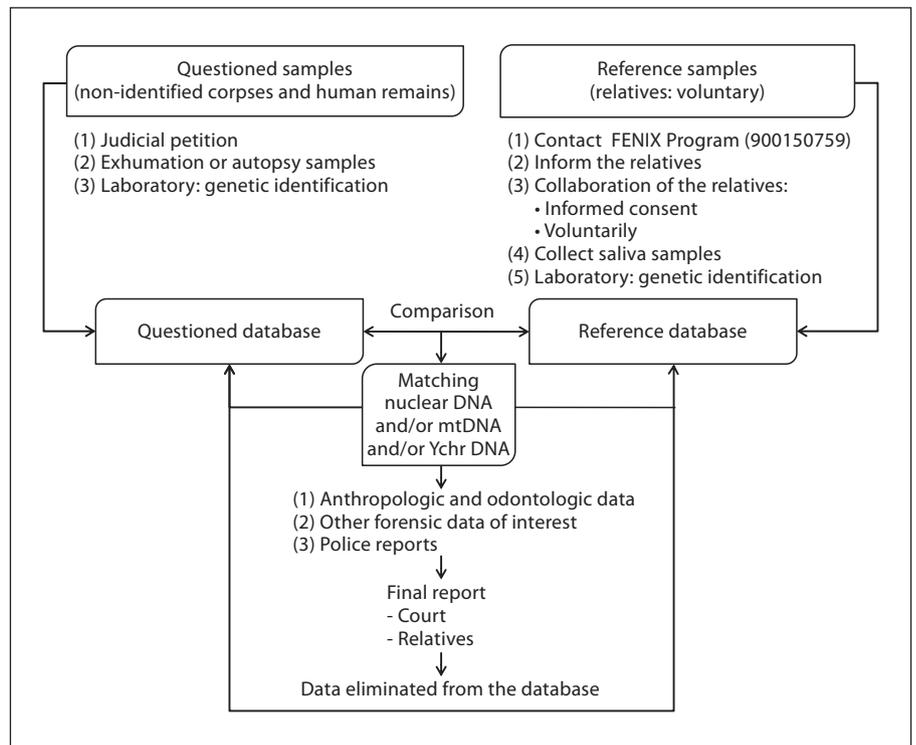


Fig. 1. Identification of corpses and human remains in the Phoenix program. Ychr = Y chromosome.

found, a second independent analysis is performed as part of the quality assurance process.

Nationally and internationally compatible protocols leading to the identification of human remains or skeletons ideally will require the use of databases that meet four basic prerequisites, as follows:

(1) Analyses have to be based on standard operating protocols and universally accepted genetic markers.

(2) Results must be reliable (laboratories and techniques are subjected to quality assurance and quality control programs).

(3) The technology should be amenable to automation, to facilitate the typing of the anticipated large volume of samples and to allow intra- and international searches and comparisons.

(4) The data provide little or no personal or confidential information about the individual(s). Proper use of the database according to national laws, dissociation of data, restricted access, informed consent from voluntary donors and court orders to handle human remains are among some of the requirements of Spanish database management.

The Phoenix program is using DNA to develop associations between relatives of missing persons and unidentified cadavers or human remains of previously unsolved

cases. Once an association is found, anthropologists, odontologists, specialists in forensic medicine and law enforcement officers work together to establish a final, positive identification and prepare a report for the Court (<http://www.guardiacivil.es/prensa/actividades/fenix/presentacion.jsp>).

The US Missing Persons Program

In the US, the Federal Bureau of Investigation (FBI) facilitates similar databases of missing persons and unidentified human remains, the NMPDD (National Missing Person DNA Database) Program. The database contains three indexes in which DNA profiles can be entered: *biological relatives of missing persons*, *unidentified human remains* and *missing persons*. There are only a handful of qualified laboratories in the US that have full capabilities to analyze missing person cases. The capabilities include typing the full battery of genetic markers, i.e. autosomal STRs, Y-chromosome STRs and mtDNA (<http://www.fbi.gov/about-us/lab/dna-nuclear>).

Nowadays, in the US, a total of 4,285 families have reported missing relatives whereas only 3,461 unidentified human remains have been found.

The largest program of missing person identification in the US, however, is in the State of Texas. The *Texas*

Missing Persons DNA Database was established in 2001 at the UNTHSC. UNTHSC in collaboration with law enforcement offers families with missing loved ones the opportunity to submit reference samples for DNA testing. The laboratory is one of only a few facilities that integrates nuclear DNA and mtDNA for analyses. Once DNA profiles are obtained, they are directly entered into the FBI Combined DNA Index System plus Mito (CODIS+mito) database.

The database began accepting samples from Texas law enforcement agencies in March 2003. Texas was the first state in the country with a missing person DNA database capable of analyzing both mitochondrial and STR systems and is the first state to participate in the federal database for missing persons (the Federal Bureau of Investigation's CODIS). The database provides a very powerful tool for investigators trying to locate missing persons or identify remains by allowing federal, state and local crime laboratories to electronically exchange and compare DNA profiles. The DNA analysis provided by the Texas missing person DNA database is at no charge to law enforcement agencies or families with missing members.

The UNTHSC in Fort Worth, Tex., USA, is home of the UNTCHI DNA laboratory established in 2001. In collaboration with law enforcement agencies, medical examiners and coroners throughout the nation, the UNTCHI has established one of the largest missing person programs in existence. The program offers the families of missing loved ones the opportunity to submit reference samples for DNA testing. The laboratory is one of only a few facilities that integrates the analysis of nuclear DNA and mtDNA. UNTCHI incorporates the use of the FBI CODIS+mito database to locate missing persons and identify human remains. The database operates at a local (LDIS), state (SDIS) and national level (NDIS). Texas was the first state in the country with a missing person DNA database. UNTCHI services are performed at no charge and have expanded to a national level. To date, UNTCHI has completed the analysis of 7,792 family reference samples, 3,461 unidentified human remains and 182 direct reference samples. Once completed, these profiles must meet eligibility requirements in order to be uploaded to the next level of CODIS. Of the eligible completed samples, a total of 5,300 family reference samples, 2,450 unidentified human remains and 140 direct reference samples have been entered into the local CODIS database. Currently, UNTCHI has made 700 associations through the use of CODIS.

Missing Children Identification

Children Trafficking and Exploitation

Children, the most innocent individuals of society who should be protected to the best of our abilities, can be subjected to many abuses. One of those abuses is human trafficking, an apparently lucrative criminal activity. According to UNICEF, 'an estimated 300 million children worldwide are subjected to violence, exploitation and abuse including the worst forms of child labor in communities, schools and institutions; during armed conflict; and to harmful practices such as female genital mutilation/cutting and child marriage'. Figures from the United States only begin to demonstrate the magnitude of the missing children problem within a country. Approximately 800,000 children are reported missing each year. Of these, approximately 360,000 are runaways, 340,000 are classified as 'missing with benign explanation', and about 100,000 are abducted either by family members or other known individuals or are lost and/or injured [13]. While these figures are disturbing, they relate to mostly domestic situations and do not represent the greater international problem where children are illegally sold for often malevolent purposes.

We must ensure that children are treated humanely and in keeping with national and international standards. In an effort to combat human trafficking and especially to protect children (and women), the United Nations (UN) launched the UN-Global Initiative to Fight Human Trafficking (UN.GIFT) in March 2007 (<http://www.unodc.org/>). Such initiatives demonstrate that governments recognize the problem of human trafficking is substantial and steps must be taken to eradicate this victimization of children. However, there are several obstacles to combating human trafficking. International agreements and policies must be in place to track, identify, communicate and share data, as well as interdict perpetrators and help victims. Worldwide political and legal coordination plays a crucial role, but the legal and social difficulties that exist for access and disclosure among countries substantially slow the progress of implementing a comprehensive and effective counter-trafficking system.

Scientists have the technology tools to assist in the identification of missing persons and yet may not be able to apply them in the international endeavor to fight human trafficking. The use of different genetic markers and/or incompatible software would obstruct this effort, which is especially troubling when children are victims. Indeed, genetic testing capabilities for identification are

readily available and implementation may seem trivial compared with overcoming legal infrastructure barriers and addressing privacy concerns. Because legal infrastructures will be slow to develop it is important that the construction of the technical infrastructure should proceed in preparation for the implementation of an appropriate international legal framework. This will facilitate efforts in thwarting human trafficking, once the policies for sharing data are addressed adequately. One technical area sufficiently developed for implementation is the use of the molecular biology analytical tools (i.e. DNA testing) and databases for the identification of missing children.

The DNA-PROKIDS Program: DNA to Fight Crime on Children

Because of the importance of the children trafficking problem and based on our experience with the Phoenix Program [14] and those at the UNTHSC and the UNTHSC Center for Human Identification (UNTHSCCHI), we have launched DNA-PROKIDS located at the University of Granada and in collaboration with UNTHSCCHI. The Program is an international effort to help identify missing children, provide support to their relatives and to contribute to efforts against human trafficking. This non-profit program is supported by the Spanish Government, the Andalusian Government and donations from private companies and foundations (BBVA, Banco Santander, CajaGranada-BMN and Life Technologies). DNA-PROKIDS is composed of three tiers. The first tier is at the National level with two genetic databases or indices per country. One index is for DNA profiles (and meta data) obtained from children who, after proper investigation, are found in an illegal situation (e.g. not living with the natural family due to abduction, kidnapping, illegal adoption or other criminal situations). The other index comprises DNA profiles (and meta data) voluntarily provided by relatives (parents, sibs and other meaningful family members, and whenever possible obtaining mothers) or from personal items of reported missing children (fig. 2). The DNA profiles in these two indices will be compared routinely to assist in identifying missing children initially within their own countries. DNA-PROKIDS first tier pilot programs are in effect in Guatemala and Mexico. Additional efforts are currently underway in Brazil and China. It would be highly desirable to coordinate these efforts. Without coordination it is feasible and somewhat tragic that systems may be developed which are incapable of data exchange because of the use of non-overlapping genetic markers or incompatible software, or both.

The power of identity testing and database searching will be most effective if DNA analysis is performed in every single case of a child being given to adoption. Furthermore, when possible, the mother (or other available biological relatives) should also be tested to confirm her relationship to the child and therefore right to relinquish the child. Implementation of this act alone would reduce crimes related to children, where the children are given to adoption not by their biological families, but after kidnapping or abduction.

The second tier is at an International level. This will allow the development of the infrastructure required to share data among countries through various existing approved links and networks as well as through the establishment of new, specific links. International cooperation requires decisions on many issues, including a common set of DNA markers to enable genetic data sharing; the meta data that should be collected; the information that can be shared; legal and privacy issues that need to be met to share data, and the development of sufficient sustainable financial support to establish such programs in various countries.

The third tier is focused on data generation. There will be a universal automatic, mandatory inclusion (with signatory countries) in the database of (1) any reported child found out of his/her family (without legitimate reasons); (2) children who are going to be adopted (and before any adoption can be made it will be necessary to confirm that the child has not been reported as missing anywhere in the world), and (3) the immediate inclusion of voluntary relatives of missing children.

DNA-PROKIDS makes use of the same robust DNA typing methods used for forensic casework and those described above for missing person identification. These markers are extensively validated, and substantial data exist that support their utility. In addition to STRs and mtDNA, single nucleotide polymorphisms (SNPs) are particularly suited for a program like DNA-PROKIDS [15–18]. SNPs are genetic variants that are the result of substitutions or insertions/deletions at one or a few bases in the genome. They occur at about 1 SNP/1,000 bp in the human genome and account for approximately 85% of human genomic variation. Millions of SNPs have been identified and a subset of these is suitable for identity testing. They would help to increase the probability or likelihood ratio in cases of positive associations, and also to overcome problems related to mutation that could occur more so with STRs and mtDNA.

The application and usefulness of DNA identity testing are already well documented. To date, DNA-



Fig. 2. Extraction kit of DNA-PROKIDS.

PROKIDS participating countries have analyzed over 2,500 cases (basically from Mexico, Guatemala, El Salvador, Paraguay, Peru, Bolivia in Latin America, and the Philippines, Thailand, Indonesia and India in Asia). DNA analyses first and subsequent application of accompanying meta data have already helped to identify >330 missing children, who have been returned to their families. If not for this intervention, it is likely these children would have been given or sold into illegal adoptions, would still be under exploitation or would have died without identification. Additionally, their respective families would still be suffering the loss of their children. Beyond the identification of these children and returning them to their families, the database could play a deterrent role. Efforts that increase the size of the database and facilitate communication among countries, such as is encountered with various criminal investigations through INTERPOL, may make criminals more reluctant to commit these heinous crimes on children (trafficking, exploitation and illegal adoptions) because at least authorities will more likely be able to identify and apprehend perpetrators.

If these programs were enacted, the ability to immediately identify reported missing children would not only permit returning them to their families, but also would begin to compromise criminal network operations. More operational data and updated information can be found at www.dna-prokids.org.

Historical Cases and DNA Identification

There have been a number of historical individuals whose identity of putative remains have been confirmed or questioned by DNA typing. These include Tsar Nicholas, St. Birgitta, Napoleon, or African-American descendants of Thomas Jefferson's line, for example. Typically, mtDNA and Y-STRs, and at times autosomal STRs, play important roles in historical analyses.

We have had the opportunity to participate in a number of historical identifications in Spain. In 1994, a set of remains thought to be those of Queen Blanca I of Navarra were found in a church in the village of Santa María de Nieva, province of Segovia, Spain, where it is known that she died on April 3rd, 1441. The anthropological analysis

did tentatively conclude that the remains could be the ones of Queen Blanca, but because of the data and changes to different graves, DNA analysis was requested. According to historians and experts, the Regional Government of Navarra (northern Spain) decided that the reference for DNA analysis should be the Prince Carlos of Viana, the son of Queen Blanca. It is also known that the Prince of Viana was buried at the Monastery of Poblet, in the province of Tarragona, Spain.

When we proceeded with the exhumation of Prince Carlos, the remains inside the coffin were found to belong to at least 3 different individuals based on skeletal morphology. This observation had a historical explanation. Poblet's monastery had been assaulted and its tombs desecrated a number of times during the XIXth century. mtDNA analysis performed on the putative remains of Queen Blanca and his son Prince Carlos [around 1994 and 1995 – at the University of Granada and at the Pennsylvania State University (Dr. Mark Stoneking and Dr. Anne Stone) – did confirm the initial findings. The coffin of the Prince of Viana contained at least 3 persons. Furthermore, none of the mtDNA sequences obtained from the remains that could have been the ones from the Prince matched the sequence generated from the remains who are thought to be the ones of Queen Blanca [unpubl. results].

Therefore, DNA analysis could not resolve the authenticity of the remains of the Queen. They did support, however, that the remains in Poblet are not likely those of Prince Carlos. Studies are on the way to find an appropriate reference sample. The process is a very difficult one because most of the reference samples belong to kings, queens and other members of royal families, and permission for sample access is often hard to obtain.

The second project where our teams have been involved is in the identification of the remains of Christopher Columbus, who died in 1506 in Valladolid (northwestern Spain). His bones reportedly were moved to Seville (Spain) in 1509 and then from Seville to Santo Domingo (Dominican Republic) in 1544; in 1795 from Santo Domingo to Havana (Cuba), and finally, in 1898, the remains of Columbus (or what it was thought to be the remains of Columbus) were sent back to Spain, and buried at the Cathedral of Seville.

Because the remains were subjected to 4 different transfers and because of some historical doubt regarding the authenticity of the remains buried in Seville, a group of historians led by Mr. Marcial Castro initiated a project (still under development) to first try to determine if the remains in Seville are the ones of Columbus. Second, he sought to undertake a study on the Y chromosome to de-

termine objectively the origins of Columbus. It is widely accepted that Columbus was an Italian sailor, but some historians propose that he was of Spanish or Portuguese nationality.

On June 3rd, 2003, the purported remains of Christopher Columbus and those from his son Hernando Columbus (Cristóbal Colón and Hernando Colon in Spanish, respectively) were exhumed in the Cathedral of Seville in Spain, and moved to the Department of Legal Medicine of the University of Granada. Prior to receiving the alleged Columbus bones (September 2002), the remains of the brother of Christopher Columbus, Diego Columbus, had been exhumed from its grave in Santiponce, a village close to Seville (fig. 3).

DNA analysis was performed in parallel by teams of the Universities of Barcelona, Santiago and Granada (Spain), University of Tor Vergata (Italy) and the Max Planck Institute for Evolutionary Anthropology. The consensus results showed a match in the sequence of mtDNA analysis of the remains of Diego Columbus with the ones thought to be of his brother, Christopher Columbus. Although the anthropological analysis was very limited because of the size and degradation of the remains, the analysis also supported that the remains could be those of Columbus.

Nevertheless, it must be mentioned that the set of remains buried in Seville and identified as the ones from Columbus are just a part, probably no more than 30–40% of the whole remains that could arise from 1 person. It is therefore logical to consider that the remains claimed to be that of Christopher Columbus and buried in Santo Domingo, Dominican Republic, could also be from the Admiral and conqueror. DNA analysis should be authorized by Dominican authorities and is pending on the samples from the Dominican Republic.

Regarding the origins of Columbus, we have performed a study on Y chromosomes from people named 'Columbus' in Northern Italy (the regions of Liguria and Lombardy), and people named 'Colom' in the regions of Catalunya, Balears and Valencia (Spain) and also some 'Colom' living in the Mediterranean French coast. The conclusion of this study [19] shows that the Y chromosome from Italian Colombo men was more diverse than the Iberian Colom ones. In Italy, the Colombo surname arose many different times. Genetic and lineage diversity were greater in Lombardy (probably because the name used to be given to orphans and foundlings in Milan) and less pronounced in Liguria and Piedmont. This suggests that Columbus may be of Catalan descent.



Fig. 3. Bone remains of Christopher Columbus relatives.

Additionally, this study shows that Colombo and Colom are two distinct surnames with no clear genealogical connection with local origins in Italy and Spain. However, Colombo is more frequent in Italy (where it is actually the most frequent surname in Lombardy) than Colom in Spain, and thus the number of Italian samples studied might not be equally representative. Since neither the Y-chromosome DNA of the remains of Christopher Columbus nor of his son Hernando has yet been analyzed, no conclusions regarding the origins of Columbus have been proffered by us.

In principle, other sets of markers can be used to attribute an individual to a population, and, given the vast number of markers, autosomal SNP arrays could be used. But the close similarity of the Catalan and North Italian general populations, as well as the poor suitability of this type of marker for ancient samples should be taken into consideration. For this reason, assessing and assigning the Columbus origin will be complex.

Future Needs and Developments in Forensic DNA Analysis

Human remains are exposed to a myriad of environmental challenges, as occurs with all forensic biological evidence. The result is that samples can be limited in

quantity, degraded and contain contaminants that affect the ability to type DNA samples. The often compromised samples that are encountered in missing person identification are bones, teeth and hair, as well as poorly preserved corpses, such as those found in water, buried or burned. It is very challenging, if not impossible, to extract genetic information from severely compromised DNA samples.

To expand DNA typing capabilities on such challenging samples, efforts are underway to extract more DNA from these materials, repair damaged DNA to generate more viable templates, improve PCR conditions to overcome troubling stochastic effects and develop alternate protocols for enhancing the sensitivity of detection. Understanding the mechanism of degradation processes can provide insight into avenues for repairing DNA. Currently, attempts to repair DNA have come primarily from the ancient DNA arena where there is a greater chance that the bases in the DNA have been chemically altered. Additionally, New England Biolabs has produced a kit for repairing DNA that is being evaluated by our laboratories [20].

As mentioned above, the standard operating procedures employ a set of STRs, as well as Y STRs and mtDNA. mtDNA typing is invaluable to missing person identifications, but it is a laborious, costly and time-consuming process. New technology exploiting electrospray

ionization mass spectrometry with the PLEX-ID system (Abbott) enables automated, high-throughput mtDNA analysis [21]. A power of discrimination, approaching that of Sanger's sequencing, is achieved but with reduced costs. In addition, some of the vagaries of mtDNA sequencing, such as the inability to analyze heteroplasmic regions and mixtures, are overcome with the PLEX-ID [21].

While the autosomal and lineage markers provide a high level of diversity and resolution for identity testing, severely compromised samples may not contain sufficiently long enough template molecules to yield results with the current format kits. Therefore, alternate genetic markers that may be more applicable to much degraded DNA are being sought. SNP variation is restricted to a small site of the genome. Thus, the amplicon generated that captures the SNP can be smaller in size than those generated for STRs. Because it is feasible to reduce the amplicon size for SNP typing to only 60–80 bp in length [22], much more degraded DNA samples can be typed than with the mainstay STRs. A single SNP is not as informative as a single STR locus; most SNPs are bi-allelic. However, technology exists to multiplex a large number of SNPs to obtain identity testing power equivalent to that afforded by using multiplex STR kits.

The SNPs can be divided into five classes [23–25] based on their application: (1) *Identity-Testing SNPs*, which are those that have the desired features of high heterozygosity and low population heterogeneity [26, 27]; (2) *Lineage-Informative SNPs*, which are sets of tightly linked SNPs that function as haplotype markers, or as pseudo-STRs, and are particularly useful for kinship analyses [28]; (3) *Ancestry-Informative SNPs*, which are SNPs that differ substantially in frequency in population

groups and can be used to reconstruct an individual's biogeographical ancestry; (4) *Phenotype-Informative SNPs*, which are SNPs that can be used to directly reconstruct an individual's phenotypic characteristics, such as skin, hair and eye pigmentation, height and facial features [29], and (5) *Pharmacogenetic SNPs*, which are SNPs that can be used to determine cause and manner of death based on an individual's genetic predisposition to triggering risk events. The forensic DNA field is investigating these markers, and we fully expect SNPs to be an important part of the forensic DNA repertoire in the not too distant future.

Lastly, the capabilities of next-generation (or actually better termed current-generation) sequencing have improved and costs have dropped dramatically. The examination of large parts of the genome will make interpretation of difficult mixtures easier and will facilitate research to identify the next generation of forensic markers. Already whole genome sequencing has been shown to be effective for microbial forensic investigations [30]. However, with these more-resolving and greater-depth tools, there is a concomitant gathering of private or personal data. Serious thought should be given to the degree of information that should be typed and/or disclosed.

The use of both new alternative genetic markers and novel technology platforms will increase abilities to type biological evidence, make stronger associations with kinship analyses and allow for greater use of genetic databases. These are exciting times for the forensic science field. The needs to enhance typing are likely to continue to drive innovation, and the field will continue to grow and become integral to more investigations.

References

- 1 Edwards H, Gotsonis C: Strengthening Forensic Science in the United States: A Path Forward. Washington, National Academy Press, 2009.
- 2 de Vargas Wolfram E, de Carvalho FM, da Costa Aguiar VR, De Nadai Sartori MP, Hirschfeld-Campolongo GCR, Tsutsumida WM, et al: Simplified buccal DNA extraction with FTA® Elute Cards. *Forensic Sci Int Genet* 2009;3:125–127.
- 3 Carpi FM, Di Pietro F, Vincenzetti S, Mignini F, Napolioni V: Human DNA extraction methods: patents and applications. *Recent Pat DNA Gene Seq* 2011;5:1–7.
- 4 Kimpton CP, Gill P, Walton A, Urquhart A, Millican ES, Adams M: Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Methods Appl* 1993;3:13–22.
- 5 Phillips C, Fernandez-Formoso L, Garcia-Magariños M, Porras L, Tvedebrink T, Amigo J, et al: Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel. *Forensic Sci Int Genet* 2011;5:155–169.
- 6 Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA: Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: The AmpFESTR® identifier® PCR amplification kit. *J Forensic Sci* 2004;49:1265–1277.
- 7 Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, et al: Validation of a 16-locus fluorescent multiplex system. *J Forensic Sci* 2002; 47:773–785.
- 8 Cotton EA, Allsop RF, Guest JL, Frazier RRE, Koumi P, Callow IP, et al: Validation of the AMPFESTR® SGM Plus™ system for use in forensic casework. *Forensic Sci Int* 2000; 112:151–161.

- 9 Alvarez JC, Johnson DLE, Lorente JA, Martínez-Espin E, Martínez-González LJ, Allard M, et al: Characterization of human control region sequences for Spanish individuals in a forensic mtDNA data set. *Leg Med (Tokyo)* 2007;9:293–304.
- 10 Carracedo A, Bär W, Lincoln P, Mayr W, Morling N, Olaisen B, et al: DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. *Forensic Sci Int* 2000;110:79–85.
- 11 Lorente JA, Entrala C, Alvarez JC, Arce B, Heinrichs B, Lorente M, et al: Identification of missing persons: the Spanish 'Phoenix' program. *Croat Med J* 2001;42:267–270.
- 12 Lorente JA, Entrala C, Alvarez JC, Lorente M, Arce B, Heinrich B, et al: Social benefits of non-criminal genetic databases: missing persons and human remains identification. *Int J Legal Med* 2002;116:187–190.
- 13 US Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention: National Incidence Studies of Missing, Abducted, Runaway, and Thrownaway Children (NISMA). 2002, www.ojjdp.ncjrs.org.
- 14 Lorente JA, Entrala C, Alvarez JC, Lorente M, Villanueva E, Carrasco F, et al: Missing persons identification: genetics at work for society. *Science* 2000;290:2257c–2258c.
- 15 Argac D, Bulbul O, Shahzad MS, Acar E, Altuncul H, Filoglu G: Optimization and validation of 10 mitochondrial DNA SNPs using SNaPshot Kit. *Forensic Sci Int Genet Suppl Ser* 2009;2:99–101.
- 16 Daniel R, Sanchez JJ, Nassif NT, Hernandez A, Walsh SJ: Partial forensic validation of a 16plex SNP assay for the inference of biogeographical ancestry. *Forensic Sci Int Genet Suppl Ser* 2009;2:477–478.
- 17 Dario P, Ribeiro T, Espinheira R, Geada H: SNPs in paternity investigation: the simple future. *Forensic Sci Int Genet Suppl Ser* 2009;2:127–128.
- 18 Fang R, Pakstis AJ, Hyland F, Wanga D, Shewale J, Kidd JR, et al: Multiplexed SNP detection panels for human identification. *Forensic Sci Int Genet Suppl Ser* 2009;2:538–539.
- 19 Martínez-González LJ, Martínez-Espín E, Álvarez JC, Albardaner F, Rickards O, Martínez-Labarga C, et al: Surname and Y chromosome in Southern Europe: a case study with Colom/Colombo. *Eur J Hum Genet* 2012;20:211–216.
- 20 New England BioLabs: NEBNext® End Repair Module. Instruction Manual. NEB #E6050S/L. Ipswich, New England BioLabs, 2010, pp 1–10.
- 21 Ivanov PL: A new approach to forensic medical typing of human mitochondrial DNA with the use of mass-spectrometric analysis of amplified fragments: PLEX-ID automated genetic analysis system (in Russian). *Sud Med Ekspert* 2010;53:46–51.
- 22 ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, et al: Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007;447:799–816.
- 23 Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, et al: Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 1987;235:1616–1622.
- 24 Richards RI, Sutherland GR: Heritable unstable DNA sequences. *Nat Genet* 1992;1:7–9.
- 25 Goedbloed M, Vermeulen M, Fang RN, Lembring M, Wollstein A, Ballantyne K, et al: Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFℓSTR® Yfiler® PCR amplification kit. *Int J Legal Med* 2009;123:471–482.
- 26 Adams SM, Bosch E, Balaesque PL, Ballereau SJ, Lee AC, Arroyo E, et al: The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *Am J Hum Genet* 2008;83:725–736.
- 27 Weber J, May P: Abundant class of human DNA polymorphism which can be typed using then polymerase chain reaction. *Am J Hum Genet* 1989;44:338–396.
- 28 Deforce DLD, Millecamps REM, Van Hoofstat D, Van den Eeckhout EG: Comparison of slab gel electrophoresis and capillary electrophoresis for the detection of the fluorescently labeled polymerase chain reaction products of short tandem repeat fragments. *J Chromatogr A* 1998;806:149–155.
- 29 Butler JM, Shen Y, McCord BR: The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J Forensic Sci* 2003;48:1054–1064.
- 30 Xue Y, Tyler-Smith C: The hare and the tortoise: one small step for four SNPs, one giant leap for SNP-kind. *Forensic Sci Int Genet* 2010;4:59–61.