CODIS AND THE PORTUGUESE DNA DATABASE: NOVEL INTERSECTIONS BETWEEN BIOLOGY, MEDICINE, INFORMATICS, MATHEMATICS, ENGINEERING AND LAW

António Amorim[1], Heloísa Afonso Costa[2], Paulo Ferreira Gomes[3], Francisco Corte-Real[4]

(1) António Amorim is geneticist at the Base de Dados de Perfis de ADN from the Instituto Nacional de Medicina Legal e Ciências Forenses, Lisboa, Portugal.
He is also professor of genetics at Department of Biology from Faculdade de Ciências da Universidade de Lisboa and teaches at Department of Biology from Universidade de Évora (e-mail: antonio.j.amorim@ inmlcf.mj.pt)
(2) Heloísa Afonso Costa is geneticist at the Base de Dados de perfis de ADN from the Instituto Nacional de Medicina Legal e Ciências Forenses, Lisboa, Portugal.
She is also PhD candidate at Department of Medicine from Universidade da Beira Interior (e-mail: heloisa.a.costa@ inmlcf.mj.pt)
(3) Paulo Ferreira Gomes is professor of law at the Universidade Autónoma de Lisboa, Lisboa, Portugal.
He is also law professor at Department of Forensic and Criminal Scienses from Instituto Universitário Egas Moniz (e-mail: psgomes@ autonoma.pt)
(4) Francisco Corte Real is professor of medicine at the Faculdade de Medicina da Universidade de Coimbra, Coimbra, Portugal. He is also President of the Instituto Nacional de Medicina Legal e Ciências Forenses and he was the first Director of Base de Dados de Perfis de ADN from the Instituto Nacional de Medicina Legal e Ciências Forenses, Lisboa, Portugal. (e-mail: francisco.m.cortereal@inmlcf.mj.pt)

ABSTRACT

At the scope of criminal and forensic investigations, the criminal authorities, health professionals, and many other intervenients can obtain different kinds of samples with biological material. The study of those samples by DNA laboratories can lead to the achievement of genetic profiles. CODIS and similar informatic DNA databases, by storing and managing those genetic profiles, can aid at criminal investigation by linking perpetrators to biological evidence samples. The same way CODIS can aid at human civil identification by linking relatives of missing persons to biological evidence samples. Thus, these novel intersections between biology, medicine, informatics, mathematics, engineering and law can lead to the resolutions of different kinds of crimes as well as other forensic matters such as human civil identification or clarification of kinship relations.

GENETIC PROFILE

Within the scope of human genetic identification, a genetic profile or the genetic fingerprint of an individual is the phenotypic description of a set of specific loci of the individual's genome with the ability to individualize that specific individual, which means that ideally a genetic profile must be unique. Those loci of the genome used to obtain genetic profiles are usually referred as genetic markers, and their phenotypic description take the form of a specific sequence of numbers, wherein each number indicate the number of repeats of a given DNA fragment of the genome. The genetic markers, of which polymorphic nature relies in the number of repeats of a given DNA fragment, are called STR, from short tandem repeats. Moreover, and in accordance with international recommendations, and particularly with recommendations of the European DNA Profiling Group (EDNAP), currently, the genetic markers that allow the definition of a genetic profile suitable for the identification of an individual are the aforementioned DNA fragments - STR -, generally with 4 to 7 nucleotides, whose variants or alleles are nominated by the number of tandem repeats they present.

GENETIC MARKERS/STR SETS

The first recommended set of STR for human identification in forensic genetics and legal medicine was introduced in 1994, included 4 STR - TH01, FES/FPS, vWA and F13A1 -, and has associated a match probability of about 1 in 10 000 individuals. In 1996 was presented a second generation of STR sets for human identification with 6 STR - TH01, vWA, FGA, D8S1179, D18S51 e D21S11 -, with expected lower match probability of about 1 in 50 000 000 individuals. In 1997, in USA, FBI Laboratory's introduced a 13 STR set, presented in table 01, called CODIS set, with a match probability higher than 1 in 1 000 000 0000 000 individuals. In alignment with USA, the European Union, in 1999, recommends a set with 7 STR - TH01, vWA, FGA, D8S1179, D18S51, D21S11, D3S1358 -, called ESS, from European Standard Set, and extended in 2009 for 12 STR, presented in table I, with a match probability higher than 1 in 1 000 000 0000 000 individuals. Currently, Portuguese official forensic laboratories use a set of 23 STR to define genetic profiles, as presented in table I. Given that the World population is about 7 000 000 000 individuals, theoretically, using at least the same number of CODIS set or ESS genetic markers to define genetic profiles, the possibility to find two individuals in world population with the same genetic profile is almost null. Thus, in human genetic identification we are working with a high level of assurance.

THE CODIS SOFTWARE, DNA DATABASES AND THE PORTUGUESE NATIONAL DNA DATABASE

FBI Laboratory's, in 1990, began a pilot software project, serving 14 state and local laboratories - the Combined DNA Index System (CODIS) software, that organizes and compare genetic profiles. Currently, this software is used by over 190 public crime laboratories in the United States and in over 50 countries worldwide that use it on their own database initiatives. Portugal is one of those countries.

In Portugal, in 2008, the law - Lei n. º 5/2008, modified by Lei n.º 90/2017 - stablished a national database of genetic profiles - Base de Dados de Perfis de ADN -, for criminal and civil purposes, that is supported by CODIS software and is housed and managed by Instituto Nacional de Medicina Legal e Ciências Forenses (INMLCF). In the role of the mentioned Base de Dados de Perfis de ADN, the Portuguese law also defined between all studied STR which of them have mandatory insertion character and which of them have complementary insertion character at CODIS. At table I we indicate which STR are of mandatory insertion character and which are complementary in Portuguese Base de Dados de Perfis de ADN.

DNA profiles from different types of biological specimens are entered into separate compartments in the CODIS software. By law, no personal data (name, date of birth, …) for any specimen is stored in the CODIS database. This ensures the privacy any person whose sample may be housed in CODIS. The DNA profiles in each index are then compared to profiles contained within the other indexes in an attempt to identify direct matches or potential relationships (genetic associations indicated not by exact profile matches, but by shared genetic data).

One of the most used and useful Index from CODIS is the Convicted Offender Index. At this Index are stored genetic profiles obtained from buccal swabs or blood stains of individuals convicted of criminal offenses. Depending on the country, and according to each country specific legal resolutions in force, only qualified convicted offenders are inserted on CODIS. In some countries the qualification arises from the kind of crime committed - theft, murder, sexual assault,…-, and in other countries the qualification arises from the penalty measure applied by the court to the offender. In Portugal, the convicted offenders housed in this CODIS Index are all convicted offenders with convictions equal or higher than 3 years. The main indexes and housed samples at CODIS software are presented at table II. CODIS software versions are updated periodically and provided to all CODIS laboratories by the FBI. As pointed previously, names and other personally identifiable information of the donners of the samples that originate genetic profiles are not stored at CODIS software although they are managed by the laboratories that provides the insertion of genetic profiles at CODIS software.

TABLE II - CODIS INDEXES

CODIS *INDEXES*	Genetic profiles
Convicted Offender	DNA profiles of individuals convicted of crimes
Arrestee	DNA profiles of arrested persons (if state law permits the collection of arrestee samples, as happens in Portugal)
Forensic	DNA profiles developed from crime scene evidence, such as semen or blood stains attributed to the putative perpetrator of the crime
Unidentified Human Remains	DNA profiles developed from unidentified human remains
Biological Relatives of Missing Persons	DNA profiles voluntarily contributed from relatives of missing persons
Missing Persons	DNA reference profiles from missing persons

TABLE I - GENETIC MARKERS INCLUDED IN DIFFERENT SETS

Genetic markers/STR	STR sets		
	PORTUGAL 2018	**ESS** 2009	**CODIS** 1997
Amelogenina (a)	•	•	•
D3S1358 (a)	•	•	•
vWA (a)	•	•	•
D16S539 (b)	•		•
CSF1PO (b)	•		•
TPOX (b)	•		•
D8S1179 (a)	•	•	•
D21S11 (a)	•	•	•
D18S51 (a)	•	•	•
TH01 (a)	•	•	•
FGA (a)	•	•	•
D5S818 (b)	•		•
D13S317 (b)	•		•
D7S820 (b)	•		•
D10S1248 (b)	•	•	
D1S1656 (b)	•	•	
D12S391 (b)	•	•	
D2S441 (b)	•	•	
D22S1045 (b)	•	•	
SE33 (b)	•		
D2S1338 (b)	•		
D19S433 (b)	•		
PentaE (b)	•		
PentaD (b)	•		

(a) Mandatory insertion
(b) Complementary insertion

CONCLUSIONS

CODIS software is a novel tool in criminal investigation and civil identification fields revolutionizing and eventually leading to refocus the paradigm of national security and criminal investigation authorities. As larger as DNA databases are, crime rates may reduce, especially in categories where forensic evidence is likely to be collected at the scene - murder, rape, assault, vehicle theft, among others -. The probability of arresting a suspect in new crimes falls as databases grow, likely due to dissuasion effect.

Forensic DNA databases have the potential to prevent and detect crime, nevertheless, there is a need for great public and policy debate as DNA databases are expanding around the world, in order to warranty some safeguards, global standards and societal engagement on the debate.