Title:

Building of the World's Largest DNA Database: the China Case

Author and affiliation:

Ausma Bernotaite; Griffith University

Abstract

Started only in 2005, China has already entered 68 million profiles into its National DNA Database (NDNAD), according to the data presented by NDNAD governing agency—the Chinese Institute of Forensic Science, Ministry of Public Security, at the Asian Forensic Sciences Network in September 2018, Beijing, China. Following the data presented by the same government agency at the International Society for Forensic Genetics in 2017, Seoul, Korea, the database grew to the said number from a reported 55 million profiles in less than one year. This number implies that by the number of profiles entered, the Chinese NDNAD is by far the largest in the world, followed by 17,530,781 profiles in the National DNA Index of the United States of America in 2018 and 6,024,032 profiles in the NDNAD of the United Kingdom in 2017. Additionally, National Missing Children DNA Database was created in 2009 to include genetic data from parents and their children. While large in the mere count of profiles included, it currently covers a relatively low percentage of approximately 4.5% of the Chinese population, and has not provided globally comparable match, or hit, rates for criminal case inquiries. Despite its rapid expansion and swift adoption of emerging genetic technologies, little effort has been made to share the developmental details of the database. This chapter serves as a short summary to trace the development timeline, goals, technological applications, main actors as well as the biggest achievements and future plans of the Chinese NDNAD. It is concluded with the emphasis for the need to ground national databases in strong socio-legal considerations.

Introduction

The DNA database has become the most important technological tool in the field of criminal technology in China. [1]

Collecting and analysis DNA evidence with the assistance of DNA databases is rapidly becoming a part of routine criminal police workflow globally. Started only in 2005, Mainland China[1] has already entered 68 million profiles into its National DNA Database (further referred to as the China's NDNAD), according to the data presented by China's NDNAD governing agency—the Chinese Institute of Forensic Science (IFS), Ministry of Public Security, at the Asian Forensic Sciences Network in September 2018, Beijing, the People's Republic of China. This makes the China's NDNAD the largest DNA database globally by the number of profiles entered, with the USA and the UK ranking second and third respectively. The largest growing portion of the database is the reference profile portion, which constitutes over 95% of all profiles preserved. Apart from the NDNAD, China also has a

[1]In this chapter we refer to "China" as the region of Mainland China without the areas of Hong Kong or Macao, as these areas have independent criminal justice governance mechanisms.

Y-Chromosome STR Haplotype Reference Database (YHRD), as well as National Missing Children DNA Database (NMCDD) founded in 2009. The IFS responsible for governing the NDNAD, the NMCDD as well as more around 600 government DNA laboratories of Provincial, Prefectural and County levels. The complex forensic database system employs a top to bottom governance mode across a growing network of forensic laboratories.

Despite of the rapid growth of the Chinese DNA databases, little is known about the governing principles of their implementation and management internationally. Nationally, there is also little debate in the country due to the authoritarian nature of its governance, where important decisions are made by the Party-state and then trickled down to provincial levels for implementation. Discussing the governance of China's genetic databases provides an interesting case study: in a country without free speech and run by one Party, which is the sole actor taking and implementing decisions for public security, and where public debate is not encouraged, to say the least. The government's ability to make quick decisions without reaching public consensus has been one, if not the most important, factor determining the speedy adoption of new technologies within the criminal justice systems.

To shed light on how the intersection of big data, genetics and criminal justice works in China, this chapter will provide a basic comprehensive overview of the history of DNA databasing internationally and in Mainland China, types of databases created, technologies used, standards followed, their relevant governing agencies and outlines the main points of contestation. Conference outcomes, articles published by government representatives and texts reflecting the views of government representatives are used where possible to give an overview of the blueprint that the Chinese government has followed in the adoption of the NDNAD. As China's forensic DNA governance is centralized and mainly hidden from the public eye, texts released by the governing agencies are the most direct reflection of what work has been completed and what developmental plans are remaining. The chapter is concluded with the overall evaluation of the forensic database governance system in China.

The international context

Technology and big data is becoming increasingly intertwined with our everyday lives, and this also applies to criminal justice. Digitalized forms of fingerprints, DNA, facial scans and criminal records—all of these examples of digitalized personal data used for policing are increasingly becoming the standard for efficient implementation of criminal justice. Following the dramatic drop in cost of the genetic analysis technologies, forensic DNA databases have boomed in popularity. A well-managed DNA fingerprinting database is a powerful and cost-effective tool for untangling crime cases, and has been proven to deter crime and reduce crime rates [2]. The most recent global report on the use of DNA profiling use published by Interpol in 2008, reported that 54 countries have national DNA databases and 120 countries use DNA profiling as evidence in criminal investigations [3]. There have been many more databases created, but there has not been a more recent global summary publication since then. A strong effort to collect comprehensive information, not only about the presence and size of DNA databases, but also on the local public discussions around them has been made by the Forensic Genetics Policy Initiative - a collaboration of three charity organizations namely, GeneWatch UK, Privacy International and the Council for Responsible Genetics. The organization estimates that there currently are sixty countries with operational and at least thirty-four with planned DNA databases globally [4]. The same source reports that there have been proposals of whole-population database creation. Tracy and Morgan called this type of DNA database expansion plans as "DNA fever", and reasoned that whole-population databases are a financial burden on the government that does not provide statistically valuable outcomes and instead creates a Big Brother state, while potentially violating constitutional rights to personal privacy [5]. Following the growth of forensic database adoption around the world, the scope of public and academic discussions on this multifaceted topic has also grown.

Just like any complex method of policing, DNA databases require to find the right balance between both ensuring the implementation of justice and individual privacy protection. While the safety and security of the people in the country is the obvious goal of all forensic DNA databases, governing authorities must take into account a variety of intersecting factors of database operation ethics. To explain this in depth, we can revert to the history of the two other DNA databases that are currently second and third after China's NDNAD in size. As mentioned in the introduction, the USA and the UK have the two largest databases in the Americas and Europe respectively, not only by their size but also by the percentage of the population covered. According to the Federal Bureau of Investigation website, the USA, the National DNA Index had a total of 17,785,074 profiles in October 2018, which is approximately 5% of the total population [6]. The National DNA Database of the United Kingdom held an estimate of 5,258,600 DNA profile records in March 2017, that number accounting for an approximate 8% population coverage; the match rate, or "the chance that a crime scene profile, once loaded onto NDNAD, matched against a subject profile stored on NDNAD" [7], of the database is high at 66% for the reporting year of 2016/2017 [7]. The National DNA Database in the UK is also the oldest one, and one that has undergone detailed scrutiny by the public. The UK's database had allowed for overly loose profile inclusion regulations, which are now partially addressed by the state: the Criminal Justice & Police Act 2001 decreased the levels of authority requirements for taking of DNA samples without consent, and eliminated the need to remove the DNA data and samples collected from the database after the end of investigation [8]. There has been a strong public backlash against the new regulations, which culminated in the case of S. and Marper v. the UK in European Court – the first of its kind. In 2008, the Grand Chamber of the European Court of Human Rights passed a judgement ruled that the law in England and Wales breached the European Convention on Human Rights. Following the judgement, the UK removed a total of 7,753,000 DNA profiles (including over 1.7 million DNA profiles taken from innocent people and from children) from its database without an observance of reduced crime detection rates [9]. The match rate of the NDNAD has seen stable increase despite the removal of a large number of profiles [7]. Following the UK NDNAD's privacy dispute, „a growing global consensus on the need for legislative provisions for the destruction of biological samples and deletion of innocent people's DNA profiles" was gaining speed [10]. A second question brought to the public eye in the recent years is the issue of racialized inequalities of DNA data inclusion and consecutive criminalization disproportionately increasing the policing of black men observed but not resolved [11]. Randerson estimated a staggering number of 77 percent of black men between the ages of 15 and 34 had a profile on the DNA database, and up to 37 percent of the total DNA database was constituted of profiles of black men as compared to 13 percent of Asian and 9 percent of white men [12]. Skinner [11] raised concerns over potential evidence of racism in DNA profile inclusion resulting in ethnic monitoring by the police.

In the USA, DNA evidence became increasingly used in the 1980s but it was not until 1994 that the Federal Bureau of Investigations formalized the Combined DNA Index System (CODIS) database, which serves as a joint platform for DNA profile sharing between states [13]. In terms of racial inequalities, the USA database has witnessed criticisms of similar nature as to those of the UK's NDNAD: Cole [14] traced archival knowledge of black policing in the USA and expressed concern over the racial discrimination in the database reproduced by historical racist biases. A further criticism of the CODIS pointed at the expansive policies of databasing and referred back to the *Marper* case against the United Kingdom and the consecutive ruling of the European Court of Human Rights:

As federal and state legislation in the United States become increasingly expansive, legislators and law enforcement need to be reminded that just because they have the technology to do something does not mean they should do it. [14]

Despite the public criticisms, the two databases have proven success in assisting the solving of crimes. In the annual UK's NDNAD report the National DNA Database Strategy Board provided the following statistics: "In 2016/17 the NDNAD provided 31,743 routine matches, including to 514 homicides and 612 rapes, and 493 urgent matches, including to 122 homicides and 141 rapes. This is an increase in the number of matches reported in the previous year, demonstrating the continued effectiveness of the NDNAD. The percentage of crime-scene profiles which matched a subject profile on load to the NDNAD (referred to as the match rate) was 66%." [7]. In the USA, the official FBI's website provides wider statistics: "As of October 2018, CODIS has produced over 440,346 hits assisting in more than 428,808 investigations" [6]. From the joint historical and modern experiences of the two largest DNA databases, following the Chinese one, suggest that DNA evidence can serve as an important tool in procedural justice by enabling the police to bring serious offenders to justice faster as well as ability to prove suspects innocent. While acknowledging the benefits of DNA databases, it is also important to recognize the fact that the use of technology does not happen in a socio-political vacuum. In the light of multifaceted considerations of DNA database ethics, it is of crucial importance that ongoing and transparent efforts are made to ensure that the genetic databases are managed ethically. To borrow the words from the Forensic Genetics Policy Initiative mission statement: "An appropriate middle ground between the legitimate needs of law enforcement and a respect for individual rights is achievable." [4]

Breaking the ground for the Chinese DNA database

The Chinese DNA databases have grown immensely with the advancement and scaling of genetic technologies. In the recent years, we have seen more publications talking about the history of this - now grandiose - public security project. The process of setting up the database has been top to bottom: led by a small group of experts from the Ministry of Public Security, and then passed down to provincial public security branches. The initial small DNA database was created to protect children from going missing - still a serious issue in modern-day China. Ge, Liu and Peng suggest that there have been four main stages of DNA database development: technological observation, active search of suitable technical DNA database implementation strategies, promoting its creation and the stage of rapid development [15]. The authors shared a few landmark events on the road to NDNAD's establishment: the idea to create a „DNA Fingerprinting Database" was first introduced in a meeting co-organized by the Chinese Ministry of Public Security and the Ministry of Science and Technology, the second time at „The First Academic Conference On Forensic Material Evidence", and was finally decidedly approved for government use for the 9th five-year plan, for the years 1996-2000. An IFS publication echoes the dates and adds commentary on what exactly was achieved in each year:

"With reference to the advanced experience of developed countries, the Physical Evidence Identification Center, the Ministry of Public Security, first proposed the establishment of an experimental DNA database in China in 1996. "Forensic DNA Quality Control Technology and DNA Database" was a complex project that combined quality control and database technologies. It set the ground for the standardization of forensic DNA testing through systematic and pioneering research. In terms of quality control, the project completed a standard named "Forensic DNA Laboratory Quality Assurance System" as well as 6 standards for forensic DNA testing and DNA database." [16]

Following the setting of basic standards, various provincial laboratories started initial work with small scale databases of their own: for example, the North Eastern city of Liaoyang built a small database of 2'211 Han Chinese people with 14 preliminary gene targets in 1999 [17]. The initial goal of the DNA database was to solve child kidnapping cases and in the year 2000, five government laboratories securely connected their database DNA data of 48'000 profiles enabling the solving of 418 missing children cases. In the year 2003, the number of profiles in the shared databases was 65'012: 11'685 of those were criminal profiles, 3'737 profile retrieved from crime scenes, 37'600 profiles of missing people and 11'990 profiles of missing persons family members [16]. The same source points out that after being set up for over two years, the database was helping solve serious crimes, such as murders, attempted murders and rapes. Up until the year 2005, preparatory work in different cities continued, followed by an ordinance from the Ministry of Public Security for database construction issued in 2005.

Thereafter, the work continued nationally: new laboratories were established, database capacity expanded and work regulations set.

Types of DNA databases in China

Over a seemingly short period of time, China has adopted a wide range of social surveillance technology measures. While it is the current owner of the largest DNA database in the world, the personal data collected in China is not limited to that. Currently available databases include: facial recognition, fingerprint, STR (direct and familial searches) and Y-STR profile (direct or mass screening) databases.

Picture 2. A police station at a crossroads at an Eastern city of Ningbo using a facial recognition and traffic control robot connected to the digital database and an LED screen displaying the faces and surnames of people crossing the street on red light.

There are three different DNA databases that are being run in China: the National DNA Database *(中国法庭科学 DNA 数据库)*, the National Missing Children DNA Database (NMCDD) *(全国失踪儿童 DNA 数据库*) and the most recently added Y-STR database*(Y-STR 数据库)* adjoined with the NDNAD. By the number of profiles, both the NDNAD as well as the Y-STR database are the biggest databases of their own kind in the world. All of the databases are run by the Ministry of Public Security of the People's Republic of China.

During the 5th National Congress of Forensic DNA Database in 2017, a presentation on the overview of forensic DNA technologies provided useful insights into the growth of forensic DNA use in the country [18]. As of September, 2017, there were 598 forensic DNA laboratories with 3’344 technical staff, and 1050 DNA analysis instruments. Data of crime cases solved with the help of the database for the years 2011 to 2016 was presented [18]:

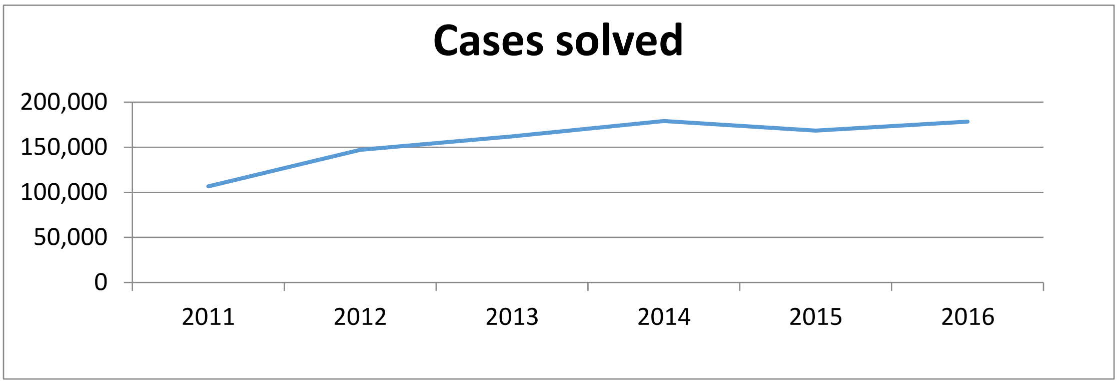

Figure 1. Cases solved with the help of the NDNAD.

The overall numbers of profiles entered into different for all the provinces of China were:

Province/Region	Number of profiles	Province/region coverage percent	Province/Region	Number of profiles	Province/region coverage percent
Beijing	217390	1.11%	Hubei	1946219	3.40%
Tianjin	213068	1.65%	Hunan	2181102	3.32%
Hebei	1352379	1.88%	Guangdong	1707206	1.64%
Shanxi	919004	2.57%	Guangxi	1593857	3.46%
Inner Mongolia	354541	1.44%	Hainan	199495	2.30%
Liaoning	930050	2.13%	Chongqing	1504827	5.22%
Jilin	1003642	3.65%	Sichuan	2746677	3.42%
Heilongjiang	1855512	4.84%	Guizhou	1553093	4.47%
Shanghai	58968	0.26%	Yunnan	879757	1.91%
Jiangsu	2997317	3.81%	Tibet	2625	0.09%
Zhejiang	1236539	2.27%	Shaanxi	1332932	3.57%
Anhui	2096514	3.52%	Gansu	981509	3.84%
Fujian	1119043	3.03%	Qinghai	224768	3.99%
Jiangxi	1596364	3.58%	Ningxia	129897	2.06%
Shandong	2392837	2.50%	Xinjiang	2851289	13.07%
Henan	4232469	4.50%			

Figure 2. Percentile population coverage in different provinces and regions of China

The total number of profiles in the database comes to 3.1% of population coverage, however it can be noted from the table that different provinces have varying coverage rates as each province is managed separately. It is worth noting once more that while increasing the percentile population coverage by the DNA database is prioritized, a very low number of those people included in the Chinese NDNAD are actual offenders [15]. According to the proceedings of the 5th National Congress of Forensic DNA Database [18], an overwhelming majority of 95% of the profiles in the DNA database are reference samples collected from target populations. The was target populations are set varies from province to province. The provinces of Xinjiang, Shandong, Henan, Jiangsu, Zhejiang, Guangdong and the autonomous region of Beijing had over 3 million profiles inserted into the database. A particularly strong percentile coverage rate of Xinjiang is by far the most notable. Xinjiang is the only case where a whole-population database was ordered by the authorities to be built [19]. The province of Xinjiang, or the Xinjiang Uyghur Autonomous Region, is one of the most politically sensitive regions in China. The region has a large population of Uyghur Muslims, and tensions between Han and Uyghur ethnicities in the region are common: reported surveillance and control, peaking in the recent Chinese government recognition of “re-education camp“ establishment for “thought reform“ of hundreds of thousands Uyghurs are justified by terrorism threats and the need to ensure the safety of the people [19]. This can be related back to the large percentile coverage of the Xinjiang overall percentile population coverage.

Picture 2. Polymerase Chain Reaction machines processing thousands of reference samples at the Center for Applied Forensic Sciences and Technologies in Beijing, China.

According to the September, 2017, presentation of IFS at the International Society for Forensic Genetics, 23.6% of crime scene samples match a person or another crime sample when entered into the database for the first time. For the China’s NDNAD, a clear goal outlined by the government on multiple occasions is to increase the database match rate between crime case profiles (match rate of evidence-individual) and profiles in the database by continuously increasing population coverage [20]. Liu, who is a representative of Institute of Forensic Science, has noted that China is lagging behind the percentile match rates of the UK and including more DNA profiles into the database would be a strong priority at IFS [20]. The author does not explain whether or not these profiles would be taken from offenders. While talking about the larger structures of running the DNA database, “the assessment of the state of public order has relied on two indicators: the number of files places on police files (li’an) and the police clear-up rate (po’an lu)” [21]. This partially explains the reasons of assigning high priority to increasing NDNAD’s match rate. Ethics and financial viability aside, the ability to quickly match a crime scene DNA profile to a known person in DNA database would dramatically cut down on the work of the police, thus raising the statistical value of police work quality.

Y-STR database was officially announced in September, 2012, as a response to the need for further DNA evidence to solve crime, and the statistical conclusion that most violent crime being committed by men. Ge et al [22] argue that since it is a well-established fact that men commit the overwhelming majority of violent crimes, male-specific Y-STR data can be used to rapidly and efficiently identify the perpetrators of crimes. Y-STR database has over 7 million entries. Four main uses of the Y-STR database have been outlined: determining connections between surnames and Y-STR heritage in regions with little migration, familial searching using autosomal and Y-STRs to determine the identity of criminals, determining the inference of source population, analyzing of mixed samples (especially rape samples with small trace of any present male DNA) and the searching of missing persons or the relatives of missing persons [23]. Determining statistically significant connections between Y-STR data and surnames, as well as being able to determine the inference of source population still need a lot published research from different areas in China. Currently, Y-STR database is identified as particularly important because genetic familial linkage information can assist the police in tracing close male relatives of a criminal in case an instant match is not available upon crime case profile search in the database. High discriminatory power and the advantage of being able to trace paternal lineage mean that while sometimes a database search might not provide a match, the related results would suggest a male family member of a crime perpetrator. In areas of China that have not seen much migration, genetic heritage is strongly connected and familial linkages can be quickly uncovered through data derived from Y-STRs, while maintaining sufficient discrimination power for direct comparisons of single-source samples as well. IFS confirms that, although increasing the number of crime scene and reference sample profiles in the database remains a continuous goal, familial searching, indeed, is being considered as a potential means of maximizing the efficiency of the database. In 2012, the work of building the Y-STR database was started in nine laboratories in six provinces, with the goal of further expanding and uniting this database. This type of database is not common in other places of the world, where most of the databases focus on non-gendered autosomal STR loci. China has taken a bold step to build a large national Y-STR database, starting with provincial crime laboratories. Since the Chinese judicial system is focused on being able to “break the case”, match rates are strongly emphasized as prime indicators of database success: they note that the gender distribution of crime statistics should be an important driving factor in Y-STR database creation [15].

The third type of the DNA database created in China is the National Missing Children DNA Database. Issues of missing children were the reason why the work of genetics was started in China is the late 90’s - NMCDD was the start starting point of all of the databases. The issue of missing children has started in the 70’s and there are around 200’000 of missing children reported each year, only 0.1% of them are found [23]. If thinking chronologically, the database for missing children was the first one to be established, however, it took a long time for the project to be approved as a stand-alone database. It was in 2009 that the Ministry of Public Security gave the go-ahead for the project.

“(...) the Fifth Bureau of the Ministry of Public Security will re-develop and construct the “National Public Security Organization DNA Database Information System for Search of Trafficked/Missing Children” and the “National Public Security Organization DNA Database Information System for

Search of Trafficked/Missing Children", which will be handed over to the Ministry of Public Security for evidence management and management. (...) Up to now, the library has accumulatively recorded 557,000 DNA data of abducted/missing children and their relatives. There are 4,420 trafficked/missing children that have been found, and the longest time that a child has been lost is 36 years" [15]

Chen notes that in the years of database running it has faced many issues ranging from ethical issues of DNA sample collection and management to the technical questions around the running of the database, and suggests that two main advancements - resolving ethical privacy issues as well as increasing the efficiency of the database - should be resolved for a more efficient running of the database [23].

China's NDNAD, Y-STR database and the NMCDD are the three main DNA databases in China. NDNAD has seen rapid growth aligned with the Party-state goals of achieving "harmonious society" and "peaceful development" of the country, further talked about in the next chapter. For the government to be able to maximize the benefits of the databases data sharing and integration has been promoted. During the Fifth Plenary Session of the 18th CPC Central Committee in 2015, the largest and most important strategic governmental session in China, a slogan to "Implement the national big data strategy, promote open sharing of data resources" was created to signal the larger direction of database development.

Forensic science governance in China: background, institutions and technologies

The People's Republic of China (PRC) is a Party-state governed by the Communist Party of China. In the Constitution, which is the main governing document of the country, it is stated that the PRC is a socialist state under the people's democratic dictatorship governed by the principle of democratic centralism [24]. This means that China's governance is highly centralized where decisions are made by the main governing body and then trickled down for implementation. Thus is also the case in criminal justice in general, as well as the management of the China's NDNAD in particular. All of the criminal justice institutions are managed by the Ministry of Public Security, which is the main police and public security authority in the People's Republic of China that ensures law enforcement, and is also the governing body of all of the Public Security Bureaus across the country.

Since the establishment the DNA database in 2009, it was run by the Ministry of Public Security. After the official establishment of the NDNAD in 2005 and National Missing Children DNA Database in 2009, the governance of these two databases was transferred to the Institute of Forensic Science in 2009. The institute, which is located in the capital city of Beijing, is part of the Ministry of Public Security and is the governing body directly responsible for the operation of DNA databases in the country as well as the independent advancement of local Chinese original DNA technology knowledge. The responsibilities that the institute officially assumes are: strategic management of the database, setting of Statement of requirements for the National DNA database, monitoring of Forensic Science laboratories, delivering DNA database services to police forces, ensuring integrity in the management and the DNA data held, development of database software, providing management information, such as standards and guidance, and delivering training. For these responsibilities, the IFS reports directly to the Ministry of Public Security.

China is also of the few countries that adjoined a research centre to develop own STR typing reagents as well as a genetic analyser. The First Research Institute of the Ministry of Public Security is the largest and the highest unit of forensic science research in the public security system. Founded in 1960, the First Research Institute is also one of the oldest scientific research institutions in China. It is the research and development base of modern police technology, and has long been responsible for the research and development of police science and technology. It plays a strategic role in China forensic science field as it pushes for China's technological independence. Accommodated within the First

Research Institute umbrella, the Center for Applied Forensic Sciences and Technologies develops and utilizes own technologies for reference samples inclusion in the NDNAD. Located on the outskirts of Beijing, the center has three functions: it is a research center, an STR kit manufacturing facility as well as a reference sample processing police DNA laboratory. During the fifteenth period, the Institute of Forensic Science of the Ministry of Public Security successfully developed domestic forensic DNA detection reagents, breaking the monopoly of foreign DNA technology products. During the 11th Five-Year Plan period, under the advocacy and care of the Ministry of Public Security, the First Institute of the Ministry of Public Security jointly developed the GA119-16A Genetic Analyzer and related analysis software and consumables with independent intellectual property rights, realized the complete supporting of domestic DNA testing reagents, domestic instruments and consumables. While not well known internationally, the analyzer is a groundbreaking invention as prior to that Thermo Fisher Scientific had been the only supplier of this laboratory equipment in the world. The genetic analyzer was developed to ensure the government's self-sufficiency, shortened investigative time and cost reduction.

Picture 3. A government representative outside of the Center for Applied Forensic Sciences and Technologies in Beijing, China.

Acquiring a software that would be able to assist the management of the DNA database has also been important to China. To be able to efficiently store, manage and share DNA data across the country, the Ministry of Public Security invested in different types of data management software, which have undergone several developmental stages. Representatives from the Ministry of Public Security Material Identification Center, have summarized the developmental pathway and main technologies used:

"In June 2009, in order to completely solve the increasingly apparent incompatibility between the DNA database system and the work requirements of the public security organs at that time, the Ministry of Public Security approved the "National DNA Database Upgrade and Renovation" [request]. In July 2014, the project passed the acceptance of the Ministry of Public Security. The project has completely solved the problems of the first generation DNA database level: many nodes, weak data flow monitoring, and slow comparison. The new generation of national DNA database structure integrates

Management Information System (MIS), Laboratory Information Management System (LIMS), DNA Data Retrieval Comparison System (DIS), and duplicate personnel information search. System, catastrophic accident identification system (DVI) and data conversion tools and a series of standard data interfaces, and design entry, storage, comparison, review, backup and other functions according to 100 million data scale. In December 2015, a new generation of "National DNA Database System" was used in 33 DNA laboratories in 9 provinces and cities including Beijing, Tianjin and Heilongjiang; in April 2016, 30 DNA laboratories in 7 provinces and cities including Shanxi, Inner Mongolia and Hunan. In June 2017, after nearly one year of actual test operation and continuous optimization and improvement, the Ministry of Public Security held a DNA Database Upgrade and Reconstruction Project meeting in Beijing, which officially launched the nation-wide construction and application of the second-generation national DNA database.The timeline for the replacement of the first-generation database is set as the end of this year. The second-generation "National DNA Database" will better support rapid DNA query, classification and other applications of massive data quantities." [15]

The DNA databases are governed by a set of regulations names "Criterion for forensic DNA database"(*法庭科学 DNA 数据库建设规范*). The set of regulations are fairly simple and even somewhat loose: while they do set the basic understanding of what elements constitute a DNA database, the document maintains ample space for the government to include, preserve and use DNA data as deemed necessary. While the size of the DNA database is growing exponentially, it is not entirely clear what standards the government is upkeeping. There are two widely internationally-accepted loci set standards: the expanded Combined DNA Index System (CODIS) loci standard is the United States [25] set by the Federal Bureau of Investigation and the European Standard Set (ESS) loci decided on by the DNA working group of the European Network of Forensic Science Institutes (ENFSI). In China the situation has been different, and it has become a source of criticism both nationally and internationally. Since the beginning of DNA database construction, the loci standards were not set. The difficulties arose in the very beginning the DNA database building as researchers discovered that due to genetic differences some loci were close to obsolete in Chinese populations, such as loci typical in Caucasian and Black populations [26]. As a result, different local and international manufacturers of the DNA technologies have followed own chosen standards, often the CODIS standard applied in the USA. Bing [20] outlined that due to lack of regulations, various DNA technology providers have come up with own versions of STR analysis kits causing a fair sense of havoc, and advised that despite the rapid growth of the database, too little thought was put into setting standards for the industry. In 2015, the Chinese NDNAD was undergoing a major technological upgrade, and grasping that opportunity, Bing urged the government to put further thought into learning from the experience of foreign scientists and considering various factors that would make a major difference in the application of the DNA database [20]. Currently, a new standard of 20 loci is being applied, although it has not been officially published.

The "outsourcing" of the DNA databases' governance responsibilities to one dedicated agency has proved to have increased the efficiency standardization. IFS has put efforts to increase the efficiency of the DNA database in terms of pushing for the development of DNA analysis technologies locally, upgrading the DNA analysis software across the country, and is aiming to unify the loci standards used by different laboratories. Further efforts should be invested toward the transparency of DNA collection rules and regulations as the current standards are loose.

Future directions

The Chinese DNA database has exhibited extreme growth and has shown voracious adoption rates of new technologies and heavy allocation of resources for innovative DNA research. The country has already implemented three types of DNA databases (NDNAD, NMCDD and Y database), fingerprint database, face image database, and has already planned for an SNP database, according to the announcement made during the 10th AFSN Annual Meeting in Beijing, China [27].

Next Generation Sequencing (NGS), also referred to as second-generation sequencing and Massive Parallel Sequencing (MPS), has risen in popularity in the Chinese forensic circles.During international presentations as well as published articles, IFS has indeed signaled that they would be ready to consider NGS technologies in the future: “Currently, MPS appears sufficiently robust to type reference samples for uploading DNA profiles into databases. With the technology evolving, it is likely that in the near term MPS will be able to offer the sensitivity of detection to analyze low quantity and quality DNA samples, and will be capable of analysis of forensic casework evidence.”[28]. Furthermore, during 2017 ISFG meeting, IFS representatives noted that the capacity of the Chinese database is ready to hold a large increase of data that using MPS technologies would require. The the AFSN event in 2018, SNP profile database was confirmed as planned for implementation.

The SNP database will be running on NGS equipment: a new type of genetic analysis equipment, also referred to as second generation sequencers, which is able to analyse large amounts of different types of genetic data at the same time. Once implemented, it will be one of the most technologically-advanced databases used around the world. In a recent publication by the IFS, the clear vision of how NGS would be used with the existing NDNAD and the biggest challenges were discussed: NGS would be incorporated into the existing NDNAD instead of replacing it and the critical challenge was to resolve compatibility issues between NGS-generated data and the currently used STR data [29]. The article comments: “From the perspective of its developmental history, the applications of forensic DNA databases have closely followed the development of DNA technologies applied in the field of biology. Thus, following the development of second-generation sequencing technologies and platforms as well as the development of many emerging biological genetic technologies such as SNP, mRNA, and whole genome sequencing, the next step in the development of DNA databases should be gradually combining the new and the currently used technologies to support better data use and retrieval” [29]. One of the main considerations prompting the adoption of NGS is the desire to make the most use of the currently existing data through increasing its efficiency. NGS could be a potential answer: as previously mentioned, it can analyze different types of data in the same reaction, thus increasing the accuracy of the profiles. It is recognized that the costs of running NGS tests as well as the need for the technology to mature and be fully compatible with the existing NDNAD as two main challenges to be solved before a large-scale adoption is undertaken [29]. The Beijing Institute of Genomics, Chinese Academy of Sciences, has taken an active role in NGS research in China, and they have identified and called for resolution of the following issues: low-template library preparation, error rate, type estimations, as well as existing problems with NGS data processing and mining [30]. Despite the existing issues and the high costs, it can be foreseen that the government will adopt the technology once all the issues are resolved.

China has successfully applied various emerging technologies to ensure the wider societal goals of ensure the stable growth and development of the country.With the NDNAD, Y-STR and NMCDD to manage and an articulated need to obtain China-specific DNA markers with high discrimination, MPS technologies do have the potential to unify all of the present needs of the police on one platform. This is one single large development, which is known to be under serious consideration by the highest governing units in China.

Issues pertinent to the DNA databases

“The problem of order is a genuinely transhistorical problem rooted in inescapable conflict between the interests and desires of individuals and the requirements of society”

Dennis Wrong, “The Problem of Order” [31]

China has made a strong push to ensure that forensic DNA technologies are an integral part of ensuring public safety and social stability. Managing around 600 laboratories across the country - their staff, buildings, technical and governance standards - is not an easy task. It is not surprising that the DNA databases have not been without issues. In a recent publication, the Physical Evidence Identification Center, the Ministry of Public Security, identified six main issues that the NDNAD is currently encountering: firstly, it is the lack of adequate management skills as well as the infrastructure and quality management capabilities. The database has been ordered to be created so quickly that personnel training has lagged behind and hard skills of DNA laboratory management have not caught up [15].

The authors also comment on the serious lack of both personnel and laboratory issues: “Some of the newly built, some of the rebuilt and some of the modernized and expanded DNA laboratories still have a low level of modernization and standardization. Some have just started work, and there is an obvious lack of functionality, laboratory room partition confusion, observed high risk for lab processes. National DNA database information quality monitoring can find contaminated data every month. Problems such as the failure of the quality management system to meet standards have seriously reduced the standardization, credibility and authority of inspection and appraisal. Among the 387 provincial and prefecture DNA laboratories participating in the 2016 Ministry of Public Security's proficiency assessment, 15 failed the results (2 provincial), and the rate of failed qualification assessment reached 3.87%.”. The second observed issue refers to the ability to collect physical evidence and handle samples. Ge at al [15] comment: “In some places, especially in areas where DNA laboratories have only recently been built, there is a significant lack of expertise, evidence handling awareness, technical preparation and operational skills in DNA technology. The on-site bio-material evidence extraction rate is far below the national average. Some provinces have more prominent problems, and the rate of on-site bio-materials extraction is less than 5%. In some provinces, the effective collection rate and sample size of DNA samples of eight key personnel such as crimes are very low, and repeated collections are at a high level for a long time. In some places, the problems of long-term backlog, failure to inspect samples or not inspecting samples in time are more prominent.”. Further, issues with ensuring adequate numbers of qualified specialists are pertinent. To comment in numbers: there were only 112 professional DNA technicians who qualified in 2006 national DNA laboratory assessment and the number of full-time DNA database administrators and DNA intelligence analysts who have mastered DNA testing techniques and surveyed on-site, familiar investigative work, and proficient in computer technology is extremely scarce [15]. With only the recent update of the DNA database management software, system security and risk management capabilities have been an issue. Although the DNA database operates on the public security network, there are still many risks in security management: in May 2017, after the outbreak of the "Eternal Blue" blackmail-born virus, DNA database server virus was found in 19 provinces and cities nationwide, causing system paralysis; the most serious consequence of DNA database virus was information loss that occurred in several individual cities [15]. Failure to upgrade the DNA software system in time was the primary reason for that, which could have been prevented with stricter management of DNA laboratories. Such large-scale database server virus infection is a serious red flag as it poses a serious security threat. The final issue has to do with the staff skills to handle incoming information and their ability to successfully use the information at hand while solving cases.

All six of the technical issues mentioned here are serious problems. It can be seen that only half of the laboratories that are being run employ staff that is fully capable to work in a forensic genetics laboratory. Consequentially, without strong support in a genetic laboratory, new staff without prior experience in the field would find the work very challenging. In 2017, the Ministry of Public Security launched “Three Strikes and One Remediation” campaign to resolve some of the most crucial of these issues; the results of this campaign are yet to be measured and evaluated.

Technical issues of the database is only one side of the coin. Ethical issues have also been pertinent to the DNA database: the steps that the country has taken around the building of its DNA database could be called somewhat radical: the country has taken big steps towards large-scale monitoring of its population without putting much effort towards transparency. Some of the measures taken to build the forensic DNA database are implements against well-researched experience from other countries. For example, in the UK the number of subject, or reference, profiles loaded to the DNA database has been decreasing from year to year but the efficiency measures by the match rate have increased. The percentile ratio of match rates as compared to the overall size of the Chinese DNA databases have not seen stable annual increases - in the year 2015 they had even decreased. Therefore, a possibility that the goal of "growing" the database by adding as many reference sample profile inclusions as possible does not result in increased efficiency. The reason of why so many reference samples are "pumped" in the name of the growth of the DNA database are unclear. Common points of critique are often also directed towards the over-simplistic governance standards that barely account for physical integrity of citizens, over-polices regular citizens without criminal records and infringes on their right to privacy.

Running the world's biggest DNA database: conclusion

The Ministry of Public Security has pushed the NDNAD to grow and become the world's largest DNA database in just over 10 years. Additionally, China's Y-STR database is now also the world's largest database of such kind. To achieve this goal, the Chinese government has taken some radical decisions in terms of fuelling the expansion of DNA laboratories without a set supply of specialists to work in them, lagging creation and implementation of standards and not publicly disclosing comprehensive uses of DNA databases. In some cases, work was started first and the relevant regulations followed after, creating concern around ethics of the sample collection, storage and use of the DNA profiles. Since there is a lack of transparency, it can be argued that an authoritarian country with so much power over its people would not necessarily only use all the data collected for the purpose of solving crime. The implications of owning and managing this size of a database are varied and extending. China must pay close attention to maintaining a balance between the legitimate goal of maintaining a safe and stable society, as identified as one of the major building blocks of the present political system, smart use of government funds as well as the ethics of forensics in the country.

In the future, the size of the NDNAD will grow not only in size but also in complexity with the use of Next Generation Sequencing technologies, as well as the addition of the Y-STR database. With the rise of facial recognition technologies, it can be speculated that at some point in the near future a crime could be solved by CCTV recognizing an offender and any DNA evidence conclusively confirming a crime. Overall, China is preparing to meet yet another five year plan in 2020 with its forensic technologies, laboratories, staff and standards mature for another challenge to come. In a way, much of the rushed growth can be explained with the goals of planned governance and central management. Only time will show how such planning would work out in the context of forensic DNA.

References

[1] Zhao, Z., et al., The Application of Next Generation Sequencing in Forensic DNA Database.

Science Discovery. 5(4)(2017), p. *257-261*

[2] Doleac, J. L., The effects of DNA databases on crime.

American Economic Journal: Applied Economics, *9*(1),(2017), p. 165-201.

[3] Interpol Global DNA Profiling Survey, 2008. http:// www.interpol.int/content/download/8994/66950/version/2/file/ GlobalDNASurvey.pdf; Full report on: http://www.dnaresource. com/ documents/2008INTERPOLGLOBALDNASURVEYREPORTV2. pdf.

(accessed 19 September 2018)

[4] The Forensic Genetics Policy Initiative.

http://dnapolicyinitiative.org/ (accessed 13 September 2018)

[5] Tracy, Paul E., and Vincent Morgan, Big brother and his science kit: DNA databases for 21st century crime control. *J. Crim. L. & Criminology* 90 (1999): 635.

[6] CODIS - NDIS Statistics. https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics (accessed 14 December 2018)

[7] National DNA Database Strategy Board Annual Report 2016/17. National Police Chiefs' Council and Home Office. (2018). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/724596/040718_new_CCS0518718592_National_DNA_Database_Strategy_Board_AR_2016-17_updates_NEW.pdf(accessed 13 September 2018)

[8] McCartney, Carole. Forensic DNA sampling and the England and Wales National DNA Database: a sceptical approach. *Critical Criminology* 12, no. 2 (2004): 157-178.

[9] Home Office National DNA database strategy board annual report. 2012-13 (2013). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/252885/NDNAD_Annual_Report_2012-13.pdf(accessed 13 September 2018)

[10] Wallace, H. M., A. R. Jackson, J. Gruber, and A. D. Thibedeau. Forensic DNA databases–Ethical and legal standards: A global review. *Egyptian Journal of Forensic Sciences* 4, no. 3 (2014): 57-63.

[11] Skinner, David. 'The NDNAD has no ability in itself to be discriminatory': Ethnicity and the governance of the UK National DNA Database. *Sociology* 47, no. 5 (2013): 976-992.

[12] Randerson, James. DNA of 37% of black men held by police. *The Guardian* 5 (2006).

https://www.theguardian.com/world/2006/jan/05/race.ukcrime (accessed 3 October 2018)

[13] Deray, Erica Solange. The Double-Helix Double-Edged Sword: Comparing DNA Retention Policies of the United States and the United Kingdom. *Vand. J. Transnat'l L.* 44 (2011): 745.

[14] Cole, David. Race, policing, and the future of the criminal law. *Hum. Rts.* 26 (1999): 2.

[15] Ge, J., F., Liu, J. Xu. History, current situation and future of DNA database construction in public security organs across the country in Ge J. (Ed.) DNA database construction, application results and prospects V. Qunzhong chubanshe, 2017, Beijing. [Chinese Mandarin]

[16] L., Hu, S., Chen, G. Zhang. Value and important issues about national DNA database, Foren Sci Technol, 6 (2003): 3-5. [Chinese Mandarin]

[17] Wang, W., H. Jia, Q. Wang, Z. Cai, L. Wei, D. Wang, A. H. Bittles. STR polymorphisms of "forensic loci" in the northern Han Chinese population. *Journal of human genetics* 48, no. 7 (2003): 337-341.

[18] Institute of Forensic Science. 2017. Working proceedings of the 5th National Congress of Forensic DNA Database. Ha'erbin, 2017 September 12-14.

[19] Human Rights Watch, China: Minority Region Collects DNA from Millions. https://www.hrw.org/news/2017/12/13/china-minority-region-collects-dna-millions# (accessed 2 October 2018)

[20] Bing, Liu, A Preliminary Evaluation of DNA Evidence in Criminal Investigations Based on National DNA Database, Foren Sci Technol. 40, no. 3 (2015): 199-203

[21] Sapio, Flora, Susan Trevaskes, Sarah Biddulph, and Elisa Nesossi, eds. *Justice: The China Experience*. Cambridge University Press, 2017.

[22] Ge, Jianye, Hongyu Sun, Haiyan Li, Chao Liu, Jiangwei Yan, and Bruce Budowle. "Future directions of forensic DNA databases." *Croatian medical journal* 55, no. 2 (2014): 163.

[23] Ge, J., J. Yan, Q. Xie, H. Sun, H. Zhou, B. Li. "Development of Chinese forensic Y-Str DNA database." *Fa yi xue za zhi* 29, no. 3 (2013): 212-5. [Chinese Mandarin]

[23] Chen, Zeyu. Application and Perfection of DNA database in the fight against child trafficking crime. PhD diss., Southwest University of Political Science and Law, 2013.

[24] Lin, Gang. Leadership transition, intra-party democracy, and institution building in China. *Asian Survey* 44, no. 2 (2004): 255-275.

[25] Hares, .Selection and implementation of expanded CODIS core loci in the United States. *Forensic Science International: Genetics*, *17*(2015): 33-34.

[26] Wang, Wei, Huiling Jia, Qin Wang, Zhiming Cai, Lihui Wei, Debing Wang, and Alan H. Bittles. "STR polymorphisms of" forensic loci" in the northern Han Chinese population." *Journal of human genetics* 48, no. 7 (2003): 337-341.

[27] Wang, Guiqiang. Forensics solving cold cases in China. 2018. 10th Asian Forensic Science Network Annual Meeting, 2018 September 4-8, Beijing, China

[28] Ge, J., Sun, H., Li, H., Liu, C., Yan, J., Budowle, B.. Future directions of forensic DNA databases. *Croatian medical journal*, *55*(2) 2014: 163.

[29] Zhao, Z., Wang, L., Li. S., Wang, T., Liu, B.The Application of Next Generation Sequencing in Forensic DNA Database. *Science Discovery.* 5(4) 2017: 257-261

[30] Yang, Y., Xie, B., Yan, J. Application of next-generation sequencing technology in forensic science. *Genomics, proteomics & bioinformatics*, *12*(5) 2014: 190-197.

[31] Wrong, D. *Problem of order*. Simon and Schuster. 1994.