# FauxDIS: an Interactive Online Forensic DNA Profile Database

**Ashley Hall PhD[1*], Jonathan Bisson, PhD[1]**

*[1]University of Chicago at Illinois, Department of Pharmaceutical Sciences, 833 S. Wood St, Chicago, IL 60612 USA.*corresponding author: amhall7@uic.edu*

**Abstract:** Forensic Science has captured our collective imaginations for generations, whether it be in the medical examiner's room with Quincy, examining blood spatter with Dexter, or in the crime lab with Forensic Files. With the right tools and applications, we can take advantage of this popularity and use forensic science as a vehicle to teach critical thinking skills and the scientific method, both of which are integral in the collection and analysis of forensic evidence. The forensic scientist makes observations, formulates hypotheses about the probative value of evidence, and tests these educated guesses by submitting crime scene samples to an operational forensic laboratory for analysis. With a DNA profile generated from crime scene evidence, the forensic scientist can conduct direct or indirect database searches in hopes of finding a match and learning the identity of the donor of the questioned sample. The U.S. national DNA database system, CODIS, contains millions of offender DNA profiles, but its use is restricted to authorized operational labs. Therefore, in this report, we introduce the FauxDIS DNA Database, a searchable online DNA profile database that is available to educators for use in experiential exercises such as mock crime scene analysis. The database currently contains autosomal profiles, but can be expanded in the future to contain other marker systems such as Y-chromosome short tandem repeats or massively parallel sequencing data.

**Keywords:** DNA database, DNA analysis, CODIS, scientific method, critical thinking

## Introduction

Forensic Science is everywhere – you can hardly search your TV offerings without encountering shows such as NCIS, one of the many iterations of CSI, or even a "so-new" Snapped. This fascination with crime and justice is not new (remember Quincy?); true crime and forensic science have captured the imagination for decades now. With the right tools and applications, we can take advantage of this popularity and use forensic science as a vehicle to teach critical thinking skills and the scientific method.

Critical thinking is the basis of all sound science. It can be defined as metacognition, logical argument analysis, and the rigorous weighing of evidence to support a claim. The scientific method is a structured mode of critical thinking that relies on hypothesis, experimentation and interpretation of the evidence (1). The collection and analysis of forensic evidence requires, among other skills, critical thinking and application of the scientific method. The forensic scientist makes observations, formulates hypotheses about the probative value of potential evidence, and tests these educated guesses by submitting samples to an operational forensic laboratory for analysis.

For DNA analysis, the forensic scientist first extracts DNA from the sample, quantifies the nucleic acid, amplifies it by polymerase chain reaction (PCR), and generates a short tandem repeat (STR) profile. The questioned crime scene profile is uploaded to a database and searched against indices of known samples for the purpose of identification. Bringing these processes to the classroom could provide experiential learning opportunities that highlight critical thinking abilities and are the scientific method in practice. Traditional approaches of DNA profiling, however, can be cost-prohibitive in a classroom setting.

Therefore, we sought to develop cost-effective analysis procedures that could increase the accessibility of these laboratory exercises (2). By avoiding the use of expensive commercial kits, cost per sample can be significantly reduced with: 1) expressing and purifying Taq DNA polymerase in-house (3); 2) quantifying DNA using a published SYBR Green method (4); 3) extracting DNA with a standard phenol:chloroform protocol (5); and 4) using in-house multiplex PCR primer mixes to amplify DNA. The DNA profile can then be searched against a profile database of known samples.

CODIS (Combined DNA Index System) is the general term used to describe the system of U.S. criminal justice DNA databases administered at the local, state and national level. CODIS is organized in separate indices containing autosomal short tandem repeat (A-STR) DNA profiles: Convicted Offender Index, Arrestee Index, Forensic Index (containing biological crime scene evidence), and unidentified human remains and voluntary samples collected from relatives of missing persons. As

of October 2021, the national arm of this database, NDIS (National DNA Index System), contained almost fifteen million offender profiles, over four and a half million arrestee profiles and over one million forensic profiles (https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics).

There are direct and indirect approaches for database searching to identify the potential source of a forensic biological sample. In a high-stringency direct search, a crime scene DNA profile is searched against the CODIS offender and/or arrestee indices for a direct match, or "hit," in which all alleles at all loci match exactly. A moderate-stringency search is useful with DNA evidence that contains a mixture, is partially degraded, or to accommodate the use of different DNA typing kits from various labs. A moderate stringency search may result in a partial match, which the FBI defines as a match between two single source profiles having at each locus all of the alleles of one sample represented in the other sample (https://www.fbi.gov/services/laboratory/biometric-analysis/codis/codis-and-ndis-fact-sheet) and may indicate a potential biological relationship between the two donors. A partial match is the spontaneous product of a regular database search and is distinct from the results of an indirect familial search (6).

Familial searching is a deliberate query of the DNA database using specially designed software for the purpose of identifying first-order biological relatives of the donor of a crime scene profile. Close relatives will share more DNA than unrelated individuals, e.g. full siblings share approximately 50% of their DNA. Familial searching begins with a query of the offender/arrestee indices for a direct match. If there are no hits, the questioned profile is searched against the database again to identify DNA profiles that are similar but not identical. The profiles are ranked in order of the probability that their donors share first-degree kinship with the person who left the crime scene DNA using the likelihood ratio and/or number of shared alleles (6,7). The top male candidates' samples are further profiled using Y-STRs to establish the familial relationship. Familial searches are not conducted at the national level. Each state must determine whether it will perform familial searching, and if so, the criteria and procedures that will govern its use. As of 2021, labs in Arkansas, California, Colorado, Florida, Michigan, Texas, Utah, Virginia, Wisconsin and Wyoming perform familial searches, while Maryland and D.C. laws specifically prohibit these searches (https://www.fbi.gov/services/laboratory/biometric-analysis/codis).

Both direct and indirect database searches can be part of experiential learning exercises in which students apply their critical thinking skills and the scientific method to solve mock crimes, ultimately searching the database and calculating match probabilities. CODIS activities are restricted to authorized government labs, therefore, a DNA profile database that can be used as a teaching tool has been established – the FauxDIS DNA Database. An earlier version of the database was previously introduced as searchable spreadsheet file (2). In the current report, we introduce the interactive, online FauxDIS DNA Database and demonstrate its function.

*FauxDIS* (https://www.https://www.fauxdis.org) is an online, interactive DNA profile database (**FIGURE 1**). It currently contains one hundred fifty-five DNA profiles, each comprising up to twenty-two STRs and one sex-informative locus. It is available for use in exchange for the submission of novel autosomal STR profiles to the database.

The FauxDIS database can be searched using a full or partial STR genotype. It is currently searchable for profiles containing any combination of the twenty CODIS loci, PentaE, PentaD and amelogenin. Database use is not restricted to a specific multiplex kit; it can support entries generated from kits such as PowerPlex 16, PowerPlex Fusion, SGM, or ProfilerPlus/Co-Filer.

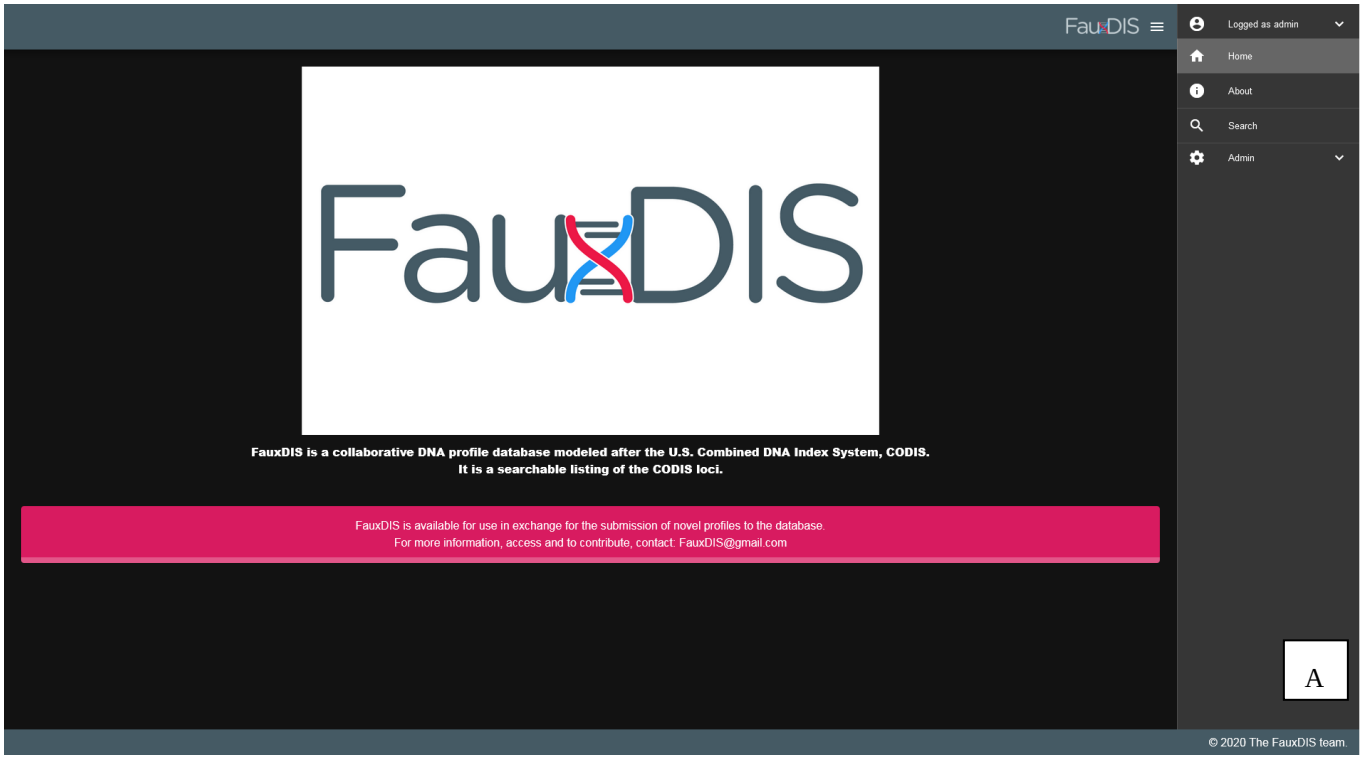
## Methods

### *The FauxDIS platform*

The back-end of the website is built in Kotlin (https://www.kotlinlang.org) with the Spring Boot framework (https://spring.io/projects/spring-boot) and uses PostgreSQL (https://www.postgresql.org/) as its database. The front-end is built using VueJS (https://www.vuejs.org) and Vuetify (https://www.vuetifyjs.com). It is deployed in Docker (https://www.docker.com) containers and deployed using Ansible (https://www.ansible.com/). The HTTPS certificates are obtained with Certbot (https://certbot.eff.org/) from Let's Encrypt (https://letsencrypt.org/).

## Results

To generate a DNA profile, we first purified the DNA using a phenol:chloroform extraction protocol, amplified with an in-house PowerPlex 16 multiplex system, and separated the amplicons by capillary electrophoresis on a 3130 Genetic Analyzer. Genemapper ID-X software (ThermoFisher Scientific) displayed the full multiplex as an electropherogram, with the x-axis delineated in units of size in base pairs (bp), and the y-axis as height in relative fluorescence units (rfu). A PowerPlex 16 allelic ladder was needed to convert the peak size in base pairs to genotype; this was used as an experiential exercise.



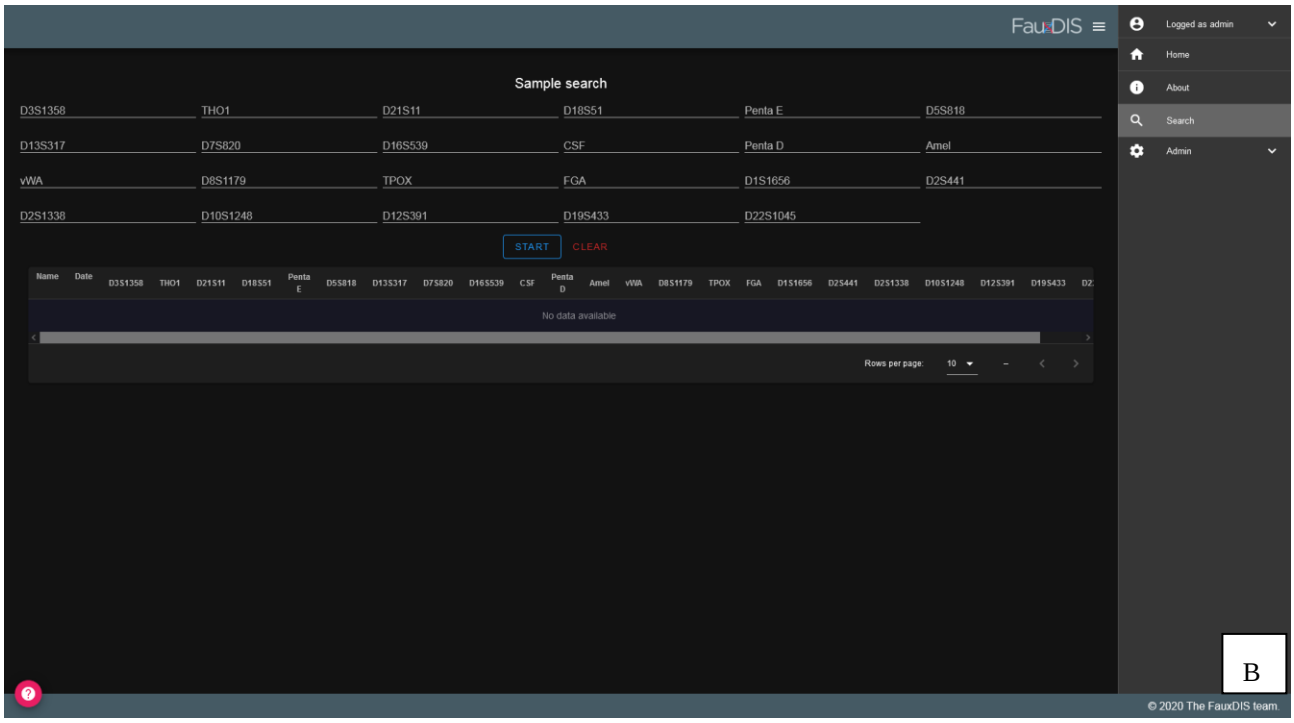




**FIGURE 1** The FauxDIS DNA Database. A) FauxDIS homehome page; and B) a clear sample search page

To construct the allelic ladder, we used the published genotype of the 2800M DNA standard (Promega) as a benchmark (8). We amplified 2800M DNA with our in-house PowerPlex 16 system and determined the size, in bp, of each of the amplified peaks. Any DNA standard with a known genotype, e.g. 9948, can be used as a benchmark in this exercise. The size of each 2800M peak was translated to its genotype, and entered in to the allelic ladder template. Then, using our understanding of structure of the STR loci, a full allelic ladder could be constructed (Supplementary Figure 1); 2800M benchmark alleles are bolded and justified left.

As an example, our amplified 2800M had two peaks, sized 232 and 236 bp at the D8S1179 locus. The known 2800M genotype at D8S1179 is 14, 15, therefore allele 14 is 232 bp and allele 15 is 236 bp. According to the PowerPlex 16 Technical Manual (8), or from the STRBase locus fact sheet (https://strbase.nist.gov/str_D8S1179.htm), we know that D8S1179 is a tetranucleotide repeat. To construct the ladder around the benchmark alleles, we can start with allele 15, which is 236 bp. Therefore allele 16 is 240 bp (236 + 4), allele 17 is 244 bp (240 + 4), and so on. This process is repeated at each locus to generate a complete allelic ladder. To ensure the most accurate measures, an allelic ladder should be generated in-house for each instrument to control for the particular environmental conditions of the space, as these affect electrophoretic mobility (9), and thus allele size.

*Direct Database Searches*. To perform a high stringency direct search, enter each allele from a full profile on the "Sample search" line corresponding to the appropriate locus. Either one or two alleles can be entered for each locus. Click "Start" and the search will be completed, typically in milliseconds. Only the samples that are a direct match, containing all alleles at all loci match exactly, will be returned (**FIGURE 2**).

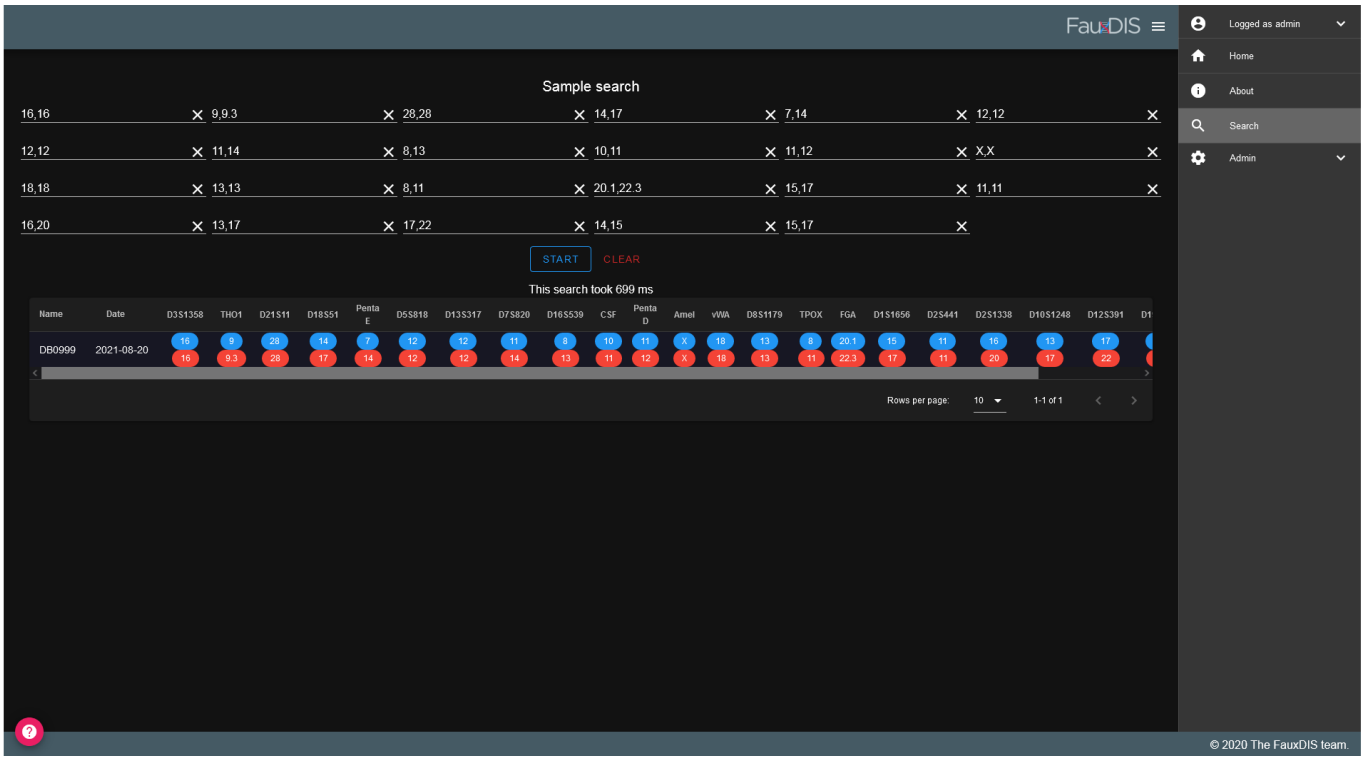



**FIGURE 2** FauxDIS Full Profile Direct Search. The 23-locus profile is found in time in the database.

A moderate stringency direct search can be simulated by entering a partial profile in the "Sample search." There is no minimum number of loci required to perform a search, and samples containing any or all of the alleles entered will be returned, that is, the database samples retrieved contain all of the alleles in the questioned sample. **FIGURE 3** demonstrates the use of a partial profile. In **FIGURE 3A**, we only entered "X,X" for amelogenin, returning 80 samples. Adding the "16,17" alleles at the D3S1358 locus, six samples were returned (**FIGURE 3B**). Including "8, 9.3" at THO1 in the query resulted in only one profile (**FIGURE 3C**). As we include additional profiles in the database, a greater number of matching loci will be necessary to identify a single profile.

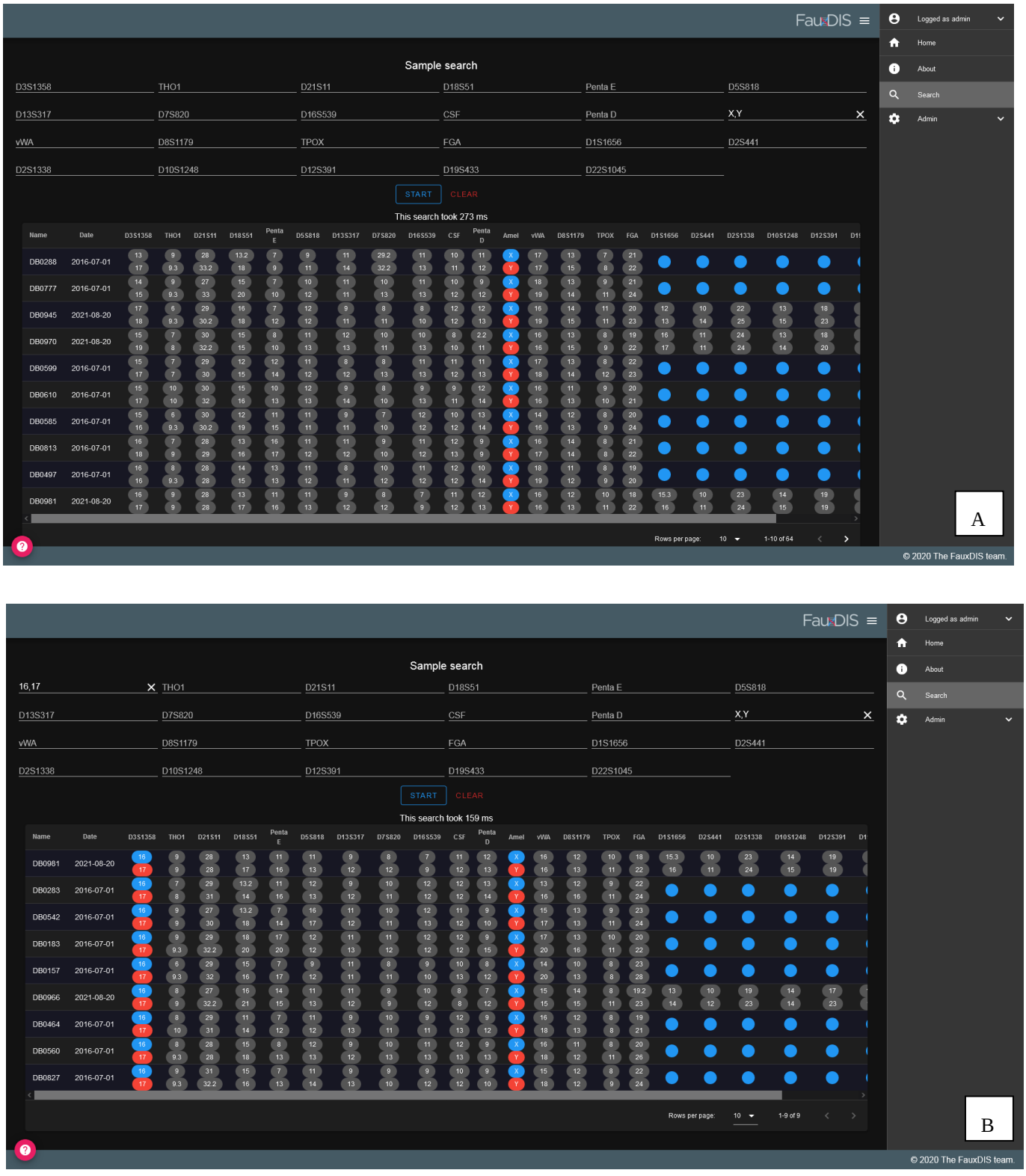



**FIGURE 3** FauxDIS Partial Profile Direct Search. A) searching with the X, X alleles at amelogenin returns 64 samples; B) adding 16, 17 at D3S1358 reduces the list to 9 samples (continued on next page)

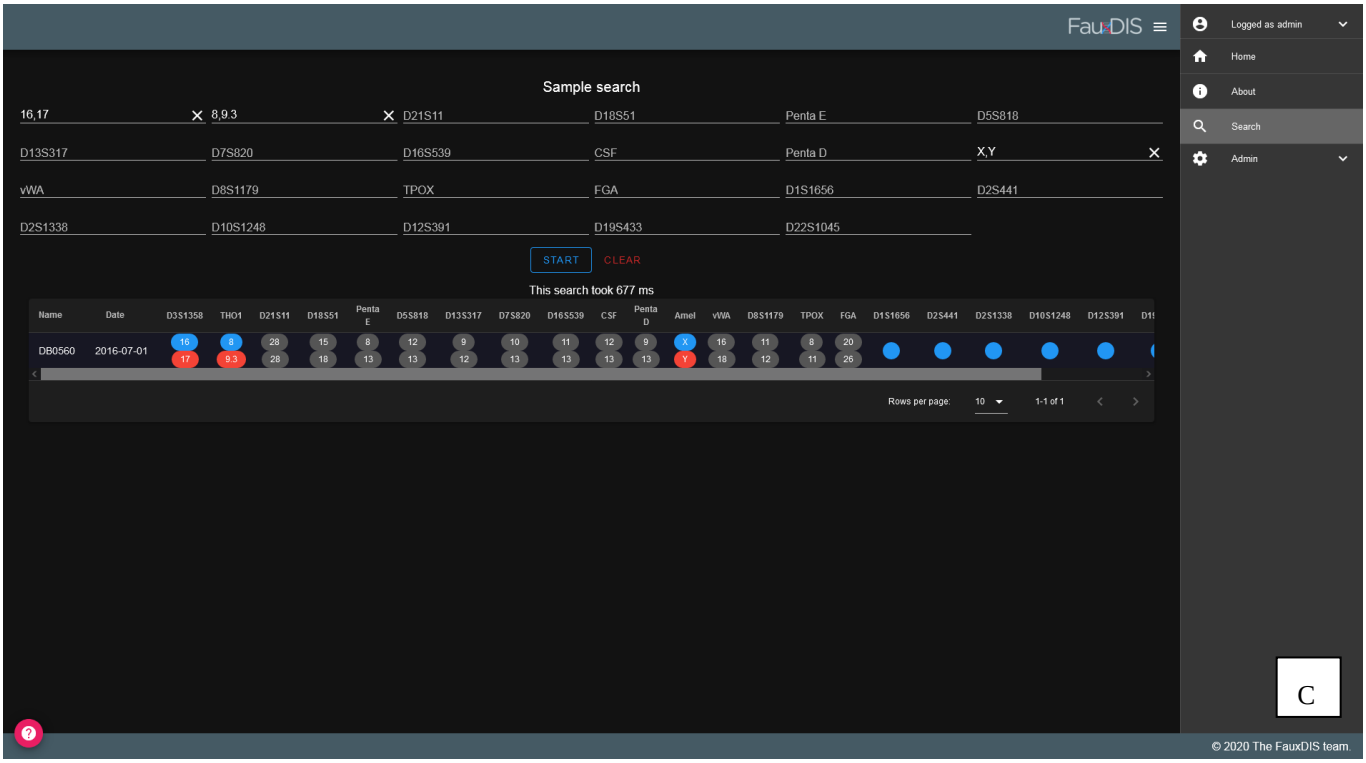



**FIGURE 3** FauxDIS Partial Profile Direct Search: continued from previous page, C) adding 8, 9.3 at THO1 results in a single profile

*Indirect Database Searches*

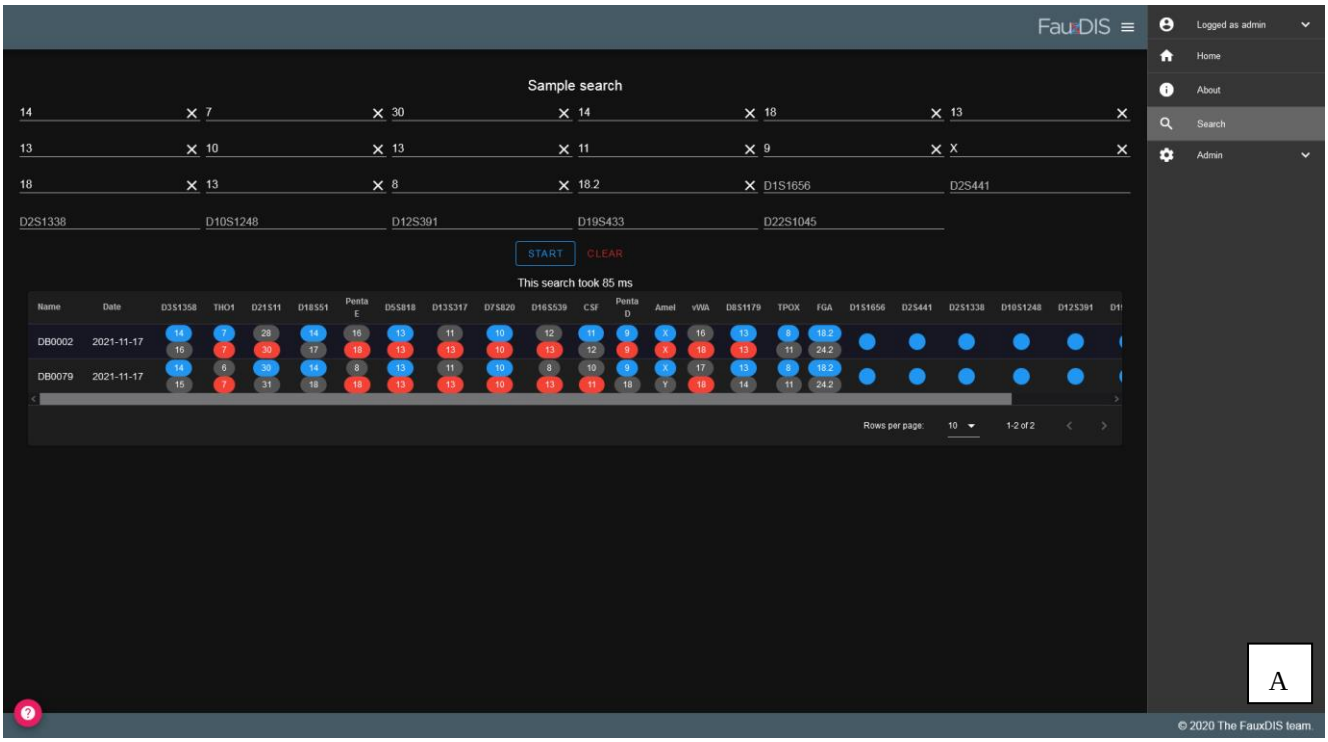
A true indirect, or familial, search requires specialized software. We cannot simulate the search exactly but can use FauxDIS to teach the principles. As of November 2021, the database contains one known family group. Their genotypes and relationships are provided as part of the worksheets in **Supplementary 2**. To demonstrate the database function, we used a partial profile comprising one allele at each locus of a known profile, DB0079. The search returned two profiles, having one common allele at each locus and indicating a parent/child relationship (**FIGURE 4A** and **Supplementary 2**).

FauxDIS can be a tool to teach the principles of allele and genotype frequency calculations and their consequence in forensic analysis. Although there are more sophisticated statistical models that educators can adopt, calculating the Random Match Probability (RMP) is a relatively straight forward demonstration of the principles. RMP is the probability that the DNA profile of a random, unrelated person in the population will match the profile generated from a crime scene sample. It can be calculated based on either observed or expected frequencies.

Genotype frequency can be estimated by direct observation using the counting method (10) as the ratio of the number of times a DNA profile is observed in the database to the total number of profiles, e.g. sample DB001 (**FIGURE 2**) has a frequency of 1 in 155 or 0.65%. Determination of genotype frequencies by counting does not rely on theoretical assumptions and, while it is a simpler method, it does not take advantage of the power of the genetic approach.

Theoretical models based on the principles of population genetics can be applied to calculate the expected allele frequencies (11). We need to make two basic assumptions about the population: 1) independence between loci (linkage equilibrium); and 2) independence between alleles (Hardy-Weinberg equilibrium). Linkage equilibrium indicates that the loci are independent and associate randomly and, with a population in Hardy-Weinberg equilibrium, allele frequency can be correlated with genotype frequency. For a heterozygous locus, frequency is calculated by: $2pipj$, where pi = the frequency of one allele and pj = the frequency of the other allele. Homozygote frequency is calculated by: $p^2 + p(1-p)\theta$, where p = allele frequency and $\theta = 0.01$ in a typical population or $\theta = 0.03$ in an isolated population. The theta correction is a measure of the effects of population substructure, or co-ancestry of alleles (12). A table of expected allele frequencies that can be used in calculations of the RMP is available in the literature (13) and online (https://www.promega.com/products/pm/genetic-identity/population-statistics/allele-frequencies/). From a forensic standpoint, having a population in both linkage and Hardy-Weinberg equilibrium means that each

matching allele is statistically independent evidence. The individual frequencies from each locus can be multiplied to calculate the RMP using the product rule. With this calculation, students can quantify the strength of the DNA match they have generated through their crime scene exercises.



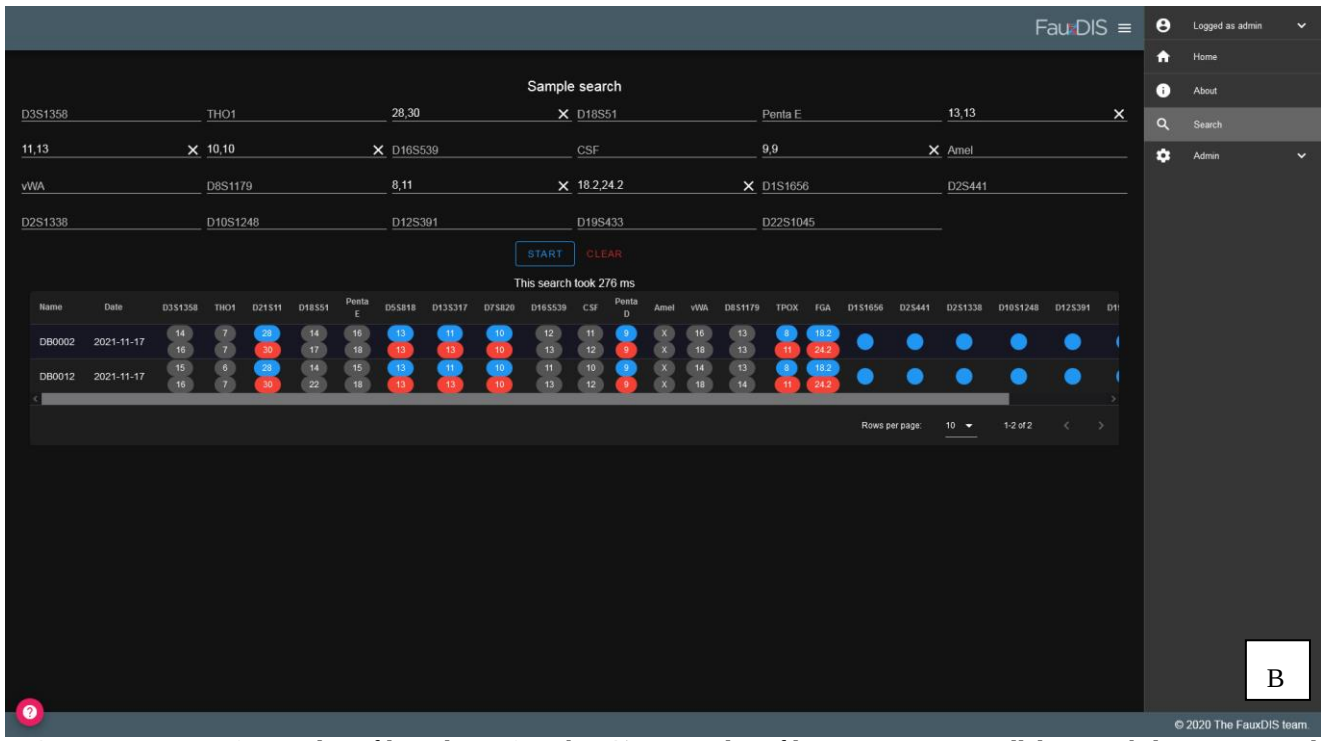




**FIGURE 4** FauxDIS Partial Profile Indirect Search. A) a partial profile comprising one allele at each locus was used to conduct an indirect search, returning two profiles, indicating a parent/child relationship; B) a partial profile consisting of two alleles at seven STR loci was searched in the database. It returned two profiles sharing all alleles at the seven loci and one allele at each of the remaining loci, and indicating a full sibling relationship.

**Discussion and Conclusion**

True crime and forensic science have captured the public's imagination for decades. With the right tools, we can take advantage of this attention and let forensic science be a vehicle for teaching critical thinking skills and the scientific method. In this report, we introduce FauxDIS, an interactive online forensic DNA profile database (www.https://www.fauxdis.org). The database can become an integral part of mock crime scene exercises that require students to apply critical thinking skills in the analysis of forensic evidence.

The FauxDIS work flow incorporates instrumentation and protocols analogous to those employed in U.S. operational crime laboratories. The database can be used to simulate both direct and indirect profile searches, demonstrating principles of genetics. It also supports experimentation with partial profiles, which can be useful in simulations of degraded and damaged samples commonly found at a crime scene. Further, with a successful database search, random match probabilities can be calculated using either observed or predicted allele frequencies. These experiential exercises teach valuable skills, and the practical experience that students gain may be attractive to potential employers.

FauxDIS currently contains 151 autosomal profiles. Growing the database with additional profiles will increase its utility. We will continue to generate profiles in-house, and will accept profiles from other educators, ensuring that the DNA profiles they use in mock crime exercises will be found in the database. We will offer access to the online system (www.https://www.fauxdis.org) in exchange for novel DNA profiles. We recognize that many colleges and universities will be limited by the availability of the necessary instrumentation to generate a DNA profile. To extend the experiential learning opportunity to as many students as possible, we will also accept single-source samples for in-house analysis. In exchange for a certain number of unique samples, we will generate profiles and deposit them in the database, as if they were collected and submitted to an operational forensic laboratory.

FauxDIS is a dynamic entity; it can be expanded to accommodate new marker systems in response to advances in forensic science. In the future, additional indices will include Y-STRs, single nucleotide polymorphisms and massively parallel sequencing data. With this database, we hope to provide a tool for experiential exercises and contribute to a collaborative network of educators.

**References**

1. Kraus S, Sears S, Burke B. Is truthiness enough? classroom activities for encouraging evidence-based critical thinking. Journal of Effective Teaching 2013;13:83-93.

2. Baranski J, Davalos-Romero K, Blum M, Foster A, Hall A. FauxDIS: a searchable forensic DNA database to support experiential learning. J Forensic Sci Educ 2020;2(1).

3. Bellin RM, Bruno MK, Farrow MA. Purification and characterization of Taq polymerase: A 9-week biochemistry laboratory project for undergraduate students. Biochem Mol Biol Educ 2010;38:11-16.

4. Nicklas JA, Buel E. Development of an Alu-based, real-time PCR method for quantitation of human DNA in forensic samples. J Forensic Sci 2003;48:936-944.

5. Comey C, Koons B, Presley K, Smerick J, Sobieralski C, Stanley DB. DNA extraction strategies for amplified fragment length polymorphism analysis. J Forensic Sci 1994;39: 1254-1269.

6. Ge J, Budowle B. Forensic investigation approaches of searching relatives in DNA databases. J Forensic Sci 2021;66:430-443.

7. NFSTC. Familial Searching. *https://projects.nfstc.org/fse/13/13-0.html*.

8. Promega. PowerPlex 16 Technical Manual.

9. Rogacs A, Santiago JG. Temperature effects on electrophoresis. Anal Chem 2013;85:5103-5113.

10. National Research Council. DNA Technology in Forensic Science. National Academy of Sciences, Washington (DC), 1992.

11. National Research Council. The Evaluation of Forensic DNA Evidence, National Academy of Sciences, Washington (DC), 1996.

12. Butler J. Forensic DNA Typing, 2nd Edition. Academic Press, 2005.

13. Steffen CR, Coble MD, Gettings KB, Vallone PM. Corrigendum to 'U.S. population data for 29 autosomal STR loci' [Forensic Sci. Int. Genet. 7 (2013) e82-e83]. Forensic Sci Int Genet 2017; 31:e36-e40.

**Supplementary FIGURE 1.** The PowerPlex allelic ladder generated using 2800M alleles as benchmarks for each locus. The 2800M genotype is bolded and justified leftSupp.

| **D3S1358** | |
|---|---|
| Allele | size (bp) |
| 12 | 110 |
| 13 | 114 |
| 14 | 118 |
| 15 | 122 |
| 16 | 126 |
| **17** | **130** |
| **18** | **134** |
| 19 | 138 |
| 20 | 142 |

| **THO1** | |
|---|---|
| Allele | size (bp) |
| 4 | 152 |
| 5 | 156 |
| **6** | **160** |
| 7 | 164 |
| 8 | 168 |
| 9 | 172 |
| **9.3** | **175** |
| 10 | 176 |
| 11 | 180 |
| 13.3 | 184 |

| **D21S11** | |
|---|---|
| Allele | size (bp) |
| 24 | 199 |
| 24.2 | 201 |
| 25 | 203 |
| 25.2 | 205 |
| 26 | 207 |
| 27 | 211 |
| 28 | 215 |
| 28.2 | 217 |
| **29** | **219** |
| 29.2 | 221 |
| 30 | 223 |
| 30.2 | 225 |
| 31 | 227 |
| **31.2** | **229** |
| 32 | 231 |
| 32.2 | 233 |
| 33 | 235 |
| 33.2 | 237 |
| 34 | 239 |
| 34.2 | 241 |
| 35 | 243 |
| 35.2 | 245 |
| 36 | 247 |
| 37 | 251 |
| 38 | 255 |

| **D18S51** | |
|---|---|
| Allele | size (bp) |
| 8 | 284 |
| 9 | 288 |
| 10 | 292 |
| 10.2 | 294 |
| 11 | 296 |
| 12 | 300 |
| 13 | 304 |
| 13.2 | 306 |
| 14 | 308 |
| 15 | 312 |
| **16** | **316** |
| 17 | 320 |
| **18** | **324** |
| 19 | 328 |
| 20 | 332 |
| 21 | 336 |
| 22 | 340 |
| 23 | 344 |
| 24 | 348 |
| 25 | 352 |
| 26 | 356 |
| 27 | 360 |

| **PentaE** | |
|---|---|
| Allele | size (bp) |
| 5 | 375 |
| 6 | 380 |
| **7** | **385** |
| 8 | 390 |
| 9 | 395 |
| 10 | 400 |
| 11 | 405 |
| 12 | 410 |
| 13 | 415 |
| **14** | **420** |
| 15 | 425 |
| 16 | 430 |
| 17 | 435 |
| 18 | 440 |
| 19 | 445 |
| 20 | 450 |
| 21 | 455 |
| 22 | 460 |
| 23 | 465 |
| 24 | 470 |

| **D5S818** | |
|---|---|
| Allele | size (bp) |
| 7 | 112 |
| 8 | 116 |
| 9 | 120 |
| 10 | 124 |
| 11 | 128 |
| **12** | **132** |
| 13 | 136 |
| 14 | 140 |
| 15 | 144 |
| 16 | 148 |

| **D13S317** | |
|---|---|
| Allele | size (bp) |
| 7 | 172 |
| 8 | 176 |
| **9** | **180** |
| 10 | 184 |
| **11** | **188** |
| 12 | 192 |
| 13 | 196 |
| 14 | 200 |
| 15 | 204 |

| **D7S820** | |
|---|---|
| Allele | size (bp) |
| 6 | 211 |
| 7 | 215 |
| **8** | **219** |
| 9 | 223 |
| 10 | 227 |
| **11** | **231** |
| 12 | 235 |
| 13 | 239 |
| 14 | 243 |

| **D16S539** | |
|---|---|
| Allele | size (bp) |
| 5 | 269 |
| 8 | 273 |
| **9** | **277** |
| 10 | 281 |
| 11 | 285 |
| 12 | 289 |
| **13** | **293** |
| 14 | 297 |
| 15 | 301 |

| **CSF1PO** | |
|---|---|
| Allele | size (bp) |
| 6 | 317 |
| 7 | 321 |
| 8 | 325 |
| 9 | 329 |
| 10 | 333 |
| 11 | 337 |
| **12** | **341** |
| 13 | 345 |
| 14 | 349 |
| 15 | 353 |

| **PentaD** | |
|---|---|
| Allele | Locus |
| 2.2 | 367 |
| 3.2 | 372 |
| 5 | 380 |
| 7 | 390 |
| 8 | 395 |
| 9 | 400 |
| 10 | 405 |
| 11 | 410 |
| **12** | **415** |
| **13** | **420** |
| 14 | 425 |
| 15 | 430 |
| 16 | 435 |
| 17 | 440 |

| **Amel** | |
|---|---|
| Allele | size (bp) |
| **X** | **106** |
| **Y** | **112** |

| **vWA** | |
|---|---|
| Allele | size (bp) |
| 10 | 124 |
| 11 | 128 |
| 12 | 132 |
| 13 | 136 |
| 14 | 140 |
| 15 | 144 |
| **16** | **148** |
| 17 | 152 |
| 18 | 156 |
| **19** | **160** |
| 20 | 164 |
| 21 | 168 |
| 22 | 172 |

| **D8S1179** | |
|---|---|
| Allele | size (bp) |
| 7 | 204 |
| 8 | 208 |
| 9 | 212 |
| 10 | 216 |
| 11 | 220 |
| 12 | 224 |
| 13 | 228 |
| **14** | **232** |
| **15** | **236** |
| 16 | 240 |
| 17 | 244 |
| 18 | 248 |

| **TPOX** | |
|---|---|
| Allele | size (bp) |
| 6 | 262 |
| 7 | 266 |
| 8 | 270 |
| 9 | 274 |
| 10 | 278 |
| **11** | **282** |
| 12 | 286 |
| 13 | 290 |

| **FGA** | |
|---|---|
| Allele | size (bp) |
| 16 | 322 |
| 17 | 326 |
| 18 | 330 |
| 18.2 | 332 |
| 19 | 334 |
| 19.2 | 336 |
| **20** | **338** |
| 20.2 | 340 |
| 21 | 342 |
| 21.2 | 344 |
| 22 | 346 |
| 22.2 | 348 |
| **23** | **350** |
| 23.2 | 352 |
| 24 | 354 |
| 24.2 | 356 |
| 25 | 358 |
| 25.2 | 360 |
| 26 | 362 |
| 27 | 366 |
| 28 | 370 |
| 29 | 374 |
| 30 | 378 |
| 31.2 | 384 |
| 43.2 | 432 |
| 44.2 | 436 |
| 45.2 | 440 |
| 46.2 | 444 |

**Supplementary 2** FauxDIS DNA Database Worksheet

**I) DIRECT SEARCH**

The genotypes for three database samples are given in the table below. They can be used to demonstrate both high and moderate stringency searches.

| Locus | Sample 1 (DB001) | Sample 2 (DB0966) | Sample 3 (DB0560) |
|---|---|---|---|
| D3S1358 | 17,17 | 16,17 | 16,17 |
| THO1 | 6, 9 | 8,9 | 8,9.3 |
| D21S11 | 28,30 | 27,32.2 | 28,28 |
| D18S51 | 13.2,15 | 16,21 | 15,18 |
| Penta E | 13,14 | 14,15 | 8,13 |
| D5S818 | 12,12 | 11,13 | 12,13 |
| D13S317 | 11,12 | 11,12 | 9,12 |
| D7S820 | 10,11 | 9,9 | 10,13 |
| D16S539 | 12,13 | 10,12 | 11,13 |
| CSF1PO | 9,12 | 8,8 | 12,13 |
| Penta D | 10,13 | 7,12 | 9,13 |
| Amelogenin | X,X | X,Y | X,Y |
| vWA | 15,17 | 15,15 | 16,18 |
| D8S1179 | 13,13 | 14,15 | 11,12 |
| TPOX | 9,9 | 8,11 | 8,11 |
| FGA | 24,25.2 | 19.2,23 | 20,26 |
| D1S1656 | | 13,14 | |
| D2S441 | | 10,12 | |
| D2S1338 | | 19,23 | |
| D10S1248 | | 14,14 | |
| D12S391 | | 17,23 | |
| D19S433 | | 14.2,14.2 | |
| D22S1045 | | 11,16 | |

Samples 1 & 2 (high stringency, direct search): enter complete genotypes: will retrieve 1 sample each from the database. Sample 1 was run with a 16-locus multiplex. Sample 2 was run with a 23-locus multiplex. The database can accommodate any combination of markers found in kits, and displays loci with no data as an empty circle.

Sample 3 (moderate stringency, direct search): To demonstrate a search with a partial profile,

a) enter X,Y at amelogenin. Sixty-six profiles are retrieved.
b) enter 15,17 at D3S1358. The field is narrowed to nine profiles.
c) enter 8, 9.3 at THO1. A single profile is returned (DB0560).

**II) INDIRECT SEARCH**

The genotypes from a family group are listed in the table below. They can be used in various combinations to demonstrate an indirect search. To use the database, select a profile to use for your search. Have the students enter only the alleles that are shared between that profile and the associated family profile(s). Examples are provided in the following pages.

| | Sibling 3 DB0002 | Sibling 2 DB0012 | Sibling 1 DB0022 | Parent 1 DB0070 | Parent 2 DB0079 |
|---|---|---|---|---|---|
| D3 | 14, 16 | 15, 16 | 14, 16 | 16 | 14,15 |
| THO1 | 7 | 6, 7 | 7 | 7 | 6, 7 |
| D21 | 28,30 | 28,30 | 28, 30 | 28, 30 | 30, 31 |
| D18 | 14,17 | 14, 22 | 18, 22 | 17, 22 | 14, 18 |
| PentaE | 16,18 | 15, 18 | 16, 18 | 15, 16 | 8, 18 |
| D5 | 13 | 13 | 13 | 13 | 13 |
| D13 | 11, 13 | 11,13 | 11 | 11 | 11, 13 |
| D7 | 10 | 10 | 10 | 10, 11 | 10 |
| D16 | 12, 13 | 11, 13 | 12, 13 | 11, 12 | 8, 13 |
| CSF | 11, 12 | 10, 12 | 9, 11 | 9, 12 | 10, 11 |
| PentaD | 9 | 9 | NA | 9 | 9, 18 |
| Amel | X | X | X | X | X,Y |
| vWA | 16, 18 | 14, 18 | 14, 18 | 14, 16 | 17, 18 |
| D8 | 13 | 13, 14 | NA | 13 | 13, 14 |
| TPOX | 8, 11 | 8, 11 | 8 | 8, 11 | 8, 11 |
| FGA | 18.2, 24.2 | 18.2, 24.2 | 18.2, 24.2 | 19.2, 24.2 | 18.2, 24.2 |

NA – no allele

| Comparison | Shared loci | Percent Match |
|---|---|---|
| Sib 3/Sib 2 | 7/15 loci | 47% |
| Sib 3/Sib 1 | 8/13 loci | 61% |
| Sib 1/Sib 2 | 5/13 loci | 38% |
| Parent1/Parent 2 | 2/15 loci | 13% |

**Question 1** To demonstrate a familial match search with a parent DNA profile, enter the following partial genotype (Parent 2, DB0079) and search:

| D3 | THO1 | D21 | D18 | PentaE | D5 | D13 | D7 | D16 | CSF | PentaD |
|---|---|---|---|---|---|---|---|---|---|---|
| 14 | 7 | 3 | 14 | 15 | 13 | 11 | 10 | 13 | 11 | 9 |

| Amel | vWA | D8 | TPOX | FGA |
|---|---|---|---|---|
| X | 18 | 13 | 8 | 18.2 |

Two profiles will be retrieved, DB0002 (Sibling 3) and DB0079 (Parent 2). The matching STR alleles are circled in the table below for reference. The parent and child share one allele at each locus

| | Sibling 3 DB0002 | Parent 2 DB0079 |
|---|---|---|
| D3 | (14), 16 | (14),15 |
| THO1 | 7, (7) | 6, (7) |
| D21 | 28, (30) | (30), 31 |
| D18 | (14), 17 | (14), 18 |
| PentaE | 16, (18) | 8, (18) |
| D5 | (13),13 | (13),13 |
| D13 | (11), 13 | (11), 13 |
| D7 | (10),10 | (10),10 |
| D16 | 12, (13) | 8, (13) |
| CSF | (11), 12 | 10, (11) |
| PentaD | (9),9 | (9),18 |
| Amel | X,X | X,Y |
| vWA | 16, (18) | 17, (18) |
| D8 | (13),13 | (13),14 |
| TPOX | (8),11 | (8),11 |
| FGA | (18.2), 24.2 | (18.2), 24.2 |

*Note*: at D13, Parent 1 has an allele 11, so the 13 allele is the obligate Parent 2 allele. At TPOX, both parents have an 8,11 so either allele could have come from the Parent 2. AT FGA, Parent 1 has a 19.2, 24.2, therefore the 18.2 allele is the obligate Parent 2 allele.

**Question 2** To demonstrate a familial match between a parent and two children, enter the following partial genotype (DB0079 Parent 2) and search:

| D3 | THO1 | D21 | D18 | PentaE | D5 | D13 | D7 | D16 | CSF | PentaD |
|---|---|---|---|---|---|---|---|---|---|---|
| No entry | 7 | 3 | 14 | 15 | 13 | 11 | 10 | 13 | No entry | 9 |

| Amel | vWA | D8 | TPOX | FGA |
|---|---|---|---|---|
| X | 18 | 13 | No entry | 18.2 |

Three profiles will be retrieved: Sibling 2 (DB0012), Sibling 3 (DB0002), and Parent 2. The matching STR alleles are circled for reference.

| | Sibling 3 DB0002 | Sibling 2 DB0012 | Parent 2 DB0079 |
|---|---|---|---|
| D3 | 14, 16 | 15, 16 | |
| THO1 | 7, 7 | 6, 7 | 7 |
| D21 | 28, 30 | 28, 30 | 30 |
| D18 | 14, 17 | 14, 22 | 14 |
| PentaE | 16, 18 | 15, 18 | 18 |
| D5 | 13, 13 | 13, 13 | 13 |
| D13 | 11, 13 | 11, 13 | 11 |
| D7 | 10, 10 | 10, 10 | 10 |
| D16 | 12, 13 | 11, 13 | 13 |
| CSF | 11, 12 | 10, 12 | |
| PentaD | 9, 9 | 9, 9 | 9 |
| Amel | X, X | X, X | X |
| vWA | 16, 18 | 14, 18 | 18 |
| D8 | 13, 13 | 13, 14 | 13 |
| TPOX | 8, 11 | 8, 11 | |
| FGA | 18.2, 24.2 | 18.2, 24.2 | 18.2 |

The siblings share the same allele with each other and the parent at 12/15 STR loci. Both share one allele with the parent at each locus.

**Question 3.** To demonstrate a familial match between full siblings, enter the following partial genotype (DB0002 Sibling 3) and search:

| D3 | THO1 | D21 | D18 | PentaE | D5 | D13 | D7 | D16 | CSF | PentaD |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 28,30 | | | 13 | 11,13 | 10 | | | 9 |

| Amel | vWA | D8 | TPOX | FGA |
|---|---|---|---|---|
| X | | | 8,11 | 18.2, 24.2 |

Two profiles will be returned: Sibling 3 (DB0002) and Sibling 2 (DB0012). The matching STR loci are circled for reference.

| | Sibling 3 DB0002 | Sibling 2 DB0012 |
|---|---|---|
| D3 | 14, 16 | 15, 16 |
| THO1 | 7 | 6, 7 |
| D21 | 28,30 | 28,30 |
| D18 | 14,17 | 14, 22 |
| PentaE | 16,18 | 15, 18 |
| D5 | 13,13 | 13,13 |
| D13 | 11, 13 | 11,13 |
| D7 | 10,10 | 10,10 |
| D16 | 12, 13 | 11, 13 |
| CSF | 11, 12 | 10, 12 |
| PentaD | 9,9 | 9,9 |
| Amel | X,X | X,X |
| vWA | 16, 18 | 14, 18 |
| D8 | 13,13 | 13, 14 |
| TPOX | 8, 11 | 8, 11 |
| FGA | 18.2, 24.2 | 18.2, 24.2 |

The siblings have the same alleles at 7/15 STR loci, and share one allele at each of the remaining loci.