Development of a DNA Database for Education and Research Purposes

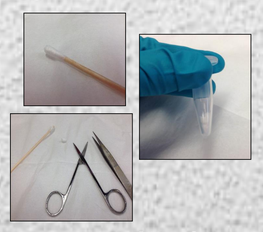
Melanie Blum, Nichole Leacock, Ashley Foster BA, Ashley Hall PhD
Department of Entomology, Forensic Science Degree Program
University of Nebraska-Lincoln

ABSTRACT

The US national DNA database, CODIS (Combined DNA Index System), CODIS contains over 10 million genetic profiles originating from convicted offenders, missing persons, and crime scene evidence and is searchable only by authorized governmental agencies. Due to the nature of the data it contains, educators do not have access to this powerful database for teaching purposes. Currently there is a lack of a searchable human DNA profile databases available for use in the research lab or classroom. If it were available, such a database could be a valuable tool in the analysis of large data sets, use in forensic laboratory exercises, and in answering questions in population genetics. Therefore, we have initiated the development of a searchable DNA database modeled after CODIS that can be used by educators nation-wide. It will allow students to gain hands-on experience, knowledge in the manipulation of large data sets and would support calculations in population genetics problems. The development of this database is a beneficial contribution to the field of forensic science because it has the potential to foster a nation-wide collaboration and can be used for graduate and undergraduate education and training. To initiate this project, human biological DNA samples containing buccal epithelial cells were collected from donors using a sterile cotton swab. DNA was isolated and purified using an organic extraction. Each sample was then quantified using the Qubit fluorometer, Multiplex polymerase chain reaction (PowerPlex 16) was used to amplify the DNA at 16 loci, which were targeted by specific, fluorescently-labeled primers. The PCR products were run through capillary electrophoresis on the 3130 Genetic Analyzer to generate a DNA profile. This profile was visualized as a plot of relative fluorescent units (RFU) against size in base pairs (bp) for each of the 16 regions (loci) amplified by PCR. Each loci has two alleles giving either one or two peaks. The size in base pairs determines which alleles are present, and using an allelic ladder each allele can be determined allowing the profile to be genotyped. The generated profiles were entered into the database. To date, a standard work-flow for processing samples has been established and 100 profiles have been added to the database. In the future, the database will be transferred to an online, searchable format and collaborators from other universities and colleges will be invited to submit samples for analysis to form a large consortium of participants.

EXPERIMENTAL DESIGN

Organic Extraction



Buccal Swabs

Quantification

Qubit Fluorometer

Photo from:
http://www.genengnews.com/elisareduxchallenge/the_sponsor.htm

Sample	Concentration (ng/µL)
Buccal Swab (positive control)	20.0
Sample 1	21.0
Sample 2	18.3
Sample 3	9.10
Sample 4	6.68
Sample 5	3.78
Sample 6	1.28
Extraction Blank	Too low OOR

Range of Quantification Concentrations

Polymerase Chain Reaction (PCR)

Thermocycler

Analyze Samples

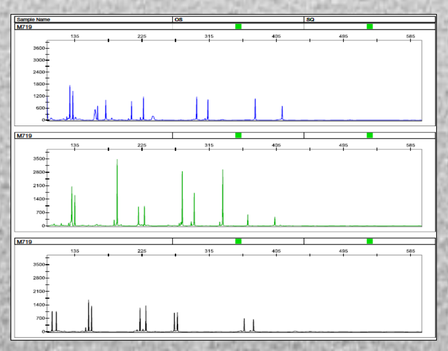
Genetic Analyzer (3130)

WORK FLOW

1) Cut cotton tips off swabs
2) Preform an organic extraction to separate the DNA
3) DNA Quantification
4) PCR
5) Analyze PCR product on a genetic analyzer (3130) to generate profiles
6) Genotype profiles
7) Add to Database

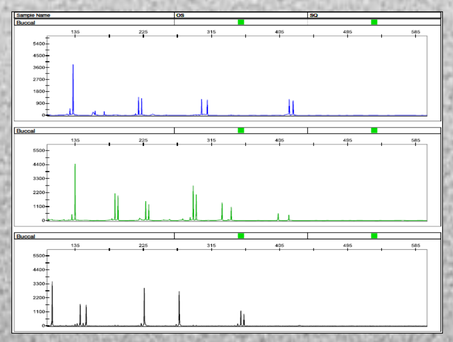
RESULTS

POSITIVE CONTROL



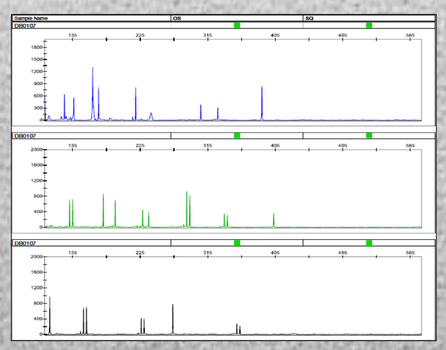
STR profile (PowerPlex 16) generated from positive control M719

POSITIVE CONTROL



STR profile (PowerPlex 16) generated from a positive control buccal swab

SAMPLE

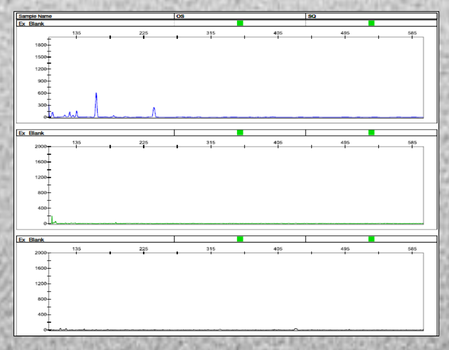


STR profile (PowerPlex 16) generated from a buccal swab collected from a donor (DB0107)

Genotype

Loci	Allele
D3S1358	15,18
THO1	6,8
D21S11	28,29
D18S51	14,20
Penta E	7,7
D5S818	11,12
D13S317	8,12
D7S820	10,12
D16S539	11,12
CSF1PO	11,12
Penta D	10,12
Amel	X,X
vWA	17,18
D8S1179	13,14
TPOX	8,8
FGA	24,25

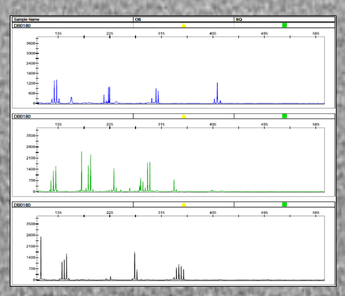
NEGATIVE CONTROL



STR profile (PowerPlex 16) generated from an extraction blank

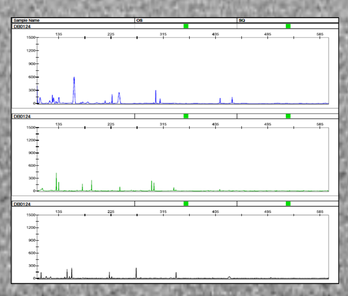
ADDITIONAL CHALLENGES & CONSIDERATIONS

SAMPLE – MIXED PROFILE



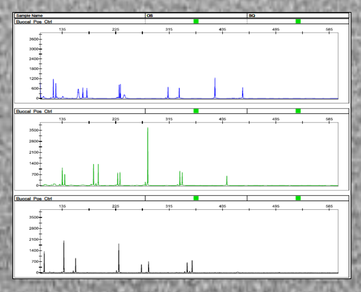
STR profile (PowerPlex 16) from a donor sample that gave multiple alleles at one loci – cannot be used in database

SAMPLE – LOW RFU



STR profile (PowerPlex 16) generated from a donor sample which gave low values -cannot be used in database

POSITIVE CONTROL



STR profile (PowerPlex 16) from a positive control buccal swab

FUTURE IMPLICATONS

Uses and Direction of Project:

1) **Continue to collect and add more profiles from donor samples.**
2) **Add software that allows the database to be searchable**
3) **Foster nation-wide collaboration : universities will be invited to submit samples for analysis and will have access to the online database upon completion.**
4) **Useful tool for graduate and undergraduate education and training.**
5) **Will be implemented into Forensic Science courses at UNL (FORS 401, FORS 408, FORS 485) and evaluated by the students.**
6) **Used for population genetic problems and analysis.**

UNIVERSITY OF NEBRASKA-LINCOLN