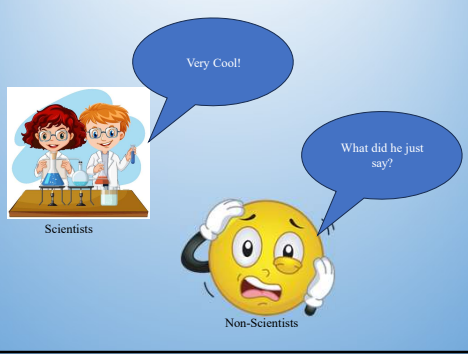
Scientific Solutions for Complex DNA Cases

Bruce Budowle
3rd Forensic DNA Symposium
Cape Town, South Africa
June 13, 2023

1



2

Forensic Science

- Application of science in the investigation of legal matters
- Scientific knowledge and technology are used to serve as witnesses in both criminal and civil matters
- Science may not offer definitive solutions for all scenarios; it does provide a special investigative role
- Goal is "attribution" – i.e., who committed the crime and who can be exonerated
- For DNA – "source attribution" – i.e., who can or cannot be a donor of the sample

3

DNA and Human Identity Testing

- Forensic cases – criminal, civil
- Parentage testing – identifying father, nursery mix-ups, immigration, inheritance
- Historical investigations
- Population studies, human diversity, anthropology
- Missing persons investigations
- Mass disasters
- Military DNA "dog tag"
- Convicted felon DNA databases
- Patient sample mix-ups

4

What are Complex Cases?

- Low quantity DNA
- Degraded DNA
- Kinship
- Mixtures
- New targets
- Lack of evidence
- Uncertainty
- Alternate explanations

5

Addressing Complex Cases

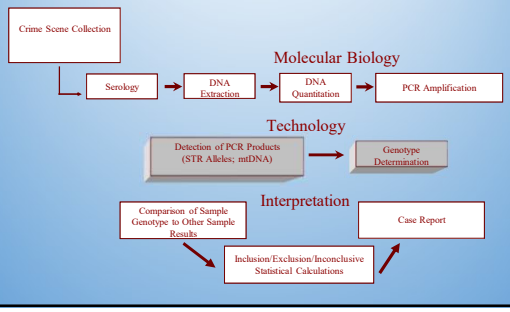
- Technology
- Information analysis
- Transparency
- Communication
- Education
- Collaboration

6

Wide Range of Areas for Addressing Challenges

7

Steps in DNA Sample Processing



8

Diversity of Swabs

(Swabs for Crime Scene Collection)

Cotton-Tip Swab

Fab Swab

Omni Swab

Bode secure swab

9

4N6FLOQSwabs™

Why consider this type of swab?

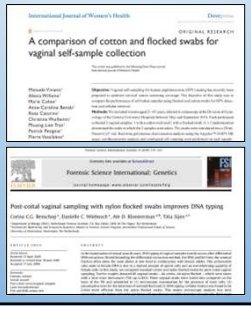
10

Physical Structure

- No internal core like cotton
  - DNA not trapped
  - FTA paper traps DNA
- All DNA on surface
- Wet or dry collection
  - Dry collection
    - less chance of degradation
- Certified Amplifiable hDNA free
- Antimicrobial treated
- ISO 18385 compliance

11

Increased DNA Yield



A comparison of cotton and flocked swabs for vaginal self-sample collection

Forensic Science International: Genetics

Post-coital vaginal sampling with nylon flocked swabs improves DNA typing

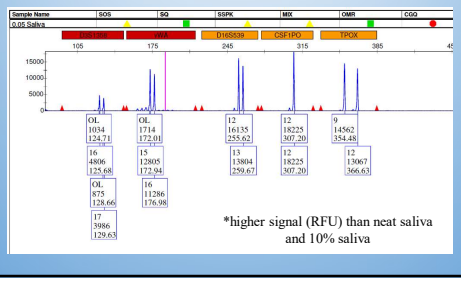
12

Enhancing Sensitivity/Workflow: microFLOQ®

- Touch DNA
- Lysed DNA on surface of collection device
- Direct PCR
- Less sample required
- Greater success rate
- Better workflow

13

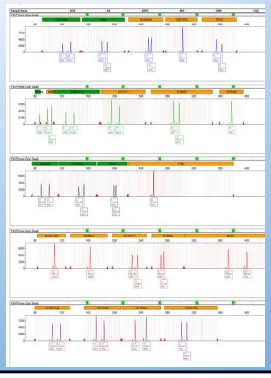
microFLOQ® Direct Amp 1:19 Saliva



*higher signal (RFU) than neat saliva and 10% saliva

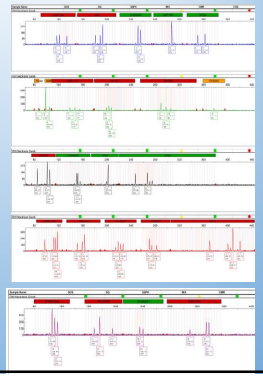
14

Phone grip



15

necklace



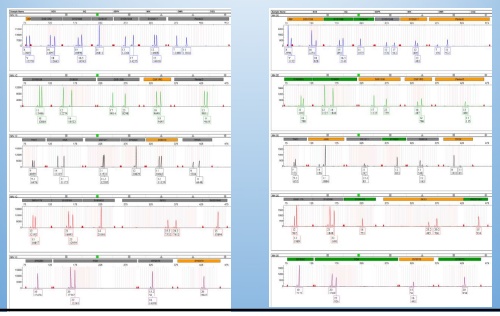
16

Sampling Stains on Cloth

- Whole blood collected in EDTA (purple top) tubes
- 1:9 dilution prepared in physiological saline
- 10 µl of 10% blood pipetted directly onto cotton cloth
- Dry overnight (sampled 1 week later)
- microFLOQ® swabs moistened with 1µl molecular grade water
- Sampling
  - 3 analysts; 10 samplings center, 10 samplings edge (n=60)
  - Direct amplification (29 cycles) with PowerPlex® Fusion 6C System

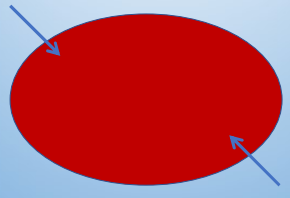
17

Center/Edge Results



18

Mixture Sampling Strategies to Improve Data Interpretation



More training on sampling

19

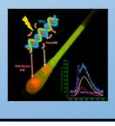
Finding Invisible DNA – Touch DNA

- Difficult to collect
- Unknown if swab contains DNA after collection unless analyzed
- Costly and laborious process

20

Fluorescence Detection

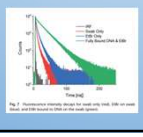
- Fluorescence methods have been reported to visualize DNA on swabs and other substrates
  - P. Kanokwongnuwut, et al Forensic Sci. Int.: Genet., 2018, 37:95 and Forensic Sci. Int., 2018, 291:115
- Background fluorescence reduces swabs or substrates that can be assessed
- Impacts sensitivity of detection

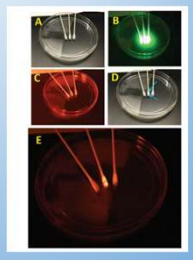


21

Fluorescence Detection

- A – swabs
- B – illuminated by 535 nm laser
- C – illuminated by 535 nm laser as seen by 575 nm long pass filter
- D – middle swab deposited DNA
- E – left swab (nothing); right swabs EtBr
  - illuminated by 535 nm laser as seen by 575 nm long pass filter

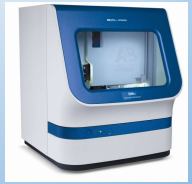




22

Capillary Electrophoresis

- Genetic Analyzer
  - Validated
  - Commonly used worldwide
  - Substantial experience
  - Used for databasing
  - Many protocols
    - Sanger sequencing -mtDNA
    - STRs
    - Microhaplotypes

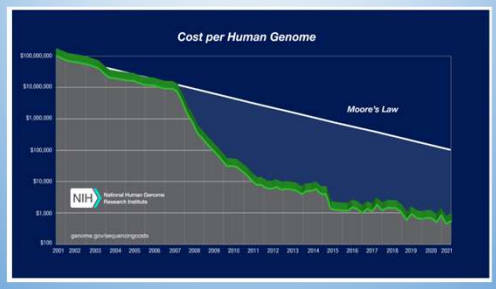


23

Technology Solutions

- Indeed, the success of PCR/CE has brought these new challenges
  - Low quantity/low quality
  - Interpretation
    - Mixtures
    - Kinship
  - Sample processing
  - Sample manipulation
  - Throughput
  - Cost-Benefit

24



25

How Will Genetics Change Our Lives Over the Next 50 Years ?

26

Massively Parallel Sequencing

- Increasingly user-friendly
- Highly automated workflow
- Compatible with numerous genetic marker types
- Available data analysis options
- Accurate, reliable data

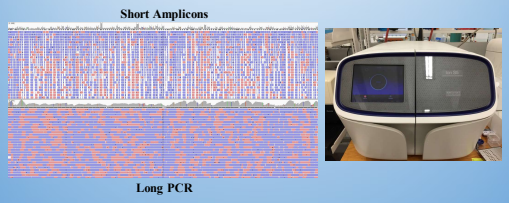
27

Sequencing and Forensic Applications

- Human ID
  - mtDNA
  - STRs
  - SNPs
  - Direct and Indirect
  - Mixtures
  - Ancestry
  - Phenotype
- Pharmacogenetics
  - Molecular autopsy
- Microbial Forensics

28

Huge Amounts of Sequence Data



[1] Churchill et al 2016
[2] King et al. 2014

29

Bioinformatics

- Unprecedented access to biological data
  - data acquisition
- Managing biological databanks with numerous contributors and users
  - store, organize, networks
- Extracting useful information from large and dense biological data
  - manipulate, visualize
- Assembling molecular pieces into predictive models of biological systems for *in silico* experiments
  - modeling, inference
  - scientific computing: multiprocessor, faster processors

30

Objectives

- Sample to answer
  - Simplicity
  - Flexibility
  - Data and information management
- What do the data mean?

31

Mitochondrial DNA

mtDNA is the most successful marker for typing degraded DNA

32

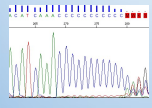
Variation Across the mtGenome
(n=283)

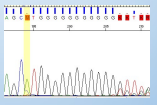
- 11,607 variants
  - defined in relation to the rCRS
- Polymorphism density clustered in HVI/HVII
  - 2,938 of the variants (25.3%)
- ~75% of variation in coding region

- Increase the value of mtDNA
- Analysis is rapid, less laborious, and less costly

33

Length Heteroplasmy
Example



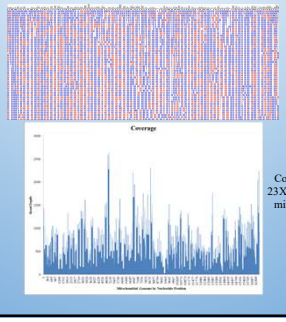


rCRS	A	A	A	A	C	C	C	C	C	T	C	C	C	C	
Length type I	A	A	A	C	C	C	C	C	C	C	C	C	C		
Length type II	A	A	A	C	C	C	C	C	C	C	C	C	C	C	
Length type III	A	A	A	A	C	C	C	C	C	C	C	C	C	C	C
Length type IV	A	A	A	C	C	C	C	C	C	C	C	C	C	C	C

Length heteroplasmy sequences detected using MPS

34

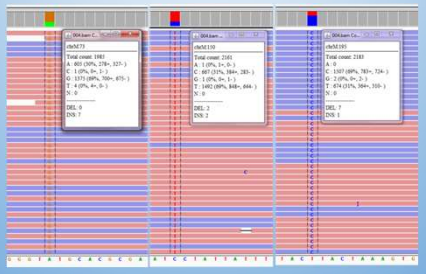
Hair Shaft



Coverage ranged from 23X to 2,634X across the mitochondrial genome

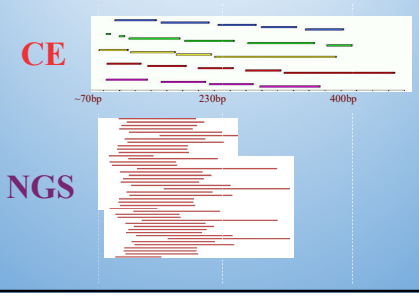
35

mtDNA Mixtures



36

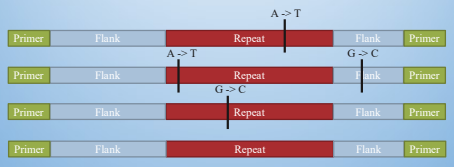
MPS and miniSTR Primer Design



37

STR Sequence Variation

- MPS provides nucleotide level data for high resolution genotyping



- Can generate size-based allele calling for backward compatibility

38

Population Studies

- Analysis of STR variation for 59 loci
- **777** individuals in **four** major population groups (African American, Hispanic, Caucasian, Chinese and Malaysian Chinese)
- **27** autosomal, **7** X chromosome and **25** Y chromosome STRs
- Library Preparation and sequencing performed using MiSeq FGx Forensic Genomics System (Illumina)
- Data analyzed using STRait Razor and in-house excel workbooks

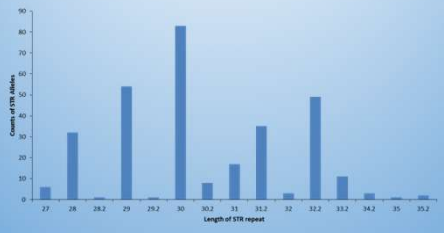
39

Population Study Results

40

D21S11 Allele Variation

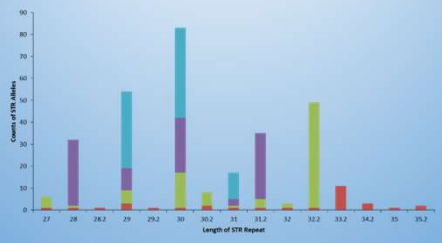
[TCTA], [TCTG] Repeat Motif



41

D21S11 Allele Variation

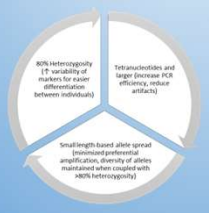
[TCTA], [TCTG] Repeat Motif

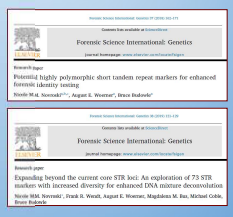


42

Search for More Resolving STRs

- 1000 Genomes Project (raw sequences, unsorted)
- STR Catalog Viewer
  - Summary of human STR variation compiled using lobSTR software




Forensic Science International: Genetics

Potential highly polymorphic short tandem repeat markers for enhanced forensic identity testing

Forensic Science International: Genetics

Expanding beyond the current core STR loci: An exploration of 73 STR markers with increased diversity for enhanced DNA mixture deconvolution

43

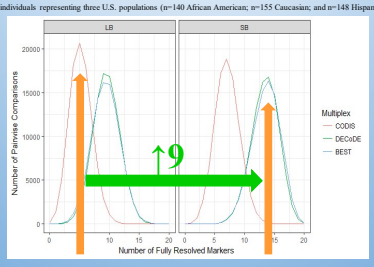
CODIS Panel Loci	Exploratory Panel Loci	BEST Panel Loci
D2S1338	D3S2406	D3S2406
D12S391	D2S1360	D2S1360
D1S1656	D7S3048	D7S3048
D21S11	D8S1132	D8S1132
D8S1179	D11S2368	D11S2368
vWA	D15S822	D15S822
D3S1358	D2N2	D2N2
D18S51	D1N10	D1N10
FGA	D12N15	D12N15
D19S433	D1N16	D1N19
D13S317	D1N19	D1N21
D5S818	D1N21	D8N23
D16S539	D8N23	D15N26
D22S1045	D15N26	D14N56
D7S820	D14N56	D3N61
D2S441	D3N61	D12S1338
CSF1PO	D4N70	D4N70
D10S1248	D11N52	D2S1338
TPOX	D17N32	D1S1656
TH01	D2N43	D11N52

44

In silico Mixtures

Top 20 Performance

N = 443 individuals representing three U.S. populations (n=140 African American; n=155 Caucasian; and n=148 Hispanic samples)

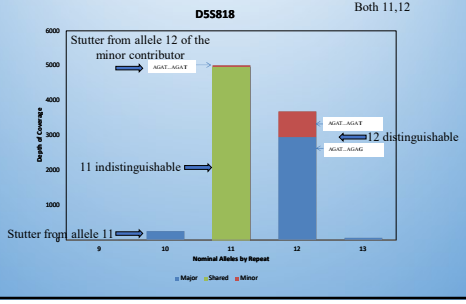


LB = length-based; SB = sequence-based

45

Minor and Major Contributor Alleles

Mixture of 2 people
Both 11,12

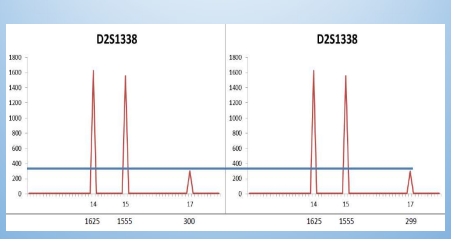


46

How to Interpret DNA Mixtures

47

Binary v Continuous



48

Binary Methods

- Assign 1 for combination of genotypes that best explain the profile
- Assign 0 for poor explanation
  - Typically these explanations are not considered

Major	Minor	$p(O\|S_j)$
13,15	12,14	1
12,13	14,15	0
12,15	13,14	0
13,14	12,15	0
14,15	12,13	0
12,14	13,15	0

49

Continuous Methods

- The probabilities are continuous
- Range between 0 and 1

Major	Minor	$p(O\|S_j)$
13,15	12,14	0.79
12,13	14,15	0.05
12,15	13,14	0.05
13,14	12,15	0.05
14,15	12,13	0.05
12,14	13,15	0.01

50

MCMC

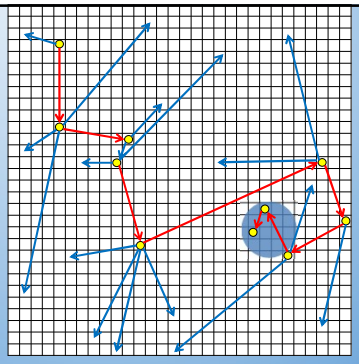
best answer

A possible answer

MCMC 'chain'

Hot guess

Cold guess

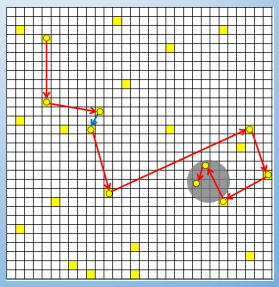


Slide Courtesy of Jo Bright (ESR)

51

MCMC Process

- 10 MCMC moves
  - Only small fraction of all possible answers need to get to best explanation
- If ran again
  - Probably different starting points
  - Would still reach best explanation
- Can run for greater number of steps

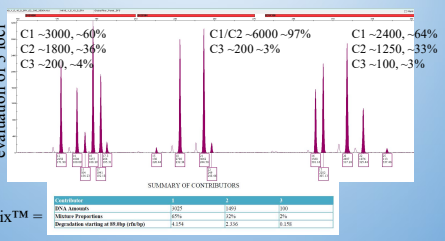


52

Casework Interpretation Example

Interpret More Data

Analyst approximate evaluation of 3 loci

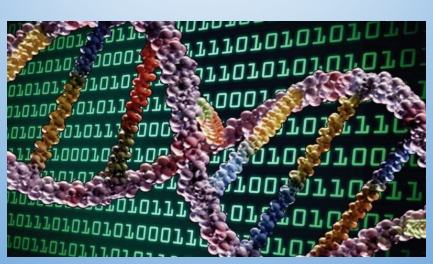


STRmix™ =

Actual = Mixture of 3 individuals, prepared as a 10:5:1, averaged unique alleles = 64% 30% 6%

53

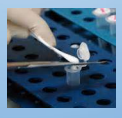
DNA Databases



54

CODIS

- CODIS blends forensic science and computer technology into an effective tool for solving violent crimes and identify the missing
- CODIS enables crime labs to exchange and compare DNA profiles electronically, thereby linking crimes to each other and to convicted offenders



55

Direct Match

Marker	Evidence	Suspect
D3S1358	15,16	15,16
vWA	17,18	17,18
FGA	19,22	19,22
D8S1179	13,15	13,15
D21S11	29,30	29,30
D18S51	14,15	14,15
D5S818	11,11	11,11
D13S317	11,12	11,12
D7S820	11,11	11,11
CSF1PO	11,13	11,13
TPOX	8,11	8,11
TH01	6,9.3	6,9.3
D16S539	11,13	11,13

56

Genetic Genealogy Approaches

- If no direct hit in database:
  - Extend investigative lead through indirect associations (some use term "hits")
  - Looking for relatives
  - Tools
    - Familial searching
    - Y-STR searching
      - Note mtDNA could be used as in missing persons identification work (but labor intensive)
    - Dense SNP data

57

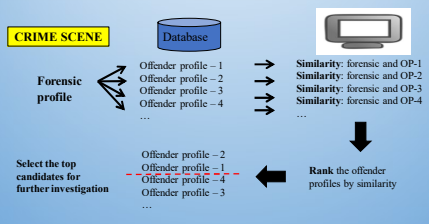
Familial Searching

Keith Davison is on a sex offenders register for life
United Kingdom case

58

Familial Searching

Familial searching is an attempt to detect potential relatives of a forensic profile in any specified database



59

Netherlands - Vaatstra Case

- 1999, Marianne Vaatstra was raped and murdered; semen found
- Y-STR test supported origin from North-Western Europe
- 2012, FS did not generate an investigative lead
- 7,600 local males asked to voluntarily provide DNA samples to build a local forensic Y-STR database; 6,600 participated
  - 2 Y-STR matches (different surnames) in the first 81 volunteers, both excluded
  - Genealogy study confirmed they shared same paternal ancestor
  - Their two surnames used to search volunteer samples
  - One of the subsequent samples had the same Y-STR and autosomal STR profiles as those of the crime scene evidence

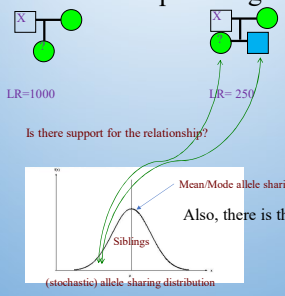
60

Real Case

- Skull
  - Only a mtDNA profile was generated
  - CODIS search has a hit with a pedigree containing a single individual, a sister of a missing person
  - LR =163
- A few years later lower half of a skeleton was found in the same county
  - Same mtDNA profile
  - Full STR profile
  - LR = 0.031
- Combined LR ~5

61

Kinship Testing



Also, there is the half-sib problem

62

Improve Kinship Resolution

- Uninformative analyses
  - More relatives – limited
  - More DNA – sample dependent
  - More markers – alternate technologies
    - Autosomal STRs are not enough
- Also can be used to develop leads in other cases

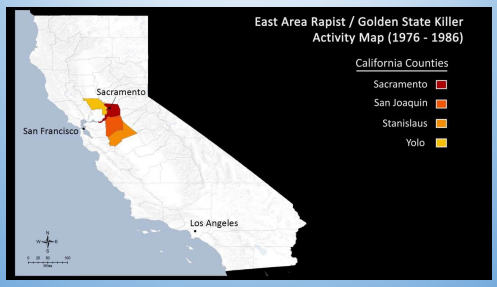
63

Joseph James DeAngelo

- One of the most notorious serial killers in California
- Years active 1974 – 1986

64

East Area Rapist (1976-1979)



65

Original Night Stalker (1979 – 1986)

DNA testing in 2001 connects EAR to ONS "Golden State Killer"

120+ Burglaries
50+ Sexual Assaults
12+ Homicides

66

Public Safety

To find alleged Golden State Killer, investigators first found his great-great-great-grandparents

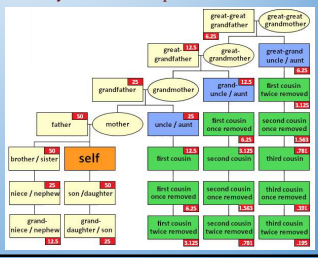
- [DNA expert Paul] Holes used DNA recovered from a crime scene to find the killer's great-great-great grandparents, who lived in the early 1800s.
- Branch by painstaking branch, he and a team created about 25 family trees containing thousands of relatives down to the present day.
- Sharing of the DNA in this case was to the level of 3rd cousins (10-20 total)

67

Genetic Genealogy

- %DNA shared
- Enhances options of family reference samples

ISOGG wiki statistics:
Chromosome sharing ---
Parent/child: 3539-3748 cMs
1st cousins: 548-1034 cMs
1st cousins 1R: 248-638 cMs
2nd cousins: 101-378 cMs
2nd cousins 2R: 43-191 cMs
3rd cousins: 43-ca 150 cMs
3rd cousins 1R: 11.5-99 cMs
More distant cousins: 5-ca 50 cMs



68

Genome Wide DNA Samples

Company	23andMe	Family Tree DNA's Family Finder test	Ancestry.com's AncestryDNA test
Price (as of 11 January 2018)	$199 for the lifetime of the platform in the U.S.	$79 for the lifetime of the platform	$69 in the U.S.
SNP chip used for testing	Customized Illumina GSA chip	Illumina OmniExpress	Customized Illumina chip
Number of autosomal SNPs tested	630,132	about 690,000	637,639
Number of people in the database (as of 5 May 2018)	**5,000,000**	**About 850,000**	**over 9,000,000**



- Feb 2019 - 26 million profiles/customers in DTC
  - Ancestry – 14 million
  - 23andme – 9 million
  - All others – 3 million +
- Ancestry 15 million (May 31, 2019)

https://www.cnbc.com/2019/02/12/privacy-concerns-rise-as-26-million-share-dna-with-ancestry-firms.html
https://blogs.ancestry.com/ancestry/2019/05/31/ancestry-surpasses-15-million-dna-customers/

69

GEDmatch

- Open-source personal genomics database and genealogy website
- Now owned by Verogen
- 270,000 opt-in (mid 2020)
- 1.9 million genetic profiles

https://blog.nebula.org/wp-content/uploads/2020/04/transfer_ancestry_dna_to_family_tree_5.png

70

ForenSeq® Kintelligence Kit

- SNP-based kinship analysis and relatedness inference can be broadly classified into two categories: (1) genome-wide relatedness methods , and (2) identity by descent (IBD)
- SNP-based* as opposed to segment based
- 12 paired-end, dual-indexed libraries for "long-range" kinship analysis
- SNPs directly compatible with GEDmatch PRO®
- Small amplicon size averaging < 150 bp
- Unique Dual Index (UDI) adapters that have distinct, unrelated adapters for both index reads, preventing repeated sequences in a plate for optimum data recovery
- Sequence up to three libraries on a MiSeq FGx® standard flow cell: one positive amplification control, one negative amplification control, and one sample

Table 1: Kit specifications

Specification	Value
Sample types	Blood, bone, buccal swabs, hair, teeth, and semen
Recommended input	1 ng gDNA per sample
Multiplexing capacity	3 libraries per run
Kit configuration	12 reactions
Number of SNPs	10,230
Mean amplicon size	< 150 bp
Total library prep time	8 hours
Hands-on library prep time	1 hour and 55 minutes
Sequencing time	28 hours

71

ForenSeq ® Kintelligence Kit

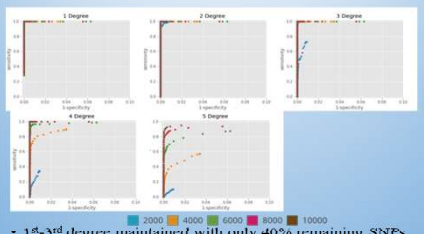
- 10,230 SNPs
- SNPs overlap with Illumina Infinium CytoSNP-850K BeadChip and Infinium Global Screening Array
- SNPs cross-referenced against the Genome Aggregation Database (gnomAD) v3.0 and Single Nucleotide Polymorphism database (dbSNP) v151
- For robust performance across global populations.
- Excludes the SNPs with known medical associations or low minor allele frequencies
- To reduce privacy concerns and protect genetic health data

Category	Number of SNPs	Percentage of Total
Ancestry SNPs	56	0.5%
Identity SNPs	94	1%
Kinship SNPs	9867	96%
Phenotype SNPs*	22	0.2%
X-SNPs	106	1.2%
Y-SNPs	85	0.8%

* Two SNPs overlap the ancestry and phenotype categories and are counted in the phenotype category only.

72

Performance of Windowed Kinship with Call Rates Between 20-100%



- 1st-3rd degree maintained with only 40% remaining SNPs
- 4th degree 80% SNPs give comparable performance with the full set and 60% is sufficient to make some calls

GEDMatch test set

73

Benefits

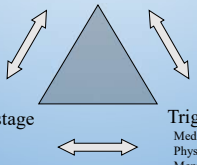
- Can extend beyond traditional search
  - Direct matches
  - Paternity trio/first order relatives
- Reduce need for Y STR or mitochondrial DNA-based reference samples
- Generate leads for difficult, old, and cold cases
- Chance of error is small
  - Depending on how error is defined
  - Some flexibility with accuracy of kinship assessment

74

Genetics of Death

Predisposition to Sudden Unexpect Death

Genetic variants



Developmental stage
- Neonates
- Infants
- Juveniles
- Adults
- Elderly

Triggering event
- Medication
- Physical stress
- Mental stress
- Climate change
- Nutritional change

75

Forensic Example

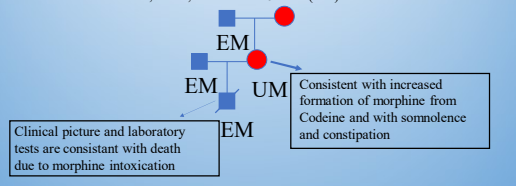
- Codeine
- Infant died of morphine overdose at 13 days old
  - Mother was prescribed Tylenol #3 (acetaminophen and codeine)
  - Codeine is metabolized into morphine
  - Mother was an ultra rapid metabolizer

http://www.ilike.com/user/codeine0yr

76

Morphine Poisoning in a Breastfed Neonate

- Genetics
- Mother: CYP2D6*2A / *2x2 (UM)
- Grandfather, father, son: CYP2D6*1/*2 (EM)



poor (PM), intermediate (IM), extensive (EM), or ultrarapid (UM) metabolizer

Koren et al. The Lancet, 2006

77

Enhancing Training

- Significant effort and resources are dedicated to developing and validating the methods.
- Often ignored is the equally important need for training and education.
- Forensic scientists must have knowledge of the theory and procedures of new and existing technologies to properly implement them and more so to properly interpret the data generated.
- Non-scientist stakeholders and customers must become sufficiently knowledgeable to understand value and implications.

78

Training Challenges

- The rapidly developing use of molecular genomics in the forensic field places a demand on suitable education and training for current employees as well as students.
- Traditionally education and training in forensic genomics rely heavily on wet lab analyses, which is labor intensive, time consuming and costly. The instrumentation and consumables required are expensive and involve specialized handling and storage.
- Hands-on training using these reagents and equipment often necessitates repetitive in-house instruction, which can be difficult when the instruments in some laboratories are fully committed to operational use.
- Most trainers in forensic laboratories have other duties.
- Since most forensic laboratory training is in house and not standardized, there can be significant variation in the quality and amount of training experiences across the country and the world.

79

Virtual Reality Enhancing Training

80

Programs

- Fundamentals of DNA
- Lexicons
- Technologies
- Human Trafficking
- Crime Scene Investigation
- Mixture Interpretation
- Sexual assault Nurse Examinations
- Topics are endless

- Consistent training
- Guaranteed fundamentals
- Cost effective
- Repetitive training
- Reinforcement
- …

81

ACKNOWLEDGMENTS

- DNAfor Africa
- ICRC
- UNODC
- Finnish Institute of Health and Welfare University of Helsinki
- Thermo Fisher Scientific
- Verogen

- Antti Sajantila
- Magdalena Bus
- Jianye Ge
- Ignacy Gryczynski
- Zygmunt Gryczynski
- John Buckleton

82

Questions?

83