

Expert Systems for Automated STR Analysis

SWGAM
Quantico, VA

Mark W. Perlin
January, 2003



DNA Is Clue to 1971 Murder: A recent, random test links a state prison inmate to the old El Dorado case. January 10, 2003

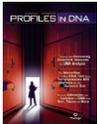
DNA Clue Cracks Open Unsolved 1979 Slaying: Colorado felon charged in San Pablo girl's death. December 4, 2002

S.F. Transient Held in Rapes of Homeless Women: DNA match led to suspect. October 19, 2002

DNA links felon to rape: The arrest marks the 100th match in state database. August 23, 2002

DNA links parolee to old rape case: Database helps authorities score 'cold hit' on suspect in attack. August 21, 2002

DNA yields arrest warrant in 1978 killing. August 14, 2002



Dealing with Increasing Casework Demands for DNA Analysis

By Joseph Varlaro¹ and Barry Duceman²
¹Boston Police Crime Laboratory Unit and ²New York State Police

The analyst will simply submit the items for DNA analysis, using the **final data interpretation step to determine relevance to the ongoing investigation ... The bottleneck becomes the interpretation of analytical results and the technical review process. ... Although these processes are currently dependent upon manual applications, software solutions are emerging that can be integrated into an automated approach.**

Collect Crime Scene Evidence

Generate DNA Data ★

Review Data

Present Results to Legal System



Collect Crime Scene Evidence

Generate DNA Data ★

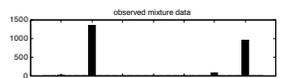
Review DNA Data ★

Present Results to Legal System

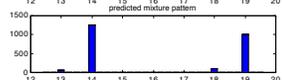
TrueAllele™



data



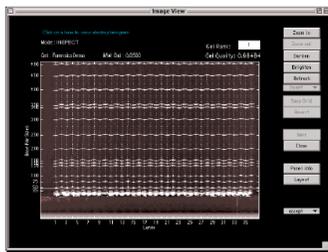
**data peaks
size + [DNA]**



model



1. Input data
2. Q/C gel run
3. Call alleles
4. Output result



The image shows a software interface with two main panels. The left panel is titled 'Ladder Values' and contains a table with the following data:

| | |
|-----------|----|
| 17 | 22 |
| 1911.1906 | |
| 91655.50 | X |
| 49162.0 | |
| 1911 | |
| 1910 | |
| 639.10 | |
| 0.75820 | |

The right panel is titled 'Basic Analysis' and contains several checkboxes and a table. The checkboxes are:

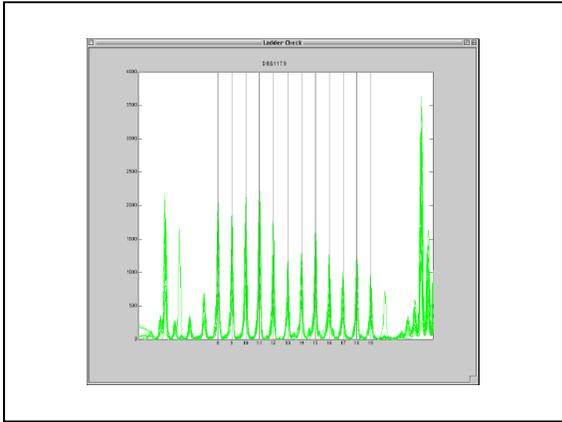
- Rules
- score distance
- ladder missing
- ladder overlap
- repeat vs peak
- peak vs missing
- primer dimer

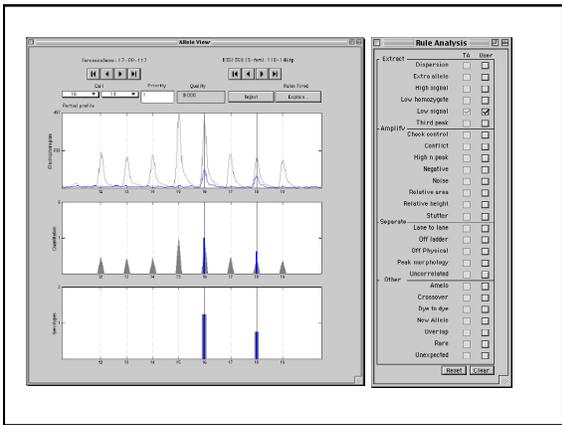
Below the checkboxes, there is a section for 'Markers that fit the ladder' with the following data:

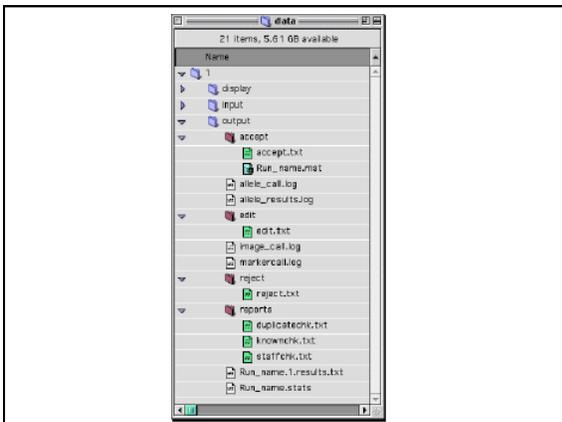
| | |
|----------|--|
| 749106 | |
| 61655.59 | |

At the bottom, there is a 'Positive Control Check' table:

| Position | Len | Result |
|----------|-----|--------|
| 99476 | 21 | Pass |
| 99476 | 29 | Pass |







FSS ABI/377 Validation

Resources

- Data: 22,000 genotypes (SGMplus)
- People: 6 reviewers + 6 managers
- Time: 8 weeks work + 4 weeks report

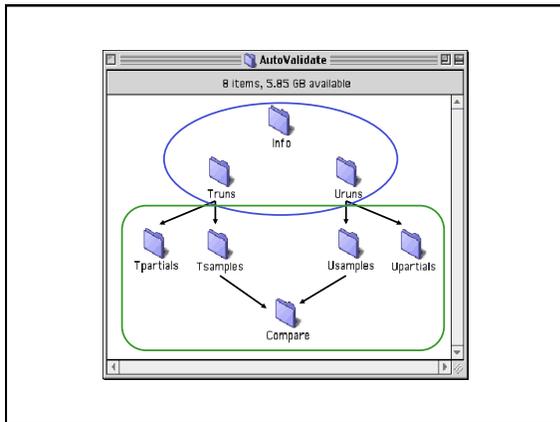
Components

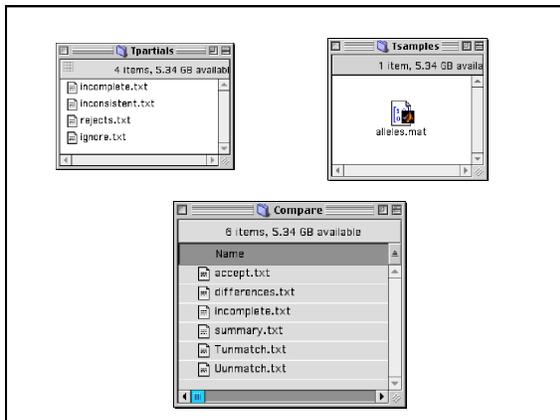
- Peak height correlation (GS vs TA)
- Establish baseline height (error-free)
- Designation accuracy (human vs TA)
- Network/computer environment
- QMS documentation

Results

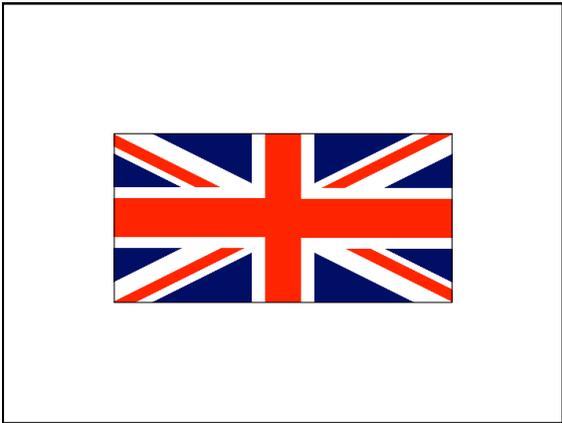
- Greater yield with TA
- No errors on quality data



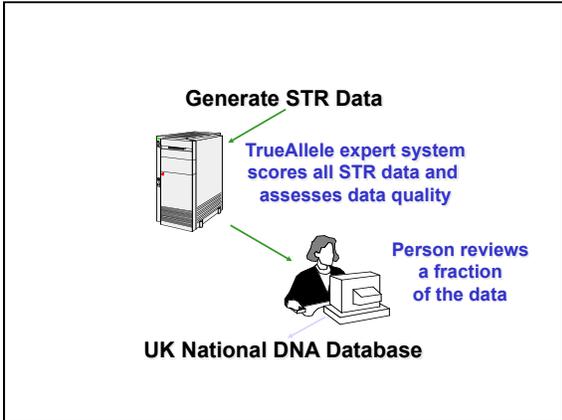


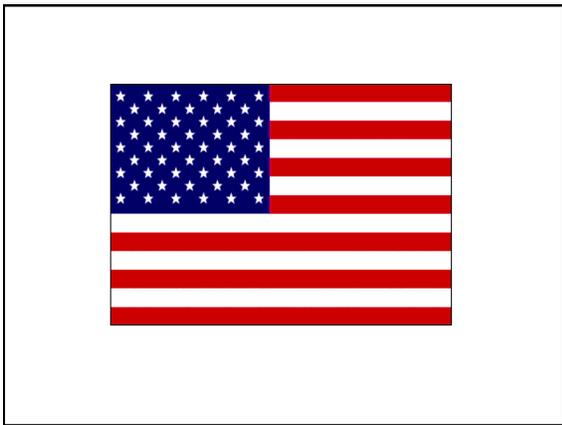




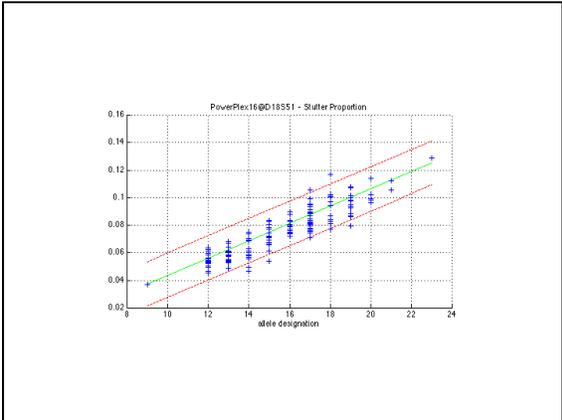


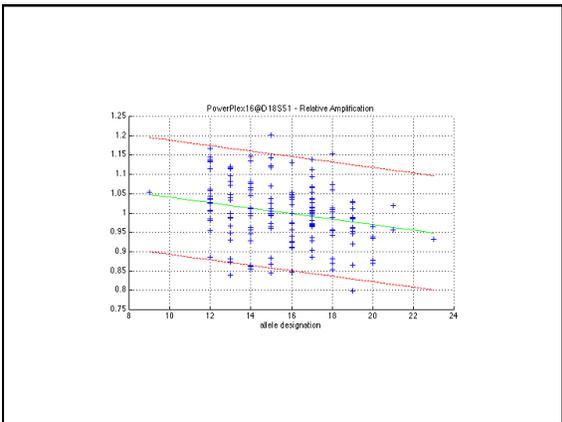


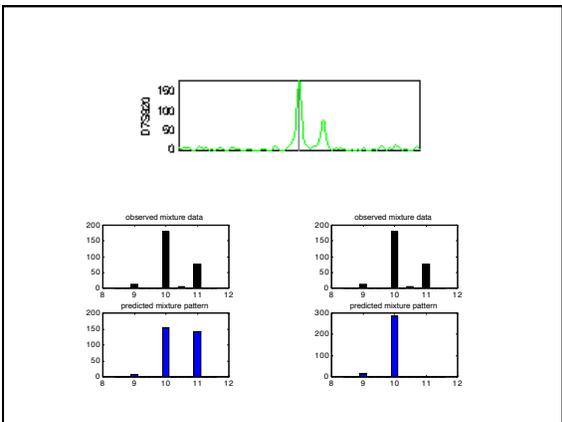




- Validation Method**
1. Obtain original data
 2. Process data in TrueAllele™ ES
(auto-setup, process run, Q/A, call alleles, apply rules, check)
computer: accept/reject/edit
 3. Review all data
one person, many computers
human: accept/reject/edit
 4. Generate results & stats







Genotype Probability

Sample @ D7S820

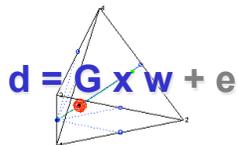
| | | | |
|----------|----|----|-------|
| Option 1 | 10 | 11 | 72.4% |
| Option 2 | 10 | 10 | 27.6% |

99% Confidence Allele Set = { 10, 11 }

 Database Searching

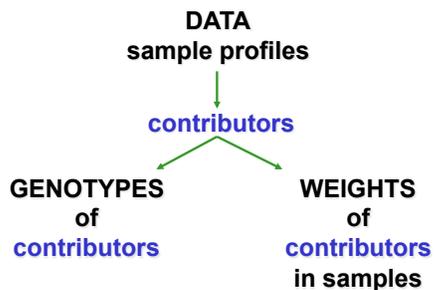
DNA Mixture Model

Linear Mixture Analysis


$$d = G \times w + e$$

M.W. Perlin and B. Szabady, "Linear mixture analysis: a mathematical approach to resolving mixed DNA samples," *Journal of Forensic Sciences*, November, 2001.

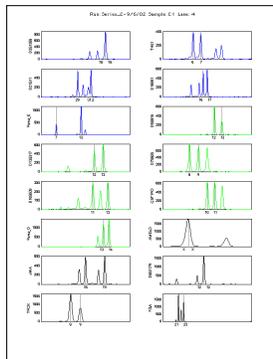
The Contributor Problem



One Sample

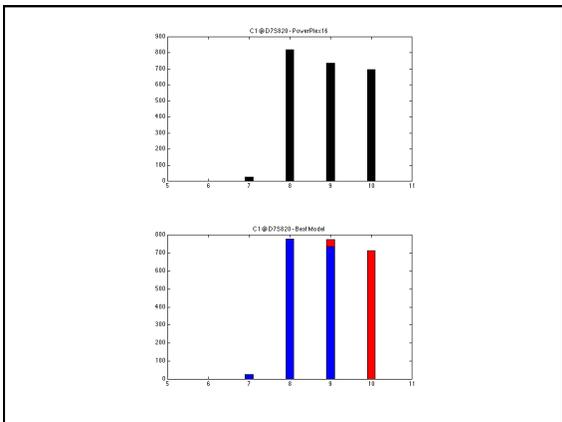
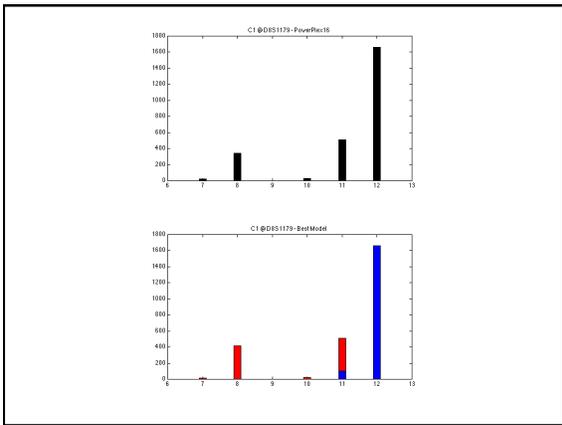
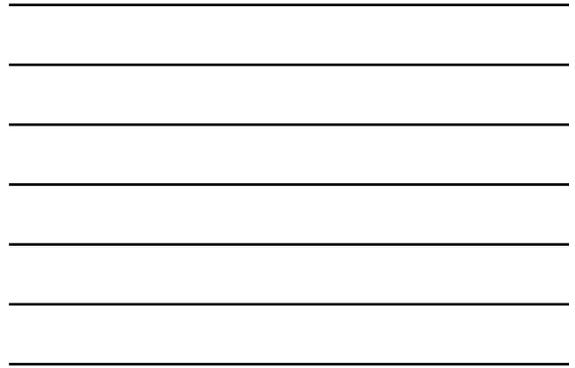
1 ng DNA
PowerPlex16
ABI/310

Sample C:
Unknown (A) 70%
Unknown (G) 30%



| | | sample |
|--------------|---|------------|
| | | mixture |
| contributors | 1 | 67% +/- 2% |
| | 2 | 33% +/- 2% |

| | | | | | |
|---------|----|-----|---------|---------|---|
| AMEL | 1 | 2 | 0.8699 | 0.8699 | 1 |
| | 2 | 2 | 0.12778 | 0.99759 | 0 |
| CSF1PO | 12 | 12 | 0.41506 | 0.41506 | 1 |
| | 10 | 10 | 0.29542 | 0.71548 | 0 |
| | 11 | 11 | 0.28851 | 1 | 0 |
| D13S217 | 9 | 13 | 0.92743 | 0.92743 | 1 |
| D16S539 | 9 | 12 | 0.81981 | 0.81981 | 1 |
| | 12 | 12 | 0.04367 | 0.86348 | 0 |
| | 11 | 11 | 0.03789 | 0.90137 | 0 |
| D18S51 | 13 | 15 | 0.99994 | 0.99994 | 1 |
| D21S11 | 30 | 31 | 0.99689 | 0.99689 | 1 |
| D2S1358 | 16 | 17 | 0.99978 | 0.99978 | 1 |
| D5S818 | 12 | 12 | 0.50257 | 0.50257 | 1 |
| | 13 | 13 | 0.49203 | 0.99966 | 0 |
| D7S820 | 10 | 10 | 0.38447 | 0.38447 | 1 |
| | 9 | 9 | 0.33296 | 0.71745 | 0 |
| | 8 | 8 | 0.28255 | 1 | 0 |
| D8S1179 | 8 | 11 | 1 | 1 | 1 |
| FGA | 21 | 22 | 1 | 1 | 1 |
| Penta_D | 12 | 14 | 1 | 1 | 1 |
| Penta_E | 7 | 14 | 0.9988 | 0.9988 | 1 |
| TH01 | 9 | 9.3 | 0.96943 | 0.96943 | 1 |
| TPDX | 9 | 9 | 0.52604 | 0.52604 | 0 |
| | 8 | 8 | 0.47396 | 1 | 1 |
| vWA | 15 | 16 | 1 | 1 | 1 |



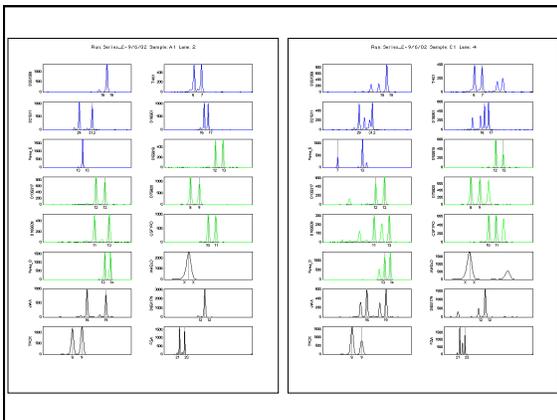
| Set 1 | | ng amplified | | | |
|---------------|----|--------------|------|-------|--|
| Mixture ratio | 1 | 0.5 | 0.25 | 0.125 | |
| 10:0 | A1 | A2 | A3 | A4 | |
| 7:1 | B1 | B2 | B3 | B4 | |
| 7:3 | C1 | C2 | C3 | C4 | |
| 5:5 | D1 | D2 | D3 | D4 | |
| 3:7 | E1 | E2 | E3 | E4 | |
| 1:9 | F1 | F2 | F3 | F4 | |
| 0:10 | G1 | G2 | G3 | G4 | |

| Set 2 | | ng amplified | | | |
|---------------|----|--------------|------|-------|--|
| Mixture ratio | 1 | 0.5 | 0.25 | 0.125 | |
| 10:0 | H1 | H2 | H3 | H4 | |
| 9:1 | I1 | I2 | I3 | I4 | |
| 7:3 | J1 | J2 | J3 | J4 | |
| 5:5 | K1 | K2 | K3 | K4 | |
| 3:7 | L1 | L2 | L3 | L4 | |
| 1:9 | M1 | M2 | M3 | M4 | |
| 0:10 | N1 | N2 | N3 | N4 | |

Two Samples

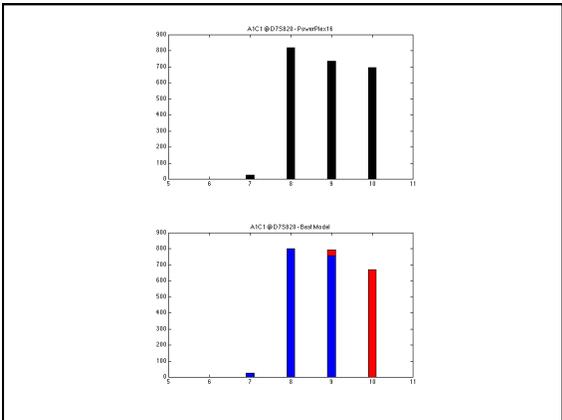
1 ng DNA
PowerPlex16
ABI/310

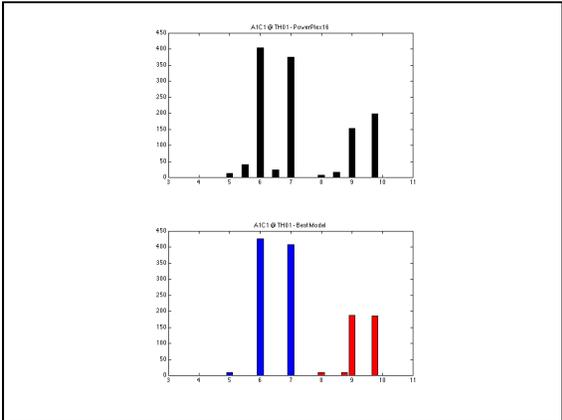
Sample A: Reference
&
Sample C:
Reference (A) 70%
Unknown (G) 30%



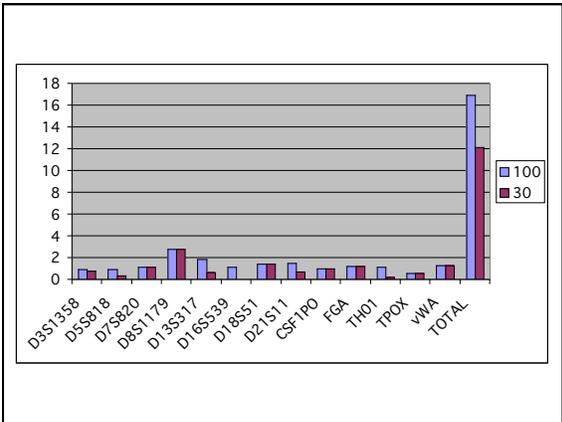
| | | samples | |
|--------------|---|---------|------------|
| contributors | | ref | mixture |
| | 1 | 100% | 68% +/- 2% |
| | 2 | 0% | 32% +/- 2% |

| | | | | | |
|---------|-----|---------|---------|---------|---|
| CSF1PO | 12 | 12 | 1 | 1 | 1 |
| D13S317 | 9 | 13 | 0.94678 | 0.94678 | 1 |
| | 12 | 13 | 0.03754 | 0.98432 | 0 |
| D16S539 | 9 | 12 | 0.89603 | 0.89603 | 1 |
| | 12 | 12 | 0.05856 | 0.95259 | 0 |
| | 12 | 13 | 0.01406 | 0.96866 | 0 |
| | 11 | 12 | 0.01028 | 0.97634 | 0 |
| | 10 | 12 | 0.00444 | 0.98138 | 0 |
| | 6,3 | 12 | 0.00423 | 0.98051 | 0 |
| 7,3 | 12 | 0.00336 | 0.98897 | 0 | |
| D18S51 | 13 | 15 | 0.99997 | 0.99997 | 1 |
| D21S11 | 30 | 31 | 0.99799 | 0.99799 | 1 |
| D3S1358 | 16 | 17 | 0.99889 | 0.99889 | 1 |
| D5S818 | 12 | 12 | 0.9947 | 0.9947 | 1 |
| D7S820 | 10 | 10 | 1 | 1 | 1 |
| DBS1179 | 8 | 11 | 1 | 1 | 1 |
| FGA | 21 | 22 | 1 | 1 | 1 |
| Festa_D | 12 | 14 | 1 | 1 | 1 |
| Festa_E | 7 | 14 | 0.99921 | 0.99921 | 1 |
| TH01 | 9 | 9,3 | 0.97793 | 0.97793 | 1 |
| | 6 | 9,3 | 0.00799 | 0.98052 | 0 |
| | 9,3 | 9,3 | 0.00314 | 0.98805 | 0 |
| TPDX | 8 | 8 | 1 | 1 | 1 |
| vWA | 15 | 18 | 1 | 1 | 1 |

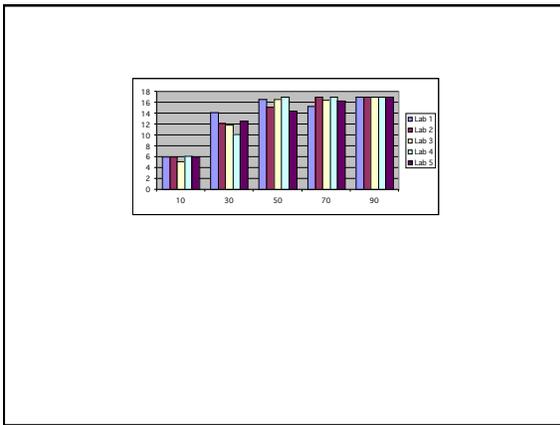


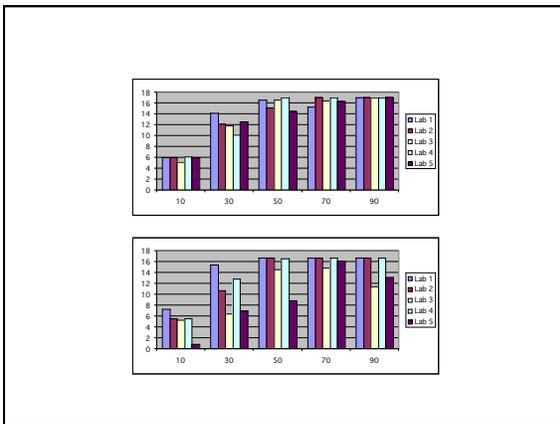


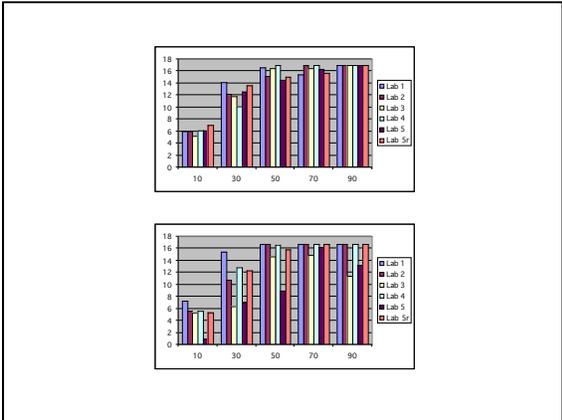
| TH01 | 6 | 7 | 9 | 9.3 |
|------|---|---|---|-----|
| 6 | X | X | X | O |
| 7 | X | X | X | X |
| 9 | X | X | X | O |
| 9.3 | O | X | O | O |

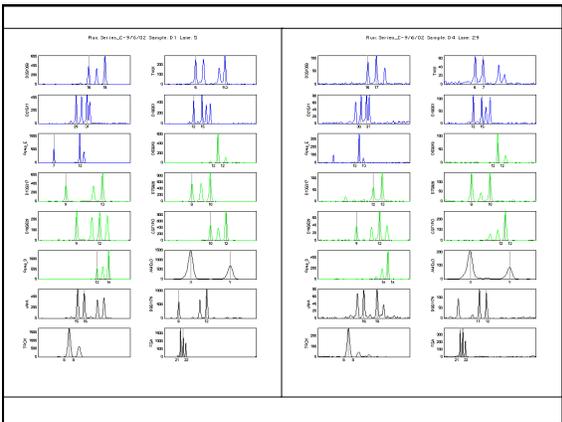


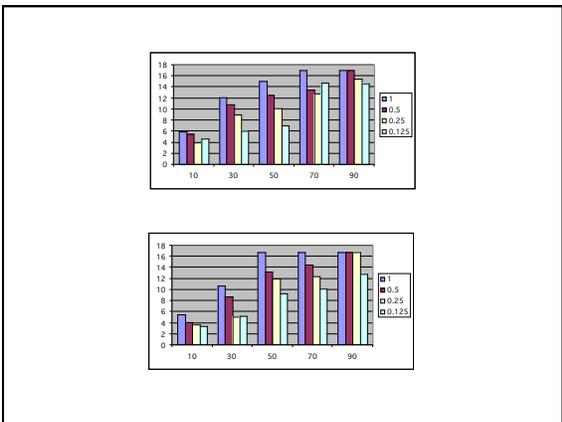










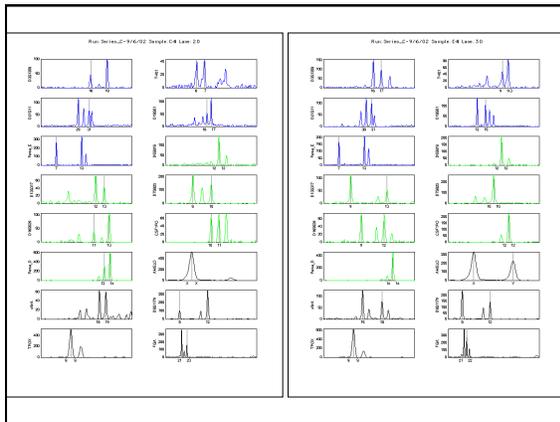


Three Samples

1, 1/2, 1/4, 1/8 ng DNA
PowerPlex16
ABI/310

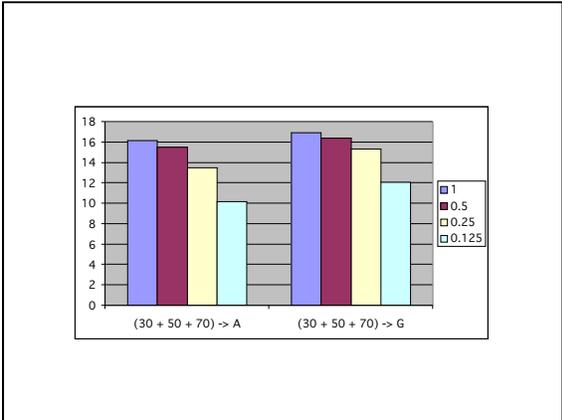
Two Contributors, No Reference

Sample C: (A) 70% (G) 30%
Sample D: (A) 50% (G) 50%
Sample E: (A) 30% (G) 70%



| | | samples | | | |
|--------------|--|---------|-----|-----|------|
| contributors | | C1 | D1 | E1 | |
| 1 | | 67% | 48% | 27% | 1 ng |
| 2 | | 33% | 52% | 73% | |

| | | samples | | | |
|--------------|--|---------|-----|-----|--------|
| contributors | | C4 | D4 | E4 | |
| 1 | | 66% | 45% | 33% | 1/8 ng |
| 2 | | 34% | 55% | 67% | |



Collect Crime Scene Evidence

Generate DNA Data ★

Review DNA Data ★

Present Results to Legal System



PROFILES IN DNA
Dealing with Increasing Casework Demands for DNA Analysis
 By Joseph Varlaro¹ and Barry Duceman²
¹Boston Police Crime Laboratory Unit and ²New York State Police

The analyst will simply submit the items for DNA analysis, using the final data interpretation step to determine relevance to the ongoing investigation ... The bottleneck becomes the **interpretation of analytical results and the technical review process**. ... Although these processes are currently dependent upon manual applications, **software solutions** are emerging that can be integrated into an **automated approach**.
