Taxonomic Identification of Traces using Non-human DNA

David A. Stoney
Stoney Forensic, Inc.
14101-G Willard Road, Chantilly, VA 20151
Taxonomic Identification of Traces using Non-human DNA

• Differences from human DNA identification applications
• Overview of the approach
• Examples of applications to predictive source attribution cases
DNA
Mutations

Wild-type gene

Transition (AT pair replaced by GC pair)

Transversion (AT replaced by TA)

Insertion (GC pair inserted)

Deletion (AT pair deleted)
More Mutations

- Inversion
- Duplication
- Deletion
- Insertion
- Translocation
The Role of Mutation and Inheritance in DNA Identification

• Mutations arise in individuals
• Mutations spread through populations
• Mutations may become fixed or polymorphic in:
  – Populations
  – Varieties
  – Species
  – Genera, etc.
Fragment Length vs. Specific Sequence

Example of Fragment Length Polymorphism:

ATATATA
AATATAAT
AATATAAAT
AAATATAAAATA

Example of Specific Sequence Comparison:

ATAGATACATTCCATATCCGG
ATAGATTCATACCTTAGGCGG
ATAGATTCATACCTTTAGGCGG
Types of Molecular Markers

- Variable Number Tandem Repeats (VNTR)
- Restriction Fragment Length Polymorphism (RFLP)
- Amplified Fragment Length Polymorphism (AFLP)
- Short Tandem Repeat (STR)
- Single Nucleotide Polymorphism (SNP)
- DNA Sequence Data
Choosing a Molecular Marker

• What is the required level of identification?
  – Family vs Genus and Species
  – Species vs Population
  – Population vs Individual

• What reference data are available?
Traditional DNA Applications

• Forensic
  – Human Identification
• Human
  – Medical Genotyping Applications
• Non-Human
  – Evolutionary and Developmental Biology
Identification at the Population or Individual Level

- Requires Population Frequency Data
  - Combined DNA Index System (CODIS) uses Short Tandem Repeats or STRs
Human Identification Markers

• STR Profile, i.e. CODIS
• Mitochondrial Sequence
• Y-STR Profile
• Region of Origin (Population Level)
  – Alu Repeats
  – Y-STRs
  – SNPs
Human Identification Markers

• STR Profile, i.e. CODIS
• Mitochondrial Sequence
• Y-STR Profile
• Region of Origin (Population Level)
  – Alu Repeats
  – Y-STRs
  – SNPs

All within one species
Non-Human Identification Markers

• DNA Sequence Data
  – Plants: rbcL, trnL, ITS
  – Vertebrates: Cytochrome B
  – Arthropods: 16S, Cytochrome Oxidase
  – Fungi: ITS, 18S
  – Bacteria: 16S

• Population Level Data
Non-Human Identification Markers

• DNA Sequence Data
  – Plants: rbcL, trnL, ITS  Among 300,000 species
  – Vertebrates: Cytochrome B
  – Arthropods: 16S, Cytochrome Oxidase
  – Fungi: ITS, 18S
  – Bacteria: 16S

• Population Level Data
Available Reference Data

• Published and Unpublished Gene Frequency Data for Populations

• GenBank DNA Sequence Database
  – National Center for Biotechnology Information (NCBI)
Available Reference Data

- Published and Unpublished Gene Frequency Data for Populations
- GenBank DNA Sequence Database
  - National Center for Biotechnology Information (NCBI)

References are abundant for species with economic, medical, or academic interest
Sample Processing Overview

- DNA Extraction
- PCR Amplification of Selected Markers
- Obtain Sequence
- BLAST Inquiry
- Phylogenetic Analysis
- Develop Inferences from DNA Analyses
Single Elements vs. Mixtures
Sample Processing Overview

1. DNA Extraction
2. PCR Amplification of Selected Markers
3. Obtain Sequence
4. BLAST Inquiry
5. Phylogenetic Analysis
6. Develop Inferences from DNA Analyses

Direct Amplification
Sample Processing Overview

1. DNA Extraction
2. PCR Amplification of Selected Markers
3. Obtain Sequence
4. BLAST Inquiry
5. Phylogenetic Analysis
6. Develop Inferences from DNA Analyses

Mixture

Direct Amplification

Stoney Forensic, Inc.  
NIJ/FBI 2011 Trace Evidence Symposium  8-11-11
Sample Processing Overview

- DNA Extraction
- PCR Amplification of Selected Markers
- Obtain Sequence
- BLAST Inquiry
- Phylogenetic Analysis
- Develop Inferences from DNA Analyses

Mixture
- Ligation
- Transformation
- Screening of Clones

Direct Amplification

Stoney Forensic, Inc.
NIJ/FBI 2011 Trace Evidence Symposium  8-11-11
DNA Extraction
DNA Quantification

• Standard protocol in many procedures
  – UV absorption at specific wavelengths
  – Comparison to standards of known concentration

• Trace samples can often be assumed to have a minimal DNA concentration
PCR Target

Areas of interest (high variability among different species)
PCR Target

Areas of interest (high variability among different species)

Flanking regions that have negligible variability among species of interest
Polymerase Chain Reaction (PCR)

30 - 40 cycles of 3 steps:

**Step 1: denaturation**
- 1 minute 94 °C

**Step 2: annealing**
- 45 seconds 54 °C
- forward and reverse primers !!!

**Step 3: extension**
- 2 minutes 72 °C
- only dNTP's

(Andy Vierstraete 1995)
Amplification of Target DNA

Exponential amplification

1st cycle

2nd cycle

3rd cycle

4th cycle

template DNA

wanted gene

$2^2 = 4$ copies

$2^3 = 8$ copies

$2^4 = 16$ copies

$2^5 = 32$ copies

$2^{36} = 68$ billion copies

(Andy Vierstraete 1999)
Electrophoresis of Products
DNA Mixture Separation via Ligation and Transformation
Restriction Digest of Cloned Mixtures to Identify Different Elements
En mass Sequencing of Clones
Dye Terminator Cycle Sequencing Reaction
DNA Sequence Interpretation

Identification of the Closest Relative
– Evaluate DNA Sequence Variation
En mass Sequencing of Clones
Dye Terminator Cycle Sequencing Reaction
DNA Sequence Interpretation

Identification of the Closest Relative

– Evaluate DNA Sequence Variation
– Comparisons to Reference Data Sequences
  • Basic Local Alignment Search Tool
Example of a Partial *Lonicera rbcL* Alignment

L.orientalis chloroplast DNA rbcL gene Length = 1402

<table>
<thead>
<tr>
<th>Unknown</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unknown</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>61</td>
<td>65</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unknown</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>121</td>
<td>125</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unknown</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>181</td>
<td>185</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unknown</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>241</td>
<td>245</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unknown</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>301</td>
<td>305</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unknown</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>361</td>
<td>365</td>
</tr>
</tbody>
</table>
### BLAST Results for *Lonicera*

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Top 20 Sequence Alignments:</th>
<th>Score (bits)</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>emb</td>
<td>X87389.1</td>
<td>LORBCLGEN</td>
<td>L.orientalis chloroplast DNA rbcL gene</td>
</tr>
<tr>
<td>emb</td>
<td>Z29672.1</td>
<td>DSCLRUBI3</td>
<td>D.sessifolia chloroplast rbcL gene f...</td>
</tr>
<tr>
<td>emb</td>
<td>Y10704.1</td>
<td>PRY10704</td>
<td>P.rupestris chloroplast rbcL gene, pa...</td>
</tr>
<tr>
<td>emb</td>
<td>Y10706.1</td>
<td>MCY10706</td>
<td>M.coulteriana chloroplast rbcL gene, ...</td>
</tr>
<tr>
<td>emb</td>
<td>Y10700.1</td>
<td>TGY10700</td>
<td>T.glandulifera chloroplast rbcL gene,...</td>
</tr>
<tr>
<td>emb</td>
<td>Y10705.1</td>
<td>NJY10705</td>
<td>N.jatamansii chloroplast rbcL gene, p...</td>
</tr>
<tr>
<td>gb</td>
<td>AF156734.1</td>
<td>AF156734</td>
<td>Scabiosa sp. Albach 39 ribulose 1,5-...</td>
</tr>
<tr>
<td>gb</td>
<td>L11202.2</td>
<td>PUQCPRBCL</td>
<td>Pittosporum japonicum ribulose 1,5-bi...</td>
</tr>
<tr>
<td>emb</td>
<td>AJ235794.1</td>
<td>SPR235794</td>
<td>Stachyurus praecox chloroplast atp...</td>
</tr>
<tr>
<td>gb</td>
<td>L13864.1</td>
<td>DIZCHLORO</td>
<td>Dipsacus sativus chloroplast ribulose...</td>
</tr>
<tr>
<td>emb</td>
<td>X87394.1</td>
<td>QVRBCLGEN</td>
<td>Q.verdonii chloroplast DNA rbcL gene</td>
</tr>
<tr>
<td>emb</td>
<td>Z29670.1</td>
<td>DSCLRUBI1</td>
<td>D.spinosa chloroplast rbcL gene for ...</td>
</tr>
<tr>
<td>gb</td>
<td>U50261.1</td>
<td>PTU50261</td>
<td>Pittosporum tobira ribulose-1,5-bispho...</td>
</tr>
<tr>
<td>emb</td>
<td>Y10698.1</td>
<td>KIY10698</td>
<td>K.intermedia chloroplast rbcL gene, p...</td>
</tr>
<tr>
<td>emb</td>
<td>AJ403005.1</td>
<td>SLO403005</td>
<td>Sphenostemon lobosporus partial ch...</td>
</tr>
<tr>
<td>emb</td>
<td>X87398.1</td>
<td>VRRBCLGEE</td>
<td>V.rhytidophyllum chloroplast DNA rbc...</td>
</tr>
<tr>
<td>gb</td>
<td>L11682.1</td>
<td>SMHCPRBCL</td>
<td>Symphoricarpos albus chloroplast rubi...</td>
</tr>
<tr>
<td>gb</td>
<td>L11177.1</td>
<td>CPJCPRBCL</td>
<td>Carpenteria californica chloroplast r...</td>
</tr>
<tr>
<td>gb</td>
<td>U50254.1</td>
<td>MAU50254</td>
<td>Melanophylla alnifolia ribulose-1,5-bi...</td>
</tr>
<tr>
<td>dbj</td>
<td>D44582.1</td>
<td>PUQCPRC31</td>
<td>Pittosporum tobira chloroplast DNA f...</td>
</tr>
<tr>
<td>emb</td>
<td>Y10703.1</td>
<td>PBY10703</td>
<td>P.bracteata chloroplast rbcL gene, pa...</td>
</tr>
</tbody>
</table>
DNA Alignments

Ulmus_americana
GGCGTCACACGCCGTTG::CCCCCCCCAAACCCCGTCGGGGGCAAGC

Ulmus_crassifolia
GGCGTCACACACCGTTGCCCCCCCCAAAAACCCCGTCGGGGCAAGG

Ulmus_rubra
GGCGTCACACACCGTTG:CCCCCCCCAAAAACCCCGTCGGGGAAAGG

Ulmus_crassifolia
GGCGTCACACACCGTTG:CCCCCCCCAAAAACCCCGTCGGGGCAAAGG

SS1-6 Clothing 67
GGCGTCACACACCATTG:CCCCCCCCAAACCCCGTCGGGGGAAAGG
DNA Sequence Interpretation

Identification of the Closest Relative

– Evaluate DNA Sequence Variation
– Comparisons to Reference Data Sequences
  • Basic Local Alignment Search Tool
– Phylogenetic Analysis
Phylogenetic Methods

• Models of Sequence Evolution
  – Maximum Likelihood
  – Parsimony
  – Distance

• Bootstrapping
Maximum Likelihood
Juniperus ITS

[Tree Diagram with species and substitutions per site]

Stoney Forensic, Inc.

NIJ/FBI 2011 Trace Evidence Symposium 8-11-11
Bootstrap Analysis
Juniperus ITS
DNA Sequence Interpretation

Identification of the Closest Relative

– Evaluate DNA Sequence Variation
– Comparisons to Reference Data Sequences
  • Basic Local Alignment Search Tool
– Phylogenetetic Analysis
DNA Sequence Interpretation

Identification of the Closest Relative

- Evaluate DNA Sequence Variation
- Comparisons to Reference Data Sequences
  - Basic Local Alignment Search Tool
- Phylogenetic Analysis
- Specificity influenced by
  - Quality of DNA
  - Variability among close relatives
  - Available reference data
DNA Sequence Interpretation

Identification of the Closest Relative

- Evaluate DNA Sequence Variation
- Comparisons to Reference Data Sequences
  - Basic Local Alignment Search Tool
- Phylogenetic Analysis
- Specificity influenced by
  - Quality of DNA
  - Variability among close relatives
  - Available reference data
- Additional Reference Data as Needed
  - Type your own nearest relatives
Predictive Source Attribution

• Use the results of analyses on one sample for inference of origin from, or exposures to geographical areas, environments, activities and processes.
Predictive Source Attribution

• Use the results of analyses on one sample for inference of origin from, or exposures to geographical areas, environments, activities and processes.

Comparative Source Attribution

• Use the results of analyses on two samples to infer the strength of association based on their degree of correspondence, relative to other possible sources.
Examples (non-mixtures)

• Thorn from shoe
  – fungal sequence

• Tiny stain on letter
  – *Citrus reticulata* (mandarin orange)

• Red-brown stain on clothing
  – *Cyamopsis tetragonoloba* (guar, cluster bean)
    commercial emulsifier (e.g. ice cream)
  – *Theobroma cacao* (cocoa)
    source of cocoa, and chocolate
Example: Vacuumed Dust from a Set of Clothing

Species-Level Identifications
Acer saccharum (Aceraceae, sugar maple)
Allium sativum (Liliaceae, garlic)
Betula alba (Betulaceae, white or paper birch)
Cannabis sativa (Cannabaceae, marijuana, or hemp)
Coriandrum sativum (Apiaceae, coriander)
Ilex cornuta (Aquifoliaceae, Chinese holly)
Juglans regia (Juglandaceae, English walnut)
Medicago sativa (Fabaceae, alfalfa)
Sinapis alba (Brassicaceae, mustard)
Vicia faba (Fabaceae, broad bean)

Genus-Level Identifications
Arachis (Fabaceae, peanut)
Camellia (Theaceae) This genus includes tea.
Cicer (Fabaceae, chick-pea)
Cucurbita (Cucurbitaceae, pumpkins and gourds)
Cucumis (Cucurbitaceae, cucumber)
Epilobium (Onagraceae; a weedy genus)
Fortunella (Rutaceae, kumquat)
Juniperus (Cupresaceae, juniper)
Lactuca (Asteraceae, lettuce)
Lykopersicon (Solanaceae) Includes tomato.
Musa (Musaceae, banana)
Pinus (Pinaceae, Pine)
Pittosporum (Pittosporaceae, ornamental shrub)
Prunus (Rosaceae, Including peaches and cherry)
Rosa (Rosaceae, rose, a large genus)

Family Level Identifications
Anacardiaceae. Includes pistachio and mango.
Betulaceae This family includes birch.
Brassicaceae A large, weedy family
Fagaceae. This family contains the oaks, beeches, and chestnuts.
Juglandaceae This family includes many nut producing species.
Ulmaceae. This family includes elm.
Poaceae Grasses. 5 or 6 different types were detected.
DNA Results from Vacuumed Dusts

36 Taxa identified
  – 10 to species level
  – 15 to sub-genus or genus level
  – 11 to family or sub-family level
DNA Results from Vacuumed Dusts

36 Taxa identified
- 13 correspond to pollen taxa
- 15 new taxa
- 8 can be attributable to foods
**Taxa Reasonably Attributable to Foods**

*Brassica oleracea*: Cabbage, brussel sprouts, broccoli, kohlrabi, cauliflower, or kale

*Cucumis*: Cantaloupe, melons, gourds, cucumber

*Foeniculum or Anethum*: Fennel or Dil

*Juglans regia*: English walnut

*Medicago sativa*: Alfalfa sprouts

*Phaseolus vulgaris*: Bean, common bean, caraota, feijao, french bean, kidney bean, haricot bean, field bean, poroto, snap bean, string bean, frijol, wax bean

*Ginkgo biloba*: Maiden-hair; Dietary supplement, herbal tea

*Nicotiana tabacum*: Tobacco
Soil on Tusks

Photo credit: Les Craven, Medford Tribune
Taxa Identified

Plants

*Colophospermum mopane* (mopane tree)
*Flueggea* (bushweed)
*Bauhinia* (orchid tree)
*Triticum* (wheat genus)
*Zea* (corn genus)
FABACEAE close to genus *Podalyria* or *Calpurnia*

Fungi

*Aspergillus penicilloides*
*Eurotium amstelodami*
*Malassezia*
Taxonomic Viability Estimates
(*Diospyros mespiliformis*)

African ebony or jackal berry
Taxonomic Occurrence Estimates
*(Colophospermum mopane)*

Mopane

Stoney Forensic, Inc.

NIJ/FBI 2011 Trace Evidence Symposium 8-11-11
Summary Points & Observations

- Botanical DNA is present in trace samples and useful for taxonomic identification
Summary Points & Observations

• Botanical DNA is present in trace samples and useful for taxonomic identification

• Botanical DNA ≠ Pollen
  – Complementary signals, both extremely useful
Summary Points & Observations

• Botanical DNA is present in trace samples and useful for taxonomic identification

• Botanical DNA ≠ Pollen
  – Complementary signals, both extremely useful

• Dusts are (much) richer in extractable, useful botanical DNA than is soil
Summary Points & Observations

• Botanical DNA is present in trace samples and useful for taxonomic identification

• Botanical DNA ≠ Pollen
  – Complementary signals, both extremely useful

• Dusts are (much) richer in extractable, useful botanical DNA than is soil

• Technology is changing rapidly
  – “Next Generation” methods avoid cloning and utilizes intense computation
  – More amenable to comparative work
Taxonomic Identification of Traces using Non-human DNA

Thanks to:

Robert Bever & Bode Technology Group
Matthew Cimino
Taxonomic Identification of Traces using Non-human DNA

David A. Stoney  david@stoneyforensic.com
Stoney Forensic, Inc.
14101-G Willard Road, Chantilly, VA 20151