



Taxonomic Identification of Traces using Non-human DNA

David A. Stoney

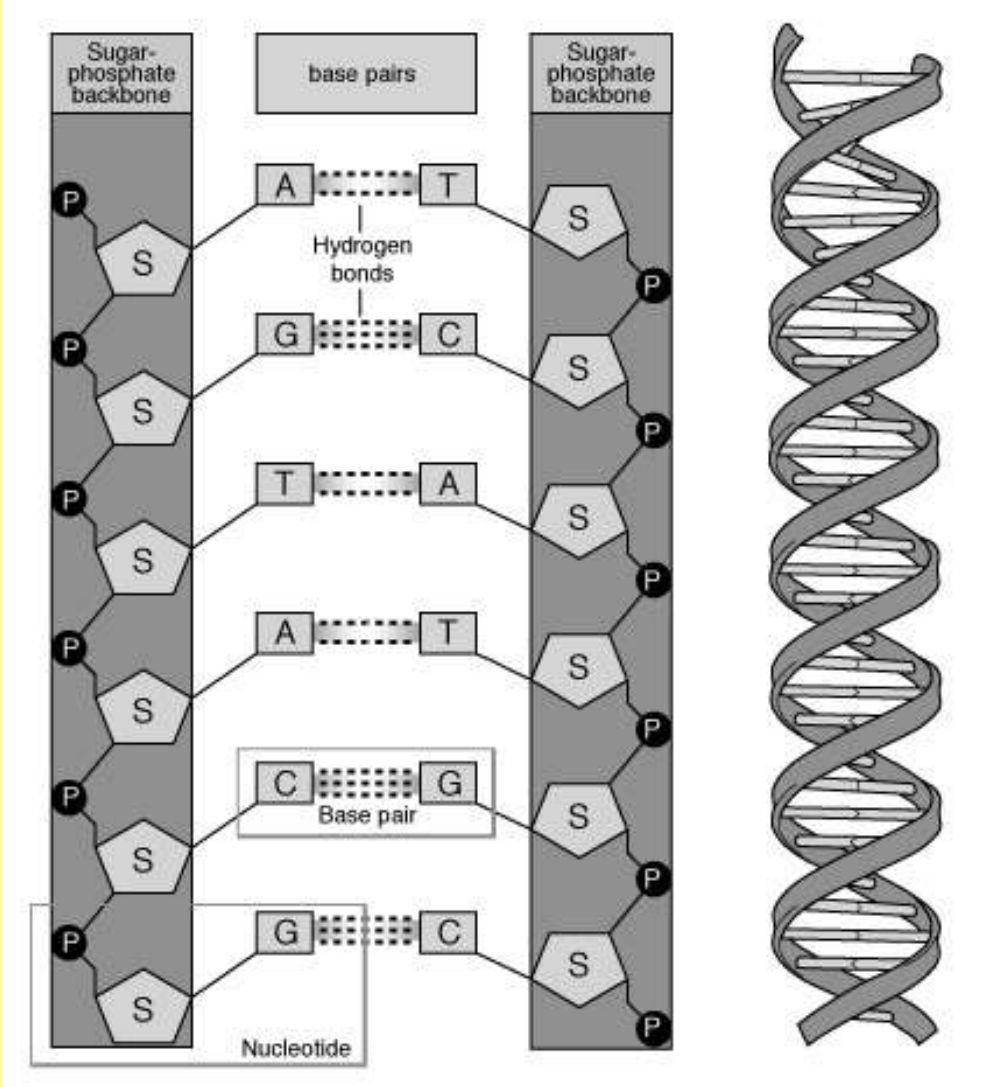
Stoney Forensic, Inc.

14101-G Willard Road, Chantilly, VA 20151

Taxonomic Identification of Traces using Non-human DNA

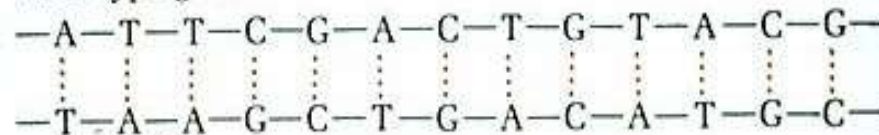
- Differences from human DNA identification applications
- Overview of the approach
- Examples of applications to predictive source attribution cases

DNA

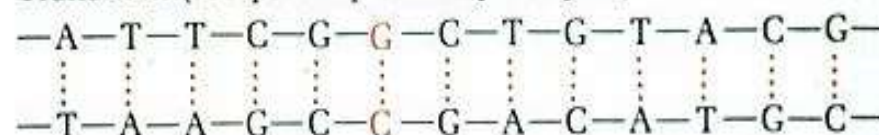


Mutations

Wild-type gene



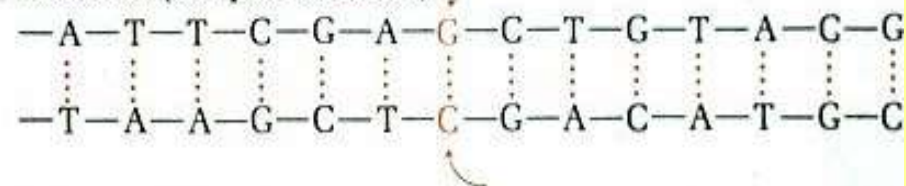
Transition (AT pair replaced by GC pair)



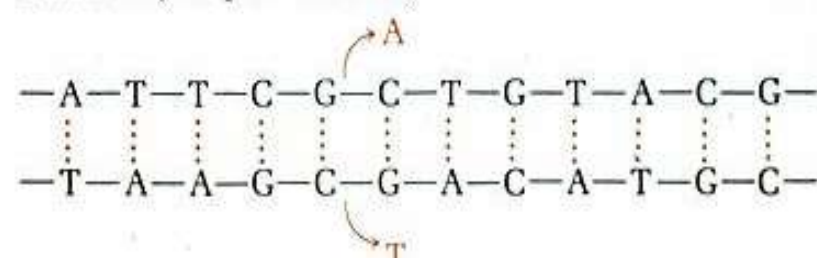
Transversion (AT replaced by TA)



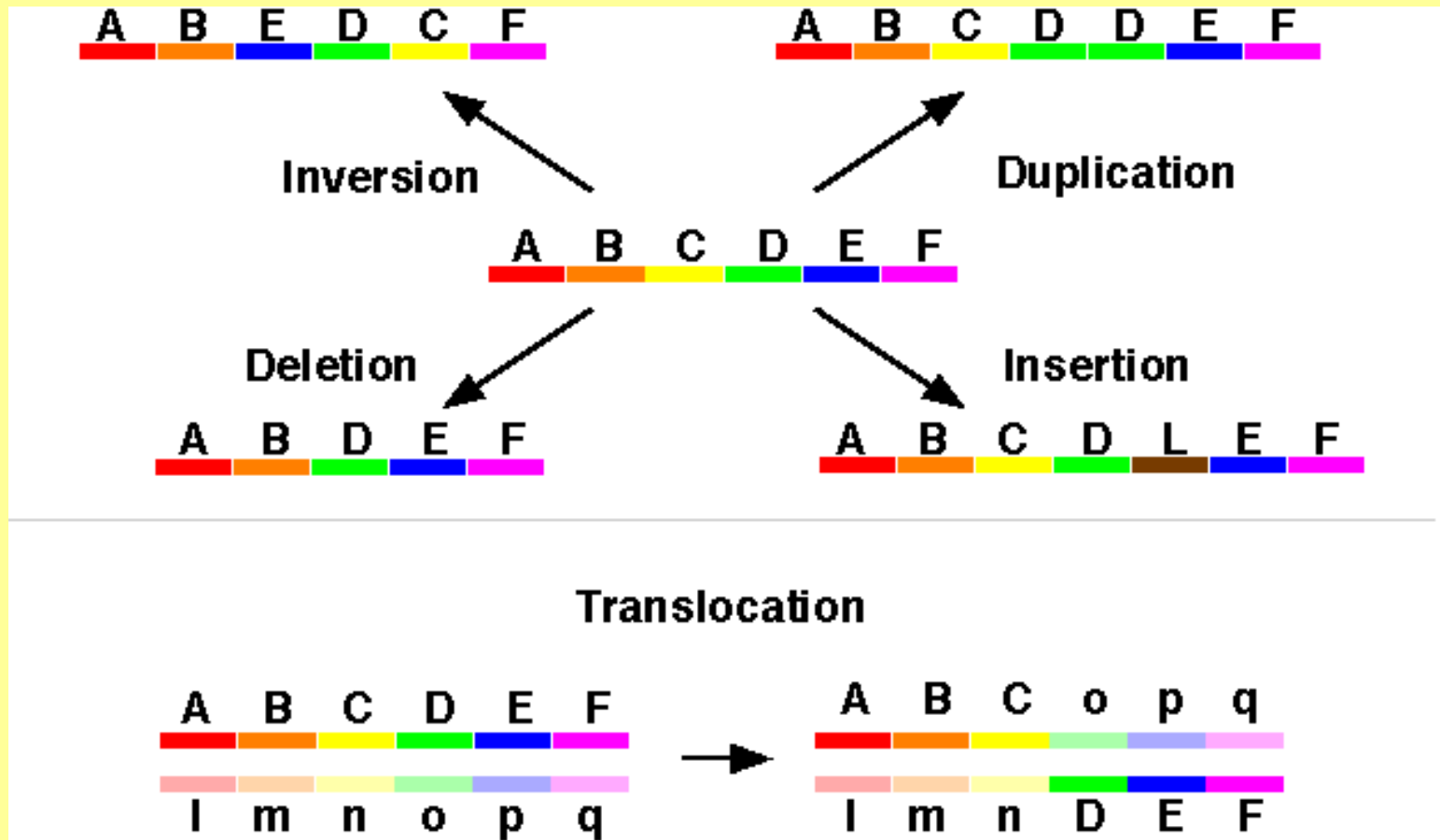
Insertion (GC pair inserted)



Deletion (AT pair deleted)



More Mutations



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

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

Locus	D3S1358	vWA	FGA	D8S1179	D21S11	D18S51	D5S818
Genotype	15, 18	16, 16	19, 24	12, 13	29, 31	12, 13	11, 13
Frequency	8.2%	4.4%	1.7%	9.9%	2.3%	4.3%	13%

Locus	D13S317	D7S820	D16S539	THO1	TPOX	CSF1PO	AMEL
Genotype	11, 11	10, 10	11, 11	9, 9.3	8, 8	11, 11	X Y
Frequency	1.2%	6.3%	9.5%	9.6%	3.52%	7.2%	(Male)

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

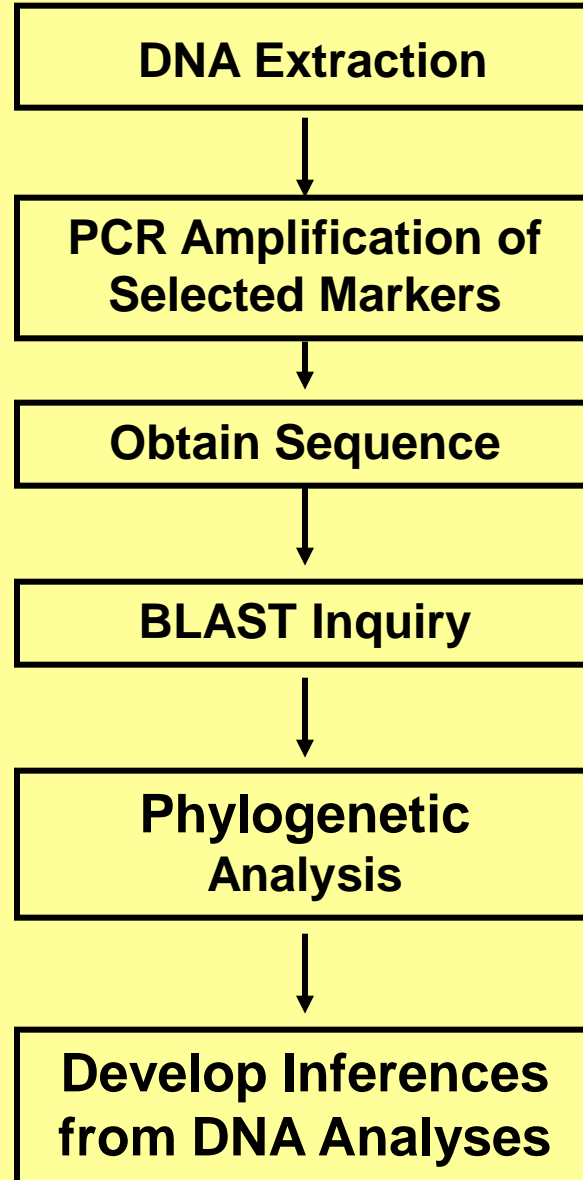
- Published and Unpublished Gene Frequency Data for Populations
- GenBank DNA Sequence Database
 - National Center for Biotechnology Information (NCBI)

Available Reference Data

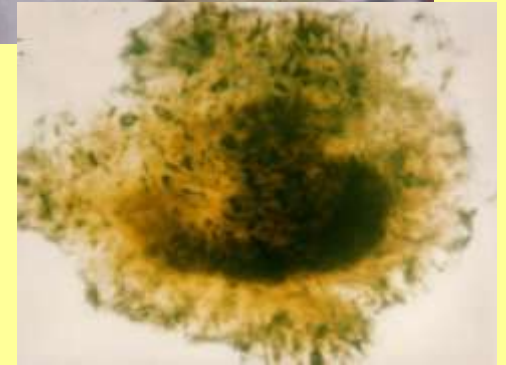
- Published and Unpublished Gene Frequency Data for Populations
- GenBank DNA Sequence Database
 - National Center for Biotechnology Information (NCBI)

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

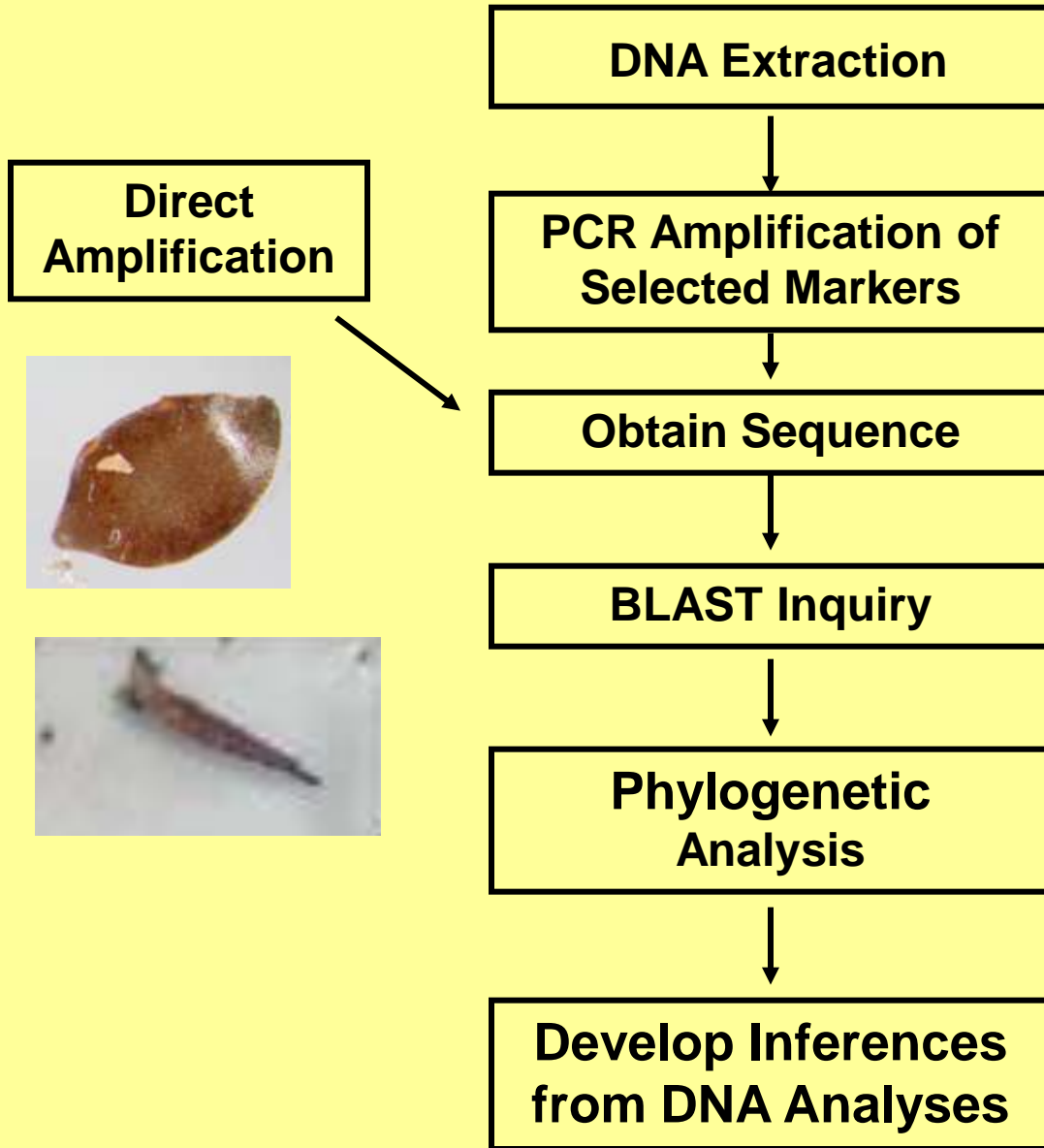
Sample Processing Overview



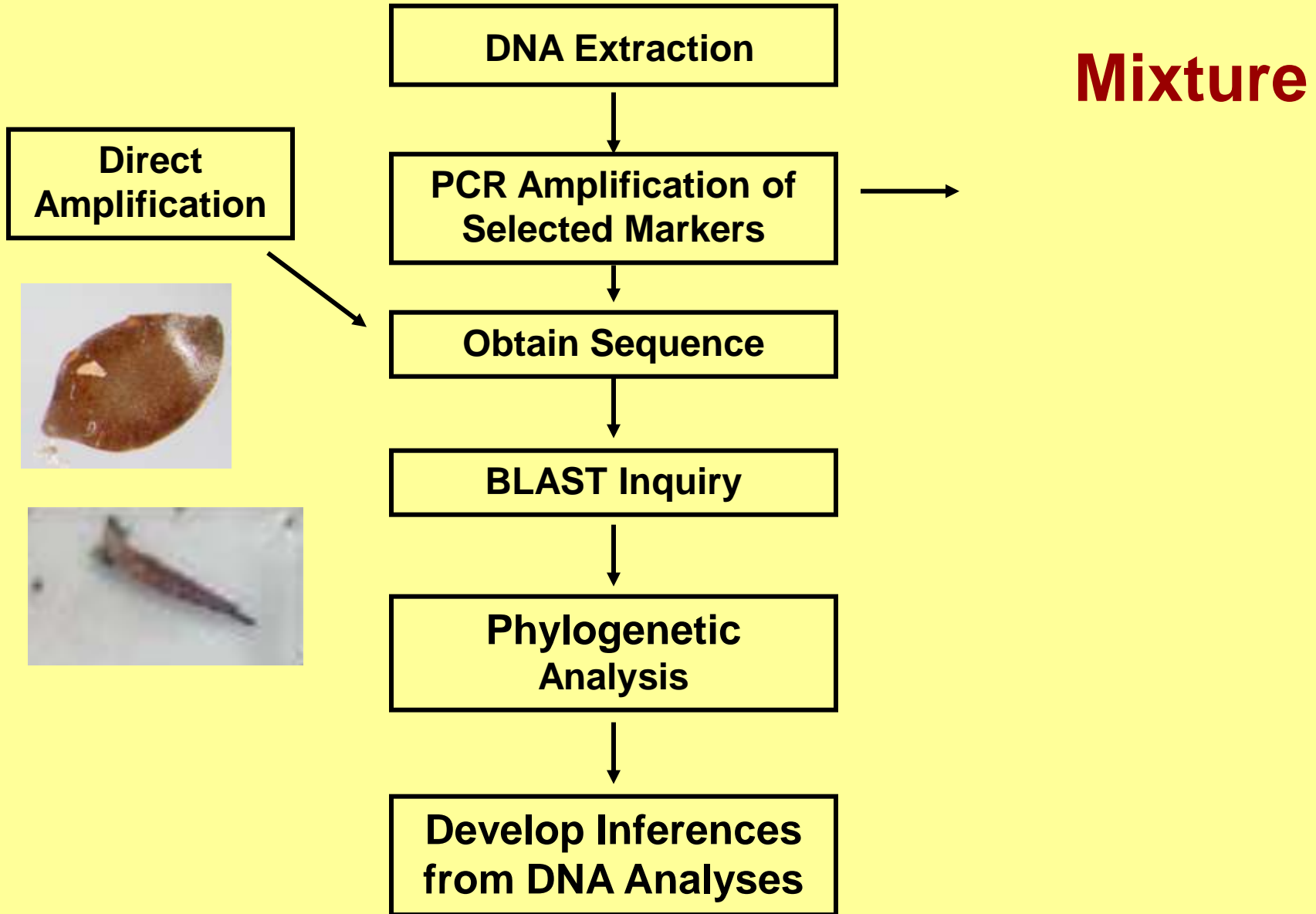
Single Elements vs. Mixtures



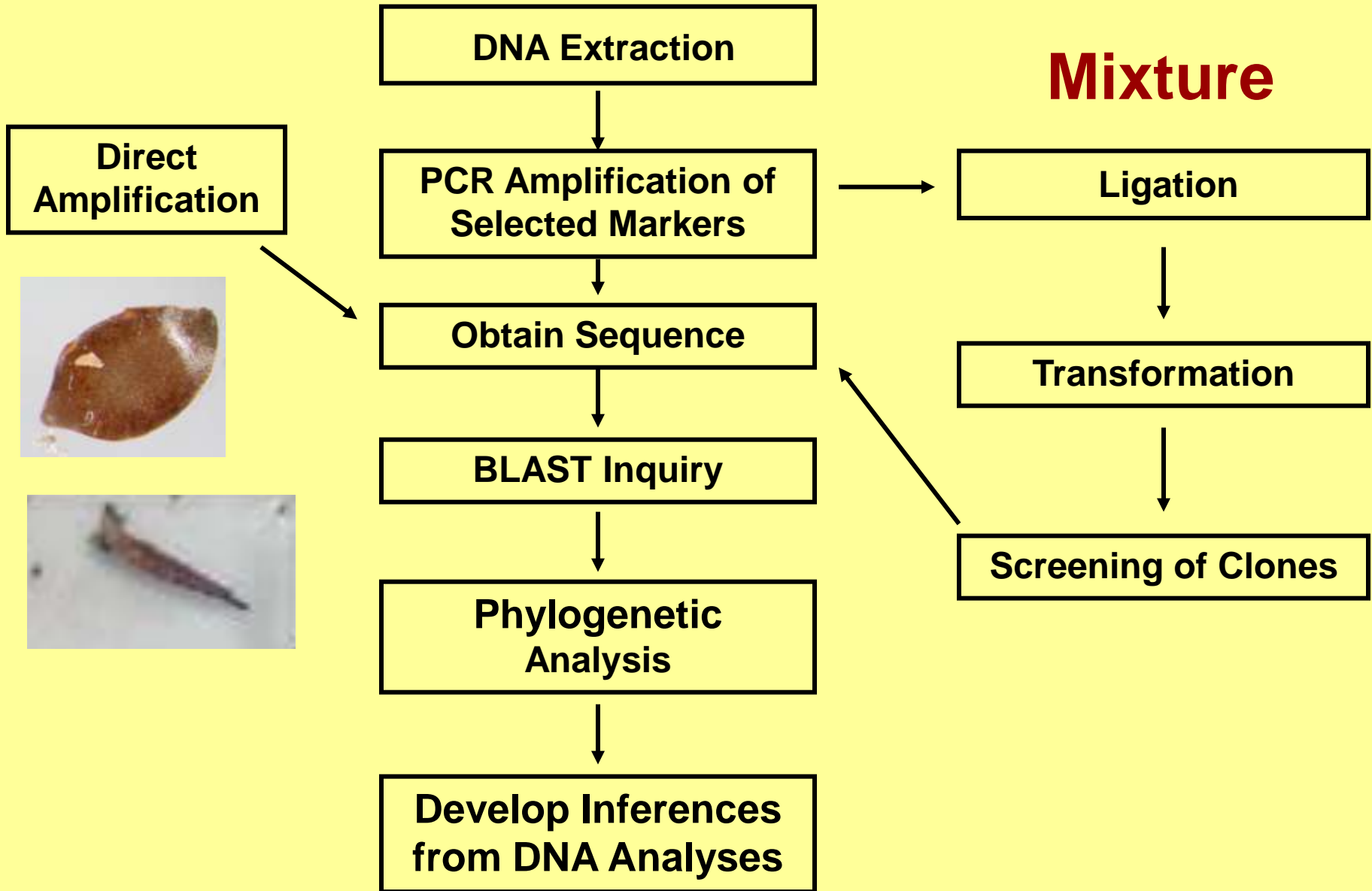
Sample Processing Overview



Sample Processing Overview

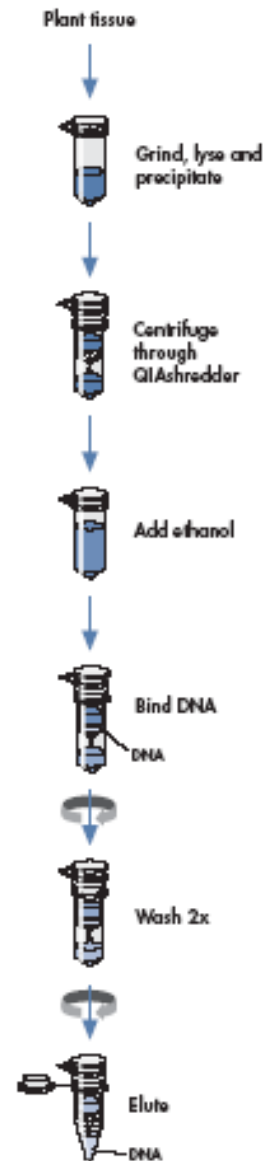


Sample Processing Overview



DNA Extraction

The DNeasy Plant Procedure

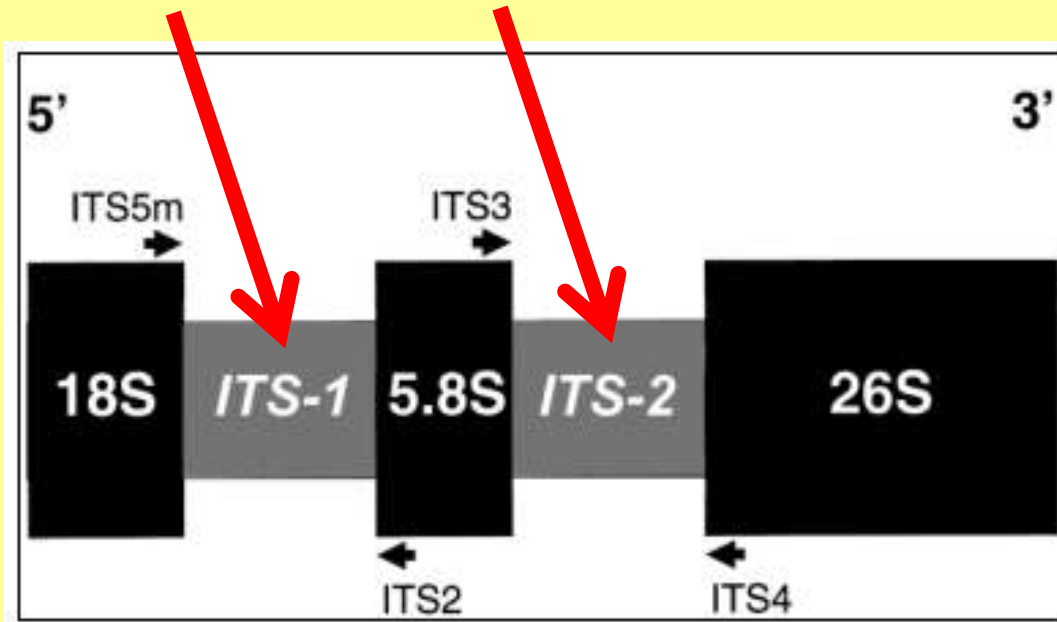


DNA Quantification

- 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

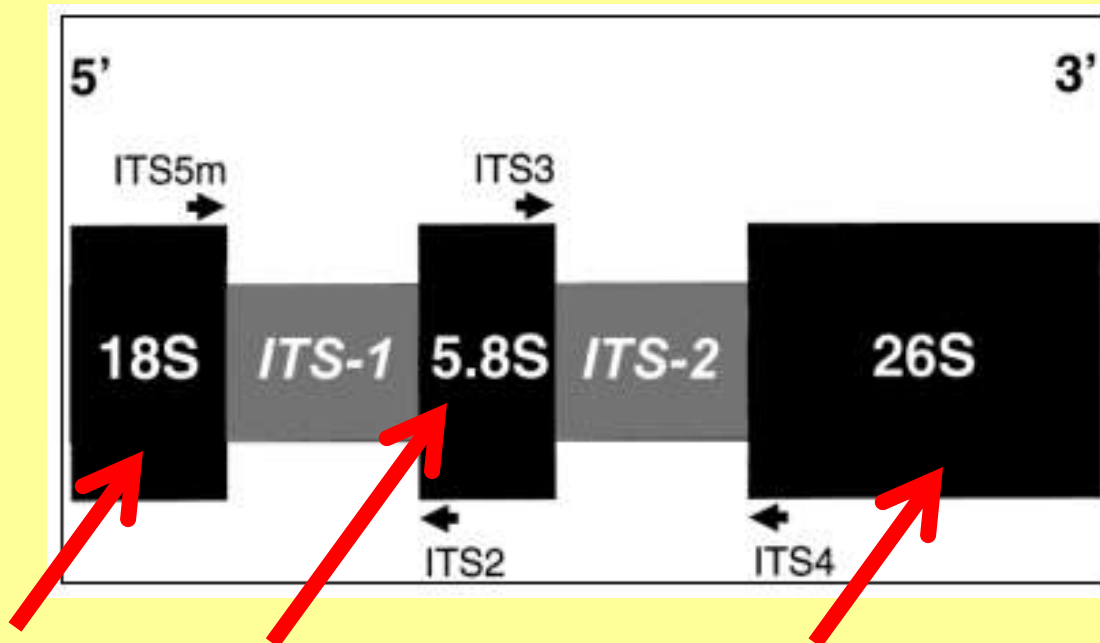
PCR Target

Areas of interest (high variability among different species)



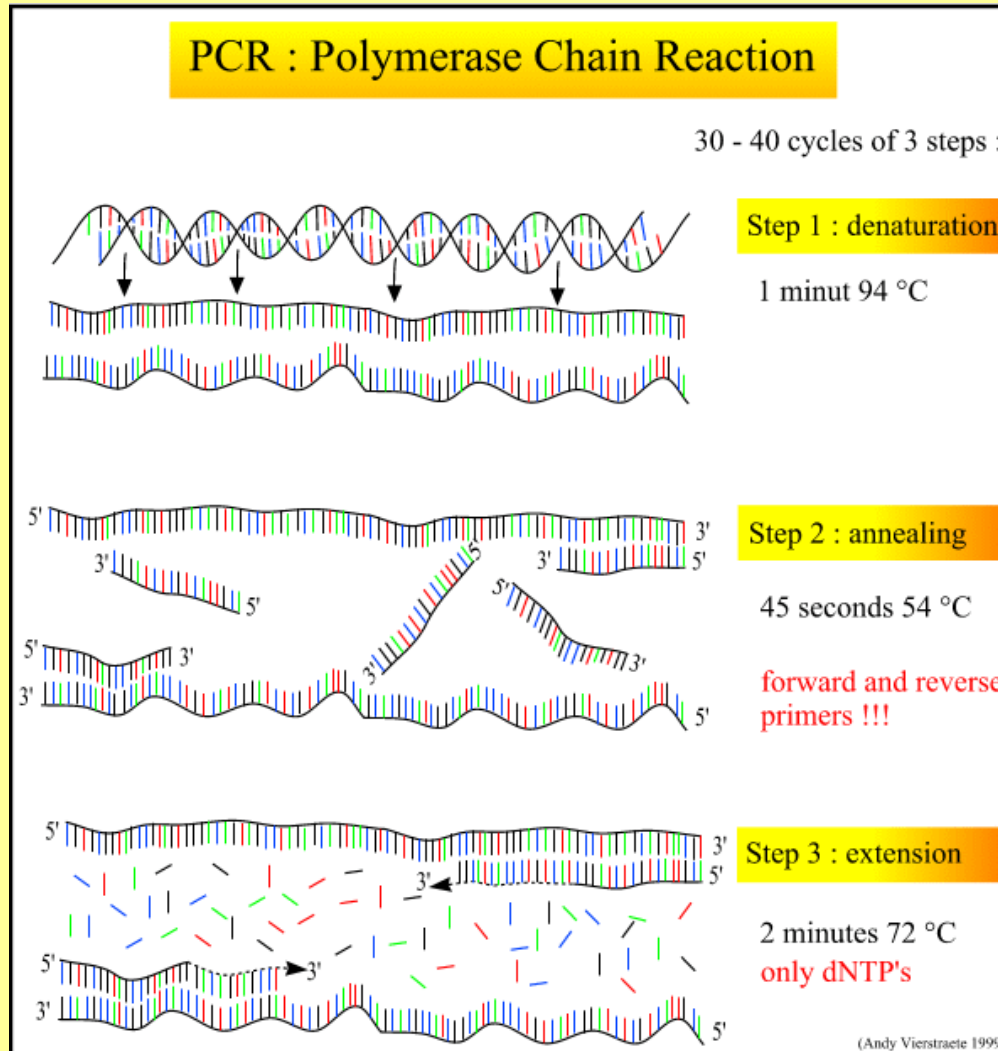
PCR Target

Areas of interest (high variability among different species)

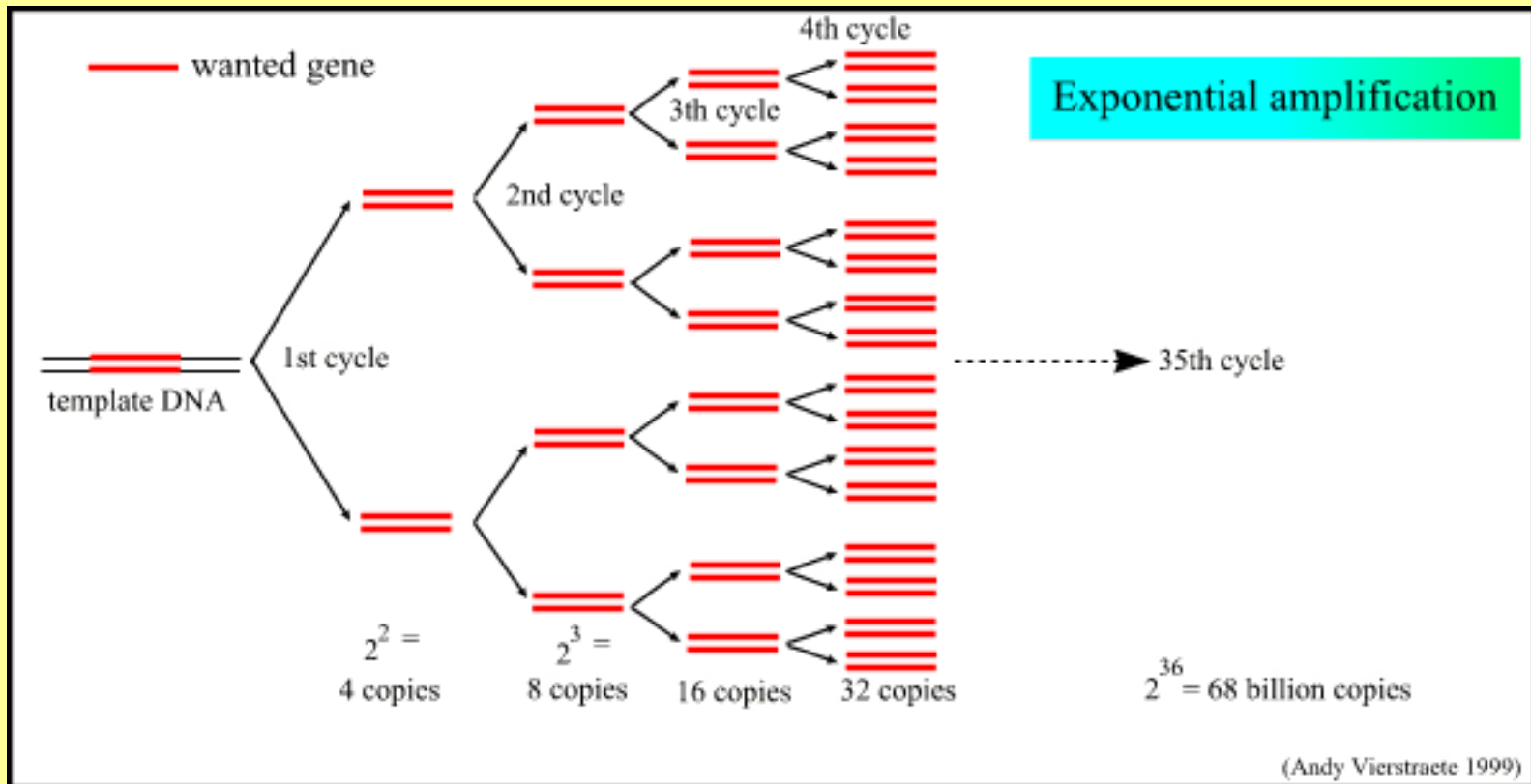


Flanking regions that have negligible variability among species of interest

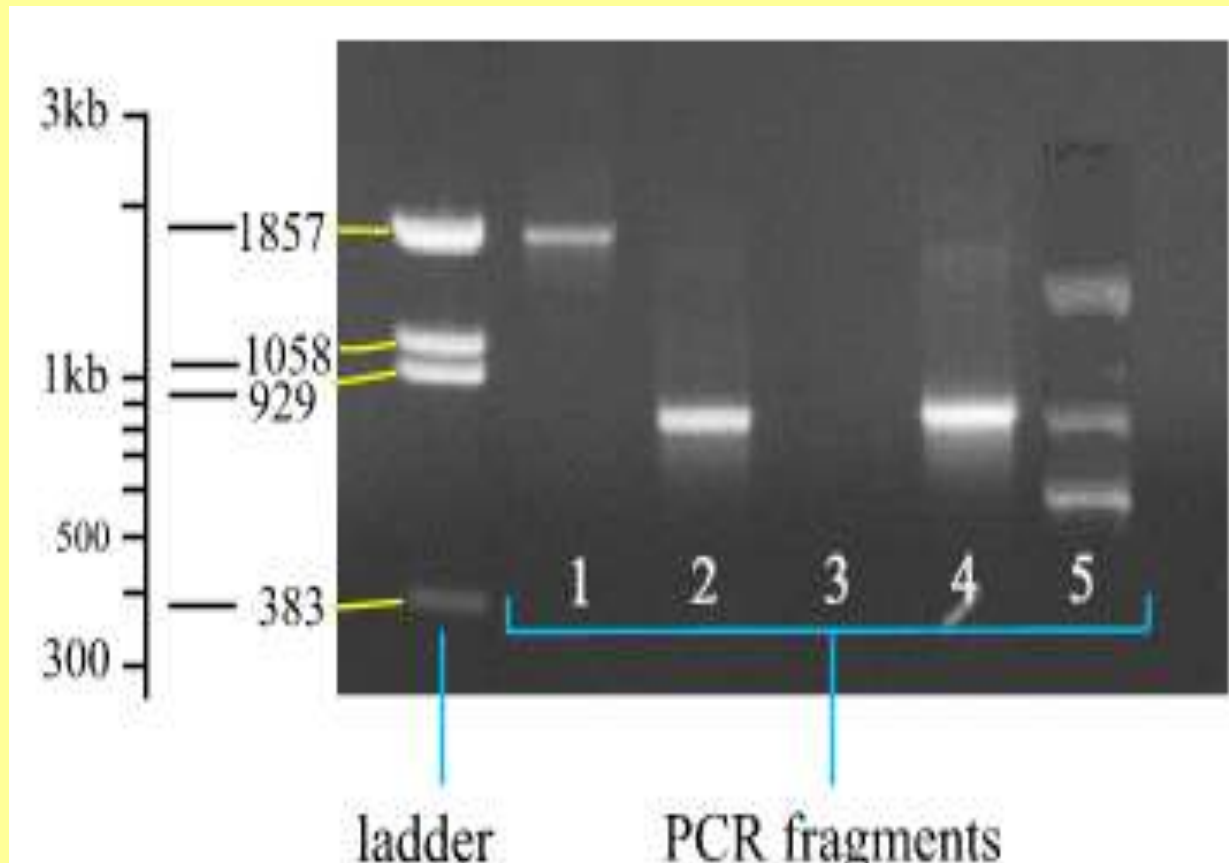
Polymerase Chain Reaction (PCR)



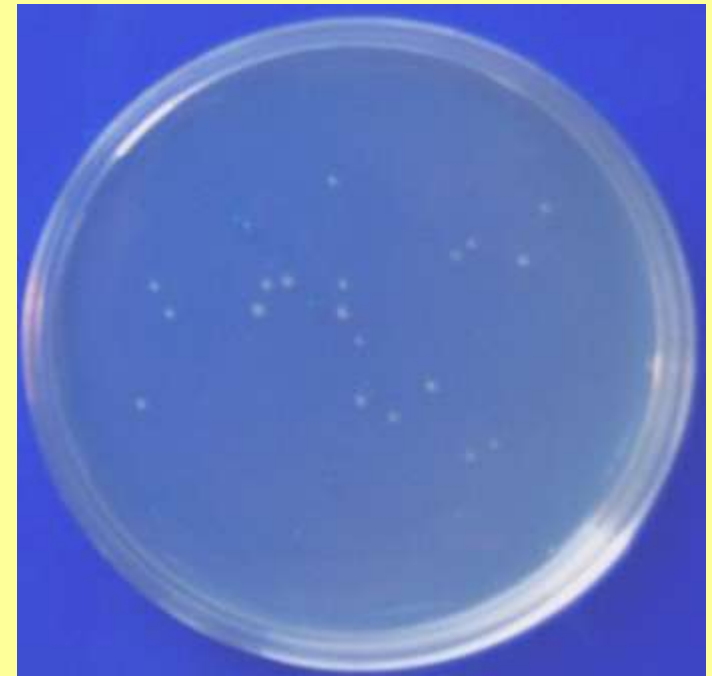
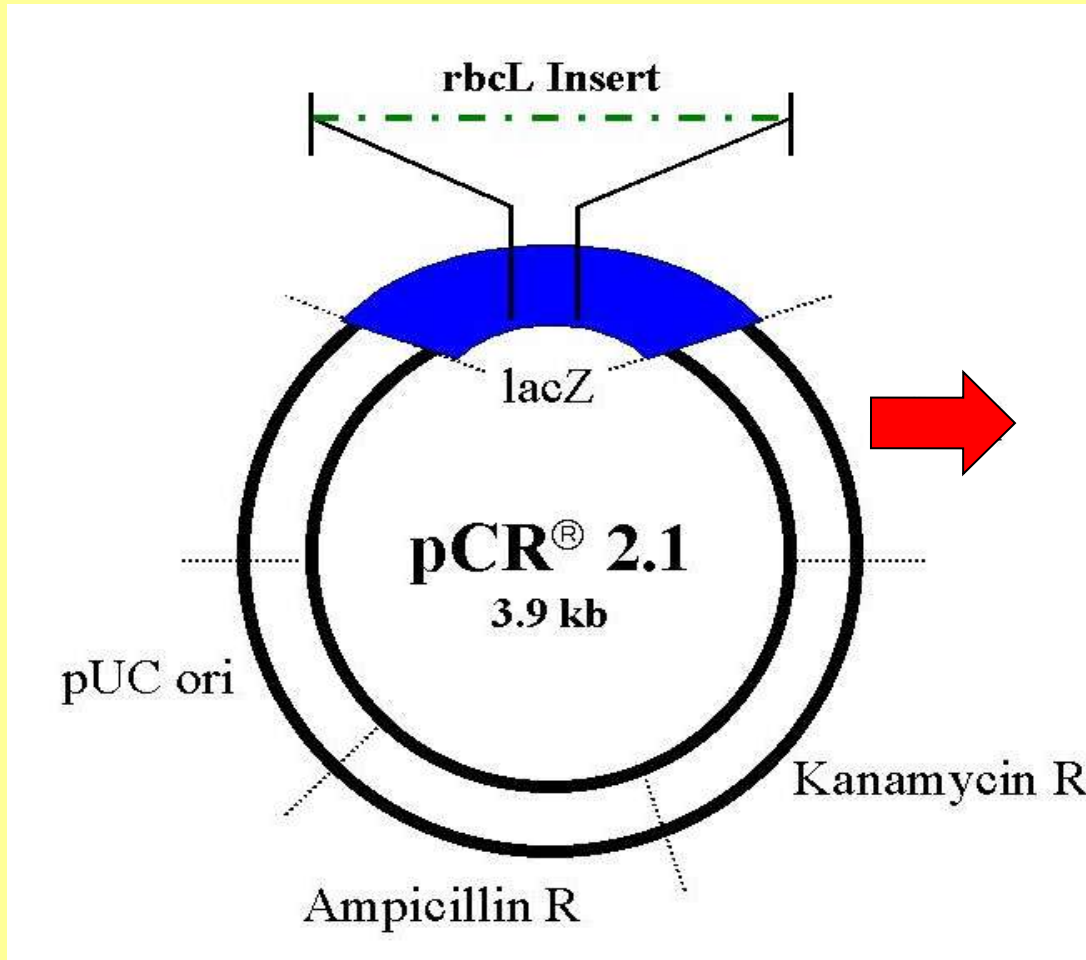
Amplification of Target DNA



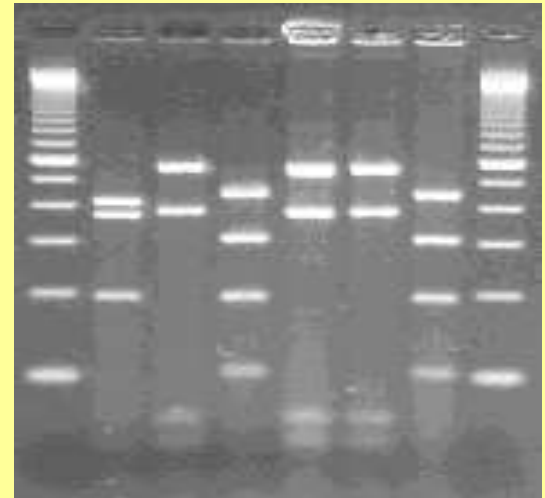
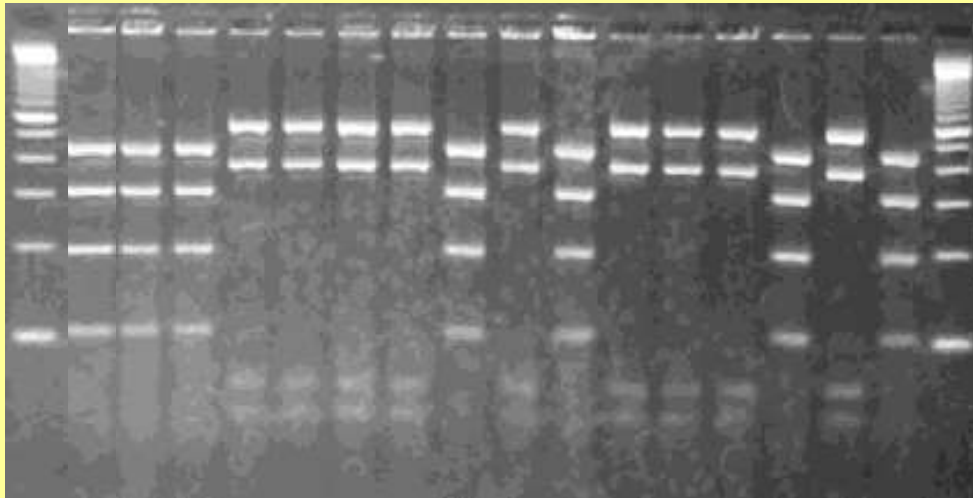
Electrophoresis of Products



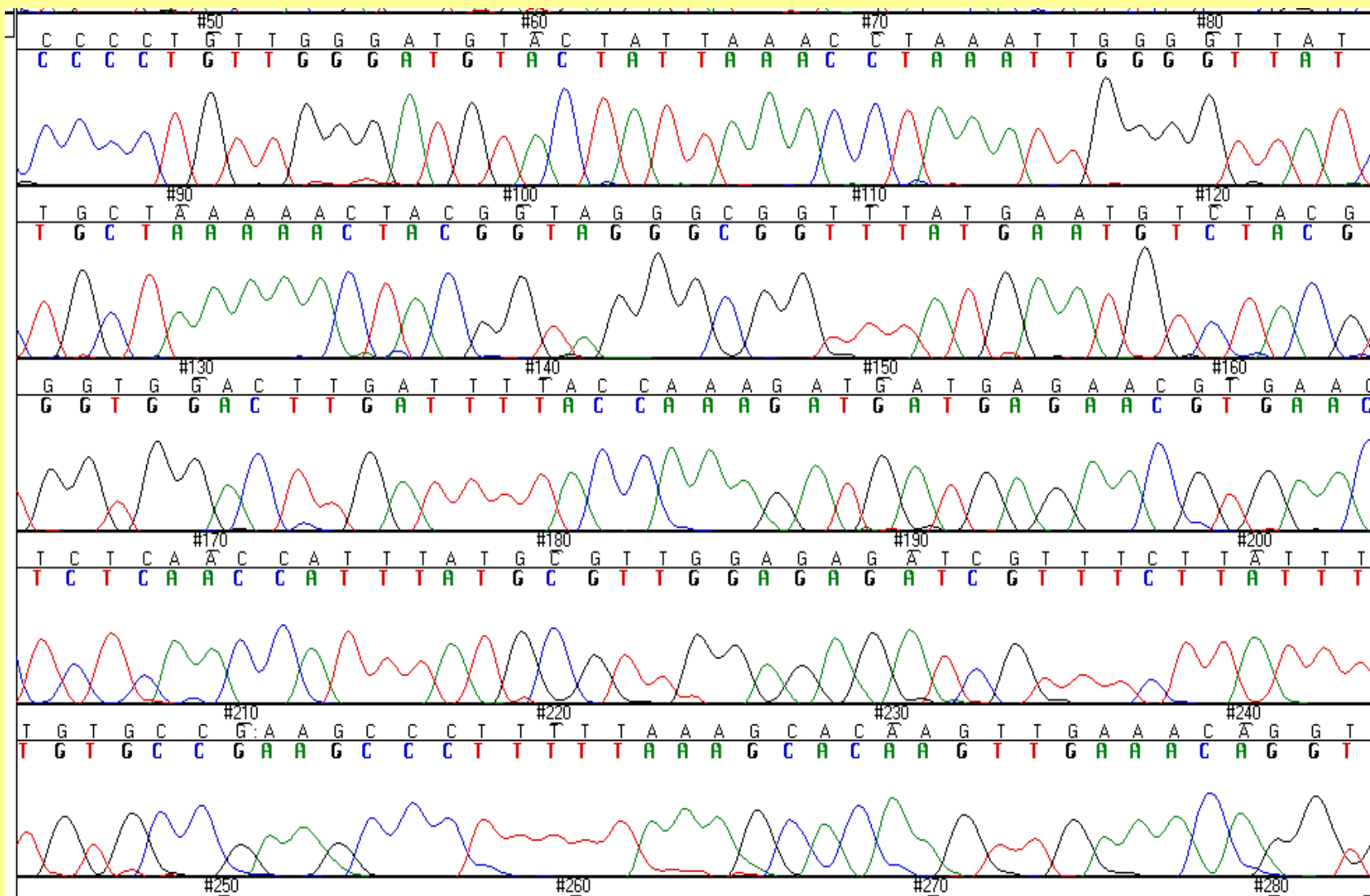
DNA Mixture Separation via Ligation and Transformation



Restriction Digest of Cloned Mixtures to Identify Different Elements



En mass Sequencing of Clones Dye Terminator Cycle Sequencing Reaction



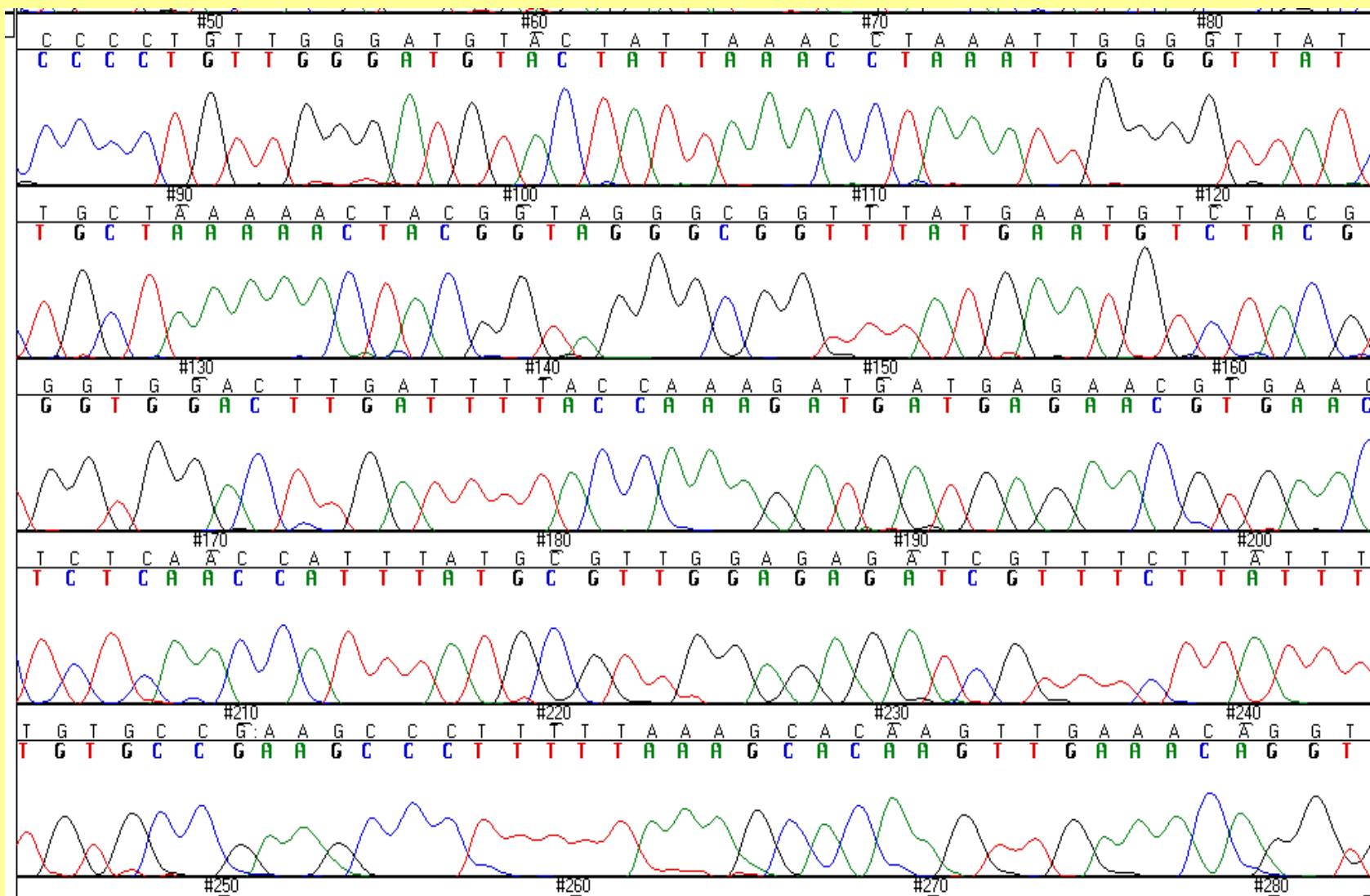
DNA Sequence Interpretation

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

>[emb|X87389.1|LORBCLGEN](#)

L.orientalis chloroplast DNA *rbcL* gene

Length = 1402

```
Unknown: 1  gttggattcaaagcgggtgttaaagattacaaattgacttattatactcctgacatgaa 60
          |
Match:    5  gttggattcaaagcgggtgttaaagattacaaattgacttattatactcctgaatgaa 64

Unknown: 61  accaaagatactgatatcttggcagcattccgagtaactcctcaaccggagttccgcct 120
          |
Match:    65  accaaagatactgatatcttggcagcattccgagtaactcctcaaccggagttccgcct 124

Unknown: 121  gaagaagcgggggcccgcggtagctgctgaatcttcaaccggtacatggacaactgtgtgg 180
          |
Match:    125  gaagaagcgggggcccgcggtagctgctgaatcttcaaccggtacatggacaactgtgtgg 184

Unknown: 181  accgatggacttaccagccttgatcgttacaaagggcgatgctaccacatcgagcccgtt 240
          |
Match:    185  accgatggacttaccagccttgatcgttacaaagggcgatgctacggcatcgagcccgtt 244

Unknown: 241  gctggagaagaaaatcaatttattgcttatgtagcttaccattagacctttttgaagaa 300
          |
Match:    245  gctggagaagaaaatcaataatattgcttatgtagcttaccattagacctttttgaagaa 304

Unknown: 301  ggttctgttactaacaatgtttacttctattgagggtaatgtatttgggttcaaagccctg 360
          |
Match:    305  ggttctgttactaacaatgtttacttctattgtgggtaatgtatttgggttcaaagccctg 364

Unknown: 361  cgcgctctacgtctggaagatctgcgaatccctgtctcttattggttaaactttccaaggc 420
          |
Match:    365  cgcgctctacgtctggaagatctgcgaatccctgctgcttattggttaaactttccaaggc 424
```

BLAST Results for *Lonicera*

<u>Accession Number</u>	<u>Top 20 Sequence Alignments:</u>	<u>Score</u>	<u>E Value</u>
emb X87389.1 LORBCLGEN	L.orientalis chloroplast DNA rbcL gene	1631	0.0
emb Z29672.1 DSCLRUBI3	D.sessifolia chloroplast rbcL gene f...	1588	0.0
emb Y10704.1 PRY10704	P.rupestris chloroplast rbcL gene, pa...	1564	0.0
emb Y10706.1 MCY10706	M.coulteriana chloroplast rbcL gene, ...	1556	0.0
emb Y10700.1 TGY10700	T.glandulifera chloroplast rbcL gene,...	1503	0.0
emb Y10705.1 NJY10705	N.jatamansii chloroplast rbcL gene, p...	1501	0.0
gb AF156734.1 AF156734	Scabiosa sp. Albach 39 ribulose 1,5-...	1493	0.0
gb L11202.2 PUQCPRBCL	Pittosporum japonicum ribulose 1,5-bi...	1489	0.0
emb AJ235794.1 SPR235794	Stachyurus praecox chloroplast atp...	1489	0.0
gb L13864.1 DIZCHLORO	Dipsacus sativus chloroplast ribulose...	1487	0.0
emb X87394.1 QVRBCLGEN	Q.verdonii chloroplast DNA rbcL gene	1483	0.0
emb Z29670.1 DSCLRUBI1	D.spinosa chloroplast rbcL gene for ...	1483	0.0
gb U50261.1 PTU50261	Pittosporum tobira ribulose-1,5-bispho...	1477	0.0
emb Y10698.1 KIY10698	K.intermedia chloroplast rbcL gene, p...	1477	0.0
emb AJ403005.1 SLO403005	Sphenostemon lobosporus partial ch...	1473	0.0
emb X87398.1 VRRBCLGEE	V.rhytidophyllum chloroplast DNA rbc...	1469	0.0
gb L11682.1 SMHCPRBCL	Symphoricarpos albus chloroplast rubi...	1467	0.0
gb L11177.1 CPJCPRBCL	Carpenteria californica chloroplast r...	1465	0.0
gb U50254.1 MAU50254	Melanophylla alnifolia ribulose-1,5-bi...	1465	0.0
dbj D44582.1 PUQCPRC31	Pittosporum tobira chloroplast DNA f...	1463	0.0
emb Y10703.1 PBY10703	P.bracteata chloroplast rbcL gene, pa...	1461	0.0

DNA Alignments

Ulmus_americana

GGCGTCACACGCCGTTG::CCCCCCCAACCCCGTCGGGGGCAAGCGGG

Ulmus_crassifolia

GGCGTCACACACCGTTGCCCCCCCAACCCCGTCGGGGGCAGAAGGG

Ulmus_rubra

GGCGTCACACACCGTTG:CCCCCCCAACCCCGTCGGGGGAAGAAGGG

Ulmus_crassifolia

GGCGTCACACACCGTTG:CCCCCCCAACCCCGTCGGGGCAAAGGG

SS1-6 Clothing 67

GGCGTCACACACCATTG:CCCCCCCAACCCCGTCGGGGGAAGAAGGG

DNA Sequence Interpretation

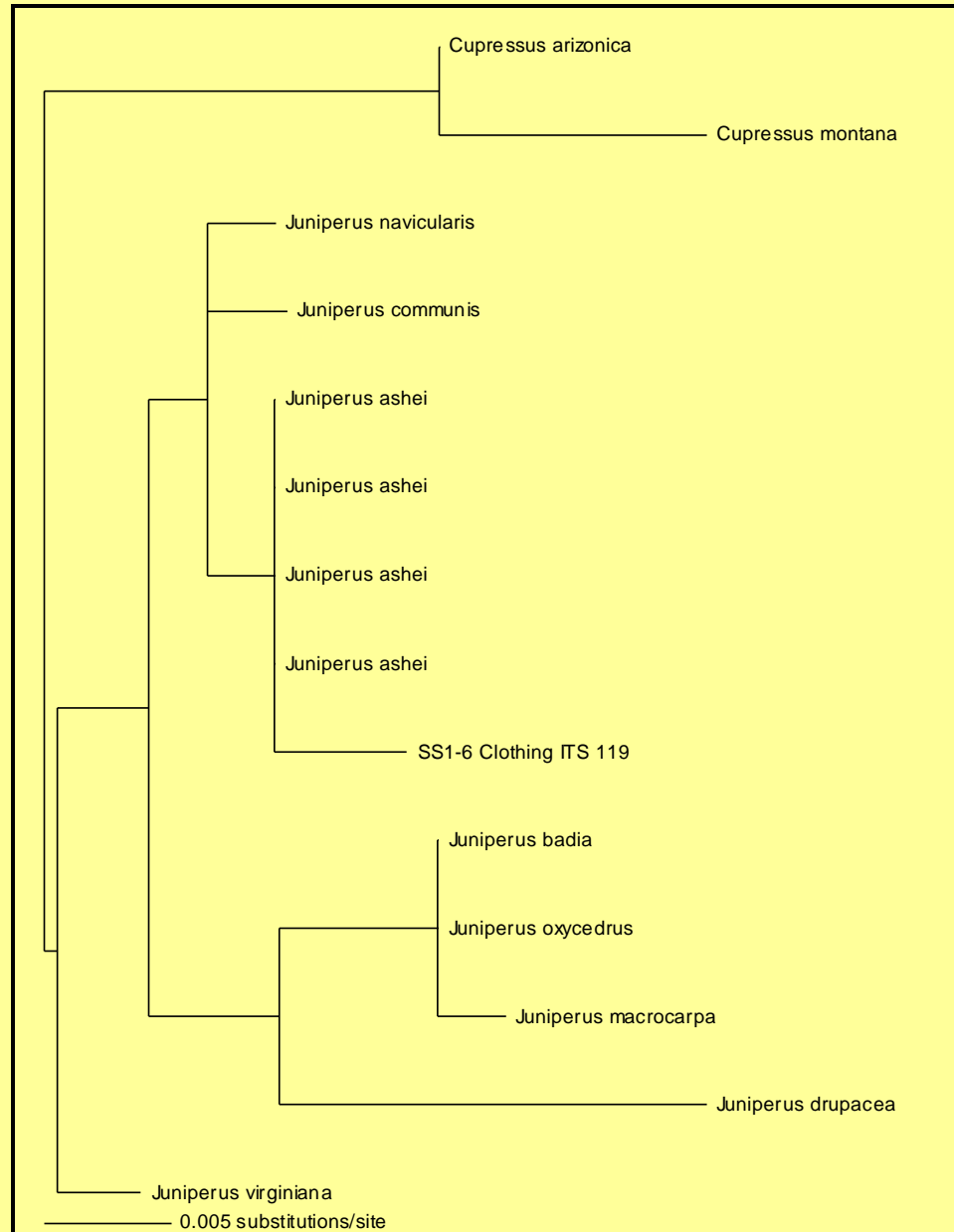
Identification of the Closest Relative

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

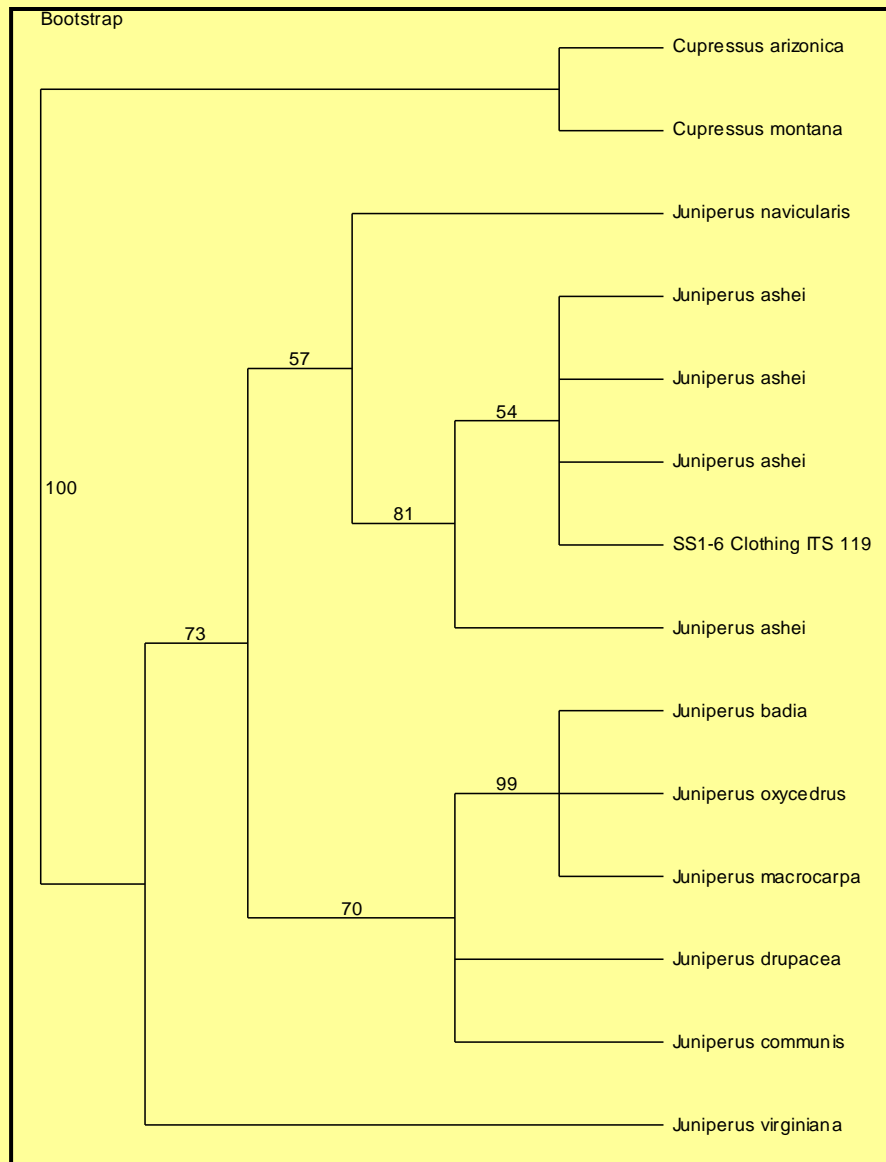
Phylogenetic Methods

- Models of Sequence Evolution
 - Maximum Likelihood
 - Parsimony
 - Distance
- Bootstrapping

Maximum Likelihood Juniperus ITS



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
- Phylogenetic 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

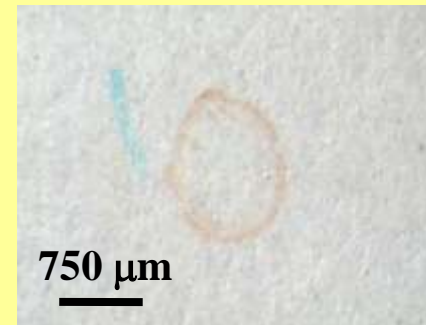
- Use the results of analyses on two samples to infer the strength of association based on their degree of correspondence, relative to other possible sources.

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)

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.

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 (Cuppressaceae, juniper)
Lactuca (Asteraceae, lettuce)
Lycopersicon (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)

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

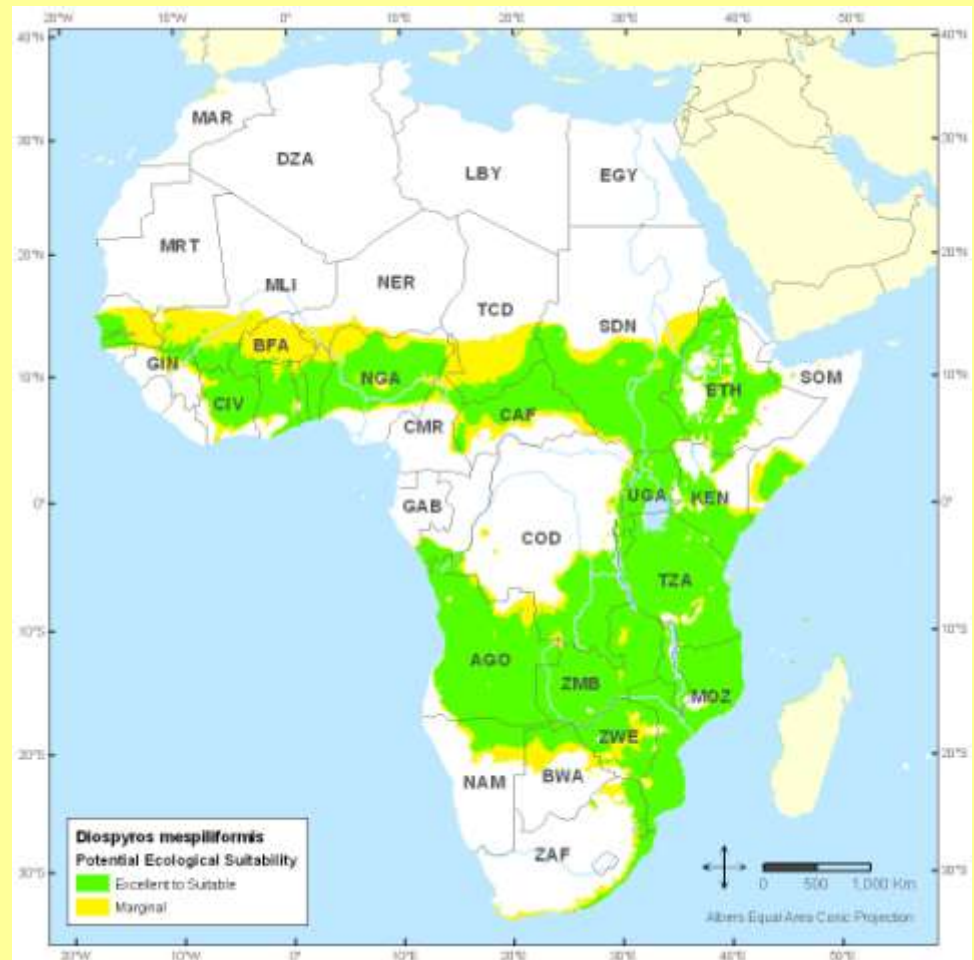
Eurotium amstelodami

Malassezia

Taxonomic Viability Estimates (*Diospyros mespiliformis*)



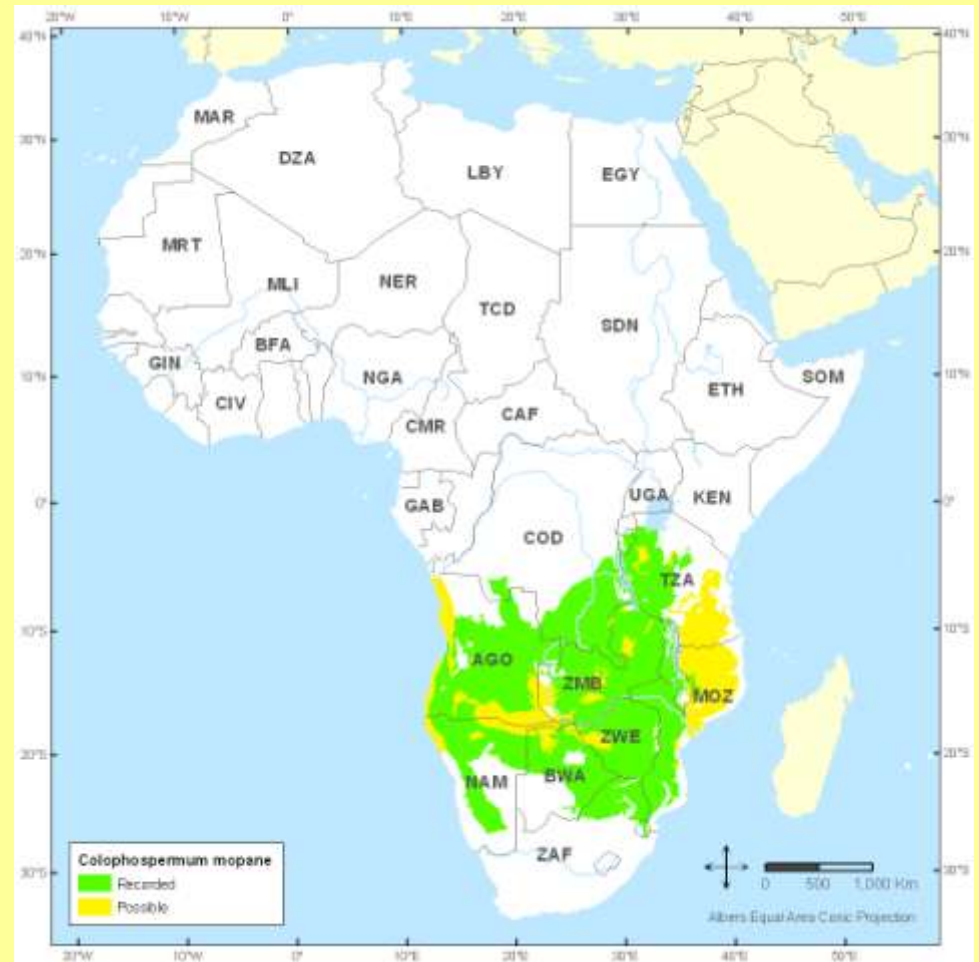
African ebony or jackal berry



Taxonomic Occurrence Estimates (*Colophospermum mopane*)



Mopane



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 \neq Pollen
 - Complementary signals, both extremely useful

Summary Points & Observations

- Botanical DNA is present in trace samples and useful for taxonomic identification
- Botanical DNA \neq 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 \neq 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