I. Announcements and Assignments
   a. Dr. John DeHaan- Fire Debris arson expert 11/29
   b. Assignments- Reading and Articles

II. Mitochondrial DNA
   a. Biology of mitochondria
   b. DNA sequencing

I. Y Chromosome markers: Intro to Y chromosomes- Types of Y polymorphism

II. Single nucleotide polymorphisms (SNPs)
   a. Why SNPs? Intro to Single Nucleotide Polymorphisms (SNPS)
   b. Applications of SNPs
   c. Detection Technologies for Y SNPs in Forensics: Primer Extension, Pyrosequencing, Light Cycling, Mass Spec
   d. Bead based assays-Luminex
   e. Universal Arrays and Bacterial Identification
   f. SNPs vs STRs or SNPs and STRs
Why mtDNA SNPs?

- Well characterized and studied (population, evolutionary, medical and forensic studies)
- Uniparental maternal inheritance missing persons
- Relatively small size (16kb) and high copy number - low quantity/quality samples (hair, bone, teeth-ancient/degraded)
- Implicated in maternally inherited diseases: diabetes, deafness, hypertrophic cardiomyopathy and myopathy
Assignments and Announcements

• Announcements-
  – Criminalist Isha Brown Weds 9th May - CA DOJ DNA Databank
  – Assignments-
  – Butler Chapters 8-11 Inman, 16, Appendices IV&V, Inman 10-11
  – Read Article Butler Y chromosome review article posted to the web - Write a 500 word summary with 3Q and 3A – FOR 5 points extra credit
  – Hand in assignment by weds 9th May
Mitochondrial DNA regions used in forensics

• Hypervariable regions- also known as D-loop or control regions involved in the replication of mtDNA

• MtDNA is in very high copy number in every cell. There are many cells per sample and therefore many more copies than nuclear DNA that has only 1 per cell

• Most forensic laboratories utilize DNA sequencing to analyze mitochondrial DNA polymorphisms
• Intro to Y chromosomes- Types of Y polymorphism

• Intro to Single Nucleotide Polymorphisms (SNPs)
  • Definitions
  • Why SNPs?
  • Applications of SNPs

• Detection Technologies for Y SNPs in Forensics
  • Primer Extension, Pyrosequencing, Light Cycling, Mass Spec
  • Bead based assays-Luminex
    • Universal Arrays and Bacterial Identification

• SNPs vs STRs or SNPs and STRs
  • Either/Or
  • Why SNPs?
Cycle sequencing:
PCR in the presence of “bad” dNTPs – dideoxynucleoside triphosphates

• Synthesize DNA in the presence of some dideoxy nucleotides without a 3-OH
• Building a railroad with some tracks that do not have connectors
• End result is a complete set of fragments that represent every base in the DNA strand
• See animation
Overview of the Y Chromosome

- Paternally inherited
- Represents 2% of the human genome
- ~60 Mb in length, 2.5 Mb on tips recombine with the X
- 95% of the Y is non-recombining
- Y SNP Consortium - Over 4193 SNPs on the Y chromosome

http://ycc.biosci.arizona.edu/

Nucleic Acids Res. 28(2), e8 (2000)
Why study the Y chromosome?

- Population Genetics\(^1\)
- Evolutionary and Genealogical studies\(^2\)
- Molecular Ecology\(^3\)
- Infertility studies\(^4\)
- Forensics\(^5\)


3 http://www.oxfordancestors.com/


YY in forensics?

- **Bad Boys:** 98% of violent crime is committed by men
- **Sexual Assault Evidence Screening:** Rapid screening of sexual assault evidence: “male specific” - so no differential
- **Mixtures:** Especially with very low copy male DNA in mixtures. May assist in determining single or multiple donors in difficult mixtures
- **Missing persons/Paternity:** Paternal lineage reference samples
Polymorphisms on the Y

– Binary (biallelic) Markers
  • SNPs (single nucleotide polymorphisms)
  • YAP (Y Alu polymorphism)

– Microsatellites – STR’s
  • Tetranucleotide repeats such as DYS19, DYS385, DYS388, DYS390, DYS391, etc.

– Minisatellites - MSY1
Definitions

What is a SNP?

Single Nucleotide Polymorphisms

**Point mutation**

- GAATCCTCCATCT
- GAATCCACCATCT

**Deletion**

- GAATCCTCCATCT
- GAATCC−CCATCT

**Insertion**

- GAATCCT−CCATCT
- GAATCCTCCCATCT

**Most study Bi-allelic SNPs**

- GAATCCTCCATCT
- GAATCCACCATCT
• Extremely Well Studied- Used in virtually every molecular field
• Huge menu: The SNP Consortium (http://snp.cshl.org/ )
• The menu of Y SNPs includes over 4193 available Y SNPs (Nature 2001. 409:928)
• Contrast to under 100 available Y STRs ( http://www.cstl.nist.gov/biotech/strbase )

• Multiplexing capability
• “Easy” to score- on/off and Automate
A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms

The International SNP Map Working Group

* A full list of authors appears at the end of this paper.

We describe a map of 1.42 million single nucleotide polymorphisms (SNPs) distributed throughout the human genome, on an average density of one SNP every 1.9 kilobases. These SNPs were primarily discovered through analysis of clone overlaps by the International Human Genome Sequencing Consortium. We estimate that the density of SNPs across the genome, in a manner broadly consistent with a standard population genetic model of human diversity, is consistent with the SNP map providing a public resource for defining haplotype variation across the genome, and should help inform important genes for diagnosis and therapy.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Length (bp)</th>
<th>All SNPs</th>
<th>SNPs</th>
<th>TSC SNPs</th>
<th>All SNPs per SNP</th>
<th>TSC SNPs per SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>214,066,000</td>
<td>129,931</td>
<td>1.85</td>
<td>75,166</td>
<td>2.85</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>222,899,000</td>
<td>103,064</td>
<td>2.15</td>
<td>70,985</td>
<td>2.90</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>186,938,000</td>
<td>93,140</td>
<td>2.10</td>
<td>63,609</td>
<td>2.94</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>169,035,000</td>
<td>84,426</td>
<td>2.00</td>
<td>65,719</td>
<td>2.57</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>170,954,000</td>
<td>117,882</td>
<td>1.45</td>
<td>63,545</td>
<td>2.69</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>165,022,000</td>
<td>96,817</td>
<td>1.71</td>
<td>53,797</td>
<td>3.07</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>149,414,000</td>
<td>71,752</td>
<td>2.09</td>
<td>42,327</td>
<td>3.53</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>125,148,000</td>
<td>57,834</td>
<td>2.16</td>
<td>42,653</td>
<td>2.93</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>107,440,000</td>
<td>62,013</td>
<td>1.73</td>
<td>43,020</td>
<td>2.50</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>127,894,000</td>
<td>61,288</td>
<td>2.09</td>
<td>42,466</td>
<td>3.01</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>129,193,000</td>
<td>84,663</td>
<td>1.53</td>
<td>47,261</td>
<td>2.71</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>125,198,000</td>
<td>59,245</td>
<td>2.11</td>
<td>38,306</td>
<td>3.28</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>93,711,000</td>
<td>53,093</td>
<td>1.77</td>
<td>35,745</td>
<td>2.62</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>89,044,000</td>
<td>44,112</td>
<td>2.03</td>
<td>29,746</td>
<td>3.00</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>73,467,000</td>
<td>37,814</td>
<td>1.94</td>
<td>25,524</td>
<td>2.77</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>74,037,000</td>
<td>38,785</td>
<td>1.91</td>
<td>23,328</td>
<td>3.17</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>73,367,000</td>
<td>34,621</td>
<td>2.12</td>
<td>19,396</td>
<td>3.78</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>73,078,000</td>
<td>45,195</td>
<td>1.62</td>
<td>27,028</td>
<td>2.70</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>56,044,000</td>
<td>25,676</td>
<td>2.18</td>
<td>11,185</td>
<td>5.01</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>63,317,000</td>
<td>29,478</td>
<td>2.15</td>
<td>17,051</td>
<td>3.71</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>33,824,000</td>
<td>20,916</td>
<td>1.62</td>
<td>9,103</td>
<td>3.72</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>33,786,000</td>
<td>28,410</td>
<td>1.19</td>
<td>11,056</td>
<td>3.06</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>181,245,000</td>
<td>34,842</td>
<td>3.77</td>
<td>20,400</td>
<td>6.43</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>21,753,000</td>
<td>4,193</td>
<td>1.08</td>
<td>1,784</td>
<td>12.19</td>
<td></td>
</tr>
</tbody>
</table>

RefSeq totals: 1,698,674

Total: 2,710,164

Length (bp) is from the public Genome Assembly of 5 September 2000. Density of SNPs on each chromosome is influenced by the amount of available genome sequence included in the Genome Assembly, depth of overlap coverage from TSC reads and clone overlaps, and the underlying heterozygosity (Table 2). Data are presented for the entire dataset of all SNPs and for those that are identically sequenced in the SNP consortium (TSC SNPs), as the latter are more evenly spaced than those from clone overlaps.
Other Applications of SNPs (aside forensics)

- **Medical Diagnostics**
  - Tissue typing- HLA DQ alpha typing
  - Cystic Fibrosis
  - Inflammatory panels
  - Neuro-psychiatric illnesses
  - Cancers
  - Chronic degenerative diseases

- **Pharmacogenomics - Predictive Pharmacology**
  - Association of genotype to drug response
  - Genetic population studies of patients and their responses to treatment
  - Personalized Medicine

- **Genetic Linkage studies- SNP Haplotyping**
Detection Technologies for Y SNPs in Forensics

- **Primer Extension**: SNapShot- aka minisequencing. Dugan et al. 2003
- **Pyrosequencing**: Ballantyne, J. 2003 AAFS
- **Quadruopole MS**: Eckenrode et al. 2003 AAFS
- **Bead based assays**: Luminex, Marligen Biosciences. Carlson et al 2002
SNPs on the Luminex

Y snp oligos attached to beads

Labeled PCR targets with labels

Beads with Hybed targets

Butler, J. et al. 2003
Bead Based Assay- Luminex 100

DNA Gumballs

- Internal Spectral Address™ R:IR ratio (gumball color) identifies each of the assays (probe?)

- Reporter fluorescence on the surface (target?) is quantified
Recap of technology

1- Beads **flow** single file past two lasers
   - **633nm** excites 2 dyes in beads
   - **532nm** excites dye on target if there

2- Detectors capture:
   - R:IR ratio → **SNP probe ?**
   - Scatter → **Single bead?**
   - Reporter fluor → **SNP target ?**

3- Digital signal processor:
   - Collects, processes and saves the data (csv)
   - Records median fluorescence intensity (mfi)
Universal Arrays: Tm Bioscience

- 100 unique tags
- No C: 75% A/T + 25% G
- 24-mers (six 4-bp motifs)
- Isothermal (± 2°C)
- Minimal cross-talk

![Diagram of allele specific PCR primer and tag interaction]
Universal arrays and Primer Extension – SNP Detection Format Alternatives

• Reproducibility: Pre-coupled capture probes eliminates any conjugation variability
• Flexibility: Bead-capture/probe sets for any loci
• Specificity: Primer-extension enhanced
• Multiplexing: 100 capture probes are isothermal (Tm ± 2°C) for Tm Bioscience beads
Primer Extension (allele specific), Tm Universal arrays and Luminex 100

Wild Type

3’

Tag 1

G

C

Mutant

3’

Tag 2

T

A

Hybe Allele-Specific Tagged Primers

Extend in the presence of Biotin dCTP

Hybridize to Bead
Detect with SA-PE

Multiplex PCR Products
<table>
<thead>
<tr>
<th>Application</th>
<th>Reference/Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterial ID</td>
<td>Ye et al. 2001 Hum Mut. 17:305</td>
</tr>
<tr>
<td>Biodefense</td>
<td>Los Alamos National Laboratory</td>
</tr>
<tr>
<td>Conservation Genetics</td>
<td>UC Davis- BML</td>
</tr>
<tr>
<td>Forensics</td>
<td>Carlson et al. 2002 ISHI</td>
</tr>
<tr>
<td>Plant Gene Expression</td>
<td>Yang et al. 2001 Genome Res. 11: 1888</td>
</tr>
<tr>
<td>Haplotyping</td>
<td><a href="http://www.polygenyx.com">www.polygenyx.com</a></td>
</tr>
<tr>
<td>HLA Testing</td>
<td><a href="http://www.onelambda.com">www.onelambda.com</a></td>
</tr>
<tr>
<td>Human Identity Testing</td>
<td><a href="http://www.marligen.com">www.marligen.com</a></td>
</tr>
<tr>
<td>Oncology</td>
<td><a href="http://www.mutlimetrix.com">www.mutlimetrix.com</a></td>
</tr>
<tr>
<td>Paternity testing</td>
<td><a href="http://www.luminexcorp.com">www.luminexcorp.com</a></td>
</tr>
<tr>
<td>Trichosporon spp</td>
<td>University of Miami/<a href="http://www.miraibio.com">www.miraibio.com</a></td>
</tr>
<tr>
<td>Thrombophilia</td>
<td><a href="http://www.luminexcorp.com">www.luminexcorp.com</a></td>
</tr>
<tr>
<td>Universal Arrays</td>
<td>Tm Bioscience</td>
</tr>
</tbody>
</table>
METHODS

Fluorescent Microsphere-Based Readout Technology for Multiplexed Human Single Nucleotide Polymorphism Analysis and Bacterial Identification

Fei Ye,1* May-Sung Li,1 J. David Taylor,1 Quan Nguyen,2 Heidi M. Colton,3 Warren M. Casey,3 Michael Wagner,2 Michael P. Weiner,1 and Jingwen Chen1

1Department of Genomic Sciences, Glaxo Wellcome Research and Development, Research Triangle Park, North Carolina
2Department of Human Genetics, Glaxo Wellcome Research and Development, Research Triangle Park, North Carolina
3Department of Analytical Sciences, Glaxo Wellcome Research and Development, Research Triangle Park, North Carolina

Ye et al. 2001. Human Mutation 17:305
Entire Issues dedicated to SNP technology and Applications
**Bacterial Identification using 16S rDNA SNPs**

Ye et al. 2001 Human Mutation. 17:305-316

**A**

Universal forward primer

![Diagram of primer binding sites](image)

**B**

**Bacteria strains**

Identification patterns generated by primer extension assays

<table>
<thead>
<tr>
<th>Bacterial Strain</th>
<th>Probe 1</th>
<th>Probe 2</th>
<th>Probe 3</th>
<th>Probe 4</th>
<th>Probe 5</th>
<th>Probe 6</th>
<th>Probe 7</th>
<th>Probe 8</th>
<th>Probe 9</th>
<th>Probe 10</th>
<th>Probe 11</th>
<th>Probe 12</th>
<th>Probe 13</th>
<th>Probe 14</th>
<th>Probe 15</th>
<th>Probe 16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacillus circulans</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Bacillus cereus</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>C</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Bacillus stearothermophilus</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>#</td>
<td>G</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Bacillus subtilis</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>#</td>
<td>G</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Listeria monocytogenes</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>#</td>
<td>T</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Staphylococcus xylosus</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>G</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>G</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Staphylococcus epidermidis</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>C</td>
<td>#</td>
<td>A</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Staphylococcus aureus</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>T</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>G</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>G</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
</tbody>
</table>

---

**Figure 4.** Design of the capture probes for multiplex ASPE or SBCE assays using 16S rDNA. **A:** Physical locations of the 16S rDNA probes. Based on the multi-alignment of different bacterial 16S rDNA sequences, 16 conserved regions were chosen for SBCE and ASPE assays. For SBCE reactions, probes are designed such that the 3' end of the primers terminates one base 5' to the variable site. For the ASPE assay, a pair of probes was designed such that the 3' end differs from each other at the variable site. The locations of the probes are listed in Table 1. **B:** Polymorphic patterns in multiplexed SBCE and ASPE assays. Bacterial species can be divided into 17 groups based on their unique readout patterns.
ASPE vs SBCE on 16S rDNA SNPs

**Staphylococcus aureus (ATCC 25923)**

<table>
<thead>
<tr>
<th>Pattern: p1-A/p2-C/p3-T/p4-C/p5-G/p6-T/p7-A/p8-G/p9-</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
</tr>
<tr>
<td>A</td>
</tr>
</tbody>
</table>

**Staphylococcus aureus (ATCC 25923)**

<table>
<thead>
<tr>
<th>Pattern: p1-A/p2-C/p3-T/p4-C/p5-G/p6-T/p7-A/p8-G/p9-</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
</tbody>
</table>

**Escherichia coli (ATCC 26)**

<table>
<thead>
<tr>
<th>Pattern: p1-G/p2-C/p3-G/p4-T/p5-A/p6-G/p7-C/p8-G/p9-</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
</tr>
</tbody>
</table>

**Escherichia coli (ATCC 26)**

<table>
<thead>
<tr>
<th>Pattern: p1-G/p2-C/p3-G/p4-T/p5-A/p6-G/p7-C/p8-G/p9-</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
</tr>
</tbody>
</table>

Ye et al. 2001 Hum. Mut.17:305-316
## SNPs vs STRs

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STRs</strong></td>
<td><strong>SNPs</strong></td>
</tr>
<tr>
<td>Higher PD</td>
<td>Lower PD-need 50-100</td>
</tr>
<tr>
<td>Multiplexes available</td>
<td>Mixture limited</td>
</tr>
<tr>
<td>Databases established</td>
<td>New instrumentation</td>
</tr>
<tr>
<td>Familiar instrumentation</td>
<td>Databases not established for forensics</td>
</tr>
<tr>
<td></td>
<td>Validated kits missing</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited abundance</td>
<td></td>
</tr>
<tr>
<td>Stutter</td>
<td></td>
</tr>
<tr>
<td>Extremely degraded samples</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Abundance</td>
<td></td>
</tr>
<tr>
<td>SNP Consortium</td>
<td></td>
</tr>
<tr>
<td>High-throughput automation</td>
<td></td>
</tr>
<tr>
<td>Highly degraded samples</td>
<td></td>
</tr>
</tbody>
</table>
SNPs and STRs

- No STR results - eg 911 samples benefit from SNP typing
- SNPs utilized as a rapid screen – Used to exclude.
- SNPs as additional markers when STRs don’t provide sufficient discrimination (mass disasters where total families are lost - relatives with high numbers of shared alleles)
- SNPSTRs?
SNPs and STRs

SNPSTRs: “Each such segment includes one or more single nucleotide polymorphisms (SNPs) and exactly one short tandem repeat (STR) locus.”

Summary

• **MtDNA** – Well studied, HV regions - degraded DNA. Maternal lineage reference samples missing persons databases - Dideoxysequencing detection.

• **Y chromosome markers** - 98% of violent crime by males, useful on mixtures and sexual assault evidence, aspermic individuals and missing persons.

• **Why Single Nucleotide Polymorphisms (SNPS)**
  • Well studied, Huge selection, multiplexed and automated.
  • Primer Extension, Pyrosequencing, Light Cycling, Mass Spec, Bead based assays - Luminex.

• **SNPs vs STRs or SNPs and STRs**
  • Either/Or Why SNPs?