Linkage Disequilibrium Analysis of D12S391 and vWA

Kristen Lewis O'Connor, Ph.D.
Applied Genetics Group
National Institute of Standards and Technology
Gaithersburg, Maryland, USA

English Speaking Working Group Meeting
Stockholm, Sweden

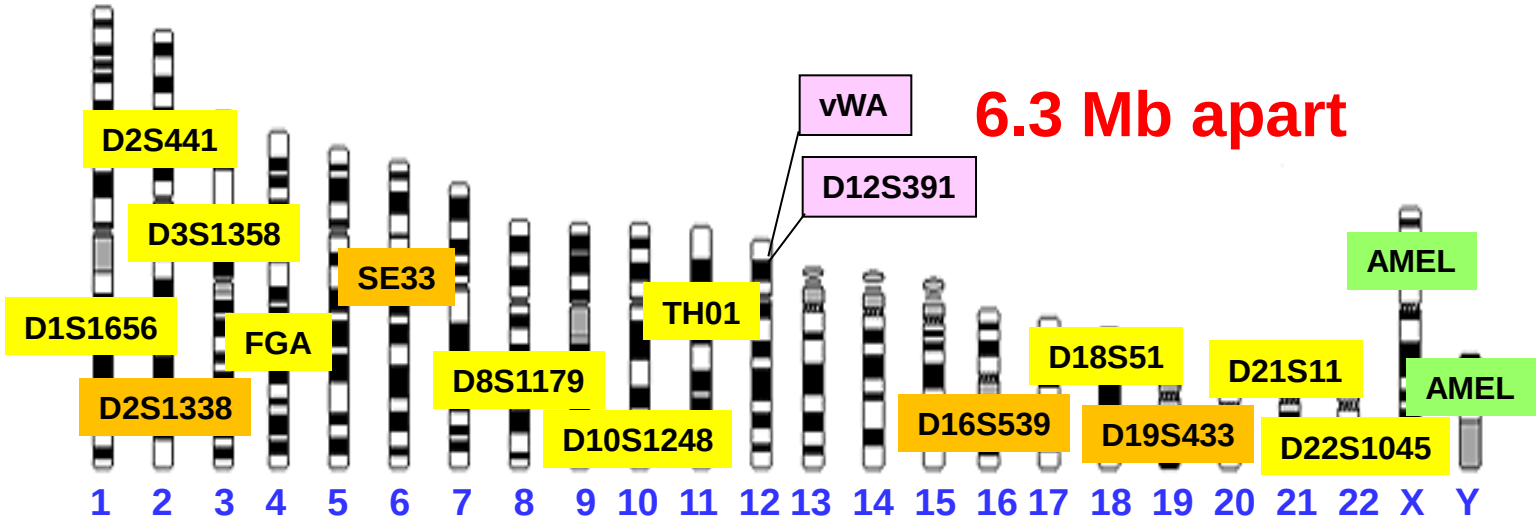
NIST
National Institute of Standards and Technology

European Standard Set of Forensic Loci

- Prior to April 2009, European Standard Set consisted of seven STR loci
  - TH01, **vWA**, FGA, D8S1179, D18S51, D21S11, and D3S1358
- In November 2009, European Union adopted five additional STRs
  - **D12S391**, D1S1656, D2S441, D10S1248, and D22S1045
- These loci are included in the next generation of multiplex PCR kits
  - PowerPlex® ESX and ESI Systems (Promega)
  - AmpF/STR® NGM™ (Applied Biosystems)
  - Investigator ESSplex and ESSplex SE Kits (Qiagen)

Chromosomal Positions for the European Standard Set

and Other Common STR Markers Used



European Standard Set + D16S539, D2S1338, D19S433, SE33

Genetic Markers for Forensic Use

- Ensure full recombination and independent inheritance
- Markers on the same chromosome should be at least 50 Mb apart (ideal for forensic use)
- CODIS loci CSF1PO and D5S818 (26 Mb apart) are considered statistically independent
  - No deviation from Hardy-Weinberg equilibrium, no linkage disequilibrium in population samples
    J.W. Bacher et al., Chromosome localization of CODIS loci and new pentanucleotide repeat loci, Progress in Forensic Genetics 8 (2000) 33–36
  - Small increased effect of linkage on match probabilities noted in full and half sibling pairs
    J. Buckleton, C. Triggs, The effect of linkage on the calculation of DNA match probabilities for siblings and half siblings, Forensic Sci. Int. 160 (2006) 193–199.

- **Are vWA and D12S391 (6.3 Mb apart) independent?**
- **Should vWA and D12S391 be used with the product rule for match probability calculations?**

Research Design

- NIST U.S. population samples
  - 254 African American, 261 Caucasian, 139 Hispanic
- U.S. father/son samples
  - 178 African American, 198 Caucasian, 190 Hispanic, 198 Asian
- Previously genotyped with PowerPlex® ESI/ESX 17
- Father/son genotypes phased to identify paternally transmitted alleles

C.R. Hill *et al*., Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex® ESX 17 and ESI 17 Systems, Forensic Sci. Int. Genet. (in press).

Statistical Tests using Arlequin v. 3.5

- Hardy-Weinberg Equilibrium (population samples)
  - Exact test = test the hypothesis that the observed genotypes are the product of a random union of gametes
- Linkage Disequilibrium (population samples, phase unknown)
  - Likelihood ratio test = likelihood of sample when hypothesis of no association between loci vs. likelihood of sample when association is allowed
- Linkage Disequilibrium (paternity samples, phase known)
  - Exact test = test for the presence of significant association between pairs of loci

L. Excoffier, G. Laval, S. Schneider, Arlequin ver. 3.0: An integrated software package for population genetics data analysis, Evolutionary Bioinformatics Online 1 (2005) 47–50.

P-values from analyses of Hardy-Weinberg equilibrium and linkage disequilibrium of the D12S391 and vWA loci using the **unrelated NIST U.S. population samples**

Population	N	Hardy-Weinberg Equilibrium		Linkage Disequilibrium
		D12S391	vWA	D12S391 and vWA
African American	254	0.0982	0.9853	0.1173
Caucasian	261	0.7814	0.1381	0.1503
Hispanic	139	0.6434	0.9329	0.8777

Significance level, $p < 0.05$.

No significant departure from HWE for D12S391 or vWA ($p > 0.05$)

No significant linkage disequilibrium between the loci ($p > 0.05$)

Consistent with results from seven worldwide populations

C. Phillips *et al*., Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel, Forensic Sci. Int. Genet. (in press).

P-value results from analysis of linkage disequilibrium of the D12S391 and vWA loci using **U.S. father/son paternity samples**

Population	N	Linkage Disequilibrium
African American	178	**0.0275**
Caucasian	198	**0.0001**
Hispanic	190	0.0915
Asian	198	**0.0031**

N = number of father/son samples. Significance level, $p < 0.05$.

Evidence of LD in African American, Caucasian, and Asian paternity samples

No significant LD detected in Hispanic paternity samples

- Population effect is possible

Linkage Disequilibrium between D12S391 and vWA

- Use of father/son pairs allowed for allelic phase to be determined
  - Significant LD was detected
  - Non-random association of alleles at D12S391 and vWA

- LD is more difficult to detect in unrelated population samples due to less power
  - Unknown allelic phase
  - Large number of possible haplotypes

Profile Probability Calculations

For casework analysis that involves unrelated or related individuals:

- Single-locus genotype probabilities of D12S391 and vWA **should not** be multiplied to determine the STR profile probability

- Possible solutions:
  1. Choose one locus for profile probability calculations
  2. Use haplotype frequencies of D12S391/vWA diplotype

A diplotype consists of two haplotypes, which are phased multilocus genotypes

Single Locus vs. Haplotype Approach

- African American allele frequencies
  - Most common allele for D12S391 = 0.267
  - Most common allele for vWA = 0.254

- African American haplotype frequencies in phased paternity samples*
  - Most common haplotype for D12S391/vWA = 0.087

Haplotype frequencies are generally rarer than the allele frequencies of a single locus

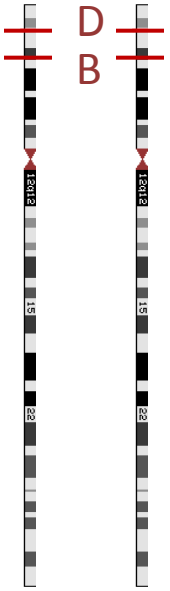
* O'Connor, K.L., et al, Linkage disequilibrium analysis of D12S391 and vWA in U.S. population and paternity samples, *Forensic Sci. Int. Genet.* (accepted)

Haplotype Approach with Unphased Alleles

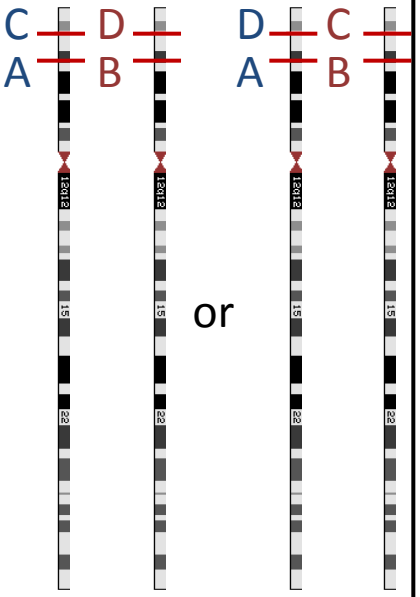
Family reference data may not be available to infer the gametic phase of alleles at D12S391 and vWA

Unphased genotype:
A,B at D12S391
C,D at vWA

vWA C D
D12S391 A B



Two possible diplotypes from four haplotype combinations



Haplotype frequency

Probability of observing either one of the possible diplotypes

If a diplotype is comprised of two different haplotypes (i.e., AC ≠ BD and AD ≠ BC), then

$$Pr(AC/BD \text{ or } AD/BC) = 2(p_{AC}p_{BD} + p_{AD}p_{BC})$$

If a diplotype is comprised of the same haplotypes (i.e., AC = BD = AD = BC), then

$$Pr(AC/BD \text{ or } AD/BC) = 2p_{AC}^2$$

Summary

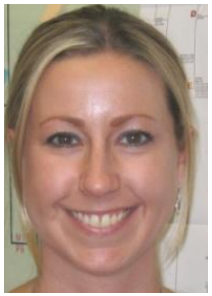
- U.S. is looking to expand the core loci to provide more international overlap (18-20 loci total)
- Single-locus genotype probabilities **should not** be multiplied for match probability calculations
- Recommend using haplotype frequencies of a D12S391/vWA diplotype for profile probability calculations
  - Allows for consideration of genotype data from both loci without statistical bias
- Further work with multi-generation families is needed to determine the actual recombination fraction between the linked D12S391 and vWA loci
  - Assess the impact of the linkage effect on match probability calculations

* O'Connor, K.L., et al, Linkage disequilibrium analysis of D12S391 and vWA in U.S. population and paternity samples, *Forensic Sci. Int. Genet.* (accepted)

Acknowledgments

Applied Genetics Group Leader

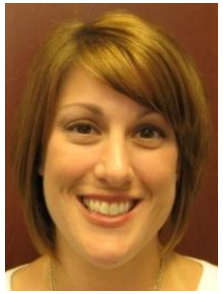
John Butler



Becky Hill

DNA Biometrics Project Leader

Peter Vallone



Kristen Lewis O'Connor

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

kristen.oconnor@nist.gov

Funding

NRC – Postdoctoral fellowship support for Kristen O'Connor

FBI – Application of DNA Typing as a Biometric Tool

NIJ – Forensic DNA Standards, Research, and Training