EDNAP and 34th ENFSI DNA WG Meeting
April 6-8, 2011 – Brussels, Belgium

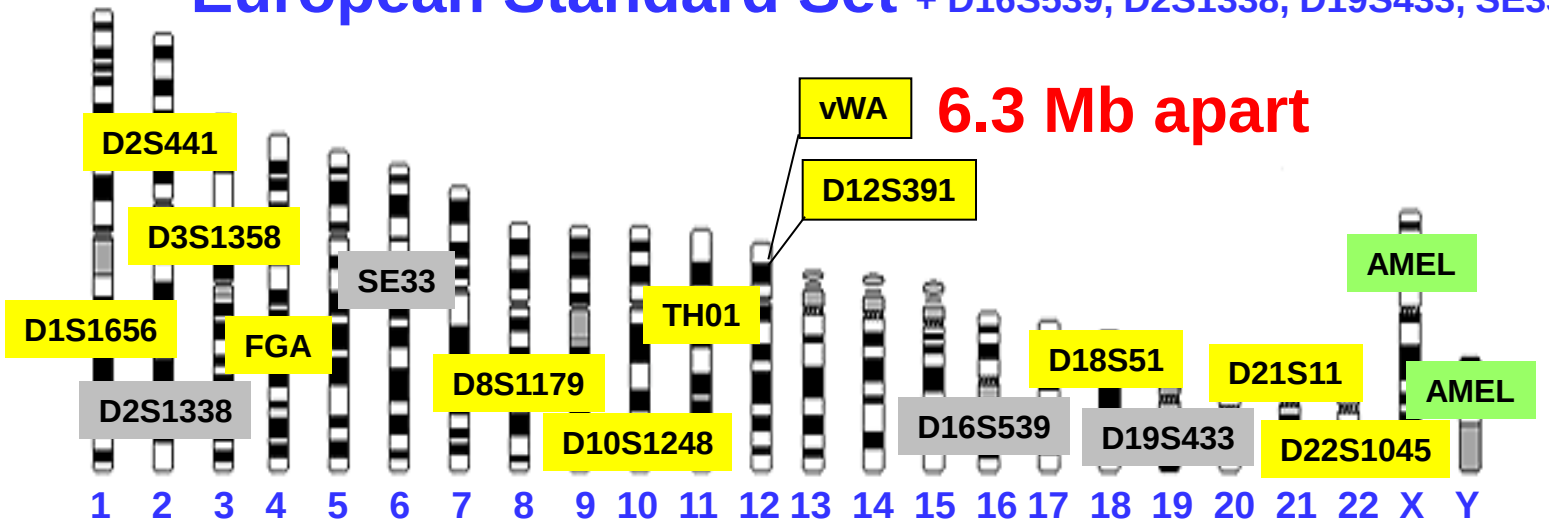
Reassessment of Linkage Disequilibrium Studies with D12S391 and vWA

Kristen O'Connor & John Butler
NIST

Common Forensic STR Loci

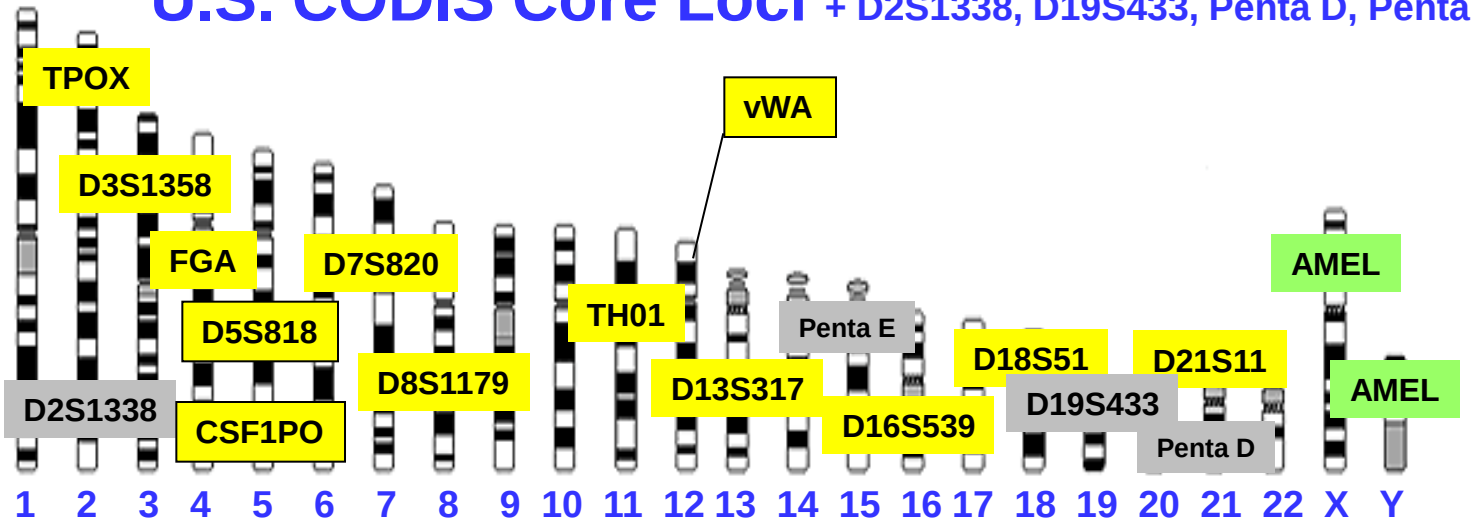
Europe

European Standard Set + D16S539, D2S1338, D19S433, SE33



United States

U.S. CODIS Core Loci + D2S1338, D19S433, Penta D, Penta E



European Standard Set of Forensic Loci

- Prior to 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*l*STR NGM (Applied Biosystems)

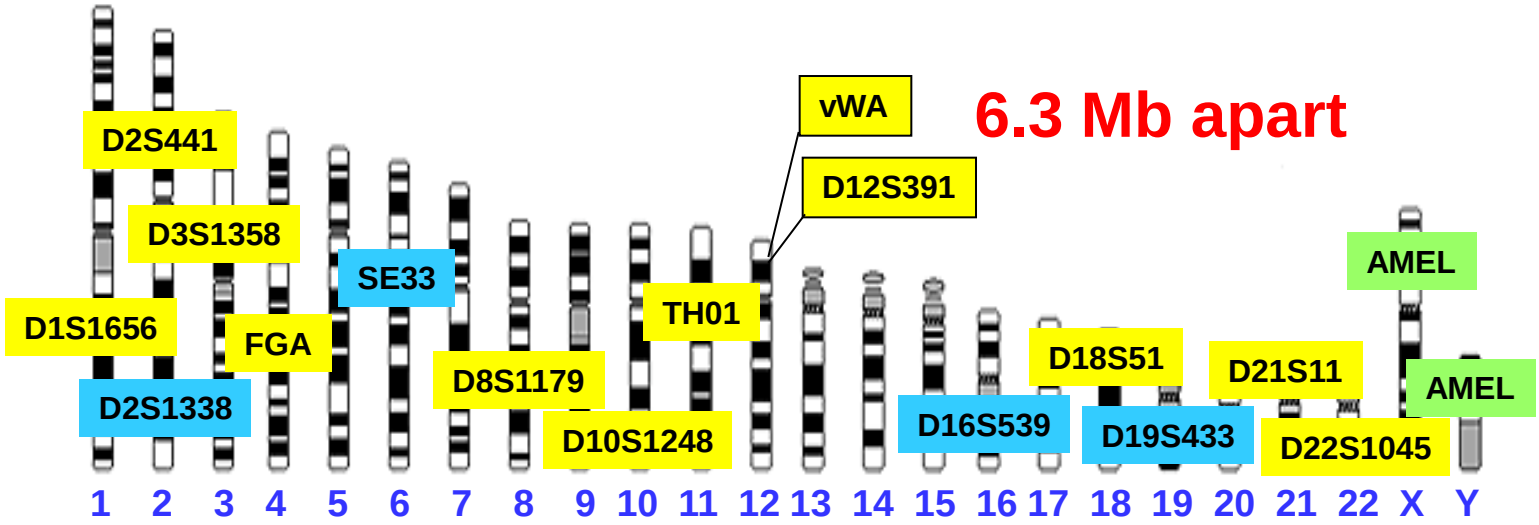
Genetic Markers for Forensic Use

- 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?**

Chromosomal Positions for the European Standard Set

and Other Common STR Markers Used



European Standard Set + D16S539, D2S1338, D19S433, SE33

Should vWA and D12S391 be used with the product rule for match probability calculations?

Locations of vWA and D12S391

vWA chr12:6,093,104 – 6,093,253

D12S391 chr12:12,449,874 – 12,450,226

6.3 megabases apart on chromosome 12

UCSC Genome Browser, Feb. 2009 assembly

Linkage and Linkage Disequilibrium (LD)

- ***Linkage*** is the co-segregation of closely located loci in a pedigree
  - Linkage refers to **physical dependency** between loci
  - Linkage can be measured between two or more genetic markers through obtaining recombination frequencies using family samples

- ***Linkage disequilibrium*** (LD) measures the allele co-segregation at a population level
  - LD relates to **allelic dependency** between alleles at different loci
  - LD is estimated from allele frequencies and haplotype frequencies measured in unrelated population samples

Tillmar, A.O., et al. (2008). Analysis of linkage and linkage disequilibrium for eight X-STR markers. *Forensic Science International: Genetics, 3,* 37-41.

Ott, J. (1999). Analysis of Human Genetic Linkage, 3rd edition, Baltimore: The Johns Hopkins University Press.

G Model
FSIGEN 643 1–3

ARTICLE IN PRESS

Forensic Science International: Genetics xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Short communication

Linkage disequilibrium analysis of D12S391 and vWA in U.S. population and paternity samples☆

Kristen Lewis O'Connor*, Carolyn R. Hill, Peter M. Vallone, John M. Butler

Biochemical Science Division, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8312, United States

K.L. O'Connor, et al., Linkage disequilibrium analysis of D12S391 and vWA in U.S. population and paternity samples, Forensic Sci. Int. Genet. (2010), doi:10.1016/j.fsigen.2010.09.003

Reassessment of these results – an erratum will be published with the article

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
- Tests for Hardy-Weinberg equilibrium and linkage disequilibrium in population samples
- Test for linkage disequilibrium in father/son samples

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. (2010), doi:10.1016/j.fsigen.2010.03.014.

Linkage Disequilibrium between D12S391 and vWA

- Population samples
  - No significant departure from HWE for D12S391 or vWA
  - No significant linkage disequilibrium detected between the loci
  - 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. (2010), doi:10.1016/j.fsigen.2010.02.003.

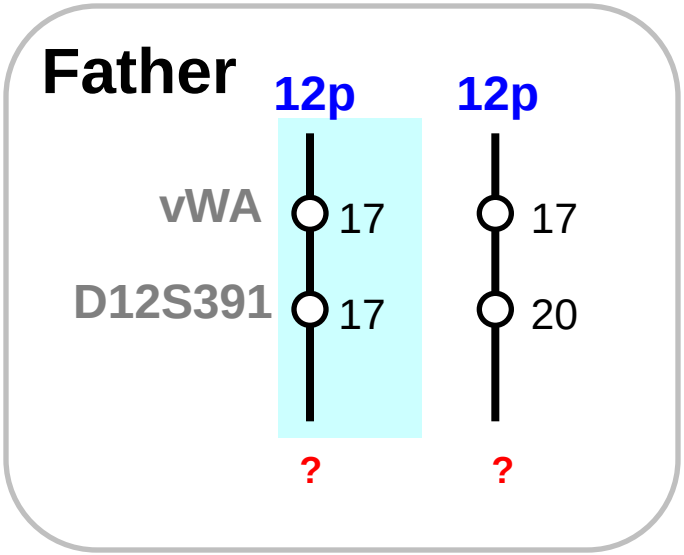
- Paternity samples with known allelic phase
  - ~~**Significant linkage**~~ between D12S391 and vWA
  - ~~**Non-random association of alleles**~~ at D12S391 and vWA

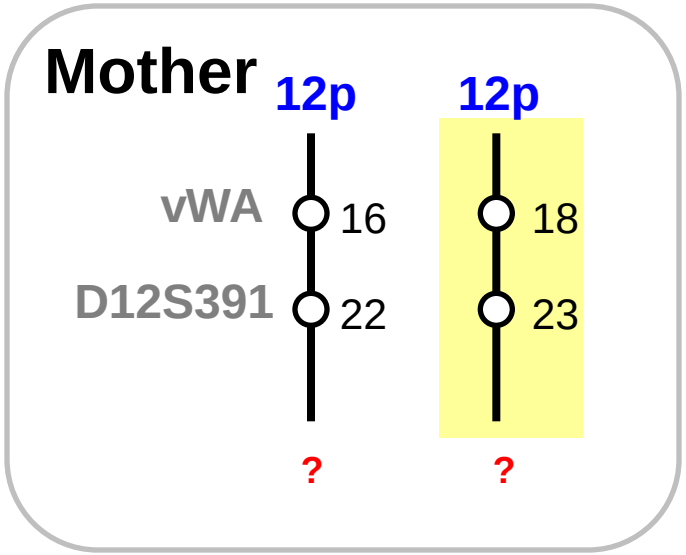
- We surmise that linkage disequilibrium is present in unrelated population samples but is more difficult to detect due to less power
  - Unknown allelic phase
  - Large number of possible haplotypes

K.L. O'Connor, et al., Linkage disequilibrium analysis of D12S391 and vWA in U.S. population and paternity samples, Forensic Sci. Int. Genet. (2010), doi:10.1016/j.fsigen.2010.09.003

Inheritance, Linkage, and Phased Haplotypes

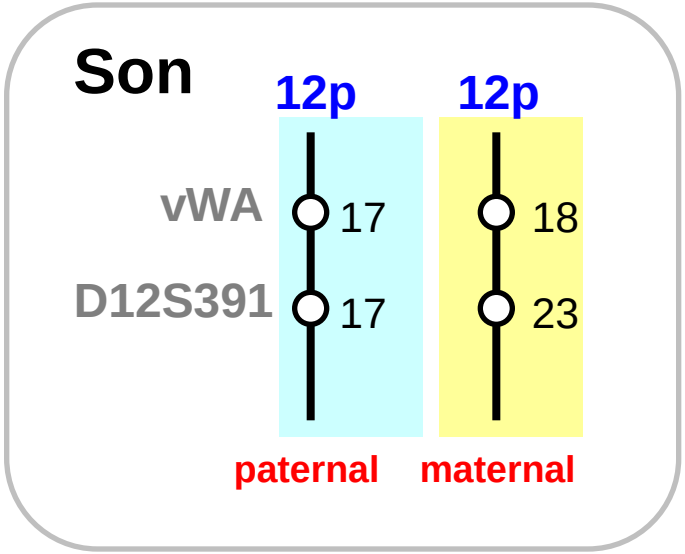
Genotypes
vWA: 17,17
D12: 17, 20





Genotypes
vWA: 16,18
D12: 22,23

Genotypes
vWA: 17,18
D12: 17, 23



Possible Haplotypes
17-17 and 18-23 or
18-17 and 17-23

Phased Haplotypes
17-17 and 18-23

Re-testing has shown no LD

G Model
FSIGEN-643; No. of Pages 3

ARTICLE IN PRESS

Forensic Science International: Genetics xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Short communication

Linkage disequilibrium analysis of D12S391 and vWA in U.S. population and paternity samples☆

Kristen Lewis O'Connor*, Carolyn R. Hill, Peter M. Vallone, John M. Butler

Biochemical Science Division, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8312, United States

Original abstract: No significant evidence of linkage disequilibrium was observed between the loci in the population samples. However, ***significant linkage disequilibrium was detected in U.S. African American, Caucasian, and Asian father/son samples with phased genotypes***.

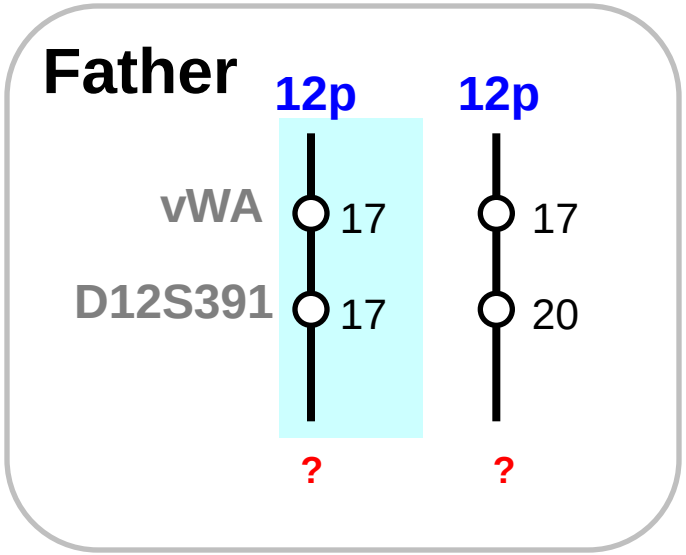
An errata will be published with this article

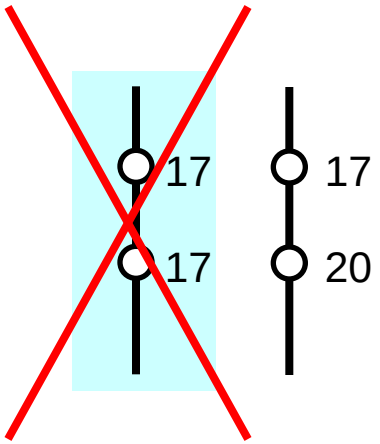
Summary of our Erratum

- **The phased father/son dataset mistakenly included both pairs of haplotypes** so that the paternally transmitted haplotype was counted twice (i.e., once for the father and once for the son).
- A transmitted haplotype is not independent between a father/son pair and should not have been counted separately for linkage disequilibrium analysis and haplotype frequency calculations.
- Also, it was determined that the phase of some father/son genotypes could not be definitively determined without the genotype of the mother, which was unavailable in this dataset.

Why an Erratum Is Necessary: Extra count of paternal haplotype

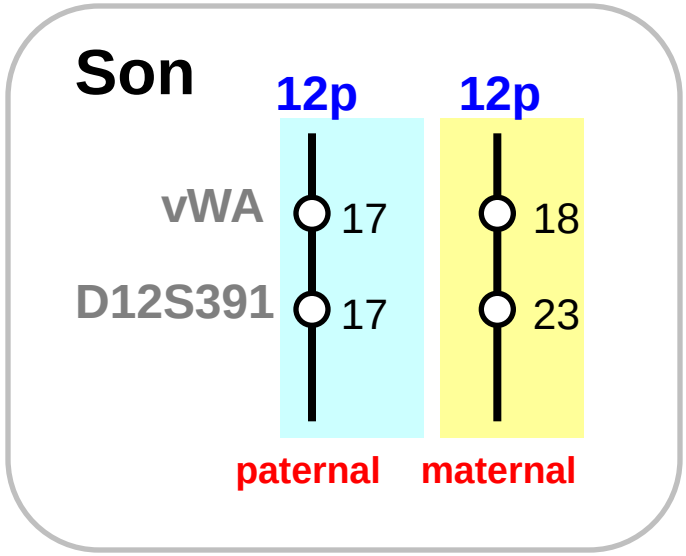
Genotypes
vWA: 17,17
D12: 17, 20

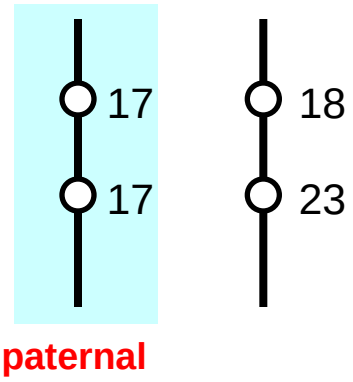




Should not have double-counted the paternally transmitted haplotype in LD studies

Genotypes
vWA: 17,18
D12: 17, 23



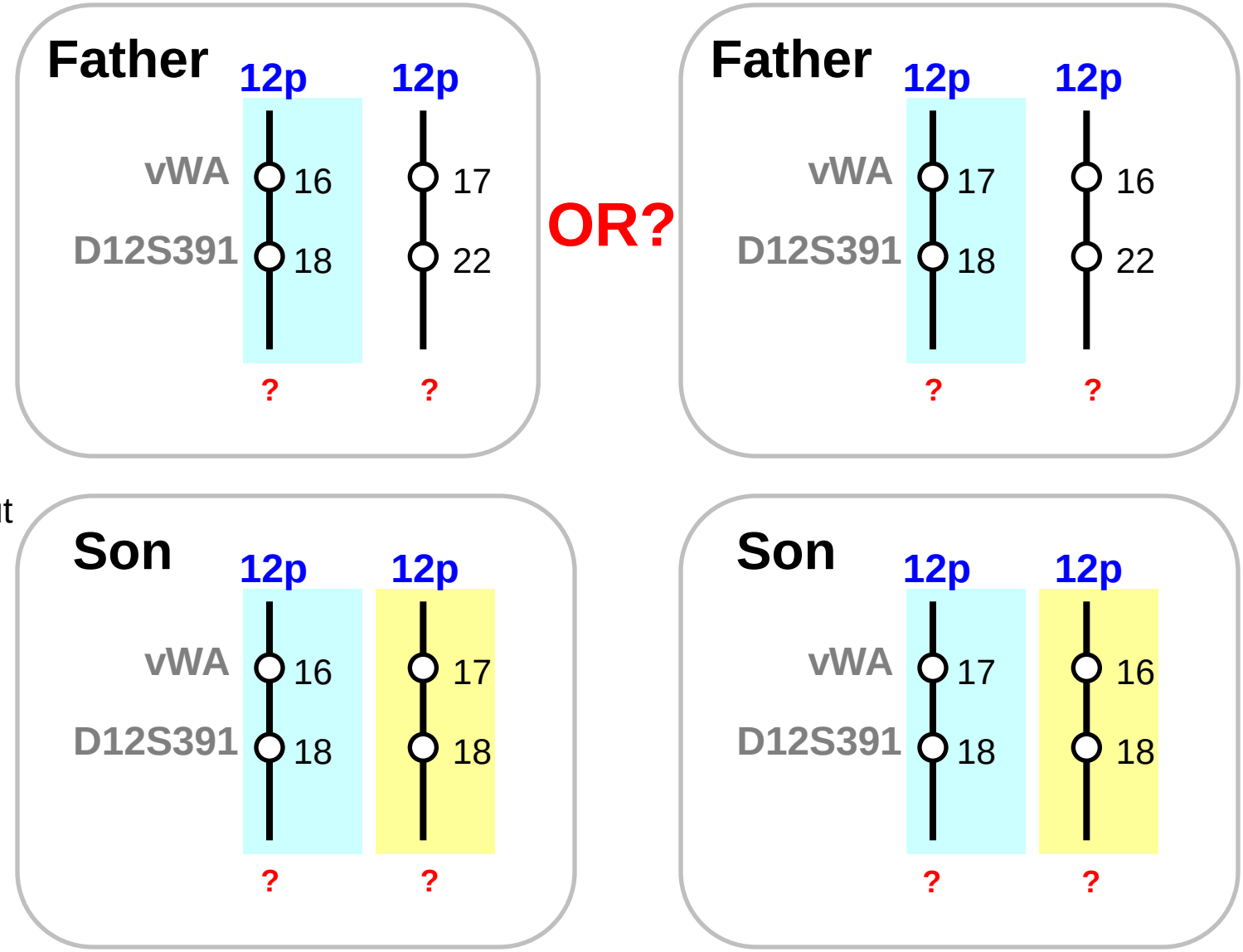


Why an Erratum Is Necessary: Ambiguous phasing of father/son alleles

Genotypes
vWA: 16,17
D12: 18, 22

The phase of some father/son genotypes could not be definitively determined without the genotype of the mother.

Genotypes
vWA: 16,17
D12: 18, 18



Re-testing has shown no LD

Two parallel methods were used to revise the D12S391/vWA haplotype data used for linkage disequilibrium analysis

Assistance from Rolf Fimmers and Andreas Tillmar

Population	Original Data		Dataset 1		Dataset 2	
	N	LD	N	LD	N	LD
African American	356	**0.0275**	214	0.4888	356	0.8756
Caucasian	396	**0.0001**	250	0.2195	396	0.1685
Hispanic	380	0.0915	228	0.7105	380	0.8857
Asian	396	**0.0031**	217	0.1317	396	0.6541

Upon revised analysis of D12S391/ vWA haplotypes, no significant evidence of LD was observed in the U.S. father/son population groups

* Consistent with results in the unrelated U.S. population samples as described in the original paper

* Conclusion of independence between D12S391 and vWA is consistent with findings from two recent articles

Dataset 1 (NIST and Andreas Tillmar)

- **The father/son dataset was reduced by removing one count of the paternally transmitted haplotype in addition to removing sample pairs if the gametic phase was ambiguous.** The revised dataset included three haplotypes per father/son pair (one haplotype for the father and two haplotypes for the son) for all unambiguously phased father/son pairs. Linkage disequilibrium analysis was performed as described in the Materials and Methods section of the original paper.

Dataset 2 (performed by Rolf Fimmers)

- The expectation-maximization (EM) algorithm was used to estimate the four parental haplotypes (two paternal and two maternal) for each son, assuming a recombination fraction of 0.108 (Budowle et al. 2011). From these estimated haplotypes, a maximum likelihood approach estimated the parental haplotype frequencies.

- Since the likelihood ratio test does not perform well if there are many haplotypes with low frequencies, alleles with frequencies below 5% were pooled for each locus prior to linkage disequilibrium analysis.

No evidence of association (LD) in population samples

G Model
FSIGEN-571; No. of Pages 15

ARTICLE IN PRESS

Forensic Science International: Genetics xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel

C. Phillips [a,b,*], L. Fernandez-Formoso [a], M. Garcia-Magariños [a], L. Porras [a], T. Tvedebrink [c], J. Amigo [b], M. Fondevila [a], A. Gomez-Tato [d], J. Alvarez-Dios [d], A. Freire-Aradas [a], A. Gomez-Carballa [a], A. Mosquera-Miguel [a], Á. Carracedo [a,b], M.V. Lareu [a]

[a] Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Galicia, Spain
[b] Genomics Medicine Group, CIBERER, University of Santiago de Compostela, Santiago de Compostela, Galicia, Spain
[c] Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark
[d] Faculty of Mathematics, University of Santiago de Compostela, Santiago de Compostela, Galicia, Spain

Phillips et al. (2010) 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)

No evidence of association (LD) in population samples

2.5. Analysis of potential association by linkage of vWA and D12S391 alleles

The ESS multiplex combines two closely positioned STRs on chromosome 12: the established marker vWA and the new D12S391. The loci are sited at nucleotide positions: 6,093,104 and 12,449,930 respectively [20: NCBI genome build 37.1] – a separation of ~6.37 Mb which is closer than any other commonly used same-chromosome STR pair. So assumptions of independence must be properly tested ahead of using these two STRs to routinely construct a cumulative frequency for ESS multiplex profiles. We made two complementary tests for association between vWA and D12S391 using the whole set of CEPH genotype pairs (i.e. disregarding population labels) to maximize the power of the tests.

Firstly, we used the χ^2 test of independence between D12S391 and vWA allele frequency distributions. As a large proportion of these combinations, involving the rarest alleles, had zero observations or very low frequencies they created considerable potential for error in χ^2 tests. Therefore we grouped certain allele classes into logical sets that had comparable repeat sizes, e.g. the rare vWA 11 and 13 repeats were combined with the common 14 repeats. Secondly, we performed simulations based on the CEPH panel allele frequency estimates for vWA and D12 to generate 10,000 simulated sample sets in order to comprehensively measure the accuracy of the p-value for the χ^2 test of independence. This allowed a comparison with the p-values obtained from the original χ^2 tests of the observed CEPH two-genotype combinations.

3.5. Tests for association by linkage of vWA and D12S391

For the unadjusted χ^2 tests of independence of vWA and D12 the p-value was 0 – rejecting a null hypothesis of independence, but as described the χ^2 test is not reliable in this case. In contrast, selective grouping of rare alleles into the nearest common repeat gave a p-value of 0.9188 strongly indicating marker independence. Grouped allele simulations gave a proportion of p-values smaller than the first test of 0.999 confirming the validity of the original χ^2 p-value.

Therefore despite some limitations imposed on a standard association test approach by a large number of rare alleles in both the chromosome-12 STRs, we found no evidence of association as a result of their close physical proximity. The handling of vWA and D12 genotypes from closely related individuals, as is normally the case in relationship testing, will need more thorough analysis with the reporting of these STR data as non-independent diplotypes most likely to be the optimum approach.

Phillips et al. (2010) 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)

No evidence for LD in unrelated individuals (population data)

Int J Legal Med (2011) 125:101–109
DOI 10.1007/s00414-010-0516-7

TECHNICAL NOTE

Population genetic analyses of the NGM STR loci

Bruce Budowle · Jianye Ge · Ranajit Chakraborty · Arthur J. Eisenberg · Robert Green · Julio Mulero · Robert Lagace · Lori Hennessy

Received: 6 August 2010 / Accepted: 14 September 2010 / Published online: 29 September 2010

Budowle et al. (2011) Population genetic analyses of the NGM STR loci. *Int. J. Legal Med.* 125: 101-109

Fine to multiply for unrelated individuals

- The test for evidence of linkage disequilibrium detected no more departures than would be expected by chance (four, six, and two out of 105 pairs of tests, in African Americans, Caucasians, and Hispanics, respectively). Furthermore, there were no departures detected for the two syntenic loci vWA and D12S391 (even before the Bonferroni correction [16]). **This observation is consistent with the findings of Phillips et al.** [12] and supports that **for identity testing multiplying the genotype frequencies is justified for the loci vWA and D12S391**. The independence between these two loci at the population level-based statistics is likely the result of the relatively high mutation rates of STR loci; **however, for kinship analyses vWA and D12S391 may be linked**; they are located on chromosome 12, at positions 6,093 kb and 12,340 kb, respectively [3]. The genetic distance is greater between these two loci at about 12 cM; however, while all physically linked loci are syntenic, physical map distances may not necessarily relate one-to-one with the recombination rate (e.g., due to presence of a recombinational hotspot). Thus, empirical testing is necessary to assess the degree of linkage effect (i.e., extent of co-segregation of alleles) for these two loci.

Budowle et al. (2011) Population genetic analyses of the NGM STR loci. *Int. J. Legal Med.* 125: 101-109

Loci are linked in relatives

- Three large pedigree CEPH families were typed for the NGM STR loci (ESM Table 5) and a linkage analysis was performed for the two biologically linked loci, vWA and D12S391. Using the software LINKAGE [10, 11], the **recombination fraction between the loci vWA and D12S391 was estimated at 0.108**. This result is close to the recombination distance (12 cM) and confirms the linkage of the loci vWA and D12S391. The recombination fraction can be used for estimating the co-inheritance of alleles at these two loci. Since the loci vWA and D12S391 are linked, independence cannot be assumed for kinship analyses.

- Therefore, there are two options recommended for the use of genetic data for kinship analyses: (1) **Incorporate the recombination rate and generate maximum likelihood estimates of the haplotype frequencies for these two loci**; or (2) use only one of the two loci in a kinship analysis, which would be the more informative of the two in a specific case.

Budowle et al. (2011) Population genetic analyses of the NGM STR loci. *Int. J. Legal Med.* 125: 101-109

Additional Studies of Linkage

- Collaboration with Andreas Tillmar
- NIST Family Samples (109 meioses)
- Recombination fraction estimated at **0.089** (95% CI = 0.044 – 0.158)
- Similar to Budowle et al. (2011) estimate of 0.108
- The confidence intervals reflects close linkage between D12S391 and vWA in family samples

Effect of Linkage on Kinship Analysis

- Increased tendency for alleles to be transmitted together during meiosis
- Important to use the appropriate haplotype frequencies and recombination fraction in likelihood calculations for kinship testing
  - M. Krawczak, Kinship testing with X-chromosomal markers: mathematical and statistical issues, Forensic Sci. Int. Genet. 1(2) (2007) 111–4.
- Ignoring linkage in likelihood calculations can lead to incorrect conclusions of kinship in some instances
  - M. Nothnagel, J. Schmidtke, M. Krawczak, Potentials and limits of pairwise kinship analysis using autosomal short tandem repeat loci. Int. J. Legal Med. 124(3) (2010) 205–15.
- Our erratum provides the maximum likelihood estimates of the D12S391/vWA haplotype frequencies

Timeline of Events

- Feb 2010 – NIST research on D12/vWA
- Mar 2010 – Phillips et al. paper appears in FSI-G (in press)
- Apr 2010 – NIST paper submitted to FSI-G

- Aug 2010 – Budowle et al. submits paper to IJLM
- Sept 2010 – Budowle et al. article appears on-line
- Nov 2010 – NIST article appears in FSI-G (in press)
- Jan 2011 – Budowle et al. article published in IJLM
- Mar 2011 – erratum on NIST article submitted from reassessment
- Apr 2011 – **presentation to EDNAP**

Updated Conclusions on D12S391 and vWA

- **Single-locus genotype probabilities** may be **multiplied** for profile match probability calculations when **unrelated individuals** are involved

- **Recombination fraction and** the maximum likelihood estimates of the D12S391/vWA **haplotype frequencies** should be used for **kinship analysis**