Forensic Science International: Genetics 6 (2012) 354–365

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

The recombination landscape around forensic STRs: Accurate measurement of genetic distances between syntenic STR pairs using HapMap high density SNP data

C. Phillips [a,*], D. Ballard [b], P. Gill [c], D. Syndercombe Court [b], Á. Carracedo [a], M.V. Lareu [a]

[a] Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Spain
[b] Academic Haematology Unit, Pathology Group, Blizard Institute, Barts and the London, Queen Mary's School of Medicine & Dentistry, London, UK
[c] Institute of Forensic Medicine, Rikshospitalet, University of Oslo, Norway

ARTICLE INFO

Article history:
Received 6 June 2011
Received in revised form 9 July 2011
Accepted 24 July 2011

Keywords:
Linkage
Linkage disequilibrium
STRs
X-STRs
Recombination
Genetic distance
Genetic maps

ABSTRACT

Family studies can be used to measure the genetic distance between same-chromosome (syntenic) STRs in order to detect physical linkage or linkage disequilibrium. However, family studies are expensive and time consuming, in many cases uninformative, and lack a reliable means to infer the phase of the diplotypes obtained. HapMap provides a more comprehensive and fine-scale estimation of recombination rates using high density multi-point SNP data (average inter-SNP distance: 900 nucleotides). Data at this fine scale detects sub-kilobase genetic distances across the whole recombining human genome. We have used the most recent HapMap SNP data release 22 to measure and compare genetic distances, and by inference fine-scale recombination rates, between 29 syntenic STR pairs identified from 39 validated STRs currently available for forensic use. The 39 STRs comprise 23 core loci: SE33, Penta D & E, 13 CODIS and 7 non-CODIS European Standard Set STRs, plus supplementary STRs in the recently released Promega CS-7™ and Qiagen Investigator HDplex™ kits. Also included were D9S1120, a marker we developed for forensic use unique to chromosome 9, and the novel D6S1043 component STR of SinoFiler™ (Applied Biosystems). The data collated provides reliable estimates of recombination rates between each STR pair, that can then be placed into haplotype frequency calculators for short pedigrees with multiple meiotic inputs and which just requires the addition of allele frequencies. This allows all current STR sets or their combinations to be used in supplemented paternity analyses without the need for further adjustment for physical linkage. The detailed analysis of recombination rates made for autosomal forensic STRs was extended to the more than 50 X chromosome STRs established or in development for complex kinship analyses.

1. Introduction

A common strategy in complex paternity analyses is to include additional short tandem repeat (STR) markers to provide the fullest possible array of polymorphic data with which to make an inference of relatedness. The same strategy is often applied to the testing of immigration claims and analysis of missing persons where pedigrees may be deficient or involve a level of population substructure/consanguinity, necessitating extra STRs. Two supplementary STR sets have recently been commercially released to help expand marker choice in such situations, these consist of the 7-plex Promega CS-7™ set with five novel STRs: LPL, F13B, FESFPS, F13A01, Penta C, and the 12-plex Qiagen Investigator HDplex™ set with nine novel STRs: D2S1360, D3S1744, D4S2366, D5S2500, D6S474, D7S1517, D8S1132, D10S2325, D21S2055 [1]. When these supplementary STRs are added to the 23 core forensic identification markers (SE33, Pentas D/E, 13 CODIS, 7 non-CODIS European Standard Set) the co-location of multiple STRs on the same chromosome (syntenic loci) becomes a certainty. Therefore it may be necessary to make statistical allowance for physical linkage between extended STR sets when used to analyse close relatives. In these cases recombination has less chance to disrupt the haplotypes of syntenic pair alleles, so they can be present at a different frequency in the pedigree to the population as a whole.

Recent concerns about handling genotype data obtained from the two closely sited STRs of vWA and D12S391 have raised the issue of how to measure and allow for physical linkage between autosomal forensic markers [2,3]. The first recourse for most analysts introducing new STRs is to refer to family studies that estimate recombination rates or aim to detect linkage disequilibrium (LD)—where loci may be associated through other factors besides physical proximity. However, such two-locus family

* Corresponding author.
E-mail address: c.phillips@mac.com (C. Phillips).

1872-4973/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2011.07.012

studies or pedigree methods have a number of shortcomings that make them inefficient for detecting both the true underlying recombination rate and LD. Firstly, a large number of meioses are needed to reliably detect recombination or LD—such studies therefore represent a major investment of money, time and effort. Secondly, unless both STRs have above-average polymorphism levels, a large proportion of pedigrees will be uninformative for recombination, with one or both parental genotypes homozygous or showing identical alleles. This factor can work the other way: a large number of rare alleles in an STR fails to provide enough examples of the rarer diplotype combinations from which to properly assess whether there is evidence of LD [4]. Thirdly, it is difficult to reliably ascertain the phase of the genotype pairs particularly when most or all pedigrees are short (i.e. involving just one or two generations), so phase often needs to be assumed and the tests lack accuracy.

For the last 6 years the genomic data of the HapMap Project [5] has provided an alternative to in-house genetic mapping, available as a freely accessible resource for the scientific community. As well as cataloguing common single nucleotide polymorphism (SNP) variation and mapping haplotype block characteristics in the human genome, HapMap has made detailed analyses of fine-scale recombination rates using SNP data at very high densities. An important point of distinction between the HapMap genome-wide recombination map and previous small-scale pedigree method maps is that HapMap's has been constructed in a completely different way by analyzing pairwise comparisons of large arrays of SNPs and the coalescent model [6–8]. Multi-point SNP data at very high densities allows the genetic distances of any one chromosome segment to be assessed at kilobase (kb) scales, while phase can be reliably inferred. Furthermore averaging across three populations reduces bias in the genome-wide estimates obtained that might arise from local variation or differences in effective population size. Most importantly, as HapMap SNP data is at four to five orders of magnitude higher density than conventional linkage marker sets, the multi-point analyses made allows a precise measurement of recombination rates from the genetic distances recorded.

We used the most recent HapMap data release 22 [9] to measure and compare centiMorgan (cM) genetic distances, and by inference fine-scale recombination rates, between 29 autosomal syntenic STR pairs identified from a total of 39 validated STRs currently available for forensic applications. The list of 39 loci includes two additional STRs: D9S1120, a marker we developed for forensic use [10] that shows less variation than most core STRs but prior to release of Penta C in CS7, was unique to chromosome 9, plus; D6S1043, a component of the Asian-informative Applied Biosystems (AB) SinoFiler™ multiplex [11]. The data we have collated provides reliable estimates of recombination rates between each STR pair that can be placed into a standard haplotype frequency calculator for short pedigrees and only requires the addition of allele frequencies. This allows all current STR sets or their combinations to be used in supplemented paternity analyses without the need for further adjustment for physical linkage.

Lastly, we have included an identical analysis of the genetic distances found on the X chromosome, which carries the largest set of syntenic forensic STRs, as well as showing a wider variation in local recombination rates than equivalent sized autosomes.

2. Materials and methods

2.1. Autosomal forensic STRs

NIST STRbase [12] provides detailed data on nearly all the 39 STRs we identified as forensic autosomal loci available in validated format and applicable to relationship testing. Relevant genomic data for each STR is detailed in Table 1. Only one STR chosen for analysis: D9S1120, is not part of a commercial multiplex kit, but this marker is used in two author laboratories (MVL and DSC), while sequenced ladders, population data and characterized repeat unit variation have been published [10].

Genome build 37.1 (GRCh37) was used to locate the chromosome positions of each STR, SNP or primer sequence. Build 37.1 is directly comparable to build 36 used to construct the HapMap

Table 1
Summary chart of the 39 forensic autosomal STRs studied. The 23 core STRs are listed consecutively by position on the left. STR locus identifier rs-numbers recently assigned by dbSNP are shown in normal type. When these have not yet been given, we identified proximal SNPs listed as repeat unit indels, e.g. [GATA/−] or by examination of closely sited substitution SNPs, both shown in italics.

Chrom.	Core STRs	dbSNP rs-number identifier for STR	Chrom.	Supplementary STRs	Kit	dbSNP rs-number identifier for STR	Chrom.	Supplementary STRs	Kit	dbSNP rs-number identifier for STR
C1	D1S1656	rs113633160	C1	F13B	Promega CS7	*rs10643350*				
C2	TPOX	rs113475620	C2	D2S1360	Qiagen HD-plex	rs113680434				
C2	D2S1338	rs112111672								
C2	D2S441	*rs10203882*								
C3	D3S1358	rs111694514	C3	D3S1744	Qiagen HD-plex	rs113865588				
C4	FGA	*rs67296980*	C4	D4S2366	Qiagen HD-plex	rs113820309				
C5	D5S818	rs112497490	C5	D5S2500	Qiagen HD-plex	rs111362704				
C5	CSF1PO	rs113729910					C6	D6S1043	AB SinoFiler	rs111544865
C6	SE33	*rs71021371*	C6	D6S474	Qiagen HD-plex	rs113991233	C6	F13A01	Promega CS7	*rs71817584*
C7	D7S820	rs112714641	C7	D7S1517	Qiagen HD-plex	rs112397288				
C8	D8S1179	*rs67563232*	C8	D8S1132	Qiagen HD-plex	*rs71307053*	C8	LPL	Promega CS7	*rs10558335*
			C9	Penta C	Promega CS7	*rs72398274*	C9	D9S1120	in-house	rs112358545
C10	D10S1248	rs113518246	C10	D10S2325	Qiagen HD-plex	no SNPs found				
C11	TH01	*rs71029110*								
C12	D12S391	rs113002069								
C12	vWA	*rs10579907*								
C13	D13S317	rs111980288								
C15	Penta E	*rs8036258*	C15	FES-FPS	Promega CS7	*rs6229*				
C16	D16S539	rs112689398								
C18	D18S51	*rs10560567*								
C19	D19S433	rs113951851								
C21	D21S11	rs113145752	C21	D21S2055	Qiagen HD-plex	rs113225349				
C21	Penta D	*rs7279663*								
C22	D22S1045	rs112790319								