Forensic Science International: Genetics 29 (2017) 29–37

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Evaluating 130 microhaplotypes across a global set of 83 populations

Kenneth K. Kidd[a,*], William C. Speed[a], Andrew J. Pakstis[a], Daniele S. Podini[b], Robert Lagacé[c], Joseph Chang[c], Sharon Wootton[c], Eva Haigh[a], Usha Soundararajan[a]

[a] Department of Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, CT, 06520-8005, USA
[b] Department of Forensic Sciences, The George Washington University, 2100 Foxhall Road, NW, Washington, D.C., 20007, USA
[c] Human Identification Group, ThermoFisher Scientific, 180 Oyster Point Blvd, South San Francisco, CA, 94080, USA

ARTICLE INFO

Article history:
Received 2 November 2016
Received in revised form 17 February 2017
Accepted 12 March 2017
Available online 16 March 2017

Keywords:
Microhaplotype
SNP
Ancestry
Massively parallel sequencing

ABSTRACT

Today the primary DNA markers used in forensics are short tandem repeat (STR) polymorphisms (STRPs), initially selected because they are highly polymorphic. However, the increasingly common need to deal with samples with a mixture of DNA from two or more individuals sometimes is complicated by the inherent stutter involved with PCR amplification, especially in strongly unbalanced mixtures when the minor component coincides with the stutter range of the major component. Also, the STRPs in use provide little evidence of ancestry of a single source sample beyond broad "continental" resolution. Methodologies for analyzing DNA have become much more powerful in recent years. Massively parallel sequencing (MPS) is a new method being considered for routine use in forensics. Primarily to aid in mixture deconvolution and avoid the issue of stutter, we have begun to investigate a new type of forensic marker, microhaplotype loci, that will provide useful information on mixtures of DNA and on ancestry when typed using massively parallel sequencing (MPS). We have identified 130 loci and estimated their haplotype (allele) frequencies in 83 different population samples. Many of these loci are shown to be highly informative for individual identification and for mixture identification and deconvolution.
© 2017 The Authors. Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The ability to "match" the multi-locus genotype of a crime scene sample directly to an individual or indirectly to an existing profile in a database has revolutionized the genetic aspect of forensics. In recent years the field has adopted a set of multi-allelic short tandem repeat (STR) polymorphic (STRP) loci. These markers and procedures have worked well over the last few years for matching evidence to suspect and database profiles. However, the use of capillary electrophoresis (CE) has imposed some limitations, especially in terms of the number of markers that could be studied simultaneously. In addition, the ability to infer biogeographic ancestry from the STRP data is limited [1]. Mixture deconvolution is an aspect of forensic practice that has become a more important part of forensic practice and can be very problematic with the standard STR markers [2]. Panels of individual SNPs have been studied by many researchers (cf. [3]) and some are capable of biogeographic ancestry inference of up to 8 to 10 distinct global regions [4,5]. While those panels of SNPs can provide very small random match probabilities (RMP) as low as or lower than the standard STR markers, they are poor at identifying mixtures and the added ancestry information has not been of high enough value in forensic practice to overcome the absence of the offender databases that exist for the STR markers. Moreover, a different methodology has been required to use the SNP panels in a forensic laboratory; that has been a significant cost impediment in equipment and training. Now, a new methodology is available and being implemented in forensic laboratories: massively parallel sequencing (MPS) [6–9]. This methodology allows simultaneous multiplexing of the existing STR markers in the offender databases along with large numbers of ancestry informative SNPs (AISNPs) and insertion-deletion polymorphisms (InDels). MPS also allows a new type of genetic marker, microhaplotypes [10–12], to be included in multiplexes with the other types of markers.

A microhaplotype locus (microhap) is defined by at least two single nucleotide polymorphisms (SNPs) within the length of a sequence read [10] and the expectation of a very low recombination rate. The alleles at the locus are defined as the haplotypes comprised, at the defining SNPs, of the specific alleles seen on the chromosomes in the population. Table 1 Illustrates that several different types of DNA markers can be used in forensics and the different detection methods that can be used. Clearly, some

* Corresponding author.
E-mail address: kenneth.kidd@yale.edu (K.K. Kidd).

http://dx.doi.org/10.1016/j.fsigen.2017.03.014
1872-4973/© 2017 The Authors. Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Table 1
Comparisons of the abilities of typing methods to genotype different types of DNA markers. Only MPS is able to genotype the three general types of markers and identify variation that is cryptic when other methods are used.

	Method used:	Capillary Electrophoresis	Chip Typing	MPS
Type of Marker:	STRP	Yes	No	Yes
	InDel	Yes	Limited	Yes
	SNP	Limited	Yes	Yes
	Microhap	No	No	Yes
	Cryptic variation	No	No	Yes

methods are not appropriate for all types of DNA markers. Only MPS is able to genotype all types of markers. Moreover, for microhaplotypes MPS will directly yield the phase, i.e., the *cis/trans* relationship between the individual SNP alleles.

We have presented the concept of microhaplotypes (microhaps) and described their values in forensics and related human population studies [10–13]. Microhaps become a very promising new type of useful forensic marker when MPS is the technology being used for marker genotyping. We have undertaken to search for microhaplotypes that are particularly informative with respect to two different criteria for which we think microhaplotypes will be particularly informative: 1) mixture detection and deconvolution and 2) identification of close biological relationships. We are now presenting the definitions and initial characteristics of 130 microhaps that we have characterized on a set of 57 populations in our lab (2611 individuals) and on the 1000 Genomes Phase 3 data (1000 Genomes Consortium [14])–for a total of 83 populations with 5115 individuals.

2. Methods

2.1. Selection of candidate loci

Our objective was to identify microhaps that would be good candidates for conversion to MPS for forensic applications. As explained in Kidd et al. [11], many sources were used to identify candidate loci: our own data on several regions previously studied in detail, the HGDP data, HapMap data, publications in the literature, etc. We found most sources provided little information on global levels of polymorphism and/or on linkage disequilibrium among SNPs within roughly 200 bp or less. TaqMan was an obvious method to determine in our population resources whether an initial candidate was worth further consideration. Our preference was for candidates with haplotypes likely to be polymorphic globally. However, that was not a criterion that could usually be applied a priori given the limited information available for some of the initial candidates. Most of the data collected for this paper were collected before the 1000 Genomes data were available.

2.2. Genotyping

The DNA used for this study was purified from lymphoblastoid cell lines that had previously been established for each individual. The individuals were typed for the individual SNPs using TaqMan assays available in the Applied Biosystems Assays on Demand catalog. Typing was done in 3 μl reactions in 384-well plates using the manufacturer's protocol. Following PCR the plates were read by an AB7900 with the SDS software. Failed reactions were repeated once. In general, data were complete for >96% of individuals for each of the 359 SNPs (on overall average the data are 98.9% complete).

2.3. Population samples

The population samples reported on in this paper include those for which the Kidd Lab has cell lines that were established over many years starting in 1984. The populations and numbers of individuals genotyped in our laboratory are listed in Supplemental Data Table S1. For the purposes of this paper we have also extracted the corresponding data for all 359 SNPs from the individuals in the 26 populations studied by the 1000 Genomes Consortium and phased (see below) them into the 130 microhaps. The locations of the 83 populations analyzed are plotted in Fig. 1. All of the individuals studied in the Kidd Lab were sampled with informed consent for studies such as this and are completely anonymous; data on them has been presented in many previously published studies.

2.4. Data analyses

Most of the computer programs utilized for our analyses are those described in earlier papers on minihaplotypes [15] and microhaps [11,13]. We elaborate here on various additions and modifications.

2.4.1. Phasing

Individuals missing any SNP typings for a particular microhap were not included in the phasing. Phasing used the program PHASE version 2.1.1 [16,17]. Individuals lacking phased genotypes for more than 20% of the 130 microhaplotypes were excluded from the Structure analyses, leaving a total of 5115 individuals.

As we emphasized in our previous paper on microhaplotypes [11] even when SNPs are typed separately, as in our case, a genotype for a haplotype can be known unambiguously if either all SNPs are homozygous or only one is heterozygous based on the minimal assumption of a co-dominant genetic system. In the current dataset, many more of the microhaps involve more SNPs and more haplotypes than in that dataset. Nonetheless, the alternative genotypes when two or more of the SNPs are heterozygous can often be resolved with near to complete certainty because of the moderately strong linkage disequilibrium present among the SNPs that limits the number of haplotypes. When one of the haplotypes that would be required for an alternative genotype was absent from the global sample the likelihood of that genotype is extremely low. Proof that a haplotype is absent is strictly impossible; the assumption is that a haplotype is absent if there is no evidence that it is present, i.e., that no genotype requires that haplotype to be present. When there are only a few different haplotypes at a locus, the proportion of resolvable genotypes can be very high. That is the case for most of the loci we are analyzing in this study. Thus, we consider the haplotype estimates to be quite accurate for each individual and very accurate for population frequencies.

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
✓ امکان دانلود نسخه ترجمه شده مقالات
✓ پذیرش سفارش ترجمه تخصصی
✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
✓ دانلود فوری مقاله پس از پرداخت آنلاین
✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات