Research article

Multiplexed SNP detection panels for human identification

Rixun Fang [a], Andrew J. Pakstis [b], Fiona Hyland [a], David Wang [a], Jaiprakash Shewale [a], Judith R. Kidd [b], Kenneth K. Kidd [b], Manohar R. Furtado [a,*]

[a] Applied Markets, Applied Biosystems/Life Technologies, 850 Lincoln Centre Dr, m/s 402, Foster City, CA 94404, United States
[b] Genetics Dept. Yale University School of Medicine, New Haven, CT 06520, United States

ARTICLE INFO

ABSTRACT

SNP profiling is a very powerful tool for human identification. Two panels of 49 and 41 SNPs were selected based on high average heterozygosity (>0.4) and low global Fst values (<0.06) as a set of highly discriminative SNPs suitable for human identification in all geographic populations. SNPs were selected based on testing samples from 44 populations across the globe using TaqMan® allelic discrimination formats. Coding SNPs, and SNPs with known functional or phenotypic manifestations were excluded. To test the performance characteristics of the two panels of SNPs and to compare their utility to STR analysis for human identification and paternity testing, we genotyped a panel of 41 individuals from three different CEPH families spanning three generations. The test samples were genotyped using the two SNP panels and a 15-loci STR kit. Further, utility of the SNP panels in human ID testing and related applications was demonstrated using degraded DNA and DNA from blood, semen and saliva samples. The development of highly multiplexed SNP detection systems enabling lower cost and higher throughput via automation will result in increased use of SNP profiling in applications like human identification, cell line authentication, species identification and bacterial strain typing.

© 2009 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

A panel of 92 SNPs was assembled for human identification after testing more than 500 potential candidate SNPs from the list available since January 2009 at the site (http://info.med.yale.edu/genetics/kkidd/contents.html). The SNPs were selected such that they have average heterozygosities of >0.4 and Fst values <0.06 on the 44 population samples (average 56 individuals per group) representing the major regions of the world [1,2]. The SNP list includes 45 essentially unlinked SNPs distributed across the human autosomes; these SNPS are an excellent panel for individual identification providing high match probabilities in the range of values comparable to those for the 13 CODIS STR markers. The unlinked status of these 45 SNPs also make them useful for resolving close biological relationships. No meaningful departures from Hardy–Weinberg ratios were seen for any SNP in the populations tested. Separating out 6 SNPs exhibiting some linkage disequilibrium (LD) left 86 SNPs in the panel with no significant pair-wise LD. This marker set is expected to have essentially the same characteristics useful for individual identification in other large human populations. In this report we describe multiplexing of 90 SNPs in two pools using the GenPlex™ assay platform [3,4] along with the amelogenin gender marker. Automation of this

format and studies to determine accuracy have been reported recently [5,6].

2. Materials and methods

40 DNA samples from three extended CEPH Utah families (1333, 1340 and 1345) each comprising three generations of European origin and sibships of 9, 4 and 7 offspring respectively were obtained from Coriell Institute for Medical Research (Camden, NJ, USA). Forensic type samples and body fluids were from Serological Research Institute (Richmond, CA) and Biochain (Hayward, CA). Degraded DNA samples were generated in the laboratory using sonication followed by digestion with 2–6 units of DNase I. The GenPlex™ assay core reagents, the amplicon specific amplification primers for PCR and the SNP specific probes for the OLA and TaqMan® allelic discrimination assays (http://www.all-genes.com) were obtained from Applied Biosystems [3].

PCR designs were optimized such that the amplicon lengths were in the range of 65–111 bp. The assay design eliminated the need for Exonuclease-SAP treatments. The details of the GenPlex™ typing system are as described previously [3–6]. TaqMan® allelic discrimination assays for 70 of the SNPs were used to test for concordance.

3. Results and discussion

Complete concordance was observed between results from the GenPlex™ assay format compared to the single-plex TaqMan®

* Corresponding author. Tel.: +1 650 554 2670; fax: +1 650 554 2774.
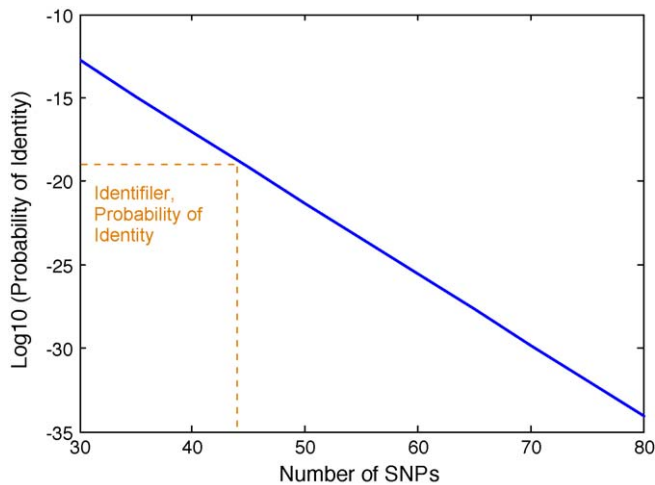E-mail address: manohar.furtado@lifetech.com (M.R. Furtado).

Fig. 1. Probability of identity with increasing number of SNPs obtained using approximately 42 SNPs is equivalent to that from a 15-STR loci kit like IdentiFiler®.

assay format when the 41 DNA samples were typed. The Log_{10} probability of identity increases with the increase in the number of SNPs typed, approaching estimated values of 10^{-35} when about 80 SNPs are included (Fig. 1). The probability of identity obtained with the 15-STR loci present in the Identifiler® kit is comparable to that observed with about 42 SNPs. Pair-wise analysis of the 40 CEPH family members analyzed using 80 SNPs gave a mean number of SNP loci with different genotypes of 44.76 with a standard deviation of ± 6.7. The minimum number of difference between pairs of samples, in related individuals (full-sibs) was around 18 and the maximum number of differences was 64. The mean value for the Identifiler® kit was 14.3 ± 1.2 with a maximum of 10.9 and a minimum of 7 loci different [3].

The SNP and STR profiles obtained using the 49 SNP panel and the Identifiler®, respectively, for the degraded DNA was evaluated. As expected, the alleles with larger amplicon size within the Identifiler® kit failed to amplify when the degraded DNA samples were used. The DNA samples degraded using 5 U of DNAse I provided results for 4 out of 15 loci (alleles with amplitude >50 RFU). The SNP panel, on the other hand, provided a complete profile for this sample. Thus, the short amplicon sizes within the

SNP set enable the generation of informative profiles with degraded DNA [7]. Additionally, the SNP panels were used to type the DNA obtained from various body fluids like blood, saliva and semen, and with DNA from cell lines with good results (data not shown).

4. Conclusions

Carefully selected SNP panels, based on high heterozygosity and low Fst values across populations, have the potential to provide informative profiles, especially from degraded DNA samples. The availability of larger panels of SNPs for genotyping can provide very high power of discrimination. SNP markers provide additional advantage in relationship analysis and samples containing degraded DNA.

Conflict of interest

Acknowledgements

References

[1] A.J. Pakstis, et al., Candidate SNPs for a universal individual identification panel, Hum. Genet. 121 (2007) 305–317.
[2] K.K. Kidd, et al., Developing a SNP panel for forensic identification of individuals, Forensic Sci. Intl. 164 (2006) 20–32.
[3] C. Phillips, et al., Evaluation of the GenPlex SNP typing system and a 49plex forensic marker panel, Forensic Sci. Intl.: Genet. 1 (2007) 180–185.
[4] J.J. Sanchez, et al., A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 27 (2006) 1713–1724.
[5] C. Tomas, et al., Typing of 48 autosomal SNPs and amelogenin with GenPlex SNP genotyping system in forensic genetics, Forensic Sci. Intl. Genet. 3 (2008) 1–6.
[6] M. Stangegaard, et al., Biomek-3000 and GenPlex SNP genotyping in Forensic Genetics, JALA 13 (2008) 297–303.
[7] E. Musgrave-Brown, et al., Forensic validation of the Genplex SNP typing system—results of an inter-laboratory study, Forensic Sci. Intl. Genet. (Suppl. 1) (2008) 389–393.