Forensic Science International: Genetics 38 (2019) 54–69

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Microhaplotypes in forensic genetics

Fabio Oldoni[a], Kenneth K. Kidd[b], Daniele Podini[a,*]

[a] Department of Forensic Sciences, The George Washington University, 2100 Foxhall Road NW, Washington, DC, 20007, United States
[b] Yale University School of Medicine, Department of Genetics, 333 Cedar Street, New Haven, CT, 06520, United States

ARTICLE INFO

Keywords:
Microhaplotypes
Single nucleotide polymorphisms
Massively parallel sequencing
Biogeographic ancestry inference
Mixture deconvolution
Human identification
Missing person and relationship identification
Probabilistic genotyping
Clinical application
Non-human DNA

ABSTRACT

Microhaplotype loci (microhaps, MHs) are a novel type of molecular marker of less than 300 nucleotides, defined by two or more closely linked SNPs associated in multiple allelic combinations. The value of these markers is enhanced by massively parallel sequencing (MPS), which allows the sequencing of both parental haplotypes at each of the many multiplexed loci. This review describes the features of these multi-SNP markers and documents their value in forensic genetics, focusing on individualization, biogeographic ancestry inference, and mixture deconvolution. Foreseeable applications also include missing person identification, relationship testing, and medical diagnostic applications. The technique is not restricted to humans.

1. Introduction: historical background of haplotype discovery

1.1. Discovery of haplotype blocks in the human genome

The term 'haplotype' was first introduced by Ruggero Ceppellini in the late 60 s to describe alleles within the human leucocyte antigen (HLA) region that are inherited together as a block [1]. Twenty years later the human genome project (HGP) launched an unprecedented international collaboration [2] foundational to the study of human genetics and biomedical research. The early work focused on mapping of human and mouse genes and sequencing the genomes of significantly smaller and easily studied organisms [3–6]. This was of paramount importance for the understanding of the hereditary architecture of diseases [7–9] and provides an essential scaffold for the assembly and annotation of the human genome. The publication that proclaimed the completion of the sequencing of the first draft of the human genome in 2001 [10,11], and completion of the full sequence in 2003 [12] allowed the investigation of polymorphisms anywhere in the human genome. In addition, the HGP contributed to the discovery that the human genome is found to have block structures of closely associated loci of variable length [13–15] and that there is a fine-scale structure of recombination "hot-and-cold spots", unevenly distributed across the highly organized genome [16].

1.2. Large-scale international haplotype map project

With the goal of establishing a resource for the study of global variation in the human genome, the HGDP-CEPH (Human Genome Diversity Panel - Centre d'Etude du Polymorphisme Humain) was established in 2002. At its core the HGDP-CEPH [17,18] is a collection of 1064 cultured lymphoblastoid cell lines from 52 populations of different parts of the world deposited at the Foundation Jean Dausset in Paris.

Later studies, aimed at understanding human evolution and population structure [19,20] focused on linkage disequilibrium (LD). Comprehensive large-scale LD studies identified distinct LD blocks, defined as haplotype blocks or specific genomic regions with a restricted number of haplotypes seemingly due to the limited number of ancestral recombination events within these regions [21,22]. With these considerations in mind, a large-scale Hap Map international project (*i.e.,* HapMap) was launched in 2002 with the main goal of providing the scientific community with a public resource for medical genetic research [23]. The project covered different exploratory phases [24] and was set to build a detailed haplotype map representative of the block-like pattern of the human genome. The purpose of this effort was to investigate the global distribution of LD blocks with regions of high and low haplotype diversity across geographically diverse populations [25]. One finding was that haplotype block structure varied significantly around the world [26]. Initial data generated from the HapMap project

* Corresponding author.
E-mail address: podini@gwu.edu (D. Podini).

https://doi.org/10.1016/j.fsigen.2018.09.009
Received 13 July 2018; Received in revised form 21 September 2018; Accepted 25 September 2018
Available online 01 October 2018
1872-4973/

played a pivotal role in the establishment of pioneering methods for the design and development of genome-wide association studies (GWASs) that enabled the identification of novel loci implicated in numerous complex diseases. The HapMap project contributed to a much deeper understanding of the global distribution and cause of recombination "hot spots" in human population groups specifically selected to reflect worldwide human diversity patterns. This project also allowed researchers to systematically catalog allele frequencies and association patterns of adjacent genetic variations for several millions of single nucleotide polymorphisms (SNPs) across different continental populations [27].

1.3. Methods for defining haplotype block-like structures

Different methods were proposed to define block-like structures in the human genome [26,28]. Ge et al. [29] suggested identifying sets of SNPs in almost-to-full-complete LD, which results in a rather low level of heterozygosity among SNP combinations, making these haplotype systems uninformative for applications such as relationship testing and human identification. In contrast, Pakstis et al. [30] focused on haplotypes with high heterozygosity with potential value for familial identification. The main criterion was to find closely linked SNPs within defined low recombination regions with multiple haplotypes. This latter strategy for the discovery of haplotype systems was also used in other non-forensic oriented studies, which led the identification of gene segments possessing multiple allelic variations with either no history of recombination or evidence of rare historic recombination events [31–34].

1.4. 1000 Genomes project

A deep and systematic exploration of human variation demands comprehensive knowledge of sequence variation across the full variety of genetic variations and allele frequencies. Inspired by the success of the HGP, different consortiums were established to catalog common human genetic variations as a foundation for analyzing the association between genotype and phenotype. Amongst them, the 1000 Genomes Project was launched in 2008, comprising various exploratory and developmental phases aimed at establishing a public reference database for DNA polymorphisms in different human population groups [35]. Multiple approaches were used in this project, including whole-genome sequencing, deep targeted-exome sequencing, and dense microarray genotyping, applied to a diverse set of individuals from multiple continental populations. This is a valuable resource specifically developed to elucidate the properties and distribution of a broad spectrum of common and rare sequence variants with allele frequencies > 1% in the human genome. The 1000 Genomes Project identified different kinds of genetic polymorphisms, including bi-allelic and multi-allelic SNPs, short insertions/deletions (INDELs/DIPs), and structural variants (SVs) [36,37].

1.5. Mini-haplotypes: multi-SNP haplotype systems

The first evidence of haplotype use was for decrypting the evolutionary history of modern humans. Jin et al. [38] reported on a locus of 565-bp on chromosome 21 and near the *MX1* gene, which could provide an explanation on the geographic distribution of modern haplotypes, thus shedding light on the multiple prehistoric human migrations. The concept of multi-SNP haplotype systems for forensics and population studies was introduced by Pakstis et al. [30]. As originally defined, a miniature haplotype locus is a short genomic region (spanning < 10 kb) entailing three or more highly heterozygous SNPs. The overall rate of recombination events among the multiple SNPs (*i.e.,* within the extent of the haplotypes) distributed across < 10 kb recombination hot spot-free regions is predicted to be on the order of the mutation rate of SNPs, but significantly lower than the mutation rate for the conventional short tandem repeat polymorphisms (STRPs) used in forensics. Throughout this paper the term STRs refers to polymorphic short tandem repeats used in forensic. These haplotype systems (*i.e.,* minihaplotyes or minihaps) were first used in anthropological studies [39–41], but could likewise provide forensically relevant information on human identity (HID) and ancestry prediction. The utility of minihaps in ancestry inference is well-exemplified by loci such as the 3-SNP AAA haplotype at the *AGT* gene and 3-SNP GAA haplotype at the *PAH* gene, which were found most frequently in Europeans and East Asians, respectively [30]. Though a set of minihaps was originally characterized, the size of such segments represents a serious limitation for their application to forensic samples. The need to statistically estimate the actual haplotypes comprising the genotypes of the individuals adds an element of uncertainty that can be a problem in forensic applications such as familial/lineage identification and mixture detection/deconvolution.

1.6. Utility of bi-allelic SNPs in forensics

The recent advances in modern DNA technology have strengthened the idea that SNPs can be valuable complementary markers to the gold standard STRs. One example is the cheap and rapid DNA array technology that enables genotyping millions of SNPs at a time but cannot genotype STRs for forensics. SNPs are characterized by a lower mutation rate (10^{-8} - 10^{-9} per site per generation) [27,42] than STRs (10^{-3} per site per generation) and are particularly suitable for the analysis of highly degraded DNA samples due to the overall small size of amplicons [42]. In addition, large panels of carefully-selected bi-allelic SNPs can provide as much uniqueness as standard CODIS STRs [43–47]. In comparison to STRs [48], SNPs can also provide significant and detailed information on ancestry as shown by the numerous AISNP (ancestry-informative SNP) assays published in the last decade [49–61]. SNPs are also efficient markers for prediction of human phenotypic features (visible traits) such as eye, hair and skin colour [62–71]. While most SNPs are not suitable markers for mixture detection and deconvolution due to their bi-allelic nature, tri- and tetra-allelic SNPs [72,73] and haplotypes of SNPs are more promising markers because they are characterized by multiple alleles similar to the conventional highly polymorphic STRs.

2. Microhaplotypes: multi- SNP markers

2.1. Microhaplotypes: a new type of genetic marker

Previous investigations have illustrated the utility of multiple closely linked SNP-based markers in anthropology for population relationship and their capacity to provide a plausible explanation for the pattern of recent human variation [39,74–78]. In addition, multi-allelic SNPs have been promoted as suitable markers for addressing relevant forensic questions such as family/clan, lineage inference, and individual identification [30,79,80]. Aiming to complement current DNA typing tools for forensics and population genetics, the Kidd laboratory proposed a novel type of genetic marker named microhaplotypes (*i.e.,* microhaps or MHs) [80]. These are short segments of DNA (< 300 nucleotides, thus "micro"), characterized by the presence of two or more closely linked SNPs that present three or more allelic combinations (*i.e.,* "haplotypes"). The short distance between SNPs implies an extremely low recombination rate among them. The level of heterozygosity of the microhaps is dependent upon different factors, including historical accumulation of allelic variants at different positions within the targeted region, incidence of rare crossover events, occurrence of random genetic drift, and/or selection [81]. Since microhaps are multi-SNP haplotypes, they can provide, on a per locus basis, a larger assembly of information than a stand-alone SNP marker.

2.2. Transition from Sanger sequencing to massively parallel sequencing

Sanger sequencing represents the "gold" standard methodology for DNA sequencing [42]. However, the main limitation of this technology is the inability to determine the *cis/trans* relationships among the individual SNP alleles (*i.e.,* the phase) from genomic DNA when two or more sites are heterozygous [82]. High throughput DNA sequencing technologies, referred to as massively parallel sequencing (MPS), essentially obviate this obstacle by effectively cloning and sequencing each strand's allele combinations separately. MPS can quickly interrogate a large battery of SNPs and several other markers in parallel, with read lengths of up to 400 nucleotides. These features make MPS useful for addressing relevant forensic DNA questions in a single laboratory analysis [83]. This sequencing approach has also changed how forensic DNA analysis can be conceptualized. Hundreds of markers can be co-amplified from a sample and individually sequenced in parallel, whereas data generated using the Sanger approach is a representation of a multitude of amplicons from two alleles, in a single source sample, or from multiple alleles in a mixture of two or more contributors. Unlike Sanger sequencing, MPS can distinguish each parental microhap allele occurring at one specific locus by clonal sequencing of each individual amplicon originating from each DNA strand present in the sample, whether from a single source or mixture [84,85]. An example illustrating the haplotype organization of alleles within a microhap locus with a four-SNP site analysed by Sanger sequencing and MPS is depicted in Fig. 1. Sixteen different haplotypes, which essentially equate to alleles, are potentially observed, and a total of $[N*(N + 1)]/2 = 136$ different genotypes (N indicates the number of haplotypes) can exist, with eight of those also being heterozygous at all four sites.

2.3. Features of microhaplotype loci

Microhaplotypes for forensic applications are single-copy multi-SNP loci distributed across the entire genome and display some advantages over conventional STR markers. The absence of a short sequence tandemly repeated multiple times prevents the occurrence of polymerase slippage, which causes artefact fragments (*i.e.,* stutter peaks) that are one repeat shorter (more rarely longer) than the original allele at a lower copy number. Stutter fragments can complicate the analysis of imbalanced mixed-source samples, especially when they are of comparable height to the allelic peaks of the minor contributor. In a single source sample this does not represent an issue, as true alleles and stutter can be easily distinguished. In mixtures however, where the amount of DNA from the different contributors is very different, true alleles from the minor contributor/s can be in the same range as peaks generated from alleles of the major/s, increasing the complexity of the interpretation. Microhaps do not generate stutter peaks. Another advantage of microhaps over STRs is that all the alleles within a locus have the same length. STR alleles within certain loci may vary in length up to 100 nucleotides. If the DNA is degraded or if the sample contains PCR inhibitors, the shorter alleles may be preferentially amplified and overrepresented compared to the longer alleles from the same contributor, complicating allele source attribution in mixtures. Another advantage of microhaps over STRs, particularly in relationship testing, is their lower mutation rate, approximately five to six orders of magnitude less [81]. A disadvantage of microhaps compared to STRs is that they generally have fewer alleles, and consequently a larger panel of markers is required to obtain a comparable power of discrimination. In addition, allele frequencies vary significantly across populations; thus, population-specific allele frequencies are necessary when calculating the random match probability (RMP) or the likelihood ratio (LR) in forensic cases as these values will consequently change across populations. One additional drawback relates to the laborious and time-consuming MPS workflow. Although relatively inexpensive when calculated per nucleotide, MPS is still expensive (compared to CE-based methods) and requires complex pipelines for assembly and analysis.

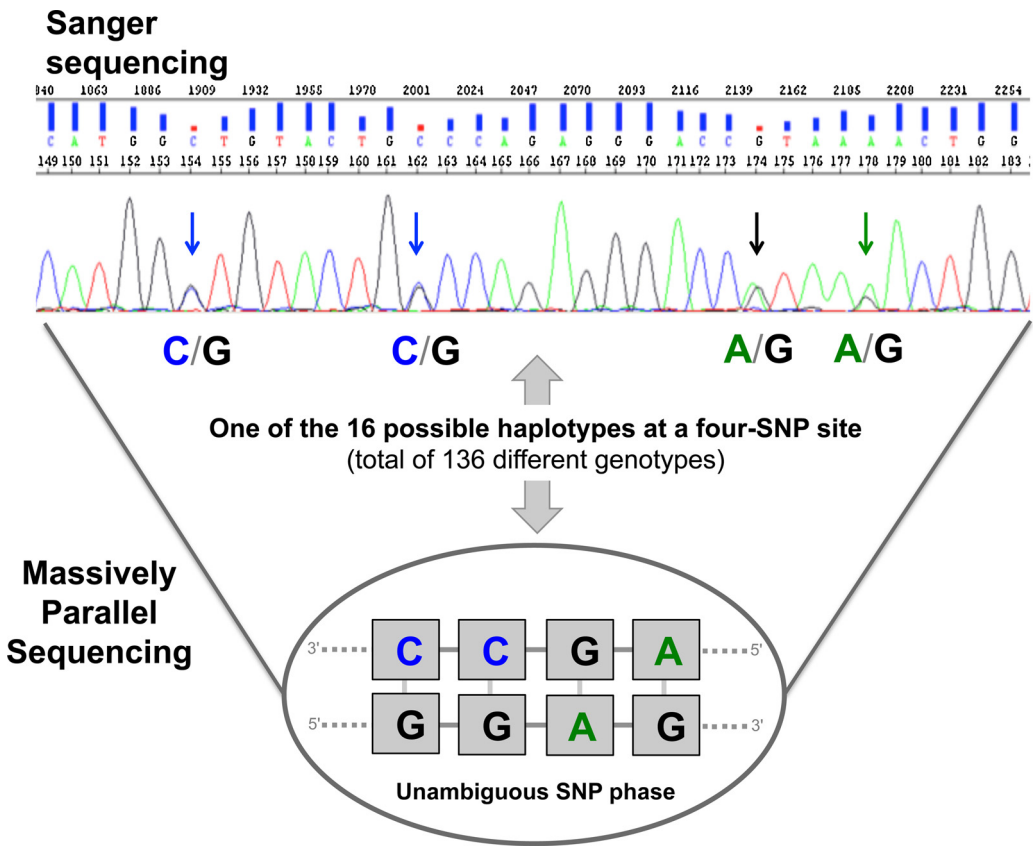

Fig. 1. Sanger sequencing *vs*. massively parallel sequencing: an example of haplotype organization of alleles within a microhaplotype locus on a chromosome.

This may reduce the rapid implementation of microhaps within the global forensic DNA community.

2.4. ALFRED: a repository of global allele frequency data

The scientific community must not only know that genetic variants exist, but also their frequencies among populations. Establishing a database of population variation, gene or allele frequencies for geographically well-defined populations is needed. Small or large genetic-oriented and specialized databases, such as dbSNP (SNPs) [86] and STRbase (forensic STRs) [87] were developed to addressing specific needs. Amongst others, ALFRED (ALlele FREquency Database) represents an extensively used domain designed for both research and education in different areas of human genetic diversity and molecular anthropology [88–90]. Launched in early 2000 and continuously supported over the years, initially by the U.S. National Science Foundation and recently by the NIJ, it is a valuable resource that meticulously catalogs data of gene frequency on human populations. ALFRED is a curated publicly accessible compilation of allele frequencies from a large variety of human nuclear DNA polymorphisms collected on a global scale of anthropologically defined populations. It is web-accessible through Kidd laboratory (https://medicine.yale.edu/lab/kidd/research/alfred/) or at https://alfred.med.yale.edu [91]. Each sample is identified using a unique identifier (UID) and related description. All data on SNP/marker panels are accessible through an additional open access teaching and web application resource named FROG-kb (Forensic Research/Reference on Genetics-knowledge base) that enables viewing, recovery and statistical calculations of SNPs of forensic interest [92,93]. To date allele frequency data on all relevant panels of SNPs/INDELs developed for different forensic purposes are available on ALFRED. These include numerous identification-informative SNPs/INDELs (IISNPs/IIINDELs), ancestry-informative SNPs/INDELs (AISNPs/AIINDELs) and phenotype-informative SNPs (PISNPs). Allele frequency data on 191 already characterized microhaplotypes are also accessible to the broad research community. The number of records in ALFRED is constantly growing as allele frequencies of newly genotyped markers and populations are added.

2.5. PHASE-based allele inference

Different SNP typing methods have been used for genotyping individuals and populations. Contrary to the MPS technology, most other methods (*i.e.,* TaqMan®, Sanger sequencing, micro-arrays, *etc.*) do not provide haplotype information (phase-known SNP data), which is essential to determine the microhap allele present in the sample. Accordingly, to be able to use SNP allele frequency databases developed with such genotyping methods (*e.g.,* ALFRED) as a source of microhap allele frequencies, statistical-based haplotype reconstruction must be performed. PHASE is a statistical tool that applies concepts from population genetics and coalescent theory to infer the likely patterns of haplotypes observed in natural populations. Notably, it utilizes a Bayesian-based approach that enables the application of *a priori* expectations to presumably reconstruct human haplotype patterns correctly using population genotype data [94–96]. PHASE also provides estimates of the uncertainty for each phase calling with the missing typing(s), included as unknown, and a highly accurate estimate of full haplotypes. This software also enables a sufficiently accurate reconstruction of haplotypes without the need of additional genotypes (*i.e.,* family members), though there is concern that phase calls may sometimes be incorrect. The specific sequence of each haplotype can be retrieved unequivocally when all SNPs of interest are homozygous; also when no more than one site is heterozygous it is possible to define both alleles irrespective of the allele frequencies. Since multi-SNPs of a microhap are within a short distance and usually in at least modest LD, most haplotypes at loci with two or more heterozygous SNPs can be accurately estimated. In contrast, MPS represents a valuable high-throughput alternative for reliable SNP typing [82], because it provides an accurate answer to the question related to the exact phasing of parental haplotypes of SNPs by enabling the direct clonal sequencing of single-DNA strand reads across the entire locus [84,85].

2.6. Microhaplotype nomenclature

It can be helpful to have a common general nomenclature for microhaplotypes that references that small segment of DNA, especially now that multiple laboratories are investigating microhaps [97–100]. It is common early in papers to define a short symbol or acronym that can be used throughout the remainder of the paper. Some of those, such as *SNP*, become generally accepted. In the case of microhaplotypes, having a common format for such symbols allows a general cataloging that, for example, distinguishes the names for such segments from gene names. A microhap is defined by the SNPs in the region using the rs numbers that are defined and cataloged in dbSNP but repetitive use of a string of rs numbers becomes tedious. Yet, the rs numbers are preferable to simply giving genomic coordinates because those coordinates change with each new genome build; the genomic coordinates of the SNPs are updated in dbSNP, so build changes need not be a concern. Based on guidelines of the HUGO Gene Nomenclature Committee [101] a microhaplotype nomenclature was proposed by Kidd [102]. The nomenclature was designed to allow each lab to choose the name for a new microhap but following a format that would allow easy reference by others and cataloging in databases. A nomenclature model is the early cataloging of DNA polymorphisms which used “D” (for DNA) followed by the chromosome number, a delimiter “S” (for site or sequence) and a consecutive catalog number assigned by the Human Gene Mapping DNA Committee to yield names such as D21S11 [103]. The proposed microhap nomenclature [102] starts with “mh” (for microhaplotype) followed by the chromosome number. That prefix is followed by a two or three letter abbreviation for the laboratory and a number that is unique in that lab for that microhap. As various labs are beginning to use this format, the microhaps are being cataloged in ALFRED (https://alfred.med.yale.edu) with the SNPs involved and haplotype frequencies available. For example, both mh05KK-022 and mh05CP-010 were independently named with evident locations on chromosome 5; the starting positions can be retrieved through ALFRED: 9,619,905 and 17,903,274 in build 37 (build 38 positions, 9,619,793 and 17,903,165, are in dbSNP).

2.7. Criteria for selecting microhaplotype loci

Different web-sources are available for screening candidate loci. These include, but are not limited to, HapMap and HGDP data, publications in the literature, and also 1000 Genomes data [104–106]. To address forensic and population genetic questions, different criteria can be applied to select microhaps optimal for different purposes. One criterion relates to the detection of at least three SNPs within a 200 bp single sequence that is currently a routine read length for MPS platforms [84]. The read length can be expanded to approximately 300 bp for certain loci to include more SNPs within the amplicon that increase the number of potential alleles. Though recombination “hot spots” within a 200 bp region can be practically ruled out, sporadic historical (ancient) recombination events will likely not affect the assumption of identity by descent (IBD) within a family. If an overall 1% recombination rate per Mb per locus and recombination hot spot-free loci are assumed, then a region < 200 bp significantly reduces the probability of recombination [81]. For selection of optimal microhap loci, two main desirable statistics are considered: effective number of alleles (A_e) and informativeness (I_n) [84]. The first metric that differentiates microhaps from “haploblocks”, as defined by Ge et al. [29], is described as the corresponding number of evenly frequent neutral alleles, based on population dynamics, that would produce the same heterozygosity as the locus with several alleles at different frequencies. This parameter

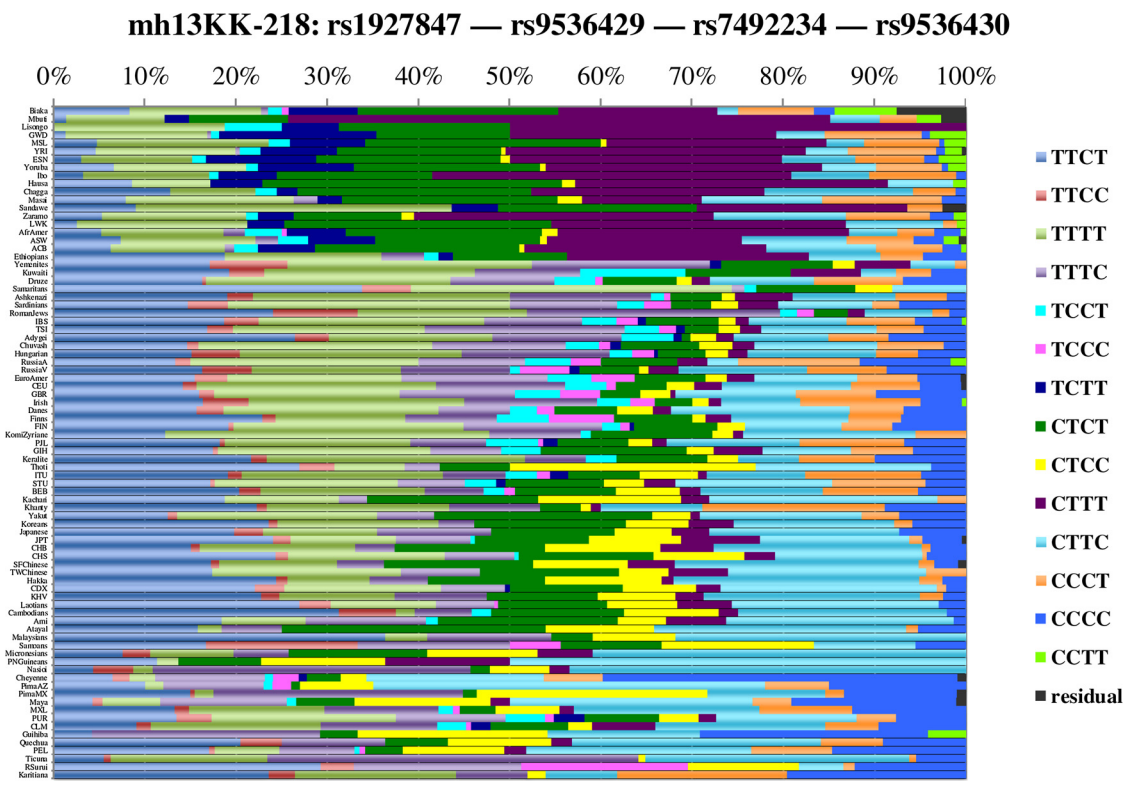

Fig. 2. Example of highly polymorphic MH locus on chromosome 13 showing multiple alleles per population (high A_e) and alleles that are very frequent in some populations and very rare in others (high I_n). For instance, allele CTTT accounts for an average of 30% of diversity in African populations while on average less than 5% in the rest of the world.

relates to the level of variation within a specific population and is associated with the marker's ability to distinguish individuals (*i.e.,* the random match probability or RMP). Therefore, larger A_e values correlate with higher capacity of a microhap locus to detect mixed-source DNA samples, deconvolute mixtures, and determine the number of contributors. Conversely, the second metric, also referred to as Rosenberg's measure of informativeness [107], provides a gauge of the variation of allele (haplotype) frequencies observed among the different populations tested. As a result, higher I_n values indicate that a microhap locus has alleles that are frequent in certain populations and rare in others. Microhaps with high I_n can help distinguish among (at least) some of the selected populations and predict the related biogeographic ancestry of a sample. The ideal locus will have both high A_e and I_n, meaning each population presents plenty of alleles and some/most/all of them vary in frequency among populations. An example of a highly polymorphic microhap locus is shown in Fig. 2

Current genome assemblies are based on short MPS reads, ranging between 50 and 150 bp, which are then reassembled bioinformatically. One of the major challenges associated with this process is the identification of short segments of DNA that occur two or more times in the genome. A potential pitfall of *in-vitro* forensic microhap selection is that multi-copy regions may be misclassified as single copy and selected as potentially effective forensic loci, which will then have to be discarded once *in-vivo* testing reveals their multi-copy nature [108–110].

3. Applications of microhaplotypes in forensic genetics

Microhaplotypes are a new promising and versatile tool that can be used to simultaneously extract many types of relevant forensic information from one or more individuals contributing to a DNA sample. The multi-faceted nature of microhaps makes them an appealing forensic marker for both DNA profiling and investigative genetics. Indeed, the advent of MPS has enhanced the value of microhaps beyond the original aim of identifying lineage-clan-family relationships by widening the spectrum of potential forensic applications [111]. This includes, but is not limited to, ancestry inference, mixture deconvolution (also through probabilistic genotyping (PG) software analysis), human, missing person, and relationship identification, non-human and clinical DNA analysis (Fig. 3). In the following sections we provide a comprehensive description of the current research on microhaps and an overview of additional foreseeable applications in forensic genetics.

3.1. Microhaplotype exploratory and discovery phases

The characterization and selection of first candidate SNPs for microhaplotypes were made possible by preexisting SNP genotype data on different densely packed genomic regions from over 2500 individuals from > 50 globally diverse populations collected over the last decades by the Kidd laboratory [81]. For optimal marker search, an average heterozygosity threshold of > 0.4 was arbitrarily chosen when screening the many different populations accessible on the worldwide web including the extensive Human Genome Diversity Project [112] and HapMap integrated three-phase datasets [24]. All Kidd multi-SNP loci developed thus far have been genotyped using the individual SNP typing TaqMan® assay and genotype data statistically phased into the corresponding haplotypes.

The first Kidd pilot study on microhaps included 31 loci [81]. This set has valuable features and includes 27 2-SNP and four 3-SNP loci for a total number of 66 SNPs distributed across 17 human chromosomes. The size range of this set was between 18 and 201 bp and the median heterozygosity level was between 0.40 and 0.63 for the 54 populations studied. These loci were selected for the high average level of heterozygosity required for kinship/lineage prediction, but not for high Fst among different populations needed for ancestry inference. The majority of loci are located on different autosomes or at least 95 Mb apart, thus making linkage disequilibrium between "adjacent" loci very

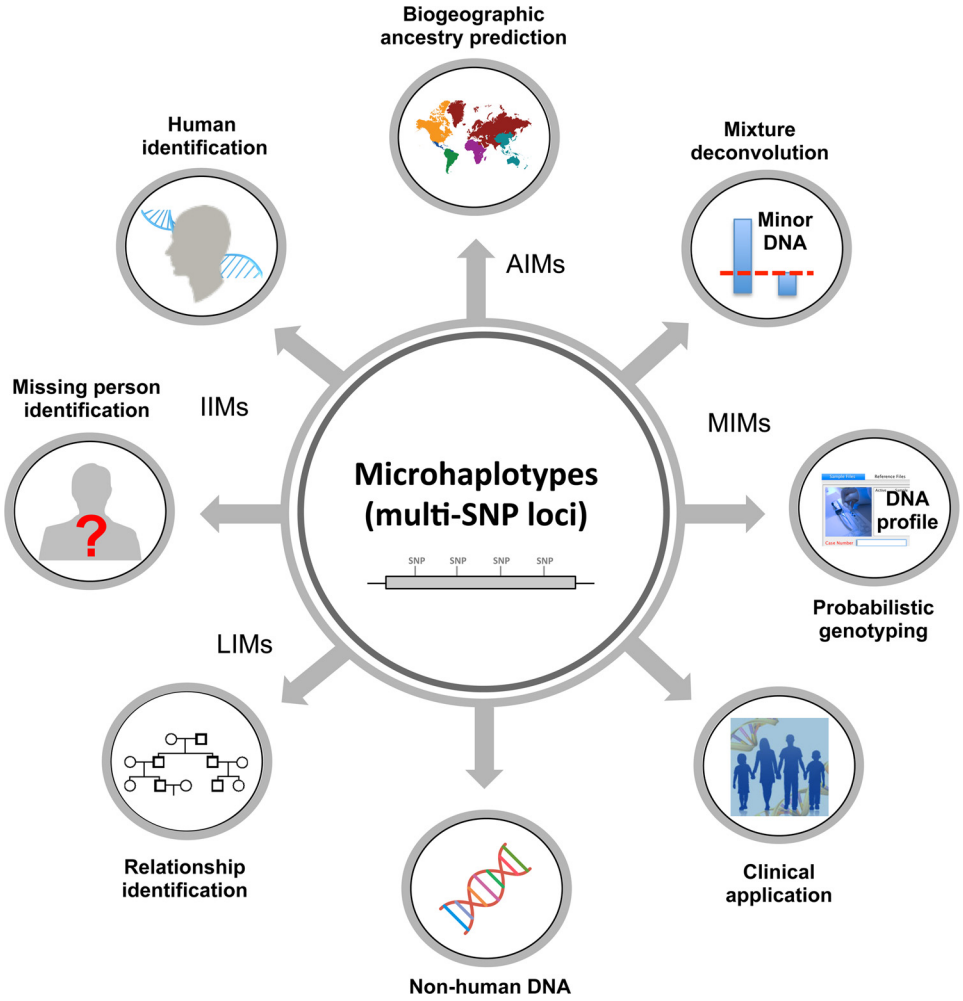

Fig. 3. An overview of current and potential applications of AIMs (ancestry informative markers), MIMs (mixture informative markers), IIMs (identification informative markers) and LIMs (lineage informative markers) in forensic genetics.

unlikely and independent segregation very likely. However, because 11 inter-microhap distances are shorter than 67 Mb, absence of independent segregation of these loci in families cannot be excluded. Overall this proof-of-concept set can be considered independent at the population level because negligible pairwise linkage disequilibrium values were observed for the 11 inter-microhap couplings of SNPs located on the same chromosome. To expand the number of microhaps useful to address forensic questions such as ancestry prediction, mixture deconvolution, and human and family-clan relationship identification, a larger panel of 130 microhap loci was further developed [82]. The microhaps in this panel are distributed across all 22 human autosomes with sizes ranging from 12 bp to 291 bp and include a total of 359 SNPs. This panel was evaluated on a set of 83 populations around the world. An additional Kidd subset of 65 loci selected from the panel of 130 loci and which includes eight new microhaps was developed [111]. This covers a total of 198 SNPs that increase the global characterization of microhaps for ancestry prediction. The 65-SNPs set was genotyped on a total of 96 populations (13 additional populations) and the allele's phase inferred to provide a broader picture of the global variation of allele frequency.

In addition to the pioneer Kidd panels, new sets of microhap loci have also been recently published by other research groups. Pu et al. developed a preliminary set of 50 loci identified from 1000 Genomes Phase 3 phased data for personal identification and ancestry inference [100]. This set shows between 3 and 10 SNPs per locus and between 4 and 47 theoretical alleles with an average heterozygosity of 0.67 over a total of 568 individuals from six global populations.

Of note, van der Gaag et al. [106] designed a multiplex set of 16 microhaplotype loci and implemented the assay on the Miseq® Sequencer® (Illumina). All candidate loci were searched in 1000 Genomes and Dutch Database (GoNL project samples) and only fragments spanning < 100 bp with six SNPs and MAF ≥ 0.1 in the relevant population were retained for testing and further development. The number of unique haplotypes observed was between 4 and 26.

Additionally, Chen et al. recently presented a new panel of 26 microhaplotypes [104]. The candidate loci were selected from 1000 Genomes Phase 3 (global population data from 832 individuals) based on different criteria that included short size (< 50 bp) and $A_e > 3$ among China (CHB) and Southern Han Chinese (CHS) population. 23 out of 26 loci included 3 SNPs and the observed allele number per locus was between 4 and 12. Sixty unrelated Chinese Han individuals were genotyped using the MH assay. The average A_e ranged from 2.54 to 3.98 and the heterozygosity from 0.38 to 0.81 among 9 populations of 26 microhaps. The length of sequences per locus ranged from 17 to 50 bp with a median value of 34 bp. Moreover, power of discrimination (PD) and power of exclusion (PE) values ranged from 0.76 to 0.87 and 0.10 to 0.63, respectively while the total probability of discrimination power (TDP) and cumulative probability of exclusion (CPE) was 0.999999999997 and 0.9997, respectively. Chen et al. also developed a set of 10 MH loci including eight 3-SNPs and two 4-SNPs for a total of 32 SNPs [105]. All selected loci containing at least four alleles (haplotypes) with frequency > 0.1 were tested on 2504 individuals from 26

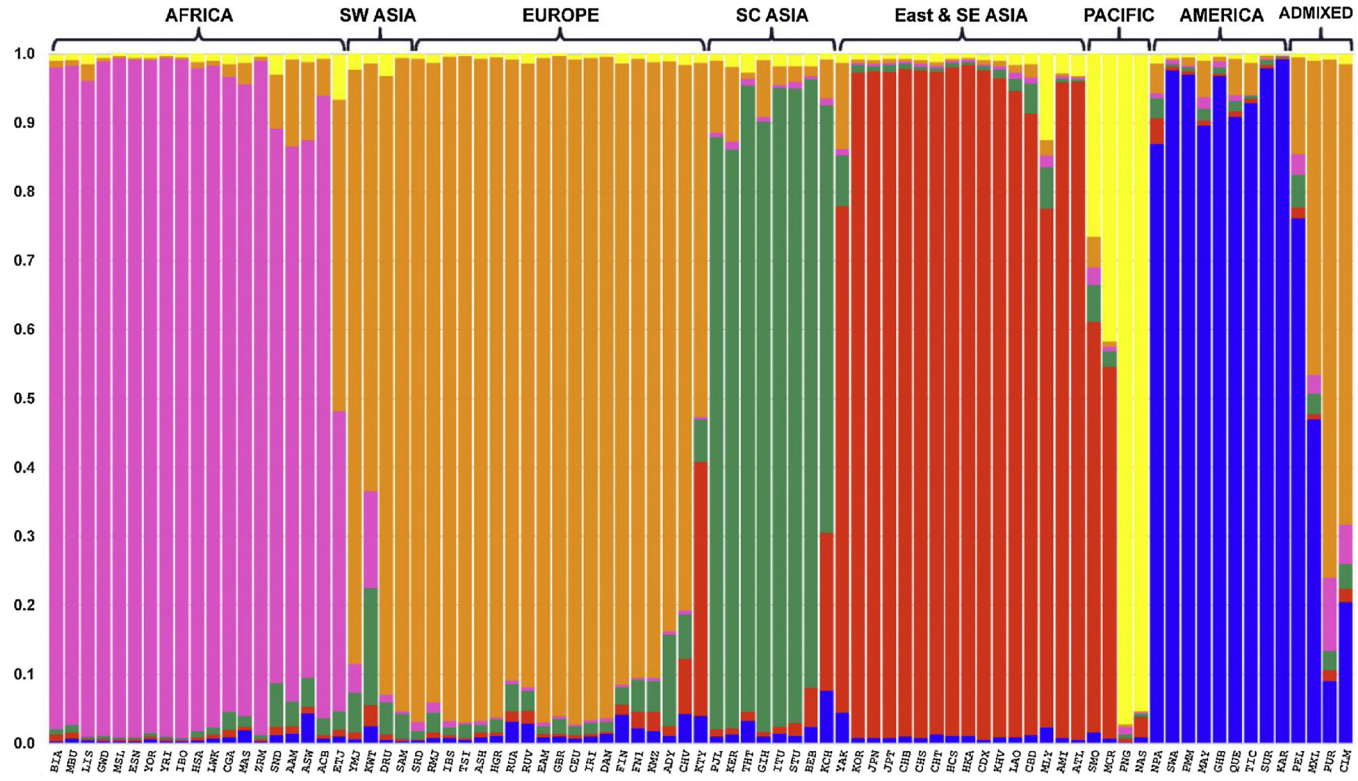

Fig. 4. An example of STRUCTURE analysis plot at K = 6 of > 5000 individuals across 83 global populations of 130 microhap loci (taken from [82]).

global populations from 1000 Genomes Phase 3. The marker size ranged between 5 and 48 bp with an average of 31.4 bp and therefore these loci are potentially suitable for the analysis of degraded DNA samples. The heterozygosity value of the 10 loci ranged from 0.2235 to 0.8958 with an average of 0.6593 while the PD from 0.3786 to 0.9242 with an average of 0.7944; the cumulative discrimination power (CDP) was between 0.99999976 and 0.999999998.

3.2. Microhaplotypes for mixture detection and deconvolution

MPS of STR alleles has been proposed as an option to address some of the issues with mixture deconvolution: (1) sequencing STR alleles enables individualization of iso-alleles (alleles that have the same length but different sequence and are indistinguishable *via* conventional CE); (2) it allows for all of the PCR primers in a multiplex to be designed to bind as closely as possible to the repeat region reducing loss of data when the DNA sample is degraded; (3) should the stutter of a major contributor have a different sequence than an overlapping true allele of a minor contributor they can be distinguished by sequencing but not by CE; (4) SNPs in the flanking sequence of the repeat region can further distinguish alleles that have the same number of repeats; (5) multiplex allows the simultaneous co-amplification of hundreds of markers, increasing the information that can be obtained from a sample [83,113–116]. Yet experimental data has shown the occurring of "poly-stuttering" of complex STR alleles, a single STR allele generating multiple types of stutter fragments [117], which can complicate interpretation. While stutter is more likely to happen in the longest uninterrupted sequence, it may occur from any part of the repeat pattern. Thus, a single allele from complex repeat structures such as D21S11 or D12S391, which tend to be the most polymorphic, can generate multiple stutter peaks. In presence of an imbalanced mixture where the minor contributor/s is/are in the stutter range of the major contributor, the sensitivity of MPS together with this 'poly-stuttering' of complex STR alleles may increase the complexity of mixture deconvolution.

Microhap amplification does not generate stutter peaks; as a consequence the presence of three or more different sequences (haplotypes) at sufficient read ratio becomes unambiguous evidence of DNA from more than one individual contributing to the mixed-source sample, and the number of haplotypes detected informs on the minimum number of contributors present in the mixture [81,111].

To evaluate the potential utility of microhaps in mixture resolution, theoretical analyses were performed using a finite number of alleles and the multinomial expansion formula for calculating the likelihood of a mixture from two random and unrelated individuals with at least three distinct alleles as a function of the related allele frequencies. The highest probability of alleles being present occurs when alleles within a locus are evenly frequent [84]. Therefore, unequally frequent alleles at a locus are converted into a value that corresponds to an integer of evenly frequent allelic variants. As a result, the potential of qualitatively identifying mixtures is translated into different integer A_e values with the cumulative probabilities of detecting mixtures reaching values higher than 95% with only five loci with average A_e values of 3. As a general rule, the criterion of minimum A_e value of > 3 often met by 3-SNP and 4-SNP loci is essential for determining the usefulness of microhaps in forensic practice. Since each locus is converted into the same "standard" form, higher A_e values [118] can better identify mixtures. Thus far, the utility in mixture deconvolution of microhaps typed by TaqMan® assay has only been evaluated theoretically. An initial evaluation was performed on the Kidd proof-of-concept set of 31 loci [81], which showed an overall average A_e range of 1.9-2.8. When considering the Kidd panel of 130 microhaps, 21 loci showed A_e values between 3.0 and 3.9, five loci between 4.0 and 4.9, and only two loci A_e > 5.0 [82] and probability of identifying a mixture at 99.99956%, 99.52% and 94.62%, respectively. Provided that 28 out of 130 loci, tested on 83 populations, with global average A_e > 3.0 are used, the theoretical probability of resolving a DNA mixture of two individuals is estimated at 99.99999857%. Based on the global average A_e values, 94 out of 130 microhaps have A_e values of 2.0 showing a higher level of heterozygosity than bi-allelic SNP loci. Albeit 38 microhaps have A_e value < 2.0 they show A_e values > 2.0 in several populations. Lastly,

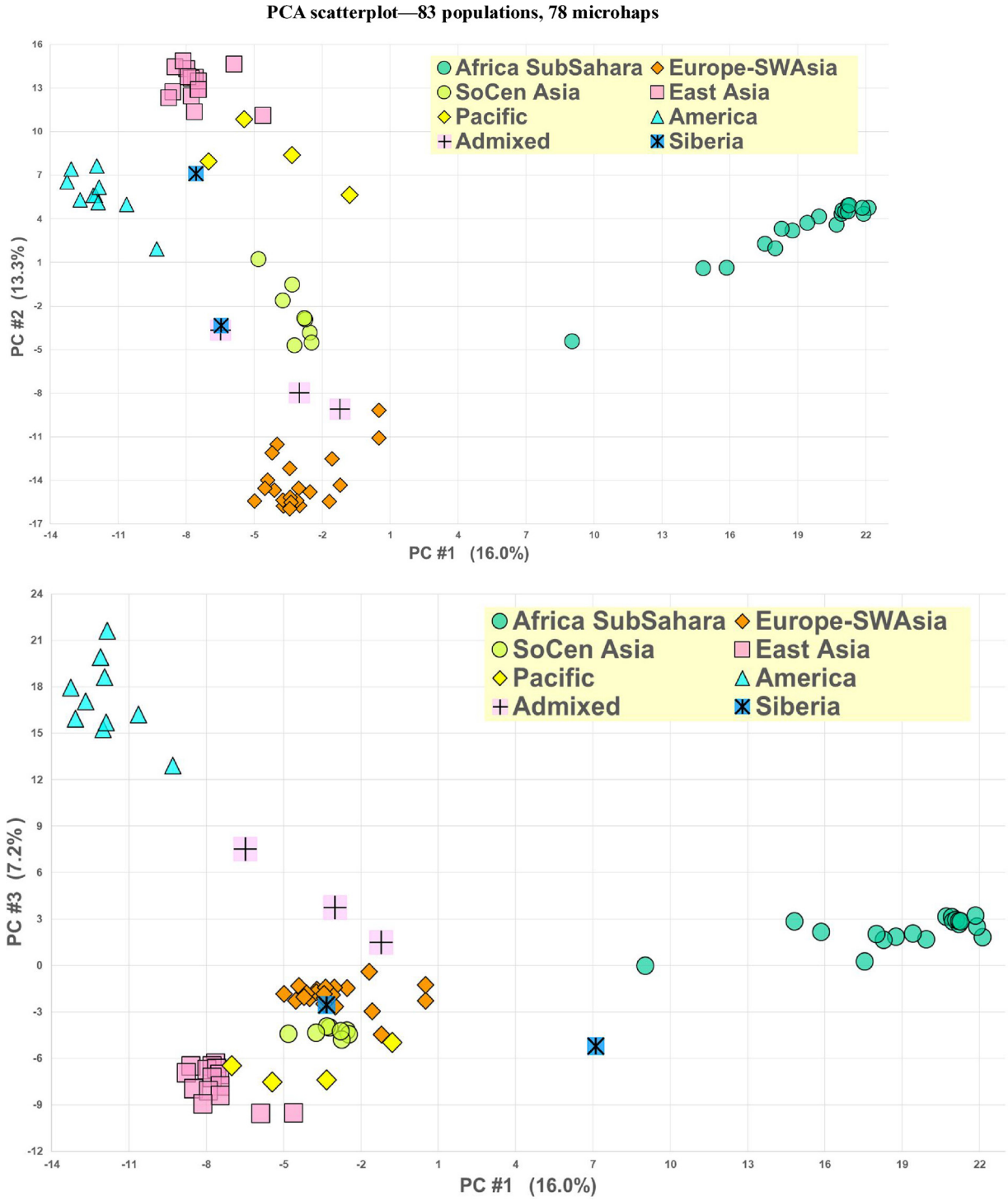

Fig. 5. Illustration of a two-dimensional PCA scatter plot (PC1 x PC2) across various global populations and the third dimension added by PC1 x PC3. The data are population allele frequencies at the 78 microhaplotyes analyzed in [124].

the Kidd subset of 65 microhaps, not developed for mixture identification, include 22 promising loci with $A_e > 3.0$ while only fewer out of 65 loci display values for $A_e < 2.0$ with various allelic variants displaying low frequencies [111].

3.3. Microhaplotypes for biogeographic ancestry inference

Frequently used methods of analysis of population relationships include STRUCTURE [119–121] and principal component analysis (PCA) [122,123]. These are helpful tools for predicting the biogeographic ancestry of an unknown sample based on testing many SNPs. An example of STRUCTURE plot analysis across 83 global populations analyzed using 130 microhap loci is illustrated in Fig. 4 (taken from [82]) and an example of a three-dimensional PCA plot (PC1 x PC2 and PC1 x PC3) is shown in Fig. 5 (based on the combined data from [124]) where individuals across different population clusters are shown in different colours and shapes. SNPs are ideal markers for predicting the biogeographic ancestry of individuals, as exemplified by the numerous AISNP panels developed so far [49–61]. As microhaplotypes are composed of combinations of SNPs, they can likewise

provide detailed ancestry information along with other useful forensic information from the same sample. As previously mentioned, the main metric used for selecting microhaplotype loci useful for ancestry is the informativeness metric: I_n. This measures the variation in allele (haplotype) frequencies among populations and therefore loci with high I_n values display greater differentiation among populations, thus making up a potentially useful marker panel for ancestry prediction. Since haplotype frequencies differ among populations in diverse regions of the world, I_n values will vary accordingly when different populations are considered distinctly. The ability of microhaps to infer the biogeographic ancestry was initially tested on the pilot set of 31 unlinked microhaps, albeit not specifically developed for such a purpose [81]. Among the selected loci tested on 2500 individuals, a microhap at the EDAR locus with the highest Fst value shows good ancestry discrimination but provides little use for individual and lineage identification outside of Africa. A microhap within the RXRA locus has a low Fst value because most populations have similar frequencies of the three haplotypes but provides good individual- and lineage- identification capabilities. Though STRUCTURE and PCA analyses of this 31 locus set [81] was less efficient in distinguishing more than five population clusters than various panels of AISNPs [57], it is still of forensic utility for ancestry inference. The different loci had different allele frequency patterns among the tested populations such that individuals from five major population clusters (k = 5), namely Africa, Europe and Southwest Asia, East Asia, the Americas and the Pacific Islands, could be clearly distinguished. The different pattern of allele frequencies and heterozygosity observed among all loci and populations is reflected in the RMP value ($< 10^{-15}$), which globally diverge more significantly among different regions than the 45 IISNPs panel [81].

For the Kidd panel of 130 microhaps, most loci show I_n values higher than 0.185 while microhaps with median I_n values < 0.185 do not offer strong ancestry prediction value, though they may potentially be improved by the inclusion of one or two additional SNPs within the locus [82]. The overall correlation between the I_n and A_e metrics is 0.53 for all loci, suggesting the potential of using this large panel of microhaps for simultaneous ancestry inference and mixture deconvolution. STRUCTURE analysis performed on a total of 5115 individuals had the highest likelihood at k = 6 while the results at higher k values showed the same pattern apart from the European populations that displayed partial ancestries. A particular cline in those specific contributors from Southwest Asia to Northern Europe is in line with what is also observed when typing the single SNP panels [55,125]. In addition, similar STRUCTURE results at k = 6 could be obtained using a subset of 50 top-ranked loci for I_n (> 0.215). Though this large panel enables the distinction of up to six different population clusters amongst 83 populations, several existing AISNP panels [52,55,57,61] can provide a better and finer ancestry inference at the biogeographical level. Nonetheless, this panel can produce greater population differentiation than current markers typed by CE when analysed by STRUCTURE.

When considering the panel of 65 loci, all of them showed greater levels of allele frequency variation among different population groups, except fewer loci that display I_n values < 0.1. When tested on 5667 individuals, the panel provided well-defined separation of the major continental population groups including Africa, Europe, South West - South Central and East Asia, America and Pacific, and a clear distinction of Native American from East Asian populations. The majority of Africans, East Asians, and Native Americans are assigned to a single cluster; however, some American populations were found highly admixed. Overall this panel provides a good separation of six global geographic regions and also offers the possibility of differentiation of up to ten population clusters. In particular, it allows further subdivision of distinct populations, including the farthest North in East Asia from East Asians as well as Sub-Saharan African populations from West Africa in comparison to East and Central Africans. Furthermore, the loose cluster of Eastern and Northern of the Eurasian populations was clearly separated from the European population, which appears to be very close and hence supporting the hypothesis that European populations occupy a geographically small area.

The preliminary investigation of the set of 50 microhaps with I_n values ranging from 0.03 to 0.29 evaluated by Pu et al [100] highlighted the ability to distinguish among three global population clusters by STRUCTURE analysis. Similarly, the 10 microhap loci developed by Chen et al. [105] were selected with levels of I_n values > 0.15 among 24 populations, and population stratification was evaluated by STRUCTURE. The small set of loci was found to be informative for differentiating African, European, East Asian and South Asian populations. However, the further development of a larger panel of shorter MH loci with AIM properties is required to provide more refined ancestry information at the regional/population level.

3.4. Development of early MPS microhaplotype panels for forensic purposes

A proof-of-concept set of 36 microhaps [126] selected from the Kidd panel of 130 loci [82] has recently been evaluated on the Ion Chef™ and Ion S5™ (Thermo Fisher Scientific). A beta version of the MPS AmpliSeq™ kit (Thermo Fisher Scientific) was evaluated for mixture deconvolution and its efficiency compared to CE analysis. Overall this initial set enabled the detection of different contributors in two, three, and four-person mixtures at 20:1, 10:1:1, and 10:1:1:1 ratios respectively, using a few hundred picograms of input DNA. In addition, a newly extended MPS Kidd panel of 74 loci implemented on the same Ion S5™ platform has recently been proposed to address questions on ancestry inference and mixture deconvolution [127]. Preliminary results confirmed the possibility to identify and deconvolute two-person DNA mixtures in presence of imbalanced allele reads within a heterozygous locus using 1–10 ng of input DNA, as similarly supported by additional preliminary investigations [98]. Results on the 74MHplex MPS assay also indicated the potential to enable successful ancestry inference of unknown testing-samples using phased inferred allele frequencies (from ALFRED) and those generated from four in-house population groups (Oldoni et al. unpublished data). The same study also showed the potential of using the 74MHplex MPS assay to simultaneously target the minor donor and predict the corresponding biogeographic ancestry (Oldoni et al. unpublished data) of the individual, as depicted in Fig. 6. This MPS investigation indicated that 74MHplex MPS assay is an effective tool to enhance biogeographic ancestry prediction, mixture deconvolution, and individual identification. Furthermore, an initial investigation was performed using a custom version of the semi-continuous probabilistic genotyping LRmixstudio (v2.1.4) software adapted to specifically accommodate microhap data. The analysis of a series of artificially made two- to four-person mixtures at different ratios and different DNA inputs, genotyped using the 74plex microhap assay, yielded probative likelihood ratio (LR) values for the correct minor contributor where the conventional autosomal STR analysis did not (data not shown).

Another study, using a panel of 89 loci selected from ALFRED tested on a population sample of 73 Italians using the Ion PGM instrument (Thermo Fisher Scientific), highlighted the potential of implementing microhaps for forensic investigations [97].

Moreover, the global variation of the 16-locus set [106] was evaluated on 276 population samples that included a total of 99 samples from the Netherlands, 87 from Asia, 90 from Africa and 97 CEPH-HGDP family samples used to confirm the inheritance of the haplotypes. STRUCTURE analysis was performed to evaluate the effectiveness of the 16-MH MPS assay for separation of populations of different ancestry, and results indicated almost complete differentiation of all three tested populations at K = 3, while a reduced separation of African and European populations was shown at higher K values. The RMP calculated for the 16 loci was higher than conventional non-STR loci kits with values of $4.4E\text{-}10^{-11}$, 1.0E- 10^{-9}, $9.2E\text{-}10^{-13}$ reported for Dutch, Asian and African populations, respectively (Table 1).

Chen et al. genotyped and multiplexed 14 out of 26 loci in a single

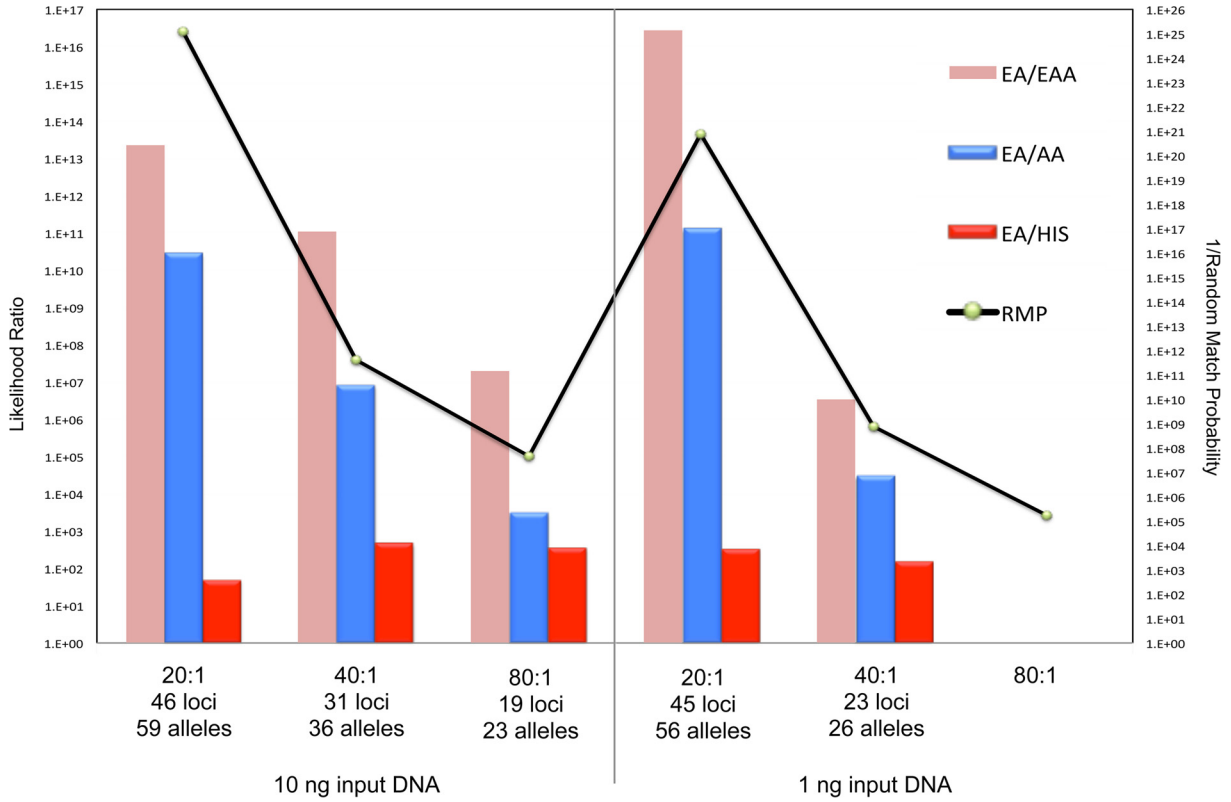

Fig. 6. An example of mixture study data on minor contributor LR-ancestry and 1/RMP. Data on two-person DNA mixtures simulated at different ratios and input amount of DNA and number of alleles/loci of the minor donor are shown. 1/RMP of European American donor and related LR-ancestry of European American (EA) individual *vs.* African American (AA), South West Hispanic (HIS) and East Asian American (EAA) are also reported (Oldoni et al. unpublished data).

Table 1
An overview of RMP values for different forensically relevant marker panels (table adapted from van der Gaag et al. [106]).

Panel	Number of loci	Type of loci	Random Match Probability	Population tested	Reference
SGM Plus kit	10	STRs	7.9E10-14	US African	Thermo Fisher Scientific
			3.0E10-13	US Caucasian	
NGM	15	STRs	1.6E10-19	US Hispanic	Thermo Fisher Scientific
			4.6E10-20	US African	
			2.2E10-19	US Caucasian	
NGM	9	STRs	3.1E10-12	US Hispanic	Thermo Fisher Scientific
			8.8E10-13	US African	
			2.6E10-12	US Caucasian	
Powerplex Fusion	24	STRs	1.6E10-28	US African	Promega
			2.4E10-27	US Caucasian	
			2.1E10-27	US Hispanic	
			1.4E10-25	US Asian	
SNPforID	52	SNPs	5.0E10-21	Euroepan	Sanchez et al [47]
			1.1E10-19	Somali	
			5.0E10-19	Asian	
IISNPs	45	SNPs	1.0E10-15 - 1.0E10-19	Global populations	Kidd et al [44]
tri-allelic SNPs	13	SNPs (tri-allelic)	3.2E10-6	Dutch	Westen et al [72]
			4.4E10-7	Dutch Antilles	
tetra-allelic SNPs	24	SNPs (tetra-allelic)	1.5E10-12	European	Phillips et al [73]
			5.2E10-10	East Asian	
			2.0E10-15	African	
Microhaplotypes	31	Micro haplotypes	1.0E10-13 - 4.0E10-21	Global populations	Kidd et al [81]
Microhaplotypes	top 50	Micro haplotypes	1.0E10-19 - 1.0E10-42 (top I_n)	Global populations	Kidd et al [124]
			1.0E10-27 - 1.0E10-50 (top A_e)		
Short hypervariable microhaplotypes	16	Micro haplotypes	4.4E10-11	Netherlands	van der Gaag et al [106]
			1.0E10-9	China/Japan	
			9.2E10-13	Kenya/Nigeria	
Microhaplotypes	74	Micro haplotypes	1.9E10-68	US African (80 samples)	Oldoni et al (unpublished)
			3.2E10-64	US Caucasian (110 samples)	
			4.9E10-67	US Hispanic (100 samples)	
			3.0E10-62	US East Asian (37 samples)	
			4.1E10-61	East Asian (62 samples)	

MPS assay on the Hiseq X® Sequencer® (Illumina) platform. The small 14-MH MPS assay [104] was tested for mixture analysis on two artificially prepared two-person mixed source samples at 1:1 ratio and 600 ng input of DNA. Preliminary results indicated that approximately one third of the tested markers showed three or more haplotypes in the simulated mixtures, thus suggesting their potential utility for mixture detection and deconvolution.

Finally, Voskoboinik et al. proposed the use of highly polymorphic haplotypes ($A_e > 15$) for mixture deconvolution [128]. They selected a set of 10 loci containing > 10 SNPs for simulation analysis with sequence reads from mixed-source samples simulated to reflect the expected reads from the emerging nanopore-based MinION instrument-sequencing platform (Oxford Nanopore). Mixed-source samples were simulated by combining a fixed number of donors and contribution ratios with each donor consisting of randomly sampled haplotypes from European haplotype frequencies from 1000 Genomes at each locus. To simulate the average substitution rates commonly observed for the MinION platform, sequencing errors were randomly introduced into the DNA strands. This simulation study highlighted the utility of these hypervariable loci to include one donor to a four or five-person mixture; however, these markers require thorough wet-laboratory development and validation.

3.5. Microhaplotypes for individual and missing person identification

SNPs are ideal markers for improving the statistics of individual identification of STR markers when they are either detected as nucleotide variants within a STR locus or used as a distinct panel of loci. Since microhaps are characterized by combinations of SNPs, they likewise display promising and valued features for human identity similarly to SNP-based genotyping assays [43–47,129]. Heterozygosity reaches its highest value when all alleles are evenly frequent. Initial evaluation of the heterozygosity of microhaps was performed on the Kidd pilot set of 31 loci albeit not specifically selected for human identifications [81]. Overall, the 31-microhap set showed either match probabilities comparable to or lower than the those obtained for 45 out of 86 mutually unlinked IISNPs (*i.e.,* RMP: 10^{-13}-10^{-18}) studied on 44 populations [44], the 52plex SNP assay (*i.e.,* RMP: 10^{-19}-10^{-21}) and conventional CODIS STRs (Table 1). Overall, the set of 31 unlinked microhaps exhibited random match probabilities < 10^{-15} (up to 10^{-20}) while only four relatively small/inbred populations showed match probabilities in the range of 10^{-13}-10^{-15} [81]. Additional panels of 65 and 130 loci include microhaps that could be much more informative due to the presence of loci with three or four highly heterozygous SNPs [82]. When extending the number of microhaps for analysis, it is plausible to assume a significant increase in the value of match probabilities beyond the range of 10^{-31}-10^{-35} calculated for 86 IISNPs specifically selected for human identity [44]. When considering the 74MHplex MPS assay the RMP value varied from 10^{-61}-10^{-68} (unpublished data). Refer to Table 1 for a comparison of RMP for different panels of forensic loci.

An alternative approach was recently presented by Bose et al. in a study investigating the potential of using a novel MPS target capture enrichment approach of nuclear SNP loci for specifically typing degraded and mixed samples [130]. A set of 411 polymorphic loci, that included 375 SNPs and 76 additional SNPs from 36 microhaps [130] was selected to address individual identification, lineage/kinship and mixture analysis. Overall the custom SNP probe capture assay was found efficient in typing mock degraded samples fragmented from 200 bp to 75 bp with > 90% reportable SNPs across different input amounts of DNA. Additionally, it successfully detected minor contributor alleles in artificially generated two-person male-male mixtures at multiple contributor ratios (5–20% minor) with 10 ng input DNA.

Lastly, current microhap panels with an overall amplicon size ranging from a few tens to a few hundred bp could be suitable for the analysis of degraded DNA samples commonly encountered in forensic practice and accordingly, potentially useful to complement conventional analysis methods for missing person and disaster victim identification.

3.6. Microhaplotypes for family and lineage relationships

The initial and leading motivation for the identification of multiallelic SNP loci was to assist in the identification of clan-and-extended family relationships [30], commonly performed using mtDNA, Y chromosome and autosomal STRs [42]. Theoretically, many alleles observed in a specific population and multiple combinations of them reduce the likelihood that two unrelated individuals share the same allele combination by chance. Therefore, the more heterozygous a locus is, the higher the likelihood that the relevant alleles are not common in the population of interest; however, they are more likely to be observed amongst close relatives than amongst random or distantly related individuals. One important parameter useful for lineage/family relationship (but also for ancestry inference) is the identity by descent (IBD) concept. For more reliable prediction of the degree of relatedness between two individuals, a panel of a few hundred (bi-allelic) SNP markers may prove sufficient, as reported by Weir et al. [131]. Moreover, a high level of global heterozygosity and Fst value either globally or among specific geographical regions of the world are also key metrics required to provide useful information on family-clan relationship and human identification [82,100]. Due to the low mutation rate of SNPs, different microhaps developed thus far can complement standard STRs in deficient family pedigrees. For the panel of 130 microhaps, only 28 loci show $A_e > 3.0$ and can therefore be of great value for lineage identification [82]. When considering the panel of 65 markers, 22 out of 65 loci with $A_e > 3.0$ show potential for identifying relatives and reduce significantly the probability that two unrelated individuals share the same microhap genotype [111]. Future studies will focus on the fine exploration of microhaps for family-clan relationship and on testing real forensic casework samples.

3.7. Non-forensic uses of microhaplotypes

A non-forensic application for which microhaplotypes have been proposed is in the field of hematopoietic stem cell transplantation. Prior to undergoing transplantation patients are subjected to myeloablative conditioning, which consist in destroying/reducing the patient's hematopoietic stem cells. The purpose of this step is (1) to suppress the immune system by preventing rejection of the new bone marrow, (2) make room in the bone marrow for the donor's stem cells to grow and, in cancer patients, (3) to kill cancer cells. After transplantation patients essentially become chimeras of themselves and the donor. In order to determine the effectiveness of the transplant and evaluate potential graft rejection, chimerism testing is performed using conventional STR testing from the blood's recipient [132,133]. The ratio between donor alleles *vs*. patient alleles is evaluated over time at set intervals and is used to formulate a prognosis and develop a treatment plan in case rejection occurs. Early detection of graft rejection improves prognosis, allowing intervention while the malignant clone is at a lower titre and the limit of detection has been shown to be in the 5% ratio range [134] with conventional STR testing. In 2014 Debaljak et al. [135] proposed the use of MPS for the analysis of contiguous SNPs (aka microhaplotypes) and developed an assay targeting a short 300 bp fragment in the HLA-A region. The assay yielded a limit of detection in the 0.01% range for the minor contributor and while conventional forensic STR with the AmpFlSTR® Profiler® and AmpFlSTR® Identifiler® kits the limit of detection was approximately 3%. In a following study Debaljak et al. developed a 5-plex amplification targeting fragments in the FARP1, HLA-A, HLA-B, MT4, and TMPRSS15 regions which allowed the detection of relapse patients on average 124 days earlier than with STRs [136]. In a similar fashion, non-forensic applications of MHs have been suggested for non-invasive parental testing of fetal cells (and cell-free

DNA) circulating in maternal blood [137].

An additional promising application of microhaps is to non-human DNA, as recently shown by Baetscher [138]. The authors reported the use of a large panel of microhaps to solve difficult pedigree inference problems on the kelp rockfish, *Sebastes atrovirens* (Actinopterygii; Sebastidae), and an ecological and culturally important nearshore Pacific fish species. The identified panel of multi-allelic microhaplotype loci showed a remarkably higher heterozygosity than the equivalent number of individual SNPs. As a whole, microhaps provided higher power for identification of single parent-offspring and full-sibling pairs. Another example is in the study of landlocked alewife populations from coastal New England lakes where a MH panel was specifically developed to identify anadromous, landlocked, and hybrid juvenile fish [139]. The use of MHs could be broadened to other species and plants overall improving conservation efforts.

3.8. Selection of microhaplotypes optimized for different purposes

The Kidd laboratory has identified and fully characterized a total of 182 microhaps in many diverse populations, and many more are under development, also by other laboratories. With this consideration in mind, different fine-tuned sets of microhaps that outperform the existing SNP assays might be optimized to provide untapped advantages for forensic purposes. The selection and ranking of optimal microhap loci based on both Rosenberg's measure of informativeness (I_n) and the effective number of allele (A_e) will be the objective of future studies aimed to address most relevant forensic DNA questions. Top-ranked microhaps for I_n would make a useful panel for ancestry prediction while loci with high A_e would make a suitable panel for mixture detection and deconvolution and also for relationship identification [124].

As microhaps are composed of combinations of SNPs, they may be found in regions associated with human phenotypic traits. As recently reported by Bulbul et al. [111], three of the new microhaps included in the panel of 65 loci, namely mk05KK-122 (rs1010872-rs28777), mk05KK-123 (rs28117-rs1423676), mk05KK-124 (rs35414-rs3756464), are reported in the region of the SLC45A2 gene, which is associated with eye, hair and skin color [63]. Typing these loci can provide pigmentation prediction information based on individual SNPs, and associations between certain phenotypes and specific MH alleles will probably be discovered in the future.

4. Concluding remarks

The very rapid transition from minihaplotype to microhaplotype markers is analogous to the early forensic shift from minisatellite to microsatellite DNA polymorphisms in the late nineties. Microhaps are a powerful and comprehensive molecular tool that provides a great wealth of information relevant to addressing different forensic and human genetic diversity questions. At least for the foreseeable future, microhaps have the potential to enhance ancestry prediction and mixture deconvolution (also by means of probabilistic genotyping software analysis), improve human and missing person identification together with relationship testing. Finally, these new markers can be effective in clinical and non-human DNA applications.

Disclosure statement

The authors declare no conflict of interest.

Acknowledgements

The authors would like to acknowledge the three anonymous reviewers for their useful comments on this manuscript. This work was supported in part by the National Institute of Justice through grant No. 2017-DN-BX-0164 awarded to Daniele Podini and grant No. 2015-DN-BX-K023 to Kenneth K. Kidd and the Swiss National Science Foundation through grant No. 2017-P2LAP3_174742 awarded to Fabio Oldoni.

References

[1] R. Ceppelini, E. Curtoni, P. Mattiuz, V. Miggiano, G. Scudeller, A. Serra, E.S. Curtoni, P.L. Mattiuz, R.M. Tosi (Eds.), Genetics of leukocyte antigens: A family study of segregation and linkage, Histocompat. Test, Munksgaard, Copenhagen, 1967, pp. 149–187.

[2] F.S. Collins, The human genome project: lessons from large-scale biology, Science (80-.) 300 (2003) 286–290, https://doi.org/10.1126/science.1084564.

[3] F.R. Blattner, G. Plunkett, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau, Y. Shao, The complete genome sequence of Escherichia coli K-12, Science 277 (1997) 1453–1462 (Accessed 17 April 2018), http://www.ncbi.nlm.nih.gov/pubmed/9278503.

[4] C. elegans sequencing consortium, genome sequence of the nematode C. elegans: a platform for investigating biology, Science 282 (1998) 2012–2018 (Accessed 17 April 2018), http://www.ncbi.nlm.nih.gov/pubmed/9851916.

[5] M.D. Adams, S.E. Celniker, R.A. Holt, C.A. Evans, J.D. Gocayne, P.G. Amanatides, S.E. Scherer, P.W. Li, R.A. Hoskins, R.F. Galle, R.A. George, S.E. Lewis, S. Richards, M. Ashburner, S.N. Henderson, G.G. Sutton, J.R. Wortman, M.D. Yandell, Q. Zhang, L.X. Chen, R.C. Brandon, Y.H. Rogers, R.G. Blazej, M. Champe, B.D. Pfeiffer, K.H. Wan, C. Doyle, E.G. Baxter, G. Helt, C.R. Nelson, G.L. Gabor, J.F. Abril, A. Agbayani, H.J. An, C. Andrews-Pfannkoch, D. Baldwin, R.M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E.M. Beasley, K.Y. Beeson, P.V. Benos, B.P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M.R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K.C. Burtis, D.A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J.M. Cherry, S. Cawley, C. Dahlke, L.B. Davenport, P. Davies, B. de Pablos, A. Delcher, Z. Deng, A.D. Mays, I. Dew, S.M. Dietz, K. Dodson, L.E. Doup, M. Downes, S. Dugan-Rocha, B.C. Dunkov, P. Dunn, K.J. Durbin, C.C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A.E. Gabrielian, N.S. Garg, W.M. Gelbart, K. Glasser, A. Glodek, F. Gong, J.H. Gorrell, Z. Gu, P. Guan, M. Harris, N.L. Harris, D. Harvey, T.J. Heiman, J.R. Hernandez, J. Houck, D. Hostin, K.A. Houston, T.J. Howland, M.H. Wei, C. Ibegwam, M. Jalali, F. Kalush, G.H. Karpen, Z. Ke, J.A. Kennison, K.A. Ketchum, B.E. Kimmel, C.D. Kodira, C. Kraft, S. Kravitz, D. Kulp, Z. Lai, P. Lasko, Y. Lei, A.A. Levitsky, J. Li, Z. Li, Y. Liang, X. Lin, X. Liu, B. Mattei, T.C. McIntosh, M.P. McLeod, D. McPherson, G. Merkulov, N.V. Milshina, C. Mobarry, J. Morris, A. Moshrefi, S.M. Mount, M. Moy, B. Murphy, L. Murphy, D.M. Muzny, D.L. Nelson, D.R. Nelson, K.A. Nelson, K. Nixon, D.R. Nusskern, J.M. Pacleb, M. Palazzolo, G.S. Pittman, S. Pan, J. Pollard, V. Puri, M.G. Reese, K. Reinert, K. Remington, R.D. Saunders, F. Scheeler, H. Shen, B.C. Shue, I. Sidén-Kiamos, M. Simpson, M.P. Skupski, T. Smith, E. Spier, A.C. Spradling, M. Stapleton, R. Strong, E. Sun, R. Svirskas, C. Tector, R. Turner, E. Venter, A.H. Wang, X. Wang, Z.Y. Wang, D.A. Wassarman, G.M. Weinstock, J. Weissenbach, S.M. Williams, K.C. WoodageT, D. Worley, S. Wu, Q.A. Yang, J.Ye Yao, R.F. Yeh, J.S. Zaveri, M. Zhan, G. Zhang, Q. Zhao, L. Zheng, X.H. Zheng, F.N. Zhong, W. Zhong, X. Zhou, S. Zhu, X. Zhu, H.O. Smith, R.A. Gibbs, E.W. Myers, G.M. Rubin, J.C. Venter, The genome sequence of Drosophila melanogaster, Science 287 (2000) 2185–2195 (Accessed 17 April 2018), http://www.ncbi.nlm.nih.gov/pubmed/10731132.

[6] R.D. Fleischmann, M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.F. Tomb, B.A. Dougherty, J.M. Merrick, Whole-genome random sequencing and assembly of Haemophilus influenzae Rd, Science 269 (1995) 496–512 (Accessed 17 April 2018), http://www.ncbi.nlm.nih.gov/pubmed/7542800.

[7] A comprehensive genetic linkage map of the human genome. NIH/CEPH Collaborative Mapping Group, Science 258 (1992) 67–86 (Accessed 17 April 2018), http://www.ncbi.nlm.nih.gov/pubmed/1439770.

[8] J.D. McPherson, M. Marra, L. Hillier, R.H. Waterston, A. Chinwalla, J. Wallis, M. Sekhon, K. Wylie, E.R. Mardis, R.K. Wilson, R. Fulton, T.A. Kucaba, C. Wagner-McPherson, W.B. Barbazuk, S.G. Gregory, S.J. Humphray, L. French, R.S. Evans, G. Bethel, A. Whittaker, J.L. Holden, O.T. McCann, A. Dunham, C. Soderlund, C.E. Scott, D.R. Bentley, G. Schuler, H.C. Chen, W. Jang, E.D. Green, J.R. Idol, V.V. Maduro, K.T. Montgomery, E. Lee, A. Miller, S. Emerling, R. Kucherlapati, R. Gibbs, S. Scherer, J.H. Gorrell, E. Sodergren, K. Clerc-Blankenburg, P. Tabor, S. Naylor, D. Garcia, P.J. de Jong, J.J. Catanese, N. Nowak, K. Osoegawa, S. Qin, L. Rowen, A. Madan, M. Dors, L. Hood, B. Trask, C. Friedman, H. Massa, V.G. Cheung, I.R. Kirsch, T. Reid, R. Yonescu, J. Weissenbach, T. Bruls, R. Heilig, E. Branscomb, A. Olsen, N. Doggett, J.F. Cheng, T. Hawkins, R.M. Myers, J. Shang, L. Ramirez, J. Schmutz, O. Velasquez, K. Dixon, N.E. Stone, D.R. Cox, D. Haussler, W.J. Kent, T. Furey, S. Rogic, S. Kennedy, S. Jones, A. Rosenthal, G. Wen, M. Schilhabel, G. Gloeckner, G. Nyakatura, R. Siebert, B. Schlegelberger, J. Korenberg, X.N. Chen, A. Fujiyama, M. Hattori, A. Toyoda, T. Yada, H.S. Park, Y. Sakaki, N. Shimizu, S. Asakawa, K. Kawasaki, T. Sasaki, A. Shintani, A. Shimizu, K. Shibuya, J. Kudoh, S. Minoshima, J. Ramser, P. Seranski, C. Hoff, A. Poustka, R. Reinhardt, H. Lehrach, International Human Genome Mapping Consortium, a physical map of the human genome, Nature 409 (2001) 934–941, https://doi.org/10.1038/35057157.

[9] W.F. Dietrich, J. Miller, R. Steen, M.A. Merchant, D. Damron-Boles, Z. Husain, R. Dredge, M.J. Daly, K.A. Ingalls, T.J. O'Connor, C.A. Evans, M.M. DeAngelis, D.M. Levinson, L. Kruglyak, N. Goodman, N.G. Copeland, N.A. Jenkins, T.L. Hawkins, L. Stein, D.C. Page, E.S. Lander, A comprehensive genetic map of the mouse genome, Nature 380 (1996) 149–152, https://doi.org/10.1038/380149a0.

[10] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, J.D. Gocayne, P. Amanatides, R.M. Ballew, D.H. Huson, J.R. Wortman, Q. Zhang, C.D. Kodira, X.H. Zheng, L. Chen, M. Skupski, G. Subramanian, P.D. Thomas, J. Zhang, G.L. Gabor Miklos, C. Nelson, S. Broder, A.G. Clark, J. Nadeau, V.A. McKusick, N. Zinder, A.J. Levine, R.J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A.E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T.J. Heiman, M.E. Higgins, R.-R. Ji, Z. Ke, K.A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G.V. Merkulov, N. Milshina, H.M. Moore, A.K. Naik, V.A. Narayan, B. Neelam, D. Nusskern, D.B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z.Y. Wang, A. Wang, X. Wang, J. Wang, M.-H. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S.C. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M.L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y.-H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N.N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J.F. Abril, R. Guigó, M.J. Campbell, K.V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y.-H. Chiang, M. Coyne, C. Dahlke, A.D. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, X. Zhu, The sequence of the human genome, Science (80-.) 291 (2001) 1304–1351, https://doi.org/10.1126/science.1058040.

[11] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, et al., International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921, https://doi.org/10.1038/35057062.

[12] I. Human Genome Sequencing Consortium, finishing the euchromatic sequence of the human genome, Nature 431 (2004) 931–945, https://doi.org/10.1038/nature03001.

[13] M.J. Daly, J.D. Rioux, S.F. Schaffner, T.J. Hudson, E.S. Lander, High-resolution haplotype structure in the human genome, Nat. Genet. 29 (2001) 229–232, https://doi.org/10.1038/ng1001-229.

[14] N. Patil, A.J. Berno, D.A. Hinds, W.A. Barrett, J.M. Doshi, C.R. Hacker, C.R. Kautzer, D.H. Lee, C. Marjoribanks, D.P. McDonough, B.T. Nguyen, M.C. Norris, J.B. Sheehan, N. Shen, D. Stern, R.P. Stokowski, D.J. Thomas, M.O. Trulson, K.R. Vyas, K.A. Frazer, S.P. Fodor, D.R. Cox, Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21, Science 294 (2001) 1719–1723, https://doi.org/10.1126/science.1065573.

[15] J.C. Stephens, J.A. Schneider, D.A. Tanguay, J. Choi, T. Acharya, S.E. Stanley, R. Jiang, C.J. Messer, A. Chew, J.-H. Han, J. Duan, J.L. Carr, M.S. Lee, B. Koshy, A.M. Kumar, G. Zhang, W.R. Newell, A. Windemuth, C. Xu, T.S. Kalbfleisch, S.L. Shaner, K. Arnold, V. Schulz, C.M. Drysdale, K. Nandabalan, R.S. Judson, G. Ruaño, G.F. Vovis, Haplotype variation and linkage disequilibrium in 313 human genes, Science 293 (2001) 489–493, https://doi.org/10.1126/science.1059431.

[16] S.B. Gabriel, S.F. Schaffner, H. Nguyen, J.M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S.N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E.S. Lander, M.J. Daly, D. Altshuler, The structure of haplotype blocks in the human genome, Science (80-.) 296 (2002) 2225–2229, https://doi.org/10.1126/science.1069424.

[17] L.L. Cavalli-Sforza, A.C. Wilson, C.R. Cantor, R.M. Cook-Deegan, M.-C. King, Call for a worldwide survey of human genetic diversity: a vanishing opportunity for the Human Genome Project, Genomics 11 (1991) 490–491, https://doi.org/10.1016/0888-7543(91)90169-F.

[18] H.M. Cann, C. de Toma, L. Cazes, M.-F. Legrand, V. Morel, L. Piouffre, J. Bodmer, W.F. Bodmer, B. Bonne-Tamir, A. Cambon-Thomsen, Z. Chen, J. Chu, C. Carcassi, L. Contu, R. Du, L. Excoffier, G.B. Ferrara, J.S. Friedlaender, H. Groot, D. Gurwitz, T. Jenkins, R.J. Herrera, X. Huang, J. Kidd, K.K. Kidd, A. Langaney, A.A. Lin, S.Q. Mehdi, P. Parham, A. Piazza, M.P. Pistillo, Y. Qian, Q. Shu, J. Xu, S. Zhu, J.L. Weber, H.T. Greely, M.W. Feldman, G. Thomas, J. Dausset, L.L. Cavalli-Sforza, A human genome diversity cell line panel, Science (80-.) 296 (2002) 261 LP-262 http://science.sciencemag.org/content/296/5566/261.2.abstract.

[19] L. Kruglyak, Prospects for whole-genome linkage disequilibrium mapping of common disease genes, Nat. Genet. 22 (1999) 139–144, https://doi.org/10.1038/9642.

[20] K.M. Weiss, A.G. Clark, Linkage disequilibrium and the mapping of complex human traits, Trends Genet. 18 (2002) 19–24 (Accessed 18 April 2018), http://www.ncbi.nlm.nih.gov/pubmed/11750696.

[21] S.L. Sawyer, N. Mukherjee, A.J. Pakstis, L. Feuk, J.R. Kidd, A.J. Brookes, K.K. Kidd, Linkage disequilibrium patterns vary substantially among populations, Eur. J. Hum. Genet. 13 (2005) 677–686, https://doi.org/10.1038/sj.ejhg.5201368.

[22] D.B. Goldstein, Erratum: Islands of linkage disequilibrium, Nat. Genet. 29 (2001) 109–111, https://doi.org/10.1038/ng1001-109.

[23] I.H. Consortium, K.A. Frazer, D.G. Ballinger, D.R. Cox, D.A. Hinds, L.L. Stuve, R.A. Gibbs, J.W. Belmont, A. Boudreau, P. Hardenbol, S.M. Leal, S. Pasternak, D.A. Wheeler, T.D. Willis, F. Yu, H. Yang, C. Zeng, Y. Gao, H. Hu, W. Hu, C. Li, W. Lin, S. Liu, H. Pan, X. Tang, J. Wang, W. Wang, J. Yu, B. Zhang, Q. Zhang, H. Zhao, H. Zhao, J. Zhou, S.B. Gabriel, R. Barry, B. Blumenstiel, A. Camargo, M. Defelice, M. Faggart, M. Goyette, S. Gupta, J. Moore, H. Nguyen, R.C. Onofrio, M. Parkin, J. Roy, E. Stahl, E. Winchester, L. Ziaugra, D. Altshuler, Y. Shen, Z. Yao, W. Huang, X. Chu, Y. He, L. Jin, Y. Liu, Y. Shen, W. Sun, H. Wang, Y. Wang, Y. Wang, X. Xiong, L. Xu, M.M.Y. Waye, S.K.W. Tsui, H. Xue, J.T.-F. Wong, L.M. Galver, J.-B. Fan, K. Gunderson, S.S. Murray, A.R. Oliphant, M.S. Chee, A. Montpetit, F. Chagnon, V. Ferretti, M. Leboeuf, J.-F. Olivier, M.S. Phillips, S. Roumy, C. Sallée, A. Verner, T.J. Hudson, P.-Y. Kwok, D. Cai, D.C. Koboldt, R.D. Miller, L. Pawlikowska, P. Taillon-Miller, M. Xiao, L.-C. Tsui, W. Mak, Y.Q. Song, P.K.H. Tam, Y. Nakamura, T. Kawaguchi, T. Kitamoto, T. Morizono, A. Nagashima, Y. Ohnishi, A. Sekine, T. Tanaka, T. Tsunoda, P. Deloukas, C.P. Bird, M. Delgado, E.T. Dermitzakis, R. Gwilliam, S. Hunt, J. Morrison, D. Powell, B.E. Stranger, P. Whittaker, D.R. Bentley, M.J. Daly, P.I.W. de Bakker, J. Barrett, Y.R. Chretien, J. Maller, S. McCarroll, N. Patterson, I. Pe'er, A. Price, S. Purcell, D.J. Richter, P. Sabeti, R. Saxena, S.F. Schaffner, P.C. Sham, P. Varilly, D. Altshuler, L.D. Stein, L. Krishnan, A.V. Smith, M.K. Tello-Ruiz, G.A. Thorisson, A. Chakravarti, P.E. Chen, D.J. Cutler, C.S. Kashuk, S. Lin, G.R. Abecasis, W. Guan, Y. Li, H.M. Munro, Z.S. Qin, D.J. Thomas, G. McVean, A. Auton, L. Bottolo, N. Cardin, S. Eyheramendy, C. Freeman, J. Marchini, S. Myers, C. Spencer, M. Stephens, P. Donnelly, L.R. Cardon, G. Clarke, D.M. Evans, A.P. Morris, B.S. Weir, T. Tsunoda, J.C. Mullikin, S.T. Sherry, M. Feolo, A. Skol, H. Zhang, C. Zeng, H. Zhao, I. Matsuda, Y. Fukushima, D.R. Macer, E. Suda, C.N. Rotimi, C.A. Adebamowo, I. Ajayi, T. Aniagwu, P.A. Marshall, C. Nkwodimmah, C.D.M. Royal, M.F. Leppert, M. Dixon, A. Peiffer, R. Qiu, A. Kent, K. Kato, N. Niikawa, I.F. Adewole, B.M. Knoppers, M.W. Foster, E.W. Clayton, J. Watkin, R.A. Gibbs, J.W. Belmont, D. Muzny, L. Nazareth, E. Sodergren, G.M. Weinstock, D.A. Wheeler, I. Yakub, S.B. Gabriel, R.C. Onofrio, D.J. Richter, L. Ziaugra, B.W. Birren, M.J. Daly, D. Altshuler, R.K. Wilson, L.L. Fulton, J. Rogers, J. Burton, N.P. Carter, C.M. Clee, M. Griffiths, M.C. Jones, K. McLay, R.W. Plumb, M.T. Ross, S.K. Sims, D.L. Willey, Z. Chen, H. Han, L. Kang, M. Godbout, J.C. Wallenburg, P. L'Archevêque, G. Bellemare, K. Saeki, H. Wang, D. An, H. Fu, Q. Li, Z. Wang, R. Wang, A.L. Holden, L.D. Brooks, J.E. McEwen, M.S. Guyer, V.O. Wang, J.L. Peterson, M. Shi, J. Spiegel, L.M. Sung, L.F. Zacharia, F.S. Collins, K. Kennedy, R. Jamieson, J. Stewart, A second generation human haplotype map of over 3.1 million SNPs, Nature 449 (2007) 851–861, https://doi.org/10.1038/nature06258.

[24] International HapMap 3 Consortium, D.M. Altshuler, R.A. Gibbs, L. Peltonen, D.M. Altshuler, R.A. Gibbs, L. Peltonen, E. Dermitzakis, S.F. Schaffner, F. Yu, L. Peltonen, E. Dermitzakis, P.E. Bonnen, D.M. Altshuler, R.A. Gibbs, P.I.W. de Bakker, P. Deloukas, S.B. Gabriel, R. Gwilliam, S. Hunt, M. Inouye, X. Jia, A. Palotie, M. Parkin, P. Whittaker, F. Yu, K. Chang, A. Hawes, L.R. Lewis, Y. Ren, D. Wheeler, R.A. Gibbs, D.M. Muzny, C. Barnes, K. Darvishi, M. Hurles, J.M. Korn, K. Kristiansson, C. Lee, S.A. McCarrol, J. Nemesh, E. Dermitzakis, A. Keinan, S.B. Montgomery, S. Pollack, A.L. Price, N. Soranzo, P.E. Bonnen, R.A. Gibbs, C. Gonzaga-Jauregui, A. Keinan, A.L. Price, F. Yu, V. Anttila, W. Brodeur, M.J. Daly, S. Leslie, G. McVean, L. Moutsianas, H. Nguyen, S.F. Schaffner, Q. Zhang, M.J.R. Ghori, R. McGinnis, W. McLaren, S. Pollack, A.L. Price, S.F. Schaffner, F. Takeuchi, S.R. Grossman, I. Shlyakhter, E.B. Hostetter, P.C. Sabeti, C.A. Adebamowo, M.W. Foster, D.R. Gordon, J. Licinio, M.C. Manca, P.A. Marshall, I. Matsuda, D. Ngare, V.O. Wang, D. Reddy, C.N. Rotimi, C.D. Royal, R.R. Sharp, C. Zeng, L.D. Brooks, J.E. McEwen, Integrating common and rare genetic variation in diverse human populations, Nature 467 (2010) 52–58, https://doi.org/10.1038/nature09298.

[25] R.A. Gibbs, J.W. Belmont, P. Hardenbol, T.D. Willis, F. Yu, H. Zhang, C. Zeng, I. Matsuda, Y. Fukushima, D.R. Macer, E. Suda, L.D. Stein, F. Cunningham, A. Kanani, G.A. Thorisson, A. Chakravarti, P.E. Chen, D.J. Cutler, C.S. Kashuk, P. Donnelly, J. Marchini, G.A.T. McVean, S.R. Myers, L.R. Cardon, G.R. Abecasis, A. Morris, B.S. Weir, J.C. Mullikin, S.T. Sherry, M. Feolo, D. Altshuler, M.J. Daly, S.F. Schaffner, R. Qiu, A. Kent, G.M. Dunston, K. Kato, N. Niikawa, B.M. Knoppers, M.W. Foster, E.W. Clayton, V.O. Wang, J. Watkin, R.A. Gibbs, J.W. Belmont, E. Sodergren, G.M. Weinstock, R.K. Wilson, L.L. Fulton, J. Rogers, B.W. Birren, H. Han, H. Wang, M. Godbout, J.C. Wallenburg, P. L'Archevêque, G. Bellemare, K. Todani, T. Fujita, S. Tanaka, A.L. Holden, E.H. Lai, F.S. Collins, L.D. Brooks, J.E. McEwen, M.S. Guyer, E. Jordan, J.L. Peterson, J. Spiegel, L.M. Sung, L.F. Zacharia, K. Kennedy, M.G. Dunn, R. Seabrook, M. Shillito, B. Skene, J.G. Stewart, D.L. Valle, E.W. Clayton, L.B. Jorde, J.W. Belmont, A. Chakravarti, M.K. Cho, T. Duster, M.W. Foster, M. Jasperse, B.M. Knoppers, P.-Y. Kwok, J. Licinio, J.C. Long, P.A. Marshall, P.N. Ossorio, V.O. Wang, C.N. Rotimi, C.D.M. Royal, P. Spallone, S.F. Terry, E.S. Lander, E.H. Lai, D.A. Nickerson, G.R. Abecasis, D. Altshuler, D.R. Bentley, M. Boehnke, L.R. Cardon, M.J. Daly, P. Deloukas, J.A. Douglas, S.B. Gabriel, R.R. Hudson, T.J. Hudson, L. Kruglyak, P.-Y. Kwok, Y. Nakamura, R.L. Nussbaum, C.D.M. Royal, S.F. Schaffner, S.T. Sherry, L.D. Stein, T. Tanaka, The international HapMap project, Nature 426 (2003) 789–796, https://doi.org/10.1038/nature02168.

[26] S. Gu, A.J. Pakstis, H. Li, W.C. Speed, J.R. Kidd, K.K. Kidd, Significant variation in haplotype block structure but conservation in tagSNP patterns among global

populations, Eur. J. Hum. Genet. 15 (2007) 302–312, https://doi.org/10.1038/sj.ejhg.5201751.
[27] International HapMap Consortium, A haplotype map of the human genome, Nature 437 (2005) 1299–1320, https://doi.org/10.1038/nature04226.
[28] H. Zhao, A.J. Pakstis, J.R. Kidd, K.K. Kidd, Assessing linkage disequilibrium in a complex genetic system. I. Overall deviation from random association, Ann. Hum. Genet. 63 (1999) 167–179.
[29] J. Ge, B. Budowle, J.V. Planz, R. Chakraborty, Haplotype block: a new type of forensic DNA markers, Int. J. Legal Med. 124 (2010) 353–361, https://doi.org/10.1007/s00414-009-0400-5.
[30] A.J. Pakstis, R. Fang, M.R. Furtado, J.R. Kidd, K.K. Kidd, Mini-haplotypes as lineage informative SNPs and ancestry inference SNPs, Eur. J. Hum. Genet. 20 (2012) 1148–1154, https://doi.org/10.1038/ejhg.2012.69.
[31] M.P. Donnelly, P. Paschou, E. Grigorenko, D. Gurwitz, S.Q. Mehdi, S.L.B. Kajuna, C. Barta, S. Kungulilo, N.J. Karoma, R.-B. Lu, O.V. Zhukova, J.-J. Kim, D. Comas, M. Siniscalco, M. New, P. Li, H. Li, V.G. Manolopoulos, W.C. Speed, H. Rajeevan, A.J. Pakstis, J.R. Kidd, K.K. Kidd, The distribution and most recent common ancestor of the 17q21 inversion in humans, Am. J. Hum. Genet. 86 (2010) 161–171, https://doi.org/10.1016/j.ajhg.2010.01.007.
[32] W.C. Speed, S.P. Kang, D.P. Tuck, L.N. Harris, K.K. Kidd, Global variation in CYP2C8-CYP2C9 functional haplotypes, Pharmacogenomics J. 9 (2009) 283–290, https://doi.org/10.1038/tpj.2009.10.
[33] W.C. Speed, B.J. O'Roak, Z. Tárnok, C. Barta, A.J. Pakstis, M.W. State, K.K. Kidd, Haplotype evolution of SLITRK1, a candidate gene for Gilles de la Tourette syndrome, Am. J. Med. Genet. B Neuropsychiatr. Genet. 147B (2008) 463–466, https://doi.org/10.1002/ajmg.b.30641.
[34] J. Yamtich, W.C. Speed, E. Straka, J.R. Kidd, J.B. Sweasy, K.K. Kidd, Population-specific variation in haplotype composition and heterozygosity at the POLB locus, DNA Repair (Amst.) 8 (2009) 579–584, https://doi.org/10.1016/j.dnarep.2008.12.005.
[35] R.M. Durbin, D.L. Altshuler, R.M. Durbin, G.R. Abecasis, D.R. Bentley, A. Chakravarti, A.G. Clark, F.S. Collins, F.M. De La Vega, P. Donnelly, M. Egholm, P. Flicek, S.B. Gabriel, R.A. Gibbs, B.M. Knoppers, E.S. Lander, H. Lehrach, E.R. Mardis, G.A. McVean, D.A. Nickerson, L. Peltonen, A.J. Schafer, S.T. Sherry, J. Wang, R.K. Wilson, R.A. Gibbs, D. Deiros, M. Metzker, D. Muzny, J. Reid, D. Wheeler, J. Wang, J. Li, M. Jian, G. Li, R. Li, H. Liang, G. Tian, B. Wang, J. Wang, W. Wang, H. Yang, X. Zhang, H. Zheng, E.S. Lander, D.L. Altshuler, L. Ambrogio, T. Bloom, K. Cibulskis, T.J. Fennell, S.B. Gabriel, D.B. Jaffe, E. Shefler, C.L. Sougnez, D.R. Bentley, N. Gormley, S. Humphray, Z. Kingsbury, P. Koko-Gonzales, J. Stone, K.J. McKernan, G.L. Costa, J.K. Ichikawa, C.C. Lee, R. Sudbrak, H. Lehrach, T.A. Borodina, A. Dahl, A.N. Davydov, P. Marquardt, F. Mertes, W. Nietfeld, P. Rosenstiel, S. Schreiber, A.V. Soldatov, B. Timmermann, M. Tolzmann, M. Egholm, J. Affourtit, D. Ashworth, S. Attiya, M. Bachorski, E. Buglione, A. Burke, A. Caprio, C. Celone, S. Clark, D. Conners, B. Desany, L. Gu, L. Guccione, K. Kao, A. Kebbel, J. Knowlton, M. Labrecque, L. McDade, C. Mealmaker, M. Minderman, A. Nawrocki, F. Niazi, K. Pareja, R. Ramenani, D. Riches, W. Song, C. Turcotte, S. Wang, E.R. Mardis, R.K. Wilson, D. Dooling, L. Fulton, R. Fulton, G. Weinstock, R.M. Durbin, J. Burton, D.M. Carter, C. Churcher, A. Coffey, A. Cox, A. Palotie, M. Quail, T. Skelly, J. Stalker, H.P. Swerdlow, D. Turner, A. De Witte, S. Giles, R.A. Gibbs, D. Wheeler, M. Bainbridge, D. Challis, A. Sabo, F. Yu, J. Yu, J. Wang, X. Fang, X. Guo, R. Li, Y. Li, R. Luo, S. Tai, H. Wu, H. Zheng, X. Zheng, Y. Zhou, G. Li, J. Wang, H. Yang, G.T. Marth, E.P. Garrison, W. Huang, A. Indap, D. Kural, W.-P. Lee, W. Fung Leong, A.R. Quinlan, C. Stewart, M.P. Stromberg, A.N. Ward, J. Wu, C. Lee, R.E. Mills, X. Shi, M.J. Daly, M.A. DePristo, D.L. Altshuler, A.D. Ball, E. Banks, T. Bloom, B.L. Browning, K. Cibulskis, T.J. Fennell, K.V. Garimella, S.R. Grossman, R.E. Handsaker, M. Hanna, C. Hartl, D.B. Jaffe, A.M. Kernytsky, J.M. Korn, H. Li, J.R. Maguire, S.A. McCarroll, A. McKenna, J.C. Nemesh, A.A. Philippakis, R.E. Poplin, A. Price, M.A. Rivas, P.C. Sabeti, S.F. Schaffner, E. Shefler, I.A. Shlyakhter, D.N. Cooper, E.V. Ball, M. Mort, A.D. Phillips, P.D. Stenson, J. Sebat, V. Makarov, K. Ye, S.C. Yoon, C.D. Bustamante, A.G. Clark, A. Boyko, J. Degenhardt, S. Gravel, R.N. Gutenkunst, M. Kaganovich, A. Keinan, P. Lacroute, X. Ma, A. Reynolds, L. Clarke, P. Flicek, F. Cunningham, J. Herrero, S. Keenen, E. Kulesha, R. Leinonen, W.M. McLaren, R. Radhakrishnan, R.E. Smith, V. Zalunin, X. Zheng-Bradley, J.O. Korbel, A.M. Stütz, S. Humphray, M. Bauer, R. Keira Cheetham, T. Cox, M. Eberle, T. James, S. Kahn, L. Murray, A. Chakravarti, K. Ye, F.M. De La Vega, Y. Fu, F.C.L. Hyland, J.M. Manning, S.F. McLaughlin, H.E. Peckham, O. Sakarya, Y.A. Sun, E.F. Tsung, M.A. Batzer, M.K. Konkel, J.A. Walker, R. Sudbrak, M.W. Albrecht, V.S. Amstislavskiy, R. Herwig, D.V. Parkhomchuk, S.T. Sherry, R. Agarwala, H.M. Khouri, A.O. Morgulis, J.E. Paschall, L.D. Phan, K.E. Rotmistrovsky, R.D. Sanders, M.F. Shumway, C. Xiao, G.A. McVean, A. Auton, Z. Iqbal, G. Lunter, J.L. Marchini, L. Moutsianas, S. Myers, A. Tumian, B. Desany, J. Knight, R. Winer, D.W. Craig, S.M. Beckstrom-Sternberg, A. Christoforides, A.A. Kurdoglu, J.V. Pearson, S.A. Sinari, W.D. Tembe, D. Haussler, A.S. Hinrichs, S.J. Katzman, A. Kern, R.M. Kuhn, M. Przeworski, R.D. Hernandez, B. Howie, J.L. Kelley, S. Cord Melton, G.R. Abecasis, Y. Li, P. Anderson, T. Blackwell, W. Chen, W.O. Cookson, J. Ding, H. Min Kang, M. Lathrop, L. Liang, M.F. Moffatt, P. Scheet, C. Sidore, M. Snyder, X. Zhan, S. Zöllner, P. Awadalla, F. Casals, Y. Idaghdour, J. Keebler, E.A. Stone, M. Zilversmit, L. Jorde, J. Xing, E.E. Eichler, G. Aksay, C. Alkan, I. Hajirasouliha, F. Hormozdiari, J.M. Kidd, S. Cenk Sahinalp, P.H. Sudmant, E.R. Mardis, K. Chen, A. Chinwalla, L. Ding, D.C. Koboldt, M.D. McLellan, D. Dooling, G. Weinstock, J.W. Wallis, M.C. Wendl, Q. Zhang, R.M. Durbin, C.A. Albers, Q. Ayub, S. Balasubramaniam, J.C. Barrett, D.M. Carter, Y. Chen, D.F. Conrad, P. Danecek, E.T. Dermitzakis, M. Hu, N. Huang, M.E. Hurles, H. Jin, L. Jostins, T.M. Keane, S. Quang Le, S. Lindsay, Q. Long, D.G. MacArthur, S.B. Montgomery, L. Parts, J. Stalker, C. Tyler-Smith, K. Walter, Y. Zhang, M.B. Gerstein, M. Snyder, A. Abyzov, S. Balasubramanian, R. Bjornson, J. Du, F. Grubert, L. Habegger, R. Haraksingh, J. Jee, E. Khurana, H.Y.K. Lam, J. Leng, X. Jasmine Mu, A.E. Urban, Z. Zhang, Y. Li, R. Luo, G.T. Marth, E.P. Garrison, D. Kural, A.R. Quinlan, C. Stewart, M.P. Stromberg, A.N. Ward, J. Wu, C. Lee, R.E. Mills, X. Shi, S.A. McCarroll, E. Banks, M.A. DePristo, R.E. Handsaker, C. Hartl, J.M. Korn, H. Li, J.C. Nemesh, J. Sebat, V. Makarov, K. Ye, S.C. Yoon, J. Degenhardt, M. Kaganovich, L. Clarke, R.E. Smith, X. Zheng-Bradley, J.O. Korbel, S. Humphray, R. Keira Cheetham, M. Eberle, S. Kahn, L. Murray, K. Ye, F.M. De La Vega, Y. Fu, H.E. Peckham, Y.A. Sun, M.A. Batzer, M.K. Konkel, J.A. Walker, C. Xiao, Z. Iqbal, B. Desany, T. Blackwell, M. Snyder, J. Xing, E.E. Eichler, G. Aksay, C. Alkan, I. Hajirasouliha, F. Hormozdiari, J.M. Kidd, K. Chen, A. Chinwalla, L. Ding, M.D. McLellan, J.W. Wallis, M.E. Hurles, D.F. Conrad, K. Walter, Y. Zhang, M.B. Gerstein, M. Snyder, A. Abyzov, J. Du, F. Grubert, R. Haraksingh, J. Jee, E. Khurana, H.Y.K. Lam, J. Leng, X. Jasmine Mu, A.E. Urban, Z. Zhang, R.A. Gibbs, M. Bainbridge, D. Challis, C. Coafra, H. Dinh, C. Kovar, S. Lee, D. Muzny, L. Nazareth, J. Reid, A. Sabo, F. Yu, J. Yu, G.T. Marth, E.P. Garrison, A. Indap, W. Fung Leong, A.R. Quinlan, C. Stewart, A.N. Ward, J. Wu, K. Cibulskis, T.J. Fennell, S.B. Gabriel, K.V. Garimella, C. Hartl, E. Shefler, C.L. Sougnez, J. Wilkinson, A.G. Clark, S. Gravel, F. Grubert, L. Clarke, P. Flicek, R.E. Smith, X. Zheng-Bradley, S.T. Sherry, H.M. Khouri, J.E. Paschall, M.F. Shumway, C. Xiao, G.A. McVean, S.J. Katzman, G.R. Abecasis, T. Blackwell, E.R. Mardis, D. Dooling, L. Fulton, R. Fulton, D.C. Koboldt, R.M. Durbin, S. Balasubramaniam, A. Coffey, T.M. Keane, D.G. MacArthur, A. Palotie, C. Scott, J. Stalker, C. Tyler-Smith, M.B. Gerstein, S. Balasubramanian, A. Chakravarti, B.M. Knoppers, G.R. Abecasis, C.D. Bustamante, N. Gharani, R.A. Gibbs, L. Jorde, J.S. Kaye, A. Kent, T. Li, A.L. McGuire, G.A. McVean, P.N. Ossorio, C.N. Rotimi, Y. Su, L.H. Toji, C. Tyler-Smith, L.D. Brooks, A.L. Felsenfeld, J.E. McEwen, A. Abdallah, C.R. Juenger, N.C. Clemm, F.S. Collins, A. Duncanson, E.D. Green, M.S. Guyer, J.L. Peterson, A.J. Schafer, G.R. Abecasis, D.L. Altshuler, A. Auton, L.D. Brooks, R.M. Durbin, R.A. Gibbs, M.E. Hurles, G.A. McVean, A map of human genome variation from population-scale sequencing, Nature 467 (2010) 1061–1073, https://doi.org/10.1038/nature09534.
[36] P.H. Sudmant, T. Rausch, E.J. Gardner, R.E. Handsaker, A. Abyzov, J. Huddleston, Y. Zhang, K. Ye, G. Jun, M. Hsi-Yang Fritz, M.K. Konkel, A. Malhotra, A.M. Stütz, X. Shi, F. Paolo Casale, J. Chen, F. Hormozdiari, G. Dayama, K. Chen, M. Malig, M.J.P. Chaisson, K. Walter, S. Meiers, S. Kashin, E. Garrison, A. Auton, H.Y.K. Lam, X. Jasmine Mu, C. Alkan, D. Antaki, T. Bae, E. Cerveira, P. Chines, Z. Chong, L. Clarke, E. Dal, L. Ding, S. Emery, X. Fan, M. Gujral, F. Kahveci, J.M. Kidd, Y. Kong, E.-W. Lameijer, S. McCarthy, P. Flicek, R.A. Gibbs, G. Marth, C.E. Mason, A. Menelaou, D.M. Muzny, B.J. Nelson, A. Noor, N.F. Parrish, M. Pendleton, A. Quitadamo, B. Raeder, E.E. Schadt, M. Romanovitch, A. Schlattl, R. Sebra, A.A. Shabalin, A. Untergasser, J.A. Walker, M. Wang, F. Yu, C. Zhang, J. Zhang, X. Zheng-Bradley, W. Zhou, T. Zichner, J. Sebat, M.A. Batzer, S.A. McCarroll, R.E. Mills, M.B. Gerstein, A. Bashir, O. Stegle, S.E. Devine, C. Lee, E.E. Eichler, J.O. Korbel, J.O. Korbel, An integrated map of structural variation in 2,504 human genomes, Nature 526 (2015) 75–81, https://doi.org/10.1038/nature15394.
[37] A. Auton, G.R. Abecasis, D.M. Altshuler, R.M. Durbin, G.R. Abecasis, D.R. Bentley, et al., A global reference for human genetic variation, Nature 526 (2015) 68–74, https://doi.org/10.1038/nature15393.
[38] L. Jin, P.A. Underhill, V. Doctor, R.W. Davis, P. Shen, L.L. Cavalli-Sforza, P.J. Oefner, Distribution of haplotypes from a chromosome 21 region distinguishes multiple prehistoric human migrations, Proc. Natl. Acad. Sci. 96 (1999) 3796–3800, https://doi.org/10.1073/pnas.96.7.3796.
[39] J.R. Kidd, F. Friedlaender, A.J. Pakstis, M. Furtado, R. Fang, X. Wang, C.M. Nievergelt, K.K. Kidd, Single nucleotide polymorphisms and haplotypes in Native American populations, Am. J. Phys. Anthropol. 146 (2011) 495–502, https://doi.org/10.1002/ajpa.21560.
[40] C.M. Schlebusch, H. Soodyall, Extensive population structure in San, Khoe, and mixed ancestry populations from southern Africa revealed by 44 short 5-SNP haplotypes, Hum. Biol. 84 (2012) 695–724, https://doi.org/10.3378/027.084.0603.
[41] M.P. Donnelly, P. Paschou, E. Grigorenko, D. Gurwitz, C. Barta, R.-B. Lu, O.V. Zhukova, J.-J. Kim, M. Siniscalco, M. New, H. Li, S.L.B. Kajuna, V.G. Manolopoulos, W.C. Speed, A.J. Pakstis, J.R. Kidd, K.K. Kidd, A global view of the OCA2-HERC2 region and pigmentation, Hum. Genet. 131 (2012) 683–696, https://doi.org/10.1007/s00439-011-1110-x.
[42] J.M. Butler, Fundamentals of Forensic DNA Typing, (2010), https://doi.org/10.1016/B978-0-12-374999-4.00016-3.
[43] C. Phillips, A. Salas, J.J. Sánchez, M. Fondevila, A. Gómez-Tato, J. Alvarez-Dios, M. Calaza, M.C. de Cal, D. Ballard, M.V. Lareu, A. Carracedo, SNPforID Consortium, inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs, Forensic Sci. Int. Genet. 1 (2007) 273–280, https://doi.org/10.1016/j.fsigen.2007.06.008.
[44] A.J. Pakstis, W.C. Speed, R. Fang, F.C.L. Hyland, M.R. Furtado, J.R. Kidd, K.K. Kidd, SNPs for a universal individual identification panel, Hum. Genet. 127 (2010) 315–324, https://doi.org/10.1007/s00439-009-0771-1.
[45] K.K. Kidd, A.J. Pakstis, W.C. Speed, E.L. Grigorenko, S.L.B. Kajuna, N.J. Karoma, S. Kungulilo, J.-J. Kim, R.-B. Lu, A. Odunsi, F. Okonofua, J. Parnas, L.O. Schulz, O.V. Zhukova, J.R. Kidd, Developing a SNP panel for forensic identification of individuals, Forensic Sci. Int. 164 (2006) 20–32, https://doi.org/10.1016/j.forsciint.2005.11.017.
[46] K.K. Kidd, J.R. Kidd, W.C. Speed, R. Fang, M.R. Furtado, F.C.L. Hyland, A.J. Pakstis, Expanding data and resources for forensic use of SNPs in individual identification, Forensic Sci. Int. Genet. 6 (2012) 646–652, https://doi.org/10.1016/j.fsigen.2012.02.012.

[47] J.J. Sanchez, C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, C.D. Harrison, E. Musgrave-Brown, A. Salas, D. Syndercombe-Court, P.M. Schneider, A. Carracedo, N. Morling, A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 27 (2006) 1713–1724, https://doi.org/10.1002/elps.200500671.
[48] B.F.B. Algee-Hewitt, M.D. Edge, J. Kim, J.Z. Li, N.A. Rosenberg, Individual identifiability predicts population identifiability in forensic microsatellite markers, Curr. Biol. 26 (2016) 935–942, https://doi.org/10.1016/j.cub.2016.01.065.
[49] C. Phillips, Forensic genetic analysis of bio-geographical ancestry, Forensic Sci. Int. Genet. 18 (2015) 49–65, https://doi.org/10.1016/j.fsigen.2015.05.012.
[50] O. Bulbul, W.C. Speed, C. Gurkan, U. Soundararajan, H. Rajeevan, A.J. Pakstis, K.K. Kidd, Improving ancestry distinctions among Southwest Asian populations, Forensic Sci. Int. Genet. 35 (2018) 14–20, https://doi.org/10.1016/j.fsigen.2018.03.010.
[51] M. Fondevila, C. Phillips, C. Santos, A. Freire Aradas, P.M. Vallone, J.M. Butler, M.V. Lareu, A. Carracedo, Revision of the SNPforID 34-plex forensic ancestry test: assay enhancements, standard reference sample genotypes and extended population studies, Forensic Sci. Int. Genet. 7 (2013) 63–74, https://doi.org/10.1016/j.fsigen.2012.06.007.
[52] R. Kosoy, R. Nassir, C. Tian, P.A. White, L.M. Butler, G. Silva, R. Kittles, M.E. Alarcon-Riquelme, P.K. Gregersen, J.W. Belmont, F.M. De La Vega, M.F. Seldin, Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America, Hum. Mutat. 30 (2009) 69–78, https://doi.org/10.1002/humu.20822.
[53] K.B. Gettings, R. Lai, J.L. Johnson, M.A. Peck, J.A. Hart, H. Gordish-Dressman, M.S. Schanfield, D.S. Podini, A 50-SNP assay for biogeographic ancestry and phenotype prediction in the U.S. population, Forensic Sci. Int. Genet. 8 (2014) 101–108, https://doi.org/10.1016/j.fsigen.2013.07.010.
[54] O. Bulbul, L. Cherni, H. Khodjet-el-khil, H. Rajeevan, K.K. Kidd, Evaluating a subset of ancestry informative SNPs for discriminating among Southwest Asian and circum-Mediterranean populations, Forensic Sci. Int. Genet. 23 (2016) 153–158, https://doi.org/10.1016/j.fsigen.2016.04.010.
[55] C.-X. Li, A.J. Pakstis, L. Jiang, Y.-L. Wei, Q.-F. Sun, H. Wu, O. Bulbul, P. Wang, L.-L. Kang, J.R. Kidd, K.K. Kidd, A panel of 74 AISNPs: improved ancestry inference within Eastern Asia, Forensic Sci. Int. Genet. 23 (2016) 101–110, https://doi.org/10.1016/j.fsigen.2016.04.002.
[56] O. Bulbul, G. Filoglu, T. Zorlu, H. Altuncul, A. Freire-Aradas, J. Söchtig, Y. Ruiz, M. Klintschar, S. Triki-Fendri, A. Rebai, C. Phillips, M.V. Lareu, Á. Carracedo, P.M. Schneider, Inference of biogeographical ancestry across central regions of Eurasia, Int. J. Legal Med. 130 (2016) 73–79, https://doi.org/10.1007/s00414-015-1246-7.
[57] K.K. Kidd, W.C. Speed, A.J. Pakstis, M.R. Furtado, R. Fang, A. Madbouly, M. Maiers, M. Middha, F.R. Friedlaender, J.R. Kidd, Progress toward an efficient panel of SNPs for ancestry inference, Forensic Sci. Int. Genet. 10 (2014) 23–32, https://doi.org/10.1016/j.fsigen.2014.01.002.
[58] C. Phillips, W. Parson, B. Lundsberg, C. Santos, A. Freire-Aradas, M. Torres, M. Eduardoff, C. Børsting, P. Johansen, M. Fondevila, N. Morling, P. Schneider, Á. EUROFORGEN-NoE Consortium, A. Carracedo, M.V. Lareu, Building a forensic ancestry panel from the ground up: the EUROFORGEN Global AIM-SNP set, Forensic Sci. Int. Genet. 11 (2014) 13–25, https://doi.org/10.1016/j.fsigen.2014.02.012.
[59] C. Phillips, A. Freire Aradas, A.K. Kriegel, M. Fondevila, O. Bulbul, C. Santos, F. Serrulla Rech, M.D. Perez Carceles, Á. Carracedo, P.M. Schneider, M.V. Lareu, Eurasiaplex: a forensic SNP assay for differentiating European and South Asian ancestries, Forensic Sci. Int. Genet. 7 (2013) 359–366, https://doi.org/10.1016/j.fsigen.2013.02.010.
[60] C.M. Nievergelt, A.X. Maihofer, T. Shekhtman, O. Libiger, X. Wang, K.K. Kidd, J.R. Kidd, Inference of human continental origin and admixture proportions using a highly discriminative ancestry informative 41-SNP panel, Investig. Genet. 4 (2013) 13, https://doi.org/10.1186/2041-2223-4-13.
[61] J.R. Kidd, F.R. Friedlaender, W.C. Speed, A.J. Pakstis, F.M. De La Vega, K.K. Kidd, Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples, Investig. Genet. 2 (2011) 1, https://doi.org/10.1186/2041-2223-2-1.
[62] M. Kayser, P. de Knijff, Improving human forensics through advances in genetics, genomics and molecular biology, Nat. Rev. Genet. 12 (2011) 179–192, https://doi.org/10.1038/nrg2952.
[63] M. Kayser, Forensic DNA Phenotyping: predicting human appearance from crime scene material for investigative purposes, Forensic Sci. Int. Genet. 18 (2015) 33–48, https://doi.org/10.1016/j.fsigen.2015.02.003.
[64] M. Kayser, P.M. Schneider, DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations, Forensic Sci. Int. Genet. 3 (2009) 154–161, https://doi.org/10.1016/j.fsigen.2009.01.012.
[65] L. Chaitanya, K. Breslin, S. Zuñiga, L. Wirken, E. Popiech, M. Kukla-Bartoszek, T. Sijen, P. De Knijff, F. Liu, W. Branicki, M. Kayser, S. Walsh, The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation, Forensic Sci. Int. Genet. 35 (2018) 123–135, https://doi.org/10.1016/j.fsigen.2018.04.004.
[66] S. Walsh, L. Chaitanya, L. Clarisse, L. Wirken, J. Draus-Barini, L. Kovatsi, H. Maeda, T. Ishikawa, T. Sijen, P. de Knijff, W. Branicki, F. Liu, M. Kayser, Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage, Forensic Sci. Int. Genet. 9 (2014) 150–161, https://doi.org/10.1016/j.fsigen.2013.12.006.
[67] S. Walsh, A. Lindenbergh, S.B. Zuniga, T. Sijen, P. de Knijff, M. Kayser, K.N. Ballantyne, Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence, Forensic Sci. Int. Genet. 5 (2011) 464–471, https://doi.org/10.1016/j.fsigen.2010.09.008.
[68] Y. Ruiz, C. Phillips, A. Gomez-Tato, J. Alvarez-Dios, M. Casares de Cal, R. Cruz, O. Maroñas, J. Söchtig, M. Fondevila, M.J. Rodriguez-Cid, A. Carracedo, M.V. Lareu, Further development of forensic eye color predictive tests, Forensic Sci. Int. Genet. 7 (2013) 28–40, https://doi.org/10.1016/j.fsigen.2012.05.009.
[69] S. Walsh, M. Kayser, A practical guide to the HIrisPlex system: simultaneous prediction of eye and hair color from DNA, Methods Mol. Biol. (2016), pp. 213–231, https://doi.org/10.1007/978-1-4939-3597-0_17.
[70] N.G. Crawford, D.E. Kelly, M.E.B. Hansen, M.H. Beltrame, S. Fan, S.L. Bowman, E. Jewett, A. Ranciaro, S. Thompson, Y. Lo, S.P. Pfeifer, J.D. Jensen, M.C. Campbell, W. Beggs, F. Hormozdiari, S.W. Mpoloka, G.G. Mokone, T. Nyambo, D.W. Meskel, G. Belay, J. Haut, H. NISC Comparative Sequencing Program, H. Rothschild, L. Zon, Y. Zhou, M.A. Kovacs, M. Xu, T. Zhang, K. Bishop, J. Sinclair, C. Rivas, E. Elliot, J. Choi, S.A. Li, B. Hicks, S. Burgess, C. Abnet, D.E. Watkins-Chow, E. Oceana, Y.S. Song, E. Eskin, K.M. Brown, M.S. Marks, S.K. Loftus, W.J. Pavan, M. Yeager, S. Chanock, S.A. Tishkoff, Loci associated with skin pigmentation identified in African populations, Science 358 (2017) eaan8433, https://doi.org/10.1126/science.aan8433.
[71] S. Walsh, F. Liu, A. Wollstein, L. Kovatsi, A. Ralf, A. Kosiniak-Kamysz, W. Branicki, M. Kayser, The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA, Forensic Sci. Int. Genet. 7 (2013) 98–115, https://doi.org/10.1016/j.fsigen.2012.07.005.
[72] A.A. Westen, A.S. Matai, J.F.J. Laros, H.C. Meiland, M. Jasper, W.J.F. de Leeuw, P. de Knijff, T. Sijen, Tri-allelic SNP markers enable analysis of mixed and degraded DNA samples, Forensic Sci. Int. Genet. 3 (2009) 233–241, https://doi.org/10.1016/j.fsigen.2009.02.003.
[73] C. Phillips, J. Amigo, Á. Carracedo, M.V. Lareu, Tetra-allelic SNPs: informative forensic markers compiled from public whole-genome sequence data, Forensic Sci. Int. Genet. 19 (2015) 100–106, https://doi.org/10.1016/j.fsigen.2015.06.011.
[74] L. Cherni, A.J. Pakstis, S. Boussetta, S. Elkamel, S. Frigi, H. Khodjet-El-Khil, A. Barton, E. Haigh, W.C. Speed, A. Ben Ammar Elgaaied, J.R. Kidd, K.K. Kidd, Genetic variation in Tunisia in the context of human diversity worldwide, Am. J. Phys. Anthropol. 161 (2016) 62–71, https://doi.org/10.1002/ajpa.23008.
[75] J.E. Brissenden, J.R. Kidd, B. Evsanaa, A.J. Togtokh, A.J. Pakstis, F. Friedlaender, K.K. Kidd, J.M. Roscoe, Mongolians in the genetic landscape of Central Asia: exploring the genetic relations among mongolians and other world populations, Hum. Biol. 87 (2015) 73–91, https://doi.org/10.13110/humanbiology.87.2.0005.
[76] S.A. Tishkoff, A. Goldman, F. Calafell, W.C. Speed, A.S. Deinard, B. Bonne-Tamir, J.R. Kidd, A.J. Pakstis, T. Jenkins, K.K. Kidd, A global haplotype analysis of the myotonic dystrophy locus: implications for the evolution of modern humans and for the origin of myotonic dystrophy mutations, Am. J. Hum. Genet. 62 (1998) 1389–1402, https://doi.org/10.1086/301861.
[77] S.A. Tishkoff, E. Dietzsch, W. Speed, A.J. Pakstis, J.R. Kidd, K. Cheung, B. Bonné-Tamir, A.S. Santachiara-Benerecetti, P. Moral, M. Krings, Global patterns of linkage disequilibrium at the CD4 locus and modern human origins, Science 271 (1996) 1380–1387 (Accessed 19 April 2018), http://www.ncbi.nlm.nih.gov/pubmed/8596909.
[78] H. Li, S. Gu, Y. Han, Z. Xu, A.J. Pakstis, L. Jin, J.R. Kidd, K.K. Kidd, Diversification of the ADH1B gene during expansion of modern humans, Ann. Hum. Genet. 75 (2011) 497–507, https://doi.org/10.1111/j.1469-1809.2011.00651.x.
[79] J.M. Butler, B. Budowle, P. Gill, K.K. Kidd, C. Phillips, P.M. Schneider, P.M. Vallone, N. Morling, Report on ISFG SNP Panel Discussion, Forensic Sci. Int. Genet. Suppl. Ser. (Progress in Forensic Genetics 12) 1 (2008) 471–472, https://doi.org/10.1016/j.fsigss.2007.10.159.
[80] K.K. Kidd, A.J. Pakstis, W.C. Speed, R. Lagace, J. Chang, S. Wootton, N. Ihuegbu, Microhaplotype loci are a powerful new type of forensic marker, Forensic Sci. Int. Genet. Suppl. Ser. 4 (2013) e123–e124, https://doi.org/10.1016/J.FSIGSS.2013.10.063.
[81] K.K. Kidd, A.J. Pakstis, W.C. Speed, R. Lagacé, J. Chang, S. Wootton, E. Haigh, J.R. Kidd, Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics, Forensic Sci. Int. Genet. 12 (2014) 215–224, https://doi.org/10.1016/j.fsigen.2014.06.014.
[82] K.K. Kidd, W.C. Speed, A.J. Pakstis, D.S. Podini, R. Lagacé, J. Chang, S. Wootton, E. Haigh, U. Soundararajan, Evaluating 130 microhaplotypes across a global set of 83 populations, Forensic Sci. Int. Genet. 29 (2017) 29–37, https://doi.org/10.1016/j.fsigen.2017.03.014.
[83] C. Børsting, N. Morling, Next generation sequencing and its applications in forensic genetics, Forensic Sci. Int. Genet. 18 (2015) 78–89, https://doi.org/10.1016/j.fsigen.2015.02.002.
[84] K.K. Kidd, W.C. Speed, Criteria for selecting microhaplotypes: mixture detection and deconvolution, Investig. Genet. 6 (2015) 1, https://doi.org/10.1186/s13323-014-0018-3.
[85] K.K. Kidd, W.C. Speed, S. Wootton, R. Lagace, R. Langit, E. Haigh, J. Chang, A.J. Pakstis, Genetic markers for massively parallel sequencing in forensics, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e677–e679, https://doi.org/10.1016/J.FSIGSS.2015.12.004.
[86] dbSNP, (n.d.). https://www.ncbi.nlm.nih.gov/SNP/index.html.
[87] STRBase, (n.d.). https://strbase.nist.gov/ (Accessed 23 April 2018).
[88] M.V. Osier, K.H. Cheung, J.R. Kidd, A.J. Pakstis, P.L. Miller, K.K. Kidd, ALFRED: an allele frequency database for diverse populations and DNA polymorphisms–an update, Nucleic Acids Res. 29 (2001) 317–319 (accessed March 26, 2018), http://www.ncbi.nlm.nih.gov/pubmed/11125124.
[89] M.V. Osier, K.-H. Cheung, J.R. Kidd, A.J. Pakstis, P.L. Miller, K.K. Kidd, ALFRED: an allele frequency database for anthropology, Am. J. Phys. Anthropol. 119 (2002) 77–83, https://doi.org/10.1002/ajpa.10094.

[90] H. Rajeevan, U. Soundararajan, J.R. Kidd, A.J. Pakstis, K.K. Kidd, ALFRED: an allele frequency resource for research and teaching, Nucleic Acids Res. 40 (2012) D1010–D1015, https://doi.org/10.1093/nar/gkr924.
[91] ALFRED: allele frequency database, (n.d.). https://alfred.med.yale.edu/alfred/ALFREDpreview.asp (Accessed 23 April 2018).
[92] H. Rajeevan, U. Soundararajan, A.J. Pakstis, K.K. Kidd, Introducing the forensic research/reference on genetics knowledge base, FROG-kb, Investig. Genet. 3 (2012) 18, https://doi.org/10.1186/2041-2223-3-18.
[93] FROG-kb, (n.d.). http://frog.med.yale.edu/FrogKB/ (Accessed 23 April 2018).
[94] M. Stephens, N.J. Smith, P. Donnelly, A new statistical method for haplotype reconstruction from population data, Am. J. Hum. Genet. 68 (2001) 978–989, https://doi.org/10.1086/319501.
[95] M. Stephens, P. Donnelly, A comparison of bayesian methods for haplotype reconstruction from population genotype data, Am. J. Hum. Genet. 73 (2003) 1162–1169, https://doi.org/10.1086/379378.
[96] J. Marchini, D. Cutler, N. Patterson, M. Stephens, E. Eskin, E. Halperin, S. Lin, Z.S. Qin, H.M. Munro, G.R. Abecasis, P. Donnelly, International HapMap Consortium, A comparison of phasing algorithms for trios and unrelated individuals, Am. J. Hum. Genet. 78 (2006) 437–450, https://doi.org/10.1086/500808.
[97] C. Turchi, M. Pesaresi, A. Tagliabracci, A microhaplotypes panel for forensic genetics using massive parallel sequencing, Forensic Sci. Int. Genet. Suppl. Ser. 6 (2017) e117–e118, https://doi.org/10.1016/J.FSIGSS.2017.09.035.
[98] J. Zhu, P. Chen, S. Qu, H. Wang, D. Chen, W. Liang, L. Zhang, Genotyping microhaplotype markers through massively parallel sequencing, Forensic Sci. Int. Genet. Suppl. Ser. 6 (2017) e314–e316, https://doi.org/10.1016/J.FSIGSS.2017.09.128.
[99] H. Wang, J. Zhu, N. Zhou, Y. Jiang, L. Wang, W. He, D. Peng, Q. Su, J. Mao, D. Chen, W. Liang, L. Zhang, NGS technology makes microhaplotype a potential forensic marker, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e233–e234, https://doi.org/10.1016/J.FSIGSS.2015.09.093.
[100] Y. Pu, P. Chen, J. Zhu, Y. Jiang, Q. Li, T. Feng, W. liang, L. Zhang, Microhaplotype: ability of personal identification and being ancestry informative marker, Forensic Sci. Int. Genet. Suppl. Ser. 6 (2017) e442–e444, https://doi.org/10.1016/J.FSIGSS.2017.09.144.
[101] HUGO Gene Nomenclature Committee, (n.d.). http://www.genenames.org/ (Accessed 1 June 2018).
[102] K.K. Kidd, Proposed nomenclature for microhaplotypes, Hum. Genomics 10 (2016) 16, https://doi.org/10.1186/s40246-016-0078-y.
[103] K.K. Kidd, A.M. Bowcock, P.L. Pearson, J. Schmidtke, H.F. Willard, R.K. Track, F. Ricciuti, Report of the committee on human gene mapping by recombinant DNA techniques, Cytogenet. Cell Genet. 49 (1988) 132–218, https://doi.org/10.1159/000132664.
[104] P. Chen, C. Yin, Z. Li, Y. Pu, Y. Yu, P. Zhao, D. Chen, W. Liang, L. Zhang, F. Chen, Evaluation of the Microhaplotypes panel for DNA mixture analyses, Forensic Sci. Int. Genet. 35 (2018) 149–155, https://doi.org/10.1016/j.fsigen.2018.05.003.
[105] P. Chen, W. Zhu, F. Tong, Y. Pu, Y. Yu, S. Huang, Z. Li, L. Zhang, W. Liang, F. Chen, Identifying novel microhaplotypes for ancestry inference, Int. J. Legal Med. (2018), https://doi.org/10.1007/s00414-018-1881-x.
[106] K.J. van der Gaag, R.H. de Leeuw, J.F.J. Laros, J.T. den Dunnen, P. de Knijff, Short hypervariable microhaplotypes: a novel set of very short high discriminating power loci without stutter artefacts, Forensic Sci. Int. Genet. 35 (2018) 169–175, https://doi.org/10.1016/j.fsigen.2018.05.008.
[107] N.A. Rosenberg, L.M. Li, R. Ward, J.K. Pritchard, Informativeness of genetic markers for inference of ancestry, Am. J. Hum. Genet. 73 (2003) 1402–1422, https://doi.org/10.1086/380416.
[108] C.W. Schmid, P.L. Deininger, Sequence organization of the human genome, Cell. 6 (1975) 345–358, https://doi.org/10.1016/0092-8674(75)90184-1.
[109] M.A. Batzer, P.L. Deininger, Alu repeats and human genomic diversity, Nat. Rev. Genet. 3 (2002) 370, https://doi.org/10.1038/nrg798.
[110] C. Alkan, B.P. Coe, E.E. Eichler, Genome structural variation discovery and genotyping, Nat. Rev. Genet. 12 (2011) 363, https://doi.org/10.1038/nrg2958.
[111] O. Bulbul, A.J. Pakstis, U. Soundararajan, C. Gurkan, J.E. Brissenden, J.M. Roscoe, B. Evsanaa, A. Togtokh, P. Paschou, E.L. Grigorenko, D. Gurwitz, S. Wootton, R. Lagace, J. Chang, W.C. Speed, K.K. Kidd, Ancestry inference of 96 population samples using microhaplotypes, Int. J. Legal Med. 132 (2017) 703–711, https://doi.org/10.1007/s00414-017-1748-6.
[112] J.Z. Li, D.M. Absher, H. Tang, A.M. Southwick, A.M. Casto, S. Ramachandran, H.M. Cann, G.S. Barsh, M. Feldman, L.L. Cavalli-Sforza, R.M. Myers, Worldwide human relationships inferred from genome-wide patterns of variation, Science 319 (2008) 1100–1104, https://doi.org/10.1126/science.1153717.
[113] K.B. Gettings, K.M. Kiesler, P.M. Vallone, Performance of a next generation sequencing SNP assay on degraded DNA, Forensic Sci. Int. Genet. 19 (2015) 1–9, https://doi.org/10.1016/j.fsigen.2015.04.010.
[114] R. Daniel, C. Santos, C. Phillips, M. Fondevila, R.A.H. van Oorschot, Á. Carracedo, M.V. Lareu, D. McNevin, A SNaPshot of next generation sequencing for forensic SNP analysis, Forensic Sci. Int. Genet. 14 (2015) 50–60, https://doi.org/10.1016/j.fsigen.2014.08.013.
[115] K.B. Gettings, K.M. Kiesler, S.A. Faith, E. Montano, C.H. Baker, B.A. Young, R.A. Guerrieri, P.M. Vallone, Sequence variation of 22 autosomal STR loci detected by next generation sequencing, Forensic Sci. Int. Genet. 21 (2016) 15–21, https://doi.org/10.1016/j.fsigen.2015.11.005.
[116] C.-S. Ku, E.Y. Loy, Y. Pawitan, K.-S. Chia, Next generation sequencing technologies and their applications, ELS, American Cancer Society, (2010), https://doi.org/10.1002/9780470015902.a0022508.
[117] R.A. Aponte, K.B. Gettings, D.L. Dueller, M.D. Coble, P.M. Vallone, Sequence-based analysis of stutter at STR loci: characterization and utility, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e456–e458, https://doi.org/10.1016/J.FSIGSS.2015.09.181.
[118] A.S. Fraser, An introduction to population genetic theory. By J. F. Crow and M. Kimura. Harper and Row, New York. 656 pp. 1970, Teratology 5 (1972) 386–387, https://doi.org/10.1002/tera.1420050318.
[119] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (2000) 945–959 (Accessed 20 April 2018), http://www.ncbi.nlm.nih.gov/pubmed/10835412.
[120] L. Porras-Hurtado, Y. Ruiz, C. Santos, C. Phillips, Á. Carracedo, M.V. Lareu, An overview of STRUCTURE: applications, parameter settings, and supporting software, Front. Genet. 4 (2013) 98, https://doi.org/10.3389/fgene.2013.00098.
[121] C. Santos, C. Phillips, A. Gomez-Tato, J. Alvarez-Dios, Á. Carracedo, M.V. Lareu, Inference of ancestry in forensic analysis II: analysis of genetic data, Methods Mol. Biol. (2016), pp. 255–285, https://doi.org/10.1007/978-1-4939-3597-0_19.
[122] A.L. Price, N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, D. Reich, Principal components analysis corrects for stratification in genome-wide association studies, Nat. Genet. 38 (2006) 904–909, https://doi.org/10.1038/ng1847.
[123] J. Edward Jackson, N. York Chichester Brisbane Toronto, A User's Guide To Principal Components, (n.d.). https://pdfs.semanticscholar.org/70a9/769e8027b48bce5a9ce3a82d0ce88e0f0730.pdf (Accessed 20 April 2018).
[124] K.K. Kidd, A.J. Pakstis, W.C. Speed, R. Lagace, S. Wootton, J. Chang, Selecting microhaplotypes optimized for different purposes, Electrophoresis. 2018 (n.d.). doi:10.1002/elps.201800092.
[125] N. Hiroaki, F. Koji, K. Tetsushi, S. Kazumasa, N. Hiroaki, S. Kazuyuki, Approaches for identifying multiple-SNP haplotype blocks for use in human identification, Leg. Med. (Tokyo). 17 (2015) 415–420, https://doi.org/10.1016/j.legalmed.2015.06.003.
[126] L.D. Bennett, K.E. Long, R. Walter, S.C. Wootton, C.-W. Chang, R. Legace, K.K. Kidd, D.S. Podini, The use of microhaplotypes in the analysis and deconvolution of mixed DNA, Samples 510 (2017) (Accessed 12 January 2018), https://www.aafs.org/wp-content/uploads/2017Proceedings.pdf.
[127] F. Oldoni, R. Hart, K. Long, K. Maddela, S. Cisana, M. Schanfield, S. Wootton, J. Chang, R. Lagace, R. Hasegawa, K. Kidd, D. Podini, Microhaplotypes for ancestry prediction, Forensic Sci. Int. Genet. Suppl. Ser. 6 (2017) 513–515, https://doi.org/10.1016/j.fsigss.2017.09.209.
[128] L. Voskoboinik, U. Motro, A. Darvasi, Facilitating complex DNA mixture interpretation by sequencing highly polymorphic haplotypes, Forensic Sci. Int. Genet. 35 (2018) 136–140, https://doi.org/10.1016/j.fsigen.2018.05.001.
[129] A.J. Pakstis, W.C. Speed, J.R. Kidd, K.K. Kidd, Candidate SNPs for a universal individual identification panel, Hum. Genet. 121 (2007) 305–317, https://doi.org/10.1007/s00439-007-0342-2.
[130] N. Bose, K. Carlberg, G. Sensabaugh, H. Erlich, C. Calloway, Target capture enrichment of nuclear SNP markers for massively parallel sequencing of degraded and mixed samples, Forensic Sci. Int. Genet. 34 (2018) 186–196, https://doi.org/10.1016/j.fsigen.2018.01.010.
[131] B.S. Weir, A.D. Anderson, A.B. Hepler, Genetic relatedness analysis: modern data and new challenges, Nat. Rev. Genet. 7 (2006) 771–780, https://doi.org/10.1038/nrg1960.
[132] C. Thiede, M. Florek, M. Bornhäuser, M. Ritter, B. Mohr, C. Brendel, G. Ehninger, A. Neubauer, Rapid quantification of mixed chimerism using multiplex amplification of short tandem repeat markers and fluorescence detection, Bone Marrow Transplant. 23 (1999) 1055, https://doi.org/10.1038/sj.bmt.1701779.
[133] S.A. Schichman, P. Suess, A.M. Vertino, P.S. Gray, Comparison of short tandem repeat and variable number tandem repeat genetic markers for quantitative determination of allogeneic bone marrow transplant engraftment, Bone Marrow Transplant. 29 (2002) 243, https://doi.org/10.1038/sj.bmt.1703360.
[134] C. Thiede, M. Bornhäuser, U. Oelschlägel, C. Brendel, R. Leo, H. Daxberger, B. Mohr, M. Florek, F. Kroschinsky, G. Geissler, R. Naumann, M. Ritter, G. Prange-Krex, T. Lion, A. Neubauer, G. Ehninger, Sequential monitoring of chimerism and detection of minimal residual disease after allogeneic blood stem cell transplantation (BSCT) using multiplex PCR amplification of short tandem repeat-markers, Leukemia 15 (2001) 293, https://doi.org/10.1038/sj.leu.2401953.
[135] M. Debeljak, D.N. Freed, J.A. Welch, L. Haley, K. Beierl, B.S. Iglehart, A. Pallavajjala, C.D. Gocke, M.S. Leffell, M.-T. Lin, J. Pevsner, S.J. Wheelan, J.R. Eshleman, Haplotype counting by next-generation sequencing for ultra-sensitive human DNA detection, J. Mol. Diagn. 16 (2014) 495–503, https://doi.org/10.1016/j.jmoldx.2014.04.003.
[136] M. Debeljak, E. Mocci, M.C. Morrison, A. Pallavajjalla, K. Beierl, M. Amiel, M. Noë, L.D. Wood, M.-T. Lin, C.D. Gocke, A.P. Klein, E.J. Fuchs, R.J. Jones, J.R. Eshleman, Haplotype counting for sensitive chimerism testing: potential for early leukemia relapse detection, J. Mol. Diagn. 19 (2017) 427–436, https://doi.org/10.1016/j.jmoldx.2017.01.005.
[137] Q. Ning-, X. Yifan, L. Haiyan, L. Hao-, L. Shaobin, H. Erwen, G. Jun-, C. Fang-, S. Yanwei, O. Xueling, Noninvasive prenatal paternity testing using targeted massively parallel sequencing, Transfusion 2018 (58) (2018) 1792–1799, https://doi.org/10.1111/trf.14577.
[138] D.S. Baetscher, A.J. Clemento, T.C. Ng, E.C. Anderson, J.C. Garza, Microhaplotypes provide increased power from short-read DNA sequences for relationship inference, Mol. Ecol. Resour. 18 (2018) 296–305, https://doi.org/10.1111/1755-0998.12737.
[139] E.P. Palkovacs, K. Reid, J.C. Garza, S. Gephard, D.M. Post, Determining the Effects of Landlocked Alewives on Anadromous Alewife Restoration, (n.d.). https://rcngrants.org/sites/default/files/final_reports/RCN 2015-1 Final Report.pdf (Accessed 10 July 2018).