Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1079 (2005) 41–49
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Review

Haplotyping by capillary electrophoresis

Eszter Szantai [a,b], Zsolt Ronai [a], Agnes Szilagyi [a], Maria Sasvari-Szekely [a], András Guttman [b,*]

[a] *Department of Medical Chemistry, Molecular Biology and Pathobiochemistry, Semmelweis University, Budapest, Hungary*
[b] *Horváth Laboratory of Bioseparation Sciences, Institute of Analytical Chemistry and Radiochemistry, University of Innsbruck, Innrain 52/A, A-6020 Innsbruck, Austria*

Available online 19 April 2005

Abstract

The investigation of the genetic background and phenotype structures of complex diseases, such as cardiovascular or psychiatric disorders and tumors, is one of the most scrutinized fields of the post genomic era. Besides the multiplex analysis of genetic markers and polymorphisms throughout the whole genome, more and more attention is focused on the interaction between the etiological factors of these traits. Haplotype determination, rather than multiplex genotyping seems to be one of the first building blocks of this endeavor. This review focuses on the importance and theoretical background of haplotyping, and summarizes the recent examples of novel and emerging haplotyping techniques by capillary gel electrophoresis based DNA fragment analysis, a powerful tool for the examination of the inheritance of complex traits.

Keywords: Haplotype; Genotype; Polymorphism; SNP; Capillary electrophoresis

Contents

* Corresponding author. Tel.: +43 512 50427072; fax: +43 512 50427066.
E-mail address: Andras.Guttman@uibk.ac.at (A. Guttman).

0021-9673/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2005.03.078

1. Introduction

The basic rules of inheritance have been known for about one and a half centuries since Gregor Mendel published his ground-braking work entitled *Experiments in Plant Hybridization* in 1865. Nearly 100 years later, Watson and Crick described their hypothesis about the double helical structure of DNA and the possible molecular structure of chromosomes [1]. Since then, efforts within the Human Genome Project revealed the sequence of the 3 billion base pairs (bp) of the human genome. Besides the vigorous development of 'in vivo' and 'in vitro' molecular biology methods, 'in silico' approaches have also emerged. Today, not only the base sequence of any human chromosomes can be downloaded from respective internet databases [2], but also information about gene function, the spatial structure of proteins or even mutations and polymorphisms can be searched. However, one of the fundamental problems investigators are facing worldwide is the analysis of genetic factors that might contribute to common diseases. It is well known that illnesses affecting the cardiovascular system, various tumors, diabetes or psychiatric disorders show accumulation in families. Genetic and environmental effects take part in the etiology of these complex traits, and in this way it is usually very difficult to detect the possible effect of a single genetic polymorphism. Unfortunately linkage studies that were successfully used in sparser monogenic diseases cannot be applied to polygenic diseases with reasonable success rates [3]. More information about the genetic background of diseases would open up new avenues towards the understanding of disease progression and concomitantly the development of diagnostics and preventive treatments. Genetic association studies seem to be more effective, when the relationship is searched between a given trait and its expected candidate gene. However, these studies necessitate reliable and high throughput haplotyping methods applied to a large number of samples (along with phenotypic analysis), especially to highly polymorphic gene regions.

2. Polymorphism and mutation

Studying the more than 3 billion base pairs in the human genome led to the striking recognition that only a very small portion of the genetic information actually codes for proteins. The approximate number of coding genes is only about 30,000 [4,5] that can apparently generate a significantly larger number of proteins through co- and post-translational modifications and alternative splicing. The more than 20,000 proteins already identified are listed in multifarious databases [6,7].

One of the major trends in post-genomic research is the exploration of genetic variations in humans. Difference in the DNA sequence between two non-related individuals is about 0.1%, resulting in an average of one in a thousand base pairs, altogether about 3 million [8]. Based on the frequency of the allelic variants, one can distinguish polymorphisms and mutations. The frequency of the minor form of a polymorphism is greater than 1%, while mutations are the rare allelic variants [9]. Polymorphisms are considered mostly neutral or having only minor phenotypic effects, while mutation usually means drop-out or serious lesion in gene function. Polymorphisms are classified into two major groups of single nucleotide polymorphisms (SNPs) and length polymorphisms. SNPs are responsible for about 90% of person-to-person variations, i.e., approximately one in every 1000–2000 bp [10,11]. Length polymorphisms are manifested as variable number of repetitions in the DNA sequence, where the length of a repeat section and the number of repeats can both vary [4]. These polymorphisms are responsible for approximately 10% of person to person variations, and can be further categorized according to the length of the repetitive DNA sections, such as short tandem repeats (STRs) that are 1–4 bp long, or longer repetitive DNA sequences that are referred to as variable number of tandem repeats (VNTR).

Most of these polymorphisms exist in parts of the genome that, in the best of our present knowledge, does not have any effects on the phenotype. However, when polymorphisms are located in the coding regions of a gene it may change the amino acid sequence of the expressed protein, and might concomitantly manifest itself in the phenotype. Polymorphisms can also result in changes in the phenotype if they are in the promoter region of a gene, thus influencing gene expression levels [12,13].

3. What is a haplotype and why is it important to detect?

Haplotype is the relative chromosomal localization of the alleles of the polymorphic loci [14]. During the search for the genetic background of complex diseases and traits it is indispensable to examine multiple polymorphisms, i.e., haplotypes, as they can serve as very effective genetic markers. To better understand the importance of haplotyping, let us consider that locus L1 is a transcription factor-binding site, which regulates the expression of gene *X* (Fig. 1). The transcription factor binds to allele *A*, but due to the polymorphism at this site not to allele *a* in L1, therefore, respectively stimulates or does not stimulate the expression of gene *X*. Furthermore, consider that in the same block, locus L2 has a polymorphism in the coding region, which causes an amino acid change in the expressed protein (allele *B* codes the normal protein, while *b* for the defective one). Thus, in case of sample 1, where *A* and *B* are on the same chromosome, a regular amount of normal protein is generated (*A*∼*B*). The other chromosome in this instance holds alleles *a* and *b*, resulting in just a few copies of the deficient protein (*a*∼*b*). In sample 2, with alleles *a* and *B*, at L1 and L2, respectively, only a small amount of normal protein is expressed (*a*∼*B*), while *A* and *b* provides a large amount of defective variation (*A*∼*b*). As Fig. 1 exhibits, determining only the genotype of the two

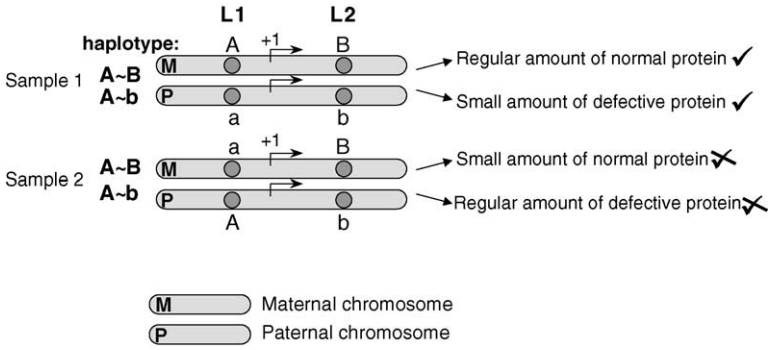

Fig. 1. Haplotype definition (+1 denotes the transcription start site).

loci in individuals 1 and 2 would not differentiate between the two heterozygote samples, as genotyping methods cannot distinguish between the '*cis*' (sample 1) and '*trans*' (sample 2) localization of the alleles. This example clearly illustrates, that haplotyping provides significantly more information on disease predisposition prediction as simple genotyping.

4. Capillary electrophoresis-based polymorphism detection methods

4.1. Length polymorphisms

Interrogation of length polymorphisms usually accomplished by amplification of the gene section of interest using polymerase chain reaction (PCR), followed by electrophoresis size separation of the resulting fragments. Two primers with annealing sites outside of the polymorphic region are utilized in the PCR, therefore the repeat number can be readily determined based on the size of the generated fragment. For example, in DNA fingerprinting the number of 4 bp long repeat sequences are investigated in the relevant sections of the genome [15]. Capillary gel electrophoresis provides more than adequate separation power and rapid analysis times for this application. More recent studies applied microfluidic devices for the separation of DNA fragments resulting in analysis times of less than a few min [16].

4.2. Single-nucleotide polymorphism

Besides the emerging use of microarrays [17] and real time PCR analysis [18,19], the most frequently used methods for single-nucleotide polymorphisms necessitate electrophoresis analysis. As capillary electrophoresis (CE) provides automation and superior resolving power even in a high throughput clinical diagnostic environment, we summarize the most frequently used CE based methods below.

4.2.1. Hybridization-based methods

One of the most frequently used methods for SNP investigation is allele-specific amplification (ASA). This technique is based on the use of an allele-specific primer in the polymerase chain reaction and subsequent electrophoresis analysis of the products [20,21]. Actually, the 3′ end of the allele-specific primer is hybridized at the SNP site, so a DNA-polymerase lacking 3′ exonuclease activity can carry out the amplification only if the primer perfectly matches with the template. Two allele-specific primers are applied to the corresponding allelic variants in two separate reactions. A novel approach allows fast and reliable genotyping of any SNP in a single tube PCR reaction followed by electrophoresis analysis [22,23].

Oligonucleotide ligation assay (OLA) applies a fluorescently labeled aspecific and an unlabeled allele-specific probe, and hybridization of the allele-specific oligo takes place only when the allelic variant of interest is present in the sample DNA. The aspecific probe is designed to hybridize to the template DNA in the adjacent position to the binding site of the specific probe. When the specific probe is complementary to the template then a nick (lack of the 3′–5′ phosphodiester bond) is formed, consequently the two DNA fragments can be ligated, resulting in the formation of a longer labeled fragment. If the allele-specific probe cannot hybridize, ligation does not occur and thus only the labeled aspecific probe is seen in the electropherogram [24]. A similar method is multiplex ligation dependent probe amplification (MLPA) where the process is initiated by the hybridization of an allele-specific and a non-specific probe, followed by a polymerase chain reaction if the two probes were ligated in the first step [25].

4.2.2. Primer extension-based methods

In primer extension-based SNP typing methods, the application of a primer ending just prior to the polymorphic region of interest is combined with a sequencing reaction using dideoxy-ribonucleotide triphosphate terminators. One approach referred to as SNaPshot [19] employs an unlabeled primer in conjunction with the four dideoxy nucleotides labeled by four different fluorescent dyes, so the color of the reaction product depicts the genotype. Multiplexing is offered in this instance by using primers with various lengths [26]. A second approach is based on the application of a fluorescent

primer in conjunction with unlabeled dideoxy and deoxy nucleotides so the genotype can be readily determined based on the DNA fragment lengths of the resulting primer extension products. Multiplexing in this instance is possible by employing different fluorophore labeled primers [27]. In both methods, the primers and the extended fragments are separated by capillary electrophoresis. DNA sequencing is also a very useful polymorphism characterization technique. In this case longer (800 bp) fragments can be sequenced applying the chain terminator fluorescently labeled dideoxy nucleotides in combination with their unlabeled deoxy counterparts [28]. It is important to note here that, although the accuracy of this method is not 100%, DNA sequencing is still considered to be the "gold standard" of SNP screening and identification by many researchers.

4.2.3. Methods based on electrophoresis mobility change

Single-stranded conformational polymorphism (SSCP) and denaturing gradient gel electrophoresis (DGGE) are the two most often used techniques that utilize the phenomenon of mobility changes during electric field mediated separations. In SSCP denatured DNA molecules are separated under carefully optimized conditions that makes possible to demonstrate the electrophoretic mobility change caused even by a single base difference [29]. The specificity and the sensitivity of the approach are significantly increased when capillary gel electrophoresis is applied. DGGE is based on the difference in electrophoretic mobilities between double-stranded (ds) DNA and partially denatured DNA. Prior to the electrophoretic analysis, the DNA fragments are denatured and re-annealed in a way that allows heteroduplex formation. In an increasingly denaturing environment along the capillary during the run, the two homoduplexes and the heteroduplexes get partially denatured at different denaturing agent concentration. Since the electrophoretic mobility of the partially denatured DNA is usually lower than that of its double stranded counterpart, the different forms are readily separated [30]. Please note that the so-called "GC-clamp" prevents complete denaturation of the fragments during capillary denaturing gradient gel electrophoresis. An alternative technique is based on the original publication of Horváth and coworkers [31] using micropellicular anion exchanger made of 3-μm rigid polystyrene-based non-porous microspheres with a covalently bound hydrophilic layer and DEAE functional groups at the surface in high-performance liquid chromatography [31].

4.2.4. Enzymatic cleavage-based methods

Restriction fragment length polymorphism can also be applied to SNP analysis. This technique utilizes type II restriction endonucleases, which recognize and cut both strands of a dsDNA at very specific 4–8 bp long palindrome sequences [14]. If the allelic variation of a single nucleotide polymorphism creates or eliminates an endonuclease recognition site, a characteristic capillary electrophoresis peak pattern is obtained, specific for the two alleles. There is a novel endonuclease based method that can identify even previously unknown SNPs. A modified repair enzyme, CEL I endonuclease, recognizes any mismatch in dsDNA and cuts the phosphodiester bond of both strands at the 3′ side. After hybridizing the sample of interest with a PCR amplicon generated from a homozygote normal sample, the presence and even the approximate position of an SNP are identified by the cleaved fragments appearing in the electropherogram [32].

5. Haplotyping methods

5.1. Family studies

One of the most obvious haplotyping methods is based on family studies, where the haplotypes are determined based on familial genotype data [33,34]. However, some multifactorial diseases are of late onset making difficult or impossible to collect DNA sample form the biological parents. Moreover family studies are not applicable in some cases, e.g., when all the members are heterozygote for the same polymorphism. These individuals must be excluded from the study, which means not only the decrease in sample size, but even the distortion of the genetic composition of the investigated population, as samples are dropped out not randomly but rather genotype dependently (i.e., mostly the heterozygotes must be excluded).

5.2. Software-based haplotype prediction

The maximum-likelihood method is effected via the expectation-maximization algorithm worked out by statisticians [35–37]. Other methods, such as the parsimony method [38], and the PHASE methods are based on Bayesian inference calculating uncertainties associated with each phase call [39,40]. In case control studies, estimated haplotype frequencies are not directly evaluated by χ^2 tests. Please note that the accuracy of haplotype estimation is not guaranteed [41] when a disease population is not under the Hardy–Weinberg equilibrium [42]. Thus, it is important to unambiguously determine individual haplotypes in a simple, direct way.

5.3. Direct haplotype determination

Direct haplotype determination, namely the application of molecular genetic methods for each sample is another fairly useful and reliable option besides family studies and the application of computer algorithms. Cloning and sequencing of DNA fragments derived from each chromosome is a classical but rather labor intensive method for molecular haplotyping. One of the reported techniques is based on double allele-specific amplification, which is suitable for the determination of haplotype of two adjacent SNPs [43]. However, when many polymorphic sites are located on the region of interest thus the number of predictable haplotypes is increased,

this method gets rather complicated. In addition, nucleotide sequences around the polymorphic sites affect the optimization of allele-specific amplification. An interesting approach is intense dilution of the DNA sample solution resulting in amplification of practically one DNA molecule, which makes possible the investigation of SNPs as far as 24 kb apart from each other [44]. Another approach by Douglas et al. [45] suggested transformation of diploid cells to haploid by construction of somatic cell hybrids. This provided an opportunity for direct haplotype determination through conventional DNA genotyping from the haploid cells. Another method suggested intra-molecular ligation of PCR fragments containing two polymorphisms of investigation. Thus, the two sites get to adjacent localization making it possible to apply allele-specific amplification [46]. Heteroduplex analysis has also been applied to molecular haplotyping [47]. Single-strand conformation polymorphism (SSCP) can be used for molecular haplotyping as well, and it is conceivable that conformation of single-stranded DNA further changes by the combination of more than one polymorphism [41].

6. Haplotyping by capillary electrophoresis

Since the importance of haplotyping become apparent, a growing number of research groups took the lead to develop rapid, sensitive and reliable methods. In this review we focus on CE-based haplotyping methods. Cook and coworkers [48] reported a CE-based SSCP haplotyping method for the investigation of two common SNPs within the gene of the EAAC1 (neuronal and epithelial glutamate transporter) that is suggested to be a functional candidate in obsessive–compulsive disorder. Fig. 2 shows six SSCP electropherograms for two

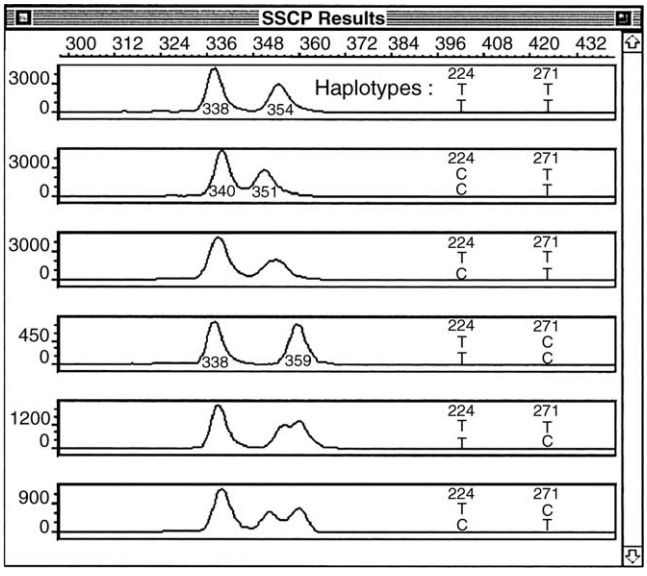

Fig. 2. Electropherograms for six SSCP patterns. Length units assigned using Genescan TAMRA 350 as internal standard. Each genotype was confirmed by cycle sequencing. With permission from [48].

SLC1A1 (the gene of the EAAC1) polymorphisms, confirmed by cycle sequencing. The peak patterns correspond to different combinations of three haplotypes of T∼T, C∼T and T∼C. The fragment size associated to each peaks are labeled on the three homozygote subjects.

Others studied possible association between SNPs and sporadic rectal cancer (SRC) [49] utilizing a multi-capillary DNA sequencing instrument and developed a method for SNPs and (micro)haplotype detection within short DNA fragments of 100–1000 bp by integrating DNA melting theory. Partial melting of DNA fragments was detected by the change in their electrophoretic mobilities. Information in melting behavior of different (micro)haplotypes relative to the wild type in combination with DNA sequencing allowed direct (micro)haplotype determination. A standard 96-capillary DNA sequencing instrument was used in cycling temperature capillary electrophoresis mode (CTCE). Fig. 3 depicts the CTCE traces representing the five (micro)haplotype combinations discovered (samples are delineated by thin, internal standards are by bold lines). Haplotypes are given as three polymorphic bases in each allele. Please note that heteroduplexes are not shown in the electropherograms. No association was found between the polymorphisms and the haplotypes in the ataxia-teleangiectasia mutated gene with respect to sporadic rectal cancer.

Serotonin-related genes in antidepressant response were investigated using a high-throughput capillary electrophoresis based SNP analysis platform for 110 SNPs and four repeat polymorphisms located in seven candidate genes by [50]. Their result revealed three SNPs in the tryptophan hydroxylase gene and one SNP in the serotonin transporter gene showing noteworthy single locus association when response to fluoxetine is compared to non-response. To investigate possible interaction of alleles from different SNPs within a gene, haplotypes were inferred using an expectation maximization algorithm and tested for association using phenotypic comparisons. The use of haplotypes, or particular combinations of alleles observed in a population, has been proved to be a powerful approach for the dissection of complex genetic traits both theoretically and empirically. Using a haplotype SNP tagging approach enabled to reduce the number of SNPs required to represent the majority of haplotypes, and generally allowed observation of even greater association between these htSNP (haplotype tagging SNP), haplotypes and phenotypes.

Association studies between the macrophage migration inhibitory factor (MIF) gene and juvenile idiopathic arthritis (JIA) by Donn et al. [51] established linkage of a 2-locus MIF promoter haplotype with JIA susceptibility using either SNaPshot ddNTP primer extension or by a fluorescently labeled primer method in conjunction with capillary gel electrophoresis. They have also demonstrated the functional consequence of this particular haplotype that was found to be transmitted in excess by measuring the plasma MIF concentration in patients with juvenile idiopathic arthritis. The application of luciferase reporter gene constructs provided

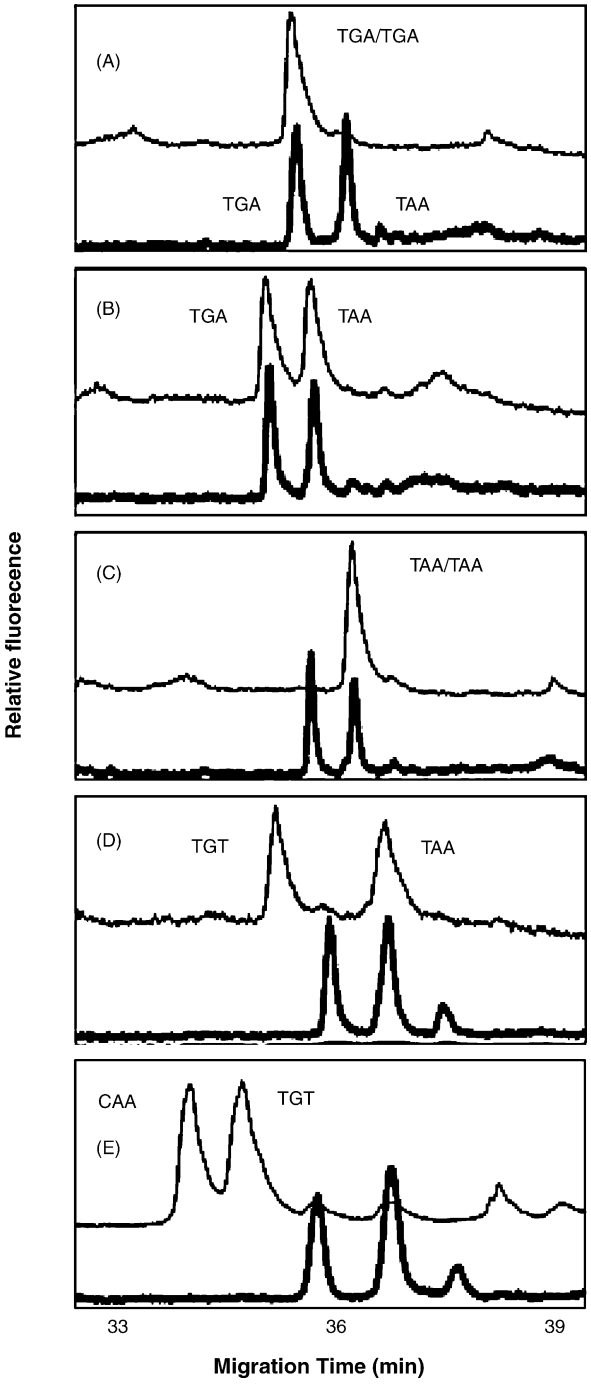

Fig. 3. Cycling temperature capillary electrophoresis of five genotypes representing five microhaplotype combinations. Sample traces are depicted with thin line and the internal standard is presented in bold. Microhaplotypes are given as the three polymorphic bases in each allele. Please note that heteroduplexes are not shown in the electropherograms. With permission from [49].

clear evidence of a functional interaction between the two sites of genetic variation within the MIF promoter, and suggested a haplotype linkage with juvenile idiopathic arthritis.

Capillary electrophoresis and PCR was utilized by Laurent-Puig and coworkers [52] in their SNP investigation

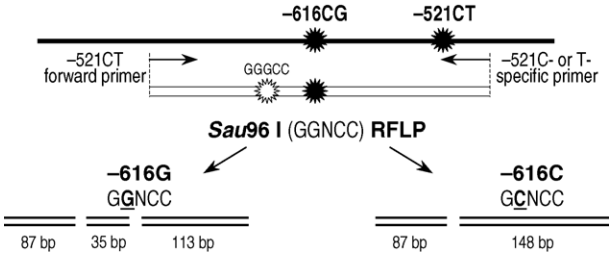

Fig. 4. Schematics of haplotype analysis using combined allele specific PCR/RFLP. Filled asterisks: position of the two SNPs, open asterisk: position of the control *Sau*96 I digestion site. With permission from [53].

of matrix metalloproteinase (MMP) 1 and 3 gene promoters as risk factors in head and neck squamous cell carcinoma. The presence of the MMP1-2G and MMP3-6A alleles seemed to be associated with decreased risk of tumor especially when the two forms were carried by the same chromosome forming a haplotype.

Our group [53] introduced a novel haplotyping approach (Fig. 4) combining allele-specific amplification and restriction fragment length polymorphism (RFLP) for the investigation of the −616CG and −521CT SNPs in the 5′ untranslated region of the Dopamine D4 Receptor gene. Fig. 5 depicts capillary gel electrophoresis haplotyping of the two SNPs of several individuals applying allele specific PCR (for the −521CT SNP) in conjunction with *Sau*96 I restriction enzyme digestion (for the −616CG SNP). The left and right panels demonstrate the digestion patterns of the PCR amplicons possessing thymine or cytidine at position −521, respectively. In sample 1, both the digested and undigested fragments are present in the −521T specific reaction, but lacks any PCR-products in the −521C specific reaction, suggesting the haplotype structure as −616C ∼ −521T/−616G ∼ −521T. Sample 2 is double heterozygote as PCR-products were created by both the −521T- and −521C-specific reactions, but as the −521C-specific product was not digested resulting in the haplotype of −616G ∼ −521T/−616C ∼ −521C. Sample 3 is also heterozygotic for site −521, as the expected PCR products were produced in both reaction mixtures and both chromosomes contain the −616G allele consequently its haplotype is −616G ∼ −521C/−616G ∼ −521T. Sample 4 is −521CC homozygote and together with the presence of digested and the undigested fragments in panel B the haplotype is assumed to be −616C ∼ −521C/−616G ∼ −521C. This study demonstrated that capillary gel electrophoresis in combination with laser induced fluorescence detection offered a sensitive and accurate tool for automated haplotyping in clinical settings.

More recently Kamio et al. [41] took the lead to compare statistical haplotyping with direct haplotype determination methods for the MUC5B (human mucin 5B gene) promoter applying capillary gel electrophoresis based single strand conformation polymorphism method. They demonstrated that SSCP can identify up to 10 common SNPs and a dinucleotide insertion/deletion site in a 2 kb long DNA.

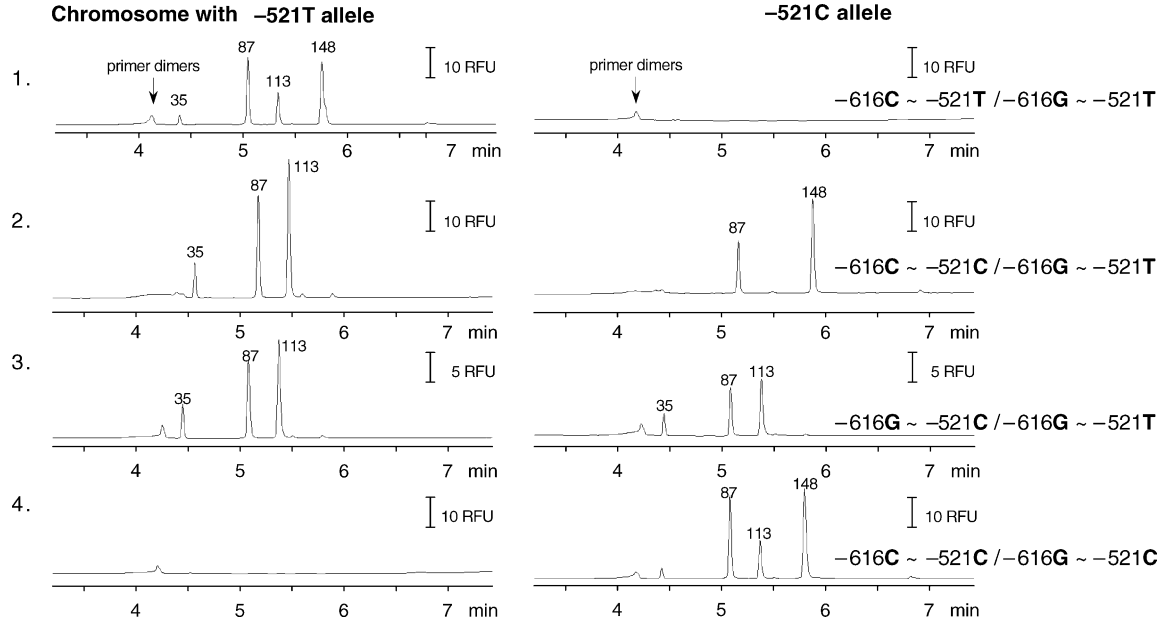

Fig. 5. Haplotype determination of the −616CG and −521CT SNPs of the Dopamine Receptor D4 gene. With permission from [53].

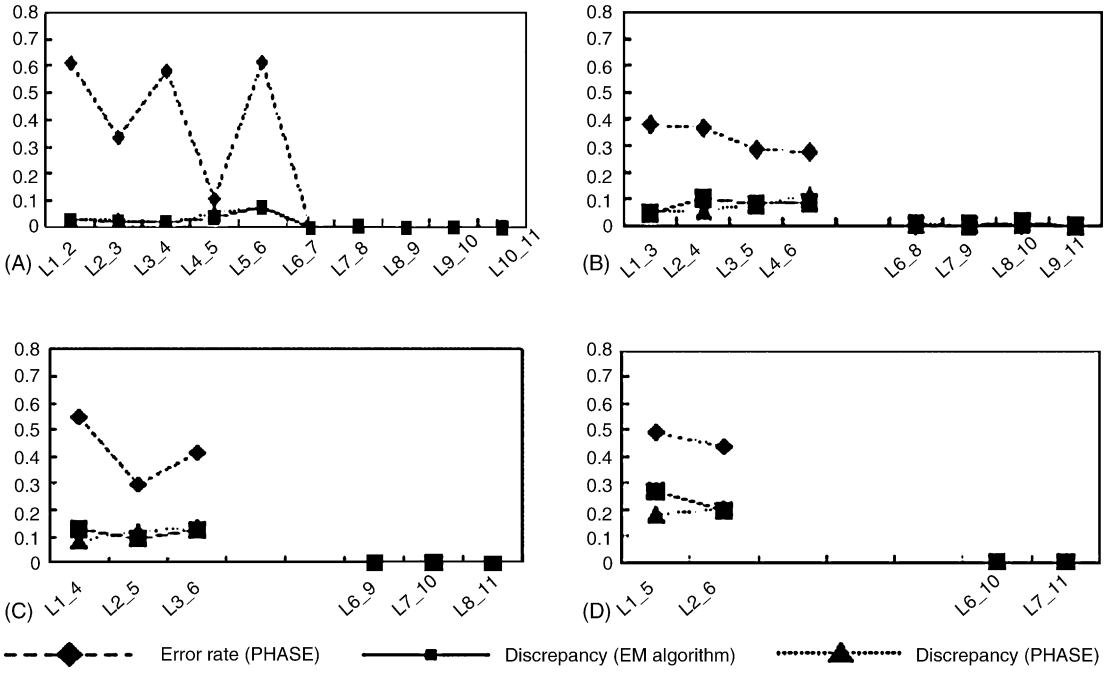

Fig. 6. Comparison of statistically estimated and experimentally determined haplotypes. Statistical estimation of haplotype frequencies was implemented for two (A), three (B), four (C), and five (D) adjacent loci in the promoter region of *MUC5B* gene. With permission from [41].

Haplotypes were reconstructed from diploid genotypes in the same region by statistical methods including expectation maximization. Fig. 6 depicts the comparison between statistical estimation and experimental haplotype determination revealing that major haplotypes containing multiple marker sites showing strong linkage disequilibrium (LD) are estimated in good accuracy however haplotypes with weak LD are not reconstructed correctly.

7. Summary

It is considered that the DNA sequence of any two human individuals is 99.9% identical. Although the frequency of all the possible variations is approximately only 0.1%, it still may affect the risk of the individuals to various diseases. Patterns of single nucleotide polymorphisms localized on the same chromosome are referred to as haplotype. In some cases

haplotypes are inherited as conserved blocks. Within these blocks, however, only a few SNPs provide appropriate haplotype information. Capillary electrophoresis and capillary array electrophoresis are playing a major role in haplotype determination, especially in clinical settings.

Acknowledgements

This work was supported by the European Community Marie Curie Chair 006733, and the national funds of T 048576, GVOP 3.1.1. 2004-05-0324/3.0, F 042730. E.S. is a PhD Student of the Classical and Molecular Genetics Program led by László Orosz at Eötvös Loránd University, Budapest.

Appendix A. The international HapMap project

To prepare a human haplotype map the so-called HapMap Project was launched by Francis Collins in 2002. This international endeavor will take approximately 3 years with the expected cost of $100 million [54]. The major goal of the project is DNA haplotype analysis of 270 individuals with different geographical origin (US Utah population with Northern and Western European ancestry with unrelated Japanese, Chinese and Nigerian individuals). It is expected that significant deviation exists among the haplotype structure of these different populations. Recombination is responsible for the development of boundary of haplotype blocks with an average of 3–150 kb in length, which can occur in a recombination 'hotspots' or can be the result of a single historical mutational event. Blocks developed in the first case are probably present in all populations, while those arouse from the second case are shared by the populations or not depending on the time passed since the recombinational event. Consequently, shorter haplotype blocks are expected in samples of African origin, since people living there had more time for genetic recombination so their blocks are probably split up more extensively. In other populations, the expected diversity is lower because those migrating from Africa had just a few genetic variants and in this way they could only proliferate these [54].

Besides analyzing haplotype blocks of the history of humankind, haplotype maps will have enormous significance in exploring the genetic background of complex traits and diseases, while the number of haplotype block variations in the populations is much less than expected from the polymorphisms forming the haplotype (2^n, where n is the number of polymorphisms in the block possessing two alleles). In this way, haplotype maps would provide a fast and cost effective way to explore genetic risk factors of multifactorial diseases. Thereby, it would only be necessary to examine just one or a few markers (SNPs) from each blocks instead of the analysis of all 3 million SNPs [2]. These results would restrict genetic analysis only for those blocks that showed linkage with the phenotype of interest [54].

Launching the HapMap project generated some concern as if genetic difference between different populations gets publicity, some groups can be stigmatized by revealing that they possess higher predisposition risk for specific illnesses. On the other hand, getting such a map of haplotype blocks is of high significance in the genetic mapping of complex diseases, in pharmacogenetic researches and in achieving personalized medicine.

References

[1] J.D. Watson, F.H.C. Crick, Nature 171 (1953) 737.
[2] ftp://ftp.ncbi.nih.gov/genomes/H_sapiens.
[3] H.M. Colhoun, P.M. McKeigue, G. Davey Smith, Lancet 361 (2003) 865.
[4] E.S. Lander, L.M. Linton, B. Birren, et al., Nature 409 (2001) 860.
[5] J.C. Venter, M.D. Adams, E.W. Myers, et al., Science 291 (2001) 1304.
[6] http://www.h-invitational.jp.
[7] http://www.hgmp.mrc.ac.uk/Bioinformatics/Databases/swissprot-help.html.
[8] R. Sachidanandam, D. Weissman, S.C. Schmidt, et al., Nature 409 (2001) 928.
[9] http://www.wellcome.ac.uk/en/genome/thegenome/hg04f001.html.
[10] J.G. Taylor, E.H. Choi, C.B. Foster, S.J. Chanock, Trends Mol. Med. 7 (2001) 507.
[11] http://www.genomenewsnetwork.org/resources/whats_a_genome/Chp4_1.shtml.
[12] M. Baron, Mol. Psychiatry 6 (2001) 143.
[13] T.R. Rebbeck, M. Spitz, X. Wu, Nat. Rev. Genet. 5 (2004) 589.
[14] B. Lewin, Genes VI, Oxford University Press, New York, 1997.
[15] P. Koumi, H.E. Green, S. Hartley, D. Jordan, S. Lahec, R.J. Livett, K.W. Tsang, D.M. Ward, Electrophoresis 25 (2004) 2227.
[16] D. Schmalzing, A. Belenky, M.A. Novotny, L. Koutny, O. Salas-Solano, S. El-Difrawy, A. Adourian, P. Matsudaira, D. Ehrlich, Nucleic Acids Res. 28 (2000) E43.
[17] E.S. Lander, Nat. Genet. 21 (1999) 3.
[18] G. Schwarz, S. Baumler, A. Block, F.G. Felsenstein, G. Wenzel, Nucleic Acids Res. 32 (2004) 24.
[19] N. Pati, V. Schowinsky, O. Kokanovic, V. Magnuson, S. Ghosh, J. Biochem. Biophys. Methods 60 (2004) 1.
[20] C.R. Newton, A. Graham, L.E. Heptinstall, S.J. Powell, C. Summers, N. Kalsheker, J.C. Smith, A.F. Markham, Nucleic Acids Res. 17 (1989) 2503.
[21] D.Y. Wu, L. Ugozzoli, B.K. Pal, R.B. Wallace, Proc. Natl. Acad. Sci. U.S.A. 86 (1989) 2757.
[22] Z. Ronai, C. Barta, A. Guttman, K. Lakatos, J. Gervai, M. Staub, M. Sasvari-Szekely, Electrophoresis 22 (2001) 1102.
[23] Z. Ronai, E. Szantai, R. Szmola, Z. Nemoda, A. Szekely, J. Gervai, A. Guttman, M. Sasvari-Szekely, Am. J. Med. Genet. B: Neuropsychiatr. Genet. 126 (2004) 74.
[24] M.M. Shi, Clin. Chem. 47 (2001) 164.
[25] J.P. Schouten, C.J. McElgunn, R. Waaijer, D. Zwijnenburg, F. Diepvens, G. Pals, Nucleic Acids Res. 30 (2002) 57.
[26] N.M. Makridakis, J.K. Reichardt, Biotechniques 31 (2001) 1374.
[27] C. Barta, Z. Ronai, M. Sasvari-Szekely, A. Guttman, Electrophoresis 22 (2001) 779.
[28] D. Schmalzing, L. Koutny, O. Salas-Solano, A. Adourian, P. Matsudaira, D. Ehrlich, Electrophoresis 20 (1999) 3066.
[29] K. Hayashi, PCR Methods Appl. 1 (1991) 34.
[30] R. Fodde, M. Losekoot, Hum. Mutat. 3 (1994) 83.
[31] Y.F. Maa, S.C. Lin, Cs. Horváth, J. Chromatogr. 508 (1990) 61.
[32] P. Qiu, H. Shandilya, J.M. D'Alessio, K. O'Connor, J. Durocher, G.F. Gerard, Biotechniques 36 (2004) 702.

[33] C.L. Barr, Y. Feng, K.G. Wigg, R. Schachar, R. Tannock, W. Roberts, M. Malone, J.L. Kennedy, Am. J. Med. Genet. 105 (2001) 84.
[34] S.E. Hodge, M. Boehnke, M.A. Spence, Nat. Genet. 21 (1999) 360.
[35] L. Excoffier, M. Slatkin, Mol. Biol. Evol. 12 (1995) 921.
[36] M.E. Hawley, K.K. Kidd, J. Hered. 86 (1995) 409.
[37] J.C. Long, R.C. Williams, M. Urbanek, Am. J. Hum. Genet. 56 (1995) 799.
[38] A.G. Clark, Mol. Biol. Evol. 7 (1990) 111.
[39] M. Stephens, N.J. Smith, P. Donnelly, Am. J. Hum. Genet. 68 (2001) 978.
[40] M. Stephens, P. Donnelly, Am. J. Hum. Genet. 73 (2003) 1162.
[41] K. Kamio, I. Matsushita, G. Tanaka, J. Ohashi, M. Hijikata, K. Nakata, K. Tokunaga, A. Azuma, S. Kudoh, N. Keicho, Genomics 84 (2004) 613.
[42] http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/H/Hardy_Weinberg.html.
[43] T.A. Greenwood, M. Alexander, P.E. Keck, S. McElroy, A.D. Sadovnick, R.A. Remick, J.R. Kelsoe, Am. J. Med. Genet. 105 (2001) 145.
[44] C. Ding, C.R. Cantor, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 7449.
[45] J.A. Douglas, M. Boehnke, E. Gillanders, J.M. Trent, S.B. Gruber, Nat. Genet. 28 (2001) 361.
[46] O.G. McDonald, E.Y. Krynetski, W.E. Evans, Pharmacogenetics 12 (2002) 93.
[47] F.M. Chang, K.K. Kidd, Am. J. Med. Genet. 74 (1997) 91.
[48] J. Veenstra-Vander Weele, S.J. Kim, D. Gonen, G.L. Hanna, B.L. Leventhal, E.H. Cook Jr., Mol. Psychiatry 6 (2001) 160.
[49] A.T. Kristensen, J. Bjorheim, J. Wiig, K.E. Giercksky, P.O. Ekstrom, Int. J. Colorectal Dis. 19 (2004) 49.
[50] E.J. Peters, S.L. Slager, P.J. McGrath, J.A. Knowles, S.P. Hamilton, Mol. Psychiatry 9 (2004) 879.
[51] R. Donn, Z. Alourfi, E. Zeggini, R. Lamb, F. Jury, M. Lunt, C. Meazza, F. De Benedetti, W. Thomson, D. Ray, Arthritis Rheum. 50 (2004) 1604.
[52] F. Zinzindohoue, H. Blons, S. Hans, M.A. Loriot, A.M. Houllier, D. Brasnu, O. Laccourreye, D.A. Tregouet, I. Stucker, P. Laurent-Puig, Anticancer Res. 24 (2004) 2021.
[53] E. Szantai, A. Szilagyi, A. Guttman, M. Sasvari-Szekely, Z. Ronai, J. Chromatogr. A 1053 (2004) 241.
[54] http://www.gnxp.com/MT2/archives/000789.html.