



Application of polymerase chain reaction-restriction fragment length polymorphism (RFLP-PCR) in the analysis of single nucleotide polymorphisms (SNPs)

BIOOPEN 2021 – POST-CONFERENCE ARTICLE

PIOTR TARACH 

University of Lodz, Faculty of Biology and Environmental Protection, Department of General Biophysics, Pomorska 141/143, 90-236 Lodz, Poland
E-mail: piotr.tarach@biol.uni.lodz.pl

ABSTRACT

Polymerase chain reaction-restriction fragment length polymorphism (RFLP-PCR) is a technique used to identify single nucleotide polymorphisms (SNPs) based on the recognition of restriction sites by restriction enzymes. RFLP-PCR is an easy-to-perform and inexpensive tool for initial analysis of SNPs potentially associated with some monogenic diseases, as well as in genotyping, genetic mapping, lineage screening, forensics and ancient DNA analysis. The RFLP-PCR method employs four steps: (1) isolation of genetic material and PCR; (2) restriction digestion of amplicons; (3) electrophoresis of digested fragments; and (4) visualisation. Despite its obsolescence and the presence of high-throughput DNA analysis techniques, it is still applied in the analysis of SNPs associated with disease entities and in the analysis of genetic variation of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). RFLP-PCR is a low-cost and low-throughput research method allowing for the analysis of SNPs in the absence of specialised equipment, and it is useful when there is a limited budget.

KEYWORDS: nucleotide polymorphisms, DNA analysis, polymerase chain reaction

Description of the polymerase chain reaction-restriction fragment length polymorphism technique and exemplary applications

Polymerase chain reaction-restriction fragment length polymorphism (RFLP-PCR) is a technique used to identify single nucleotide polymorphisms (SNPs) based on the recognition of restriction sites by restriction enzymes (Saiki *et al.*

1985). These enzymes are used to digest specific fragments of DNA, which are then separated electrophoretically on an agarose or polyacrylamide gel and visualised (Figure 1). RFLP-PCR is an easy-to-perform and inexpensive tool for initial analysis of SNPs potentially associated with some monogenic diseases, such as sickle cell anaemia (Saiki *et al.* 1985), β thalassaemia (Pramoonjago

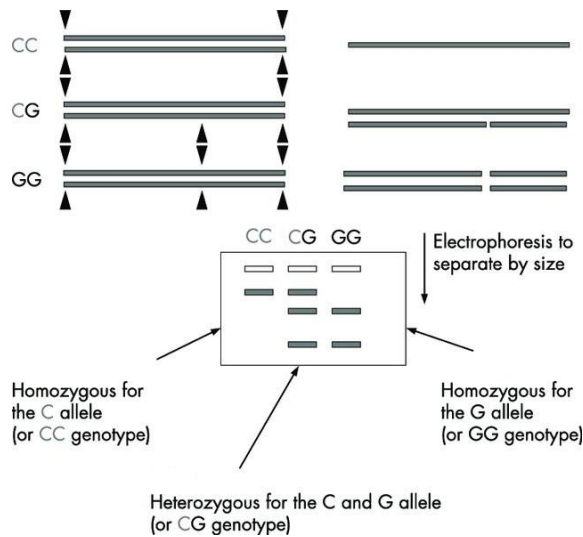


Figure 1. Restriction fragment length polymorphism (RFLP) and detection of alleles. Restriction enzyme digestion of DNA occurs at specific DNA sequences, indicated by the arrow. If a polymorphism (change in DNA sequence) occurs in a restriction enzyme site near a gene of interest, different sized molecules (corresponding to the alleles) will be produced, which are resolved during electrophoresis and subsequently visualised. Adapted with the permission of Harding (2007).

et al. 1999), coeliac disease (Endreffy *et al.* 1992, Catamo *et al.* 2015), phenylketonuria (Meijer *et al.* 1993, Kozák *et al.* 1995) or haemophilia (Křepelová *et al.* 1993, Backfisch *et al.* 1994, Stankovic *et al.* 2005, Herrmann *et al.* 2008, Tasleem Raza *et al.* 2009). RFLP-PCR is also used in genotyping, genetic mapping, pedigree testing and forensics. One of the most interesting applications of this technique is the study of ancient DNA (aDNA) from long extinct organisms, mummified tissues or preserved plant tissues (Hagelberg *et al.* 2015, Orlando *et al.* 2021).

The RFLP-PCR method comprises four steps. In the first step, the genetic material is isolated and the investigated DNA fragment of known length is pre-amplified by using PCR with a pair of specific primers. In the second step, the DNA fragment undergoes restriction digestion carried out by restriction enzymes, which recognise a 4–8 base pair restriction site. In the third step, the digested

amplicons are separated via electrophoresis. Depending on the equipment and reagents at hand, the most common type of electrophoresis is slab gel electrophoresis with either agarose or polyacrylamide as the molecular separation matrix (Berg 2012). In the fourth step, the restriction-enzyme-treated amplicons can be visualised by DNA fragment complexation with ethidium bromide (Laber *et al.* 1994) or silver (Budowle *et al.* 1991).

Restriction enzymes are a type of endonuclease that are part of the anti-infection system in bacteria. They differ from one another in the DNA sequences they recognise. Bacteria often possess several restriction enzymes, each specific to a particular short DNA sequence. Type II restriction enzymes are one of the most commonly used in RFLP. These enzymes cleaves the DNA strand at its centre, forming a blunt end, or in a staggered cut, leaving overhangs called sticky ends (Pingoud and Jeltsch 2001). Rebase® is

the most comprehensive repository of restriction enzymes, providing a list of the enzymes, accompanied by information on the restriction enzyme type, the restriction site in a given nucleotide sequence or the presence of isoschizomers. An example list of type II restriction enzymes is shown in Table 1. Chang *et al.* (2006, 2010) developed a database allowing restriction enzyme mining for SNPs in genomes. The SNP-RFLP analysis provides the SNP contig position, heterozygosity, function, protein residue and amino acid position for coding SNPs (cSNPs) as well as commercial and non-commercial restriction enzymes.

One of the many practical examples of the application of RFLP-PCR in SNP genotyping is that performed by Alavian *et al.* (2018) for the detection of the rs1127354 and rs7270101 polymorphisms associated with the inosine triphosphate pyrophosphatase (ITPA)

gene. rs1127354 and rs7270101 SNPs are associated with a functional impairment in ITPase, resulting in anaemia protection in patients with chronic hepatitis C virus (HCV) infection undergoing ribavirin (RBV)-dependent regimens. In a given study, 100 Iranian patients with chronic hepatitis C were genotyped for the detection of rs1127354 and rs7270101 polymorphisms with the help of RFLP-PCR and Sanger sequencing method to validate the results (Figure 2). The results showed that all of the 100 samples tested with PCR-RFLP and sequencing had exactly the same results, with 100% concordance. This demonstrates the utility of RFLP-PCR in studying SNPs with as high efficiency and reliability as Sanger sequencing, but at a reduced cost (Alavian *et al.* 2018).

Despite the presence of high-throughput screening techniques, RFLP-PCR is still applicable for studying the genetic variability of viruses, for

Table 1. Examples of type II restriction enzymes from the Rebase® database. The arrows point to the restriction cut site, creating blunt or sticky ends in the amplicon sequence.

Enzyme	REBASE Number	Source	Recognition sequence	Cut	Isoschizomers
AccBSI	2733	<i>Acinetobacter calcoaceticus</i> BS	5' CCGCTC	5' CCG [↓] CTC 3' 3' GGC [↑] GAG 5'	BsrBI, BstD102I, Bst31NI, MbiI
AflII	39	<i>Anabaena flosaquae</i>	5' CTTAAG	5' CTTAA [↓] G 3' 3' GAA [↑] TC 5'	BfrI, BspTI, Bst98I, BstAFI, BstPZ740I, Esp4I, MspCI, Vha464I
BamHI	185	<i>Bacillus amyloliquefaciens</i> H	5' GGATCC	5' G [↓] GATCC 3' 3' CCT [↑] AGG 5'	AccEBI, AII, ApaCI, AsiI, Bce751I, Bsp98I, Bsp4009I, BspAAIII,
EcoRI	993	<i>Escherichia coli</i> RY13	5' GAATTC	5' G [↓] AATTC 3' 3' CTT [↑] AAG 5'	Bci528I, Eco82I, Eco228I, FunII, Kpn49kI, Ppu111I
KpnI	1180	<i>Klebsiella pneumoniae</i> OK8	5' GGTACC	5' G [↓] GATCC 3' 3' CCT [↑] TGG 5'	Acc65I, AhaB8I, Asp718I, SthI
HindIII	1151	<i>Haemophilus influenzae</i> Rd	5' AAGCTT	5' A [↓] AGCTT 3' 3' T [↑] TCGA 5'	Asp52I, Asp3065I, BspLAIH, Cfr32I, HinJCI, LlaCI,

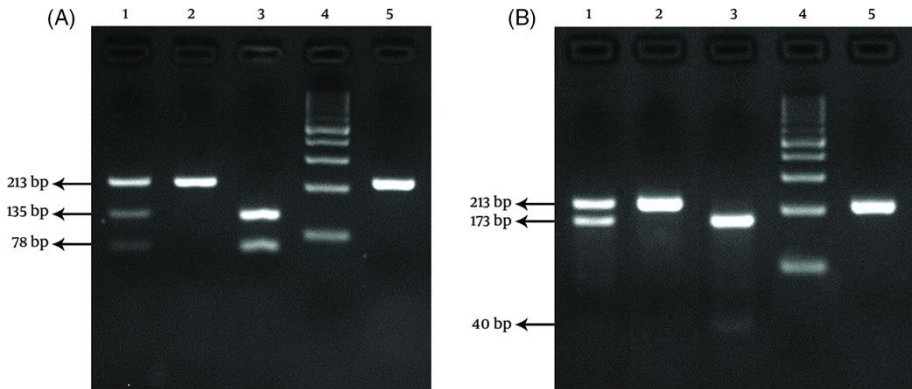


Figure 2. The result of polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) products after digestion by the XceI and MboII enzymes. (A) Gel electrophoresis results of PCR-RFLP products after digestion for rs1127354 by the XceI enzyme. Lanes 1, 2 and 3 were genotyped as CA, CC, and AA, respectively. Lane 4 indicates the 100-bp gene ruler. Lane 5 is non-digested PCR product. (B) Gel electrophoresis results of PCR-RFLP products after digestion for rs7270101 by the MboII enzyme. Lanes 1, 2, and 3 were genotyped as AC, AA, and CC, respectively. Lane 4 indicates the 100-bp molecular gene ruler. Lane 5 is non-digested PCR product. Reprinted with permission from Alavian *et al.* (2018).

example, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the cause of the COVID-19 pandemic. The D614G mutation is characterised by an amino acid substitution of aspartic acid for glycine at position 614 of the spike glycoprotein (S) amino acid sequence. This leads to increased binding to the human cell-surface receptor angiotensin-converting enzyme 2 (ACE2), resulting in increased viral replication in upper respiratory tract cells, thus increasing viral transmissibility (Plante *et al.* 2020). The D614G variant is currently the most prevalent SARS-CoV-2 variant in the world (Zhou *et al.* 2021).

Hashemi *et al.* (2020) developed an RFLP-PCR protocol to detect the D614G mutation in SARS-CoV-2 using bioinformatics and software tools. DNA samples from 144 SARS-CoV-2-positive patients were evaluated for the presence of the D614G mutation. In the first step, the spike (S) glycoprotein sequence of SARS-CoV-2 was used to find a compatible restriction endonuclease and primer design. The S-D type is character-

ised by the presence of a T nucleotide at position 1845, which encodes aspartic acid at position 614 of the amino acid chain. If a T to G mutation has occurred at this position, aspartic acid is replaced by glycine at position 614 of the amino acid chain, which is referred to as the S-G type. The size of the PCR product is 590 base pairs (bp). The enzymatic digestion produces two fragments of 433 bp and 157 bp in length if the T nucleotide is at position 1845. If nucleotide G is at this position, the digestion has no effect on the PCR product, and after agarose gel electrophoresis, one 590 bp fragment is visible. Out of 144 samples, 127 (88.2%) samples belonged to type S-D, 13 (9%) samples were S-G and 4 (2.8%) samples had mixed bands related to both the S-D and S-G types. The results of the given study were consistent with the findings of other investigations examining the frequency of D614G mutations (Bhattacharyya *et al.* 2020). The results of the given study demonstrated the efficiency and effectiveness of the RFLP-PCR

technique in studying the frequency of D614G mutations.

Concluding remarks

The RFLP-PCR technique, despite its obsolescence, is still applicable for studying the frequency of SNPs, especially when there is a severely limited budget and a lack of specialised laboratory equipment.

References

- Alavian, S.E., Sharafi, H., Shirmast, P., Alavian, S. M., Behnava, B., Pouryasyn, M., Keshvari, M., Pouryasyn, A. 2018. A facile PCR-RFLP method for genotyping of ITPA rs1127354 and rs7270101 polymorphisms. *Journal of Clinical Laboratory Analysis*, 32: e22440.
- Backfisch, W., Neuenschwander, S., Giger, U., Stranzinger, G., Pliška, V. 1994. Carrier detection of ovine hemophilia a using an RFLP marker, and mapping of the factor VIII gene on the ovine X-chromosome. *Journal of Heredity*, 85: 474–478.
- Berg Rasmussen, H. 2012. Restriction fragment length polymorphism analysis of PCR-amplified restriction fragment length polymorphism fragments (PCR-RFLP) and gel electrophoresis – valuable tool for genotyping and genetic fingerprinting. In: Magdeldin, S. (ed.) *Gel Electrophoresis – Principles and Basics*, InTech, London.
- Bhattacharyya, C., Das, C., Ghosh, A., Singh, A., Mukherjee, S., Majumder, P., Basu, A., Biswas, N. 2020. Global spread of SARS-CoV-2 subtype with spike protein mutation D614G is shaped by human genomic variations that regulate expression of TMPRSS2 and MX1 genes. *bioRxiv*, 2020.05.04.075911.
- Budowle, B., Chakraborty, R., Gusti, A.M., Eisenberg, A.J., Allen, R.C. 1991. Analysis of the VNTR locus DIS80 by the PCR followed by high-resolution PAGE. *American Journal of Human Genetics*, 48: 137–144.
- Catamo, E., Zupin, L., Segat, L., Celsi, F., Crovella, S. 2015. HLA-G and susceptibility to develop celiac disease. *Human Immunology*, 76: 36–41.
- Chang, H.W., Cheng, Y.H., Chuang, L.Y., Yang, C.H. 2010. SNP-RFLPping 2: an updated and integrated PCR-RFLP tool for SNP genotyping. *BMC Bioinformatics*, 11: 173.
- Chang, H.W., Yang, C.H., Chang, P.L., Cheng, Y.H., Chuang, L.Y. 2006. SNP-RFLPping: restriction enzyme mining for SNPs in genomes. *BMC Genomics*, 7: 30.
- Endreffy, E., Várkonyi, Á., Kaiser, G.I., Raskó, I. 1992. Association of altered RFLP with coeliac disease among Hungarian families. *Journal of Pediatric Gastroenterology and Nutrition*, 14: 118–119.
- Hagelberg, E., Hofreiter, M., Keyser, C. 2015. Ancient DNA: the first three decades. *Philosophical Transaction of the Royal Society B: Biological Sciences*, 370: 20130371.
- Harding, D. 2007. Impact of common genetic variation on neonatal disease and outcome. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, 92: F408–F413.
- Hashemi, S.A., Khoshi, A., Ghasemzadeh-Moghadam, H., Ghafouri, M., Taghavi, M., Namdar-Ahmadabad, H., Azimian, A. 2020. Development of a PCR-RFLP method for detection of D614G mutation in SARS-CoV-2. *Infection, Genetics and Evolution*, 86: 104625.
- Herrmann, F.H., Wehnert, M., Wulff, K. 2008. RFLP analysis for diagnosis of haemophilia A in the German Democratic Republic. *Clinical Genetics*, 37: 12–17.
- Kozák, L., Kuhrová, V., Blažková, M., Fajkusová, L., Dvořáková, D., Romano, V., Pijáčková, A. 1995. Phenylketonuria mutations and their relation to RFLP haplotypes at the PAH locus in Czech PKU families. *Human Genetics*, 96: 472–476.
- Křepelová, A., Brdicka, R., Vorlová, Z. 1993. Factor VIII gene mutations and RFLP analysis in hemophilia A. *Stem Cells*, 11: 72–76.
- Laber, T.L., Giese, S.A., Iverson, J.T., Liberty, J.A. 1994. Validation studies on the forensic analysis of restriction fragment length polymorphism (RFLP) on LE agarose gels without ethidium bromide: effects of contaminants, sunlight, and the electrophoresis of varying quantities of deoxyribonucleic acid (DNA). *Journal of Forensic Sciences*, 39: 13649J.
- Meijer, H., Jongbloed, R.J.E., Hekking, M., Spaapen, L.J.M., Geraedts, J.P.M. 1993. RFLP haplotyping and mutation analysis of the phenylalanine hydroxylase gene in Dutch phenylketonuria families. *Human Genetics*, 92: 588–592.
- Orlando, L., Allaby, R., Skoglund, P., Sarkissian, C. Der, Stockhammer, P.W., Ávila-Arcos, M.C., Fu, Q., Krause, J., Willerslev, E., Stone, A.C., Warinner, C. 2021. Ancient DNA analysis. *Nature Reviews Methods Primers* 1: 1–26.
- Pingoud, A., Jeltsch, A. 2001. Structure and function of type II restriction endonucleases. *Nucleic Acids Research*, 29: 3705–3727.
- Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muruato, A.E., Zou, J., Fontes-Garfias, C.R., Mirchandani, D.,

- Scharton, D., Bilello, J.P., Ku, Z., An, Z., Kalveram, B., Freiberg, A.N., Menachery, V.D., Xie, X., Plante, K.S., Weaver, S.C., Shi, P.Y. 2020. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*, 592: 116–121.
- Pramoonjago, P., Harahap, A., Taufani, R.A., Setianingsih, I., Marzuki, S. 1999. Rapid screening for the most common β thalassaemia mutations in south east Asia by PCR based restriction fragment length polymorphism analysis (PCR-RFLP). *Journal of Medical Genetics*, 36: 937–938.
- Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., Arnheim, N. 1985. Enzymatic amplification of β -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, 230: 1350–1354.
- Stankovic, M., Rakicevic, L., Mikovic, D., Jankovic, G., Nikolic, A. 2005. Indirect diagnosis of haemophilia B by multiplex PCR/RFLP. *Clinical & Laboratory Haematology*, 27: 145–146.
- Tasleem Raza, S., Husain, N., Kumar, A. 2009. Screening for hemophilia A carriers: Utility of PCR-RFLP-based polymorphism analysis. *Clinical and Applied Thrombosis*, 15: 78–83.
- Zhou, B., Thi Nhu Thao, T., Hoffmann, D., Taddeo, A., Ebert, N., Labroussaa, F., Pohlmann, A., King, J., Steiner, S., Kelly, J.N., Portmann, J., Halwe, N.J., Ulrich, L., Trüeb, B.S., Fan, X., Hoffmann, B., Wang, L., Thomann, L., Lin, X., Stalder, H., Pozzi, B., Brot, S. de, Jiang, N., Cui, D., Hossain, J., Wilson, M., Keller, M., Stark, T.J., Barnes, J.R., Dijkman, R., Jores, J., Benarafa, C., Wentworth, D.E., Thiel, V., Beer, M. 2021. SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature*, 592: 122–127.