DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
**177**

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
**177**

**KAAREL KRJUTŠKOV**

# Arrayed Primer Extension-2 as a multiplex PCR-based method for nucleic acid variation analysis: method and applications

TARTU UNIVERSITY
PRESS

Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu

Dissertation is accepted for the commencement of the degree of Doctor philosophiae (in molecular diagnostics) on 9.03.2010 by the Council of the Institute of Molecular and Cell Biology, University of Tartu.

Supervisor: Professor Andres Metspalu, PhD, M.D
(University of Tartu, Estonia)

Opponent: Dr. Ivo Glynne Gut, PhD (Centro Nacional de Analisis Genomico, Parc Cientific de Barcelona, Spain)

Commencment: Room No 217, 23 Riia str., Tartu, on April 23th, at 10.00

The publication of this dissertation is granted by the University of Tartu.

The doctoral studies and the publication of the current thesis were supported by the European Union Sixth Framework Program project Molecular Tools, Enterprise Estonia, ESF DoRa program activity 6 and the Graduate School of Biomedicine and Biotechnology created under the auspices of European Union Social Fund structural funds measure *1.1 Educational System Supportive the Flexibility and Employability of the Labor force and Providing Opportunities of Lifelong Learning for All.*

ISSN 1024–6479
ISBN 978–9949–19–333–2 (trükis)
ISBN 978–9949–19–334–9 (PDF)

Autoriõigus Kaarel Krjutškov, 2010

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr 122

# TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS AND PATENT ........ 6

LIST OF ABBREVIATIONS ........ 7

INTRODUCTION ........ 8

1. REVIEW OF THE LITERATURE ........ 11
   1.1. Polymerase chain reaction ........ 11
   1.2. Nucleic acid as a target for PCR ........ 12
   1.3. From single amplicon PCR to multiplex-PCR ........ 13
   1.4. Multiplex PCR for large-scale sequencing ........ 13
   1.5. SNP genotyping methods ........ 15
      1.5.1. Primer extension ........ 17
         1.5.1.1. Single Base Extension ........ 18
         1.5.1.2. iPLEX® from Sequenom, Inc. ........ 19
         1.5.1.3. SNPstream® from Beckman Coulter, Inc. ........ 19
         1.5.1.4. Fluorescence polarization after SBE ........ 20
         1.5.1.5. Solid-phase minisequencing ........ 21
         1.5.1.6. Arrayed Primer Extension ........ 21
         1.5.1.7. SNaPshot® from Applied Biosystems ........ 22

2. PRESENT INVESTIGATIONS, RESULTS AND DISCUSSION ........ 23
   2.1. APEX-2 primer construction and reaction protocol ........ 24
      2.1.1. APEX-2 primer construct and working schema (Ref. I and II) 24
      2.1.2. Multiplex PCR ........ 26
         2.1.2.1. PCR components (Ref. II and IV) ........ 26
         2.1.2.2. Studied DNA amount and quality (Ref. II, IV and V) 27
         2.1.2.3. Cycling conditions (Ref. II, III and IV) ........ 28
   2.2. Quality parameters of the APEX-2 assay (Ref. II, IV) ........ 29
   2.3. APEX-2 applications and feedback (Ref. III–V) ........ 30
   2.4. Closing remarks ........ 32

CONCLUSIONS ........ 34

REFERENCES ........ 35

SUMMARY IN ESTONIAN ........ 42

ACKNOWLEDGMENTS ........ 44

PUBLICATIONS ........ 45

CURRICULUM VITAE ........ 117

ELULOOKIRJELDUS ........ 119

# LIST OF ORIGINAL PUBLICATIONS AND PATENT

The current dissertation is based on the following publications and a patent referred to in the text by their Roman numerals:

I. US patent US 2009/0048119 A1 “A method to determine single nucleotide polymorphisms and mutations in nucleic acid sequence”. December 29 2009. Applicant: Estonian Biocentre and inventors: **Kaarel Krjutškov** and Andres Metspalu
II. **Krjutškov K**, Andreson R, Mägi R, Nikopensius T, Khrunin A, Mihailov E, Tammekivi V, Sork H, Remm M, Metspalu A (2008). Development of a single tube 640-plex genotyping method for detection of nucleic acid variations on microarrays. Nucleic Acids Research, 36(12), e75 doi:10.1093/nar/gkn357
III. Khrunin A, Mihailov E, Nikopensius T, **Krjutškov K**, Limborska S, Metspalu A (2009). Analysis of Allele and Haplotype Diversity Across 25 Genomic Regions in Three Eastern European Populations. Human Heredity 68 35–44, doi:10.1159/000210447
IV. **Krjutškov K**, Viltrop T, Palta P, Metspalu E, Tamm E, Suvi S, Sak K, Merilo A, Sork H, Teek R, Nikopensius T, Kivisild T, Metspalu A (2009). Evaluation of the 124-plex SNP typing microarray for forensic testing. Forensic Science International: GENETICS 4: 43–48, doi:10.1016/j.fsigen.2009.04.007
V. Viltrop T, **Krjutškov K**, Palta P, Metspalu A (2009). Comparison of DNA extraction methods for multiplex PCR. Analytical Biochemistry 398: 260–262, doi:10.1016/j.ab.2009.11.026

My contribution to the articles referred in the current thesis is as follows:

Ref. I co inventor of the concept, performed proof of principle experiments and participated in the drafting of application manuscript;
Ref. II conducted the study, carried out all laboratory experiments in the optimization and validation phases, participated in the creation of the primer design algorithm and was responsible for drafting the manuscript. First author;
Ref. III performed assay optimization tests, participated in the primer design process and analysis software modification. Co-author;
Ref. IV conducted and designed the study, carried out most laboratory experiments in the optimization and validation phases, participated in the design of the primers, and was responsible for drafting the manuscript. First author;
Ref. V designed the study and was a co-author in drafting the manuscript. Shared first authorship;

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ABI | Applied Biosystems |
| APEX | Arrayed Primer EXtension |
| APEX-2 | Improved version of APEX |
| DHPLC | Denaturing High Performance Liquid Chromatography |
| DMPS | Direct Multiplex PCR Sequencing |
| HRM | High Resolution Melting |
| INDEL | INsertion and DELetion |
| kb | Kilobase |
| LMP | Ligation-Mediated PCR |
| MALDI-MS | Matrix-assisted laser desorption/ionization mass spectrometry |
| MALDI-TOF | MALDI time-of-flight mass spectrometry |
| MDA | Multiple Displacement Amplification |
| MIP | Molecular Inversion Probe |
| mtDNA | Mitochondrial DNA |
| SBE | Single Base Extension |
| SGS | Second-Generation Sequencing |
| SNP | Single Nucleotide Polymorphism |
| *tag* sequence | Bar-coding part in oligonucleotides |
| *ctag* sequence | Complementary sequence to *tag* |
| $T_m$ | Primer melting temperature |
| WGA | Whole Genome Amplification |
| Y-DNA | Y chromosome DNA |

# INTRODUCTION

In 2004, there was an assumption that genome-wide association studies required genotyping several hundreds to thousands of single nucleotide polymorphisms (SNPs) in disease-associated gene mapping studies. In 2005, A.C. Syvänen stated in a Nature Genetics genotyping supplement that efficient and cost-effective SNP genotyping methods are required for routine clinical applications once disease-predisposing genes have been identified (Syvänen, 2005). She continued: "*A key technical obstacle is the PCR amplification step. /…/ The multiplexing level that can be achieved in PCR does not match that of current microarray-based methods, making PCR the limiting step in the assays*."

A.C. Syvänen's group was one of many labs in which new genotyping method development and multiplex PCR *tuning* were the main subjects of research. In our laboratory, Kurg and colleagues have published the arrayed primer extension (APEX) protocol (Kurg *et al.*, 2000) as an update of a original APEX paper (Shumaker *et al.*, 1996) and thereafter only minor changes were introduced. In the APEX protocol, the complexity of genomic DNA was reduced by single (or up to five) – plex PCR, making it a very useful method that has been employed in many studies in our department (Kõks *et al.*, 2006; Canova *et al.*, 2009) as well as by other groups/laboratories (Jaakson *et al.*, 2003; Tebbutt *et al.*, 2004; Schrijver *et al.*, 2005; Zernant *et al.*, 2005; Lahermo *et al.*, 2006). Concurrently, a large amount of different PCR protocols were developed to increase multiplex size. For example, a 25-plex SNP typing assay for the Y chromosome DNA (Y-DNA) (Sanchez *et al.*, 2003), a 39-plex genotyping system for distinguishing blood group loci (Inagaki *et al.*, 2004) and a 45-plex genotyping assay, based upon careful primer design using the *tag-complementary tag (tag-ctag)* system (Kaderali *et al.*, 2003), were developed. The main hindrance of such multiplex assays was the necessity for extensive optimization; the differences in primer concentrations varied up to 120 times (Inagaki *et al.*, 2004) and due to the complexity of the reaction, 50-plex was the approximate upper limit for any of the above-mentioned protocols. In the beginning of 2000, the first papers, with hidden technical details, were published by Oliphant and colleagues regarding the Illumina BeadArray platform (Oliphant *et al.*, 2002). In 2003, ParAllele BioScience described for the first time the scope of molecular inversion probe (MIP) genotyping, wherein multiplex analysis of more than 1,000 probes in a single tube was performed (Hardenbol *et al.*, 2003). Six months later, Affymetrix introduced their large-scale whole-genome genotyping technology (Kennedy *et al.*, 2003), although they have published a 512-SNParray, that was combined into a 46-plex PCR with carefully chosen primer pairs and hybridization-based SNP detection, years ago (Wang *et al.*, 1998). This era, as we realize now, was the beginning of an extensive growth in the field of genotyping.

In 2004, our laboratory received an invitation to participate in the FP6 MolTools project[1], in which one of the objectives was to develop an advanced

[1] European Union Sixth Framework Program project Molecular Tools (LSHG-CT-2004-503155)

multiplex PCR method for existing APEX technology. Our goal was to exploit the direct selection of studied loci by microarray hybridization. Genomic DNA was digested by DNA restrictase(s), universal linkers were ligated to the end of the fragments (Lin *et al.*, 1996; Matsuzaki *et al.*, 2004) and the enrichment of specific sequences was performed on a microarray. After capturing the hybridized regions, PCR with a universal primer system was used to amplify all selected genomic fragments (Fig. 1). The amplification results after enrichment were promising, but single base extension (SBE) on a microarray revealed low call rates and specificity. These problems hindered further progress in the development of this technology. Later, in collaboration with NimbleGen Systems Inc, three papers describing microarray-based genomic selection for second-generation resequencing (SGS), were published simultaneously (Fig. 1) (Albert *et al.*, 2007; Hodges *et al.*, 2007; Okou *et al.*, 2007).

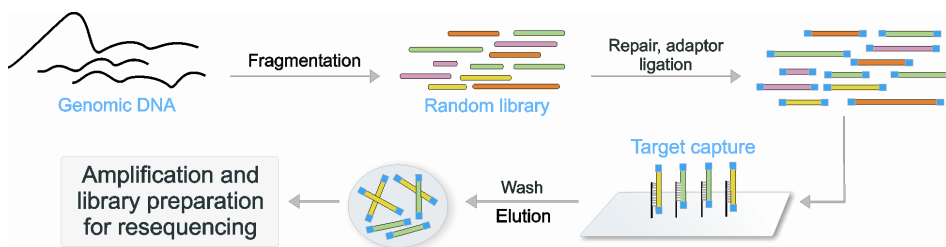



**Figure 1.** Microarray-based genomic selection and resequencing of complex genomes. A similar principle was followed in our enrichment experiment in 2004.

Successful capturing of targeted DNA apparently relied upon the amount of starting DNA, that was many times higher (20 µg) than we had used (<2 µg). Also, oligonucleotide probes in our research did not facilitate sufficiently specific and stable hybridization due to their short length (25 bp, whereas NimbleGen probes were in the range of 60–90 bp). Today, such hybridization based target enrichment systems[2,3] have made great progress and are also being used in whole exons (exome) SGS (Ng *et al.*, 2009) or targeted research (Gnirke *et al.*, 2009). Using HybSelect as a microarray based sequence capture application, can reduce the starting DNA requirement (1.5 µg), hybridization time (16 h) and specific probe length (50 bp) before SGS (Bau *et al.*, 2009; Summerer *et al.*, 2009). One can conclude that, in order to ensure a highly specific capture system for APEX multiplex amplification, longer capturing probes ought to been used in a much smaller volume in the specific microfluidic hybridization system.

---

[2] Sequence capture 2.1M Human Exome Array from Roche NimbleGen, Inc. (Madison, USA)

[3] SureSelect DNA Capture array from Agilent Technologies (Santa Clara, USA)

When our first effort did not succeed, a new concept for multiplex PCR was still on elaboration agenda. First, an idea worth testing was PCR with SBE primers themselves (APEX array-attached SBE primers), where only the nucleotide of interest would be amplified from the genomic DNA. In order to get a positive outcome in amplification and on an array, a universal *tag* was added to the 5' end of each oligonucleotide. As a result, ten improved APEX (now APEX-2) primer pairs were incorporated in one primer mixture and on May 20, 2005, the first correct signals were detected on a microarray. Later, 66-plex was tested for the first time and at the end of the year, 1200 APEX-2 oligonucleotides for a real biological application[4] were ordered. In summary, APEX-2 was created at a time when medium-scale multiplex PCR applications were demanded. The developments were funded by FP6 (MolTools project) and Enterprise Estonia[5].

Why is multiplex-PCR development relevant if high-density genotyping commercial microarrays and relatively cheap resequencing is currently available? First, the application of multiplex PCR and basecalling on arrays are meant for SNP rescreening, (e.g replication phase of the genome-wide association studies or massive resequencing). The second use of our APEX-2 method is in molecular diagnostics, where tens to hundreds of genomic loci (mutations) are analyzed in diagnostic assays. Thirdly, custom-made microarrays and APEX-2 multiplex PCR are affordable for many research or diagnostic laboratories. This is very important, because a robust, cost-effective, and flexible genotyping assay is a basic tool in any genomic laboratory now and in the near future.

The main goals of the present study were (i) to investigate the factors affecting multiplex PCR, (ii) create an effective methodology to overcome multiplexing limitations, and (iii) elaborate a simple protocol for the simultaneous analysis of 10 to 1,000 genetic markers for scientific and public research.

1. To create a multiplex-PCR protocol to amplify 10–1,000 loci of interest in a single tube and detect them on an oligomicroarray. To test the primer constructs in different genomic regions (SNP-, AT- or GC-rich) and find out optimal conditions and solutions for cases in which close polymorphisms or INDELs are being studied.
2. To experiment with different DNA sources simultaneously (genomic, mitochondrial, and Y-chromosome DNA) as templates in one assay and discover an appropriate saliva DNA extraction method as a promising alternative to venous blood in everyday diagnostic or public research.
3. To apply the given protocol in population genetics research, genealogical studies and in case-control genetic analysis wherein different plex sizes are used.

---

[4] Obesity candidate gene SNP based association study (not published)

[5] Enterprise Estonia grant through Estonian Biocentre (EU19955)

# 1. REVIEW OF THE LITERATURE

## 1.1. Polymerase chain reaction

Skipping the origin of the polymerase chain reaction (PCR), we can generalize that in 1993, when the PCR invention received the Nobel Prize in Chemistry, it was ready for different applications. In PubMed, contains 400,000 references citing PCR. The citations come from medical, infectious disease, forensic, environmental science, and research applications. In clinical applications, there are presently over 1,200 genetic tests[6] available and the number is predicted to increase by 25% annually (Allingham-Hawkins, 2008). Also, PCR has revolutionized the characterization and detection of organisms in infectious diseases. A pioneering study described the detection of human immunodeficiency virus (HIV) by PCR (Kwok *et al.*, 1987). Many other viruses and their subtypes (e.g. influenza and human papillomavirus) have reliable PCR-based tests available (Brouqui *et al.*, 2009; Weissenborn *et al.*, 2010). Short tandem repeats and miniSTR multiplex detection (Grubwieser *et al.*, 2006), ABO genotyping (Doi *et al.*, 2004; Inagaki *et al.*, 2004), and SNP marker profiling (Sanchez *et al.*, 2006) are used concurrently in forensic applications such as genetic fingerprinting.

PCR itself is based on the very unique molecular mechanism of replication. A DNA molecule can be copied from generation to generation remaining (almost) identical to the initial molecule. The enzyme that catalyzes the natural reaction is DNA polymerase, and the thermostable DNA polymerase from *Thermus aquaticus*, for example, is a key compound in PCR. Other components are 2'-deoxynucleotide 5'-triphoshates (dNTPs), DNA molecules of interest and short oligonucleotides (primers) that specify the region of amplification and offer the free 3' hydroxyl group (3'-OH) for DNA polymerase as a starting position for synthesis (Saiki *et al.*, 1988). A standard PCR involves the following three steps: (i) double-stranded DNA denaturation to produce single-stranded DNA; (ii) primer hybridization to specific sites and (iii) enzymatic primer extension from the annealed primer. Such a reaction order is repeated 20–35 times. Since the amount of amplified PCR product is doubled in each cycle, the studied DNA region (amplicon) accumulates exponentially.

There are plenty of PCR modifications that have developed into new PCR-based methods: real-time PCR, allele-specific PCR, competitive oligopriming PCR, digital PCR and SBE are just a fraction of the list. In other words, as Nathan Blow, technology editor for Nature and Nature Methods, said in 2007: *"PCR - the workhorse of modern molecular biology – is charging forward…"*

---

[6] Diagnostic testing, carrier testing, predictive and presymptomatic testing, prenatal testing and newborn screening.

## 1.2. Nucleic acid as a target for PCR

The sensitivity and reliability of PCR increased to a level where even a single genomic copy was sufficient to carry out genetic testing. This allows to analyze single cells, for example DNA removed from embryos at the preimplantation stage (Handyside *et al.*, 1990). Subsequently, single-cell PCR has been used for the molecular diagnostics of cystic fibrosis mutations (Drury *et al.*, 2000), diagnostics of human T-lymphotropic virus type I DNA in lymphocytes (Miyagi *et al.*, 1998) and β-thalassaemia mutation detection for preimplantation diagnosis (el-Hashemite *et al.*, 1997). These reports demonstrated the possibility of analyzing a single or a very restricted number of studied positions in the genome. To overcome the scarcity of genomic DNA, one can use whole genome amplification (WGA), which was developed in 1992 and then used for the degenerated oligonucleotide primed (DOP) PCR system (Telenius *et al.*, 1992). The most recently developed WGA methods are based on ligation-mediated PCR (LMP), the T7-based linear amplification of DNA (TLAD) and multiple displacement amplification (MDA), offering many prospects for DNA analysis. The usefulness of these new WGA methods has been demonstrated in applications like MDA amplification from a single sperm (Jiang *et al.*, 2005) and a single-cell whole genome MDA protocol (Spits *et al.*, 2006a; Spits *et al.*, 2006b).

One of the first challenges that PCR-based methods in modern forensic testing faced at the beginning of the next millennium was the identification of victims in the catastrophe of September 11$^{th}$ in 2001. The number and identity of the victims were unknown and many victims were represented only by bone and tissue fragments (~20,000 pieces of human remains were found). The identification began initially by the Combined DNA Index System 13-locus STR panel (CODIS, 500 persons identified) (Budimlija *et al.*, 2003), continued with mitochondrial DNA (mtDNA) analysis (57% of 2792 persons identified). The medical examiner's office was reopened when Bode Technology Group developed a new methodology of DNA extraction that required much less sample material for analysis than before. Therefore, a trial for genetic engineering in forensic application and in computer analysis took place because there was a shortage of target DNA, and often it was also seriously obstructed (Lovgren, 2003). This dramatic accident of 9/11 thus created a link between forensic science, data analysis and public interest.

Fragmented DNA may push molecular diagnostics or genetic forensic tests to their limits, whereas PCR using ancient or degraded input DNA (comprised only of small (<200 bp) fragments) should amplify specific regions. In order to do that, it is possible to (i) partially reconstruct the target DNA and amplify larger fragments, so that the initial DNA is self-priming in the first phase and then primer-specific PCR is carried out (Golenberg *et al.*, 1996); (ii) use very short amplicons (40–70 bp) for SNP analysis, universal *tags* in primers' 5' ends, labeled universal primers in a gel-based system for allele detection (Asari *et al.*, 2009); or (iii) to use a carefully optimized genotyping strategy for highly degraded DNA to perform a two-stage multiplex assay whereby multiple

fragments are first amplified in a single exponential reaction and the products of this PCR are added to a linear single-base-extension reaction (Sanchez and Endicott, 2006). The feasibility of size-reduced amplicons in cases of degraded nuclear- or mtDNA has been proven in one study wherein high-quality results were achieved from multiplex-PCR utilizing as little as 25 pg of target DNA (von Wurmb-Schwark *et al.*, 2009).

## 1.3. From single amplicon PCR to multiplex-PCR

Current genetic testing in molecular diagnostics mostly addresses certain causative alleles in disease genes. For example, more common genetic disorders like color blindness, cystic fibrosis, haemophilia, phenylketonuria, and sickle-cell anemia are caused mostly by point mutations within the transcribed area[7].

In the research sector, there is a clear understanding that cost-effective and more informative methods of analysis are always welcome. Multiplex level PCR protocols, optimized to decrease the cost and hands-on work as much as possible, have been put into practice in a huge number of independent and replication studies. The first multiplex PCR (2-plex) application was directed toward screening mutations of the Duchenne muscular dystrophy loci in 1988 (Chamberlain *et al.*, 1988). Years later, this multiplex PCR technique had spread to pathogen identification, gender screening, linkage analysis, and forensic studies. Template molecule quantification and genetic disease diagnostics were also found in the list of multiplex PCR usages (Edwards and Gibbs, 1994). This area is described in detail by (Shuber *et al.*, 1995; Henegariu *et al.*, 1997; Elnifro *et al.*, 2000) and there are numerous protocols for customizing multiplex PCR assays, if commercial solutions[8,9,10] do not match with one's demands.

## 1.4. Multiplex PCR for large-scale sequencing

Multiplex PCR was originally developed for the detection of SNPs, mutations, or short tandem repeats using different techniques (described in Chapter 1.5). SGS initiated the publications of several original multiplex PCR techniques in the last few years. Multiplex PCR in SGS is focused on reducing the complexity of the genome studied, for example, by sequencing only certain genomic regions, genes, exons, or mtDNA.

Three major publications demonstrate the impact of SGS on multiplex PCR. Technical details of these papers are presented here because of their interesting molecular engineering solutions to amplify many kilobases from distant regions

7 Online Mendelian Inheritance in Man database at http://www.ncbi.nlm.nih.gov/omim/
8 PowerPlex assay from Promega (Madison, USA)
9 SNaPshot kit from Applied Biosystems (Foster City, USA)
10 Multiplex PCR Kit from QIAGEN (Hilden, Germany)

simultaneously. Firstly, Varley and colleagues have published a protocol called Nested Patch PCR which enables one to amplify many kilobases (kb) of genomic DNA (in their case, colon cancer candidate gene exons) for further analysis by sequencing (Varley and Mitra, 2008). Ninety-four non-overlapping primer pairs were designed to amplify 94 exons from six genes, which were located across four chromosomes and ranged in length from 74 to 438 bp (21.6 kb total). All primer pairs were mixed together in equal concentrations (50 nM each), and 10 cycles of multiplex PCR were performed using 1 μg of input DNA. The PCR primers contained uracil substituted for thymine. In the next step, primers were cleaved from the amplicons by the addition of uracil DNA glycosylase and made blunt-ended using admixtures of other enzymes (Fig. 2A). Oligonucleotides were then annealed to the target amplicons and served as a patch between the correct amplicons and universal primers. Subsequently, ligation and simultaneous PCR amplification using universal primers was carried out. Using the above-described strategy, 96% of the targets (90 of 94) were amplified and thereby several germline SNPs and one colon tumor specific nonsense mutation were successfully detected. A great advantage of this method is its high specificity, ensured by specific PCR and ligation reactions, as the thermostable ligase is sensitive to mismatched bases near the ligation junction site. Thus, the sequence determining the specificity of the process is 40 to 60 bp on each end of the product (primer plus patch oligonucleotides), making it a total of 80 to 120 bp per amplicon.

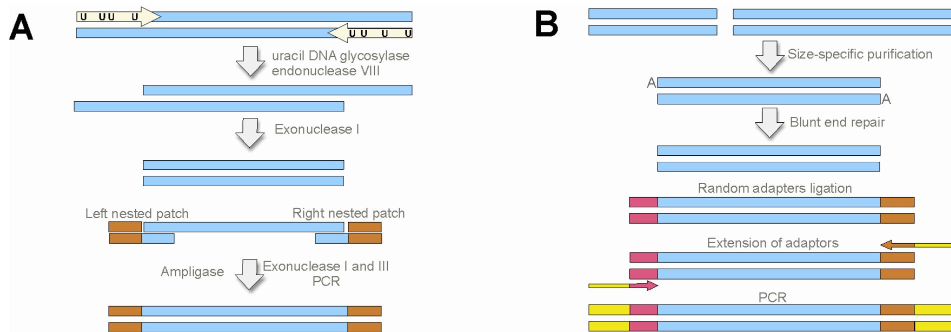



**Figure 2.** Two principles of multiplex PCR in second-generation sequencing studies. (**A**) Nested Patch PCR and (**B**) Direct MultiPlex Sequencing (DMPS). More detailed description can be found in the text above.

The second paper presents direct multiplex PCR sequencing (DMPS), a novel method for targeted SGS of ancient and degraded DNA, primarily mtDNA. It combines standard multiplex PCR with sample bar-coding[11] and SGS in a straightforward way (Stiller *et al.*, 2009). The aim of the study was to sequence

[11] bar-code is a part of the linker, essential in SGS when different samples are sequenced simultaneously

mitochondrial genomes of 31 cave bears, a species that became extinct 25,000 years ago, using SGS. They designed 128 non-overlapping primer pairs, with amplicons ranging in length between 150 and 180 bp, and performed 20-cycle multiplex PCR in two equal separate reactions (64-plex PCR, 150 nM concentration of each primer). Next, a relatively stringent product length cutoff separation and purification method[12] was used, amplicon ends were polished to blunt ends, and adapters were ligated to enable later bar-coding and universal primer PCR (Fig. 2B). After the second round of sequencing, the average coverage of the mitochondrial targets was 95.7%.

The third study, which demanded more labor force, was designed to amplify seven CMT[13] disease associated genes (~8 kb of coding region, 14.4 kb total) by multiplex PCR and SGS sequencing (Goossens *et al.*, 2009). All 63 primer pairs encompassed fragments ranging from 150 to 250 bp with adapter sequences needed for the sequencing reaction. Multiplex PCRs were performed in seven different reactions, using primer concentrations in the range of 50 to 200 nM and 100 ng of input DNA. Next, all products were mixed, purified, concentrated by column, and prepared for sequencing.

## 1.5. SNP genotyping methods

SNP genotyping methods have shown rapid progress throughout many years to achieve higher throughput with inexpensive costs. Knowledge in molecular engineering and new possibilities for detection instruments have unleashed the development of genotyping methods in different ways. There are tens of different well-known genotyping methods, which are based on the dissimilar detection principle. These methods start with the less-complex classic restriction fragment length polymorphism (RFLP) analysis, Sanger *gold standard* dideoxy sequencing, pyrosequencing, a high-resolution melting-based method, and finish with ultra high-throughput SNP genotyping microarrays (Affymetrix and Illumina) or (capture using) SGS. Subsequently, I briefly discuss different SNP genotyping solutions which may offer interest and creates frameworks/material for discussion of similar methods for small- and medium scale genotyping projects. Also, I attempt to explain and argue why primer extension-based (in addition to other enzyme-based) methods are still one of the most promising from this selection.

Widely exploited microtiter plate-based fluorescent readout systems use different commercial DNA probes in detection, such as Taqman, KASPar, Molecular Beacons, Scorpion Probes and fluorescence polarization (Chapter 1.5.1.4). They all can be performed as an endpoint assay in a completely

---

[12] AMPure® PCR purification system from Agencourt Bioscience, a Beckman Coulter Company

[13] Charcot-Marie-Tooth disease, the most common inherited disorder of the peripheral nervous system.

homogeneous reaction. All the reagents and genomic DNA are mixed at the beginning, and the fluorescent signal is read after the thermocycling step. There is neither a separate pre-amplification step nor intermediate processing, making them the simplest assay formats possible. However, only low-level (up to 3 or 4) multiplex is possible, restricting its potential for an even higher throughput. This lack of multiplexing capability is largely due to the difficulty in discriminating signals from different dyes with overlapping spectra, limiting the number of compatible dye combinations.

Microarray-based assays can be divided into two major categories: (i) hybridization based (Affymetrix) or (ii) hybridization and SBE based methods (more specific and require fewer oligonucleotides per locus detection, discussed in Chapter 1.5.1.1). A clear advantage of both approaches is the ability to genotype a large number of loci in parallel. The throughput depends on the analyzed DNA preparation method before detection: (i) *classical PCR*; (ii) restriction-ligation based enrichment (Matsuzaki *et al.*, 2004); or (iii) WGA before probe hybridization and SBE (Steemers *et al.*, 2006)), and microarray manufacturing possibilities, especially the density of oligonucleotides on the array.

Mass spectrometry based genotyping assays (Chapter 1.5.1.2) have a steady and very promising position for future developments due to their simple process and low running costs. Although these assays require amplification, purification, and detection as separate phases, the system is robust, enables simultaneous amplification, and the analysis is largely automated.

The flow cytometry-based commercial genotyping method, Luminex's xMAP®, uses microspheres, lasers, and allele-specific primer extension to detect up to 50 loci simultaneously. DNA of interest is amplified by single or multiplex PCR, mixed with an allele-specific primer (containing a 5' *tag* sequence) and primer extension using biotinylated dCTPs or an oligonucleotide ligation reaction is performed (Bruse *et al.*, 2008). Next, *tag* sequence extended chains are hybridized onto a microsphere-bounded *ctag* and the resulting complex is analyzed by flow cytometry. The assay setup requires much optimization (primers, dye combinations, and *tag-ctags*) and preceding PCR amplification, but the advantages include high throughput, simplicity of operation, and low running cost.

An original genotyping method which enables one to detect loci without pre-amplification is an endonuclease based Invader® assay (Lyamichev *et al.*, 1999). The Invader assay uses a structure-specific flap endonuclease to cleave a three-dimensional complex formed by hybridization of allele-specific overlapping oligonucleotides. This assay targets DNA containing a SNP site and requires at least 50 ng of genomic DNA or pre-amplified product. Cleavage can be detected by three different approaches: (i) the cleavage product triggers a secondary cleavage reaction on a fluorescence resonance energy transfer to release a fluorescent signal, (ii) the cleavage can be detected directly by use of fluorescence polarization probes, or (iii) by mass spectrometry. The main obstacle of this Invader technology is the throughput, because the chemical principle allows single-locus detection in one vessel. A meaningful develop-

ment in the context of multiplex PCR is the link between 100-plex PCR and Invader genotyping (Ohnishi *et al.*, 2001). After testing the length and specificity of these 100 primer pairs *in silico* (Andreson *et al.*, 2006), we found that the majority of the primers were specific, but product length varied between 1000 and 1600 bp. In addition, equal primer concentrations (50 pmol of each primer in an unknown volume, which means an immense amount of primers per reaction), *tag*-free primers (Wang *et al.*, 1998), and the knowledge that high-level specific multiplex can be successfully performed by short amplicons (Sanchez *et al.*, 2006; Podder *et al.*, 2008) suggest that the 100-plex level achieved in this report is a misleading result.

There are four post-amplification methods based on the physical properties of DNA: (i) single strand conformation polymorphisms, where single stranded DNA is analyzed by gel electrophoresis, which separates the different conformations (not discussed here); (ii) temperature gradient gel electrophoresis (not discussed here); (iii) denaturing high performance liquid chromatography (DHPLC); and (iv) high-resolution melting (HRM) of the entire amplicon. DHPLC, only one type of high performance liquid chromatography, is a well-known commercialized method[14], which provides a useful platform for high-throughput mutation or SNP discovery, but is not used in SNP calling. Genotyping by an HRM-based technology uses two PCR primers per locus and a generic, saturating DNA dye that detects heteroduplexes as well as homoduplexes. Heterozygous genotypes have a characteristic melting curve shape and a broader width than homozygous genotypes. Modern instruments[15] have combined PCR and detection into a single procedure that enables 4-plex analysis (Seipp *et al.*, 2009). The assay validation is relatively time-consuming and sensitive for any other polymorphic site in the PCR amplicon. However, the method is cost-effective, simple, and appropriate if only a few loci and thousands of DNA samples are under study.

### 1.5.1. Primer extension

Primer extension is a comprehensive term in molecular engineering. Primer extension refers to the following: (i) As a method it is known as the RNA or DNA 5' end mapping technique, wherein labeled oligonucleotides ($^{32}$P) are used to determine potential transcriptional start sites and localize promoters or TATA boxes. (ii) In sequencing, primer extension and random termination with ddNTPs in simultaneous polymerase reactions are the main principles of the Sanger method. In addition, SGS techniques like sequencing by synthesis from Roche (Margulies *et al.*, 2005; Droege and Hill, 2008; Wheeler *et al.*, 2008) and from Illumina (Bentley *et al.*, 2008) (one exception is the ABI ligation based SOLiD™ system) use primer extension for detecting nucleotides of interest.

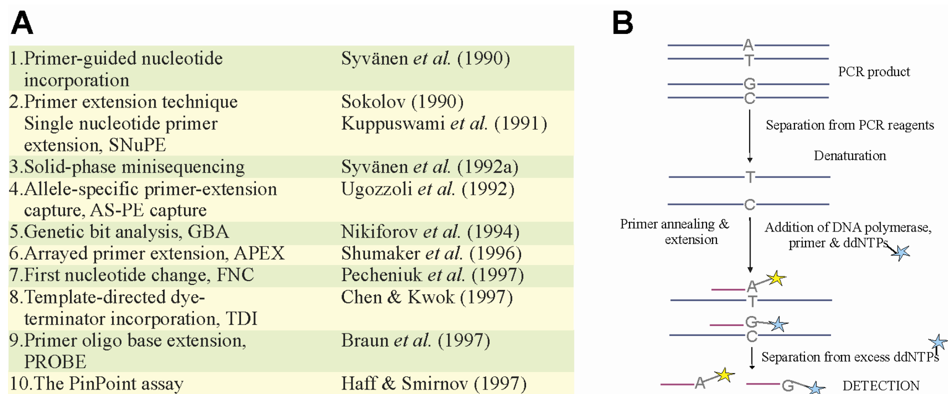
[14] Agilent Technologies, PerkinElmer, Beckman Coulter, Hitachi, Transgenomics and others

[15] LightScanner-32 from Idaho Technology (Salt Lake City, USA)

(iii) Modern genotyping systems are fully dependent on primer extension, with minor exceptions: restriction fragment length polymorphisms, short tandem repeats, and some mass-spectrometry solutions analyze the PCR product for genotyping (Pitterl *et al.*, 2008). Consequently, hereinafter I focus on primer extension in the field of genotyping. Particularly SBE on primers' 3' ends as a nucleotide identification technique if small- to medium-scale numbers of loci are under study. Different SBE-based genotyping methods are described, and advantages and limitations are highlighted.

#### 1.5.1.1. Single Base Extension

Single base extension, also called as minisequencing or APEX, was published in the beginning of the 1980s by three independent laboratories (Sokolov, 1990; Syvänen *et al.*, 1990; Kuppuswamy *et al.*, 1991) as a very promising reaction concept for analyzing point mutations and SNPs. Later, different strategies and assay formats by A.C. Syvänen were suggested, and these are listed in Fig. 3A (Syvänen, 1999). In the same year, the template-directed dye-terminator incorporation assay with fluorescence polarization detection (FP-TDI)[16] was published (Chen *et al.*, 1999), providing a basis for high-throughput analysis in a simple, fully-automated format.



**A**

| Name | Reference |
|---|---|
| 1.Primer-guided nucleotide incorporation | Syvänen *et al.* (1990) |
| 2.Primer extension technique<br>Single nucleotide primer extension, SNuPE | Sokolov (1990)<br>Kuppuswami *et al.* (1991) |
| 3.Solid-phase minisequencing | Syvänen *et al.* (1992a) |
| 4.Allele-specific primer-extension capture, AS-PE capture | Ugozzoli *et al.* (1992) |
| 5.Genetic bit analysis, GBA | Nikiforov *et al.* (1994) |
| 6.Arrayed primer extension, APEX | Shumaker *et al.* (1996) |
| 7.First nucleotide change, FNC | Pecheniuk *et al.* (1997) |
| 8.Template-directed dye-terminator incorporation, TDI | Chen & Kwok (1997) |
| 9.Primer oligo base extension, PROBE | Braun *et al.* (1997) |
| 10.The PinPoint assay | Haff & Smirnov (1997) |



**Figure 3.** (**A**) Names and acronyms for the single base extension assay (Syvänen, 1999). (**B**) Reaction steps of the single base extension assay to analyze a G→A transition, with minor modifications from Syvänen, 1999.

In SBE, a DNA synthesis reaction catalyzed by DNA polymerases is used to distinguish between sequence variants. The concept of the assays is to allow a detection primer to anneal to the nucleic acid sequence immediately 3' of the

[16] Commercially available from PerkinElmer Inc. (Waltham, USA)

nucleotide position to be analyzed, and to extend this primer with a single labeled 2'-dideoxynucleotide 5'-triphoshate (ddNTP) that is complementary to the nucleotide to be detected using DNA polymerase (Fig. 3B) (Syvänen, 1999).

Currently, many SBE-based assays are successfully in use, and offer such commercial products as iPLEX (Sequenom), SNPstream (Beckman Coulter), a kit solution for SBE detection like SNaPshot from Applied Biosystems, or template-directed dye-terminator incorporation with fluorescence polarization detection systems (Chen *et al.*, 1999; Kwok, 2002). Custom solutions like *tag ctag* SBE from the Syvänen group (Lindroos *et al.*, 2002; Lovmar and Syvänen, 2005) and APEX (Shumaker *et al.*, 1996; Kurg *et al.*, 2000; Tõnisson *et al.*, 2002), are often applicable and easily available due to their flexibility and cost-effectiveness.

#### 1.5.1.2. iPLEX® from Sequenom, Inc.

Several versions of primer extension assays using MALDI-MS as the detection platform have been developed and commercialized (Haff and Smirnov, 1997; Wenzel *et al.*, 2003), but perhaps the most widely applied SNP genotyping assay using MALDI-MS for detection is provided by Sequenom. Two enzymatic reactions are involved: (i) locus-specific multiplex PCR amplification and (ii) SBE using allele-specific primers and acyclic mass-modified ddNTP terminators, which create mass differences large enough to differentiate between all four bases (Tost and Gut, 2006) (Fig. 4A). Using MALDI-TOF mass spectrometry, the distinct mass of the one nucleotide-extended primer identifies the SNP allele (Gabriel *et al.*, 2009). The iPLEX Gold assay provides multiplexing and high throughput, enabling researchers to perform large fine-mapping and SNP validation studies, or any routine applications that employ focused SNP panels. The system's chemistry enables one to create up to 40-plex reactions and analyze up to 384 samples with one multiplex assay or up to 384 different multiplexed assays with one DNA sample. In practice, smaller plex sizes are more reliable and assay optimization is more labor-intensive.

#### 1.5.1.3. SNPstream® from Beckman Coulter, Inc.

The SNPstream genotyping system provides an automated scalable system capable of performing from 4,600 to over 3,000,000 SNP genotypes per day. By combining it with multiplexing and operational flexibility, this enables one to conduct a wide range of studies[17]. The assay enables one to create up to 48-plex reactions by carefully designed locus-specific primers. SBE requires a third primer with a *tag* sequence that enables *tag-ctag* hybridization in the detection step on *tag* microarrays in a glass-bottom microtiter plate (Fig. 4B). All components of the SNPstream system are provided by Beckman Coulter and are optimized for most applications.

---

[17] www.beckmancoulter.com/Literature/BioResearch/T-2050A.pdf

### 1.5.1.4. Fluorescence polarization after SBE

The template-directed dye-terminator incorporation assay with fluorescence polarization detection (FP-TDI) is a highly sensitive and robust method for the simultaneous analysis of large numbers of samples, and uses one reaction tube per SNP. PCR is usually run in 384-well microtiter plates, followed by enzymatic inactivation of PCR primers/nucleotides and cyclic primer extension reactions with fluorescently labeled ddNTPs (Fig. 4D). The presence of one or both SNP allele(s) is determined by measuring fluorescence polarization using a fluorometer. In this instrument, plane-polarized light excites the fluorophores, and the degree of polarization in the emitted light is measured. The differences between polarized and depolarized fluorescence from each sample is displayed in a scatter plot, with three separated clusters representing each of the three genotypes (Chen *et al.*, 1999). The principle of discrimination between single base extended/unextended primers lies in the fact that incorporated labeled ddNTP molecules rotate (depolarize) faster than primer-bounded molecules.

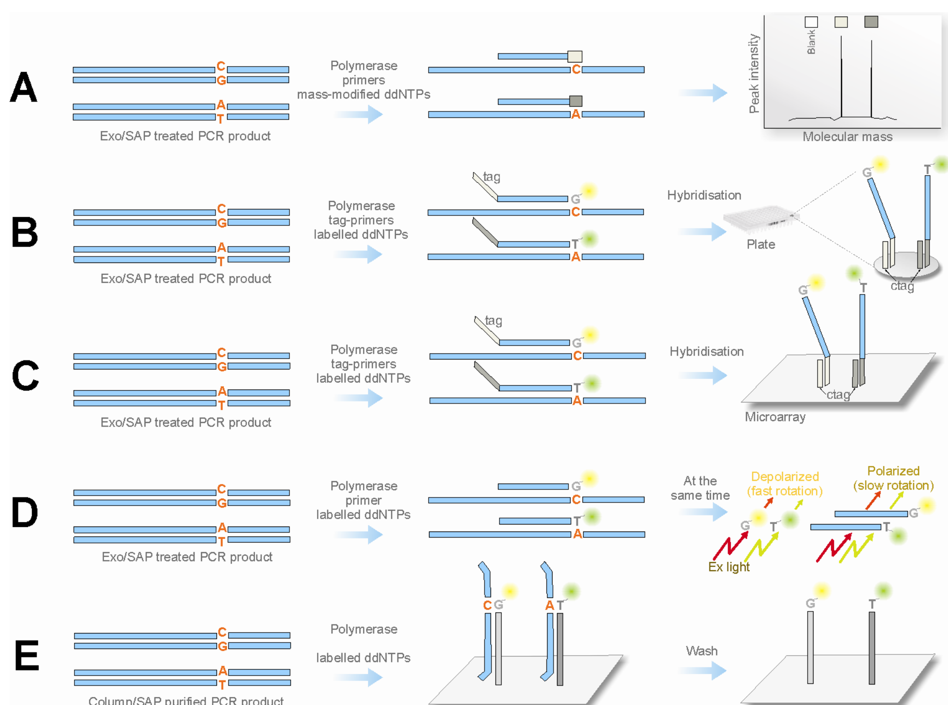



**Figure 4.** Different applications of primer extension. (**A**) iPLEX® MALDI mass-spectrometry based commercial platform, (**B**) commercial *tag-ctag* based SBE and detection on the botton of the reaction plate vessel, (**C**) *tag-ctag* principle using micro-array, (**D**) fluorescence polarization after SBE and (**E**) array-based SBE (APEX).

### 1.5.1.5. Solid-phase minisequencing

Ten years after SBE in solid-phase was published by Syvänen in the beginning of the 1990s (Syvänen *et al.*, 1992), more effective and robust four-color microarray-based SBE assays were developed by the same group (Hirschhorn *et al.*, 2000; Lindroos *et al.*, 2002). In a study by Lindroos *et al.,* 125 candidate SNPs were amplified in three-plex PCR reactions, amplicons were pooled together, and unused primers and dNTPs were removed using Exonuclease I and shrimp alkaline phosphatase, respectively. A cyclic SBE reaction was performed, using primers with 5'-*tag* sequences, followed by capture of the products on microarrays by hybridization to *ctag* oligonucleotides (Fig. 4C). Before *tag-ctag* hybridization and signal detection, cyclic SBE was performed in liquid phase and repeated 34 times. The advantage of this method lies in the signal amplification during the cycling procedure and in parallel DNA analysis on a microarray, using a custom chamber system. Nevertheless, as a disadvantage, the method demands a substantial amount of hands-on work.

### 1.5.1.6. Arrayed Primer Extension

Arrayed primer extension (APEX) principle was primarily described by Shumaker and colleagues (Shumaker *et al.*, 1996) and Pastinen (Pastinen *et al.*, 1997). It was called SBE in solid-phase, wherein template DNA was annealed to an oligonucleotide array, extended with $\alpha^{32}$-P dNTPs and analyzed by a Phosphor Imager. The grid position of the oligonucleotide identified the mutation site and the extended base identified the mutation (Shumaker *et al.*, 1996). Later, when fluorescently labeled ddNTPs were used (Kurg *et al.*, 2000), simultaneous four-color multiplex primer extension on an array was available because the mutation was detected by a change in the color code of the primer sites, where each ddNTP is provided with a different fluorescent label (Fig. 4E). Although multiplex SBE on an array enabled a high signal-to-noise ratio (40:1), in the analysis of tens to hundreds of positions, a key technical obstacle was the PCR amplification step in a very low plex level. The exception was resequencing APEX assays, where many exons can be amplified simultaneously by three PCR reactions and the PCR throughput was not a limitation (Tõnisson *et al.*, 2002).

In January of 2008, a summarizing report was published about the APEX assay and using multiplex PCR, wherein the completion rate, call rate and the accuracy of APEX were objectively evaluated (Podder *et al.*, 2008). Fifty SNPs across the hundreds of HapMap Coriell DNA samples were amplified (products 100 to 700 bp) in a total of 7 multiplex PCRs. Good results for 41 SNPs, with 99.8% genotype concordance, at a call rate of 94.9% (Podder *et al.*, 2008) were obtained. The call rate and accuracy increasing factor were presented in the same paper. In their second experiment, involving modifications in the initial DNA amplification step, a single 50-plex PCR was achieved. 50-plex amplification (products 100 to 200 bp) across 49 DNA samples demonstrated 100% assay completion rate, 100% call rate and >99.9% accuracy using 5 ng of input

DNA. The modifications involved shortening the amplicon lengths to less than 200 bp, avoiding of SNPs at the primer-site (Tebbutt and Ruan, 2008), and the addition of a common linker sequence (*tag*) at its 5′ end (5'-TACGACTC ACTTAGGGAG-3' 18 bp, $T_m$ 54ºC) for each of the left hand primers and at its 3' end (5'-CGATGTAGGTGACACTAG-3' 18 bp, $T_m$ 54ºC) for each of the right hand PCR primers. The *tags* have many properties: a balanced GC content (50%) to increase the melting temperature of the primer in AT rich regions, reduced primer-dimer formation during PCR, and a unique sequence which is not found in the human DNA template (Brownie *et al.*, 1997; Wang *et al.*, 1998; Wang *et al.*, 2005; Podder *et al.*, 2008). In the reaction, after the first cycles of PCR, the *tag* sequence becomes incorporated into the amplicon and is amplified together with the template sequence.

#### 1.5.1.7. SNaPshot® from Applied Biosystems

The SNaPshot® multiplex kit is a primer extension system that enables multiplexing of up to 10 SNPs at known and previously amplified locations and offers a one-tube SBE reagent to label DNA fragments. Labeled fragments are analyzed with a (gel-based) capillary electrophoresis instrument using the size standard reference marker. Due to an unknown set of ddNTP conjugates, this fully commercial system also requires Applied Biosystems' capillary analyzers for detection. In summary, although ABI analyzers are widely used in research and clinical labs and the SNaPshot® kit has been successful, the system has two disadvantages: (i) The customer is fully dependent on ABI, which reflects in its price and (ii) sometimes the assay needs additional optimization for high quality results (Pati *et al.*, 2004).

# 2. PRESENT INVESTIGATIONS, RESULTS AND DISCUSSION

A four-color APEX genotyping method was published in 2000, but years later it became clear that despite the brilliant SBE solution, in which multiplex was achievable on a microarray, the template preparation procedure still required considerable labor and laboratory resources. The available multiplex-PCR protocols were modified for APEX applications but the throughput of sample preparation did not satisfy expectations: reagent and plastic costs stayed almost at the same level, each multiplex (2 to 5-plex) demanded individual optimization, products required visualization on agarose gels, and there was a lack of reliability. The limitations of multiplex PCR merited ongoing scientific attention and listed it among the top of problems demanding molecular engineering solutions (FP6 MolTools).

There were different scientific and technical obstacles which demanded solutions in order to develop a high-level multiplex PCR for mutation and SNP calling using the APEX principle:

1. Customized *in silico* primer design and product prediction software together with a unique (non-binding in the human genome) and highly efficient universal primer were required.
2. Reaction conditions for multiplex PCR. To ensure low side-product levels in multiplex reactions, all parameters for PCR mixture and cycling conditions needed systematic optimization. The effect of high-level multiplex PCR conditions on primer hybridization stringency had not been previously characterized *in situ*.

The primer design was performed by the modified version of the GenomeTester 1.3 software (Andreson *et al.*, 2006). Specificity analysis was performed by analysis of the results from first 10- and 66-plex reactions. To overcome the multiplex PCR limitations, we created a new primer construct, which could be used simultaneously in the template DNA amplification and detection step on an array. Microarray-specific SBE primers were supplied with universal sequences. Two separate amplification reactions in a single tube, specific- and universal primer PCR, were performed. Such primer constructs and optimized protocols enabled us to uphold the principle that one DNA sample could be amplified in a single tube. We performed studies wherein 124- (Ref. IV) and 640-plex (Ref. II) PCR was applied to analyze the respective number of genetic markers.

The name APEX originally meant only *arrayed primer extension* where "traditional" PCR was used in the sample preparation step. Here, the described multiplex PCR principle is an improvement on the basic APEX amplification protocol, and indicates that (an improved) high-level and specific multiplex PCR and SBE detection are incorporated into the second version of APEX (APEX-2).

## 2.1. APEX-2 primer construction and reaction protocol

### 2.1.1. APEX-2 primer construct and working schema (Ref. I and II)

APEX-2 is a nucleic acid variation detection method, which uses multiplex PCR as template amplification and SBE on an array as the detection step. The APEX-2 protocol consists of three main parts – primer constructs, nested-PCR-like cycling, and optimal reaction conditions. Two primers for the amplification of one locus are used and they differ from each other in by the specific region. The universal sequence and amino-modified 5' end are common to all primers, like those shown in Table 1 in Ref. II. The $T_m$ of the specific regions is in the range of 57–62ºC[18] and the $T_m$ of the 21 bp universal sequence is 58ºC[19].

The primer binding location and the amplification schema are depicted in Fig. 5 (Fig. 1 in Ref. I). APEX-2 primers bind specifically to genomic DNA sequences immediately up- or downstream of the position of interest (Fig. 5A). After primer extension, the synthesized product contains the complementary sequence of the respective APEX-2 primer and the position of interest (SNP/mutation). The second cycle of primer extension generates a complementary strand for the universal sequence. Next, during template amplification (Fig. 5B), the universal primer hybridizes to the 3' end of the previously generated product, followed by microarray genotyping using SBE. It can be noted that APEX-2 primers also have a 5'-amino modification, enabling spotting and immobilization on the microarray. The purified phase 2 PCR product hybridizes to the immobilized APEX-2 primers and genotyping as a four-color SBE reaction is performed (Fig. 5C). The maximum amplification multiplex level which we have achieved is 640-plex (Ref. II) and many smaller plexes in other studies, such as 124-plex in a forensic testing assay (Ref. IV), 600-plex in a cleft-lip and palate case-control study (manuscript in review), 350-plex in a coronary artery disease case-control study (manuscript in review), and 68-plex in a practical course of biotechnology for undergraduate students (unpublished protocol).

[18] Calculated using Primer3 software from http://biotools.umassmed.edu/bioapps/primer3_www.cgi
[19] Basic calculation from www.basic.northwestern.edu/biotools/oligocalc.html

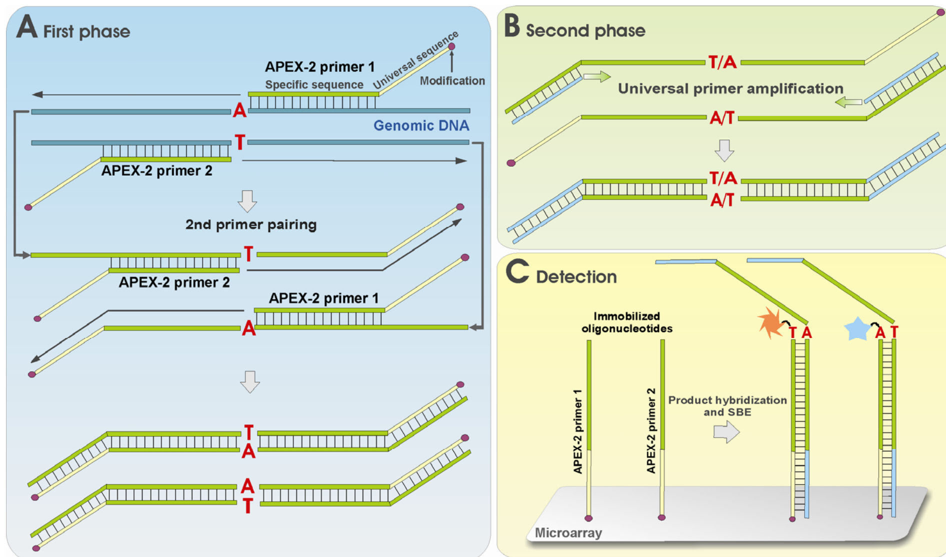



**Figure 5.** The principle of the APEX-2 assay. APEX-2 primers bind to genomic DNA immediately upstream of the position of interest (SNP/mutation/INDEL). After primer extension, the synthesized sequence contains the complementary strand of the respective APEX-2 primer and the position of interest. (**B**) The universal primer hybridizes to the 3’ end of the previously-generated product during template amplification. (**C**) The purified universal amplification phase PCR products hybridizes to the immobilized APEX-2 primers that have a 5’-amino modification. The 5’ amino group enables spotting and immobilization of the primers on the array surface. Genotyping is performed as a four-color single-base extension reaction. Modification from Fig. 1 in Ref. II.

Advantages of such primer constructs, where specific and universal sequences are incorporated and supplied with amino modification, are time and cost effective and enable to create 10- to 100-fold higher multiplexes than previously demonstrated (Lahermo *et al.*, 2006; Sanchez *et al.*, 2006; Meuzelaar *et al.*, 2007; Podder *et al.*, 2008). The foundation of this advancement include: (i) generation of short products (~100 bp) due to the close proximity of hybridizing APEX-2 primers (primers are only separated by the position of interest) and (ii) incorporation of the universal primer in the APEX-2 oligonucleotide sequence, allowing phase two PCR amplification under optimal conditions, which is not feasible using PCR primers containing only unique sequences. Four core factors of APEX-2 design allow to increase signal specificity and strength compared with alternative techniques: (i) APEX-2 primers anneal to both strands of genomic DNA within 26 bp up- and downstream of the position of interest; (ii) universal primer PCR amplification favors a high yield of short products (~100 bp); (iii) the universal primer has no homologous binding sites in published human genome sequences. Therefore, templates lacking the universal

primer sequence will not be amplified and (iv) prior to SBE, the probe-target hybridization control step on a solid surface ensures a high signal-to-noise ratio. While primer specificity is assured by a short region of interaction with genomic DNA (median of 26 bp), microarray hybridization involves a longer region of interaction (median 48 bp). Therefore, microarray specificity is influenced by probe length and distance from the 3'end of the primer (Suzuki *et al.*, 2007), as well as distance between the complementary region and the solid surface of the array.

APEX-2 analysis is unsuitable only for analysis of SNPs and mutations, (i) where multiple short amplicons could be generated during PCR using one praimer pair, (ii) for very similar flanking sequences, due to the cross-hybridization considerations, and (iii) for DNA methylation analysis. Methylation analysis is complicated only in case of methylated sites locate close together and there is "no room" for primers.

## 2.1.2. Multiplex PCR

### 2.1.2.1. PCR components (Ref. II and IV)

The optimization of reaction components in the first, specific PCR step required systematic research to find out appropriate *Taq* polymerase, buffer, primer and $Mg^{++}$ concentrations for high quality specific multiplex PCR. The second phase, universal primer PCR, is a fast-moving reaction producing high yield and reliable amplicons for SBE (Page 3 in Ref. II and page 3 in Ref. IV). TrueStart[20] polymerase has been used in all APEX-2 applications because of its high specificity and sensitivity. This enzyme provides the convenience of room temperature reaction set-up, has a 1 minute activation time and high stability in long-lasting in PCR. Our multiplex assay also requires $K^+$ and $NH_4^+$ in buffer at a 50–70 mM concentration each and 4.75 mM[21] $Mg^{++}$. The effect of KCl on multiplex PCR is previously characterized (Henegariu *et al.*, 1997) and the concentration we used correlates with prior published reports where high-level PCR has been achieved using 50 mM KCl (Wang *et al.*, 2005; Sanchez *et al.*, 2006; Luo *et al.*, 2009). In addition, commercial producers[22,23] have designed their multiplex mastermixes at the same level. The other buffer component, $NH_4^+$, is also widely used at a regular concentration in the range of 5 to 15 mM (15 mM in the APEX-2 assay). Using concurrently offers a balance in reaction specificity and product yield for multiplex PCR and available premade master-

[20] TrueStart™ Hot Start *Taq* DNA Polymerase from Fermentas Ltd. (Vilnius, Lithuania)
[21] May vary in different applications and depend on plex size.
[22] NewEngland Biolabs multiplex mastermix: www.neb.com/nebecomm/products/productM0284.asp
[23] Epicentre MasterAmp PCR PreMix: www.epibio.com/f5_3/f5_3mb.asp

mixes[24,25] follows the same principle, albeit the precise concentrations of those are not available.

Minimal specific primer concentrations can start from 30 nM of each primer in the case of larger multiplexes (Ref. II, IV) and up to ~200 nM for tens of plexes (not published) which is similar to the primer concentration in "traditional" PCR (100 nM and higher).

#### 2.1.2.2. Studied DNA amount and quality (Ref. II, IV and V)

The required amount of starting DNA was investigated in Ref. IV and Ref. II. Six hundred forty-plex APEX-2 and 124-plex forensic assays require 150 ng per reaction (0.25 ng/locus, 42 human genome equivalents per locus) and 50 ng (0.4 ng/locus) of total DNA, respectively. Dr. Podder and colleagues have published 50-plex PCR and APEX-based genotyping (Podder *et al.*, 2008) which another group has used for 65-plex sperm-based genotyping (Luo *et al.*, 2009) using only 5 ng of starting DNA (~0.1 ng/locus). Outstanding results in that field have been achieved in genetic forensics where the DNA amount is limited and the reaction conditions are pushed to their limits. Fifty two-plex amplification and gel-based detection were performed using as low as 500 pg of starting DNA (0.001 ng/locus) (Sanchez *et al.*, 2006). Remarkable in this research was the fact that the dNTP concentration was 700 µM, and $MgCl_2$ was 8 mM, although the manufacturer-recommended $MgCl_2$ optimum is 1–4 mM. However, if dNTP concentration is more than three times higher than recommended, it is well balanced with higher $MgCl_2$ concentration. The positive effect of high dNTP concentration is unknown, but after careful optimization of other reaction components, it is possible to use this to relieve the scarcity of template DNA (Sanchez *et al.*, 2003; Sanchez and Endicott, 2006; Sanchez *et al.*, 2006; Sanchez *et al.*, 2008). Unfortunately, the influence of higher dNTP concentration on the other PCR enzymes in multiplex applications has not been studied.

Using the genealogical APEX-2 assay (Ref. IV), we compared six DNA extraction methods for obtaining DNA from whole blood and saliva and published the results in Ref. V. Saliva is an alternative to blood DNA samples because saliva collection is a painless procedure with no risk of disease transmission and no requirements for specialized medical personnel. Also, saliva collection allows wider population sampling because it is possible to collect DNA samples by mail (Freeman *et al.*, 1997; Hansen *et al.*, 2007). First, we were curious about DNA quality parameters between commercially available saliva DNA extraction kits. Secondly we had the hypothesis that different DNA extraction methods can yield different ratios of mtDNA and autosomal

[24] Gene Choice multiplex mastermix: www.geneseesci.com/ProductInfo.php?productID=42-301
[25] Qiagen booklet: www1.qiagen.com/literature/brochures/pcr/pdf/pcrcha35.pdf

DNA. In Fig. 1, Ref. V we demonstrated all measured parameters and reported different mitochondrial/genomic DNA ratios than in one previous study (Guo *et al.*, 2009) where similar results were measured and significant differences were detected. In the context of PCR, the copy number of target DNA, especially Y chromosome DNA, may influence the multiplex PCR outcome if different targets are used simultaneously. Also it can influence primer pair dropout at a higher multiplex level or lower template DNA concentrations and has an impact if the DNA extraction method is replaced in an already working assay. In addition, we detected significant differences in control-system PCR efficiency and total yield of extracted DNA (especially in the case of one saliva DNA extraction kit). Although there were variations between the obtained DNA extraction methods, no significant differences in the efficiency of multiplex APEX-2 PCR and oligomicroarray signals after SBE were detected. Thus, DNA extraction from saliva is a user-friendly and reliable alternative to DNA extraction from blood to obtain DNA for the use in further applications, such as the APEX-2 genotyping assay in science, molecular diagnostics and in genealogical studies.

#### 2.1.2.3. Cycling conditions (Ref. II, III and IV)

To perform such complicated reaction as multiplex-PCR, flexible and stringent cycling conditions are required to assure high success rate within different amplicons and minimize side-products. Diverse PCR cycle principles, which have been utilized in different studies, are summarized in Fig. 6. Variant **A,** optimized for a 100-plex study (Ref. III) using low ramp speed between annealing and extension, was chronologically the first. Variant **B** was developed to control the annealing step more precisely, but the lowest annealing point after denaturation step may cause nonspecific hybridization (Ref. II, IV). Version **C** is used in commercial assays[26] and the latest, variant **D,** has been used in a practical student course and seems promising for future studies. In all cases, hybridization is done by using a touchdown amplification protocol, where the lowest annealing temperature is easily controllable and offers a nice balance between a high success rate and a low amount of side products.

Working with relatively low specific primer concentrations (30 nM in 640-plex in Ref. II) the annealing time is a key parameter in the multiplex PCR cycle. Primers have various hybridization efficiencies, due to their thermodynamic parameters ($\Delta H$, $\Delta S$ and $\Delta G$ (Sugimoto *et al.*, 1996)) and genomic context (number of binding sites on genome). In a multiplex reaction, all oligonucleotides are mixed and they hybridize with their specific part simultaneously, producing amplicons with different efficiencies. The conditions are similar to "less primer method" multiplex where the decreased primer con-

[26] Asper Biotech Ltd. (Tartu, Estonia)

centration (2.5–5% from initial) is compensated with longer annealing time (5 min) and using 40 cycles (Kane *et al.*, 2007). As a result, smaller amount of primers with a higher number of PCR cycles permits the specific amplification due to the plateau effect: efficiently amplified loci reach to the plateau during early PCR cycles, the remaining PCR cycles employ to the production of lower efficient loci. Similar results were observed in APEX-2 optimization experiments where 4 to 5 minutes annealing time was sufficient to achieve high success rate within different amplicons. In addition, annealing stringency was ensured by the increasing/decreasing temperature gradient, which minimized the yield of side-products. The temperature gradient effect in multiplex conditions in annealing phase has still not in depth been characterized.

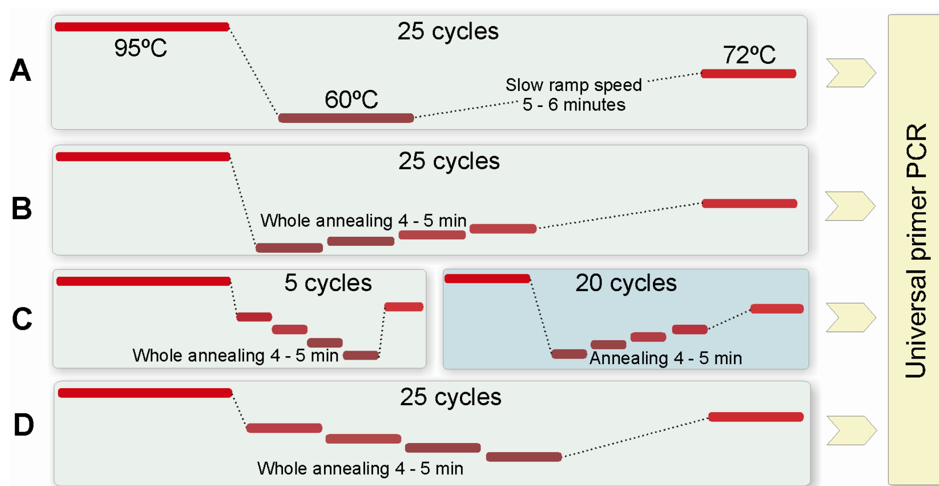



**Figure 6.** Different PCR cycling principles for the APEX-2 reaction. The Y axis depicts the annealing temperature in different PCR cycles.

## 2.2. Quality parameters of the APEX-2 assay (Ref. II, IV)

The specificity representing "correctly identified negative calls" is generally not under question, as missing calls can be caused by reasons such as a polymorphic site under the primer or microarray production errors, which do not reflect a failure of the detection method. To test the sensitivity of APEX-2, we performed a concordance study with the Illumina genotyping method, family trio Mendelian heredity testing, and Y-chromosome DNA and mtDNA haplogroup determination using direct Sanger sequencing as a control method.

For the Illumina concordance study (Ref. II, page 5) 173 SNPs in 19 individuals were observed. From a total of 3242 comparisons, 3195 matches and 47

genotyping errors were detected. To analyze discordant calls with a third method, we chose 4/29 discordant SNPs (in 10 error-causing loci) in which case the polymorphic position was located at the restriction cleavage site. The restriction results showed that APEX-2 had generated six errors, the Illumina platform made two mistakes, and in two cases the calls of both methods differed from the results obtained from the restriction analysis.

In Ref. IV (page 3–4), Mendelian heredity testing was performed to compare family genotypes by Y-DNA patrilineal, mitochondrial matrilineal, and autosomal analyses. Family scanning of 1783 positions exposed three Mendelian inconsistencies in autosomal loci where the heterozygosity of the child had been left undetected by the genotyping software. In the same report, internal control DNA samples (wherein genotyping was performed previously with APEX) were under investigation and one typing error was detected within 562 comparisons. Hence, the autosomal SNP concordance rate across family trios and internal controls was estimated as 99.83%.

The third possible criterion of genotyping accuracy, which was used in Ref. IV to characterize the sensitivity of APEX-2, was mtDNA and Y-DNA haplogroup determination. This can be used to define the genetic population origin of studied DNA of interest. Y-DNA is passed solely along the patrilineal line, from father to son, while mtDNA is passed down the matrilineal line, from mother to offspring. Neither recombines, and thus Y-DNA and mtDNA change only by chance mutations in each generation with no intermixture between parents' genetic material. Each haplogroup is the collection of well-characterized genetic loci (SNPs) and all deviations were analyzed using direct sequencing. Six markers, which needed a confirmation of the assay in haplogroup analysis, were studied. No discordance between APEX-2 genotyping and direct sequencing results were found.

## 2.3. APEX-2 applications and feedback (Ref. III–V)

APEX-2 has found usage in many projects and has been tested in various conditions and in different plex sizes. In our first research project (Ref. III), a total set of studied markers (624 SNPs) was divided into six individual 104-plex PCR reactions and analyzed on 347 individuals. Using first preliminary APEX-2 protocol, the allele drop-out rate of 624-plex was ~35% that did not satisfy our expectations. This has led us to crate 104-plex groups. Finally, after amplification, all six PCR-plexes were combined and prepared for oligomicroarray hybridization as a separate step before SBE (Ref. III, Supplement X). That was a labour-intensive procedure, due to the hybridization and multi-step washing. As an alternative, column purification and a regular APEX SBE protocol could have been used. In 2006, due to the lack of commercial purification columns, which can only recover short PCR products (75–120 bp), this strategy was not

employed. Later, appropriate purification columns[27], providing a giving high-quality template for SBE with minimal time consumption, simplified the PCR product purification procedure.

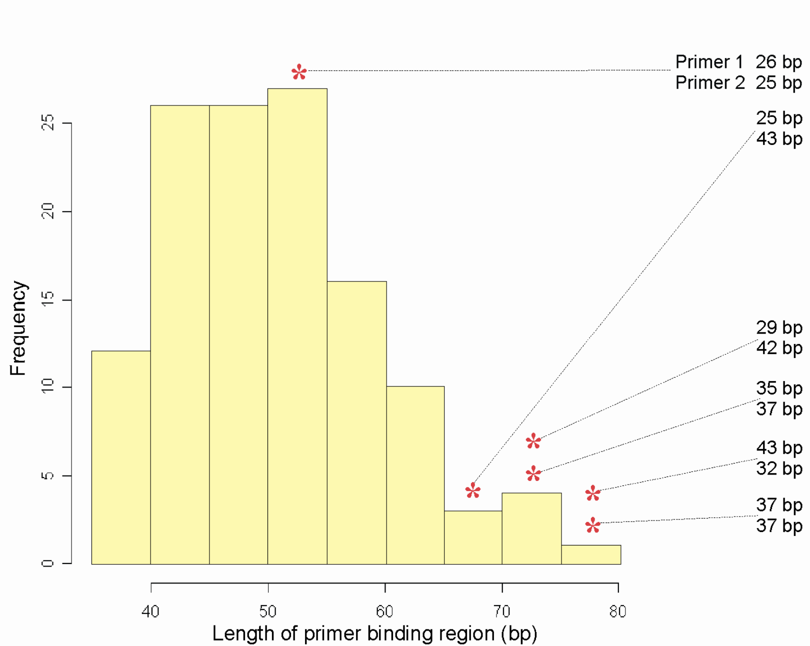



**Figure 7.** Histogram illustrating the distribution of 124 specific regions' lengths and autosomal drop out positions after validation (one star represents a single locus). 124 amplicons are divided into groups according to the length of primer binding region and displayed as frequency of representation (y-axis). The primer binding region, on x-axis, is calculated as a sum of both lengths of APEX-2 primer specific sequences plus the locus of interest (SNP = 1 bp). Specific region lengths of each locus oligonucleotide are displayed on the right.

The original APEX-2 heredity assay (Ref. IV) in which three different targets (autosomal DNA, Y-DNA and mtDNA) were detected in a single tube was developed in 2008. The 124-plex assay was designed for the simultaneous genotyping of SNPs from single copy loci (46 autosomal and 29 Y chromosomal genetic markers) and SNPs from the mitochondrial genome (49 markers). As the mitochondrial genome is present in up to thousands of copies per cell in different tissue types, the concentrations of specific primers in the first PCR phase have to compensate for this abundance. For that assay design, our knowledge about primer design and reaction conditions had improved and the assay success rate was 94%. The initial number of loci was 132 and 124 of them passed validation. A relatively high failure rate (6 of 8) in autosomal positions

[27] NucleoSpin® Extract II from Macherey-Nagel (Düren,Germany)

may have been caused by the fact that these loci were taken from the SNP*for*ID Consortium database[28] (Sanchez *et al.*, 2006). Validation of these loci was performed using gel-based length separation after SBE. The GC% around SNPs diverged greatly and caused many APEX-2 primer pairs to fail in AT-rich regions (long products). The failed products are shown in Fig. 7 as red stars on the primer binding region length diagram (modified from Ref. IV Fig. I; the drop out information can be found in Supplementary Table 1). In addition to Ref. IV, other APEX-2 studies have indicated that the optimal length of the specific region is approximately 25 bp, but should not be shorter than 20 bp or exceed 30 bp level and should fall between 58–65ºC $T_m$, if possible (unpublished data).

## 2.4. Closing remarks

APEX-2 is a flexible, highly multiplex and SBE based genotyping method. The primer design depends greatly on candidate gene/region or replication study loci, which may be located at SNP-, AT- or GC-rich regions (Ref. IV) making the analysis more complicated. Fig. 9 in Ref I and Fig. 3 in Ref. II depict the primer design concept in SNP-rich regions, where the polymorphic site is located under the primer binding sequence, and demonstrates how to analyze insertion or deletion polymorphisms in such cases (details in Ref. IV, Supplementary Table 1). It has become clear that for successful analysis of a certain locus, one of the APEX-2 primers needs to be redesigned if basecalling of genetic marker calling fails in the first validation step. Therefore, in order to increase the specificity of the primer, another binding site ought to be chosen for that certain locus. After that, only one primer is left for SBE detection on a microarray (unpublished data). Comparing APEX-2 as a medium-scale specific multiplex PCR method with some previously published protocols (Wang *et al.*, 2005; Fredriksson *et al.*, 2007), one can see that APEX-2 is robust, needs less molecular engineering steps, and requires only two oligonucleotides per locus detection.

The microarray era was roughly a dozen years old, when manufacturers made initial efforts to create arrays, which now enable one to analyze hundreds of thousands of SNPs simultaneously. An astounding expansion of arrays to many research fields, has brought an enormous amount of knowledge to the field of human genomics. The peak of the microarray expansion took place in the summer of 2007 when Affymetrix and Illumina released high-density microarrays (LaFramboise, 2009). This is when SGS platforms were introduced and the microarrays, previously the most promising genomic research tools, lost the leading position. Although the long-term future of large-scale SNP arrays is unclear, the cost considerations still ensure their relevance for at least the next

[28] www.snpforid.org/

several years. However, there is still great potential for (i) replication study tools, where meaningful loci after whole-genome analysis need confirmation, (ii) different clinical genetic testing tools (e.g mutation specific of copy-number variation arrays), and (iii) public study tools for genealogy and personalized medicine testing, which seems to be in the beginning of an expansion as a new field (Ikediobi, 2009). APEX-2 is a robust and reliable method, which has the potential to be cost-effective if the number of loci of interest is in the range of tens to a thousand and 50 ng of starting genomic DNA is attainable. Its flexibility does not only mean the ability to design original primers to specific sites, but also the fact that custom-made arrays, where loci exchange or supplementation is not complicated, can be used.

# CONCLUSIONS

I. I have developed, optimized and applied a novel and flexible genotyping tool, APEX-2, which enables one to amplify hundreds of studied genetic loci in a single reaction vessel and to analyze the products on a single microarray.

II. Nineteen hundred APEX-2 primer pairs were tested in various genome contexts (SNP-, AT- or GC-rich) in order to work out basic rules for the APEX-2 assay development. The assay enables one to detect closely located polymorphisms or INDELs without primer pair drop-out or lower calling quality. Different DNA sources (genomic, mitochondrial, and Y chromosome DNA) have been tested successfully for simultaneous analysis in APEX-2 application (Ref. IV). Additionally, DNA extraction methods from blood and saliva, as an initial step of the genotyping procedure, have been studied from several aspects (Ref. V).

III. APEX-2 has been used as a genotyping method in four published original reports. The method was described in an international patent (2009) (Ref. I) and later published as a research paper (Ref. II). Forensic testing in a 124-plex application assay (Ref. IV) characterized the influence of calling signals in the multiplex reaction if target DNA concentrations are not equal in the multiplex reaction. An allele and haplotype diversity study, (Ref. III), is an example wherein methods like APEX-2 can be applied.

# REFERENCES


1. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ, Weinstock GM, Gibbs RA. 2007. Direct selection of human genomic loci by microarray hybridization. Nat Methods 4(11): 903–5.
2. Allingham-Hawkins D. 2008. Successful Genetic Tests Are Predicated on Clinical Utility. Genetic Engineering & Biotechnology News 28(14).
3. Andreson R, Reppo E, Kaplinski L, Remm M. 2006. GENOMEMASKER package for designing unique genomic PCR primers. BMC Bioinformatics 7:172.
4. Asari M, Watanabe S, Matsubara K, Shiono H, Shimizu K. 2009. Single nucleotide polymorphism genotyping by mini-primer allele-specific amplification with universal reporter primers for identification of degraded DNA. Anal Biochem 386(1): 85–90.
5. Bau S, Schracke N, Kranzle M, Wu H, Stahler PF, Hoheisel JD, Beier M, Summerer D. 2009. Targeted next-generation sequencing by specific capture of multiple genomic loci using low-volume microfluidic DNA arrays. Anal Bioanal Chem 393(1): 171–5.
6. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ and others. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456(7218):53–9.
7. Brouqui P, Vu Hai V, Nougairede A, Lagier JC, Botelho E, Ninove L, Zandotti C, Charrel RN, De Lamballerie X, Raoult D. 2009. Improving the diagnostic efficiency of H1N1 2009 pandemic flu: analysis of predictive clinical signs through a prospective cohort. PLoS Curr Influenza:RRN1120.
8. Brownie J, Shawcross S, Theaker J, Whitcombe D, Ferrie R, Newton C, Little S. 1997. The elimination of primer-dimer accumulation in PCR. Nucleic Acids Res 25(16):3235–41.
9. Bruse S, Moreau M, Azaro M, Zimmerman R, Brzustowicz L. 2008. Improvements to bead-based oligonucleotide ligation SNP genotyping assays. Biotechniques 45(5): 559–71.
10. Budimlija ZM, Prinz MK, Zelson-Mundorff A, Wiersema J, Bartelink E, MacKinnon G, Nazzaruolo BL, Estacio SM, Hennessey MJ, Shaler RC. 2003. World Trade Center human identification project: experiences with individual body identification cases. Croat Med J 44(3):259–63.
11. Canova C, Hashibe M, Simonato L, Nelis M, Metspalu A, Lagiou P, Trichopoulos D, Ahrens W, Pigeot I, Merletti F, Richiardi L, Talamini R, Barzan L, Macfarlane GJ, Macfarlane TV, Holcatova I, Bencko V, Benhamou S, Bouchardy C, Kjaerheim K and others. 2009. Genetic associations of 115 polymorphisms with cancers of the upper aerodigestive tract across 10 European countries: the ARCAGE project. Cancer Res 69(7):2956–65.
12. Chamberlain JS, Gibbs RA, Ranier JE, Nguyen PN, Caskey CT. 1988. Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. Nucleic Acids Res 16(23):11141–56.
13. Chen X, Levine L, Kwok PY. 1999. Fluorescence polarization in homogeneous nucleic acid analysis. Genome Res 9(5):492–8.

14. Doi Y, Yamamoto Y, Inagaki S, Shigeta Y, Miyaishi S, Ishizu H. 2004. A new method for ABO genotyping using a multiplex single-base primer extension reaction and its application to forensic casework samples. Leg Med (Tokyo) 6(4):213–23.
15. Droege M, Hill B. 2008. The Genome Sequencer FLX System--longer reads, more applications, straight forward bioinformatics and more complete data sets. J Biotechnol 136(1–2):3–10.
16. Drury KC, Liu MC, Zheng W, Kipersztok S, Williams RS. 2000. Simultaneous single-cell detection of two mutations for cystic fibrosis. J Assist Reprod Genet 17(9):534–9.
17. Edwards MC, Gibbs RA. 1994. Multiplex PCR: advantages, development, and applications. PCR Methods Appl 3(4):S65–75.
18. el-Hashemite N, Wells D, Delhanty JD. 1997. Single cell detection of beta-thalassaemia mutations using silver stained SSCP analysis: an application for preimplantation diagnosis. Mol Hum Reprod 3(8):693–8.
19. Elnifro EM, Ashshi AM, Cooper RJ, Klapper PE. 2000. Multiplex PCR: optimization and application in diagnostic virology. Clin Microbiol Rev 13(4):559–70.
20. Fredriksson S, Baner J, Dahl F, Chu A, Ji H, Welch K, Davis RW. 2007. Multiplex amplification of all coding sequences within 10 cancer genes by Gene-Collector. Nucleic Acids Res 35(7):e47.
21. Freeman B, Powell J, Ball D, Hill L, Craig I, Plomin R. 1997. DNA by mail: an inexpensive and noninvasive method for collecting DNA samples from widely dispersed populations. Behav Genet 27(3):251–7.
22. Gabriel S, Ziaugra L, Tabbaa D. 2009. SNP genotyping using the Sequenom MassARRAY iPLEX platform. Curr Protoc Hum Genet Chapter 2:Unit 2 12.
23. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 27(2):182–9.
24. Golenberg EM, Bickel A, Weihs P. 1996. Effect of highly fragmented DNA on PCR. Nucleic Acids Res 24(24):5026–33.
25. Goossens D, Moens LN, Nelis E, Lenaerts AS, Glassee W, Kalbe A, Frey B, Kopal G, De Jonghe P, De Rijk P, Del-Favero J. 2009. Simultaneous mutation and copy number variation (CNV) detection by multiplex PCR-based GS-FLX sequencing. Hum Mutat 30(3):472–6.
26. Grubwieser P, Muhlmann R, Berger B, Niederstatter H, Pavlic M, Parson W. 2006. A new "miniSTR-multiplex" displaying reduced amplicon lengths for the analysis of degraded DNA. Int J Legal Med 120(2):115–20.
27. Guo W, Jiang L, Bhasin S, Khan SM, Swerdlow RH. 2009. DNA extraction procedures meaningfully influence qPCR-based mtDNA copy number determination. Mitochondrion 9(4):261–5.
28. Haff LA, Smirnov IP. 1997. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. Genome Res 7(4):378–88.
29. Handyside AH, Kontogianni EH, Hardy K, Winston RM. 1990. Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification. Nature 344(6268):768–70.
30. Hansen TV, Simonsen MK, Nielsen FC, Hundrup YA. 2007. Collection of blood, saliva, and buccal cell samples in a pilot study on the Danish nurse cohort:

comparison of the response rate and quality of genomic DNA. Cancer Epidemiol Biomarkers Prev 16(10):2072–6.
31. Hardenbol P, Baner J, Jain M, Nilsson M, Namsaraev EA, Karlin-Neumann GA, Fakhrai-Rad H, Ronaghi M, Willis TD, Landegren U, Davis RW. 2003. Multiplexed genotyping with sequence-tagged molecular inversion probes. Nat Biotechnol 21(6):673–8.
32. Henegariu O, Heerema NA, Dlouhy SR, Vance GH, Vogt PH. 1997. Multiplex PCR: critical parameters and step-by-step protocol. Biotechniques 23(3):504–11.
33. Hirschhorn JN, Sklar P, Lindblad-Toh K, Lim YM, Ruiz-Gutierrez M, Bolk S, Langhorst B, Schaffner S, Winchester E, Lander ES. 2000. SBE-TAGS: an array-based method for efficient single-nucleotide polymorphism genotyping. Proc Natl Acad Sci U S A 97(22):12164–9.
34. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, McCombie WR. 2007. Genome-wide in situ exon capture for selective resequencing. Nat Genet 39(12):1522–7.
35. Ikediobi ON. 2009. Personalized medicine: are we there yet? Pharmacogenomics J 9(2):85.
36. Inagaki S, Yamamoto Y, Doi Y, Takata T, Ishikawa T, Imabayashi K, Yoshitome K, Miyaishi S, Ishizu H. 2004. A new 39-plex analysis method for SNPs including 15 blood group loci. Forensic Sci Int 144(1):45–57.
37. Jaakson K, Zernant J, Külm M, Hutchinson A, Tõnisson N, Glavac D, Ravnik-Glavac M, Hawlina M, Meltzer MR, Caruso RC, Testa F, Maugeri A, Hoyng CB, Gouras P, Simonelli F, Lewis RA, Lupski JR, Cremers FP, Allikmets R. 2003. Genotyping microarray (gene chip) for the ABCR (ABCA4) gene. Hum Mutat 22(5):395–403.
38. Jiang Z, Zhang X, Deka R, Jin L. 2005. Genome amplification of single sperm using multiple displacement amplification. Nucleic Acids Res 33(10):e91.
39. Kaderali L, Deshpande A, Nolan JP, White PS. 2003. Primer-design for multiplexed genotyping. Nucleic Acids Res 31(6):1796–802.
40. Kane M, Masui S, Nishi K. 2007. Application of less primer method to multiplex PCR. Forensic Science International: Genetics Supplement Series 1(1):41–43.
41. Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, Liu G, Su X, Cao M, Chen W, Zhang J, Liu W, Yang G, Di X, Ryder T, He Z, Surti U, Phillips MS, Boyce-Jacino MT, Fodor SP, Jones KW. 2003. Large-scale genotyping of complex DNA. Nat Biotechnol 21(10):1233–7.
42. Kuppuswamy MN, Hoffmann JW, Kasper CK, Spitzer SG, Groce SL, Bajaj SP. 1991. Single nucleotide primer extension to detect genetic diseases: experimental application to hemophilia B (factor IX) and cystic fibrosis genes. Proc Natl Acad Sci U S A 88(4):1143–7.
43. Kurg A, Tõnisson N, Georgiou I, Shumaker J, Tollett J, Metspalu A. 2000. Arrayed primer extension: solid-phase four-color DNA resequencing and mutation detection technology. Genet Test 4(1):1–7.
44. Kwok PY. 2002. SNP genotyping with fluorescence polarization detection. Hum Mutat 19(4):315–23.
45. Kwok S, Mack DH, Mullis KB, Poiesz B, Ehrlich G, Blair D, Friedman-Kien A, Sninsky JJ. 1987. Identification of human immunodeficiency virus sequences by using in vitro enzymatic amplification and oligomer cleavage detection. J Virol 61(5):1690–4.

46. Kõks S, Nikopensius T, Koido K, Maron E, Altmäe S, Heinaste E, Vabrit K, Tammekivi V, Hallast P, Kurg A, Shlik J, Vasar V, Metspalu A, Vasar E. 2006. Analysis of SNP profiles in patients with major depressive disorder. Int J Neuropsychopharmacol 9(2):167–74.
47. LaFramboise T. 2009. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic Acids Res 37(13): 4181–93.
48. Lahermo P, Liljedahl U, Alnaes G, Axelsson T, Brookes AJ, Ellonen P, Groop PH, Hallden C, Holmberg D, Holmberg K, Keinanen M, Kepp K, Kere J, Kiviluoma P, Kristensen V, Lindgren C, Odeberg J, Osterman P, Parkkonen M, Saarela J and others. 2006. A quality assessment survey of SNP genotyping laboratories. Hum Mutat 27(7):711–4.
49. Lin Z, Cui X, Li H. 1996. Multiplex genotype determination at a large number of gene loci. Proc Natl Acad Sci U S A 93(6):2582–7.
50. Lindroos K, Sigurdsson S, Johansson K, Ronnblom L, Syvänen AC. 2002. Multiplex SNP genotyping in pooled DNA samples by a four-colour microarray system. Nucleic Acids Res 30(14):e70.
51. Lovgren S. 2003. Rewriting the Science on DNA Mass Identification. National Geographic News.
52. Lovmar L, Syvänen AC. 2005. Genotyping single-nucleotide polymorphisms by minisequencing using tag arrays. Methods Mol Med 114:79–92.
53. Luo M, Cui X, Fredman D, Brookes AJ, Azaro MA, Greenawalt DM, Hu G, Wang HY, Tereshchenko IV, Lin Y, Shentu Y, Gao R, Shen L, Li H. 2009. Genetic structures of copy number variants revealed by genotyping single sperm. PLoS One 4(4):e5236.
54. Lyamichev V, Mast AL, Hall JG, Prudent JR, Kaiser MW, Takova T, Kwiatkowski RW, Sander TJ, de Arruda M, Arco DA, Neri BP, Brow MA. 1999. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. Nat Biotechnol 17(3):292–6.
55. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC and others. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057):376–80.
56. Matsuzaki H, Loi H, Dong S, Tsai YY, Fang J, Law J, Di X, Liu WM, Yang G, Liu G, Huang J, Kennedy GC, Ryder TB, Marcus GA, Walsh PS, Shriver MD, Puck JM, Jones KW, Mei R. 2004. Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. Genome Res 14(3):414–25.
57. Meuzelaar LS, Lancaster O, Pasche JP, Kopal G, Brookes AJ. 2007. MegaPlex PCR: a strategy for multiplex amplification. Nat Methods 4(10):835–7.
58. Miyagi T, Murakami K, Sawada T, Taguchi H, Miyoshi I. 1998. A novel single cell PCR assay: detection of human T lymphotropic virus type I DNA in lymphocytes of patients with adult T cell leukemia. Leukemia 12(10):1645–50.
59. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. Nature.

60. Ohnishi Y, Tanaka T, Ozaki K, Yamada R, Suzuki H, Nakamura Y. 2001. A high-throughput SNP typing system for genome-wide association studies. J Hum Genet 46(8):471–7.
61. Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, Zwick ME. 2007. Microarray-based genomic selection for high-throughput resequencing. Nat Methods 4(11):907–9.
62. Oliphant A, Barker DL, Stuelpnagel JR, Chee MS. 2002. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. Biotechniques Suppl:56–8, 60–1.
63. Pastinen T, Kurg A, Metspalu A, Peltonen L, Syvanen AC. 1997. Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. Genome Res 7(6):606–14.
64. Pati N, Schowinsky V, Kokanovic O, Magnuson V, Ghosh S. 2004. A comparison between SNaPshot, pyrosequencing, and biplex invader SNP genotyping methods: accuracy, cost, and throughput. J Biochem Biophys Methods 60(1):1–12.
65. Pitterl F, Niederstatter H, Huber G, Zimmermann B, Oberacher H, Parson W. 2008. The next generation of DNA profiling--STR typing by multiplexed PCR--ion-pair RP LC-ESI time-of-flight MS. Electrophoresis 29(23):4739–50.
66. Podder M, Ruan J, Tripp BW, Chu ZE, Tebbutt SJ. 2008. Robust SNP genotyping by multiplex PCR and arrayed primer extension. BMC Med Genomics 1:5.
67. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239(4839):487–91.
68. Sanchez JJ, Borsting C, Balogh K, Berger B, Bogus M, Butler JM, Carracedo A, Court DS, Dixon LA, Filipovic B, Fondevila M, Gill P, Harrison CD, Hohoff C, Huel R, Ludes B, Parson W, Parsons TJ, Petkovski E, Phillips C and others. 2008. Forensic typing of autosomal SNPs with a 29 SNP-multiplex--results of a collaborative EDNAP exercise. Forensic Sci Int Genet 2(3):176–83.
69. Sanchez JJ, Borsting C, Hallenberg C, Buchard A, Hernandez A, Morling N. 2003. Multiplex PCR and minisequencing of SNPs--a model with 35 Y chromosome SNPs. Forensic Sci Int 137(1):74–84.
70. Sanchez JJ, Endicott P. 2006. Developing multiplexed SNP assays with special reference to degraded DNA templates. Nat Protoc 1(3):1370–8.
71. Sanchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N. 2006. A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27(9):1713–24.
72. Schrijver I, Oitmaa E, Metspalu A, Gardner P. 2005. Genotyping microarray for the detection of more than 200 CFTR mutations in ethnically diverse populations. J Mol Diagn 7(3):375–87.
73. Seipp MT, Durtschi JD, Voelkerding KV, Wittwer CT. 2009. Multiplex amplicon genotyping by high-resolution melting. J Biomol Tech 20(3):160–4.
74. Shuber AP, Grondin VJ, Klinger KW. 1995. A simplified procedure for developing multiplex PCRs. Genome Res 5(5):488–93.
75. Shumaker JM, Metspalu A, Caskey CT. 1996. Mutation detection by solid phase primer extension. Hum Mutat 7(4):346–54.
76. Sokolov BP. 1990. Primer extension technique for the detection of single nucleotide in genomic DNA. Nucleic Acids Res 18(12):3671.

77. Spits C, Le Caignec C, De Rycke M, Van Haute L, Van Steirteghem A, Liebaers I, Sermon K. 2006a. Optimization and evaluation of single-cell whole-genome multiple displacement amplification. Hum Mutat 27(5):496–503.
78. Spits C, Le Caignec C, De Rycke M, Van Haute L, Van Steirteghem A, Liebaers I, Sermon K. 2006b. Whole-genome multiple displacement amplification from single cells. Nat Protoc 1(4):1965–70.
79. Steemers FJ, Chang W, Lee G, Barker DL, Shen R, Gunderson KL. 2006. Whole-genome genotyping with the single-base extension assay. Nat Methods 3(1):31–3.
80. Stiller M, Knapp M, Stenzel U, Hofreiter M, Meyer M. 2009. Direct multiplex sequencing (DMPS)--a novel method for targeted high-throughput sequencing of ancient and highly degraded DNA. Genome Res 19(10):1843–8.
81. Sugimoto N, Nakano S, Yoneyama M, Honda K. 1996. Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. Nucleic Acids Res 24(22):4501–5.
82. Summerer D, Wu H, Haase B, Cheng Y, Schracke N, Stahler CF, Chee MS, Stahler PF, Beier M. 2009. Microarray-based multicycle-enrichment of genomic subsets for targeted next-generation sequencing. Genome Res 19(9):1616–21.
83. Suzuki S, Ono N, Furusawa C, Kashiwagi A, Yomo T. 2007. Experimental optimization of probe length to increase the sequence specificity of high-density oligonucleotide microarrays. BMC Genomics 8:373.
84. Syvänen AC. 1999. From gels to chips: “minisequencing” primer extension for analysis of point mutations and single nucleotide polymorphisms. Hum Mutat 13(1):1–10.
85. Syvänen AC. 2005. Toward genome-wide SNP genotyping. Nat Genet 37 Suppl:S5–10.
86. Syvänen AC, Aalto-Setala K, Harju L, Kontula K, Soderlund H. 1990. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. Genomics 8(4):684–92.
87. Syvänen AC, Ikonen E, Manninen T, Bengtstrom M, Soderlund H, Aula P, Peltonen L. 1992. Convenient and quantitative determination of the frequency of a mutant allele using solid-phase minisequencing: application to aspartylglucosaminuria in Finland. Genomics 12(3):590–5.
88. Zernant J, Külm M, Dharmaraj S, den Hollander AI, Perrault I, Preising MN, Lorenz B, Kaplan J, Cremers FP, Maumenee I, Koenekoop RK, Allikmets R. 2005. Genotyping microarray (disease chip) for Leber congenital amaurosis: detection of modifier alleles. Invest Ophthalmol Vis Sci 46(9):3052–9.
89. Tebbutt SJ, He JQ, Burkett KM, Ruan J, Opushnyev IV, Tripp BW, Zeznik JA, Abara CO, Nelson CC, Walley KR. 2004. Microarray genotyping resource to determine population stratification in genetic association studies of complex disease. Biotechniques 37(6):977–85.
90. Tebbutt SJ, Ruan J. 2008. Combining multiple PCR primer pairs for each amplicon can improve SNP genotyping accuracy by reducing allelic drop-out. Biotechniques 45(6):637–8, 640, 642 passim.
91. Telenius H, Carter NP, Bebb CE, Nordenskjold M, Ponder BA, Tunnacliffe A. 1992. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. Genomics 13(3):718–25.
92. Tost J, Gut IG. 2006. DNA analysis by mass spectrometry-past, present and future. J Mass Spectrom 41(8):981–95.

93. Tõnisson N, Zernant J, Kurg A, Pavel H, Slavin G, Roomere H, Meiel A, Hainaut P, Metspalu A. 2002. Evaluating the arrayed primer extension resequencing assay of TP53 tumor suppressor gene. Proc Natl Acad Sci U S A 99(8):5503–8.
94. Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D and others. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280(5366):1077–82.
95. Wang HY, Luo M, Tereshchenko IV, Frikker DM, Cui X, Li JY, Hu G, Chu Y, Azaro MA, Lin Y, Shen L, Yang Q, Kambouris ME, Gao R, Shih W, Li H. 2005. A genotyping system capable of simultaneously analyzing >1000 single nucleotide polymorphisms in a haploid genome. Genome Res 15(2):276–83.
96. Varley KE, Mitra RD. 2008. Nested Patch PCR enables highly multiplexed mutation discovery in candidate genes. Genome Res 18(11):1844–50.
97. Weissenborn SJ, Wieland U, Junk M, Pfister H. 2010. Quantification of beta-human papillomavirus DNA by real-time PCR. Nat Protoc 5(1):1–13.
98. Wenzel T, Elssner T, Fahr K, Bimmler J, Richter S, Thomas I, Kostrzewa M. 2003. Genosnip: SNP genotyping by MALDI-TOF MS using photocleavable oligonucleotides. Nucleosides Nucleotides Nucleic Acids 22(5–8):1579–81.
99. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y and others. 2008. The complete genome of an individual by massively parallel DNA sequencing. Nature 452(7189):872–6.
100. von Wurmb-Schwark N, Preusse-Prange A, Heinrich A, Simeoni E, Bosch T, Schwark T. 2009. A new multiplex-PCR comprising autosomal and y-specific STRs and mitochondrial DNA to analyze highly degraded material. Forensic Sci Int Genet 3(2):96–103.

# SUMMARY IN ESTONIAN

## Geenikiibil põhinev genotüpiseerimise meetod, APEX-2, kui uudne ja paindlik lahendus DNA variatsioonide määramiseks

DNA mutatsioonide ja polümorfismide määramise meetodid on muutunud laialdaselt kasutatavaks nii geneetika- kui ka molekulaardiagnostika laboratooriumites. Välja on töötatud mitmeid protokolle akadeemilise sektori poolt, mille kasutamine on suhteliselt lihtne ning ei nõua spetsiaalset ja kulukat aparatuuri. Näiteks restriktsiooni põhine DNA fragmentide pikkuste analüüs või alleelspetsiifiline PCR. Erinevad firmad pakuvad keerukamaid ja suurema läbilaskevõimega lahendusi, kus tarbijale avanevad laiemad võimalused, kuid suureneb ka kulutus analüüsitava lookuse kohta. Numbriliste kvaliteedinäitajate kõrval peetakse oluliseks veel paindlikkust, inimtöö hulka ja lihtsust projekti alustamise ja valideerimise etapis. Kokkuvõtlikult, on mitmeid toimivaid lahendusi, kuid kuna genotüpiseerimise või mutatsioonide määramise iseloom on mitmekesine siis ühte kindlat ja domineerivat meetodit ei ole siiani välja kujunenud.

Käesoleva doktoritöö kirjanduse ülevaade keskendub PCR reaktsiooni arengule ühe-reaktsiooni formaadist kuni keerukama multipleks PCRini, kus ainsas reaktsioonituubis toimuvad üheskoos kümned ja sajad paralleelsed reaktsioonid. Lisaks antakse ülevaade kaasaegsetest genotüpiseerimise meetoditest ja tuuakse välja sealsed molekulaar-tehnilised lahendused, plussid ja kitsaskohad.

Antud doktoritöö eesmärgiks oli luua paindlik, kasutajasõbralik ja hästi töötav genotüpiseerimise meetod kuni tuhande erineva lookuse samaaegseks analüüsiks. Lisaks leida reaktsiooni tingimused ja praimerite disaini kriteeriumid, mis tagaksid võimalikult madala analüüsitavate alleelide väljalangemise valimist ja kõrge tulemuste usaldusväärsuse. Lisaks uurida erinevaid DNA eraldusmeetodeid verest ja süljest, teostada DNA proovide võrdlevad analüüsid ja testida analüüsitavaid DNA proove APEX-2 protokolliga.

APEX-2 võimalikud rakendused saavad tulevikus olla molekulaardiagnostilised DNA testid, kordusuuringud tähendusrikaste polümorfismi alleelide kinnitamiseks ja genealoogilised testid nii isiklikuks kui ka perekondlikuks analüüsiks.

Doktoritöö põhineb neljal eel-retsenseeritud publikatsioonil ja ühel USA patendil. Artiklid kirjeldavad APEX-2 meetodi olulisi komponente: praimerite ülesehitus multipleks PCR jaoks, optimaalseid reaktsiooni- ja paljundustsükli tingimusi, vajamineva DNA kogust ja olukorda, kus alg-DNA hulk ei ole ühes reaktsioonis kõigile uuritavatele lookustele võrdne. Kaks artiklit kirjeldavad APEX-2 rakendust erinevas genoomika uuringus. Patent (Ref. I) toob välja APEX-2 reaktsiooni põhimõtte, erisused ja uudsuse võrreldes olemasolevate meetoditega. Ref. II kirjeldab detailselt reaktsioonitingimusi ja praimerite disainimise põhimõtteid ning tõestab katseliselt 640-plex PCR töötamist ja signaalide lugemist geenikiibil. Ref. III on ühest küljest näide APEX-2 raken-

dusest, kus genoomse DNA uurimisega on läbi viidud populatsioonide võrdlus, kuid teisalt annab antud artikkel edasi APEX-2 arengut. Selles töös kasutasime 104-plex süsteeme ja ühe DNA analüüsimiseks viisime läbi kuus paralleelset reaktsiooni. Produktide puhastus- ja kontsentreerimise etapp põhines DNA väljasoolamisel ning enne ühenukleotiidilist detekteerimise reaktsiooni geenikiibil teostati spetsiifiline hübridisatsioon ja pesuetapid. Ref. IV 124-plex APEX-2 demonstreeris üheaegset genotüpiseerimist edukalt mitmelt erinevalt DNA allikalt (mitokondriaalne-, autosomaalne ja Y kromosoom) ning iseloomustab olukorda, kus uuritavate DNA proovide kontsentratsioon reaktsioonisegus erineb oluliselt, peegeldades nii reaalset olukorda. Ref. V võrdleb erinevaid DNA eraldusmeetodeid nii verest kui ka süljest, et leida sobivaim APEX-2 süsteemile. Nii molekulaardiagnostikas kui ka (tulevikus) personaalses meditsiinis on oluline kasutada võimalikult edukalt DNA allikat, mille hankimine ei põhjusta valu, vähe aeganõudev ning ei eelda meditsiinitöötaja osavõttu. Lahenduseks on näiteks DNA eraldus süljest.

# ACKNOWLEDGMENTS

First, I would like to thank my supervisor Prof Andres Metspalu who has guided me through many years, since 2003, and who was supportive or sobering if new ideas were raised. He helped me to stay in focus, gave valuable guidelines and encouraged to concentrate on what is important.

I would like to thank all the co-authors of the papers on which this dissertation is based on. Especially Dr Andrey Khrunin who's teamwork was very productive in 2006, Dr Toomas Kivisild who supported me a lot in my forensic assay study, Dr Reidar Andreson for vast bioinformatics support and Ms Triin Viltrop who has helped me in all technical aspects during the last years. Also, I would like to thank Dr Neeme Tõnisson and Miss Chelsea Anne Stern for the careful reading of the thesis.

I wish to thank all my present and former colleagues in the Department of Biotechnology, Department of Bioinformatics, Asper Biotech and BioAtlas for all the support and advice.

I owe my huge gratitude to my great parents and to my wonderful grandmother for all kind of support during my studies in the university from the beginning to the current point and I'm very pleased that they have believed in my venture all the time.

Finally, I would like to thank my lovely Helena who stands with two legs on the Earth and has helped me to stay on the same ground. She and our newborn daughter Liisa gave me the best reason and a lot of energy to compose my doctoral thesis in a timely manner and move forward.

# PUBLICATIONS

# CURRICULUM VITAE

## Kaarel Krjutškov


| | |
|---|---|
| Date and place of birth: | 14th March 1981, Kose, Estonia |
| Address: | Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu<br>Riia str 23, 51010, Tartu, Estonia |
| Phone: | +372 51 26 416 |
| E-mail: | karlakrj@ut.ee |


### Education and professional employment

| | |
|---|---|
| 1996–2000 | Tartu Tamme Gymnasium (Biology class) |
| 2000–2004 | *BSc* in biotechnology and biomedicine, Chair of Biotechnology, IMCB, University of Tartu |
| 2004–2006 | *MSc* in molecular engineering, Chair of Biotechnology, IMCB, University of Tartu |
| 2006– | *PhD* student in molecular engineering, Chair of Biotechnology, IMCB, University of Tartu |
| 2005–2006 | Specialist of molecular engineering, Estonian Biocentre |
| 2006–2010 | Scientist, Estonian Biocentre |
| 2007–2009 | APEX-2 specialist, Asper Biotech |
| 2009 | Visiting scientist at Beijing Genomics Institute, Shenzhen Sequencing Centre, China |
| 2010– | Scientist, Estonian Genome Centre University of Tartu |

### Teaching

| | |
|---|---|
| 2005– | Practical training course in molecular biotechnology (LOMR.01.003). Microarray-based genotyping using APEX-2 technology. |
| 2005– | Two weeks practical training course in molecular biotechnology for high-school students |

### Scientific work

My study project is focused on research and development to increase multiplex PCR level from ten to hundreds level. After when the APEX-2 principle was published in 2008, I have focused on different applications contributing as a technical manager of these projects.

# LIST OF PUBLICATIONS

1. US patent US 2009/0048119 A1 „*A method to determine single nucleotide polymorphisms and mutations in nucleic acid sequence*". December 29th 2009 and the applicant is Estonian Biocentre.
   Inventors: **Kaarel Krjutškov** and Andres Metspalu
2. **Krjutškov K**, Andreson R, Mägi R, Nikopensius T, Khrunin A, Mihailov E, Tammekivi V, Sork H, Remm M, Metspalu A (2008). *Development of a single tube 640-plex genotyping method for detection of nucleic acid variations on microarrays.* Nucleic Acids Research 36(12) e75, doi:10.1093/nar/gkn357
3. Khrunin A, Mihailov E, Nikopensius T, **Krjutškov K**, Limborska S, Metspalu A (2009). *Analysis of Allele and Haplotype Diversity Across 25 Genomic Regions in Three Eastern European Populations*. Human Heredity 68 35–44, doi:10.1159/000210447
4. **Krjutškov K**, Viltrop T, Palta P, Metspalu E, Tamm E, Suvi S, Sak K, Merilo A, Sork H, Teek R, Nikopensius T, Kivisild T, Metspalu A (2009). *Evaluation of the 124-plex SNP typing microarray for forensic testing*. Forensic Science International: GENETICS 4 43–48, doi:10.1016/ j.fsigen.2009.04.007
5. Viltrop T, **Krjutškov K**, Palta P, Metspalu A (2009). *Comparison of DNA extraction methods for multiplex PCR*. Analytical Biochemistry 398 260–262, doi: 10.1016/j.ab.2009.11.026
6. Theodoraki E.V, Nikopensius T, Suhorutsenko J, Peppes V, Fili P, Kolovou G, Papamikos V, Dimitrios Richter D, Zakopoulos N**, Krjutškov K,** Metspalu A, Dedoussis V.G (2010). *Fibrinogen beta variants confer protection against coronary artery disease in a Greek case-control study.* BMC Medical Genetics 11:28, doi:10.1186/1471–2350–11–28
7. Tõnisson N, Oitmaa E, **Krjutškov K**, Pullat J, Lind I, Leego M, Kurg A, Metspalu A (2009). Molecular Diagnostics 2nd Revised edition, Chapther 17: *Arrayed Primer Extension Microarrays for Molecular Diagnostics*, Academic Press Inc., ISBN-10: 0123745373, ISBN-13: 9780123745378

# ELULOOKIRJELDUS

## Kaarel Krjutškov

Sünniaeg ja koht: 14. märts 1981, Kose, Eesti
Aadress: Biotehnoloogia õppetool, Molekulaar- ja Rakubioloogia instituut, Tartu Ülikool
Riia 23, 51010, Tartu, Eesti
Telefon: +372 51 26 416
E-mail: karlakrj@ut.ee

### Haridus ja erialane teenistuskäik

1996–2000 Tartu Tamme Gümnaasium (Bioloogia erikallak)
2000–2004 *BSc* biotehnoloogia ja biomeditsiini erialal, Biotehnoloogia õppetool, TÜ MRI, Tartu Ülikool
2004–2006 *MSc* geenitehnoloogia erialal, Biotehnoloogia õppetool, TÜ MRI, Tartu Ülikool
2006– Doktorant geenitehnoloogia erialal, Biotehnoloogia õppetool, TÜMRI, Tartu Ülikool
2005–2006 Geenitehnoloogia spetsialist, Eesti Biokeskus
2006–2010 Teadur, Eesti Biokeskus
2007–2009 APEX-2 spetsialist, Asper Biotech
2009 Külalisteadlane Pekingi Genoomi Instituudis, Shenzheni Sekveneerimise Keskus, Hiina Rahvavabariik
2010– Teadur, Eesti Geenivaramu Tartu Ülikool

### Juhendamine

2005– Praktiline kursus molekulaarses biotehnoloogias (LOMR.01.003). Genotüpiseerimine geenikiibiga.
2005– Kahenädalane praktiline kursus molekulaarses biotehnoloogias gümnaasiumi õpilastele, juhendaja

### Teadustegevus

Teadustöö on suunatud arendustegevusele, mille käigus töötati välja multipleks PCR protokoll, mis võimaldab seniselt 10 multipleks tasemest liikuda 640-pleks tasemele. Peale APEX-2 idee avaldamist 2008 aastal olen keskendunud erinevatele APEX-2 rakendustele ning olnud vastutav projektide tehnilise poole eest.

# PUBLIKATSIOONID

1. US patent US 2009/0048119 A1 „*A method to determine single nucleotide polymorphisms and mutations in nucleic acid sequence*". 29. detsember 2009. Taotleja Eesti Biokeskus.
   Leiutajad: **Kaarel Krjutškov** ja Andres Metspalu
2. **Krjutškov K**, Andreson R, Mägi R, Nikopensius T, Khrunin A, Mihailov E, Tammekivi V, Sork H, Remm M, Metspalu A (2008). *Development of a single tube 640-plex genotyping method for detection of nucleic acid variations on microarrays.* Nucleic Acids Research 36(12) e75, doi:10.1093/nar/gkn357
3. Khrunin A, Mihailov E, Nikopensius T, **Krjutškov K**, Limborska S, Metspalu A (2009). *Analysis of Allele and Haplotype Diversity Across 25 Genomic Regions in Three Eastern European Populations*. Human Heredity 68 35–44, doi:10.1159/000210447
4. **Krjutškov K**, Viltrop T, Palta P, Metspalu E, Tamm E, Suvi S, Sak K, Merilo A, Sork H, Teek R, Nikopensius T, Kivisild T, Metspalu A (2009). *Evaluation of the 124-plex SNP typing microarray for forensic testing*. Forensic Science International: GENETICS 4 43–48, doi:10.1016/ j.fsigen. 2009.04.007
5. Viltrop T, **Krjutškov K**, Palta P, Metspalu A (2009). *Comparison of DNA extraction methods for multiplex PCR*. Analytical Biochemistry 398 260–262, doi: 10.1016/j.ab.2009.11.026
6. Theodoraki E.V, Nikopensius T, Suhorutsenko J, Peppes V, Fili P, Kolovou G, Papamikos V, Dimitrios Richter D, Zakopoulos N**, Krjutškov K,** Metspalu A**,** Dedoussis V.G (2010). *Fibrinogen beta variants confer protection against coronary artery disease in a Greek case-control study.* BMC Medical Genetics 11:28, doi:10.1186/1471–2350–11–28
7. Tõnisson N, Oitmaa E, **Krjutškov K**, Pullat J, Lind I, Leego M, Kurg A, Metspalu A (2009). Molecular Diagnostics 2$^{nd}$ Revised edition, Chapther 17: *Arrayed Primer Extension Microarrays for Molecular Diagnostics*, Academic Press Inc., ISBN-10: 0123745373, ISBN-13: 9780123745378

## DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

1. **Toivo Maimets**. Studies of human oncoprotein p53. Tartu, 1991, 96 p.
2. **Enn K. Seppet**. Thyroid state control over energy metabolism, ion transport and contractile functions in rat heart. Tartu, 1991, 135 p.
3. **Kristjan Zobel**. Epifüütsete makrosamblike väärtus õhu saastuse indikaatoritena Hamar-Dobani boreaalsetes mägimetsades. Tartu, 1992, 131 lk.
4. **Andres Mäe**. Conjugal mobilization of catabolic plasmids by transposable elements in helper plasmids. Tartu, 1992, 91 p.
5. **Maia Kivisaar**. Studies on phenol degradation genes of *Pseudomonas* sp. strain EST 1001. Tartu, 1992, 61 p.
6. **Allan Nurk**. Nucleotide sequences of phenol degradative genes from *Pseudomonas sp.* strain EST 1001 and their transcriptional activation in *Pseudomonas putida.* Tartu, 1992, 72 p.
7. **Ülo Tamm**. The genus *Populus* L. in Estonia: variation of the species biology and introduction. Tartu, 1993, 91 p.
8. **Jaanus Remme**. Studies on the peptidyltransferase centre of the *E.coli* ribosome. Tartu, 1993, 68 p.
9. **Ülo Langel**. Galanin and galanin antagonists. Tartu, 1993, 97 p.
10. **Arvo Käärd**. The development of an automatic online dynamic fluorescense-based pH-dependent fiber optic penicillin flowthrought biosensor for the control of the benzylpenicillin hydrolysis. Tartu, 1993, 117 p.
11. **Lilian Järvekülg**. Antigenic analysis and development of sensitive immunoassay for potato viruses. Tartu, 1993, 147 p.
12. **Jaak Palumets**. Analysis of phytomass partition in Norway spruce. Tartu, 1993, 47 p.
13. **Arne Sellin**. Variation in hydraulic architecture of *Picea abies* (L.) Karst. trees grown under different enviromental conditions. Tartu, 1994, 119 p.
13. **Mati Reeben**. Regulation of light neurofilament gene expression. Tartu, 1994, 108 p.
14. **Urmas Tartes**. Respiration rhytms in insects. Tartu, 1995, 109 p.
15. **Ülo Puurand.** The complete nucleotide sequence and infections *in vitro* transcripts from cloned cDNA of a potato A potyvirus. Tartu, 1995, 96 p.
16. **Peeter Hõrak**. Pathways of selection in avian reproduction: a functional framework and its application in the population study of the great tit (*Parus major*). Tartu, 1995, 118 p.
17. **Erkki Truve**. Studies on specific and broad spectrum virus resistance in transgenic plants. Tartu, 1996, 158 p.
18. **Illar Pata**. Cloning and characterization of human and mouse ribosomal protein S6-encoding genes. Tartu, 1996, 60 p.
19. **Ülo Niinemets**. Importance of structural features of leaves and canopy in determining species shade-tolerance in temperature deciduous woody taxa. Tartu, 1996, 150 p.

20. **Ants Kurg**. Bovine leukemia virus: molecular studies on the packaging region and DNA diagnostics in cattle. Tartu, 1996, 104 p.
21. **Ene Ustav**. E2 as the modulator of the BPV1 DNA replication. Tartu, 1996, 100 p.
22. **Aksel Soosaar**. Role of helix-loop-helix and nuclear hormone receptor transcription factors in neurogenesis. Tartu, 1996, 109 p.
23. **Maido Remm**. Human papillomavirus type 18: replication, transformation and gene expression. Tartu, 1997, 117 p.
24. **Tiiu Kull**. Population dynamics in *Cypripedium calceolus* L. Tartu, 1997, 124 p.
25. **Kalle Olli**. Evolutionary life-strategies of autotrophic planktonic microorganisms in the Baltic Sea. Tartu, 1997, 180 p.
26. **Meelis Pärtel**. Species diversity and community dynamics in calcareous grassland communities in Western Estonia. Tartu, 1997, 124 p.
27. **Malle Leht**. The Genus *Potentilla* L. in Estonia, Latvia and Lithuania: distribution, morphology and taxonomy. Tartu, 1997, 186 p.
28. **Tanel Tenson**. Ribosomes, peptides and antibiotic resistance. Tartu, 1997, 80 p.
29. **Arvo Tuvikene**. Assessment of inland water pollution using biomarker responses in fish *in vivo* and *in vitro.* Tartu, 1997, 160 p.
30. **Urmas Saarma**. Tuning ribosomal elongation cycle by mutagenesis of 23S rRNA. Tartu, 1997, 134 p.
31. **Henn Ojaveer**. Composition and dynamics of fish stocks in the gulf of Riga ecosystem. Tartu, 1997, 138 p.
32. **Lembi Lõugas**. Post-glacial development of vertebrate fauna in Estonian water bodies. Tartu, 1997, 138 p.
33. **Margus Pooga**. Cell penetrating peptide, transportan, and its predecessors, galanin-based chimeric peptides. Tartu, 1998, 110 p.
34. **Andres Saag**. Evolutionary relationships in some cetrarioid genera (Lichenized Ascomycota). Tartu, 1998, 196 p.
35. **Aivar Liiv**. Ribosomal large subunit assembly *in vivo*. Tartu, 1998, 158 p.
36. **Tatjana Oja**. Isoenzyme diversity and phylogenetic affinities among the eurasian annual bromes (*Bromus* L., Poaceae). Tartu, 1998, 92 p.
37. **Mari Moora**. The influence of arbuscular mycorrhizal (AM) symbiosis on the competition and coexistence of calcareous crassland plant species. Tartu, 1998, 78 p.
38. **Olavi Kurina**. Fungus gnats in Estonia (*Diptera: Bolitophilidae, Keroplatidae, Macroceridae, Ditomyiidae, Diadocidiidae, Mycetophilidae*). Tartu, 1998, 200 p.
39. **Andrus Tasa**. Biological leaching of shales: black shale and oil shale. Tartu, 1998, 98 p.
40. **Arnold Kristjuhan.** Studies on transcriptional activator properties of tumor suppressor protein p53. Tartu, 1998, 86 p.

41. **Sulev Ingerpuu.** Characterization of some human myeloid cell surface and nuclear differentiation antigens. Tartu, 1998, 163 p.
42. **Veljo Kisand.** Responses of planktonic bacteria to the abiotic and biotic factors in the shallow lake Võrtsjärv. Tartu, 1998, 118 p.
43. **Kadri Põldmaa.** Studies in the systematics of hypomyces and allied genera (Hypocreales, Ascomycota). Tartu, 1998, 178 p.
44. **Markus Vetemaa.** Reproduction parameters of fish as indicators in environmental monitoring. Tartu, 1998, 117 p.
45. **Heli Talvik.** Prepatent periods and species composition of different *Oesophagostomum* spp. populations in Estonia and Denmark. Tartu, 1998, 104 p.
46. **Katrin Heinsoo.** Cuticular and stomatal antechamber conductance to water vapour diffusion in *Picea abies* (L.) karst. Tartu, 1999, 133 p.
47. **Tarmo Annilo.** Studies on mammalian ribosomal protein S7. Tartu, 1998, 77 p.
48. **Indrek Ots.** Health state indicies of reproducing great tits (*Parus major*): sources of variation and connections with life-history traits. Tartu, 1999, 117 p.
49. **Juan Jose Cantero.** Plant community diversity and habitat relationships in central Argentina grasslands. Tartu, 1999, 161 p.
50. **Rein Kalamees.** Seed bank, seed rain and community regeneration in Estonian calcareous grasslands. Tartu, 1999, 107 p.
51. **Sulev Kõks.** Cholecystokinin (CCK) — induced anxiety in rats: influence of environmental stimuli and involvement of endopioid mechanisms and erotonin. Tartu, 1999, 123 p.
52. **Ebe Sild.** Impact of increasing concentrations of $O_3$ and $CO_2$ on wheat, clover and pasture. Tartu, 1999, 123 p.
53. **Ljudmilla Timofejeva.** Electron microscopical analysis of the synaptonemal complex formation in cereals. Tartu, 1999, 99 p.
54. **Andres Valkna.** Interactions of galanin receptor with ligands and G-proteins: studies with synthetic peptides. Tartu, 1999, 103 p.
55. **Taavi Virro.** Life cycles of planktonic rotifers in lake Peipsi. Tartu, 1999, 101 p.
56. **Ana Rebane.** Mammalian ribosomal protein S3a genes and intron-encoded small nucleolar RNAs U73 and U82. Tartu, 1999, 85 p.
57. **Tiina Tamm.** Cocksfoot mottle virus: the genome organisation and translational strategies. Tartu, 2000, 101 p.
58. **Reet Kurg.** Structure-function relationship of the bovine papilloma virus E2 protein. Tartu, 2000, 89 p.
59. **Toomas Kivisild.** The origins of Southern and Western Eurasian populations: an mtDNA study. Tartu, 2000, 121 p.
60. **Niilo Kaldalu.** Studies of the TOL plasmid transcription factor XylS. Tartu 2000. 88 p.

61. **Dina Lepik.** Modulation of viral DNA replication by tumor suppressor protein p53. Tartu 2000. 106 p.
62. **Kai Vellak.** Influence of different factors on the diversity of the bryophyte vegetation in forest and wooded meadow communities. Tartu 2000. 122 p.
63. **Jonne Kotta.** Impact of eutrophication and biological invasionas on the structure and functions of benthic macrofauna. Tartu 2000. 160 p.
64. **Georg Martin.** Phytobenthic communities of the Gulf of Riga and the inner sea the West-Estonian archipelago. Tartu, 2000. 139 p.
65. **Silvia Sepp.** Morphological and genetical variation of *Alchemilla L.* in Estonia. Tartu, 2000. 124 p.
66. **Jaan Liira.** On the determinants of structure and diversity in herbaceous plant communities. Tartu, 2000. 96 p.
67. **Priit Zingel.** The role of planktonic ciliates in lake ecosystems. Tartu 2001. 111 p.
68. **Tiit Teder.** Direct and indirect effects in Host-parasitoid interactions: ecological and evolutionary consequences. Tartu 2001. 122 p.
69. **Hannes Kollist.** Leaf apoplastic ascorbate as ozone scavenger and its transport across the plasma membrane. Tartu 2001. 80 p.
70. **Reet Marits.** Role of two-component regulator system PehR-PehS and extracellular protease PrtW in virulence of *Erwinia Carotovora* subsp. *Carotovora*. Tartu 2001. 112 p.
71. **Vallo Tilgar.** Effect of calcium supplementation on reproductive performance of the pied flycatcher *Ficedula hypoleuca* and the great tit *Parus major,* breeding in Nothern temperate forests. Tartu, 2002. 126 p.
72. **Rita Hõrak.** Regulation of transposition of transposon Tn*4652* in *Pseudomonas putida*. Tartu, 2002. 108 p.
73. **Liina Eek-Piirsoo.** The effect of fertilization, mowing and additional illumination on the structure of a species-rich grassland community. Tartu, 2002. 74 p.
74. **Krõõt Aasamaa.** Shoot hydraulic conductance and stomatal conductance of six temperate deciduous tree species. Tartu, 2002. 110 p.
75. **Nele Ingerpuu.** Bryophyte diversity and vascular plants. Tartu, 2002. 112 p.
76. **Neeme Tõnisson.** Mutation detection by primer extension on oligonucleotide microarrays. Tartu, 2002. 124 p.
77. **Margus Pensa.** Variation in needle retention of Scots pine in relation to leaf morphology, nitrogen conservation and tree age. Tartu, 2003. 110 p.
78. **Asko Lõhmus.** Habitat preferences and quality for birds of prey: from principles to applications. Tartu, 2003. 168 p.
79. **Viljar Jaks.** p53 — a switch in cellular circuit. Tartu, 2003. 160 p.
80. **Jaana Männik.** Characterization and genetic studies of four ATP-binding cassette (ABC) transporters. Tartu, 2003. 140 p.
81. **Marek Sammul.** Competition and coexistence of clonal plants in relation to productivity. Tartu, 2003. 159 p

82. **Ivar Ilves.** Virus-cell interactions in the replication cycle of bovine papillomavirus type 1. Tartu, 2003. 89 p.
83. **Andres Männik.** Design and characterization of a novel vector system based on the stable replicator of bovine papillomavirus type 1. Tartu, 2003. 109 p.
84. **Ivika Ostonen.** Fine root structure, dynamics and proportion in net primary production of Norway spruce forest ecosystem in relation to site conditions. Tartu, 2003. 158 p.
85. **Gudrun Veldre.** Somatic status of 12–15-year-old Tartu schoolchildren. Tartu, 2003. 199 p.
86. **Ülo Väli.** The greater spotted eagle *Aquila clanga* and the lesser spotted eagle *A. pomarina*: taxonomy, phylogeography and ecology. Tartu, 2004. 159 p.
87. **Aare Abroi.** The determinants for the native activities of the bovine papillomavirus type 1 E2 protein are separable. Tartu, 2004. 135 p.
88. **Tiina Kahre.** Cystic fibrosis in Estonia. Tartu, 2004. 116 p.
89. **Helen Orav-Kotta.** Habitat choice and feeding activity of benthic suspension feeders and mesograzers in the northern Baltic Sea. Tartu, 2004. 117 p.
90. **Maarja Öpik.** Diversity of arbuscular mycorrhizal fungi in the roots of perennial plants and their effect on plant performance. Tartu, 2004. 175 p.
91. **Kadri Tali.** Species structure of *Neotinea ustulata*. Tartu, 2004. 109 p.
92. **Kristiina Tambets.** Towards the understanding of post-glacial spread of human mitochondrial DNA haplogroups in Europe and beyond: a phylogeographic approach. Tartu, 2004. 163 p.
93. **Arvi Jõers.** Regulation of p53-dependent transcription. Tartu, 2004. 103 p.
94. **Lilian Kadaja.** Studies on modulation of the activity of tumor suppressor protein p53. Tartu, 2004. 103 p.
95. **Jaak Truu.** Oil shale industry wastewater: impact on river microbial community and possibilities for bioremediation. Tartu, 2004. 128 p.
96. **Maire Peters.** Natural horizontal transfer of the *pheBA* operon. Tartu, 2004. 105 p.
97. **Ülo Maiväli.** Studies on the structure-function relationship of the bacterial ribosome. Tartu, 2004. 130 p.
98. **Merit Otsus.** Plant community regeneration and species diversity in dry calcareous grasslands. Tartu, 2004. 103 p.
99. **Mikk Heidemaa.** Systematic studies on sawflies of the genera *Dolerus, Empria,* and *Caliroa* (Hymenoptera: Tenthredinidae). Tartu, 2004. 167 p.
100. **Ilmar Tõnno.** The impact of nitrogen and phosphorus concentration and N/P ratio on cyanobacterial dominance and $N_2$ fixation in some Estonian lakes. Tartu, 2004. 111 p.
101. **Lauri Saks.** Immune function, parasites, and carotenoid-based ornaments in greenfinches. Tartu, 2004. 144 p.

102. **Siiri Rootsi.** Human Y-chromosomal variation in European populations. Tartu, 2004. 142 p.
103. **Eve Vedler.** Structure of the 2,4-dichloro-phenoxyacetic acid-degradative plasmid pEST4011. Tartu, 2005. 106 p.
104. **Andres Tover.** Regulation of transcription of the phenol degradation *pheBA* operon in *Pseudomonas putida*. Tartu, 2005. 126 p.
105. **Helen Udras.** Hexose kinases and glucose transport in the yeast *Hansenula polymorpha*. Tartu, 2005. 100 p.
106. **Ave Suija.** Lichens and lichenicolous fungi in Estonia: diversity, distribution patterns, taxonomy. Tartu, 2005. 162 p.
107. **Piret Lõhmus.** Forest lichens and their substrata in Estonia. Tartu, 2005. 162 p.
108. **Inga Lips.** Abiotic factors controlling the cyanobacterial bloom occurrence in the Gulf of Finland. Tartu, 2005. 156 p.
109. **Kaasik, Krista.** Circadian clock genes in mammalian clockwork, metabolism and behaviour. Tartu, 2005. 121 p.
110. **Juhan Javoiš.** The effects of experience on host acceptance in ovipositing moths. Tartu, 2005. 112 p.
111. **Tiina Sedman.** Characterization of the yeast *Saccharomyces cerevisiae* mitochondrial DNA helicase Hmi1. Tartu, 2005. 103 p.
112. **Ruth Aguraiuja.** Hawaiian endemic fern lineage *Diellia* (Aspleniaceae): distribution, population structure and ecology. Tartu, 2005. 112 p.
113. **Riho Teras.** Regulation of transcription from the fusion promoters generated by transposition of Tn*4652* into the upstream region of *pheBA* operon in *Pseudomonas putida*. Tartu, 2005. 106 p.
114. **Mait Metspalu.** Through the course of prehistory in india: tracing the mtDNA trail. Tartu, 2005. 138 p.
115. **Elin Lõhmussaar.** The comparative patterns of linkage disequilibrium in European populations and its implication for genetic association studies. Tartu, 2006. 124 p.
116. **Priit Kupper.** Hydraulic and environmental limitations to leaf water relations in trees with respect to canopy position. Tartu, 2006. 126 p.
117. **Heili Ilves.** Stress-induced transposition of Tn*4652* in *Pseudomonas Putida.* Tartu, 2006. 120 p.
118. **Silja Kuusk.** Biochemical properties of Hmi1p, a DNA helicase from *Saccharomyces cerevisiae* mitochondria. Tartu, 2006. 126 p.
119. **Kersti Püssa.** Forest edges on medium resolution landsat thematic mapper satellite images. Tartu, 2006. 90 p.
120. **Lea Tummeleht.** Physiological condition and immune function in great tits (*Parus major* l.): Sources of variation and trade-offs in relation to growth. Tartu, 2006. 94 p.
121. **Toomas Esperk.** Larval instar as a key element of insect growth schedules. Tartu, 2006. 186 p.

122. **Harri Valdmann.** Lynx (*Lynx lynx*) and wolf (*Canis lupus*) in the Baltic region: Diets, helminth parasites and genetic variation. Tartu, 2006. 102 p.
123. **Priit Jõers.** Studies of the mitochondrial helicase Hmi1p in *Candida albicans* and *Saccharomyces cerevisia*. Tartu, 2006. 113 p.
124. **Kersti Lilleväli.** Gata3 and Gata2 in inner ear development. Tartu, 2007. 123 p.
125. **Kai Rünk.** Comparative ecology of three fern species: *Dryopteris carthusiana* (Vill.) H.P. Fuchs, *D. expansa* (C. Presl) Fraser-Jenkins & Jermy and *D. dilatata* (Hoffm.) A. Gray (Dryopteridaceae). Tartu, 2007. 143 p.
126. **Aveliina Helm.** Formation and persistence of dry grassland diversity: role of human history and landscape structure. Tartu, 2007. 89 p.
127. **Leho Tedersoo.** Ectomycorrhizal fungi: diversity and community structure in Estonia, Seychelles and Australia. Tartu, 2007. 233 p.
128. **Marko Mägi.** The habitat-related variation of reproductive performance of great tits in a deciduous-coniferous forest mosaic: looking for causes and consequences. Tartu, 2007. 135 p.
129. **Valeria Lulla.** Replication strategies and applications of Semliki Forest virus. Tartu, 2007. 109 p.
130. **Ülle Reier**. Estonian threatened vascular plant species: causes of rarity and conservation. Tartu, 2007. 79 p.
131. **Inga Jüriado**. Diversity of lichen species in Estonia: influence of regional and local factors. Tartu, 2007. 171 p.
132. **Tatjana Krama.** Mobbing behaviour in birds: costs and reciprocity based cooperation. Tartu, 2007.
133. **Signe Saumaa.** The role of DNA mismatch repair and oxidative DNA damage defense systems in avoidance of stationary phase mutations in *Pseudomonas putida.* Tartu, 2007. 172 p.
134. **Reedik Mägi**. The linkage disequilibrium and the selection of genetic markers for association studies in european populations. Tartu, 2007. 96 p.
135. **Priit Kilgas.** Blood parameters as indicators of physiological condition and skeletal development in great tits (*Parus major*): natural variation and application in the reproductive ecology of birds. Tartu, 2007. 129 p.
136. **Anu Albert**. The role of water salinity in structuring eastern Baltic coastal fish communities. Tartu, 2007. 95 p.
137. **Kärt Padari.** Protein transduction mechanisms of transportans. Tartu, 2008. 128 p.
138. **Siiri-Lii Sandre.** Selective forces on larval colouration in a moth. Tartu, 2008. 125 p.
139. **Ülle Jõgar.** Conservation and restoration of semi-natural floodplain meadows and their rare plant species. Tartu, 2008. 99 p.
140. **Lauri Laanisto.** Macroecological approach in vegetation science: generality of ecological relationships at the global scale. Tartu, 2008. 133 p.
141. **Reidar Andreson**. Methods and software for predicting PCR failure rate in large genomes. Tartu, 2008. 105 p.

142. **Birgot Paavel.** Bio-optical properties of turbid lakes. Tartu, 2008. 175 p.
143. **Kaire Torn.** Distribution and ecology of charophytes in the Baltic Sea. Tartu, 2008, 98 p.
144. **Vladimir Vimberg.** Peptide mediated macrolide resistance. Tartu, 2008, 190 p.
145. **Daima Örd.** Studies on the stress-inducible pseudokinase TRB3, a novel inhibitor of transcription factor ATF4. Tartu, 2008, 108 p.
146. **Lauri Saag.** Taxonomic and ecologic problems in the genus *Lepraria* (*Stereocaulaceae*, lichenised *Ascomycota*). Tartu, 2008, 175 p.
147. **Ulvi Karu.** Antioxidant protection, carotenoids and coccidians in greenfinches – assessment of the costs of immune activation and mechanisms of parasite resistance in a passerine with carotenoid-based ornaments. Tartu, 2008, 124 p.
148. **Jaanus Remm.** Tree-cavities in forests: density, characteristics and occupancy by animals. Tartu, 2008, 128 p.
149. **Epp Moks.** Tapeworm parasites *Echinococcus multilocularis* and *E. granulosus* in Estonia: phylogenetic relationships and occurrence in wild carnivores and ungulates. Tartu, 2008, 82 p.
150. **Eve Eensalu.** Acclimation of stomatal structure and function in tree canopy: effect of light and $CO_2$ concentration. Tartu, 2008, 108 p.
151. **Janne Pullat**. Design, functionlization and application of an *in situ* synthesized oligonucleotide microarray. Tartu, 2008, 108 p.
152. **Marta Putrinš.** Responses of *Pseudomonas putida* to phenol-induced metabolic and stress signals. Tartu, 2008, 142 p.
153. **Marina Semtšenko.** Plant root behaviour: responses to neighbours and physical obstructions. Tartu, 2008, 106 p.
154. **Marge Starast.** Influence of cultivation techniques on productivity and fruit quality of some *Vaccinium* and *Rubus* taxa. Tartu, 2008, 154 p.
155. **Age Tats.** Sequence motifs influencing the efficiency of translation. Tartu, 2009, 104 p.
156. **Radi Tegova.** The role of specialized DNA polymerases in mutagenesis in *Pseudomonas putida.* Tartu, 2009, 124 p.
157. **Tsipe Aavik.** Plant species richness, composition and functional trait pattern in agricultural landscapes – the role of land use intensity and landscape structure. Tartu, 2008, 112 p.
158. **Kaja Kiiver.** Semliki forest virus based vectors and cell lines for studying the replication and interactions of alphaviruses and hepaciviruses. Tartu, 2009, 104 p.
159. **Meelis Kadaja.** Papillomavirus Replication Machinery Induces Genomic Instability in its Host Cell. Tartu, 2009, 126 p.
160. **Pille Hallast.** Human and chimpanzee Luteinizing hormone/Chorionic Gonadotropin beta (*LHB/CGB*) gene clusters: diversity and divergence of young duplicated genes. Tartu, 2009, 168 p.

161. **Ain Vellak.** Spatial and temporal aspects of plant species conservation. Tartu, 2009, 86 p.
162. **Triinu Remmel.** Body size evolution in insects with different colouration strategies: the role of predation risk. Tartu, 2009, 168 p.
163. **Jaana Salujõe.** Zooplankton as the indicator of ecological quality and fish predation in lake ecosystems. Tartu, 2009, 129 p.
164. **Ele Vahtmäe.** Mapping benthic habitat with remote sensing in optically complex coastal environments. Tartu, 2009, 109 p.
165. **Liisa Metsamaa.** Model-based assessment to improve the use of remote sensing in recognition and quantitative mapping of cyanobacteria. Tartu, 2009, 114 p.
166. **Pille Säälik.** The role of endocytosis in the protein transduction by cell-penetrating peptides. Tartu, 2009, 155 p.
167. **Lauri Peil.** Ribosome assembly factors in *Escherichia coli.* Tartu, 2009, 147 p.
168. **Lea Hallik.** Generality and specificity in light harvesting, carbon gain capacity and shade tolerance among plant functional groups. Tartu, 2009, 99 p.
169. **Mariliis Tark.** Mutagenic potential of DNA damage repair and tolerance mechanisms under starvation stress. Tartu, 2009, 191 p.
170. **Riinu Rannap.** Impacts of habitat loss and restoration on amphibian populations. Tartu, 2009, 117 p.
171. **Maarja Adojaan.** Molecular variation of HIV-1 and the use of this knowledge in vaccine development. Tartu, 2009, 95 p.
172. **Signe Altmäe.** Genomics and transcriptomics of human induced ovarian folliculogenesis. Tartu, 2010, 179 p.
173. **Triin Suvi.** Mycorrhizal fungi of native and introduced trees in the Seychelles Islands. Tartu, 2010, 107 p.
174. **Venda Lauringson.** Role of suspension feeding in a brackish-water coastal sea. Tartu, 2010, 123 p.
175. **Eero Talts.** Photosynthetic cyclic electron transport – measurement and variably proton-coupled mechanism. Tartu, 2010, 121 p.
176. **Mari Nelis.** Genetic structure of the Estonian population and genetic distance from other populations of European descent. Tartu, 2010, 97 p.