

**„Multiplex-Genotypisierung forensisch
relevanter SNPs unter Verwendung von
Microarrays auf chemisch aktivierten
Glasflächen“**

Dissertation
zur Erlangung des Grades
„Doktor der Naturwissenschaften“

am Fachbereich Biologie
der Johannes Gutenberg-Universität Mainz

Magdalena Bogus
geb. am 18.12.1975 in Waldenburg

Mainz, 2010

Dekan:

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung: 07.02.2011

Meinen Eltern
in Liebe und Dankbarkeit

1	ZUSAMMENFASSUNG	1
2	EINLEITUNG	3
2.1	FORENSISCH RELEVANTE POLYMORPHISMEN	3
2.1.1	ERSTE GENETISCHE MARKER	3
2.1.2	REPETITIVE DNA-SEQUENZEN	3
2.1.3	SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs)	4
2.1.3.1	Y-chromosomale SNPs	5
2.1.3.2	Mitochondriale SNPs (mtSNPs)	6
2.2	VERFAHREN ZUR SNP-DETEKTION	7
2.2.1	KLASSIFIKATION UND GRUNDLAGEN	7
2.2.1.1	Restriktionsfragment-Längen-Polymorphismus (RFLP)-Analyse	9
2.2.1.2	Physikalische Grundlagen der Fluoreszenzdetektion	10
2.2.2	DNA-POLYMERASE-BASIERTE GENOTYPISIERUNGSMETHODEN	11
2.2.2.1	Single Base Extension-Technologie - Minisequenzierung	11
2.2.2.2	SNaPshot™ Multiplex Kit-Reaktion	12
2.2.3	LIGATIONSBASIERTE GENOTYPISIERUNGSMETHODEN	13
2.2.4	HYBRIDISIERUNGSBASIERTE GENOTYPISIERUNGSMETHODEN	16
2.2.4.1	Microarray-Technologie	16
2.2.4.2	SBE-TAG-Reaktion an Glasobjektträgern	19
2.2.4.3	„Array of arrays“-Prinzip	20
2.2.5	ANDERE ENZYMKATALYSIERTE GENOTYPISIERUNGSVERFAHREN AUF FESTEN TRÄGERMATERIALIEN	21
2.3	SNPFORID-PROJEKT (EU-PROJEKT)	22
2.3.1	ARBEITSTHEMA 1: SNP-AUSWAHL	23
2.3.2	ARBEITSTHEMA 2: ENTWICKLUNG DER MULTIPLEXMETHODEN	25
2.3.3	ARBEITSTHEMA 3: ENTWICKLUNG DER DETEKTIONSMETHODEN	26
2.3.4	ARBEITSTHEMA 4: FORENSISCHE VALIDIERUNG	27
2.4	ZIELSETZUNGEN DER ARBEIT UND DES PROJEKTS	27
3	MATERIAL UND METHODEN	29
3.1	VORBEREITENDE VERFAHREN ZUR PROBENANALYSE MITTELS MULTIPLEXREAKTIONEN	29
3.1.1	SNP-SELEKTION	29
3.1.2	PCR-PRIMERDESIGN	30
3.1.3	MULTIPLEX-PCR-REAKTION	30
3.1.4	ENZYMATISCHE PROBENAUFREINIGUNG	31
3.2	VORVALIDIERUNG DER 23 SNP MULTIPLEXREAKTION MITTELS MINISEQUENZIERUNG	32
3.2.1	DNA-PROBEN	32
3.2.2	PROBENAUFBEREITUNG MITTELS WHOLE GENOME AMPLIFICATION (WGA)	33
3.2.3	PROBENDETEKTION	33
3.2.3.1	SBE-Primerdesign	33
3.2.3.2	Minisequenzierungsreaktion	33
3.2.3.3	Kapillarelektrophorese	34
3.2.4	AUSWERTUNG DER POPULATIONSDATEN	35
3.2.4.1	Hardy-Weinberg-Gleichgewicht	35
3.2.4.2	Exakter Test auf Unabhängigkeit	36
3.2.4.3	F-Statistiken	37
3.2.4.4	AMOVA	37
3.2.4.5	Korrelationskoeffizient	38
3.2.4.6	Ausschlusswahrscheinlichkeit	38
3.3	VALIDIERUNG DER 52 SNP MULTIPLEXREAKTION MITTELS MINISEQUENZIERUNG	39
3.3.1	SNP-SELEKTION UND PCR-PRIMERDESIGN	39
3.3.2	DNA-PROBEN	39

3.3.3	PROBENDETEKTION	40
3.3.3.1	SBE-Primerdesign	40
3.3.3.2	Sequenzierungsreaktion und Kapillarelektrophorese	41
3.3.3.3	Signalnormalisierung	42
3.3.3.4	Reproduzierbarkeits- und Sensitivitätsüberprüfung	42
3.3.4	AUSWERTUNG DER POPULATIONS DATEN	42
3.3.4.1	Linkage Disequilibrium	43
3.3.4.2	F-Statistiken und Korrelationskoeffizient	43
3.3.4.3	Clusteranalyse	43
3.3.4.4	Forensische Biostatistik	44
3.3.5	STR-TYPISIERUNG UND SEQUENZIERUNG	44
3.4	MICROARRAY-ANALYSE DER 33 SNP MULTIPLEXREAKTION	45
3.4.1	DNA-PROBEN	45
3.4.1.1	Proben zur Optimierung der Microarray-Analyse	45
3.4.1.2	Proben für die Sensitivitätsstudie	45
3.4.2	REFERENZMETHODEN	46
3.4.2.1	Sequenzierung	47
3.4.2.2	Fällung der Sequenzierung	47
3.4.3	ARRAYDESIGN	48
3.4.3.1	Versuchsaufbau 1	48
3.4.3.2	Versuchsaufbau 2	49
3.4.4	TAG-DESIGN	52
3.4.4.1	Spotten der ctags	53
3.4.4.2	Blocken der Oberfläche	54
3.4.5	SONDENDESIGN	54
3.4.6	PROBENDETEKTION UND AUSWERTUNG	55
3.4.6.1	Sequenzierungsreaktion für Microarrays	55
3.4.6.2	Hybridisierung der fluoreszenzmarkierten Oligonukleotide	55
3.4.6.3	Detektion der Signale	56
3.4.6.4	Quantifizierung der Signalintensitäten	57
3.4.6.5	Normalisierung der Messdaten	57
3.4.7	VISUALISIERUNG DER DATEN UND GENOTYPISIERUNG	58
3.4.8	KLASSIFIZIERUNG DER DATEN MITTELS DES SILHOUETTE-KOEFFIZIENTEN	61

4 ERGEBNISSE **63**

4.1	VORVALIDIERUNG DER 23 SNP MULTIPLEXREAKTION	63
4.1.1	PROBENAUFBEREITUNG MITTELS WGA	63
4.1.2	AUSWERTUNG DER POPULATIONS DATEN	65
4.1.2.1	Typisierungserfolg	65
4.1.2.2	Allelverteilung in vier Populationen	65
4.1.2.3	Hardy-Weinberg-Gleichgewicht	68
4.1.2.4	F-Statistiken	68
4.1.2.5	AMOVA	69
4.2	MICROARRAY-ANALYSE DER 33 SNP MULTIPLEXREAKTION	71
4.2.1	ENTWICKLUNG UND OPTIMIERUNG DER PROBENDETEKTION MITTELS MICROARRAY-ANALYSE	71
4.2.1.1	Tag- und Sondendesign	71
4.2.1.2	Reaktionskontrollen	72
4.2.1.3	Umfang der Multiplexreaktion	73
4.2.1.4	Etablierung einer 33 SNP Multiplexreaktion	74
4.2.1.5	Spotreinheit	76
4.2.1.6	Optimierung der Microarray-Reaktion durch Anwendung unterschiedlicher Sequenzierenzyme	81
4.2.2	PROBENDETEKTION UND EVALUIERUNG	82
4.2.2.1	Visualisierung der Daten und Einzelbetrachtung der ausgewählten Sonden	82
4.2.2.1.1	Gut diskriminierende SNPs	84

4.2.2.1.2	SNPs mit mittlerer Clusterqualität	89
4.2.2.1.3	Systeme mit schlechter Clusterqualität	93
4.2.3	AUSWERTUNG DER DATEN	96
4.2.3.1	Referenzmethoden	96
4.2.3.2	Typisierungserfolg	98
4.2.4	KLASSIFIZIERUNG DER DATEN MITTELS SILHOUETTE-KOEFFIZIENTEN	103
4.2.5	ERGEBNISSE DER SENSITIVITÄTSSTUDIE	105

5 DISKUSSION **107**

5.1	DISKUSSION ZUR VORVALIDIERUNG DER 23-SNP MULTIPLEXREAKTION	107
5.1.1	PROBENUMFANG UND WHOLE GENOME AMPLIFICATION (WGA)	107
5.1.2	TECHNISCHEN VORRAUSSETZUNGEN	109
5.1.3	DISKUSSION ZUR AUSWERTUNG POPULATIONSGENETISCHER DATEN	110
5.1.4	FAZIT	113
5.2	VALIDIERUNG DER 52 SNP MULTIPLEXREAKTION – ERGEBNISSE UND DISKUSSION DER GEMEINSAMEN PROJEKTARBEIT	114
5.2.1	TECHNISCHE VORAUSSETZUNGEN UND TYPISIERUNGSERFOLG INNERHALB DER INSTITUTE	114
5.2.2	ANALYSE DER 52-SNP MULTIPLEXREAKTION	116
5.2.2.1	Allelverteilung in neun Populationen	118
5.2.2.2	Clusteranalyse	120
5.2.2.3	Forensische Biostatistik	121
5.3	DISKUSSION ZUR MICROARRAY-ANALYSE DER 33 SNP MULTIPLEXREAKTION	123
5.3.1	TECHNISCHE VORAUSSETZUNGEN	123
5.3.1.1	Versuchskonzeption und Optimierung	123
5.3.1.2	„Array of arrays“-Prinzip	124
5.3.1.3	Kontamination	124
5.3.2	SIGNALDETEKTION	125
5.3.2.1	Signalüberlappung	125
5.3.2.2	Untersuchungen des Bindeverhaltens der tag/ctag-Sequenzen während der Hybridisierung	125
5.3.2.3	Limitierende Faktoren der Minisequenzierungsreaktion	126
5.3.2.3.1	Signalstärke und Einbauverhalten	127
5.3.2.3.2	Einbauverhalten und Fehleinbauten	127
5.3.3	TYPISIERUNGSERFOLG UND VERGLEICHENDE ANALYSE DER REFERENZMETHODEN	128
5.3.4	QUALITÄTSPRÜFUNG MITTELS SILHOUETTE-KOEFFIZIENTEN	131
5.3.5	SENSITIVITÄT DER MICROARRAY-MULTIPLEXREAKTION	133
5.3.6	ABSCHLIEßENDE BEWERTUNG DER METHODE	134
5.4	VERGLEICHENDE BETRACHTUNG ANDERER TYPISIERUNGSMETHODEN UND AUSBLICK	136

6 LITERATURVERZEICHNIS **144**

ANHANG

1 Zusammenfassung

Die Analyse tandem-repetitiver DNA-Sequenzen hat einen festen Platz als genetisches Typisierungsverfahren in den Bereichen der stammesgeschichtlichen Untersuchung, der Verwandtschaftsanalyse und vor allem in der forensischen Spurenkunde, bei der es durch den Einsatz der Multiplex-PCR-Analyse von *Short Tandem Repeat*-Systemen (STR) zu einem Durchbruch bei dem Nachweis und der sicheren Zuordnung von biologischen Tatortspuren kam. Bei der Sequenzierung des humanen Genoms liegt ein besonderes Augenmerk auf den genetisch polymorphen Sequenzvariationen im Genom, den SNPs (*single nucleotide polymorphisms*). Zwei ihrer Eigenschaften – das häufige Vorkommen innerhalb des humanen Genoms und ihre vergleichbar geringe Mutationsrate – machen sie zu besonders gut geeigneten Werkzeugen sowohl für die Forensik als auch für die Populationsgenetik.

Ziel des EU-Projekts „SNPforID“, aus welchem die vorliegende Arbeit entstanden ist, war die Etablierung neuer Methoden zur validen Typisierung von SNPs in Multiplexverfahren. Die Berücksichtigung der Sensitivität bei der Untersuchung von Spuren sowie die statistische Aussagekraft in der forensischen Analyse standen dabei im Vordergrund. Hierfür wurden 52 autosomale SNPs ausgewählt und auf ihre maximale Individualisierungsstärke hin untersucht. Die Untersuchungen der ersten 23 selektierten Marker stellen den ersten Teil der vorliegenden Arbeit dar. Sie umfassen die Etablierung des Multiplexverfahrens und der *SNaPshot*TM-Typisierungsmethode sowie ihre statistische Auswertung. Die Ergebnisse dieser Untersuchung sind ein Teil der darauf folgenden, in enger Zusammenarbeit der Partnerlaboratorien durchgeführten Studie der 52 SNP Multiplexmethode.

Ebenfalls im Rahmen des Projekts und als Hauptziel der Dissertation erfolgten die Etablierung und Evaluierung des auf der Microarray-Technologie basierenden Verfahrens der Einzelbasenverlängerung auf Glasobjektträgern. Ausgehend von einer begrenzten DNA-Menge wurde hierbei die Möglichkeit der simultanen Hybridisierung einer möglichst hohen Anzahl von SNP-Systemen untersucht. Die Auswahl der hierbei eingesetzten SNP-Marker erfolgte auf der Basis der Vorarbeiten, die für die Etablierung des 52 SNP Multiplexes erfolgreich durchgeführt worden waren.

Aus einer Vielzahl von Methoden zur Genotypisierung von biallelischen Markern hebt sich das Assay in seiner Parallelität und Einfachheit des experimentellen Ansatzes durch eine erhebliche Zeit- und Kostenersparnis ab. In der vorliegenden Arbeit wurde das „array of arrays“-Prinzip eingesetzt, um zur gleichen Zeit unter einheitlichen Versuchsbedingungen zwölf DNA-Proben auf einem Glasobjektträger zu typisieren. Auf der Basis von insgesamt 1419 typisierten Allelen von 33 Markern konnte die Validierung mit einem Typisierungserfolg

von 86,75 % abgeschlossen werden. Dabei wurden zusätzlich eine Reihe von Randbedingungen in Bezug auf das Sonden- und Primerdesign, die Hybridisierungsbedingungen sowie physikalische Parameter der laserinduzierten Fluoreszenzmessung der Signale ausgetestet und optimiert.

2 Einleitung

2.1 Forensisch relevante Polymorphismen

2.1.1 Erste genetische Marker

Bereits am Anfang des 20. Jahrhunderts untersuchte und beschrieb Karl Landsteiner den ersten polymorphen molekularen Marker (Landsteiner, 1901). Es war das humane AB0-System der Blutgruppen und Landsteiner wurde 1930 für diese Entdeckung mit dem Nobelpreis geehrt. Fünfzig Jahre später, folgend auf die Blutgruppensysteme, wurden Polymorphismen der Serumproteine in die Abstammungsbegutachtung miteinbezogen (Smithies, 1955). Weitere serologische Untersuchungswerkzeuge kamen mit der Entdeckung der erythrozytären Enzym polymorphismen (Hopkinson *et al.* 1963, 1964, Spencer *et al.*, 1964) sowie mit der Etablierung des HLA-Systems (*Human Leucocyte Antigen*), das ebenfalls mit einem Nobelpreis (1980) ausgezeichnet wurde und auf die Entdeckung von Jean Dausset zurückgeht (Dausset, 1954), hinzu.

Seither sind viele weitere genetische Variationen beschrieben worden. Als forensisch besonders relevant erwiesen sich hierbei DNA-Polymorphismen, die bereits seit den frühen 1980er Jahren verfügbar waren. Diese werden im nächsten Kapitel genauer beschrieben, wobei die größte Aufmerksamkeit auf die arbeitsrelevanten Einzelbasenpolymorphismen gerichtet ist.

2.1.2 Repetitive DNA-Sequenzen

Unter Minisatelliten fasst man repetitive DNA-Bereiche mittlerer Länge zusammen. Die sich wiederholenden Sequenzabschnitte können eine Größe zwischen 10 und 100 *bp* besitzen und sich über eine Länge zwischen hundert und mehreren tausend Basenpaaren wiederholen. Sie kommen häufig in Telomer- oder Centromerregionen vor und sind als Marker zur Kopplungsanalyse nicht gut geeignet (Vergnaud *et al.* 1991), können jedoch durch ihre Variabilität innerhalb der Populationen und die stabile Vererbung zur Bestimmung des genetischen Fingerabdrucks unter Verwendung von Multi- und Singlelocussonden genutzt werden (Jeffreys *et al.* 1987).

Neben den Minisatelliten treten auch Mikrosatelliten, besser bekannt als STRs (*short tandem repeats*), in nichtcodierenden Bereichen der DNA auf. Diese werden standardisiert als Werkzeuge zur Identifikation und zur Bestimmung der Abstammung hinzugezogen (Queller, 1993, Blouin *et al.* 1995). Sie wurden im Genom von *Drosophila virilis* bereits 1984 von Tautz und Renz entdeckt (Tautz und Renz, 1984a, 1984b) und 1989 ebenfalls von Tautz als geeignete Marker zur Identifizierung, für Kopplungsanalysen, zur Kartierung sowie zu populationsgenetischen Studien benutzt.

Der Vollständigkeit halber sollen hier auch die transponiblen, repetitiven DNA-Sequenzen erwähnt werden. Sie unterteilen sich in vier DNA-Klassen: 1) SINEs (*short interspersed nuclear elements*), 2) LINEs (*long interspersed nuclear elements*), 3) LTR-Elemente (*long terminal (direct) repeats*) und 4) DNA-Transposons. Es sind allesamt mobile Elemente mit unterschiedlichen Eigenschaften sich innerhalb des Genoms zu bewegen (Smit *et al.* 1995, Smit und Riggs, 1996, Prak und Kazazian 2000). Zu ihnen gehören u. a. die *Alu*-Elemente aus der SINEs-Klasse.

2.1.3 Single Nucleotide Polymorphisms (SNPs)

Neben Mikrosatelliten oder den mobilen *Alu*-Elementen gehören Einzelbasenpolymorphismen, die so genannten *single nucleotide polymorphisms* (SNPs), zu den Sequenzpolymorphismen, die am häufigsten im menschlichen Genom vorkommen. Sie sind zufällig auf der DNA verteilt, können jedoch an bestimmten Stellen häufiger vorkommen als an anderen (Venter *et al.* 2001). Wie Untersuchungen zeigen, liegen sie im Genom alle 1000 bis 2000 Nukleotide vor (Sachidanandam *et al.* 2001). Bei SNPs handelt es sich grundsätzlich um biallelische Marker, vereinzelt können jedoch auch drei Allele detektiert werden (Phillips *et al.* 2004a).

Ausgehend von weit über drei Milliarden Basenpaaren eines kompletten menschlichen Genoms nehmen die variablen Marker einen Anteil von 0,1% ein. Sie beschreiben die genetische Diversität innerhalb der unterschiedlichen Populationen und lassen auf die vorangegangenen evolutiven Schritte zurückblicken (Hacia *et al.* 1999, Stoneking 2001).

Tritt ein Basenaustausch mit einer Häufigkeit von über einem Prozent innerhalb einer Population auf, so handelt es sich nicht mehr nur um eine Variante (Wang *et al.* 1998), sondern um ein stabil vererbtes Allel (SNP), das zusammen mit weiteren Polymorphismen eine populationsgenetische Aussage liefern kann.

Abhängig davon, an welcher Stelle der Austausch im Genom stattfindet, kann man SNPs in drei Kategorien aufteilen: a) codierende SNPs, diese liegen in Bereichen der Gene (Exons) und haben mit einer höheren Wahrscheinlichkeit Einfluss auf die Funktion eines Proteins, b) Gen-assoziierte SNPs, welche in unmittelbarer Nachbarschaft zu Genen oder in den Introns der Gene vorliegen. Sie können *enhancer* oder Promotorenbereiche beeinflussen, indem sie das Ablesen eines Gens unterdrücken oder beschleunigen. Die beiden Kategorien werden ebenfalls als phänotyprelevante SNPs zusammengefasst (Halushka *et al.* 1999). c) *random* SNPs – zufällige SNPs. Unter die letzte Kategorie fallen die meisten SNPs, denn es handelt sich hierbei um Polymorphismen, die innerhalb nichtcodierender genomischer Bereiche

liegen, welche ca. 90% unseres Genoms ausmachen. Sie haben in der Regel keine funktionalen Konsequenzen, können jedoch zu Kartierungszwecken genutzt werden¹.

SNPs der letzten Gruppe stellen mit ihren Eigenschaften ideale Marker zur Identifizierung von Individuen dar. Sie besitzen eine geringere Mutationsrate (2×10^{-8} pro Base pro Generation) im Vergleich zu den bisher benutzten Analysesystemen (STR oder VNTR) und sind daher stabiler (Brinkmann, 1998, Sachidanandam *et al.* 2001). So lassen SNPs die Detektion an kurzen DNA-Bruchstücken zu und können auch dann erfolgreich analysiert werden, wenn biologisches Material in geringer Menge, nach starker, altersbedingter Zersetzung oder nach Zerstörung durch biologische oder chemische Einflüsse, stark degradiert vorliegt. Das macht SNPs zu idealen Markern für populationsgenetische, anthropologische und forensische Untersuchungen. Zusätzlich steht der SNP-Analyse ein breites Spektrum an Analysemethoden zur Verfügung und das Interesse an der Entwicklung neuer und effizienterer Methoden hat noch kein Ende erreicht (Kapitel 2.2).

2.1.3.1 Y-chromosomale SNPs

Populationsgenetische und forensische Marker sind nicht nur auf autosomalen DNA-Abschnitten zu finden. Besonders das geschlechtsspezifische und nur bei männlichen Individuen vorhandene Y-Chromosom ergänzt die DNA-Analyse. Die Vererbung verläuft ausschließlich in paternaler Linie. Das Chromosom wird vom Vater auf die männlichen Nachkommen unverändert als ein so genannter Y-Haplotyp vererbt (Jobling *et al.* 1999) und ist aus Mangel eines Partnerchromosoms bis auf zwei pseudoautosomale Regionen (PAR1 und PAR2) der Telomere nicht rekombinant (Rappold, 1993).

Aufgrund der hohen Mutationsrate der STR-Marker und der damit verbundenen Unzulänglichkeit Generationsfolgen genau zu bestimmen, werden Y-SNPs mit Vorzug zur kombinierten Haplotypisierung hinzugezogen. Die Y-chromosomalen SNPs stehen mit einer Mutationsrate von 2×10^{-8} per Base pro Generation (Underhill *et al.* 1997, Thomson *et al.* 2000) den repetitiven Markern entgegen. Für diese wurden die Mutationsraten durchschnittlich auf 2×10^{-3} per Locus in einer Generation (Heyer *et al.* 1997, Jobling *et al.* 1999) berechnet. Neben den bereits länger etablierten Typisierungsmöglichkeiten (Y-Chromosome Consortium - YCC, 2002) mit 243 binären Markern und 153 definierten Haplogruppen bieten aktuelle Sets der binären Marker neue Möglichkeiten zur Haplotypisierung (Karafet *et al.* 2008). Die Anzahl der Polymorphismen hat sich auf 600 und die der definierten Haplogruppen auf 311 erhöht. Das YCC hat alle Haplogruppen klassifiziert und in 18 Hauptgruppen unterteilt, die einer alphabetischen Nomenklatur folgen und von A bis R reichen. Mit der Neuklassifizierung der Haplogruppen im Jahr 2008 wurden diese um zwei Gruppen ergänzt (S und T). Durch die Bestimmung der Haplogruppen können

¹ http://www.rochetest.com/de/sci_gene_snps_d.pdf

demographische Expansionen von Populationen verfolgt sowie Ursprung und Diversität spezifischer Populationen untersucht werden. In Studien wurde gezeigt, dass eine Untersuchung von 29 Y-SNPs in der Lage ist, die meisten weltweit verbreiteten Haplogruppen in einer Multiplexreaktion zu definieren (Bríón *et al.* 2005, Bríón, 2005). Für genauere Haplotypisierung gewünschter regionaler Gruppen sind bereits kommerzielle Reaktionen auf dem Markt erhältlich (*Signet™ Y-SNP Genotyping kit*, Marligen Biosciences, Ijamsville, USA)².

Auf populationsgenetischer Ebene lassen die Y-STR Marker nur eine Zugehörigkeitsbestimmung zu einer Familienlinie zu (Roewer, 2001). Die Y-chromosomalen Marker bieten sich jedoch hervorragend zur Vaterschaftsdiagnostik z. B. bei einem Vater/Sohn Verwandtschaftsfall an (Roewer und Epplen, 1992) sowie im Verbrechensfall zur Unterscheidung zwischen einer weiblich/männlichen Mischspur, wie sie im Falle eines Sexualdelikts vorkommen kann (Prinz *et al.* 1997). Hierbei hat die Spur einen überwiegenden DNA-Anteil des weiblichen Opfers. Durch die chemische Analysetechnik der differentiellen Lyse können die Spermien von den Vaginalepithelzellen der weiblichen Person zwar getrennt werden aber nicht in jedem Fall reicht die so gewonnene Menge an Spurenmaterial aus, um ein Profil des Spurenlegers zu erhalten. In einem solchen Fall kann eine Untersuchung mit Y-chromosomalen Markern behilflich sein. Hierfür muss keine Trennung der Sekrete durchgeführt werden, da nur die männliche DNA amplifiziert wird.

2.1.3.2 Mitochondriale SNPs (mtSNPs)

Wie in der Kern-DNA, kommen Sequenzpolymorphismen auch in der zirkulär organisierten, maternal vererbten mitochondrialen DNA vor. Der geschlossene DNA-Doppelstrang ist zwar sehr klein, liegt aber in jeder Zelle in einer ca. 1000fachen Kopienzahl vor. Zudem ist die DNA sehr stark aufgewickelt (*supercoiled*), was ihr zusätzlichen Schutz gegenüber Umwelteinflüssen gibt. Sie steht Analysen auch dann noch zur Verfügung, wenn kein erfolgreiches Typisieren der Kern-DNA mehr möglich ist, so z.B. bei altem Knochenmaterial oder aber bei telogenen, wurzellosen Haaren (Gill *et al.* 1994, Holland und Parsons, 1999).

Den größten Teil der mitochondrialen DNA machen deren Gene aus. Nur ein sehr kleiner Teil des ringförmigen Moleküls ist nichtcodierend und kann für forensische Zwecke genutzt werden. Aus der, Anfang der 1980er Jahre von Anderson beschriebenen Gesamtsequenz mitochondrialer DNA, die 16 569 *bp* umfasst (Anderson *et al.* 1981), sind für forensische Aspekte besonders die hypervariablen Regionen I und II von Bedeutung. Es handelt sich dabei um die mt-DNA-Kontrollregion, die sich über 1122 *bp* erstreckt und den so genannten D-Loop (*displacement loop*) beinhaltet. Diese Region des mitochondrialen Genoms weist u. a. zwei Besonderheiten auf: 1) innerhalb des D-Loops kommt es durch eine Duplikation

² <http://www.marligen.com/y-snp-genotyping-kits.html>

des äußeren Stranges (schwerer Strang, *heavy chain*) zu einer Dreifachstrangstruktur (Strachan, 2005), 2) die Kontrollregion beinhaltet nichtcodierenden Bereiche und weist eine hohe Sequenzvariabilität auf (Stoneking *et al.* 1991). Die Bedeutung dieses Abschnitts zeigt sich bei Untersuchungen von Verwandtschaftsverhältnissen bzw. wird bei Identifizierungen im Sinne einer Ausschlusskonstellation verwendet (Carracedo *et al.* 2000).

Faktoren wie das Vorhandensein von Heteroplasmien und eine größere Kontaminationsgefahr bei der Extraktion von mtDNA machen eine Analyse dieses genetischen Materials besonders sensibel. Bei der Besonderheit der Heteroplasmie handelt es sich um das potentielle Vorhandensein unterschiedlicher mtDNA-Moleküle innerhalb der Zelle. Unterscheiden kann man zwischen zwei Arten von Heteroplasmien der Sequenz- und der Längenheteroplasmie. Ihr Vorkommen kann gewebespezifisch oder auch zellspezifisch sein (MacMillan *et al.* 1993, Lightowlers *et al.* 1997, Jansen, 2000) und kann die Auswertung erschweren. Sie kann aber auch die Beweiskraft der Analyse erhöhen, wie sich am Beispiel der sterblichen Überreste des russischen Zaren Nikolaus II. zeigt, bei dem erst eine Heteroplasmie zu seiner sicheren Identifizierung führte (Gill *et al.* 1994). Häufiger jedoch wirken sich die Sequenzunterschiede innerhalb eines Individuums erschwerend auf dessen Identifizierung aus.

Wie die Y-chromosomalen Marker, spielen die mitochondrialen Marker eine Rolle bei der Untersuchung von Migrationsereignissen. Die Herkunft lässt sich hierbei anhand von Haplogruppen ermitteln (Allard *et al.* 2002). Im forensischen Bereich kann die Herkunftsbestimmung im Fall eines unbekannten Spurenlegers bei der Ermittlungsarbeit hilfreich sein. Es gibt zuverlässige technische Methoden, die es erlauben, mehrere mtSNPs einer Haupthaplogruppe zusammen zu analysieren und ihre Untergruppen genauer zu differenzieren (Vallone *et al.* 2004, Brandstätter *et al.* 2006).

2.2 Verfahren zur SNP-Detektion

2.2.1 Klassifikation und Grundlagen

Unterschiedliche Ansatzmöglichkeiten zur Genotypisierung von SNPs erlauben eine Unterteilung der Verfahren in Kategorien, so zum Beispiel nach der Notwendigkeit einer Präamplifikation, nach der Diskriminierung oder nach der Detektionsmethode. Eine übersichtliche Klassifikation nach Diskriminierungsreaktionen ist in Tabelle 1 zusammengefasst.

Tabelle 1: Klassifikation der SNP-Detektmethode nach Diskriminierungsreaktion.

Diskriminierungsmethode	Reaktionsformat	Detektmethode
i. enzymatisches Schneiden (RFLP)	i. homogene Formate	i. Lumineszenz
ii. SBE	ii. feste Oberflächen (Mikrotiterplatten, Beads, Microarrays)	ii. Fluoreszenz
iii. Ligation		iii. zeitauflösende Fluoreszenz
iv. Hybridisierung		iv. FRET
		v. Fluoreszenz-Polarisation
		vi. Massenspektrometrie
		vii. elektronische Detektion

In der folgenden Darstellung wird nach den Reaktionsprinzipien unterschieden (Abbildung 1). Sie gibt übersichtlich wieder, über welche Reaktionsschritte und Detektmethode die einzelnen Reaktionsformate untereinander kombinierbar sind.

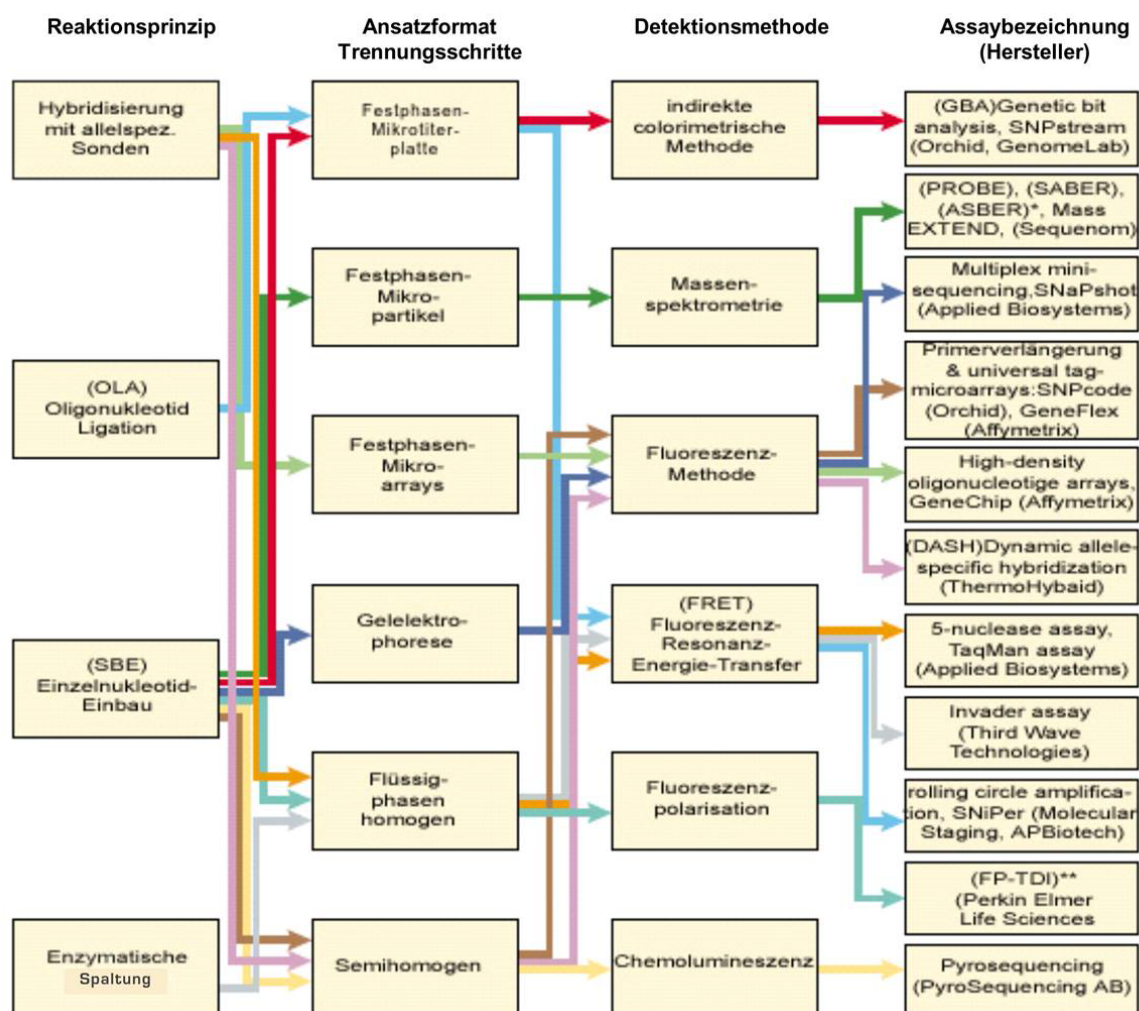


Abbildung 1: Baukastenprinzip-Darstellung einiger, teilweise miteinander kombinierbarer SNP-Genotypisierungsansätze. Farbige Pfeile verbinden die Reaktionsprinzipien, die Ansatzformate und die Detektmethode und geben an, welche Firma den Ansatz als kommerzielles Kit anbietet. Die Abbildung bietet einen Überblick über automatisierte SNP-Analysemethoden und hat keinen Anspruch auf Vollständigkeit (modifiziert nach Syvänen, 2001).

*SABER- single-allele base extension reaction, ASBER- allele-specific base extension reaction

**FP-TDI- fluorescence polarization template-directed dye incorporation

Die vier Reaktionsprinzipien, auf denen die folgenden Detektionsmethoden beruhen, werden im Textverlauf erläutert. Drei der oben genannten Reaktionsprinzipien fanden in Form unterschiedlicher Methoden im Gesamtprojekt Anwendung und/oder waren arbeitsrelevant (Kapitel 2.2.2, 2.2.3 und 2.2.4). Das Wirkungsprinzip der enzymatischen Spaltung wird zur Vervollständigung des Überblicks ergänzt und kurz beschrieben (Kapitel 2.2.1.1).

Da bei allen vier Reaktionsprinzipien die Detektion von Fluoreszenzsignalen eine hervorgehobene Rolle spielt, wird im Folgenden ein kurzer Überblick über die Fluoreszenzdetektion gegeben (Kapitel 2.2.1.2).

2.2.1.1 Restriktionsfragment-Längen-Polymorphismus (RFLP)-Analyse

Bei dieser vergleichsweise alten und gebräuchlichen Methode zur SNP-Bestimmung handelt es sich um eine Kombination des enzymatischen Schneidens und einer Gelelektrophorese. Die Regionen des bekannten Polymorphismus werden zunächst mittels PCR amplifiziert, um anschließend mit sequenzspezifischen Restriktionsenzymen geschnitten zu werden. Dazu werden Endonukleasen gewählt, die an der entsprechenden SNP-Stelle zum Schneiden ansetzen. Dabei entsteht die Schnittstelle mit dem Vorliegen oder Fehlen des Polymorphismus. Trägt man das verdauten Produkt auf ein Gel auf, so entstehen typische Bandenmuster für das jeweilige Allel (Wildtyp oder Mutante). Ist der Genotyp heterozygot, liegen sowohl die geschnittenen als auch die ungeschnittenen Fragmente vor (Abbildung 2).

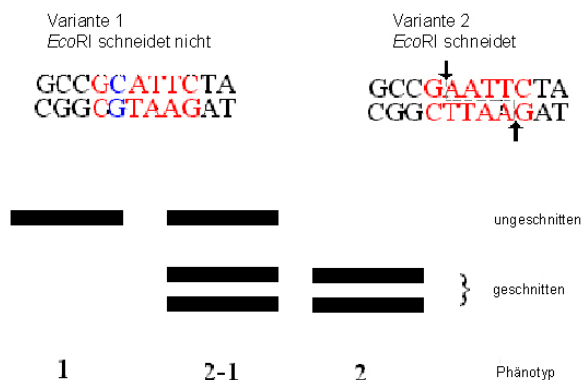


Abbildung 2: Prinzip der RFLP-Analyse. Schnittstelle des Restriktionsenzym *EcoRI* und sich aus dem Verdau ergebendes Muster. 1 homozygoter Wildtyp, 2-1 heterozygote und 2 homozygote Mutante. Nach: http://www.ucl.ac.uk/~ucbhjow/b241/biochemical_2.html.

Die Vorteile dieser Methode liegen in dem sehr geringen Aufwand bei der Durchführung. Zusätzlicher Vorteil ist die kostengünstige Realisierung der Analysen. Dennoch ist die Methode nicht immer exakt genug. Die Eindeutigkeit der Ergebnisse hängt vom Schneideverhalten der Enzyme ab. Des Weiteren kann die Methode nur auf bekannte polymorphe Bereiche angewendet werden, da nur dann die Restriktionsstellen dem passenden Enzym zugeordnet werden können. Zudem sollten keine weiteren passenden

Restriktionsstellen im unmittelbaren Bereich der untersuchten Sequenz liegen. Diese würden ebenfalls enzymatisch verdaut werden und zusätzliche Signale ergeben.

Die Methode ist weiterhin nicht hochdurchsatz- und nicht multiplexfähig, so dass der Zeitaufwand nur eine Bestimmung einzelner SNPs rechtfertigt (Syvänen, 2001).

Ein Verfahren, welches es ermöglicht, auf der Basis des enzymatischen Restriktionsverdaus, mehrere polymorphe Stellen parallel zu bestimmen, ist das *Flap*-Endonukleaseverfahren (*Invader Assay*) und ist bei Kwok (2001) nachzulesen.

2.2.1.2 Physikalische Grundlagen der Fluoreszenzdetektion

Das Phänomen der Fluoreszenz ist eine der vielen Formen der Lumineszenz, zu welcher jede Art von Strahlungsemission im sichtbaren, infraroten und ultravioletten Spektralbereich gehört. Diese Art von Strahlungsemission kann bei gasförmigen Stoffen, Flüssigkeiten oder Feststoffen (außer Metall) nach Anregung durch externe Energiequellen (z.B. Elektronenstahlen) freigesetzt werden. Ist der Energiegehalt der externen Quelle größer, werden die Elektronen aus ihren Orbitalen auf ein höheres Energieniveau gehoben. Charakteristisch für die Fluoreszenz ist, dass das angeregte Molekül sich in einem so instabilen Zustand befindet, dass es innerhalb von 10^{-10} bis 10^{-7} Sekunden unter Abgabe von Strahlung (Emission) in den Ursprungszustand zurückfällt. Die Emission erfolgt immer im langwelligeren Bereich, verglichen zur Absorption. Die Übersicht über die energetischen Vorgänge gibt das Jablonski Diagramm anschaulich wieder (Abbildung 3).

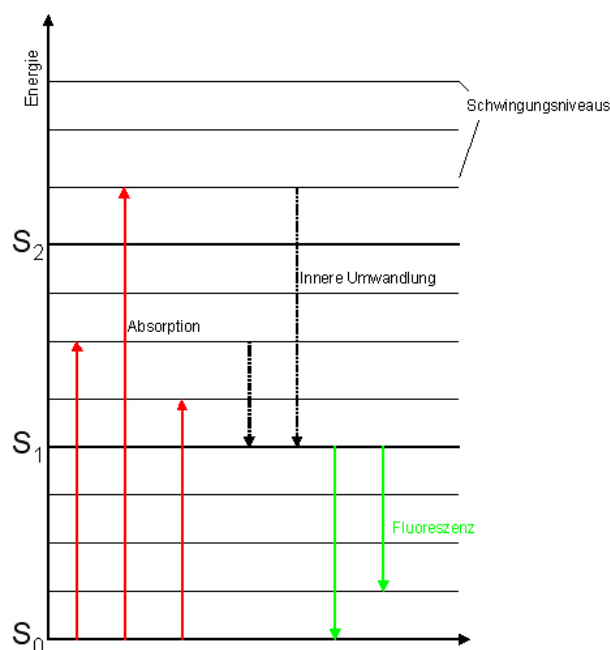


Abbildung 3: Jablonski Diagramm: Energieschema für ein Molekül mit den potentiellen energetischen Übergängen. S_0 – Grundzustand des Moleküls (Singulett-Zustand), S_1 – erster angeregter Zustand, S_2 – zweiter angeregter Zustand.

Die im Folgenden beschriebenen Verfahren zur SNP-Detektion beruhen zum größten Teil auf der Messung eines Signals. Dabei spielen Licht und Energie eine besondere Rolle. Die am häufigsten verwendete Detektionsmethode für SNP-Analysen ist die Fluoreszenzdetektion, weswegen auf das Prinzip hier näher eingegangen wird. Infolgedessen kann durch die Verwendung von Fluorochromen in vielen Analyseverfahren die Markierung mit radioaktiven Isotopen vermieden werden.

Die direkte Markierung der einzelsträngigen DNA beruht auf einer kovalenten Bindung des Farbstoffs am 5'- oder 3'-Ende. Der potentielle Absorptions- und Emissionsbereich der Farbstoffe erstreckt sich dabei über den gesamten Spektralbereich und wird nach dem Detektionsspektrum der Messgeräte gewählt. Die Messgeräte verfügen zudem über Laser mit unterschiedlichen Wellenlängen oder über Weißlichtquellen mit Filtersätzen zur Anregung der Fluoreszenzfarbstoffe.

Verfahren wie *SNaPshot*[™], Microarray und das ligationsbasierte *SNPlex*[™] mit der Variante des *Genplex* werden im weiteren Verlauf beschrieben (Kapitel 2.2.2.2, 2.2.3 und 2.2.4).

2.2.2 DNA-Polymerase-basierte Genotypisierungsmethoden

2.2.2.1 Single Base Extension-Technologie - Minisequenzierung

Ein Reaktionsprinzip zur SNP-Genotypisierung mittels eines hoch spezifischen DNA-Polymeraseenzym ist die Primerverlängerung durch einen Einbau eines einzelnen Nukleotids (*single base extension* – SBE). Es handelt sich dabei um eine häufig angewendete und hoch standardisierte SNP-Diskriminierungsmethode, welche durch die Kapillarelektrophorese (Pastinen *et al.* 1996) ausgelesen, mittels Massenspektroskopie (Haff und Smirnov, 1997) detektiert sowie zur Microarray-Detektion (Pastinen *et al.* 1997, Hirschhorn *et al.* 2000, Lindroos *et al.* 2002) angewendet werden kann. Aus der Literatur ist diese Methode unter weiteren Namen bekannt, so z.B. *single nucleotide base extension* (SNBE), *single nucleotide primer extension* (SNuPE), *primer extension* oder *minisequencing*. Das Prinzip der *single base extension* beruht auf einer thermalen, zyklischen Reaktion, bei der ein sequenzspezifischer Primer auf dem gereinigten PCR-Produkt unmittelbar *downstream* an den zu detektierenden Polymorphismus bindet (Abbildung 4).

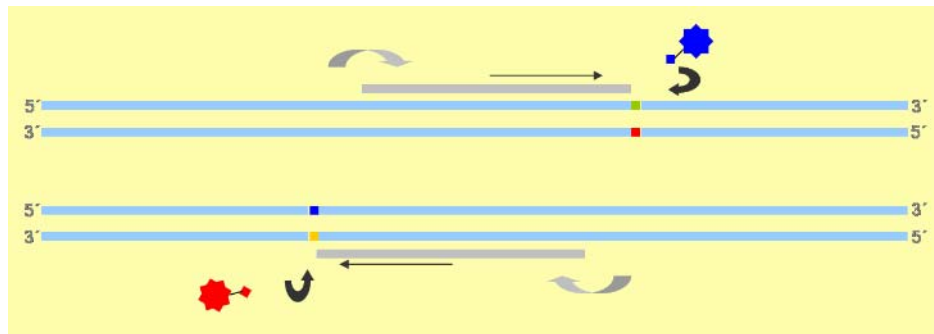


Abbildung 4: Wirkungsprinzip der Minisequenzierung. Der sequenzspezifische Primer (grauer Balken) hybridisiert an das PCR-Produkt in direkter Nachbarschaft zum SNP. Mit Hilfe einer Sequenzierpolymerase erfolgt eine Bindung der komplementären SNP-Base (ddNTP), an die ein Fluoreszenzfarbstoff gekoppelt ist (blauer und roter Stern). Es folgt ein Reaktionsabbruch.

Mit Hilfe einer DNA-Sequenzierpolymerase erfolgt am 3'-Ende der Primersequenz eine Verlängerung des Oligonukleotids um ein einzelnes komplementäres fluoreszenzgekoppeltes ddNTP und es folgt ein initiiertes Reaktionsabbruch. Nach einer Aufreinigung und Denaturierung können die fluoreszenzmarkierten Oligos mittels einer Vielzahl von Nachweisreaktionen auf die homozygoten und heterozygoten Formen der ausgesuchten Loci hin detektiert und analysiert werden (siehe Kapitel 2.2.1, Abbildung 1).

2.2.2.2 SNaPshot™ Multiplex Kit-Reaktion

Die *SNaPshot*™ Kit Reaktion basiert auf dem Prinzip der Einzelbasenverlängerungs-Reaktion (SBE), bei welcher unter Anwesenheit von fluoreszenzmarkierten Dideoxynukleotiden (ddNTPs) und einer DNA-Polymerase ein sequenzspezifischer Einbau der ddNTPs stattfindet. Mit diesem Reaktionsprinzip können die in unterschiedlichen Längen vorliegenden Primersequenzen elektrophoretisch aufgetrennt werden. Bedingt durch ihr unterschiedliches Molekulargewicht wandern die markierten Sequenzen entlang einer Kapillare, vorbei an einem Detektionsfenster, wobei eine laserinduzierte Detektion stattfindet und die elektronisch umgewandelten Signale in einem Elektropherogramm mit fünf Fluoreszenzkanälen ausgelesen werden können.

Es hat sich gezeigt, dass diese Methode hochgradig multiplexfähig ist (Sanchez *et al.* 2006). In den im Projekt etablierten Reaktionssets von mehreren Dutzend unterschiedlichen Produkten werden Analyseprimer in unterschiedlichen Längen in die Reaktion hineingegeben und verlängert, um in einer einzelnen Reaktion die Signalantworten aller Marker zu erhalten. Für die Analyse technisch determinierende Faktoren sind dabei die kleinste Länge eines Analyseprimers von 17 Nukleotiden (nt) und ca. 90 nt als die maximale Sequenzlänge, die effizient und kostenorientiert synthetisiert werden kann. Die *SNaPshot*™ Kit Reaktion (Abbildung 5), mit welcher die Detektion der SNPs stattfand, beinhaltet ein Set aus vier an unterschiedliche Fluoreszenzfarbstoffe gekoppelten Dideoxynukleotide (ddNTPs):

- dR6G (ddATP)

- dTAMRA™ (ddCTP)
- dR110 (ddGTP)
- dROX™ (ddTTP)

Zusätzlich wird ein Längenstandard (*GeneScan™-120 LIZ™ Size Standard*, Applied Biosystems, Foster City, USA) in die Reaktion appliziert, welcher im fünften Farbkanal zur Auswertung hinzugezogen wird.

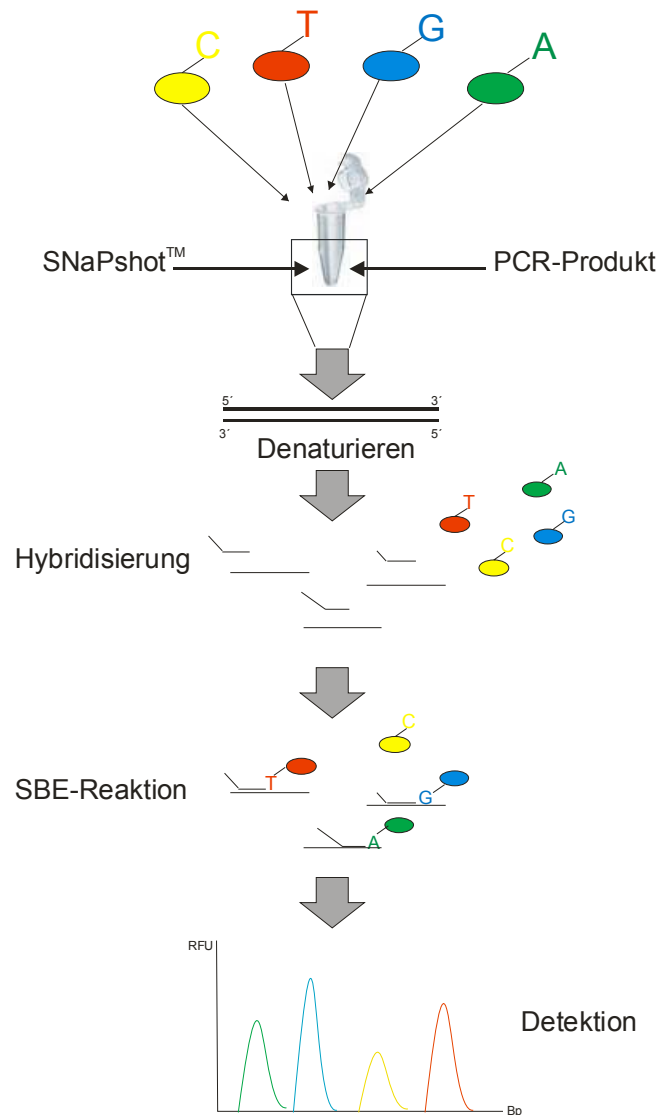


Abbildung 5: Schematische Darstellung des Reaktionsprinzips vom *SNaPshot™ Reaction Kit*.

2.2.3 Ligationsbasierte Genotypisierungsmethoden

Die ligationsbasierte Methode macht sich die Fähigkeiten des DNA-Ligase-Enzyms zu Nutze, welches in der Lage ist DNA-Stränge über Phosphodiesterbindungen zu verknüpfen. Der von Landegren *et al.* (1988) zum ersten Mal beschriebene Typisierungsansatz des *oligo-*

nucleotide ligation assays (OLA) besteht aus drei Sonden: einer so genannten gemeinsamen Sonde und zwei allelspezifischen Sonden, wobei eine am 3'-Ende das komplementäre Nukleotid des Wildtypallels und die andere Sonde das komplementäre Nukleotid des Mutantenallels trägt (*match/mismatch*-Prinzip). Die mit Fluoreszenzfarbstoff markierte gemeinsame Sonde bindet an die Zielsequenz und die allelspezifischen Sonden konkurrieren um die Bindung. Durch die Zugabe der DNA-Ligase wird die Bindung zwischen der allelspezifischen und der gemeinsamen Sonde gebildet und nach der Denaturierung kann die Signalauswertung z.B. durch die Kapillarelektrophorese stattfinden (Abbildung 6). Hierbei wird, wie bei der Minisequenzierung die Durchlaufgeschwindigkeit in der Kapillare vom Molekulargewicht bestimmt. Dieses wird durch eine unspezifische Sequenz einer vorgegebenen Länge an den allelspezifischen Sonden festgelegt.

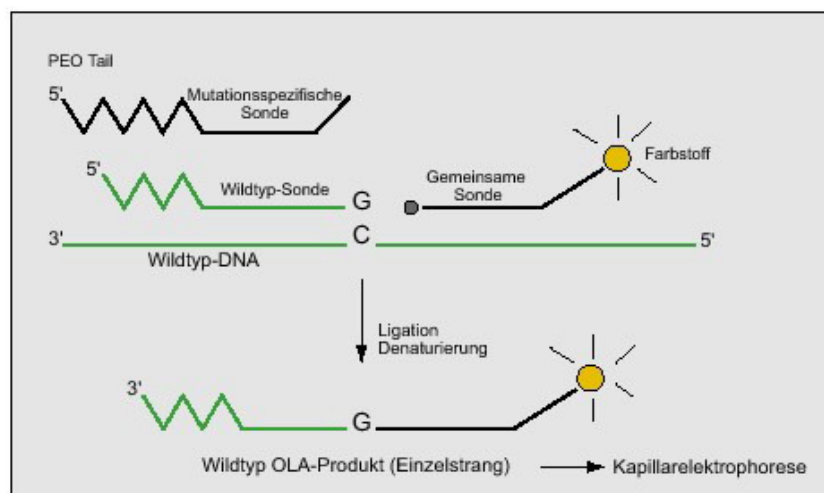


Abbildung 6: Schematische Darstellung der Ligation allelspezifischer Nukleotide (OLA) mit dem Beispiel für eine Wildtyp-Sonde. Die Detektion erfolgt nach Längenauftrennung durch Kapillarelektrophorese der fluoreszenzmarkierten Produkte. Nach: <http://www.medizinische-genetik.de/index.php?id=2001>.

Der Ligationsansatz ist eine sehr flexible Detektionsmethode für Einzelbasenveränderungen. Die Signale können nach Immobilisierung einer Sonde im Microarray-Format detektiert werden (Gerry *et al.* 1999) oder mittels einer kombinierten DNA-Polymerase assistierten und ligasevermittelten Markerdetektion, der sog. *real-time* PCR, unterzogen werden. Hierbei erfolgt die Detektion nach dem FRET-Prinzip (*fluorescence resonance energy transfer*, Chen *et al.* 1998).

Ein ebenfalls ligationsvermitteltes neueres Verfahren ist die *SNPlex*TM-Technologie (Applied Biosystems), die hier genauer beschrieben wird, da sie eine der Vergleichsmethoden für die vorliegende Arbeit darstellt.

Die *SNPlex*TM-Technologie kombiniert ebenfalls die Amplifikations- und die Ligationsreaktion, wobei die Ligation von drei unmarkierten Sonden direkt an genomischer DNA durchgeführt wird. In der folgenden PCR-Reaktion wird ein biotinylierter Primer verwendet,

welcher anschließend an einer mit Streptavidin beschichteten Platte bindet und so das PCR-Produkt in seine Einzelstränge auftrennt. Jetzt erst folgt das Auslesen der Genotypen über fluoreszenzmarkierte und sequenzspezifische Sonden, die sog. *ZipChute*[®]-Sonden. Diese hybridisieren allelspezifisch an die immobilisierten Sequenzen und werden anschließend durch ein Kapillarelektrophorese-Gerät detektiert. Eine Weiterentwicklung dieser Technologie für die Anwendungsmöglichkeiten im forensischen Bereich stellt das *Genplex*-Verfahren dar (Phillips *et al.* 2007a). Hierbei findet die Amplifikation vor der Ligation statt, durch welche die Notwendigkeit umgangen wird, große Mengen genomischen Materials in die Reaktion einzusetzen. Anschließend erst findet die Ligation der drei Sonden statt und die Reaktion wird wie beschrieben durchgeführt (Abbildung 7).

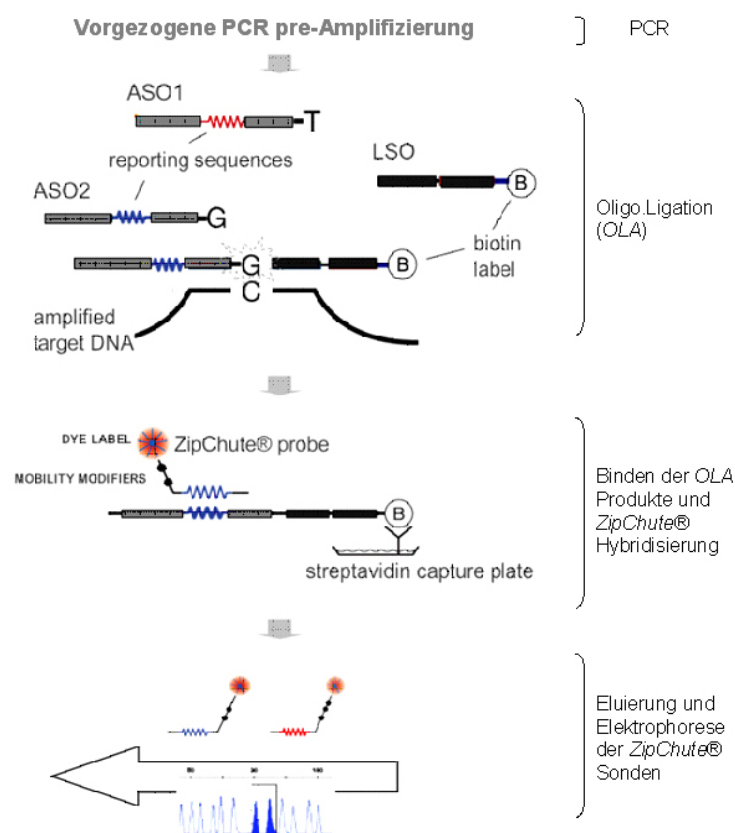


Abbildung 7: Schematischer Ablauf des *Genplex*-Verfahrens. Die Reaktion besteht aus einer vorgezogenen Amplifikation der gewünschten Sequenzen. Anschließend findet die *oligonucleotide ligation assay* (OLA) statt. Die Detektion erfolgt durch die Kapillarelektrophorese nach einer Markierung der OLA-Produkte mit *ZipChute*[®]-Sonden. Modifiziert nach Phillips *et al.* 2007a.

Die Multiplexfähigkeit der beiden beschriebenen Verfahren beschränkt sich gegenwärtig noch auf 48 Marker, die parallel mit zwei Fluoreszenzfarbstoffen genotypisiert werden können.

2.2.4 Hybridisierungsbasierte Genotypisierungsmethoden

Thermale Stabilität der Verbindung von zwei komplementären Sequenzen ist die Ausgangsbasis der Hybridisierung mit allelspezifischen Oligonukleotidsonden (Wallace *et al.* 1979). Die Differenzierung erfolgt bei diesem Prinzip über die Herstellung von Sonden, die an der Stelle des untersuchten Markers ein komplementäres oder ein nicht-komplementäres Nukleotid besitzen können. Meistens wird das *match/mismatch*-Nukleotid in der Mitte der Sondensequenz platziert, da sie dort stärker destabilisierend wirken und so stärker diskriminieren als endständig gelegene Nukleotidunterschiede. Liegt eine vollständige Basenpaarung vor, ist die Verbindung von Sonde und Ziel-DNA stabil und liegt bei der gewählten Schmelztemperatur (T_m). Der T_m -Wert verringert sich in der Regel pro Prozent einer Fehlbindung um 1°C. Mit den längeren Nukleinsäuresonden (ca. 100 nt) können unter weniger stringenten Reaktionsbedingungen auch bei relativ stark abweichender Zielsequenz stabile Bindungen erbracht werden. Die Stabilität bei kurzen Oligonukleotidsonden ist dagegen bei einer nicht-komplementären Base nicht mehr gegeben. Nicht weniger wichtig sind aber auch die Faktoren, wie Salzkonzentration in der Lösung, die flankierenden Sequenzen und die Sekundärstrukturen der Sequenzen, die zusammengekommen die Stringenz der Reaktion definieren (Mir und Southern, 1999).

Ihren Ursprung hat die Hybridisierungsmethode im diagnostischen Bereich. Ihre Anwendung umfasst die Messung von Transkripten bestimmter Gene (Johnston, 1998). Sie umfasst jedoch eine ganze Bandbreite unterschiedlicher Verfahren, die ihren Zukunftsausblick im Nanobereich haben.

Der Schwerpunkt dieser Arbeit liegt in der Analyse der Multiplexfähigkeit der DNA-Microarrays im Hinblick auf die Anwendung im forensischen Bereich. Um die Thematik genauer beleuchten zu können folgt vorab ein kurzer Überblick über die Entwicklung der Microarray-Technologie.

2.2.4.1 Microarray-Technologie

Die Microarray-Technologie ist in der Forschung und innerhalb der Diagnostik eine weit verbreitete Analyseplattform. Die DNA-Array-Technologie macht sich dabei das Prinzip der Basenpaarung nach dem Watson-Crick-Modell zunutze, bei der sich die komplementären Nukleinsäurestränge über Wasserstoffbrückenbindungen aneinanderlagern. Das Reaktionsprinzip wurde erstmals durch den Erfinder des *Southern Blots* angewendet (Southern, 1975). Er hybridisierte vorab markierte Oligonukleotide an ihre komplementären Nukleinsäurestränge, um diese zu analysieren.

Die Technologie macht sich Trägermaterialien zur Nutze, auf denen rasterartig Sondenmoleküle angeordnet sind, die in hoher Anzahl und Dichte immobilisiert werden können.

Diese rasterartige Anordnung führt zu der englischen Bezeichnung „array“ und ihre miniaturisierte Form schließlich zu dem in der Forschung häufig anzutreffenden Namen des „*microarray*“. Die häufigste Anwendung finden Microarrays - wie bereits beschrieben - in den Forschungs- oder Diagnosebereichen, die sich mit Expression von Genen beschäftigen. Zu Zwecken der Genotypisierung werden Oligonukleotid-Microarrays eingesetzt. Sie beruhen auf dem Prinzip der reversen Hybridisierung. Hierbei wird nicht die Sonde markiert und an die Zielsequenz hybridisiert, sondern die unmarkierte Sonde wird an ein Trägermaterial fixiert und die Markierung erfolgt an der Zielsequenz (Nur *et al.* 1990), wobei der Begriff so zu verstehen ist, dass eine Sequenz, die bekannt ist und die der Forscher artifiziell herstellen lassen, den Begriff der reversen Hybridisierung prägt.

Die Trägermatrizen der Arrays können unterschiedlich sein. Aufgrund der geringen Eigenfluoreszenz, die einen unnötig hohen Hintergrund liefern kann, wird neben Gel (Khrapko *et al.* 1991), Nylon- bzw. Nitrocellulose-Membranen (Southern, 1975, Maier *et al.* 1994; z.B.: *SuperArray*, Biomol; *Serum Biomarker Chip High-Density Profiling*, Whatman), Quartz-Objektträgern (z.B.: *Affymetrix SNP Genotyping Service*, ATLAS Biolabs GmbH) oder Plastikchips (z.B.: *TaqMan Array*, Applied Biosystems), vorwiegend Glas als Träger eingesetzt (Guo *et al.* 1994; z.B.: *Affymetrix SNP Genotyping Profiling Service*, KFB; *OpArrays*, Operon).

Bezüglich der Art und Länge der aufgetragenen Sequenzen (Sonden) unterscheidet man zwischen Oligonukleotid- oder cDNA-Arrays, wenn Nukleinsäuren eingesetzt werden. Peptide, Proteine oder Antikörper können ebenfalls auf Arrays immobilisiert werden, werden aber in der vorliegenden Arbeit nicht ausführlicher beschrieben.

Unterschiede gibt es zudem in der Art und Weise, wie Sonden auf das Trägermaterial aufgetragen werden. Dieses kann durch eine *in situ*-Synthese erfolgen, bei welcher DNA in einem speziellen Verfahren mit Hilfe von Masken direkt auf dem Trägermaterial synthetisiert wird (Fodor *et al.* 1991; Lipshutz *et al.* 1999). Dabei werden lichtempfindliche Schutzgruppen durch Licht aktiviert (Abbildung 8). Das Substrat, auf welchem sich diese befinden, wird mit der jeweiligen, ebenfalls lichtempfindlichen DNA-Base geflutet und diese bindet gemäß der vorab beleuchteten Arraystellen. Die Grenze dieser Oligosynthese ist durch Qualität und Ausbeute gesetzt. Die Vorteile jedoch liegen in der hohen Moleküldichte, die auf diese Weise erreicht wird (Pease *et al.* 1994).

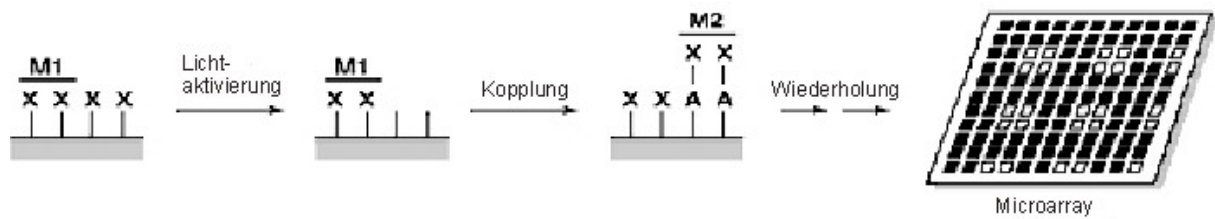


Abbildung 8: Schematische Darstellung des von Fodor *et al.* 1991 vorgestellten photolithographischen Verfahrens in Kombination mit der *in situ*-Synthese zur Herstellung von Oligonukleotidarrays. Die Belichtung erfolgt durch definierte Masken (M1-M2), die nur an exakt bestimmten Stellen das Licht durchlassen. Syntheseschritte erfolgen in Einzelnukleotidschritten vom 3'-Ende. Jedes neue Nukleotid (hier: A) enthält eine photolabile Schutzgruppe (x) und erst wenn diese entfernt ist, findet der nächste Syntheseschritt statt (Abbildung modifiziert nach Schena *et al.* 1998).

Die gängigere Methode zur Herstellung von Microarrays ist das so genannte *spotting* (Abbildung 9a und b). Hierbei werden die Moleküle vorsynthetisiert und mit Hilfe von Mikrodispensierungssnadeln im Kontakt (Abbildung 9a) oder im kontaktfreiem Verfahren (Abbildung 9b) in einer Pufferlösung auf das beschichtete Trägermaterial aufgetragen.

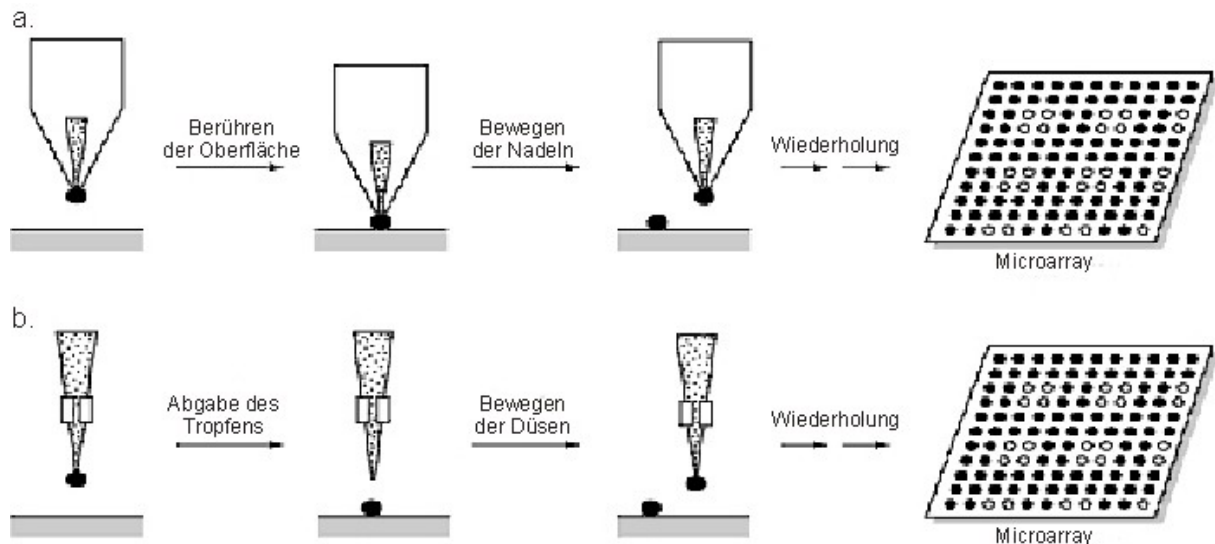


Abbildung 9: Schematische Darstellung von mechanischen *spotting*-Vorgängen a) Kontaktverfahren mittels Nadeln. Durch Kapillarkräfte wird die Probe in den Hohlraum der Nadel angesaugt und beim Absetzen der Nadel durch Berührung an der Objektträgeroberfläche abgesetzt. b) kontaktfreies Verfahren mittels Düsen. Die Probe wird in der Düse aufgenommen und über der Kontaktstelle durch einen Impuls an der Oberfläche abgesetzt (Abbildung modifiziert nach Schena *et al.* 1998).

Die Sonden sind in diesem Fall oft mit einem *spacer* versehen, einer gemeinsam synthetisierten, der eigentlichen Sequenz vorangehenden Molekulanordnung. Mit einer chemischen Struktur dient er der Kopplung auf Oberflächen und erleichtert durch einen sterischen Vorteil die Hybridisierung. Die Kopplung erfolgt durch eine chemische Bindung, die dann zustande kommt, wenn die Oberflächenbeschichtung, der darauf ausgerichtete Kopplungsmechanismus und der Puffer, in welchem die Sonden gelöst vorliegen, mit ihren chemischen Eigenschaften aufeinander abgestimmt sind.

2.2.4.2 SBE-TAG-Reaktion an Glasobjektträgern

Den Schwerpunkt der Arbeit bildet die Genotypisierung von biallelischen Polymorphismen basierend auf dem Microarray-Verfahren. Dazu werden sich die beiden beschriebenen Verfahren der Minisequenzierung und der Hybridisierung auf Glasobjektträgern zunutze gemacht. Zwei unterschiedliche Ansatzverfahren stehen hierzu dem Experimentator zur Verfügung, die je nach Zielsetzung gewählt werden können. Beide Methoden werden mittels DNA-Polymerasen katalysiert.

Shumaker *et al.* (1996) beschrieben die Methode der „*arrayed primer extension*“ (APEX), der Minisequenzierung an Festphasen und Pastinen *et al.* (1997) konnten für diese Methode eine um das Zehnfache erhöhte Effektivität und Spezifität gegenüber der allelspezifischen Hybridisierung nachweisen. Wie der Name bereits andeutet, werden hier Minisequenzierungsprimer als Sonden eingesetzt und auf einer Oberfläche immobilisiert. Die Bindung erfolgt über das 5'-Ende, so dass nach der Hybridisierung des PCR-Produkts mit dem gesuchten Polymorphismus unmittelbar an der polymorphen Stelle diese mittels fluoreszenz- oder radioaktiv markierter ddNTPs verlängert wird.

Mit dem im Folgenden ausführlich dargestellten Verfahren der SBE-*tag*-Reaktion wurden die Daten für diese Arbeit generiert. Eine entscheidende Veränderung führt hier zu der größeren Flexibilität des Ansatzes, denn sie hört auf SNP-spezifisch zu sein und erlaubt in der Theorie eine parallele SNP-Typisierung in einem sehr großen Umfang.

Hierbei findet die SBE-Reaktion nicht auf dem Array, sondern in getrennter Reaktion statt. Die Besonderheit an diesem Ansatz sind die bifunktionalen Primer, die zum einen die Sequenzspezifität eines SBE-Primers besitzen, zum anderen mit einer *tag*-Sequenz synthetisiert werden, die universell ist und dessen komplementäre Sequenz (*ctag*) in 3'-5'-Richtung an der beschichteten Glasobjektträgeroberfläche gekoppelt wurde (Hirschhorn *et al.* 2000, Fan *et al.* 2000). Durch den Hybridisierungsschritt erfolgt eine Sortierreaktion auf dem Array, welche im weiteren Verlauf der Arbeit als der Begriff des „Demultiplexen“ weitergeführt wird (Abbildung 10).

Die *tag*-Sequenzen sind so konzipiert, dass sie keine Sequenzhomologien aufweisen. Sie werden unter ähnlichen thermodynamischen Eigenschaften ausgewählt, so dass jede Sonde einen Nachweis ausschließlich über ihren komplementären Partner erweisen sollte. Der Nachweis der Fluoreszenzsignale ist SNP-spezifisch und erfolgt durch eine laserbasierte Scananalyse über einen Microarray-Scanner (Kapitel 3.4.6.3).

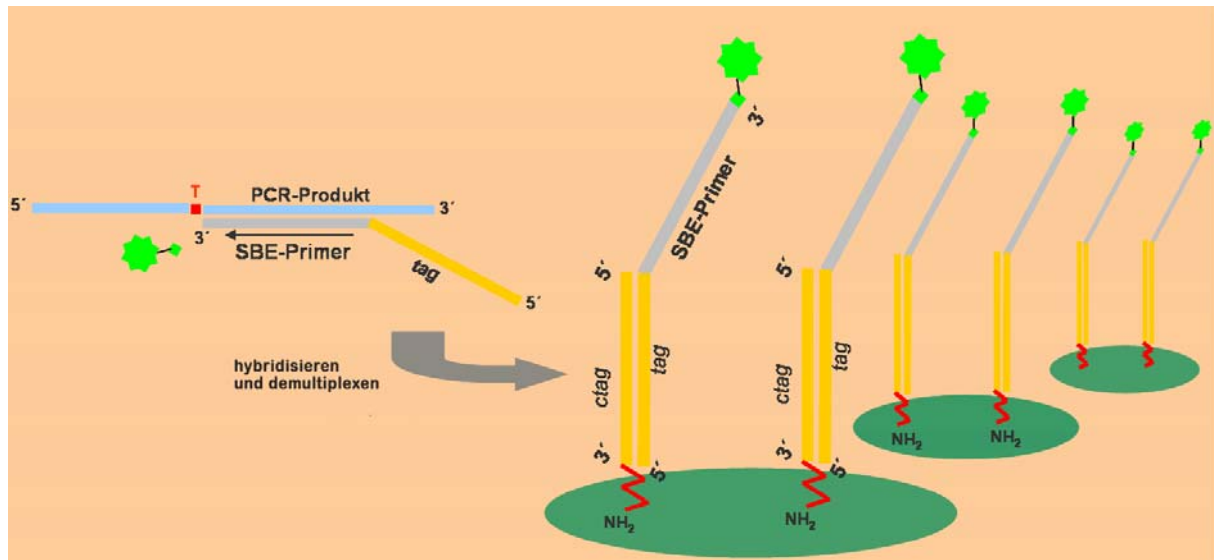


Abbildung 10: Prinzip der SBE-Reaktion in Flüssigkeit und anschließendes hybridisierungsbasiertes Sortieren (Demultiplexen) der SBE-Produkte an die komplementären tags (ctags) auf dem Glasobjektträger.

2.2.4.3 „Array of arrays“- Prinzip

Das Ziel, den Durchsatz der Proben maximal zu erhöhen, konnte durch die Entwicklung der Microarray-Technologie weit vorangetrieben werden. Bei DNA-Microarrays kann die Anzahl der Spots pro Chip bei einer Spotdichte von ca. 200 μm bis zu 20 000 Spots betragen. Im Hinblick auf die parallele Genotypisierung möglichst großer Individuenzahlen auf Glasobjektträgern wurde eine denkbar einfache technische Lösung etabliert (Pastinen *et al.* 2000). Das Format des „array of arrays“ besteht aus einer Zwischenlage aus Silikon, die auf dem Objektträger aufliegt und eine Maske darstellt, welche die Glasoberfläche in Kompartimente unterteilt (Abbildung 11).

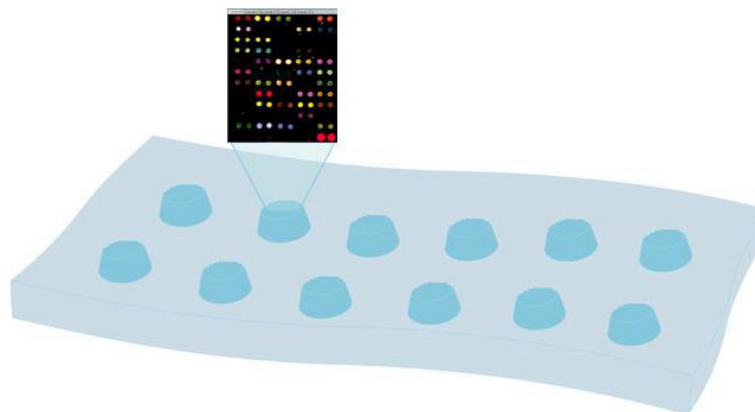


Abbildung 11: Genotypisierungsmöglichkeit mittels „array of arrays“-Formats. Zwischen Glasobjektträger und Deckel der Hybridisierungskammer (Abb im Kapitel 3.4.3.2) wird eine Zwischenlage aus Silikon platziert, die das gleiche Format der Spotanordnung besitzt. Pro Array kann dadurch ein Individuum typisiert werden.

Die Anordnung der Öffnungen orientiert sich an ihrer Vorlage: den beiden gängigen Formaten der Mikrotiterplatten (96-well oder 384-well). Auf dem Glasobjektträger befindet

sich eine vergleichbare Anordnung der einzelnen Arrays und im Verschlussdeckel der Hybridisierungskammer ebenfalls eine solche Anordnung der Öffnungen. Verschließt man den Deckel, entstehen kleine Einzelreaktionskammern, in welchen jeweils ein Individuum analysiert werden kann.

Dies ermöglicht das gleichzeitige Typisieren von bis zu 80 Individuen. Dabei spielt die Anzahl der zu untersuchenden Marker bzw. der zu untersuchenden Individuen eine Rolle. Benutzt man hierzu das Format der 384-well Mikrotiterplatte, so kann man pro *subarray* bei einem Abstand von 200 µm (zwischen den Zentren zweier Spots) 182 Spots in einer Ausrichtung von 14 x 13 spotten. Soll eine größere Anzahl von Markern pro Individuum analysiert werden, kann man die Größe der *subarrays* dem Durchmesser eines *well*s der 96-well Mikrotiterplatte anpassen. In diesem Fall werden 14 x 576 Spots einer gleichzeitigen Analyse unterzogen (Pastinen *et al.* 2000, Liljedahl *et al.* 2004).

2.2.5 Andere enzymkatalysierte Genotypisierungsverfahren auf festen Trägermaterialien

Neben DNA-Polymerasen können ligasevermittelte Typisierungsreaktionen mit hoher Genauigkeit auf Objektträgern durchgeführt werden (Gunderson *et al.* 1998). Dazu werden 20 *nt* lange unspezifische Ankersequenzen mit dem 3'-Ende an das Array gekoppelt. Diese sind mit einer spezifischen Sequenz versehen. Im nächsten Schritt findet die Hybridisierung einer komplementären Ankersequenz mit einem freien 3'-Ende an ihre Zielsequenz statt und eine markierte DNA-Sonde kann in einem Ligationsschritt sequenzspezifisch angelagert werden.

Eine effiziente und stark multiplexfähige Methode stellt auch der Ansatz der SNP-Detektion mit Hilfe einer reversen Transkriptase dar (Pastinen *et al.* 2000). Aus dieser Arbeit ist auch das im oberen Abschnitt beschriebene Prinzip des "array of arrays" hervorgegangen. Der Ansatz ist stark in den Arbeitsschritten reduziert. Nach einer Multiplex-PCR folgt ohne einen zusätzlichen Aufreinigungsschritt die *in vitro*-Transkription und die allelspezifische Anlagerung des fluoreszenzmarkierten Nukleotids ausschließlich bei einer vollständigen Komplementarität.

2.3 SNPforID-Projekt (EU-Projekt)

Das SNPforID-Projekt wurde von vier europäischen Arbeitsgruppen aus dem Bereich der forensischen Genetik beantragt und durchgeführt. Das der Europäischen Union obliegende *EU Growth* Programm hat das Vorhaben über einen Zeitraum von drei Jahren unterstützt.

Im Vordergrund der gemeinsamen Arbeit der vier kooperierenden Institute stand die Idee, die Anwendbarkeit von SNPs zur Identifizierung biologischer Spuren im forensischen Zusammenhang zu etablieren und unterschiedliche SNP-Analysemethoden auf ihre Eignung in diesen Bereich zu prüfen. Zusätzliches Bestreben war es, biallelische Marker des Y-chromosomalen und des mitochondrialen Genoms für phylogenetische Studien zu validieren. Die Ziele des Projekts ergaben sich aus den Anforderungen an die schwierige Arbeit mit Spurenmaterial - schwierig in Bezug auf die Handhabung durch die begrenzte Menge und geringe Qualität der Spuren-DNA sowie die Analyse und Interpretation der Ergebnisse.

Es sollte eine Zusammenstellung von diskriminatorischen SNPs entwickelt werden, welche in ihrer Unterscheidungskraft den bisher angewandten STR-Identifikationskits gleich kommt. Da eine kleine Auswahl biallelischer Marker keine biostatistische Effizienz, vergleichbar mit der von 15 STR-Systemen besitzen kann, muss die Anzahl von mindestens 50 SNPs typisiert werden, um eine sichere Identifizierung einer Person unbekannter Herkunft zu erreichen (Gill, 2001). Aus den ständig wachsenden SNP Datenbanken galt es, eine entsprechende Auswahl zu treffen.

Die weiteren Schwerpunkte des Projekts lagen in der Etablierung von effizienten Multiplex-Amplifikationsmöglichkeiten sowie den Typisierungsmethoden für die selektierten SNP-Panels. Diese Schritte bedurften einiger strikter Vorgaben an die Selektion und Durchführbarkeit sowie an die forensische Validierung.

Eines der Ziele des Projekts war eine parallele Typisierung einer großen Anzahl von Markern. Hierfür musste gewährleistet werden, dass diese nicht codierend sind, in keinem Kopplungsgrad zu bekannten und in der forensischen Genetik bereits etablierten Mikrosatelliten sowie Genen stehen und sich in einem ausreichend großen Abstand zueinander befinden. Die Zielsetzungen des methodischen Teils des Projekts wurden mit der Zuverlässigkeit und Reproduzierbarkeit der Ergebnisse festgelegt. In zweiter Linie galten die Vorgaben der Kosteneffizienz und der Höhe des Durchsatzes. Wünschenswert war die Gewährleistung der Durchführbarkeit der entwickelten Vergleichsmethoden in den Routinelabors. Das erforderte die Analyse an Geräten, die zur Standardtypisierung in allen beteiligten Laboren bereits vorhanden waren.

Hierzu wurde die Projektentwicklung in vier Themenbereiche aufgeteilt:

- Teilprojekt 1: SNP-Auswahl
- Teilprojekt 2: Entwicklung der Multiplexmethoden
- Teilprojekt 3: Entwicklung der Detektionsmethoden
- Teilprojekt 4: Forensische Validierung

Eine kurze Beschreibung der einzelnen Arbeitsthemen soll einen Überblick über den zeitlichen und inhaltlichen Verlauf des Projekts geben.

2.3.1 Arbeitsthema 1: SNP-Auswahl

Im Laufe des Projekts wurden vier unterschiedliche SNP-Sets festgelegt, auf denen der Schwerpunkt der Analysen liegen sollte. Diese beinhalteten die **a)** autosomalen, diskriminatorischen SNPs, **b)** autosomalen, populationsspezifischen SNPs, **c)** Y-chromosomale SNPs sowie **d)** mitochondriale SNPs.

Zusammen mit der angegebenen Auswahl wurden Kriterien festgelegt, die dem spezifischen forensischen Ansatz der anstehenden Analysen entsprachen. Dabei wurde auf die Anforderung der besonderen Sensitivität der Loci Wert gelegt. Diese wurden nicht nur unter optimalen analytischen Bedingungen untersucht, sondern in Reaktionen mit DNA von schlechter, degradierter Qualität sowie geringer Konzentration. Diesen Kriterien gegenüber wurde die Anforderung an die Höhe der Variabilität des einzelnen Markers herabgestuft.

Grundsätzliche Überlegung zur Auswahl forensischer Marker war es, eine adäquate Anzahl von SNPs aus nichtcodierenden Genomregionen zu selektieren, welche jedoch die notwendige Variabilität besitzen, den Grad der Leistungsfähigkeit der Unterscheidung zwischen zwei Personen maximal zu erhöhen. Der Grad der Heterozygotie ist hierfür ausschlaggebend. Binäre Marker mit einer Allelfrequenz von 0,5:0,5 entsprechen dabei der bestmöglichen Differenzierung. Der Gradmesser der forensischen Aussagekraft ist der Diskriminierungsindex [DI] (Jones, 1972). Er ist abhängig von der Allelfrequenz und variiert am geringsten bei Allelfrequenzen zwischen 0,5 und 0,3. Diejenigen Marker, deren Heterozygotiegrad sich innerhalb dieser Spannbreite bewegte, fielen in diese Auswahl.

Nach statistischen Parametern sind mindestens 40 binäre Marker notwendig, um zwei zufällig aus einer Population ausgewählte Personen zu identifizieren, wenn diese nicht verwandt sind. Die diskriminatorische Kraft muss jedoch weiter erhöht werden, will man erstgradig verwandte Personen oder Individuen aus heterogenen Populationen analysieren. Die höhere diskriminatorische Kraft ist ebenfalls notwendig, um mögliche Mischspuren oder Kontaminationen aufzudecken. Hierfür muss auch die Anzahl der SNPs ansteigen. Um die gleiche biostatistische Aussagekraft von 15 STRs (wie sie in kommerziellen Multiplexkits

kombiniert werden³⁾ zu erreichen, müssen also ca. 50 gut balancierte Einzelpolymorphismen parallel typisiert werden. Erst dann wird die maximale Ausschlusswahrscheinlichkeit (*probability of exclusion*) von 0,99997 erreicht (Gill, 2001).

Tabelle 2: Ausschlusswahrscheinlichkeiten beim Vergleich des Mutter/Kind-Genotyps (aus Gill, 2001).

Frequency allele A	No. of loci			
	1	10	20	50
0.1	0.08190	0.57450	0.81895	0.98605
0.2	0.13440	0.76386	0.94424	0.99927
0.3	0.16590	0.83700	0.97343	0.99988
0.4	0.18240	0.86652	0.98218	0.99996
0.5	0.18750	0.87462	0.98428	0.99997
0.6	0.18240	0.86652	0.98218	0.99996
0.7	0.16590	0.83700	0.97343	0.99988
0.8	0.13440	0.76386	0.94424	0.99927
0.9	0.08190	0.57450	0.81895	0.98605

Ein weiteres, sehr wichtiges Kriterium zur Auswahl der forensischen Marker ist die Abwesenheit jeglicher Kopplung zwischen den untersuchten Polymorphismen. Das Vermeiden der Kopplung stellt jedoch mit der steigenden Anzahl der Marker eine größere Herausforderung dar, da mehrere Marker auf gleichen Chromosomen ausgewählt werden müssen.

Anders als bei krankheitsassoziierten Studien, bei denen die Kopplung kausaler Gene mit benachbarten Genorten und deren gemeinsame Vererbung für die Analyse eine Rolle spielen, muss in der forensischen DNA-Analyse eine gekoppelte Weitergabe der untersuchten Informationen ausgeschlossen werden. Es entspricht der Unabhängigkeit aller Allele auf Populationsebene innerhalb unterschiedlicher Positionen auf dem Genom (Li und Stephens, 2003). Ein forensisches Untersuchungsprofil erfordert die Unabhängigkeit der Marker, um eine kombinierte Genotyp-Frequenz nach der Produktregel berechnen zu können. Es erfüllt so die rechtlichen Kriterien mehrerer europäischer Staaten, nach welchen durch forensische Marker keine Rückschlüsse auf die phänotypischen Merkmale erzielt werden dürfen (Ausnahme: Bestimmung des Geschlechts).

Das letzte Kriterium zielte auf die Vielfalt der Detektionsmethoden ab mit denen die Marker analysiert werden konnten. Um das Spektrum der technischen Anwendungen maximal zu erhöhen, musste gewährleistet werden, dass die Sequenzqualität die Untersuchungen an weiteren Analyseplattformen zulassen wird. Der Anspruch an die flankierende Sequenz war, dass diese sowohl für die PCR-Primer als auch für die Minisequenzierungsprimer eine gute Zielsequenz darstellt. Zu den Selektionsfaktoren zählten hierbei der Gehalt der GC-

³⁾ z.B.: *SE-Filter Plus Kit* (ABI); *Identifiler Kit* (ABI), *Powerplex 16 Kit* (Promega). Beim *Powerplex 16 Kit* werden folgende 15 Genorte analysiert: Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA, Penta D, CSF1PO, D16S539, D7S820, D13S317 und D5S818. Beim sechzehnten Marker handelt es sich um den Amelogenin-Marker, der zur Geschlechtsbestimmung dient.

Nukleotide sowie das Fehlen repetitiver Sequenzen in den benachbarten Bereichen der SNPs.

2.3.2 Arbeitsthema 2: Entwicklung der Multiplexmethoden

Die erstmalige Beschreibung der PCR-Methode durch Kary Mullis' Mitarbeiter im Jahre 1985 (Saiki *et al.* 1985, Mullis *et al.* 1986) hat auf dem Gebiet der molekularen Genetik zu zahlreichen neuen Anwendungsmöglichkeiten im Bereich der Labortechnik geführt. Im Bereich der forensischen Genetik kam 1990 das erste PCR-Kit zur Identifizierung von Menschen auf dem Markt. Dieses basierte auf der Typisierung des HLA-DQ Alpha-Gens (Erich *et al.* 1990). Im Bereich der Spurenkunde findet seit ca. 1992 die Analyse von STRs zur Identifikationszwecken statt (Kapitel 2.1.2). Aus den ca. 30 forensisch validierten STR-Systemen werden heutzutage bis zu 15 autosomale STR-Loci simultan mittels Multiplexkits analysiert (siehe Fußnote ³).

Wie bereits in den Kapiteln 2.1.3 und 2.3.1 beschrieben, musste zur Steigerung der biostatistischen Effizienz die Anzahl der SNP-Marker innerhalb einer Multiplexreaktion angepasst werden. Dennoch lag die Anforderung der Multiplexentwicklung nicht ausschließlich in der Kombination einer entsprechenden Anzahl von Markern in der PCR-Reaktion. Mit der Einhaltung der vorab formulierten Kriterien wurden in mehreren Schritten auf beiden Armen der autosomalen Chromosomen die Marker selektiert und zunächst zwei Multiplexreaktionen gebildet, die im Verlauf des Projekts in einem einzigen Reaktionsansatz amplifiziert werden konnten. Für die Y-chromosomale Multiplexreaktion galten ebenfalls die oben genannten Kriterien der Markerselektion. Es konnte eine Multiplexreaktion entwickelt werden, die von Anfang an in einer Amplifikationsreaktion durchgeführt wurde (Bríón *et al.* 2005). Andere Anforderungen wurden wiederum an die Entwicklung der mitochondrialen Multiplexreaktion gestellt. Aufgrund der geringen Flexibilität bei der Markerwahl und der hierdurch eingeschränkten Auswahl von Zielsequenzen für die Minisequenzierungsreaktion musste die Analyse entsprechend angepasst werden (Brandstätter *et al.* 2006). Alle Multiplexreaktionen besaßen ähnliche Kriterien bezüglich der Reaktionstringenz und der erwünschten Kürze der Amplikons.

Wie an die Entwicklung der PCR-Multiplexreaktionen, wurden auch entsprechende Anforderungen an die Evaluierung der generellen Typisierungsreaktion mittels der SBE-Multiplexreaktion gestellt. Die Herausforderungen lagen hierbei in der Wahl der Minisequenzierungsprimer bezüglich ihrer Länge und der *annealing*-Temperatur und anschließend in der Anpassung der Reaktionskonzentration. Das gewünschte Ergebnis war ein Elektropherogramm mit ausbalancierten Signalen, die sowohl in kürzeren als auch in längeren Analysebereichen möglichst homogene Signalstärken aufwiesen und die eindeutige Auftrennung der Allele darboten.

2.3.3 Arbeitsthema 3: Entwicklung der Detektionsmethoden

Wie im Kapitel 2.2 dargestellt, wurden im Verlauf der Jahre unterschiedliche Plattformen entwickelt, auf welchen eine SNP-Genotypisierung stattfinden kann.

Die Standardmethode (Minisequenzierungsreaktion mit dem *SNaPshot*TM-Kit) wurde für alle weiteren Systeme als Vergleichsmethode genutzt: zum einen wegen ihrer hohen Multiplexfähigkeit und zum anderen wegen der zuverlässigen, automatisierten Typisierung und eindeutigen Interpretation der Ergebnisse.

Die MALDI-TOF Massenspektrometrie-Methode (*Matrix-Assisted Laser Desorption/Ionisation – Time of Flight*) erwies sich ebenfalls als multiplexfähig und wurde für eine Reihe der Multiplexreaktionen zu einer viel versprechenden Typisierungsmethode. Die Probenherstellung basiert auf der SBE-Reaktion, in der durch die Kopplung des spezifischen Nukleotids an den Primer dieser seine spezifische Masse annimmt. Für die Detektion erfolgt zunächst eine Überführung der Produkte in die Ionisierungsphase mittels eines Stickstofflasers. Anschließend findet die Messung der Molekülmasse in Flugzeitmassenspektrometern statt.

Zwei microarraybasierte Detektionsmethoden wurden im Mainzer Laboratorium evaluiert. Die *NanoChip*[®]-Plattform (Nanogen Inc.) basiert auf elektronisch adressierten Microarrays, auf welche die PCR-Produkte direkt appliziert werden können. Die PCR-Produkte können dabei entweder direkt am Array immobilisiert werden (*amplicon down strategy*, Børsting *et al.* 2004, 2005), oder es wird eine so genannte *capture*-Probe auf dem Chip appliziert (*capture down strategy*). Diese bindet das PCR-Produkt als Zielsequenz und ist multiplexfähig (Dissertation Balogh, 2009).

Die andere microarraybasierte Methode beruht auf der Detektion von SBE-Produkten auf einem konventionellen Glasobjektträger. Diese Methode ist Thema dieser Arbeit und wird im weiteren Verlauf ausführlich besprochen. Als zweite Referenzmethode wurde neben der *SNaPshot*TM-Detektion das ligationsbasierte und auf dem *SNPlex*TM-System basierende *Genplex*-Verfahren, welches im Kapitel 2.2.3 kurz dargestellt wird, genutzt.

Zusätzlich wurde die Detektion der selektierten SNPs auf weiteren Plattformen erprobt. In einer frühen Phase des Projekts wurde im Mainzer Labor die auf Chemilumineszenz basierte Technologie der Pyrosequenzierung erfolgreich an 22 der anfänglich selektierten SNPs getestet. Diese wurden zunächst in Einzel-PCRs, anschließend in Triplexen und 6plexen typisiert (Bender *et al.* 2006). Anschließend wurde die Methode anhand einer weiterentwickelten Plattform im Londoner Laboratorium untersucht (Harrison *et al.* 2006).

Mit Hilfe des Unternehmens Orchid Cellmark, wurde im Rahmen des Projekts eine quantitative Methode entwickelt, die sich einer seltenen Untergruppe von SNPs behilft. Diese sind unter dem Begriff der *SWaP*TM SNPs (*S/W amid Palindromes*) zusammengefasst. Die G/C-, bzw. A/T-Transversionen, für die im Versuchsansatz die Buchstaben S oder W stehen,

befinden sich innerhalb eines Sequenzpalindroms. Die Analyse dieser Polymorphismen kann bei der Interpretation von Mischspuren behilflich sein (Musgrave-Brown *et al.* 2006).

2.3.4 Arbeitsthema 4: Forensische Validierung

Ziel der Entwicklung der 52-SNP-Multiplexmethode war es, neben einer geeigneten Markerauswahl und der Entwicklung und Etablierung der Multiplexmethode, auch eine Evaluierung und Validierung des Verfahrens auf forensischer Ebene durchzuführen. Gerade im Hinblick auf eine potentielle methodische Ergänzung der STR-Typisierungssysteme musste gezeigt werden, dass das entwickelte Verfahren den Anforderungen in der Spurenanalytik gerecht wird. Untersucht wurde dabei, ob die entwickelte Multiplexmethode genetisches Material von geringer Quantität, stark degradiertes Spurenmaterial sowie Mischspuren von mehr als einer Person, verlässlich zu typisieren vermag.

Zu diesem Zweck wurde Spurenmaterial unter kontrollierten Bedingungen hergestellt, DNA aus etablierten Zelllinien in Verdünnungsreihen angesetzt sowie Mischspuren mit bekannter Zusammensetzung gebildet und innerhalb aller teilnehmenden Institute unter gleichen Bedingungen untersucht. Als Vergleichsstandard wurde das etablierte *Powerplex®16* (Promega) – STR-Typisierungssystem genutzt. Alle Ergebnisse der forensischen Validierung der 52-SNP Multiplexreaktion auf der Basis der *SNaPshot™* Referenzmethode sind nachzulesen bei Musgrave-Brown *et al.* (2007).

2.4 Zielsetzungen der Arbeit und des Projekts

Die Untersuchungsmöglichkeiten in der forensischen Genetik sind mit der Etablierung der Typisierung humaner Spuren mittels STRs noch nicht vollständig erschöpft. Zwar bieten die Mikrosatelliten in der spurenkundlichen Routine die höchste biostatistische Aussagekraft bezüglich der Übereinstimmung der Merkmale, ihre Grenzen erreichen diese genetischen Werkzeuge jedoch bei stark degradiertem Untersuchungsmaterial und bei niedrigen Ausgangskonzentrationen. Hier eröffnet sich für die Forschung die Möglichkeit, die forensische DNA-Analyse mit Hilfe von Einzelbasenpolymorphismen zu unterstützen.

In der niedrigen Mutationsrate sowie der größeren Stabilität gegenüber Mechanismen, die degradierend auf das genetische Material wirken können, liegen die Vorteile der SNPs. Ihr Nachteil und gleichzeitig die Herausforderung des gesamten Projekts und der daraus resultierenden Dissertation, ist die geringere biostatistische Leistungsfähigkeit, die sie aufgrund ihrer biallelischen Form aufweisen. Diese ist, gegenüber der von Mikrosatelliten, um das Fünffache reduziert und muss durch eine höhere Anzahl von Systemen in einer Reaktion ausgeglichen werden. In der Praxis müssen die für eine SNP-Analyse entwickelten Verfahren die niedrigere diskriminatorische Kraft durch eine erhöhte Anzahl der parallel

untersuchten Marker aufwiegen. STR-Untersuchungen werden standardmäßig in zusammengefassten, so genannten Multiplexverfahren durchgeführt. Die Menge der parallel analysierbaren Marker umfasst bei kommerziellen STR-Kits zwischen drei und 17⁴. Obwohl SNPs den entscheidenden Vorteil einer kürzeren Amplikonlänge in der PCR-Reaktion meistens erfüllen können, sind im weiteren Analysenverlauf limitierende Faktoren möglich, die meistens technischen oder kostenbedingten Ursprungs sind. In der Konzeption und der Durchführung der Arbeit müssen diese Faktoren berücksichtigt werden.

Die hier beschriebenen Kriterien sollen in der Entwicklung einer Referenzmethode, in der mehr als 50 SNPs parallel untersucht werden können, umgesetzt werden. Projektübergreifend soll ein validiertes Standardverfahren etabliert werden, welches in jedem forensischen Labor mit vergleichbarem Zeitaufwand und einer höchst möglichen biostatistischen Zuverlässigkeit durchführbar sein soll.

Als Hauptziel der vorliegenden Dissertation steht die Anpassung des Standardverfahrens auf eine microarraybasierte Hybridisierungsanalyse auf Glasobjektträgern im Mittelpunkt. Die Vorteile der Unabhängigkeit von Fragmentlängen der Minisequenzierungsprodukte sowie die durch die Anwendung des „array of arrays“-Verfahrens erzielte Parallelität, sollen vergleichend dem Standardverfahren (*SNaPshot*[™]) und einer weiteren Referenzmethode (*Genplex*) gegenübergestellt und untersucht werden.

⁴ Kommerzielle Kits, wie das *PowerPlex*[®] *ESX 17* und *PowerPlex*[®] *ESI 17* (Promega) mit zusätzlichen von ENFSI und EDNAP empfohlenen Systemen D10S1248, D22S1045, D2S441, D12S391 und D1S1656 zusätzlich zu den innerhalb von Europa untersuchten elf Systemen D8S1179, D18S51, D21S11, FGA, TH01, vWA, D2S1338, D3S1358, D16S539, D19S433 und Amelogenin.

3 Material und Methoden

3.1 Vorbereitende Verfahren zur Probenanalyse mittels Multiplexreaktionen

Aufgrund der allgemeinen Gültigkeit der folgenden Parameter für alle verwendeten Typisierungsplattformen wurden die Kriterien der SNP-Auswahl und des PCR-Primer Designs unabhängig beschrieben und beziehen sich auf alle durchgeführten Experimente.

Weiterführende Kriterien wie Sondendesign und die Auswertung der Proben werden in den jeweiligen Kapiteln aufgeführt. Bei Kriterien, die im Zuge der Zusammenarbeit in Partnerlaboratorien entwickelt wurden, wird der Kontext kurz beschrieben und ggf. auf die jeweilige Publikation verwiesen.

3.1.1 SNP-Selektion

Alle durchgeführten Analysemethoden basierten auf einer SNP-Auswahl, die im Rahmen des SNPforID-Projekts gemeinsam festgelegt wurde. Die Kriterien der SNP-Auswahl wurden durch das Projekt-*Consortium* bestimmt und umgesetzt (Phillips *et al.* 2004c). Zusammengefasst sind diese Kriterien: a) Die Größe des Amplikons sollte im Hinblick auf die forensische Ausrichtung des Projekts unter 120 *bp* liegen. b) Die Heterozygotenrate eines Markers sollte bei mindestens einer der Populationen wenigstens bei 30% liegen (niedrigste Allelfrequenz bei 0,28) und bei den weiteren drei Populationen eine Heterozygotenrate von mindestens 20% aufweisen (niedrigste Allelfrequenz 0,17). c) Die Zusammenstellung eines SNP-Sets sollte aus Markern von distalen Enden der *p*- und *q*-Arme der autosomalen Chromosomen erfolgen. d) Eine Mindestentfernung von 100 *kb* zum benachbarten SNP oder einem Gen sollte in beiden Leserichtungen (*downstream* und *upstream*) eingehalten werden. e) Es durfte keine Kopplung mit bekannten STR-Systemen vorliegen, die zur forensischen Analyse verwendet werden. Zusätzlich war es wichtig, dass die flankierende Sequenz des Markers verbindlich publiziert wurde und dass sich innerhalb dieser und speziell in den Primerbindungsstellen keine weiteren Polymorphismen, sowie keine Polybasen befanden. Die Auswahl der Marker-SNPs erfolgte aus den folgenden kommerziellen und öffentlich zugänglichen Datenbanken:

a) *Celera Discovery Systems Database* (CDS) ⁵

b) *Applied Biosystems Assay-on-Demand* (ABI) ⁶

c) *NCBI dbSNP Database* ⁷

⁵ www.celeradiscoverysystem.com

⁶ www.appliedbiosystems.com/products/assay

- d) *The SNP Consortium (TSC) Database*^{8 9}
- e) *HapMap project data*¹⁰
- f) *SNP Browser*^{TM 11}

(Alle Richtlinien zur Selektion der Marker *in silico* finden sich bei Phillips *et al.* 2004c).

Es erfolgte eine Prüfung der flankierenden Regionen auf benachbarte SNPs oder einer Sequenz mit geringer Komplexität, um Probleme beim Primerdesign zu vermeiden. Aus einem „pool“ von ca. 400 Kandidaten-SNPs, welche in kleineren Gruppen mit vergleichbarer Variabilität und Position auf dem entsprechenden Chromosom gesammelt wurden, wurde in mehreren Analyseschritten ein Set von 52 SNPs ausgewählt (Sanchez *et al.* 2006).

3.1.2 PCR-Primerdesign

PCR-Primer wurden aus der den Marker umgebenden Sequenz so gewählt, dass die PCR-Produkte eine Amplikongröße zwischen 59 und 115 *bp* aufwiesen. Sie erfüllten die festgelegten Kriterien der Länge, Schmelztemperatur bei der erforderlichen Salzkonzentration und der Basenausgewogenheit sowie der Vermeidung von Sekundärstrukturen. In seltenen Fällen wurden benachbarte SNPs innerhalb der flankierenden Region zugelassen. Dieser Sequenzabschnitt wurde bei der Primerwahl jedoch ausgelassen (Sanchez *et al.* 2006). Die PCR-Primer wurden unverändert für die Microarray-Methode übernommen. Eine Liste aller Primer und ihrer in der Multiplexreaktion verwendeten Konzentrationen befindet sich im Anhang (Table S1 zu Sanchez *et al.* 2006).

3.1.3 Multiplex-PCR-Reaktion

In der vorliegenden Arbeit wurde für alle beschriebenen Verfahren eine modifizierte PCR-Reaktion angewendet – die so genannte Multiplex-PCR. In dieser wurden mehrere Primerpaare eingesetzt, so dass mehrere kurze Amplikons nach der Reaktion für weitere Analysen zur Verfügung standen. Eine bestmögliche Ausgewogenheit der Menge der PCR-Produkte konnte durch die Ausbalancierung der Primerkonzentrationen erreicht werden.

In Tabelle 3 wird der Ansatz einer Multiplex-PCR in 25 μ l Volumen beschrieben. Dieser Ansatz bezieht sich auf die in der vorliegenden Arbeit erstellten 23-, 52- und 33-SNP Multiplexreaktionen.

⁷ www.ncbi.nlm.nih.gov/entrez

⁸ www.snp.cshl.org

⁹ Die in der vorangegangenen Fußnote angegebene Webseite leitet zwischenzeitlich auf die Webseite des HapMap Projekts um. Die Datenbankeinträge des Consortiums sind unter den angegebenen Nummern des TSC nicht mehr aufrufbar.

¹⁰ www.hapmap.org

¹¹ www.allsnps.com/snpbrowser

Tabelle 3: Zusammenstellung aller Reagenzien und ihrer Konzentrationen sowie die Einsatzmenge der Reagenzien in einer Multiplex-PCR.

<i>Reagenzien</i>	<i>Konzentration</i>	<i>Endkonzentration</i>	<i>25 µl Ansatz</i>
PCR-Reaktionspuffer [X]	10 x	1x	2,5 µl
MgCl ₂ [µM]	25 µM	6,5 µM	6,5 µl
dNTPs [µM]	10 µM	0,7 µM	1,75 µl
Primermix (23plex) [µM]	25 µM	0,1 µM	6,5 µl
Primermix (29plex) [µM]	25 µM	0,1 µM	6,5 µl
Taq-Polymerase [U/µl]	5 U	2 U	0,4 µl
Genomische DNA [ng/µl]	20 ng	0,5 ng	0,6 µl
HPLC-H ₂ O			0,25 µl
Summe Reagenzien			25 µl

Das Temperaturprofil für die PCR-Reaktion war: 95°C für 10 Min. (einmalige Denaturierung), 95°C für 30 Sek., 35 Zyklen bei 60°C für 45 Sek., 65°C für 30 Sek. und 65°C für 6 Min.

3.1.4 Enzymatische Probenaufreinigung

Die PCR-Produkte wurden vor der SBE-Reaktion enzymatisch aufgereinigt (Tabelle 4). Eine Exonuclease (ExoI) baut dabei die überschüssigen PCR-Primer ab und die nicht inkorporierten dNTPs werden mittels einer alkalischen Phosphatase (SAP) dephosphoryliert, wodurch sie keiner weiteren Verlängerungsreaktion im folgenden SBE-Ansatz zur Verfügung stehen können.

Tabelle 4: Zusammenstellung der Reagenzien und Angaben der Konzentrationen für den enzymatischen Verdau.

<i>Reagenzien</i>	<i>Konzentration</i>	<i>Endkonzentration</i>	<i>21 µl Ansatz</i>
SAP Puffer [X]	10 x	1,07 x	2,24 µl
ExoI [U/µl]	10 U	4 U	0,4 µl
SAP [U/µl]	1 U	3 U	3 µl
HPLC-H ₂ O			0,36 µl
PCR-Produkte			15 µl
Summe Reagenzien			6 µl

Das Temperaturprofil für den enzymatischen Verdau war: 37°C für 60 Min. und 75°C für 15 Min.

3.2 Vorvalidierung der 23 SNP Multiplexreaktion mittels Minisequenzierung

3.2.1 DNA-Proben

Zunächst fand eine Etablierung der ersten Multiplex-PCR-Reaktion (23plex) innerhalb des SNPforID-Projekts statt. Um diese durchzuführen wurden Blutproben (neun humane Proben und eine canine Probe als Kontrolle) auf FTA[®]-Papier (Whatman) aufgetragen. Aus dem Labor der Rechtsmedizin in Kopenhagen wurden diese Proben sowie PCR- und Analyseprimer an die Partnerlabore verschickt und in jedem Labor doppelt analysiert. Weitere Blutproben von fünf männlichen und weiblichen Probanden wurden im Mainzer Labor entnommen, extrahiert und quantifiziert. Die Extraktion und Aufreinigung erfolgte mittels PEQLAB E.Z.N.A.[®] *Blood DNA Kit II*.

Im Anschluss an die Analyse der Proben im Mainzer Labor wurden diese DNA-Extrakte sowie deren Amplifikate und die Minisequenzierungsreaktionen, sowie die Minisequenzierungsreaktionen der FTA-Proben nach Kopenhagen versandt und dort nochmals typisiert. Nach der erfolgreichen Etablierung der Multiplexreaktion wurden in jedem Labor Proben aus den Hauskollektiven typisiert. In Mainz erfolgte die Typisierung in der Gesamtheit von n=146 asiatischen Proben, davon 40 japanischen Proben, 57 chinesischen Proben und 49 Proben aus Thailand (Waiyawuth *et al.* 1998). Zusätzlich wurden 64 Proben aus deutschen Vaterschaftstrios typisiert. Die Menge der in die PCR eingesetzten genomischen DNA betrug 6-24 ng.

Folgende DNA-Mengen wurden in den extrahierten Proben aus Mainz gemessen:

- LM1f (~20 ng/µl)
- LM2m (~35 ng/µl)
- LM3f (~33 ng/µl)
- LM4f (~40 ng/µl)
- LM5f (~23 ng/µl)

Die Quantifizierung erfolgte mittels Bio-Rad *VersaFluor*[™] Fluorometer 2 µl DNA in 2 ml H33258 Hoechst (1 mg/ml).

Anschließend wurden die Extrakte auf eine einheitliche Menge von 20 ng/µl eingestellt und in dieser Verdünnung in die PCR eingesetzt.

3.2.2 Probenaufbereitung mittels Whole Genome Amplification (WGA)

Von den 40 Proben aus dem japanischen Kollektiv wurden zehn vorab mit dem Verfahren der *Whole Genome Amplification* amplifiziert (Dean *et al.* 2002, Hosono *et al.* 2003). Dabei handelt es sich um ein Verfahren zur unspezifischen Vervielfältigung genomischer DNA.

Die 95 DNA Proben japanischen Ursprungs aus dem Hauskollektiv waren zum größten Teil eingetrocknet und wurden anfänglich in 20 µl Wasser aufgenommen und über Nacht gelöst. Anschließend wurden die DNA-Mengen gemessen. Sie lagen zwischen 0 und 150 ng/µl.

40 DNA-Proben, bei welchen DNA-Konzentrationen von weniger als 14 ng/µl gemessen wurden, wurden ausgewählt. Zehn dieser Proben wurden zur weiteren Analyse mit dem *GenomiPhi™ DNA Amplification Kit* (GE Healthcare Life Sciences, früher Amersham Biosciences, Freiburg) bearbeitet. Zur Erzeugung einzelsträngiger DNA wurde zunächst 1 µl des Ausgangsmaterials für 3 Min. bei 95°C in 9 µl Puffer (*sample buffer*) denaturiert. Durch die folgende Abkühlungsphase auf ca. 4°C wird den im Puffer befindlichen so genannten *random*-Primern erlaubt an den Matrizenstrang zu hybridisieren. Anschließend erfolgten die Zugabe von 1 µl der *Phi29*-DNA-Polymerase und 9 µl eines Reaktionspuffers mit den notwendigen dNTPs für die Amplifikation bei 30°C für 16–18 Std.

Die Vorteile der *Phi29*-DNA-Polymerase ist ihre sehr hohe Prozessivität und zusätzlich die „*strand displacement*“-Aktivität, sodass die in der Reaktion generierten DNA-Moleküle sich selbst als Matrize dienen. Dieses Prinzip lässt eine beträchtliche Vervielfältigung der Ausgangs-DNA zu. Auf diese Weise können je nach Einsatzkonzentration der Ausgangs-DNA 4–7 µg DNA generiert werden. (Dean *et al.*, 2001, Lasken und Egholm, 2003).

3.2.3 Probendetektion

3.2.3.1 SBE-Primerdesign

Die Primer für die Minisequenzierung der ersten Multiplexreaktion - der 23plex - wurden von Kopenhagen aus als 100 µM Stammlösungen an die Laboratorien verschickt (Anhang zu Sanchez *et al.* 2006: Table S2). Die Länge der Primer lag zwischen 18 und 78 Nukleotiden. Nachdem zwei Primer (rs729172 und rs1493232) aus dem Set bereits nach einer Analyse degradiert waren, wurden bereits vorhandene Primer aus einer internen Studie genutzt.

3.2.3.2 Minisequenzierungsreaktion

In der Minisequenzierungsreaktion wirkt das Prinzip der Einzelbasenverlängerung. Sie wurde mittels des kommerziellen ABI PRISM® *SNaPshot™* Multiplex Kits protokollgemäß durchgeführt. Es handelt sich dabei um eine thermale, zyklische, Polymerase-vermittelte Einzelbasen-Extensions-Reaktion (SBE = *single base extension*), bei welcher der Analyseprimer mit dem 3'-Ende eine Base vor dem zu identifizierenden SNP bindet und

anschließend in Abhängigkeit von der SNP-Position ein fluoreszenzmarkiertes Dideoxynukleotid (ddNTP) angelagert wird. Analyseprimer wurden auf eine Stammkonzentration von 50 μM justiert und anschließend zum Primermix zusammengefügt, sodass die Endkonzentration wie in der Anhangstabelle vorgegeben in der Reaktion erreicht werden konnte (Tabelle 5).

Tabelle 5: Zusammenstellung aller Reagenzien und Konzentrationen für die Minisequenzierungsreaktion mit dem *SNaPshot*TM-Kit. Bei der Zugabe beider Primermixe in die 52plex Reaktion wurde 1 μl H₂O weniger hinzugegeben.

Reagenzien	Konzentration	Endkonzentration	8 μl Ansatz
<i>SNaPshot</i> TM -Mix			4 μl
Primermix (23/29 autSNPs) [μM]	50 μM	div. *	1 μl
HPLC-H ₂ O			2 μl
Aufgereinigte PCR-Produkte			1 μl
Summe Reagenzien			7 μl

*die Primer wurden in unterschiedlichen Konzentrationen in den Primermix pipettiert und erreichten damit unterschiedliche Endkonzentrationen.

Das Temperaturprofil für die SBE-Reaktion war: 96°C für 10 Sek., 30 Zyklen bei 55°C für 5 Sek., 60°C für 30 Sek.

Anschließend wurden zur Entfernung der überschüssigen ddNTPs in jede Reaktion je 1 μl SAP gegeben. Der Verdau erfolgte nach dem selben Zeit- und Temperaturprofil, wie im Absatz 3.1.4 beschrieben.

3.2.3.3 Kapillarelektrophorese

Ein Teil des Ansatzes wurde anschließend durch eine automatische Kapillarelektrophorese analysiert, bei welcher die nukleotidgekoppelten Fluoreszenzfarbstoffe nach Auftrennung der Fragmente durch einen Laser angeregt werden und deren emittierendes Licht durch eine CCD-Kamera detektiert wird. Die Auftrennung erfolgt physikalisch nach Molekulargewicht.

Die Typisierungsplattform für die Multiplexreaktion war ABI PRISM[®] 310 *Genetic Analyzer*. Bei diesem Gerät handelt es sich um ein vollautomatisches Kapillarelektrophorese-System, das sowohl die DNA-Sequenzierung als auch die DNA-Fragmentanalyse ermöglicht. Das Funktionsprinzip folgt der Methode der Gelelektrophorese mit einem Polymer als Geläquivalent. Das Polymer besteht aus einem Cellulose-Harnstoffgemisch und besitzt unterschiedliche Dichtegrade (POP: *performance optimized polymer*). Zur Analyse wurde POP 4 verwendet. Die Typisierung der gewonnenen Daten wurde mit der *Genotyper 3.7 Software* (Applied Biosystems) durchgeführt.

3.2.4 Auswertung der Populationsdaten

3.2.4.1 Hardy-Weinberg-Gleichgewicht

Die statistische Auswertung der Daten erfolgte mittels Berechnung der Allelfrequenzen der typisierten Populationen aus dem Hauskollektiv. Als Vergleichsdaten dienten die Typisierungsergebnisse der im Rahmen des *HapMap* Projekts¹² analysierten europäischen und asiatischen Populationen, basierend auf den Einträgen in der *Entrez Nucleotide Database* des *National Center for Biotechnology Information* (NCBI)¹³. Der Datenumfang betrug bei den typisierten Individuen aus Mitteleuropa $n=116-120$ und bei Han Chinesen $n=88-90$. Es wurden die Allelfrequenzen des ancestralen Allels notiert. Für einen Marker (rs2016276, hier A15) ist das ursprüngliche Allel nicht bekannt. Hierfür wurde die Frequenz des seltenen Allels (*minor allele*) notiert.

Die Validierung der Daten erfolgte mittels eines Exakten Tests zur Prüfung des Hardy-Weinberg Gleichgewichts mit der *DNAVIEW™ Software* 27.17 (C. Brenner¹⁴). Der Exakte Test basiert auf einer Monte Carlo Simulation, die mit 10.000 Iterationen durchgeführt wurde (Guo und Thompson, 1992).

Als grundlegende Berechnung einer populationsstatistischen Datenanalyse gilt die Ermittlung der jeweiligen Allelfrequenzen in den untersuchten Populationen. Die Allelfrequenz f wird mittels folgender Formel errechnet:

$$f_{ij} = \frac{n_{ij}}{2N_j} \quad ,$$

dabei ist f_{ij} die Frequenz des Allels i des Locus j , n_{ij} ist die Gesamtheit der Allele des Locus j und N_j gibt die Größe der Population an.

Der nächste notwendige Auswertungsschritt ist die Überprüfung, ob es innerhalb der Populationen zu einer Abweichung vom Hardy-Weinberg-Gleichgewicht (HWG) kommt.

Hierbei wird davon ausgegangen, dass innerhalb einer ausreichend großen Population ein Gleichgewicht im Hinblick auf die wirkenden evolutiven Kräfte herrscht. In der Theorie beschreibt man das Gleichgewicht in einer Population als die Summe der Genotypfrequenzen, die sich gemäß der folgenden quadratischen Gleichung aufteilen müssen:

$$p^2 + 2pq + q^2 = 1.$$

¹² <http://hapmap.ncbi.nlm.nih.gov/index.html.en>

¹³ <http://www.ncbi.nlm.nih.gov/nucleotide/>

¹⁴ <http://dna-view.com/>

3.2.4.2 Exakter Test auf Unabhängigkeit

Mit Hilfe des folgenden Tests sollte nachgewiesen werden, dass keine Kopplung zwischen den Loci auftritt. Die Kopplung konnte im vorliegenden Datenset der 23 SNP Multiplex vorwiegend ausgeschlossen werden, da alle Loci, bis auf die SNPs rs735155 und rs826472 auf getrennten Chromosomen lokalisiert waren.

Die Prüfung auf Unabhängigkeit erfolgte mit Hilfe des im Internet¹⁵ frei erhältlichen Programms GENEPOP 4.0 (Rousset und Raymond, 1995).

Zur Berechnung des genaueren Signifikanzniveaus wurde mit dem Exakten Test nach Fischer geprüft (1012 Durchläufe [4 Populationen x 253 Locikombinationen]). Dieser benutzt zur Überprüfung der Testvariablen die exakte Stichprobenverteilung.

$$p_{(a)} = \frac{\binom{a+c}{a} \cdot \binom{b+d}{b}}{\binom{N}{a+b}},$$

wobei N die Anzahl der Beobachtungen wiedergibt.

Der p -Wert wurde für alle Tests mit $p < 0,05$ nach Korrektur des α -Fehlers nach Bonferroni festgelegt. Eine erhöhte α -Fehler-Wahrscheinlichkeit durch das vielfache Testen der Grundgesamtheit war zu erwarten, deshalb wurde eine Multiplikation des p -Wertes mit der Anzahl der Vergleiche durchgeführt (Bonferroni, 1935).

Einen weiteren populationsgenetischen Parameter stellt die Berechnung des prozentualen Anteils an heterozygoten Individuen dar. Zusätzlich werden der erwartete Heterozygotengrad (H_{exp}) und der beobachtete Heterozygotengrad (H_{obs}) betrachtet und zueinander in Verhältnis gesetzt.

$$H_{exp} = 1 - \sum_{i=1}^k p_i^2,$$

wobei k Anzahl der Allele und p_i die Frequenz des i -ten von k Allelen wiedergibt.

Für den beobachteten Heterozygotengrad gelten:

$$H_{obs} = \sum_{j=1}^i \frac{h_j}{i},$$

wobei H_{obs} die mittlere beobachtete Heterozygotierate aller untersuchten Loci widerspiegelt und i die Anzahl der Loci, j der Locus und h_j die Heterozygotenrate für den j -ten Locus darstellen.

¹⁵ <http://kimura.univ-montp2.fr/~rousset/Genepop.htm>

Der Zusammenhang zwischen den populationsgenetischen Parametern wie Fitness, Anpassung und dem Paarungsverhalten einer Population erklärt eine mögliche Abweichung vom H_{exp} .

3.2.4.3 F-Statistiken

Ausgehend von der Heterozygotenrate, die den Grad der genetischen Variation innerhalb einer Population darstellt, wurden weitere statistische Variablen errechnet. Zur Betrachtung der möglichen Störungsprozesse des HWG wurde die F-Statistik (Fixationsindizes, Wright, 1951) angewendet.

Dabei wurde der F_{IS} -Wert ermittelt, der dazu dient, das Verhältnis der Allele zwischen den Individuen gleicher Subpopulation zu berechnen, sowie der F_{ST} -Wert zur Bestimmung des Varianzgrades von Allelfrequenzen zwischen der Gesamtpopulation und ihren Subpopulationen.

Beide Fixationsindizes werden wie folgt berechnet:

$$F_{IS} = \frac{H_s - H_l}{H_s},$$

wobei H_s die mittlere erwartete Heterozygotie H_{exp} in den zufallsgepaarten Subpopulationen darstellt und H_l die mittlere beobachtete Heterozygotie in einer Subpopulation bedeutet.

$$F_{ST} = \frac{H_T - H_s}{H_T},$$

wobei H_T der Mittelwert der erwarteten Heterozygotie in der Gesamtpopulation ist.

Die Ermittlung der F_{ST} -Werte und die Berechnung der Assoziationen zwischen den Allelen der unterschiedlichen Loci erfolgte mittels der frei zugänglichen Software FSTAT Version 2.9.3.2¹⁶ (Goudet *et al.* 2002).

3.2.4.4 AMOVA

Eine Erweiterung der Wright'schen F-Statistiken bilden die Φ -Statistiken. Mit ihnen lassen sich die genetischen Distanzen innerhalb und zwischen den untersuchten Populationen darstellen. Mit Hilfe des durch Schneider *et al.* (2000) entwickelten ARLEQUIN v. 3.1¹⁷ Programms wurde die Analyse der molekularen Varianz durchgeführt, engl. **Analysis of Molecular Variance** – AMOVA (Excoffier *et al.* 1992). In Einzelschritt-Mutationsunterschieden wird die mittlere Entfernung des Genotyps eines Individuums von jedem anderen im Versuch analysierten Individuum ermittelt. Diese wird durch die Gesamtvarianz σ^2 beschrieben, welche sich aus den einzelnen Varianzkomponenten σ_c^2 , σ_b^2 und σ_a^2 zusammensetzt. Dabei stellt σ_c^2 die Variabilität der Allelfrequenzen innerhalb einer Population dar. Die übrigen

¹⁶ <http://www2.unil.ch/popgen/softwares/fstat.htm>

¹⁷ <http://lgb.unige.ch/arlequin/>

Kovarianzen entsprechend den Variabilitäten zwischen Populationen einer Gruppe (σ_b^2) und zwischen den Gruppen (σ_a^2).

Mit Hilfe der Kovarianzen wurde für die Allelfrequenzen die Φ -Statistik nach folgenden Formeln ermittelt:

$$\Phi_{CT} = \frac{\sigma_a^2}{\sigma_T^2}, \quad \Phi_{ST} = \frac{\sigma_a^2 + \sigma_b^2}{\sigma_T^2} \quad \text{und} \quad \Phi_{SC} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_c^2}.$$

Die Kovarianzen durchlaufen einen Signifikanztest, welcher auf dem Vergleich der beobachteten Werte mit denen einer Nullverteilung basiert. Der Test wurde mittels 1023 Permutationen durchgeführt.

3.2.4.5 Korrelationskoeffizient

Der Korrelationskoeffizient nach Pearson quantifiziert den linearen Zusammenhang zwischen zwei intervallskalierten Variablen. Er ergibt sich aus der Division der Kovarianz (cov) durch die Standardabweichungen ($s_{x,y}$) der beiden Variablen (x, y),

$$r = \frac{cov(x, y)}{s_x s_y} = \frac{\frac{1}{n} \sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n} \sum (X_i - \bar{X})^2} \cdot \sqrt{\frac{1}{n} \sum (Y_i - \bar{Y})^2}}.$$

Korrelationskoeffizienten (r) nach Pearson wurden mittels SPSS-Software Version 11.5 berechnet.

3.2.4.6 Ausschlusswahrscheinlichkeit

Im Rahmen der forensischen Statistik wurde die kombinierte Ausschlusswahrscheinlichkeit ($cEXP$) für alle untersuchten Loci mittels der DNAVIEW™ Software 27.17 errechnet. Hierbei gilt:

$$cEXP = 1 - \prod_{i=1}^1 (1 - EXP),$$

wobei $\prod_{i=1}^1$ die Produktsumme und i die Anzahl der Loci ist.

3.3 Validierung der 52 SNP Multiplexreaktion mittels Minisequenzierung

3.3.1 SNP-Selektion und PCR-Primerdesign

Aus einem Pool von insgesamt 345 primär ausgesuchten SNPs (Kapitel 3.1.1) wurde eine weitere Selektion durchgeführt. Die distalen Bereiche der *p*-Chromosomenarme bildeten hierbei die erste Gruppe der autosomalen Loci, aus der 46, d.h. ca. 2-3 SNPs pro Chromosom ausgesucht wurden. Ausgenommen davon war das Chromosom 19. Der zweite Selektionsbereich befand sich innerhalb der distalen Bereiche der *q*-Arme. Hier wurden 3-4 Loci pro Chromosom selektiert, was der Anzahl von 67 SNPs entsprach. Auch in diesem Fall konnte aufgrund der Gendichte auf dem Chromosom 19 kein Marker ausgewählt werden. Die dritte Gruppe bildeten 25 Loci, die aus den Bereichen der großen Autosome selektiert wurden.

Die abschließende Selektion fand im Zuge der Primerentwicklung statt. Hier fiel die Auswahl der Loci auf die mit der kürzesten Amplikonlänge und der dadurch höheren Wahrscheinlichkeit in einer Multiplexreaktion amplifiziert zu werden. 42 Loci wurden aus den ersten beiden Gruppen ausgewählt. 12 Marker wurden aus der dritten Gruppe hinzugenommen. Im Zuge dessen wurde auch ein geeigneter SNP auf dem Chromosom 19 etabliert. Die Auswahl ergab eine Summe von 54 Markern, welche erfolgreich in Singleplexen amplifiziert wurden. Eine Aufteilung der 54plex ergab zwei gesonderte Multiplexreaktionen von 23 und 31 Markern, die nacheinander etabliert wurden.

Nach einer Testphase beider Reaktionen (23 SNP Multiplexreaktion als Teil dieser Arbeit) kam es zu fünf Änderungen der Analyseprimer: in drei Fällen erfolgte eine Verlagerung der Primer auf den komplementären Strang und in zwei Fällen wurde eine Längenänderung vollzogen. Zwei der selektierten Marker wurden aus der Reaktion herausgenommen, so dass am Ende eine 52 SNP Multiplexreaktion bereit stand (Kapitel 5.2). Alle Änderungen erfolgten im Hinblick auf eine Effizienzerhöhung und Ausbalancierung der Reaktionen.

3.3.2 DNA-Proben

Zur Validierung der 52 SNP Multiplexanalyse unter Verwendung der Minisequenzierung mittels *SNaPshot*TM wurden in Mainz insgesamt 152 Proben zweifach typisiert. Sie gehörten zu einem Pool von insgesamt 700 Proben, die in vier Partnerlaboren analysiert wurden. Bei den in Mainz untersuchten Proben handelte es sich um folgende Populationen: Deutsche (n=49), Chinesen (n=63), Thailänder (n=33) und Japaner (n=7). Die Proben gehörten zum Hauskollektiv der Rechtsmedizin in Mainz und wurden im Rahmen früherer Studien des

Instituts gesammelt (Zhang *et al.* 1993, Waiyawuth *et al.* 1998, Schneider *et al.* 1998, Seo *et al.* 1998). Die Extraktion der Proben aus Vollblut wurde mittels der Aussalzungsmethode, nach dem Protokoll von Miller *et al.* (1988) oder mit dem QIAamp *DNA Blood Mini Kit* (Qiagen) durchgeführt. Des Weiteren wurden in der gesamten Studie 156 Individuen aus Dänemark, 149 aus Grönland, 104 aus Somalia, 96 aus der Türkei und 43 aus Taiwan typisiert.

Zur Feststellung des ancestralen Allels wurden DNA-Proben von 29 unverwandten Schimpansen, drei Gorillas, einem Orang-Utan, zwei Pavianen, fünf Rhesusaffen und sechs Cynomolgus Affen (Makaken) typisiert (Mack *et al.* 2004). Zur Kontrolle wurden außerdem Proben einer Katze, eines Hundes sowie von zwei Pferden untersucht.

3.3.3 Probendetektion

3.3.3.1 SBE-Primerdesign

Die SBE-Multiplexreaktionen setzten sich zusammen aus der bereits vorvalidierten 23plex Reaktion und einer neu entwickelten Multiplexreaktion aus 29 SBE-Primern. Die Primerlänge lag zwischen 16 und 92 Nukleotiden für beide Multiplexreaktionen. Vorgaben für die Abstandsintervalle der Analyseprimer wurden bei Primern kürzer als 35 Nukleotide auf fünf bis sechs Nukleotide festgelegt sowie auf vier Nukleotide Abstand für Primern länger als 35 Nukleotide. Es war darauf zu achten, dass komplementäre Allelkombinationen bei Möglichkeit im gleichen Längenintervall analysiert wurden (Sanchez *et al.* 2006, Abbildung 12).

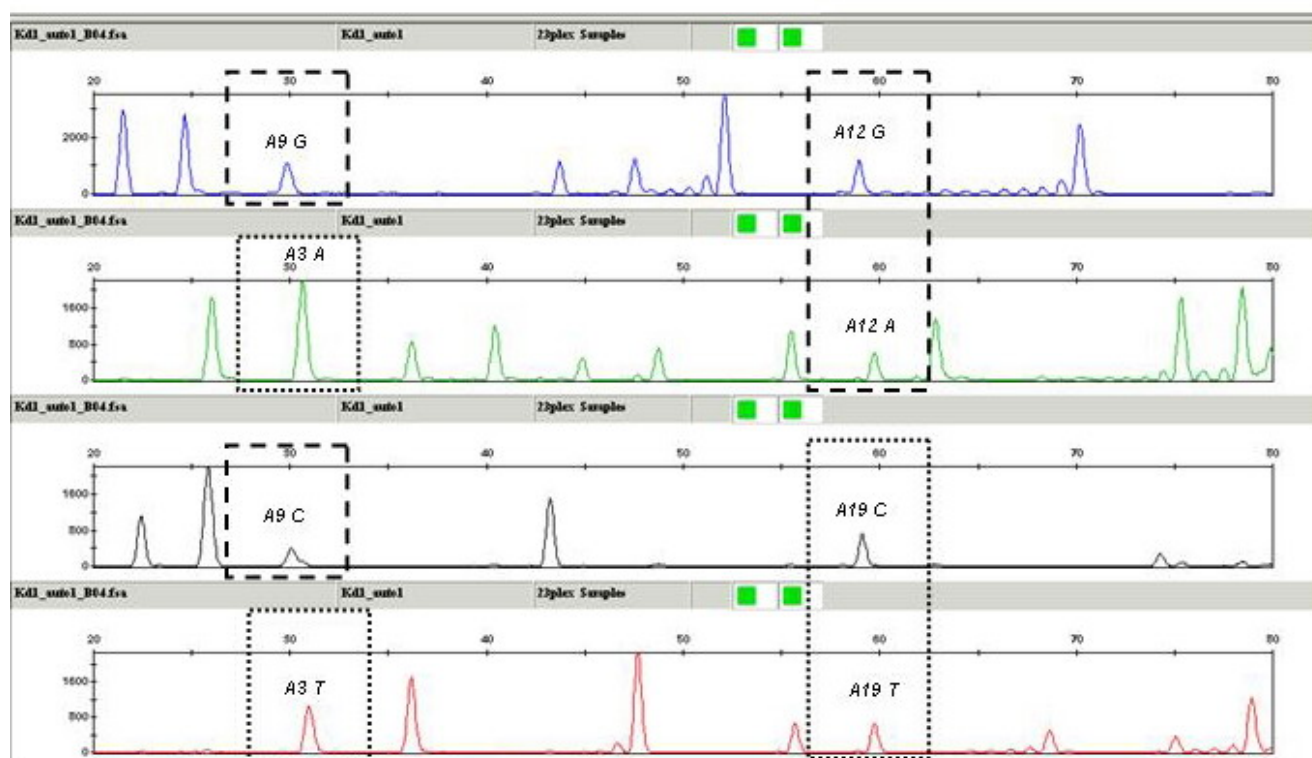


Abbildung 12: Darstellung eines Elektropherogramms. Allele, die mit SBE-Primern gleicher Länge amplifizieren, können durch die vier Farben gleichzeitig dargestellt werden, da sie in einem unterschiedlichen Spektrum emittieren. Die zusammengehörigen Polymorphismen sind jeweils in Blöcken dargestellt (gestrichelte Linie, bzw. gepunktete Linie).

Das so festgelegte Primerdesign ermöglichte eine Detektion von zwei SNPs innerhalb eines Analysefensters von nur 4 nt. Die Begrenzung der Primerlängen auf maximal 90 Nukleotide ergab sich aus dem Qualitäts- und Kostenfaktor. Längere Primer können durch eine unvollständige Synthese bei einer schlechten Aufreinigung so genannte Schulterpeaks aufweisen. Ein weiterer Faktor ist, bedingt durch unsachgemäße Lagerung, eine höhere Abbaurate der Nukleotide, die zum gleichen Phänomen im Elektropherogramm führt. Lange SBE-Primer sind zudem in ihrer Synthese sehr kostenintensiv. Aus diesen Gründen musste die Detektion auf zwei Reaktionen aufgeteilt werden (s.a. Kapitel 3.3.3.2).

Zusätzlich wurden die Mixansätze zur Erhaltung der maximalen Qualität der Minisequenzierungsprimer pro Reaktion in 160 mM Ammoniumsulfat verdünnt.

3.3.3.2 Sequenzierungsreaktion und Kapillarelektrophorese

Bei Sanchez *et al.* (2006) wurde das Analyseverfahren zur Detektion von SBE-Produkten beschrieben, das im Kopenhagener Partnerlabor angewendet wurde. Dieses beinhaltete ein verändertes Modul in den Methodendateien der *GeneScan Analysis 3.7* Software. Beide Multiplexreaktionen wurden in einem Detektionsschritt nach dem Prinzip der „doppelten Injektion“ analysiert. Das Verfahren wurde später nach ISO17025 akkreditiert (Børsting *et al.* 2009).

Im den restlichen Laboratorien konnte diese Modifikation aufgrund einer neueren Version der ABI 3100 *Data Collection* Software nicht eingesetzt werden. Die Detektion wurde daher in zwei getrennten Schritten durchgeführt. Die aufgereinigten Amplifikate wurden in zwei *SNaPshot*TM Ansätze überführt, in denen sich jeweils die Analyseprimer der beiden SBE-Reaktionen befanden: für die 23plex mit der Bezeichnung – *auto1*, und die darauf folgende Multiplexreaktion mit 29 SNP Markern – *auto2*. Beide Reaktionen wurden nacheinander im ABI PRISM[®] 3100 *Genetic Analyzer* aufgetrennt und detektiert.

Ausgelesen und typisiert wurden beide Reaktionen mit Hilfe der GeneMapper *ID* 3.0 Software. Die automatisierte Genotypisierung nach der Überprüfung der *bins* erfolgte durch den Datenexport und die Umwandlung der Tabelle in Textformat sowie der Generierung einer umformatierten Tabelle im *Excel*-Format mit Hilfe eines Access-Makros¹⁸.

3.3.3.3 Signalnormalisierung

Die unterschiedliche Signalstärke der Fluoreszenzfarbstoffe machte es notwendig, die Intensitäten zueinander in ein Verhältnis zu stellen. Durchschnittlich verhielt sich das Verhältnis der Signalintensität (RFU) der Fluorophore dR110 (blau), dR6G (grün), *TAMRA* (gelb) und dROX (rot) bei allen Geräten 4:2:1:1. Die Schwellenwerte für die Erkennung der homo- und heterozygoten Allele wurden diesem Verhältnis entsprechend angeglichen.

Bei schwachen Signalen einzelner SNPs in einigen der Proben wurde diesen SNPs der Status eines fehlenden Signals erteilt. In 700 doppelt typisierten Individuen haben bei 20% der Proben die Signale von bis zu fünf Loci in einem der Durchgänge gefehlt. In diesen Fällen wurde ein dritter Durchlauf der Probe mit erhöhter Menge DNA durchgeführt.

3.3.3.4 Reproduzierbarkeits- und Sensitivitätsüberprüfung

Die Validierungsstudie erfolgte nach den Richtlinien der „*Scientific Working Group on DNA Analysis Methods*“¹⁹.

Die Beurteilung der Multiplexleistung erfolgte nach Analysen zweier Individuen in einer Verdünnungsreihe von 0,07; 0,14; 0,27; 0,55; 1,09; 2,19; 4,37; 8,75; 17,50; 35 und 70 ng genomischer DNA. Die Signalintensitätswerte (RFUs) jedes Fluorophors wurden zusammengefasst und durch Teilung der Homozygotenwerte durch zwei normalisiert.

3.3.4 Auswertung der Populationsdaten

Die Berechnung der populationsgenetischen Statistik erfolgte wie in Kapitel 3.2.4 beschrieben. Abweichende oder zusätzliche Berechnungen sowie Analysen, die mit anderen statistischen Programmen durchgeführt wurden, werden hier vermerkt.

¹⁸ Erhältlich auf der Projekthomepage <http://www.snpforid.org>

¹⁹ <http://www.fbi.gov/hq/lab/fsc/backissu/july2004/index.htm>

3.3.4.1 Linkage Disequilibrium

Die Ermittlung des LD-Koeffizienten sowie die Berechnung der vorab aufgeführten Faktoren erfolgten für die Validierung der 52 SNP Multiplexreaktion mittels des *SNP Assistant*™ v. 1.0.9 Programms²⁰.

Zur Ermittlung der Kopplung wurden drei Verfahren gewählt: der χ^2 -Test, der pL10-Koeffizient und der Exakte Test.

Zur Überprüfung der Allelhäufigkeit wurde von der genutzten Software das Modell der Anpassungsgüte (*goodness-of-the-fit statistic*) eingesetzt.

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - e_{ij})^2}{e_{ij}},$$

wobei n_{ij} die errechnete Frequenz für Allele i und j und e_{ij} die erwartete Frequenz für beide Allele ist (Weir 1979).

Die Untersuchung der LD-Hypothese unter der Annahme der Panmixie behilft sich des Modells der unbeschränkten Genotypbildung durch zufällige Paarung der Allele. Diese Größe wird durch die *log-likelihood*-Funktion beschrieben ($\ln L_1$). Die *likelihood-ratio*-Prüfung beruht dabei auf folgender Teststatistik:

$$LR = 2(\ln L_1 - \ln L_0),$$

wobei $\ln L_0$ der natürliche Logarithmus der *likelihood* des restriktierten Modells ist. Diese Statistik folgt der χ^2 -Verteilung und die Zahl der Freiheitsgrade entspricht der Zahl der Restriktionen (pL10-Koeffizient nach Weir, 1979).

3.3.4.2 F-Statistiken und Korrelationskoeffizient

Die Ermittlung der F_{ST} -Werte erfolgte durch das frei erhältliche Programm *Genetic Data Analysis Software*²¹ (Lewis und Zaykin, 2002).

Der Korrelationskoeffizient (r) wurde zur statistischen Auswertung der 52 SNP Multiplexreaktion mit Hilfe von *Excel* berechnet.

3.3.4.3 Clusteranalyse

Die von der SNPforID-Projektgruppe ausgewählte Methode der Clusteranalyse war der von Pritchard *et al.* (2000) implementierte Ansatz der Individuenzuordnung anhand der Genotypen, basierend auf ungekoppelten Markern, unter der Annahme des Hardy-Weinberg-Gleichgewichts innerhalb der Populationen. Der Ansatz basiert auf dem Bayesschen Theorem, mit dessen statistischem Ansatz unbekannte Parameter geschätzt werden.

²⁰ www.biodata.ee/SNPAssistant.htm

²¹ <http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php>

Die Analyse wurde im Labor in Santiago de Compostela mittels der *Structure ver. 2.0 Software* durchgeführt und dargestellt.

3.3.4.4 Forensische Biostatistik

Zur Berechnung der Diskriminationsstärke von autosomalen Systemen wurde die *PowerStats Software* (Promega) zur Hilfe genommen. Diese berechnet die Wahrscheinlichkeit, mit welcher zwei zufällig ausgewählte Individuen aus einer Population unterschiedliche Genotypen aufweisen nach folgender Formel (Jones, 1972):

$$1 - 2(\sum p_i^2)^2 + \sum p_i^4.$$

Mit der so genannten *matching probability* P_M (Brenner, 1993) wird die Wahrscheinlichkeit errechnet, mit welcher ein zweites, nicht verwandtes Individuum zufällig die gleiche genetische Merkmalskombination aufweist. Die Wahrscheinlichkeit verringert sich mit der Anzahl der untersuchten Systeme. Hierbei werden die Einzelwahrscheinlichkeiten der Systeme multipliziert.

Die Ausschlusswahrscheinlichkeit (*EXP*) für einen Locus, sowie die kombinierte Ausschlusswahrscheinlichkeit (*cEXP*) für alle untersuchten Loci wurden nach der Vorlage aus Kapitel 3.2.4 berechnet.

Zuletzt wurde auch der Vaterschaftsindex (engl. *paternity index* - PI), bei welchem das Verhältnis eines DNA-Markers, welchen der Putativvater dem Kind vererbt hat, gegenüber dem gleichen Marker einer zufälligen männlichen Person errechnet (Brenner und Morris, 1989).

Alle oben genannten Parameter wurden im Kopenhagener Labor mittels der DNAVIEW™ Software 27.19 berechnet.

3.3.5 STR-Typisierung und Sequenzierung

Die Prüfung der Sensitivität der Multiplexanalyse fand ebenfalls im Kopenhagener Labor statt. Hierbei wurden von sieben Individuen *post mortem* entnommene Proben aus unterschiedlichen Geweben wie Knochen, Muskeln u. ä. extrahiert. Die Proben befanden sich in unterschiedlichen Zersetzungszuständen und entstammten alle einer vorangegangenen rechtsmedizinischen Untersuchung. Es wurden verschiedene Verdünnungen der Proben hergestellt. Sie wurden zum Abgleich mit dem *AmpFISTR® Identifier®* oder dem *AmpFISTR® SGM Plus* PCR-Amplifikationskit (Applied Biosystems) untersucht.

Die Sequenzierung von mindestens drei Individuen mittels *Big Dye® Terminator v.3.1 Cycle Sequencing Kit* (Applied Biosystems) wurde an 14 PCR-Produkten der SNPs rs1028528,

rs1029047, rs1979255, rs2016276, rs2046361, rs2056277, rs2107612, rs2831700, rs354439, rs733164, rs735155, rs901398, rs907100 und rs938283 durchgeführt.

3.4 Microarray-Analyse der 33 SNP Multiplexreaktion

Parallel zur Validierung der oben beschriebenen Referenzmethode fanden die ersten Versuche zur Evaluierung der Multiplexmethode mittels Microarray-Technologie auf Glasobjektträgern statt. Bis auf eine heterogene DNA-Probenauswahl, begründet durch eine höhere Anzahl an Referenzmethoden, nahm die Probenvorbereitung bis zur enzymatischen Aufreinigung der Amplifikate den gleichen Verlauf wie in Kapitel 3.1 beschrieben und wird in den folgenden Kapiteln nicht näher ausgeführt.

3.4.1 DNA-Proben

Zur vergleichenden Untersuchung der Microarray-Methode wurden genomische DNA-Proben aus einer standardisierten Kontrollplatte der Firma Applied Biosystems verwendet (Abbildung A-1 im Anhang). Diese diente als Referenzpanel für das *SNPlex*TM System und enthielt 44 doppelt applizierte DNA-Proben aus dem *European Collection of Cell Cultures*²².

3.4.1.1 Proben zur Optimierung der Microarray-Analyse

Die ersten Experimente, die zur Optimierung der 33plex auf Glasobjektträgern dienten, wurden anhand bereits ausgewerteter DNA-Proben vollzogen. Hinzugezogen wurden bereits mehrfach typisierte Proben deutscher Herkunft aus dem Hauskollektiv sowie eine weitere Kontroll-DNA, die ebenfalls bereits mit der 52 SNP Multiplexreaktion typisiert worden war.

Die genaue Bezeichnung der Proben lautete:

- PV19 (Konzentration ~20 ng/μl)
- Kd11 (Konzentration ~20 ng/μl)
- BSB (Konzentration ~20 ng/μl)

3.4.1.2 Proben für die Sensitivitätsstudie

Nach abgeschlossener Optimierung wurde eine Verdünnungsreihe erstellt, um die Sensitivität der Methode zu untersuchen. Die hierfür benutzten Proben stammen aus dem oben beschriebenen Probenpool (Kapitel 3.4.1).

Die genaue Bezeichnung der Proben war:

- E34a
- E9046a

²² www.ecacc.org.uk

Die DNA-Konzentration der Proben wurde vorab bestimmt. Die Verdünnungsreihe enthielt nach Ansatz mit HPLC-H₂O folgende Konzentrationen: 1 ng, 500 pg, 200 pg, 100 pg, 50 pg und 20 pg.

Die Konditionen der PCR-Reaktion und der weiterführenden Reaktionen blieben unverändert. Verwendet wurden frisch angesetzte Primermixe.

Die Mittelwerte der Signalintensitäten mit ihren Standardfehlern wurden für jeden gekoppelten Fluoreszenzfarbstoff innerhalb der einzelnen Proben ermittelt und gegen die eingesetzte Menge DNA in PCR-Reaktion aufgetragen. Die Darstellung der Ergebnisse erfolgte anhand eines Diagramms (Diagramm 10).

3.4.2 Referenzmethoden

Die DNA-Proben wurden in drei SNPforID Laboratorien mit der *Genplex*-Methode genotypisiert. Zusätzlich wurden die Proben in einem teilnehmenden Labor mit dem *SNaPshot*TM Kit zwei Mal typisiert. Mit den Referenzmethoden *Genplex* und *SNaPshot*TM wurden 48 SNPs analysiert. Mit der *Genplex*-Methode erfolgte die Typisierung der Probenserien 1 und 2. Mit *SNaPshot*TM wurde die Probenserie 1 typisiert. Vier der ursprünglich 52 Marker wurden nicht in die Genotypisierung einbezogen. Dabei handelte es sich um folgende Marker: rs2016276, rs826472, rs2830795 und rs1028528. Eine zweite Typisierung mittels *SNaPshot*TM erfolgte an allen Markern anhand der Probenserie 1. Aus der Probenserie 2 wurden zwei Proben typisiert (E34b und E35b). In der Vergleichsanalyse dieser Arbeit wurde als Nomenklatur die Typisierung der 52 SNPs mittels *SNaPshot*TM als *SNaPshot*TM I und die Typisierung der 48 SNPs des *Genplex*-SNP-Panels als *SNaPshot*TM II geführt.

Das *MassARRAY*[®] *iPLEX* System (SEQUENOM) war eine weitere Typisierungsplattform, mit der ein Teil der 52plex untersucht wurde. Die hier erzielten Daten stellten keine Referenz für die Auswertung der Microarray-Ergebnisse dar. Sie wurden dennoch in die Beurteilung der schwer sequenzierbaren Genotypen einbezogen. Mit dem *iPLEX*[®]-System wurden 39 SNPs typisiert. Folgende Marker wurden nicht in diese Methode einbezogen: rs2016276, rs826472, rs2830795, rs1028528, rs1024116, rs1355366, rs1382387, rs1493232, rs1528460, rs251934, rs735155, rs737681 und rs873196.

Mit der Microarray-Methode wurden zunächst alle 52 SNPs anhand von DNA-Proben aus dem Hauskollektiv analysiert. Im weiteren Verlauf wurde die Anzahl der Marker auf 33 reduziert. Herausgenommen wurden Marker, die schwache, bzw. keine Signale lieferten sowie SNPs, bei denen es auf den Spots zu Kreuzreaktionen kam. Die übrigen Marker wurden, um die Ergebnisse vergleichend darstellen zu können, ebenfalls anhand der Kontrollplatte (Applied Biosystems) analysiert.

3.4.2.1 Sequenzierung

Kam es im Vergleich der Ergebnisse mit den Referenzproben zu regelmäßigen Abweichungen, wurden Stichproben sequenziert, um den Nachweis der korrekten Sequenz zu erhalten.

Die PCR wurde nach den in Kapitel 3.1.3 genannten Vorgaben angesetzt.

Die Cyclesequenzierung wurde in 20 µl Gesamtvolumen durchgeführt, wobei das Matrizenvolumen, je nach beobachteter Bandenstärke zwischen 1-2 µl lag. Die Reaktion wurde mit dem *Big Dye[®] Terminator v.3.1 Cycle Sequencing Kit* (Applied Biosystems) nach folgendem Protokoll durchgeführt (Tabelle 6):

Tabelle 6: Zusammenstellung aller Reagenzien und ihrer Konzentrationen, sowie die Einsatzmenge der Reagenzien in einer Cyclesequenzierung.

Reagenzien	Konzentration	Endkonzentration	20 µl Ansatz
PCR-Reaktionspuffer [X]	10 x	1,75 x	3,5 µl
<i>Dye[®] Terminator-Mix</i>			1 µl
Primer [µM]	10 µM	0,5 µM	1 µl
HPLC-H ₂ O			ad 20 µl
Aufgereinigte PCR-Produkte			1-2 µl
Summe Reagenzien			18-19 µl

Die Sequenzierungsbedingungen waren: 96°C für 1 Min. (Initialdenaturierung), 25 Zyklen bei 96°C für 10 Sek. 50°C für 5 Sek. 60°C für 4 Min.

3.4.2.2 Fällung der Sequenzierung

Gefällt wurde das gesamte Reaktionsvolumen nach dem in Tabelle 7 beschriebenen Protokoll.

Tabelle 7: Zusammenstellung aller Reagenzien und ihrer Konzentrationen sowie die Einsatzmengen der Reagenzien und die Reihenfolge der Vorgehensweise in der Fällungsreaktion nach einer Cyclesequenzierung.

Reagenzien	Konzentration	363,5 µl Ansatz
Sequenzierprodukt		20 µl
HPLC-H ₂ O		80 µl
EDTA [M]	0,5 M	2,5 µl
Mischen und in 1,5 ml Reaktionsgefäß mit folgenden Reagenzien überführen		
Ethanol [%]	100%	250 µl
<i>Blue Dextran</i> [mg/ml]	50 mg/ml	1 µl
NaAcetat pH 4,8 [M]	3 M	10 µl
10 Min. bei RT stehen lassen. 25 Min. bei 13 000 U/min zentrifugieren. Anschließend Überstand abnehmen		
Ethanol [%]	70 %	300 µl
5 Min. bei 13 000 U/min zentrifugieren		

Der Überstand wurde im Anschluss mit größter Sorgfalt abgenommen und das Pellet bei 45-50°C 30-45 Min. vollständig getrocknet. Gelöst wurden die Probe in 20-25 µl HiDi-Formamid und anschließend in 0,5 ml Reaktionsgefäße überführt.

Die Sequenzierung erfolgte auf dem *ABI Prism 310 Genetic Analyser* mittels Kapillarelektrophorese bei 15 Min. Laufzeit in POP-6. Die Daten wurden mit der *Data Collection Software 3.1* gesammelt und die Analyse erfolgte unter Zuhilfenahme der *Sequencing Analysis Software 3.1*.

3.4.3 Arraydesign

Die Durchführung des experimentellen Teils in Bezug auf das Arraydesign, die verwendeten Fluoreszenzfarbstoffe und die für die Analyse zur Verfügung stehenden Geräte wurden durch eine Reihe nicht vorhersehbarer Entwicklungen sehr stark beeinflusst. Dies führte zu unterschiedlichen konzeptionellen Änderungen in den Versuchsansätzen, auf die im Folgenden eingegangen wird.

3.4.3.1 Versuchsaufbau 1

Das wichtigste Kriterium bei der Auswahl der Glasobjektträger war die Eigenschaft, durch spezifische Oberflächenbeschichtung Aminogruppen (NH₂-Gruppen) zu binden. Bei den mit Epoxysilan beschichteten Objektträgern erwies sich das Bindeverhalten als besonders hoch (Lindroos *et al.* 2001).

Zu Beginn der Arbeit wurde die Wahl getroffen, die Synthese der DNA-tags mit ihren Kopplungsgruppen (siehe Kapitel 3.4.4) sowie die Herstellung von fertig gespotteten Glasobjektträgern an ein Unternehmen abzugeben, das beide Vorgaben durchführen konnte. Das Kerngeschäft des Unternehmens MWG Biotech AG lag zum damaligen Zeitpunkt bei der Herstellung von Microarray-Produkten. Dies änderte sich jedoch im Laufe des Projekts, da das Unternehmen die Microarraysparte verkaufte. Im weiteren Verlauf übernahm MWG Biotech AG (jetzt Eurofins MWG Operon) nur die Synthese der DNA-tags. Die Produktion der Objektträger wurde daraufhin vom Institut für Genetik der Universität Mainz durchgeführt. Hierfür wurden *CodeLink™ Activated Slides* verwendet (GE Healthcare, früher Amersham Biosciences).

Ziel des Designs der DNA-Mikroarrays auf den Glasobjektträgern war, mehrere Proben gleichzeitig auf eine bestimmte Anzahl von beliebig auszuwählbarer SNPs zu untersuchen. Dabei kam hier das Prinzip des „array of arrays“ zur Anwendung.

In der ersten Versuchsanordnung wurden auf jedem beschichteten Glasobjektträger zunächst vier Subarrays platziert. Der Abstand der einzelnen Arrays entsprach dem Abstand der Vertiefungen einer 96er Mikrotiterplatte.

Die Array-Anordnung entsprach 10 *spots* horizontal und 6 *spots* vertikal (Abbildung 13). Alle *spots* wurden in doppelter Anordnung angelegt.

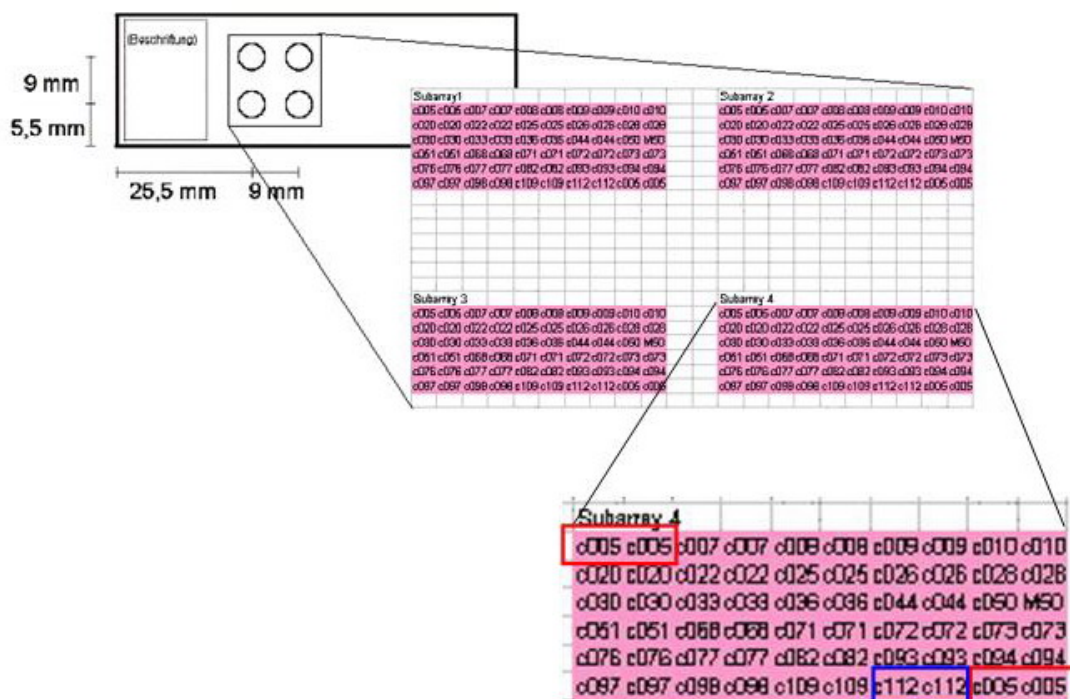


Abbildung 13: Schematische Darstellung der ersten Glasobjektträger, hergestellt für die im Text beschriebene Applikation. Die pink gefärbten Felder entsprechen den vier *subarrays* (auf dem Glasobjektträger dargestellt als Kreise) mit der Anordnung der komplementären *tags* (*ctags*). Der rote Rahmen ist um die Felder der Hybridisierungskontrolle gelegt. Der blaue Rahmen umrandet die beiden *spots* der Reaktionskontrolle.

3.4.3.2 Versuchsaufbau 2

In der weiteren Entwicklung der Methode wurde die Subarrayanzahl sukzessive auf zwölf und die *spot*-Anzahl im Subarray auf 120 Felder erhöht. Sie wurden in der Anordnung 10 *spots* horizontal und 12 *spots* vertikal angelegt. Die Felder für die Hybridisierungskontrollen blieben jeweils an den Rändern des Subarrays, den RC-Spots (*reaction control* – siehe unten) wurde ein neuer Ort zugewiesen (Abbildung 14).

Kontrollsonden:

Von der äußeren linken Ecke über die Diagonale lagen die Felder der Hybridisierungskontrollen (Abbildung 13, roter Rahmen). Die Felder mit der Bezeichnung c112 wurden für die Reaktionskontrollen vorgesehen (Abbildung 13, blauer Rahmen).

Hybridisierungskontrollen:

Die Felder der Hybridisierungskontrollen fungierten als Positionskontrollen zum korrekten Anbringen der Auswertungsraster (Abbildung 14 und Kapitel 3.4.6.4). Die gespotteten *ctag*-Sequenzen wurden während der Hybridisierungsreaktion (Kapitel 3.4.6.2) mit einem Cy5-markierten Nukleotid gekoppelt.

Reaktionskontrollen:

Die Komponenten für die Reaktionskontrolle bestanden aus vier artifiziellen Sequenzen, welche in der Minisequenzierungsreaktion (Kapitel 3.4.6) neben den PCR-Produkten als Vorlage (*template*) dienten. Sie unterschieden sich ausschließlich in einer Base in der Mitte der Sequenz [N]:

P25hz : 5' CGAGCACAATTCTATTT[N]CAGGCAGGTTTCAGGCAG 3'
[]GTCCGTCCAAAGTCCGTC - tag 5' : SBE_P25hz

Die *template*-Sequenzen wurden in gleichen Konzentrationen in der Reaktion eingesetzt. Die SBE_P25hz-Sequenz wird in der Minisequenzierungsreaktion in allen vier markierten ddNTPs zu gleichen Teilen verlängert und bindet in der Hybridisierungsreaktion an die immobilisierte *ctag*-Sequenz am Array.

Im Versuchsaufbau besaß die Reaktionskontrolle zwei Funktionen: **1)** Sie fungierte als eine optische Positivkontrolle zur Beurteilung des Erfolges der Minisequenzierungsreaktion. **2)** Mittels der Messung der Signalintensitäten aller auf den Kontrollfeldern des Arrays elongierter und gebundener SBE-Primer, wurde die Normalisierung der Daten durchgeführt (Kapitel 3.4.6.5).

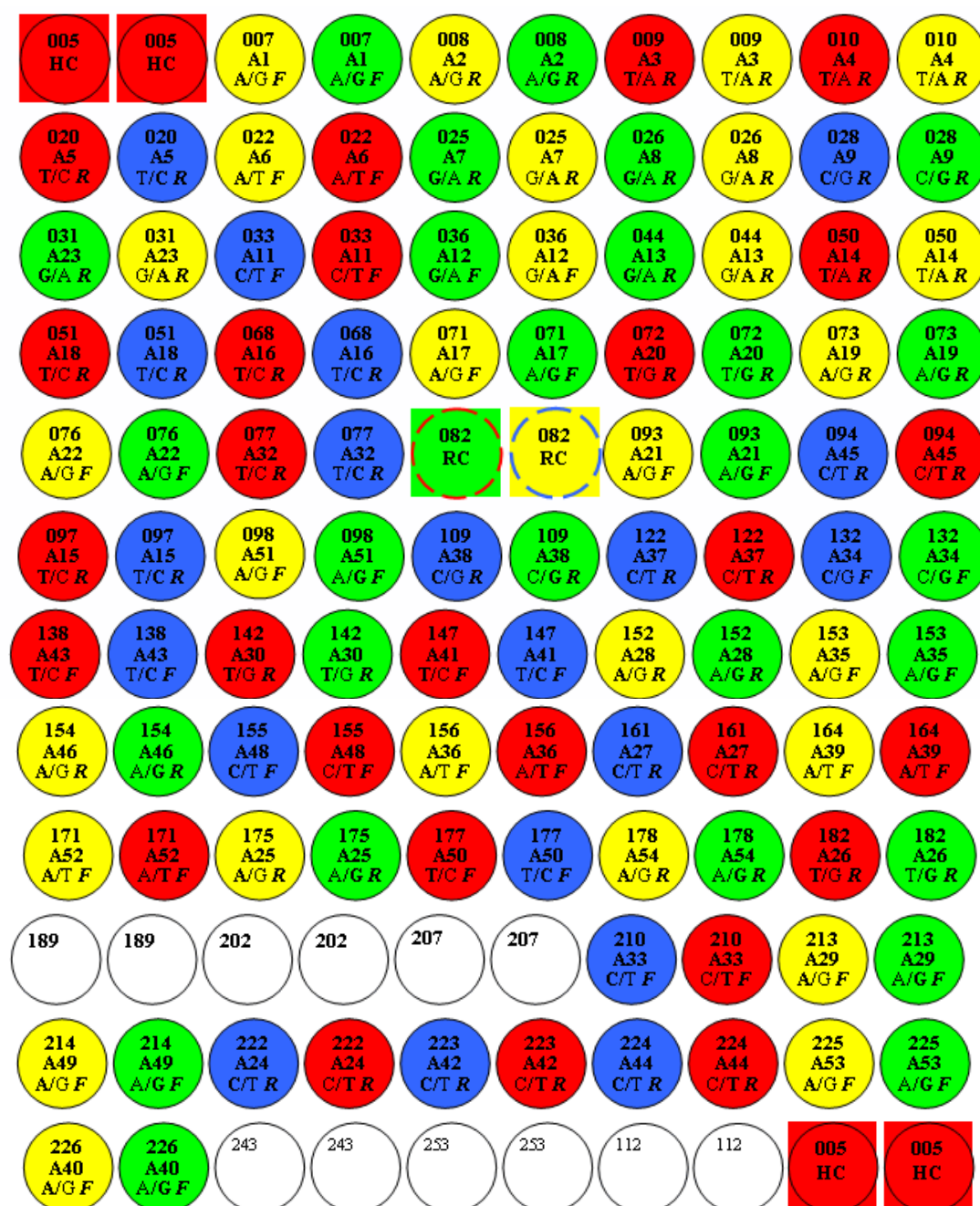


Abbildung 14: Schematische Abbildung eines Arrays aus der erweiterten Versuchsanordnung. Die *spot*-Menge ist auf die 52 SNP Multiplexreaktion ausgerichtet. Die Beschriftung der Felder ist an der Reihenfolge der drei letzten Ziffern der *ctag*-Probe, der SNP-Bezeichnung, dem Polymorphismus und der Primerrichtung ausgerichtet. Farben der *spots* entsprechen jeweils einem der beiden Fluoreszenzfarbstoffe der polymorphen Stelle. Die Anordnung der Hybridisierungskontrollen (005 HC) an den äußeren Rändern des Arrays blieb unverändert zu der ursprünglichen Versuchsanordnung. Die Reaktionskontrolle wurde in die Mitte des Arrays versetzt (gestrichelte Kreislinie, 082 RC).

Um bis zu zwölf unabhängige Proben in einem Hybridisierungsexperiment analysieren zu können, wurde eine Hybridisierungskammer (freundliche Schenkung: Dr. Steffen Schmitt²³, Zytometrielabor, Universität Mainz, Angaben über den Hersteller sind nicht bekannt) modifiziert. Der Plexiglasdeckel wurde durch einen zweiten ersetzt. In diesem befanden sich die Deckelöffnungen in der gleichen, durch die Lage der Subarrays vorgegebenen Anordnung (Abbildung 15).

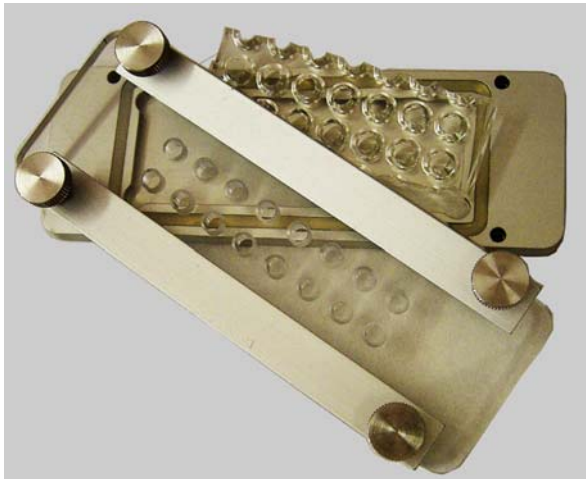


Abbildung 15: Hybridisierungskammer mit einem Silikonadapter und präpariertem Deckel.

Zwischen dem Glasobjektträger und dem Deckel wurde in jedem Experiment ein aus Silikon gegossener Adapter (Abbildung 15) gelegt und die Schrauben der Hybridisierungskammer wurden fest angezogen, so dass die Einzelproben durch die so entstandenen Kanäle auf den einzelnen Subarrays aufgetragen werden konnten, ohne dass die Gefahr einer Kontamination bestand.

3.4.4 Tag-Design

Im anfänglichen Verlauf der Arbeit wurden 29 nicht kreuzreagierende und nicht homologe *tag*-Sequenzen aus dem Genom der λ Bakteriophagen gewählt (Hirschhorn *et al.* 2000)²⁴ und mit einem *spacer* bestehend aus 15 Thymin-Basen (*Oligo-d[T]-Spacer*) und einem Aminolinker am 3'-Ende als komplementäre *tags* (*ctags*) synthetisiert (Tabelle 8). Diese wurden anschließend auf beschichtete *CodeLink*TM Glasobjektträger gespottet.

²³ Dr. Steffen Schmitt ist zur Zeit Leiter der Zytometrie *Core Facility* (FACS) des Deutschen Krebsforschungszentrums, Heidelberg

²⁴ www.genome.wi.mit.edu/publications/SBE-Tags

Tabelle 8: Auswahl der nicht homologen Sequenzen, die als komplementäre *tags* mit einem Aminolinker und einem T-spacer auf dem Glasobjektträger gespottet wurden. ** Beispiel für komplettes *ctag*-Oligo (1): 5'-AATGCTCGGGAAGGCTACTCTTTTTTTTTTTTTTTT-3'NH₂.

Nr.	Name	<i>ctag</i> 5' - 3'	Modification
1	Hybridization Control	cProbeSet00005 **	AATGCTCGGGAAGGCTACTC
2		cProbeSet00007	ATCCCGTGAGTCGATGGTTT
3		cProbeSet00008	CGCACCGCAGTTTGGTCAAT
4		cProbeSet00009	CACGCGGCAGTCGAGTTAAT
5		cProbeSet00010	AGCCCGGTCTCATCGTTGTT
6		cProbeSet00020	ATCGTACTTGGCACTGGAGT
7		cProbeSet00022	GCCCAAATAAGACGTGAGCC
8		cProbeSet00025	TCCGACGCAACAATAGGGCA
9		cProbeSet00026	CCTGCTCGACAAC TAGAAGA
10		cProbeSet00028	CCGCGACCAGAAATTAGATTA
11		cProbeSet00031	GCCTCGTCGAAATTATCACA
12		cProbeSet00033	GCCGAAGCAGACTTAATCAC
13		cProbeSet00036	ACAGGTTTCAGAACTCCTCGAC
14		cProbeSet00044	GCACGTCCCAGATTCGTTAG
15		cProbeSet00050	ACACGCACGGAATAGTATCC
16		cProbeSet00051	ATATGCTCGGAACGTGTGCG
17		cProbeSet00068	ATAGCACGGAATAAGGCC
18		cProbeSet00071	CCATCGCACGAATGTAGTAC
19		cProbeSet00072	CTCATAGACGAAGCGAGGAC
20		cProbeSet00073	TGCACTCACGAAGACGAGAC
21		cProbeSet00076	GCGGCATACGTTCTCAAAT
22		cProbeSet00077	CCAGGCGACGTTCCGAATAT
23		cProbeSet00082	CAGACTCAGCTTTAGGAACG
24		cProbeSet00093	TCGCATAACTAACGGTCAAC
25		cProbeSet00094	CTATGCAACGTAAGCGAACC
26		cProbeSet00097	TCAGCTCGGTAAC TTGCACG
27		cProbeSet00098	ATGAGCTGTCTAACGCGACG
28		cProbeSet00109	ATCGGCCTGTAATCGGATCG
29	Reaction Control	cProbeSet00112	AAGCTGATCTAAGCCGCGAG

Im weiteren Verlauf der Arbeit wurden weitere 30 *tag*-Sequenzen gewählt und in den Versuchsaufbau integriert (Tabelle A-1 im Anhang).

3.4.4.1 Spotten der *ctags*

Die synthetisierten und bei -20°C gelagerten *ctag*-Oligonukleotide wurden vor dem Spotten auf Glasobjektträger gemäß des Herstellerprotokolls in 50 mM Natriumphosphat-Spottinglösung auf 25 µM verdünnt und gemäß der Anordnung auf dem Chip in eine 96-well Microtiterplatte pipettiert. Das Spotten erfolgte mit Hilfe eines DNA-Microarray-Spotters (*OmniGrid Spotter*, GeneMachines, San Carlos, CA, USA) durch das IMSB (Institut für Molekulargenetik, gentechnologische Sicherheitsforschung und Beratung, Johannes Gutenberg Universität Mainz). Das Immobilisieren der Oligonukleotide auf den Objektträgern erfolgte über Nacht bei Raumtemperatur in einer Inkubationskammer mit gesättigter NaCl-Lösung.

3.4.4.2 Blocken der Oberfläche

Nach der zur Immobilisierung notwendigen Inkubationszeit erfolgte das Blocken der Glasträgeroberfläche. Damit keine unerwünschten, unspezifischen Bindungen der gespotteten Oligonukleotide stattfinden konnten, wurde die beschichtete Oberfläche mit NHS (N-hydroxysuccinimidester) deaktiviert. In einer auf 50°C vorgewärmten Blockierlösung (enthält 50 mM Ethanolamin und 0,1 M Tris pH 9) wurden die überschüssigen reaktiven Gruppen inaktiviert. Durch das folgende Waschen mit HPLC-H₂O und 4x SSC, 0,1% SDS Puffer, wurden nicht gebundene Oligonukleotide und der überschüssige Spottingpuffer entfernt. Es war während des kompletten Blockierverfahrens streng darauf zu achten, dass die Glasobjektträger nicht trocken wurden. Erst im letzten Schritt wurden die Objektträger in einer Aufbewahrungsbox platziert, für 3 Min. bei 800 U/min zentrifugiert und für die weitere Verarbeitung bei trockenen Bedingungen aufbewahrt.

3.4.5 Sondendesign

Die SBE-Primer aus der 52-Multiplexreaktion wurden entsprechend für die Multiplexreaktion des Glasarray-Ansatzes verändert. Da keine Längentrennung für die Reaktion erforderlich war, wurde die Primerlänge verkürzt. Die hergestellten Sonden wurden mit einer doppelten Funktion versehen. Sie fungierten als Minisequenzierungsprimer in der SBE-Reaktion, indem sie 3' am polymorphen Nukleotid hybridisierten und, versehen mit der *tag*-Sequenz, als Andockmoleküle für das Demultiplexen auf dem Glasobjektträger dienten. Für die Funktion als Minisequenzierungsprimer wurde eine locuspezifische Region für den jeweiligen SNP angepasst. Richtlinie hierfür war eine Länge zwischen 17 und 66 Nukleotiden, eine mittlere Schmelztemperatur von ca. 59°C und keine inadäquate Primerdimer-Formation. Die Schmelztemperatur und die Dimerstruktur wurden durch ein Oligoanalyseprogramm (*OligoAnalyzer 3.1*, IDT SciTools) ermittelt²⁵.

Um starken Haarnadelstrukturen vorzubeugen wurden komplette Primersequenzen mit der zugehörigen *tag*-Sequenz auf ihre Sekundärstruktur hin untersucht²⁶ (*mfold*-Programm, Michael Zuker).

Eine *tag*-Sequenz wurde am 5'-Ende des Minisequenzierungsprimers synthetisiert. Sie entsprach der komplementären *tag*-Sequenz, die auf der Glasträgeroberfläche gespottet wurde und deren Hilfe im weiteren Reaktionsverlauf die Hybridisierungsreaktion stattfand.

²⁵ <http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/>

²⁶ <http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/dna-form1.cgi>

3.4.6 Probendetektion und Auswertung

3.4.6.1 Sequenzierungsreaktion für Microarrays

Bei den SBE-Primern handelt es sich um bifunktionale Minisequenzierungsprimer mit einer locuspezifischen Sequenz und einer Sequenz analog zum *ctag*. In der Minisequenzierungsreaktion wurde die SNP-spezifische Sequenz durch ein fluoreszenzmarkiertes ddNTP verlängert. Die im ersten Versuchsansatz verfügbaren Farbstoffe waren Cy3 (gekoppelt an ddATP und ddCTP), *TAMRA* bzw. *Rox* (gekoppelt an ddGTP) und Cy5 (gekoppelt an ddUTP [Äquivalent zu ddTTP]). Die markierten ddNTPs wurden von Perkin Elmer erworben. Die Umstellung von *TAMRA* auf *Rox* erfolgte innerhalb des ersten Versuchsansatzes.

Die Versuchsanordnung sah vor, mit den drei Fluoreszenzfarbstoffen die Marker zu genotypisieren, wobei A/C-Polymorphismen hierbei nicht detektiert werden konnten und in diesem Fall die Primer auf die komplementären Stränge umgelegt wurden.

In der zweiten Versuchsanordnung wurde mit folgender Farbstoffzusammenstellung gearbeitet: *TAMRA* (gekoppelt an ddGTP), Cy5 (gekoppelt an ddUTP), *Fluorescein-12* (gekoppelt an ddCTP) und *Texas Red* (gekoppelt an ddATP). Durch das Nutzen von vier Fluoreszenzfarbstoffen konnte jeder Basenaustausch typisiert werden.

Die Reaktion wurde mit 30 Zyklen (96°C für 30 Sek., 60°C für 5 Sek. und 60°C für 10 Sek.) nach folgendem Protokoll durchgeführt (Tabelle 9):

Tabelle 9: Zusammenstellung aller Reagenzien und ihrer Konzentrationen sowie die Einsatzmenge der Reagenzien in einer Minisequenzierungsreaktion für Microarray-Experimente.

Reagenzien	Konzentration	Endkonzentration	15 µl Ansatz
Puffer [x]	10x	1x	1,50 µl
ThermosequenasePol/Therminator Polymerase [U/µl]	10/1 U/µl	3 U	1,30 µl
00112_RC_2_tag [µM]	20 µM	1 µM	0,75 µl
RC2_template [µM]	25 µM	0,35 µM	0,21 µl
MgCl ₂ [µM]	50 µM	2 µM	0,60 µl
Triton [%]	20%	1%	0,75 µl
H ₂ O			1,92 µl
ddNTPs [µM]	100/1000 µM	0,3/0,6 µM	0,21 µl
SBE-Primer [µM]	2,5 µM	0,13 µM	0,80 µl
PCR-Produkt			7,00 µl
Summe der Reagenzien			15,00 µl

3.4.6.2 Hybridisierung der fluoreszenzmarkierten Oligonukleotide

In der Hybridisierungslösung (Tabelle 10) banden die *tag*-Anhänge der SBE-Primer mit den auf den Glasobjektträgern immobilisierten komplementären *tags* (*ctags*). Die Multiplex-Sequenzierungsreaktion wurde mit Hilfe des Arrays in Einzelinformationen geordnet (= „demultiplexen“) und ermöglichte so den nachfolgenden Schritt der Genotypisierung.

Tabelle 10: Zusammenstellung aller Reagenzien und ihrer Konzentrationen sowie die Einsatzmenge der Reagenzien für die Hybridisierungsreaktion.

Reagenzien	Konzentration	Endkonzentration	22 µl Ansatz
SSC [x]	20 x	5 x	5,50 µl
SDS [%]	2%	1%	1,10 µl
BSA [mg/ml]	10 mg/ml	0,1 mg/ml	0,22 µl
Hybr.Contr. [µM]	1 µM	0,005 µM	0,11 µl
H ₂ O			0,67 µl
Miniseq.Produkt [µl]			15,00 µl
Summe der Reagenzien			22,00 µl

Die Hybridisierung erfolgte bei 42°C für 2,5-3 Stunden.

Die Waschschrte im Anschluss an die Hybridisierungsreaktion waren: 4x SSC bei RT für 1 Min., 2-fach in 2x SSC/0,1% SDS Lösung bei 45°C für 5 Min. und 2-fach in 0,2x SSC Lösung bei RT für 1 Min. Anschließend wurden die Glasobjektträger in der Zentrifuge für 15 Min. bei 900 U/min getrocknet.

3.4.6.3 Detektion der Signale

Die Detektion der Marker nach der Hybridisierungsreaktion erfolgte ebenfalls in zwei unterschiedlichen Versuchsansätzen. Zunächst erfolgte die Typisierung an einem konfokalen Fluoreszenzscanner (*Affymetrix® 428™ Array Laser-Scanner*, Affymetrix, Santa Clara, CA, USA, jetzt *Eurofins MWG Operon*), welcher zwei Laser zur Anregung von Cy3- und Cy5-Farbstoffen besaß. Der Scanner war mit drei Emissionsfiltern ausgestattet (551 nm, 570 nm und 665 nm). Diese Filterkombination reichte jedoch nicht aus, um die Spektren der Farbstoffe Cy3 und *TAMRA* zu differenzieren, so dass das *TAMRA*-gekoppelte Nukleotid durch ein *Rox*-gekoppeltes Nukleotid ersetzt wurde. Hierzu erforderte es zusätzlicher Emissionsfilter, die in das Gerät integriert werden konnten. Der Verkauf der Lizenzrechte des Laserscanners an die damalige *MWG Biotech AG* hatte jedoch zur Folge, dass keine Weiterentwicklung der Gerätesoftware stattfand und das ursprünglich für sechs Emissionsfilter entwickelte Steuerrad des Geräts nur drei Filterpositionen bedienen konnte. Eine weitere Beschränkung der Farbstoffauswahl bestand in den Patentrechten für die Farbstoffe und damit für die angebotenen Farbstoffkopplungen. Eine Lösung hierfür war ein Umstieg auf ein Vierfarb-Laserscanner Gerät. Der Microarray-Scanner *GeneTAC™ LSIV*, *Genomic Solutions* wurde vom Institut für Genetik an der Universität zu Köln für die Ermittlung der Daten zur Verfügung gestellt (mit freundlicher Erlaubnis und Unterstützung vom Prof. Tautz²⁷ und seiner Arbeitsgruppe). Das Gerät besaß drei interne Laser für die Anregungswellenlängen zwischen 532 und 633 nm und einen energiereichen externen Laser für den Wellenlängenbereich von 488 nm. Unterschiedliche Filtereinsätze konnten dem

²⁷ Seit 2007 ist Diethard Tautz Direktor am Max-Planck-Institut für Evolutionsbiologie.

Emissionsbereich angepasst werden (Tabelle 11) Die gescannten Dateien wurden im *.tiff* Format ausgegeben und zur weiteren Quantifizierung verwendet.

Tabelle 11: Verwendeten Fluoreszenzfarbstoffe und die physikalischen Eigenschaften des *GeneTAC™ LSIV Microarray*-Scanners.

	Fluoreszenzfarbstoffe	Emission/ Absorption	Anregungswellenlänge des Lasers	Emissionsfilter mit dem Durchlassbereich
G	TAMRA	546 nm/576 nm	532 nm Solid State laser (green, 20 mW)	560/25 Band Pass
T	Cy5	649 nm/670 nm	633 nm Helium Neon laser (red 15 mW)	675/25 Band Pass
A	Texas Red	583 nm/603 nm	594 nm Helium Neon laser (yellow, 2 mW)	615/25 Band Pass
C	Fluorescein-12	494 nm/517 nm	488 nm Argon Ion laser (blue, 10 to 20 mW)	512/25 Band Pass

3.4.6.4 Quantifizierung der Signalintensitäten

Die Quantifizierung der Signalintensitäten der Spots erfolgte mit der *ImaGene™ 5.5 Standard Software* (BioDiscovery).

Die Darstellung der detektierten Scanbilder erfolgte in Falschfarben. Jedem Farbstoff wurde dabei eine Farbe zugeordnet (*TAMRA*/grün, *Cy5*/rot, *Texas Red*/gelb, Fluorescein-12/blau). In der Folgeauswertung wurde über jedes Subarray ein Raster (*grid*) gelegt und die einzelnen Spots wurden in der Größe und der genauen Position ausgerichtet. Anschließend erfolgten die Quantifizierung und der Export der Daten in Form von Zahlenwerten. Bei den Export-Dateien handelte es sich um vier **.txt*- Dateien. Jede beinhaltete die Signalintensitätswerte, Hintergrundintensitätswerte sowie die jeweiligen metrischen Parameter (Mittelwert, arithmetisches Mittel etc.).

Vor der weiterführenden Analyse wurden die Textdokumente umgewandelt, indem alle Dezimal-Punktzeichen gegen Kommata ersetzt wurden. Die Dateien konnten dann in ein *Excel*-Dokument importiert werden. Dies erfolgte durch ein *Excel*-Makro (geschrieben von Matthias Herrgen, Institut für Anthropologie). In vier *Excel*-Tabellenblättern wurden die Daten der vier Farbkanäle eingefügt. Es erfolgte der Abzug der ermittelten mittleren Hintergrundintensitätswerte von den mittleren Signalwerten. Anschließend wurden die „*Signal Mean*“-Werte aller *spots* in eine *Master*-Tabelle importiert. Daraus wurden nur die zur Analyse notwendigen Werte extrahiert in eine Export-Tabelle im separaten Tabellenblatt überführt. Die Export-Datei besaß das für den Export in den folgenden Analyseschritt vordefinierte Format.

3.4.6.5 Normalisierung der Messdaten

Die erste Normalisierung der Daten erfolgte bereits am Laserscanner. Hierzu wurde die Farbwiedergabe des Ablesevorgangs während des Vorscans invertiert. Nach dieser

Visualisierungsmöglichkeit wurden die Laserintensitäten eines jeden Kanals abhängig von der Signalintensität justiert. Diese Einstellung erfolgte unabhängig von der fixen Einstellung der Photovervielfacherröhren (engl. *photomultiplier tube* - PMT). Die Angleichung wurde akzeptiert, wenn kein Spot übersättigt erschien (weißer Spot) und die Signale der jeweiligen Fluorophore möglichst gleichmäßig waren. Dabei war es von Vorteil sich an der Sättigung der Reaktionskontrolle zu orientieren.

Ein weiterer Normalisierungsschritt ermöglichte eine vergleichende Analyse der Daten. Mit dem nachfolgend beschriebenen Verfahren sollten mögliche Artefakte und systematische Fehler in den Werten eliminiert werden. Die Normalisierung wurde für jede Probe auf dem Array einzeln durchgeführt.

Ähnlich dem Prinzip der globalen Normalisierung wurde aus den Intensitätswerten der Reaktionskontrolle für jeden Farbstoff das arithmetische Mittel gebildet. Da die Reaktionskontrolle in jedem Array in gleicher Qualität und Quantität vorhanden war, wurde sie zur Referenzprobe in jedem Experiment. Die Quotienten aus dem arithmetischen Mittel der Reaktionskontrolle jedes Farbkanals und dem Mittelwert der Reaktionskontrolle mit dem höchsten Intensitätswert bildeten einen Skalierungsfaktor.

$$F_{skal} = \frac{\bar{X}_{RC_n}}{\bar{X}_{RC_{max}}}$$

Mit dem errechneten Skalierungsfaktor für jeden der vier Farbkanäle wurden die Signalintensitäten multipliziert, um diese auf ein einheitliches Skalierungsniveau zu bringen. Dieser Schritt wurde ebenfalls durch das vorab beschriebene *Excel*-Macrodokument durchgeführt und konnte parallel zur Quantifizierung der Daten stattfinden.

Die Datenanalyse fand aus Vergleichsgründen sowohl mit als auch ohne eine Normalisierung statt. Als eine alternative Bezeichnung der Normalisierung wird der Name der Faktorkorrektur verwendet.

Die Proben wurden doppelt typisiert. Die Typisierungsdurchgänge wurden in der Analyse für den ersten bzw. zweiten Arbeitsdurchgang als **Serie 1** und **Serie 2** bezeichnet. Die einzelnen Proben bekamen zur Unterscheidung die Buchstaben **a** bzw. **b** (Bsp.: L34a, L34b). Eine Analyse der Daten mit und ohne die Faktorkorrektur erfolgte jeweils für die beiden Analysedurchgänge.

3.4.7 Visualisierung der Daten und Genotypisierung

Die endgültige Genotypisierung wurde mit der *SNPSnapper Software* durchgeführt. Die Bereitstellung der Software erfolgte durch Juha Saharinen, *National Public Health Institute Helsinki*²⁸.

²⁸ www.Bioinfo.helsinki.fi/snpnapper/start.htm

Bevor die Genotypisierung durchgeführt werden konnte, waren weitere Formatierungsschritte zur Visualisierung der Daten notwendig. Die Tabellen mit den quantifizierten Intensitätswerten wurden aus dem *Excel*-Format zurück in ein Textformat umgewandelt (Abbildung 16 a). Dabei war zu beachten, dass die Zuordnung der Intensitätswerte des potentiellen Polymorphismus stets durch eine bestimmte „Codierung“ eingehalten wurde, die es der *SNPSnapper-Software* möglich machte, die Zuordnung der jeweiligen Farbkanäle vorzunehmen.

Diese war wie folgt aufgebaut: *Hyb121***SNP1_1=AG* und enthielt die Information über die fortlaufende Hybridisierungsnummer, den SNP-Namen und die Basenkombination des Polymorphismus.

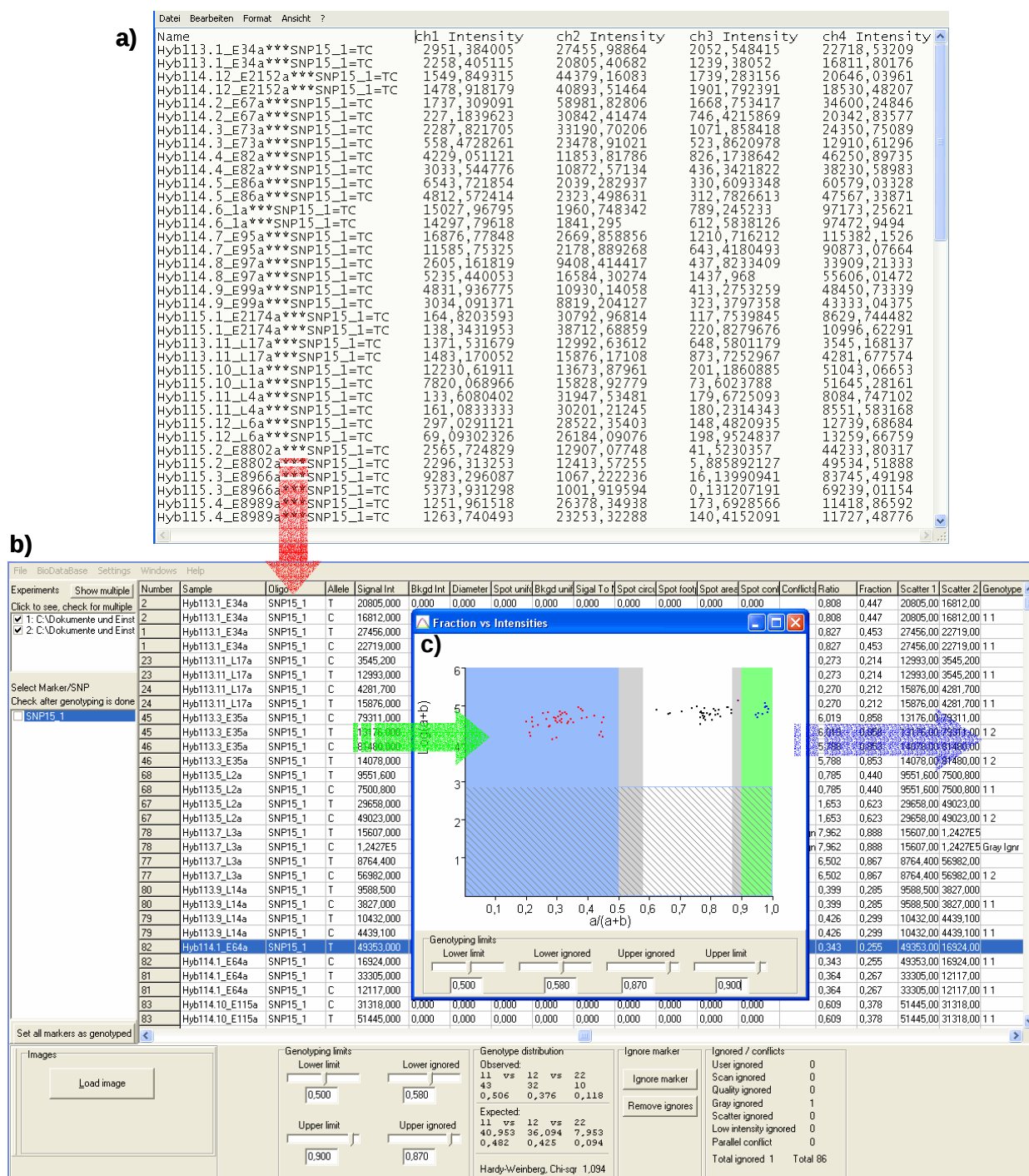


Abbildung 16: Verkürzte Darstellung des Datenexportweges und der Genotypisierung mittels *SNPSnapper-Software*. a) Die Datenquantifizierung folgt einer Umwandlungskette bis zum vorgegebenen Format als Textdatei. b) Diese wird durch die Textcodierung in der ersten Spalte in die *SNPSnapper-Software* eingefügt, c) dort werden aus den Signalintensitäten beider Farbstoffe das Verhältnis und der Bruch berechnet und in einer Logarithmusfunktion gegen den Intensitätsquotienten aufgetragen. Dies führt zu einer Clusterteilung und zur eigentlichen Genotypisierung (blauer Pfeil).

Die Darstellung konnte sowohl pro Array als auch pro SNP erfolgen. Zur Umsortierung der Daten aus den Einzelexperimenten wurden alle Quantifizierungstabellen in ein *Excel*-Datenblatt exportiert und mittels der *Pivot*-Tabellenfunktion wurden anschließend die Daten der SNPs in einzelne Tabellen extrahiert. Die Gruppierung der SNP-Daten führte zu einer Clusterdarstellung im Streudiagramm (Abbildung 16 c). Die Einstellungen der Flächen für

homozygote, bzw. heterozygote Resultate konnten mittels Wertefenster eingestellt werden. Eine graue Fläche markierte ambivalente Signalpunkte. Eine zusätzliche Markierung möglicher Fehlsignale durch eine schraffierte Fläche erlaubte deren Eliminierung aus der Betrachtung. Nach Anpassung der Plots erfolgte der Export der Daten in Tabellen und deren Gruppierung.

3.4.8 Klassifizierung der Daten mittels des *Silhouette-Koeffizienten*

Eine Klassifizierung der Cluster erfolgte mittels des *ClusterA*-Programms²⁹. Das Programm basiert auf einem Algorithmus zur Berechnung der Distanzen im Signalplot (Rousseeuw, 1987). Die *Silhouette*-Kalkulation berechnet die durchschnittlichen Entfernungen der Datenpunkte im internen Cluster zueinander ($a_{(i)}$) und bestimmt deren Position in Bezug auf die durchschnittliche Entfernung aller Datenpunkte zum „nächstmöglichen“ Cluster ($b_{(i)}$).

$$s_{(i)} = \frac{b_{(i)} - a_{(i)}}{\max\{a_{(i)}, b_{(i)}\}},$$

wobei $b_{(i)} = \min\{b_{1(i)}, b_{2(i)}\}$.

Die Klassifikation erlaubt den Rückschluss auf die Qualität des Versuchsansatzes. Die Rangbreite erstreckt sich von +1,0 bei sehr einheitlichen Clustern bis -1,0 für Cluster mit einem unausgewogenen Verhältnis, bzw. mit willkürlicher Verteilung der Signalpunkte.

Obwohl der Koeffizient für beide Dimensionen der graphischen Plotdarstellung berechnet werden kann, wurden hier die Distanzen entlang der x-Achse berücksichtigt, da der Quotient,

berechnet aus $\frac{Signal_{Allel_2}}{(Signal_{Allel_1} + Signal_{Allel_2})}$ den Genotypen determiniert. Die Dimension der y-

Achse, den der Logarithmus der Summe aller Signale darstellt, wird genutzt um den Grenzwert der Signalstärke festzustellen.

Signalpunkte unterhalb dieses Wertes wurden bei der Koeffizientenberechnung nicht berücksichtigt. Um den *Silhouette*-Wert nicht zu verzerren, wurden diese Werte aus der entsprechenden Tabelle entfernt.

Die Darstellung der Werte erfolgte über ein Verbundliniendiagramm, in welches die Koeffizientenwerte aller Vergleichsanalysen aufgetragen wurden (Abbildung 17).

²⁹ Programm ist auf der Homepage der medizinischen Fakultät der Universität Uppsala zugänglich:
<http://www.medsci.uu.se/molmed/software.htm>

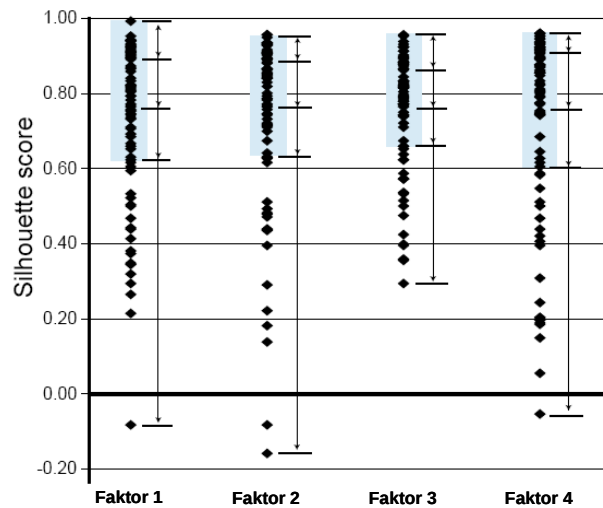


Abbildung 17: Beispiel eines Verbundliniendiagramms. Die *Silhouette*-Werte befinden sich auf der *y*-Achse. Entlang der *x*-Achse gruppieren sich die Versuchsreihen. Jede Raute signalisiert einen Wertepunkt der Untersuchung. Innerhalb der blauen Balken sind die 75% der Plots mit dem höchsten *Silhouette*-Wert platziert. (Verändert nach Lovmar *et al.* 2005).

4 Ergebnisse

Für die Entwicklung und Etablierung neuer SNP-Typisierungsmethoden wurde im Rahmen des *SNPforID* EU-Projekts eine große Reihe von Versuchen unternommen, von denen die Validierung der autosomalen 23 SNP Multiplexreaktion im nachfolgenden Kapitel 4.1 beschrieben wird. Die Ergebnisse der weiterführenden Entwicklung der Multiplexreaktion auf den Untersuchungsumfang von 52 SNPs wird in einem Teil der Diskussion zusammenfassend dargestellt und erörtert (Kapitel 5.2). Das Ziel des Projekts war, eine Methode zu entwickeln, die unterstützend für forensische Zwecke durch jedes Labor, welches mit einem Kapillarelektrophorese-Gerät ausgestattet ist, genutzt werden kann.

Der Hauptteil der vorliegenden Arbeit umfasst die Entwicklung einer hybridisierungsbasierten Typisierungsmethode auf der Microarray-Plattform. Im Folgenden wird gezeigt, dass die Methode klare physikalisch begründete Grenzen bezüglich des Hochdurchsatzes sowie der Sensitivität mit sich bringt. Letzteres ist jedoch essentiell im Hinblick auf die forensische Anwendbarkeit. Etabliert wurde ein Assay, das 33 der vorab ausgewählten 52 SNP-Marker reproduzierbar typisieren kann und welches in der Lage ist, zwölf Proben parallel unter gleichen Konditionen zu analysieren. Die Methode erscheint durch ihre Handhabung, die bis zur Detektion fast vollständig ohne komplizierte elektronische Geräte auskommt, robust und einfach bedienbar. Sie birgt ihre Schwierigkeiten jedoch in der Konzeption und der Analyse sowie in der Sensitivität der molekularen Interaktion der einzelnen Komponenten. Die erforderliche Abstimmung der Reaktionsbestandteile sowie die Validierung der Methode sind in den Kapiteln 4.2 bis 4.2.3.2 dargestellt.

4.1 Vorvalidierung der 23 SNP Multiplexreaktion

4.1.1 Probenaufbereitung mittels WGA

Das für die Validierung genutzte Probenkollektiv umfasste neben Proben deutscher Individuen drei asiatische Stichproben (Kapitel 3.2.1). Nach der Feststellung der geringen Menge und Qualität der japanischen Proben wurde die viel versprechende Methode der *Whole Genome Amplification* dazu genutzt, die Proben weitestgehend wiederherzustellen. Zehn von vierzig der Proben japanischen Ursprungs wurden zunächst nach dem Standardprotokoll amplifiziert (Kapitel 3.2.2). Aus 1 ng/µl DNA, eingesetzt in die WGA-Reaktion, wurden Konzentrationen zwischen 320 und 437 ng/µl DNA erzielt.

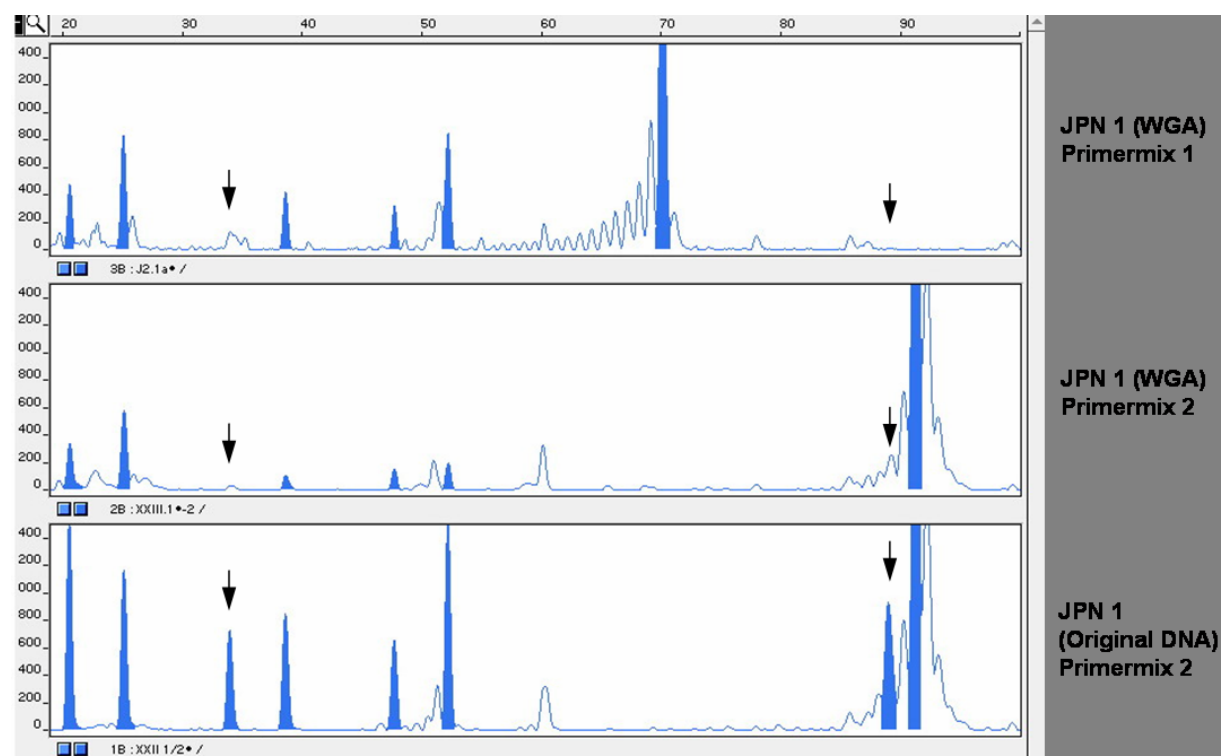


Abbildung 18: Darstellung von drei Analysen (jeweils des G-Nukleotids mit dem blau emittierenden Fluoreszenzfarbstoff dR110) der Probe JPN1 aus dem japanischen Kollektiv. Oberes und mittleres Elektropherogramm zeigt die WGA-analysierten Proben. Bei der dritten Probe handelt es sich um Original-DNA. Pfeile in den beiden oberen Proben deuten auf fehlende Allele hin. Entsprechende Pfeile in der unteren Darstellung markieren die Allele. Der Primer an Position 70 (oben) wurde ersetzt und auf insgesamt 90 nt verlängert.

Für die PCR der zehn ausgewählten Proben wurden 20 ng/µl der WGA-amplifizierten DNA eingesetzt. In der Auswertung der WGA-Proben konnten keine höheren Signalintensitäten erzielt werden. Die Allele waren im Vergleich zu Proben, die mit Original-DNA amplifiziert worden waren stark unbalanciert und zeigten eine deutliche Tendenz zum *allelic dropout* (Abbildung 18, Pfeile). Anfängliche Probleme mit degradierten Primern für die SNPs 16 und 18, die ein höheres Hintergrundrauschen verursachten, konnten mit neuen, verlängerten Primern behoben werden (Abbildung 18).

Im Anschluss an die Versuche, WGA-Proben für die folgenden Untersuchungen des japanischen Kollektivs zu etablieren, wurden zunächst zehn Originalproben zufrieden stellend amplifiziert. Insgesamt 40 Individuen, bei denen eine ausreichende DNA-Konzentration messbar war, wurden in der Validierungsstudie mit 23 SNPs verwendet. In der weiterführenden Methodenetablierung der 52 SNP Multiplexreaktion konnten jedoch nur in sieben japanischen DNA-Proben alle Marker erfolgreich amplifiziert werden. WGA-amplifizierte Proben japanischer Individuen wurden zur Analyse Y-chromosomaler SNPs eingesetzt und konnten dort gute Ergebnisse erzielen (nachzulesen in der Dissertation von Balogh, 2009).

4.1.2 Auswertung der Populationsdaten

4.1.2.1 Typisierungserfolg

Beim Einsatz der für diesen Teil der Arbeit verwendeten *SNaPshot*TM-Methode betrug der durchschnittliche Typisierungserfolg innerhalb der asiatischen Populationen 99,87%. Die Profile der deutschen Individuen konnten vollständig typisiert werden (Tabelle 12). Die höchste Anzahl fehlender Genotypen wiesen die thailändischen Proben auf. Bei der nochmaligen Kapillarelektrophorese kam es zudem zu einer Abweichung in einem Genotyp, was den Typisierungserfolg in der Übereinstimmung mit den Replikaten auf 99,91% senkte. Bei dem abweichenden Polymorphismus handelte es sich um den Locus rs1493232 (A18), welcher eine stärkere Analyseproblematik aufgrund eines hohen Schulterpeaks aufwies (siehe auch Abbildung 18, rechter Pfeil). Eine manuelle Nacheditierung der Ergebnisse erfolgte in Fällen, bei denen die Peakhöhe abweichend war oder einer der Peaks zu klein ausfiel. Die Korrektur erfolgte nach einheitlichen Kriterien (siehe Anhang Tabelle A-1).

Tabelle 12: Zusammengefasste Ergebnisse der Typisierung von vier Kollektiven mittels 23 SNP Multiplexreaktion.

Population	N	Σ typisierter SNPs	Σ Genotyp- paare	Σ typisierter Genotypen	Σ ausgefallener Genotypen	Σ abweichender Genotypen	Σ editierter Ergebnisse	% Typisierungs- erfolg	% Übereinstim- mung mit Replikaten
Deutsche	64	23	1472	1472	0	0	2	100	100
Japaner	40	23	920	919	1	0	3	99,89	100
Chinesen	57	23	1311	1311	0	0	4	100	100
Thailänder	49	23	1127	1124	3	1	9	99,73	99,91

4.1.2.2 Allelverteilung in vier Populationen

Die Allelfrequenzen der aus den oben beschriebenen vier Populationen stammenden 210 Proben wurden ermittelt und den publizierten Frequenzen der Vergleichspopulationen gegenübergestellt (Tabelle A-3 im Anhang). Aufgrund fehlender Daten für die SNPs A21 und A23 bei Mitteleuropäern und die SNPs A15, A21 und A23 für die Han Chinesen wurden nachträglich Daten aus dem *SNPforID browser*³⁰ hinzugezogen. Diese wurden mit der 52 SNP Multiplexreaktion zu einem späteren Zeitpunkt ermittelt und standen erst nach der abgeschlossenen Validierung zur Verfügung. Sie ergänzen in der vergleichenden Darstellung der Allelfrequenzen die fehlenden Werte.

³⁰ Die Frequenzen wurden aus der *SNPforID* Datenbank unter der Adresse: <http://spsmart.cesga.es/snpforid.php?dataSet=snpforid52> entnommen und basieren auf den Daten für die mitteleuropäische Bevölkerung (n=809) und für Han Chinesen (n=44).

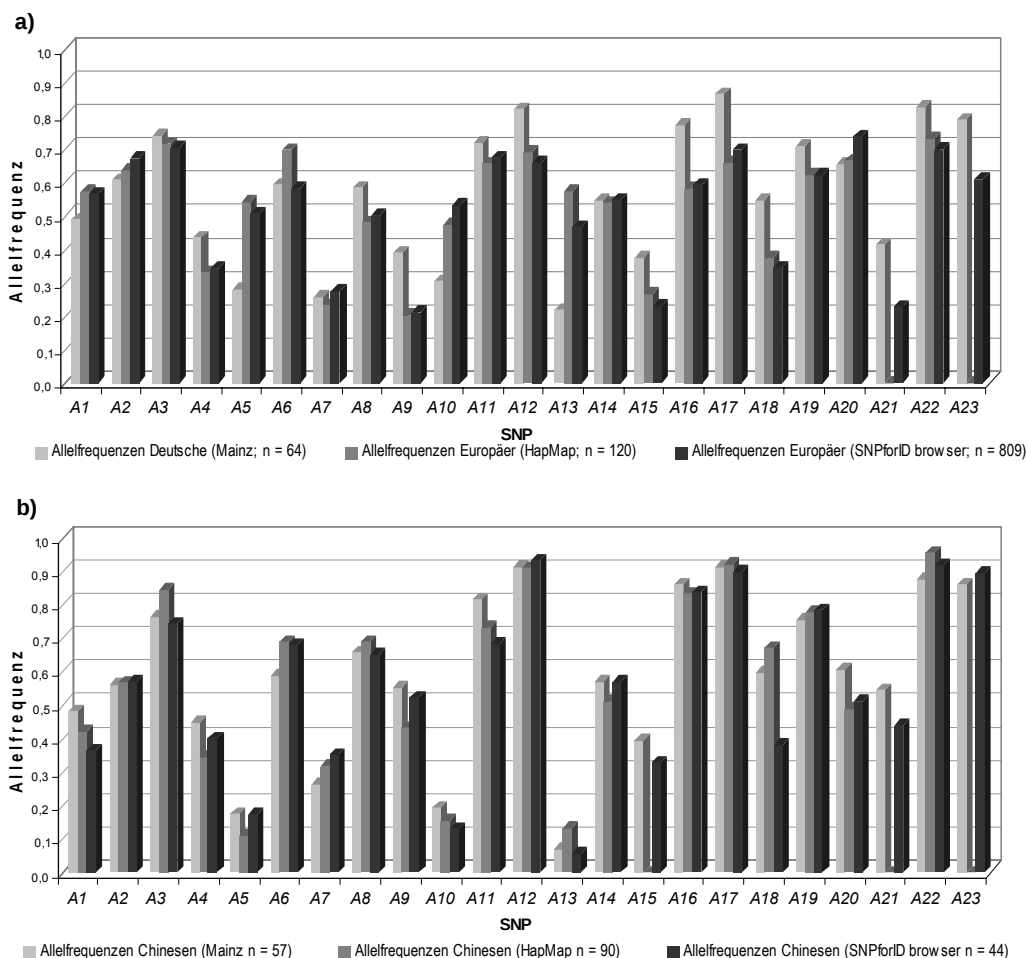


Diagramm 1: Balkendiagramm der Allelfrequenzen aus der vorliegenden Arbeit (hellgrau) und der Allelfrequenzen der *HapMap* Datenbank (mittelgrau). In dunkelgrau sind die Daten aus der *SNPforID* Datenbank dargestellt. Diagramm a) mitteleuropäische Proben vs. Proben deutscher Bevölkerung, Diagramm b) chinesische Proben.

Für die europäischen Individuen lag die Abweichung der Allelfrequenzen im Bereich zwischen 0,005 (Marker A14) für die niedrigste und 0,356 (Marker A13) für die höchste Differenz. Für die beiden Marker (A21 und A23), für die es keine Vergleichswerte gab, konnte eine Abweichung zwischen 0,178 und 0,186 nach einem Vergleich mit den Datenbankwerten des *SNPforID browsers* ermittelt werden.

Die Werte der chinesischen Proben zeigten eine deutliche geringere Spannweite der Frequenzunterschiede. Sie lagen zwischen 0,001 (Marker A12) und 0,120 (Marker A9). Die Abweichungen der Werte für die fehlenden Marker lagen im mittleren Bereich zwischen 0,35 (A23), 0,65 (A15) und 0,105 (A21). Eine größere Abweichung konnte zwischen den Werten der Vergleichspopulationen und den der Datenbank des *SNPforID browsers* festgestellt werden. Sie lag für die Allelfrequenz des Mainzer Kollektivs bei 0,289.

Zur besseren Darstellung der Vergleichbarkeit der errechneten und publizierten Frequenzen wurden Korrelationskoeffizienten (r) nach Pearson berechnet (Tabelle 13).

Tabelle 13: Bestimmung des Korrelationskoeffizienten für die Populationsgruppen in der untersuchten 23 SNP Multiplexreaktion gegen die Populationen aus der Datenbank.

Korrelationskoeffizient (r)	Populationsgruppen Projekt vs. Datenbank
0,682*	Deutsche vs. Europäer (<i>HapMap</i>)
0,965*	Japaner, Thailänder und Chinesen, vs. Asiaten (<i>HapMap</i>)

* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Für die Beurteilung der Abweichungen zwischen den erwarteten und beobachteten Heterozygotenraten (H_{exp} ; H_{obs}) wurden die Daten der analysierten Populationen für die 23 Marker in einem Boxplot gruppiert dargestellt (Diagramm 2). Auf diese Weise war die Streuung der detektierten Heterozygotenwerte zwischen den Populationen anschaulich zu erkennen. Der niedrigste sowie der höchste gemessene Wert traten in der thailändischen Stichprobe auf. Für den SNP Marker rs740910 (A17) lag der Minimalwert bei 0,061. Der Maximalwert von 0,674 wurde durch den SNP rs1493232 (A18) verursacht. Beide Werte sind im Diagramm durch jeweils nach unten oder nach oben von den Boxen abgehende Linien dargestellt. Die größte Streuung der Werte wurde bei der chinesischen Stichprobe beobachtet. Durch die Position der Mediane (dunkle Balken), werden die Unterschiede zwischen den erwarteten und den beobachteten Heterozygotenwerte dargestellt. Hier zeigte sich die größte Differenz innerhalb der thailändischen Stichprobe. Die Werte variierten um 0,08 Datenpunkte und erreichten mit 0,367 den niedrigsten Zentralwert.

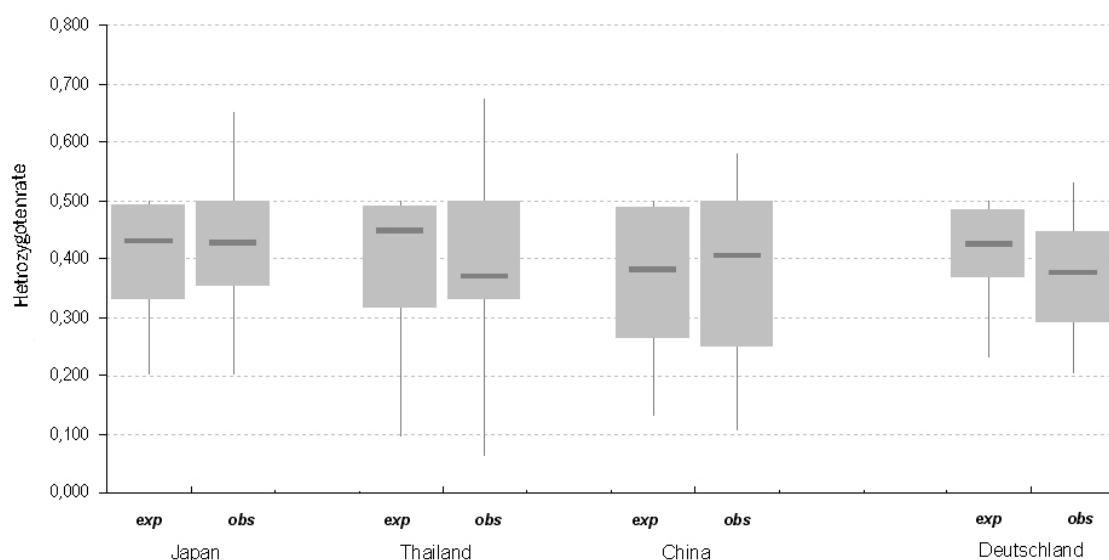


Diagramm 2: Boxplotdarstellung der Heterozygotenraten analysierter Stichproben der vier Populationen. Die Boxen gruppieren sich in erwartete (*exp*) und beobachtete (*obs*) Heterozygotenraten. Aufgetragen wurden Median, unteres und oberes Quartil sowie der Minimal- und der Maximalwert der 23 SNPs.

Die Werte innerhalb der deutschen Stichprobe zeigten eine geringere Streuung und wiesen deutlich weniger Ausreißer auf, als die Stichproben asiatischer Gruppen. Die beobachteten

Heterozygotenraten lagen zwischen 0,203 für den SNP A17 und dem Höchstwert von 0,531 für den SNP A14. Die Abweichungen zwischen den Heterozygotenraten H_{exp} ; H_{obs} waren jedoch auch innerhalb dieser Stichprobe stark und lagen für den Marker rs2016276 (A15) bei 0,157 Datenpunkten (Tabelle A-4 im Anhang).

4.1.2.3 Hardy-Weinberg-Gleichgewicht

Signifikante Abweichungen der Genotypen vom Hardy-Weinberg-Gleichgewicht wurden für acht SNPs festgestellt. Nach Bonferroni-Korrektur lagen diese jedoch alle im nicht-signifikanten Bereich. Die Abweichungen traten nie in mehreren Populationen für den gleichen SNP auf. Lediglich die Genotypenverteilung der Japaner erfüllte in allen Systemen die Erwartungen des Gleichgewichts.

Tabelle 14: Loci mit einer mit dem Exakten Test errechneten signifikanten Abweichung vom HWG (vor der Bonferroni-Korrektur).

SNP	Locus	Thailänder	Chinesen	Deutsche
A1	rs1490413	P < 0.05	P < 0.05	P < 0.05
A5	rs717302			
A7	rs917118	P < 0.05	P < 0.05	P < 0.05
A8	rs763869			
A11	rs901398	P < 0.05	P < 0.05	P < 0.05
A15	rs2016276			
A18	rs1493232	P < 0.05	P < 0.05	P < 0.05
A20	rs1031825	P < 0.05	P < 0.05	P < 0.05

Die abweichende Verteilung der Genotyphäufigkeiten in den acht in Tabelle 14 aufgeführten Markern konnte mit Hilfe der im Exakten Test berechneten *odds* beurteilt werden. Lagen diese $<0,0$ wurden zu wenige der erwarteten Genotypen typisiert. Bis auf die Marker rs1493232 (A18) und rs1031825 (A20), die beide in der thailändischen Population abweichende Genotypverteilung lieferten und zu viele Homozygote aufwiesen, lag die Ursache der Abweichungen bei den Markern A1 und A8 in einer deutlich höheren Typisierung heterozygoter Proben. Die chinesischen sowie deutschen Proben wiesen in den betroffenen Loci erhöhte Anteile homozygoter Individuen auf (Tabelle A-6).

4.1.2.4 F-Statistiken

Die Varianzuntersuchung aller typisierten Marker erfolgte in der deutschen Population mittels des F_{IS} -Index. Für die drei asiatischen Populationen wurden F_{ST} -Werte berechnet. Die Verteilung der F_{IS} -Werte lag innerhalb der deutschen Population zwischen -0,18 und 0,39, innerhalb der asiatischen Population zwischen -0,01 und 0,07. Vergleich man die Varianzen aller vier Populationen, waren die Werte unverändert zu den Werten der asiatischen Proben (Tabelle A-7 im Anhang).

Tabelle 15: Analyse der Varianzen beider Populationsgruppen für die untersuchten 23 Marker. Sie basiert auf der Verteilung der F_{IS} -Werte für die deutsche Population und der Verteilung der F_{ST} -Werte für die asiatischen Populationen.

a)

Deutsche	
Anzahl der Marker	23
Mittelwert	0,11
Median	0,107
Std. Abweichung	0,137
Minimum	-0,18
Maximum	0,39
Schiefte	0,129

b)

Asiaten	
Anzahl der Marker	23
Mittelwert	0,008
Median	0,002
Std. Abweichung	0,019
Minimum	-0,01
Maximum	0,07
Schiefte	1,519

4.1.2.5 AMOVA

Neben den F-Statistiken wurde die molekulare Varianz anhand der Φ -Statistiken berechnet. Diese Analyse ermöglicht zusätzlich, direkt Populationsunterschiede auf der Basis molekularer Daten zu berechnen.

Hierbei werden mögliche hierarchische Ebenen aus den Werten für die Varianzkomponenten berechnet. Die Gesamtvarianz wird prozentual den geographischen Einheiten zugeordnet. Sie beschreibt dabei das Verhältnis der Differenzierung zwischen den Gruppen (V_a und Φ_{CT}), zwischen den Populationen innerhalb der Gruppen (V_b und Φ_{SC}) sowie die Gesamtdifferenzierung der Populationen (V_c und Φ_{ST}). Die Varianz ermittelt sich aus der Quadratsumme der Abweichungen der Messwerte von den Mittelwerten, dividiert durch die Anzahl der Freiheitsgrade (n-1).

Tabelle 16: Ergebnisse der molekularen Varianzanalyse (AMOVA) zwischen drei asiatischen Populationen und der deutschen Population.

Abweichung	FG	Quadratsumme	Varianz	Varianzanteil (%)	Φ -Werte	p -Werte
innerhalb der Gruppen	1	8,680	0,00240	0,05	$\Phi_{CT}=0,00052$	$\geq 0,417$
zwischen den Populationen						
innerhalb der Gruppen	2	15,094	0,03075	0,67	$\Phi_{SC}=0,00666$	$\geq 0,004$
innerhalb der Populationen	411 - 416	1905,263	4,58527	99,28	$\Phi_{ST}=0,00718$	$\leq 0,000$

Mittels der Varianzanalyse konnte eine sehr geringe genetische Strukturierung nachgewiesen werden. Lediglich 0,05% der Gesamtvarianz erklärten die Differenzierung innerhalb der Gruppen (Tabelle 16). Die genetische Varianz innerhalb der Populationen war mit 99,28% am höchsten. Die Berechnung der genetischen Distanz ergab für alle hierarchischen Ebenen sehr geringe Werte, wobei der Maximalwert für Φ_{ST} bei 0,00718 lag. Diese und die Differenzierung zwischen den Populationen innerhalb der vorgegebenen Gruppen (Chinesen, Japaner, Thailänder vs. Deutsche, 0,67%) waren signifikant ($p < 0,05$), während die Varianz innerhalb der Gruppen keine Signifikanz aufwies.

In Ergänzung zu den populationsgenetischen Statistiken wurde die kumulative Ausschlusschance für die Populationen errechnet. Diese lag bei allen Populationen bei über 99,999%.

4.2 Microarray-Analyse der 33 SNP Multiplexreaktion

4.2.1 Entwicklung und Optimierung der Probendetektion mittels Microarray-Analyse

In der vorliegenden Arbeit lag der Schwerpunkt auf der Entwicklung eines multiplexfähigen Ansatzes basierend auf dem Microarray-Prinzip. Bei den ersten theoretischen Überlegungen zu Beginn der Arbeit waren viele der Faktoren noch nicht bekannt, die zur späteren Umgestaltung der eingesetzten Methoden geführt haben. Im Verlauf dieses Kapitels werden deshalb die Schritte von der Entwicklung bis zum fertigen Ansatz dargestellt. Zu jedem erreichten Etablierungsschritt waren zahlreiche Versuche notwendig. Dargestellt werden im Folgenden die wesentlichen Versuche, die zum Ausschluss eines experimentellen Ansatzes führten sowie jene, die zu den besten Versuchsbedingungen geführt haben und der Weiterentwicklung der Methode dienten.

Die Optimierung und Anpassung der Reaktionsbedingungen erfolgte für folgende Parameter:

- Sondenauswahl und Primerdesign
- Umfang der Multiplexreaktion
- Farbstoffauswahl und -zusammensetzung
- Primerkonzentration
- ddNTP-Konzentration
- Sequenzierenzyme
- Hybridisierungstemperatur
- Behandlung der Glasobjektträger nach der Hybridisierung

Die PCR-Bedingungen wurden gegenüber der vorab beschriebenen Methode nicht verändert. Es zeigte sich jedoch, dass die Allelproportionen mit der Menge des jeweiligen PCR-Produkts korrelierten. Um dieses zu beeinflussen, wurden die Konzentrationen in den PCR-Primermixen neu eingestellt und für einzelne Loci erhöht.

Soweit nicht anders beschrieben, beziehen sich die folgenden Angaben auf die im Methodenteil beschriebenen Versuchsabläufe. Angeführte Signalintensitäten beziehen sich immer auf die errechneten Mittelwerte aus den Spotreplikaten. Abbildungen der Arrays sind in Echtfarben dargestellt und entsprechen dem Farbspektrum der zugehörigen Fluorophore. Bild Darstellungen der einzelnen Kanäle werden in Graustufen wiedergegeben.

4.2.1.1 Tag- und Sondendesign

Die Auswahl der als Auffangmoleküle fungierenden *tags* erfolgte sukzessive aus der Liste der nicht homologen *tag*-Sequenzen (Hirschhorn *et al.* 2000, Kap. 3.4.4). Sie wurden in

komplementärer Sequenzabfolge in der Reihenfolge der Liste (Tabelle 8) auf die Glasobjektträger gespottet. Die Minisequenzierungsprimer (Sonden) wurden unter Berücksichtigung der thermodynamischen Vorgaben manuell ausgewählt und mittels Computersoftware überprüft und angepasst. Intra- und intermolekulare Basenpaarungen sollten weitestgehend vermieden werden.

Bei einigen Minisequenzierungsprimern war keine optimale Anpassung an die physikalischen Vorgaben aufgrund nachteiliger, SNP-naher Sequenzabfolge möglich. Eine potenzielle Rückfaltung der Primer mit der 5'-verbundenen *tag*-Sequenz musste aufgrund der erreichten Länge von >38 *nt* zusätzlich überprüft werden, um internen Baseneinbau und damit eine unspezifische Signalbildung zu vermeiden.

4.2.1.2 Reaktionskontrollen

Zur eindeutigen Identifizierung der Position der *spots* innerhalb der Arrays wurden Kontrollsonden definiert. Als Reaktionskontrolle wurde eine artifizielle Sonde in die Ansatzlösung gebracht. Sie diente innerhalb jedes Arrays als eine Positivkontrolle der Minisequenzierungsreaktion, mit der sichtbar wurde, ob der Einbau aller fluoreszenzmarkierten Dideoxynukleotide erfolgreich war. Mit Hilfe der Reaktionskontrolle war es zudem möglich, die Einbauraten der jeweiligen Farbstoffe zu ermitteln und die Ergebnisse an den so ermittelten Werten innerhalb der Reaktion zu normalisieren. Die stärksten mittleren Einbauraten erzielte das, am Guanin-Nukleotid gekoppelte *TAMRA*-Fluorophor. Der maximale Wert lag bei 49 704 RFUs. Der maximale Wert für den deutlich schwächer emittierenden *Fluorescein*-12-Farbstoff lag bei 10 810 RFUs. Die maximalen RFU-Werte der beiden anderen Farbstoffe, *Cy5* und *Texas Red*, bewegten sich in den Bereichen von 32 192 bis 34 478, wobei *Texas Red* die deutlich größere Spannweite der Intensitätswerte aufzeigte und somit der im Mittel schwächer emittierende Farbstoff war.

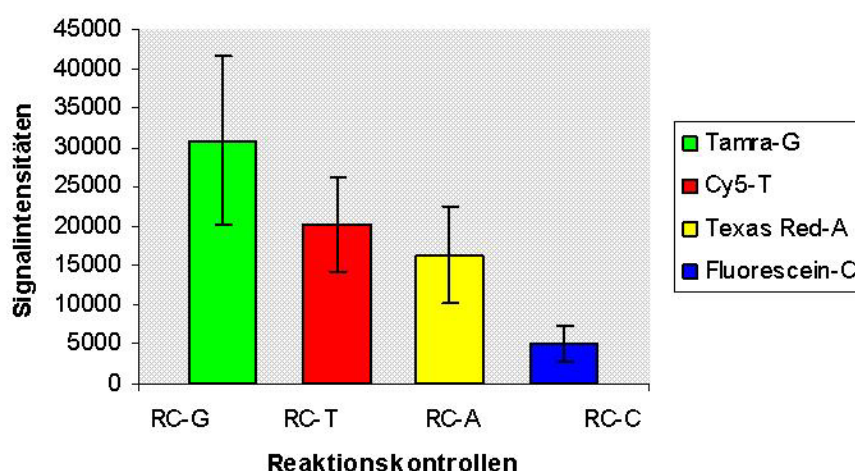


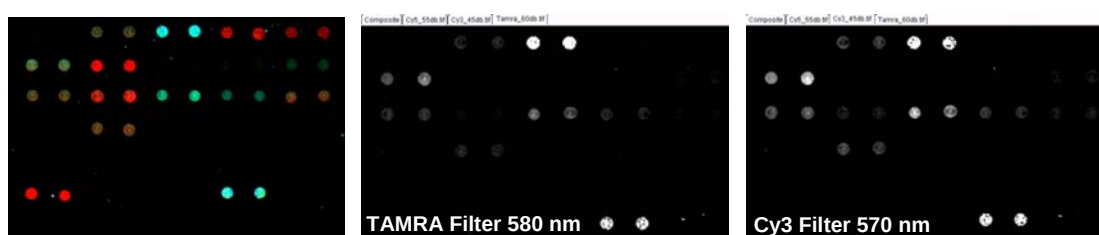
Diagramm 3: Darstellung der mittleren Signalintensitäten und ihrer Standardabweichung für die Reaktionskontrollen der vier fluoreszenzmarkierten Nukleotide.

Für die Analyse mussten die detektierten Werte den variierenden Einbauraten der Farbstoffe angeglichen werden. Durch die Multiplikation der erzielten Werte mit den errechneten Faktoren, um welche die Werte sich zur höchsten gemessenen Intensität unterschieden, konnte eine Normalisierung durchgeführt werden (vgl. Kapitel 3.4.6.5).

4.2.1.3 Umfang der Multiplexreaktion

Die ersten Versuche, eine Multiplexreaktion auf Microarrays zu etablieren, beschränkten sich auf 16 Loci. Relativ schnell konnte nachgewiesen werden, dass es auf dem zur Verfügung stehenden Arrayscanner (*Affymetrix 428*), welcher nur über zwei Anregungslaser verfügte, nicht möglich sein würde, nur mit Filtersätzen die Fluoreszenzfarbstoffe *TAMRA* und *Rox* von *Cy3* bzw. *Cy5* optisch zu trennen (Abbildung 19 a bzw. b). Wie aus der Abbildung deutlich wird, überlappten die Emissionsspektren, was in der Konsequenz eine verfälschte Detektion zur Folge hatte.

a)



b)



Abbildung 19: a) *composite scan* und die Abbildung des grünen Kanals mit einem *TAMRA*-Filter (580 nm) und des grünen Kanals mit dem Filter für *Cy3* (570 nm); b) *composite scan* der Minisequenzierungsreaktion mit *SNaPshot™* Kit Farbstoffen. Nachweis, dass der *Rox*-Farbstoff mit dem Emissionsspektrum des *Cy5*-Farbstoffs überlappt.

Anschließend kamen sechs weitere Loci hinzu und die Multiplexreaktion entsprach der etablierten 23 SNP Multiplexreaktion bis auf einen Locus (A23), welcher erst im weiteren Verlauf hinzukam. Dieser Reaktionsumfang konnte auf beiden verwendeten Glasobjektträgertypen ausgetestet werden und in diesem Etablierungsstadium erfolgte auch die Umstellung des Versuchsansatzes (Kapitel 3.4.3.2). Ein Teil der Farbstoffe wurde ersetzt und das Scannen wurde auf dem *GeneTAC™* LMIV-Laserscanner ausgeführt, welcher in der Lage war, mit vier Anregungslasern einen breiten Spektralbereich abzudecken.

Dadurch konnte die spektrale Trennung aller Farbkanäle gewährleistet werden. Jeder Fluoreszenzfarbstoff (3.4.6.3) wurde von einem Laser angeregt und emittierte in einem separaten Kanal (Abbildung 20).

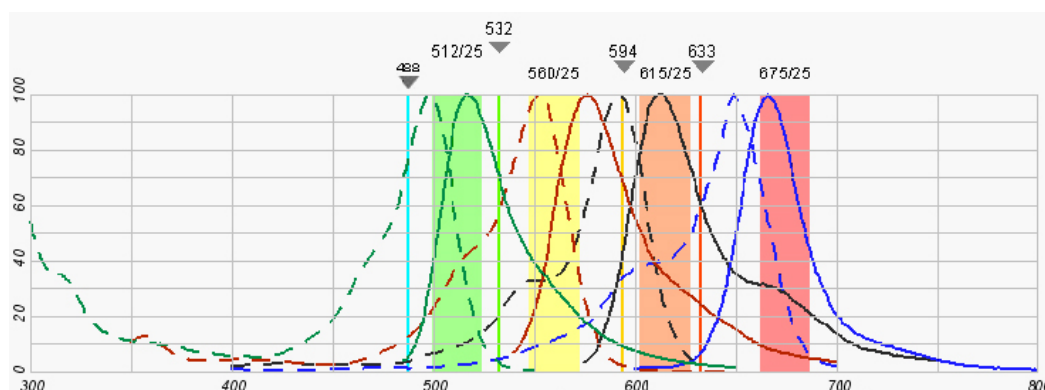


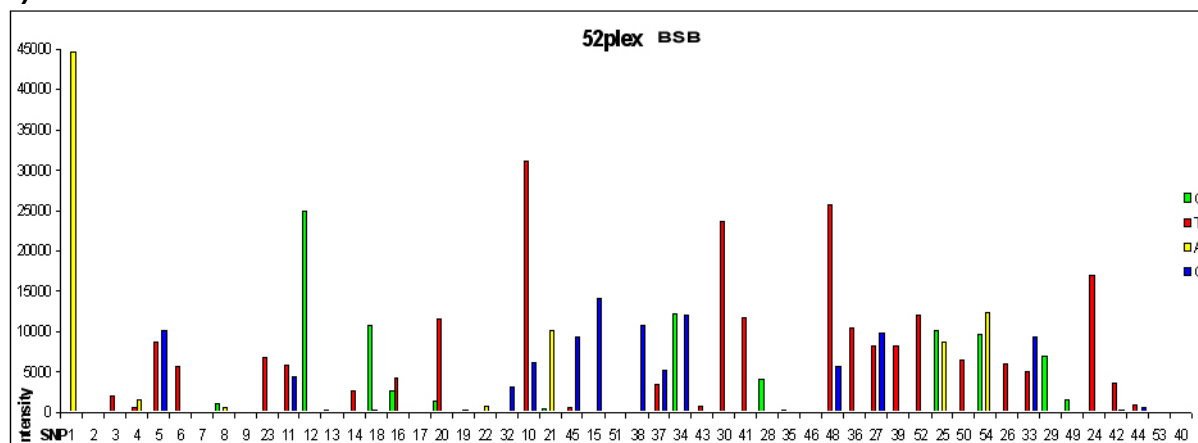
Abbildung 20: Darstellung der Extinktions- und Emissionsspektren der vier Fluoreszenzfarbstoffe im neuen Versuchsaufbau. Auf der x-Achse ist die Wellenlänge in nm, auf der y-Achse der prozentuale Anteil der Fluorophore angegeben. Unterbrochene Linien zeigen die Extinktions-, durchgehende Linien die Emissionsbereiche. Farbige Balken zeigen die eingesetzten Filter an, die den Detektionsbereich in der jeweils hinter dem Schrägstrich angegebenen Zahl (nm) nach rechts und links begrenzen. Die Zahlen oberhalb der grauen Dreiecke sowie die farbigen vertikalen Linien geben die Anregungswellenlänge des entsprechenden Lasers wieder.

Mit dem Anfertigen einer weiteren Charge Glasobjektträger wurden weitere *ctags* synthetisiert und gespottet (Tabelle A-8, im Anhang). Parallel zur Validierungsstudie der 52 SNP Marker wurde die Zahl der SNPs im Microarray-Ansatz erhöht.

4.2.1.4 Etablierung einer 33 SNP Multiplexreaktion

Es konnten nicht alle selektierten Marker für die Detektion mittels Microarrays verwendet werden. Der Ansatz wurde von 52 auf 33 SNPs reduziert. Das große Gesamtvolumen der Reaktion minimierte die Sensitivität und Effizienz der Analyse (Diagramm 4). Die bereits im Vorfeld festgelegte Menge und Abfolge der gespotteten *ctags* verhinderte zudem die Etablierung einer 52 SNP Multiplexreaktion auf Microarray-Glasobjektträgern mangels Ausweichmöglichkeiten auf andere Auffangmoleküle. Ausserdem zeigten zwei getrennt durchgeführte Minisequenzierungsreaktionen (22plex+30plex), die auf dem Glasobjektträger zusammengefügt wurden, keine optimierten Ergebnisse. Das höhere Gesamtvolumen führte zu einer höheren Hintergrundentwicklung und zu einer um 25% erhöhten Ausfallrate gegenüber einer Reaktion im einzelnen Ansatz.

a)



b)

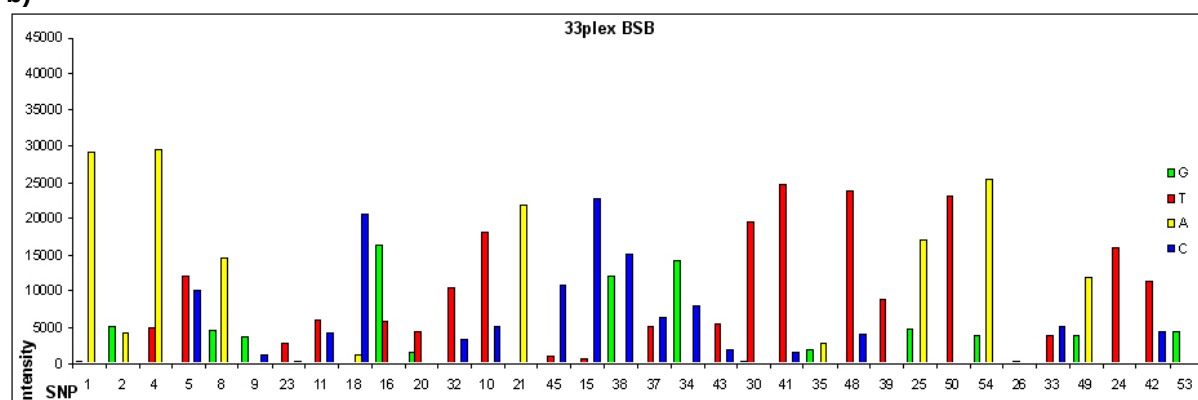


Diagramm 4: Vergleichende graphische Darstellung von a) 52 und b) 33 SNP Multiplexreaktionen.

Wie im Diagramm 4 dargestellt, konnte nach der Reduzierung des Reaktionsumfangs eine deutlich bessere Ausgewogenheit der Signale erzielt werden. Die Einbauraten der einzelnen Farbstoffe und besonders des gelb dargestellten *Texas Red*-Farbstoffs ergaben eine gleichmäßigere Verteilung und einen effizienteren Einbau (Diagramm 4 a und b; Abbildung 21 a und b). Mehrere Versuche zeigten vorab konstante Ausfälle der SNPs A7, A40 und A51 in der 52plex (Abbildung 22 a, rote Rahmen). Diese SNPs wurden aus der Reaktion herausgenommen. SNPs, die in Vorversuchen einen schwachen Einbau zeigten (A4, A35, A43, A49) konnten in der 33plex bestätigt werden (Abbildung 22 a und b, grüne Rahmen). Nicht konstant ausfallende SNPs wurden in die 33plex übernommen (A2, A9, A35, A53, Abbildung 22 a und b, blaue Rahmen). Die 33plex-Reaktion war reproduzierbar und die heterozygoten Loci konnten typisiert werden.

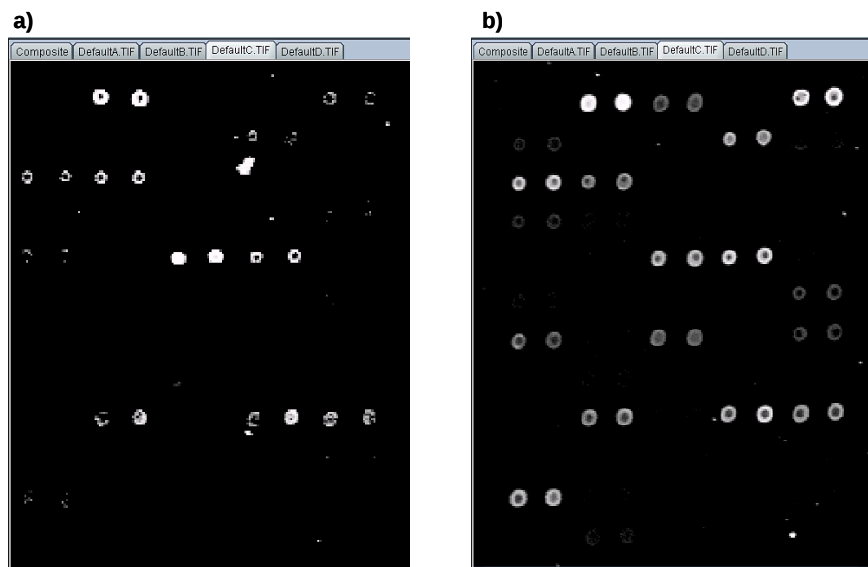


Abbildung 21: Abbildung von einer a) 52 SNP und b) einer 33 SNP Multiplexreaktion. Die Signale zeigen den Einbau des *Texas Red*-Farbstoffs. Die Reaktionsbedingungen unterscheiden sich nur in der Gesamtanzahl der Marker.

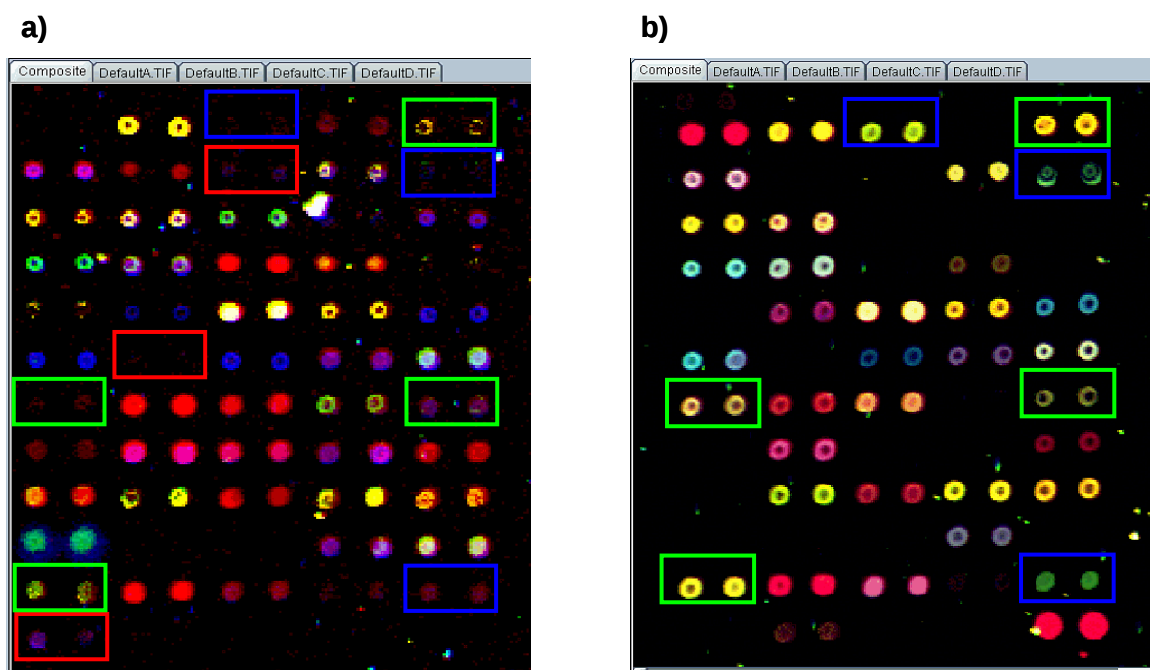


Abbildung 22: Darstellung der Microarray-Scans einer a) 52 SNP Multiplexreaktion und b) eines reduzierten Versuchsansatzes mit 33 Markern. Die roten Rahmen in Abb. a) zeigen SNPs, die konstante Ausfälle aufwiesen und nicht in die 33plex Reaktion übernommen wurden. Die grünen Rahmen zeigen einige Marker, die in der kleineren Reaktion eine Steigerung der Signalintensitäten aufwiesen. Die blauen Rahmen zeigen einige Marker, bei denen es zu Ausfällen in der 52 SNP Multiplexreaktion kam und die in der reduzierten Reaktion reproduziert wurden.

4.2.1.5 Spotreinheit

Ein nur schwer zu eliminierendes Problem war die Kontamination der Spots durch die benachbarte Sequenz der *ctags* sowie ein ungeklärter Einbau von Farbstoffen auf sequenzfremden Spots. Aufgefallen war das Phänomen durch eine stark gesteigerte

Einbaureate der Hybridisierungskontrolle in den für die Hybridisierung der Minisequenzierungsprodukte vorgesehenen Bereichen (Abbildung 23).

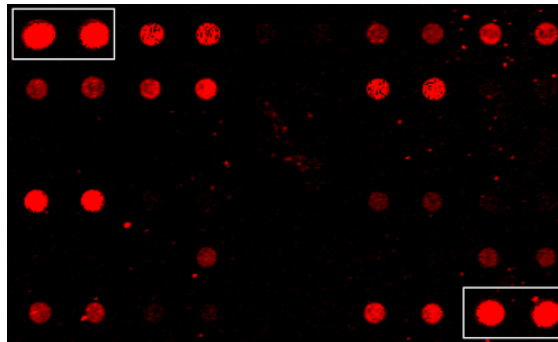


Abbildung 23: Array nach einer Hybridisierungsreaktion, die im Reaktionspuffer ohne Minisequenzierungsprodukte und damit ohne zusätzliche gekoppelte ddNTPs durchgeführt wurde. Die Ausgangskonzentration der Hybridisierungskontrolle betrug 1 μ M. Die durch die komplementäre Sequenz determinierten Positionen im Array sind in der Darstellung weiß umrandet.

Die Funktionsweise der Hybridisierungskontrolle ist die Hybridisierung der komplementären *ctag*-Sequenz mit dem am 3'-Ende gekoppelten Cy5-Fluoreszenzfarbstoff. Diese Reaktion erforderte keine vorangehende Minisequenzierung. Aufgrund der nichthomologen Sequenz der gespotteten *tags* darf keine Kreuzreaktion zwischen den Spots auftreten. Eine Verunreinigung der *tag*-Arrays vor oder während des Spottingsschritts war eine mögliche Erklärung des auftretenden Problems. Im Falle der Hybridisierungskontrolle konnte das Phänomen der Kreuzreaktion oder einer möglichen Kontamination durch eine starke Verdünnung umgangen werden (Abbildung 24).

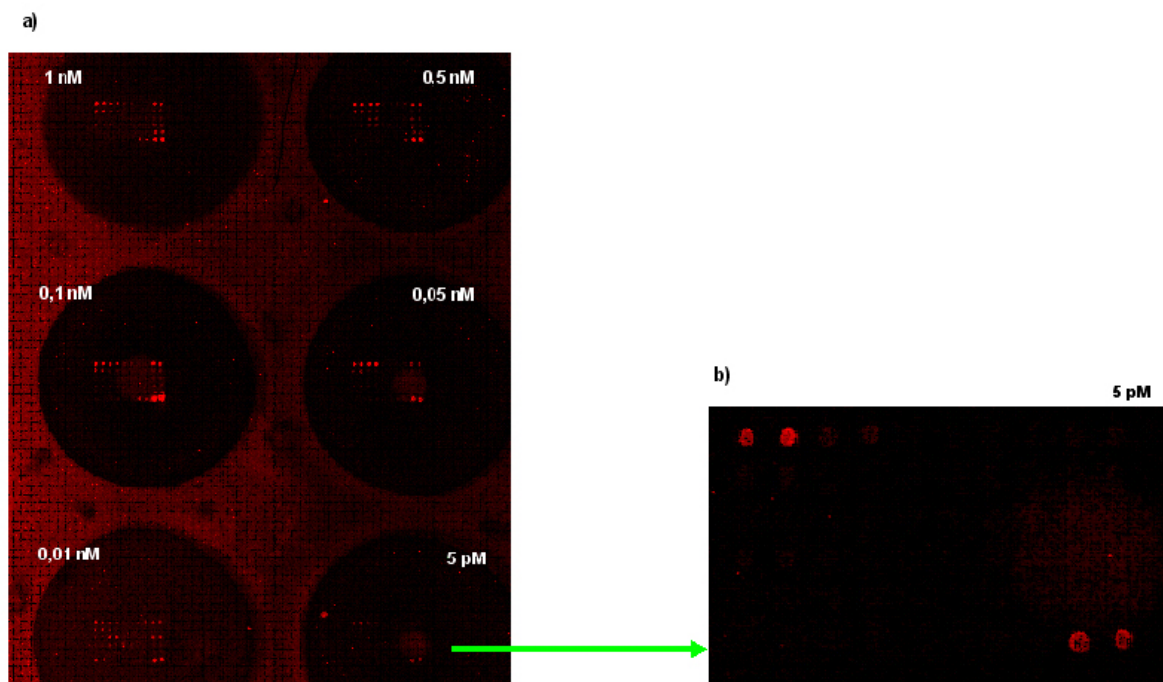


Abbildung 24: a) Zu Optimierungszwecken wurde eine Verdünnungsreihe der Hybridisierungskontrolle durchgeführt. b) Diese zeigte erst bei einer Konzentration von 5 pM keine Kreuzreaktionen mit benachbarten wie auch spotfernen Bereichen.

Diffiziler gestaltete sich die Festlegung der Grenze für das Hintergrundsignal innerhalb der Analysespots. Der Verdacht entstand, als in einer neuen Charge Glasobjektträger Signale auf Spots detektiert wurden, die keinen komplementären Hybridisierungspartner besaßen. Der erste Nachweisversuch erfolgte über die Betrachtung der *spots* der Reaktionskontrolle. Es verstärkte sich der Verdacht, dass vor oder während des Spottens eine Kontamination mit dem Oligovorrat aus der benachbarten Vertiefungen auf der Platte stattgefunden haben musste, die zu einer unspezifischen Übertragung der *ctags* geführt hatte (Abbildung 25).

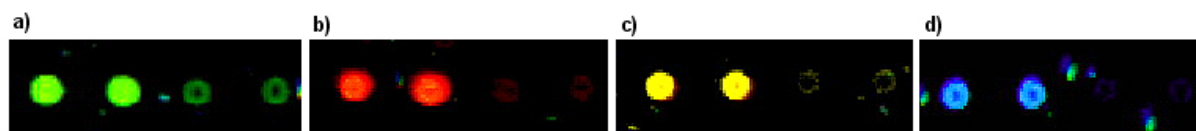


Abbildung 25: Nachweis von Kontaminationen der benachbarten Spots durch eine mögliche Verschleppung der komplementären *tag*-Sequenzen beim Spotten. a) bis d) Doppelspots der Reaktionskontrollen RC-G, RC-T, RC-A und RC-C (jeweils links) und die Doppelspots für die *ctags* 00093_21 (jeweils rechts) mit starker Kontamination. In allen Fällen fand beim Spotten eine Verschleppung der *ctags* der Reaktionskontrollen auf die Nachbarspots statt und anschließend die Hybridisierung der *tag*-Oligos der jeweiligen Reaktionskontrollen.

Ein erneutes Spotten mit frisch im Puffer angesetzten *ctags* konnte den sichtbaren Übertrag verhindern.

Ein weiterer Verdacht auf unspezifisches Bindeverhalten von Hybridisierungssonden konnte jedoch nicht entkräftet werden und spielte bei der Auswertung der Proben eine erhebliche Rolle. Mehrere Versuche zeigten, dass entweder eine reproduzierbare Bindungsaffinität der *tags* auf falschen *spots* vorlag oder eine fehlerhafte Inkorporation von Farbstoffen in der Minisequenzierungsreaktion stattgefunden hat. Wie durch das Beispiel in der Abbildung 26 belegt wird, kam es zu Signalüberlagerungen von bis zu vier Farben für einen Marker (SNP A24 und SNP A48 Abbildung 26 a – e). Häufig traten auch Überlagerungen eines zusätzlichen Farbstoffs auf (SNP A23 und SNP A54 Abbildung 26 a – d). Der Marker A4 (Abbildung 26 a, c und d) stellte hierbei einen der wenigen Marker dar, bei denen es zu keiner zusätzlichen Farbstoffinkorporation bzw. einer nichtspezifischen Bindung von fremden *tags* kam. Die Intensitätswerte der Farbstoffe in falschen Kanälen konnten in Einzelfällen stärker sein als die erwarteten Signale. Je größer der Umfang der parallel hybridisierten *tag*-Oligos war, umso stärker waren die Farbüberlagerungen, die detektiert wurden.

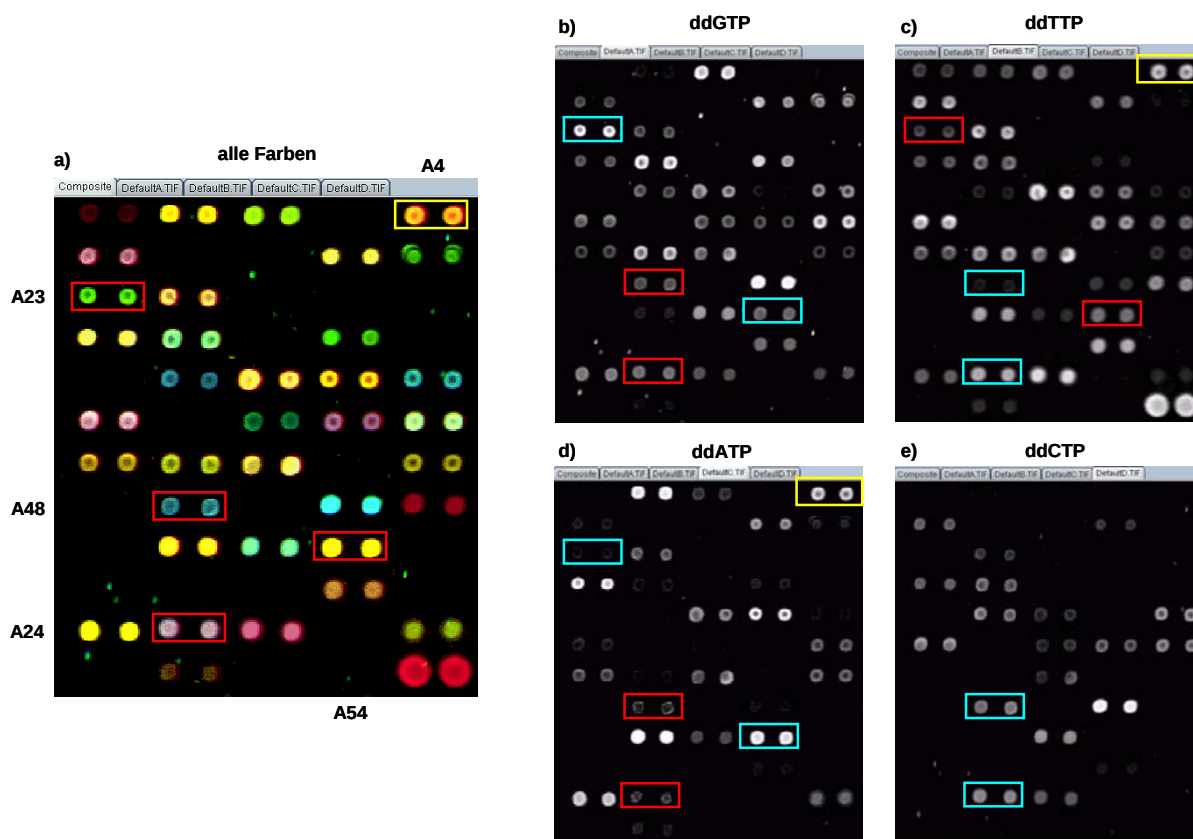


Abbildung 26: a) Kombinationsdarstellung und b) – e) Darstellung der Einzelkanäle für die Probe L35a zur Verdeutlichung der Signalübertragung. Beispiele mit unterschiedlicher Ausprägung. c) und d) SNP A4 – ein T/A-Polymorphismus (heterozygot, gelber Rahmen), korrekt dargestellt in den entsprechenden Farbkälen. b) bis d) SNP A23 (hier homozygot) und SNP A54 (hier heterozygot): G/A-Polymorphismen (blaue Rahmen) und die Signalübertragung in den Farbkälen des roten Lasers (rote Rahmen). Abb. b) bis e) SNP A48 (hier homozygot) und SNP A24 (hier heterozygot): T/A-Polymorphismen (blaue Rahmen) und die Signalübertragung in die beiden übrigen Farbkälen des grünen und des gelben Lasers (rote Rahmen).

Die beschriebenen Signalüberlagerungen konnten auch in sehr kleinen Reaktionsansätzen nachgewiesen werden. Es wurden mehrere Tri- und Quadruplexe durchgeführt, in welchen falsche Signale in einem oder in beiden Farbkälen detektiert worden sind (Abbildung 27). Die geringste Anzahl falscher Signale war im Farbkälen des gelben Lasers zu beobachten (vgl. Abbildung 27 d). Dieses kann auf eine geringere Tendenz einer Fehlinkorporation vom ddATP hinweisen. Im Fall einer Fehlhybridisierung von *tag*-Oligos könnten größere sterische Störungen einen Behinderungsgrund darstellen.

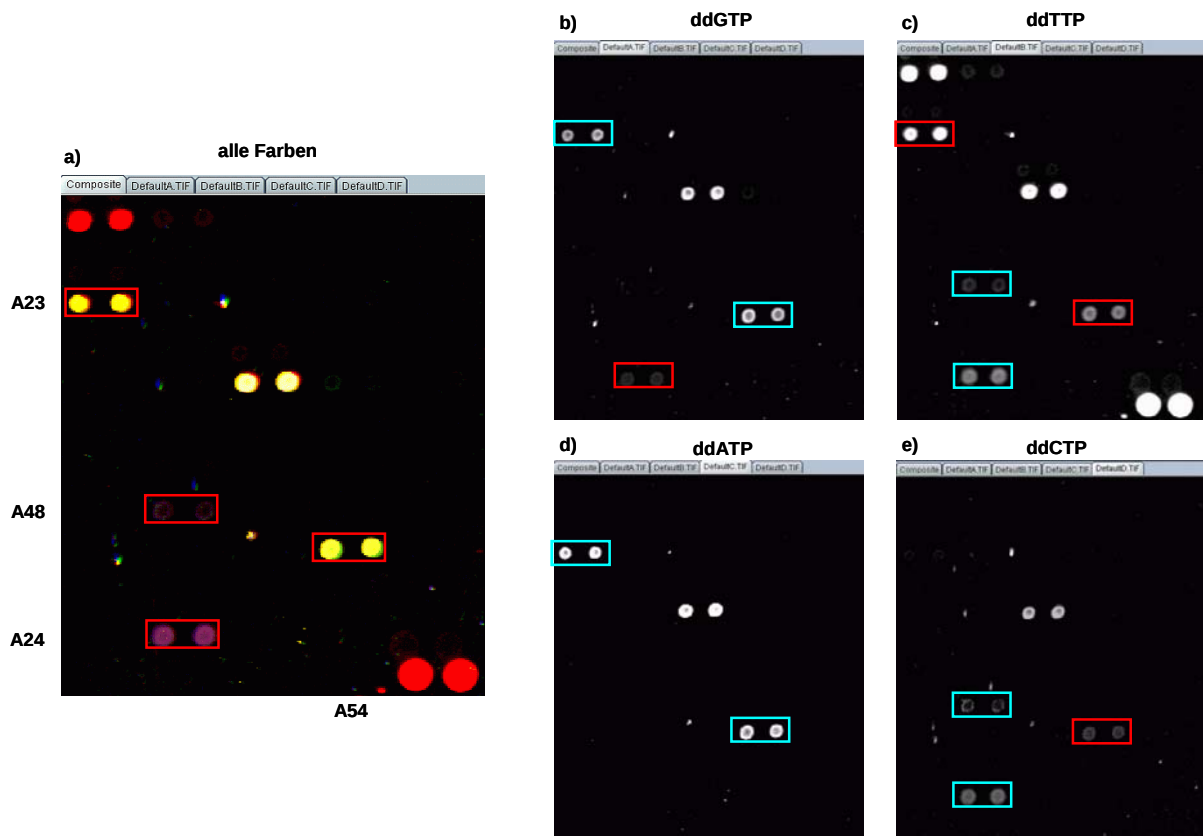


Abbildung 27: Darstellung einer Quadruplex-Reaktion mit den Markern A23, A24, A48 und A54. a) Kompositionsdarstellung aller Farbkanäle. Rahmen befinden sich um die Marker-Doppelspots. Bei nicht umrahmten Spots handelt es sich um die Hybridisierungskontrollen und die Reaktionskontrolle. Abb. b) bis e) SNP A23 und SNP A54 (beide heterozygot): G/A-Polymorphismen (blaue Rahmen) und die Signalübertragung in den Farbkanal des roten Lasers (rote Rahmen) sowie e) Signaldetektion im blauen Farbkanal beim SNP A54 (roter Rahmen). Abb. b), c) und e) SNP A24 (hier heterozygot): c) T/A-Polymorphismus (blauer Rahmen) und die Signalübertragung in den Farbkanal des roten Lasers (rote Rahmen). Abb. c) und e) Bei SNP 48 (hier heterozygot) werden ausschließlich die korrekten Signale wiedergegeben.

Lediglich sechs Marker (A4, A9, A20, A38, A30 und A39) wiesen keine falsch positiven Signale auf.

Kam es zu einer Signalüberlagerung außerhalb des Farbkanals der zu detektierenden SNPs, hatte dies keine unmittelbare Auswirkung auf das Ergebnis der Genotypisierung, denn die Auswertung erfolgte ausschließlich über die beiden zugeordneten Kanäle. Es drängten sich jedoch zwei Vermutungen auf:

- 1) Das Phänomen der Überlagerung mit falsch eingebauten Farbstoffen kann auch innerhalb der korrekten Farbkanäle auftreten. Dies kann in einem solchen Fall zu falsch positiven Signalen führen und das Gesamtergebnis verfälschen.
- 2) Der Fehleinbau führt zu einer verstärkten Konkurrenz der Signale, so dass die Signalintensitäten nicht ihre vollständige Stärke erreichen können und es zu einseitig verschobenen Verhältnissen kommt. Die Ausbalancierung der Signale wird durch das Konkurrenzverhalten der markierten Oligos auf den Spots beeinträchtigt.

Die Ergebnisse müssen zwangsläufig unter Berücksichtigung beider Aspekte analysiert und diskutiert werden.

4.2.1.6 Optimierung der Microarray-Reaktion durch Anwendung unterschiedlicher Sequenzierenzyme

Anlehnend an die Literaturvorgaben (Lindroos *et al.* 2002, Liljedahl *et al.* 2004 und Lovmar und Syvänen, 2005) wurden die besten Typisierungsergebnisse mit der Sequenzierpolymerase *ThermoSequenase* (Amersham) erzielt. Mit der Umstellung der Fluoreszenzfarbstoffe und der Erhöhung der Auswahl der Marker innerhalb einer Reaktion kam es jedoch zu einem deutlichen Rückgang der Intensitätswerte (Abbildung 28 a.1). Die Signale im gelben Kanal verschwanden fast komplett (Abbildung 28 a.2). Dieses erhöhte die Gefahr einer Fehlinterpretation der daraus hervorgehenden falschen Homozygoten, im Speziellen für den am häufigsten vorkommenden G/A Polymorphismus.

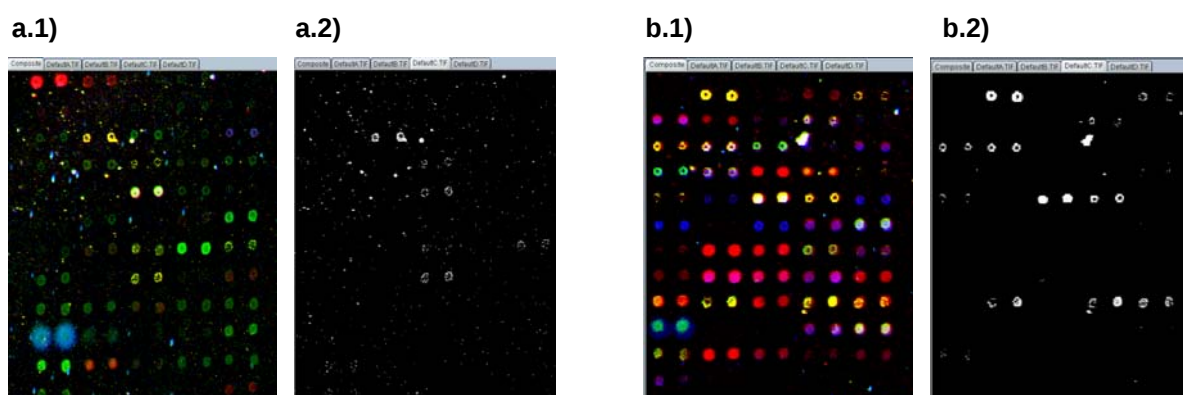


Abbildung 28: a.1) Darstellung einer Multiplexreaktion durchgeführt mit der *ThermoSequenase* Polymerase. Die Signalintensitäten waren stark herabgesetzt und die Reaktion konnte nicht analysiert werden, a.2) besonders deutlicher Signalverlust im gelben Farbkanal. b.1) Gleiche Reaktionsbedingungen unter Zuhilfenahme der Polymerasen *ThermoSequenase* und *Therminator DNA*. Deutliche Verstärkung der Signalintensitäten in allen Farbkanälen – besonders im gelben Farbkanal, Abb. b.2).

Die Hinzunahme eines weiteren Sequenzierenzym (*Therminator DNA Polymerase*, New England Biolabs) führte zu einer deutlichen Verbesserung der Einbauraten des *Texas Red*-Fluoreszenzfarbstoffs. Einzeln in der Reaktion eingesetzt, zeigte die Polymerase jedoch eine Einbauschwäche des *Cy5*-Farbstoffs und führte damit zu einer Fehlinterpretation der T/A, T/G und T/C Polymorphismen.

Wurden beide Enzyme in der Minisequenzierungsreaktion eingesetzt, resultierte dies in einer Angleichung der Einbauraten aller Fluoreszenzmoleküle (Abbildung 28 b.1), insbesondere die des *Texas Red*-Fluorophors (Abbildung 28 b.2). Die besten Ergebnisse wurden mit dem Einsatz von 3 Units *Therminator DNA Polymerase* und 3 Units *ThermoSequenase Polymerase* erzielt.

Bei der Überlegung einer Verkleinerung des Umfangs des Multiplexansatzes spielte das Einbauverhalten der Sequenzierpolymerasen ebenfalls eine entscheidende Rolle. In Versuchen mit 52 SNPs wurde eine schwächere Einbaurate der Moleküle nachgewiesen

(Abbildung 29 a.2). Demgegenüber konnten alle potentiellen Signale für *Texas Red*-ddATP im kleineren Multiplexansatz mit 33 Markern sichtbar gemacht werden (Abbildung 29 b.2).

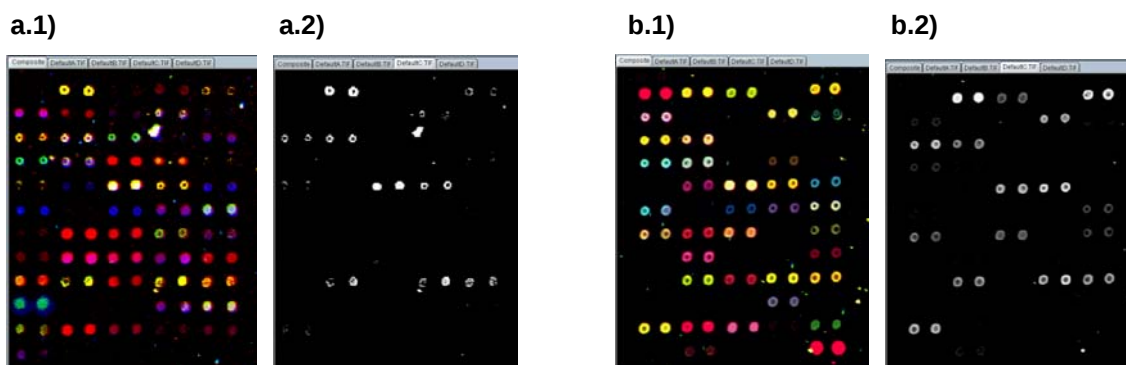


Abbildung 29: Darstellung des Hybridisierungsverhaltens in einer a.1.) 52 SNP Multiplexreaktion, mit der Detailabbildung des Hybridisierungsverhaltens des gelben *Texas Red*-Farbstoffs, Abb. a.2. b.1.) Deutliche Zunahme der Signalintensitäten nach der Reduzierung der SNP-Anzahl in der Multiplexreaktion. b.2.) Detailabbildung des gelben Farbkanales.

4.2.2 Probendetektion und Evaluierung

4.2.2.1 Visualisierung der Daten und Einzelbetrachtung der ausgewählten Sonden

Die folgenden Ergebnisse zeigen, dass nicht alle SNPs gleichwertig hybridisierten. Um die Qualität der Sonden zu beurteilen, wurde das Hybridisierungsverhalten jedes einzelnen SNP in einer Clusteranalyse gemeinsam ausgewertet.

Vor der deskriptiven Auswertung des Genotypisierungserfolges war diese Vorgehensweise hilfreich. Sie erlaubte die Lesbarkeit der Daten zu beurteilen und half abzuschätzen, inwieweit eine zusätzliche Normalisierung zur besseren Darstellung notwendig war.

Die Clusterdarstellung stellte eine Qualitätskontrolle dar. Mit Hilfe der Cluster konnte verdeutlicht werden, ob die Oligosonden in den einzelnen Experimenten gleichwertig hybridisierten und deren Ergebnisse somit vergleichbar waren.

Dazu wurden die Sonden in drei unterschiedliche Kategorien eingeteilt und unter dem qualitativen Aspekt ausgewertet.

Fast die Hälfte der 33 ausgewählten SNPs konnte anhand der *SNPSnapper Software* in eindeutigen Clustern dargestellt und anschließend genotypisiert werden. Sieben SNPs wiesen eine schlechte Clusterseparation auf und konnten nur unvollständig typisiert werden. Zehn der 33 SNPs wurden als unzureichend klassifiziert. Hier war keine automatisierte Zuordnung der Genotypen möglich und die Inkorporation der Farbstoffe fiel vollständig oder zum großen Teil aus.

Zwischen den Typisierungsdurchgängen **Serie 1** und **Serie 2** gab es geringe Unterschiede, wobei diese sich in der Streuung der Werte bemerkbar machten. In wenigen Fällen kam es

zu Ausfällen einzelner Proben in beiden DNA-Pools. Dieser Bereich wurde in der Plotdarstellung durch eine schraffierte Fläche markiert und aus der weiteren Analyse ausgeschlossen. In der Plotdarstellung befinden sich die Graphen der Serie 1 und Serie 2 jeweils untereinander, wobei links, wenn nicht zusätzlich anders beschrieben, die normalisierten Cluster dargestellt werden.

Die Darstellung der Ergebnisse erfolgt im Folgenden anhand ausgewählter Beispiele. Die vollständige Darstellung aller Clustergraphen findet sich im Anhang (Abbildung A-2 bis Abbildung A-34).

Es muss vorab darauf hingewiesen werden, dass verschiedene Faktoren und physikalische Größen auf das Gesamtergebnis Einfluss nehmen. Diese werden im Diskussionsteil einzeln besprochen und kritisch beleuchtet. Im Ergebnisteil werden die SNPs rein deskriptiv betrachtet und tabellarisch oder in Plots dargestellt.

4.2.2.1.1 Gut diskriminierende SNPs

Die hier beschriebenen SNPs zeigten alle in mindestens einem Analysedurchgang eine eindeutige Untergliederung in drei Cluster, die jeweils einem der drei Genotypen entsprachen. Dies war gewährleistet, wenn der heterozygote Genotyp in einem Bereich um 0,5 und die beiden homozygoten Genotypen um die Werte 0 bzw. 1 auf der x-Achse angeordnet waren. Dieses traf auf zwei der SNPs zu (A2 und A50, Abbildung 30 und Abbildung 31). Bei den restlichen SNPs wurden die Genotypisierungsbegrenzungen (in der Plotdarstellung als blaue oder grüne Flächen) stärker angepasst.

Die Cluster des SNP A2 zeigten eine maximale Trennung auf der Skala von 0 bis 1 auf. Bei diesem SNP handelt es sich um einen A/G Polymorphismus, wobei die meisten Individuen im Versuch den homozygoten Genotyp GG aufwiesen. Bemerkenswert war die Gruppierung der homozygoten Proben dicht um den entsprechenden Wert auf der Skala, während die heterozygoten A/G-Individuen eine Streuung aufzeigten. Die durchgeführte Normalisierung ergab keine Veränderung in der Gruppierung der heterozygoten Werte (Abbildung 30 a). Alle Proben im heterozygoten Bereich, bis auf die Probe L45a, konnten eindeutig genotypisiert werden (Pfeil). Im Typisierungsdurchgang der b-Proben zeigte sich eine ähnliche Separation der Genotypen und ein Ausfall der Probe L44b. L45b konnte hingegen genauer bestimmt werden und zeigte einen heterozygoten Genotyp mit einem Wert um 0,5 auf (Abbildung A-3).

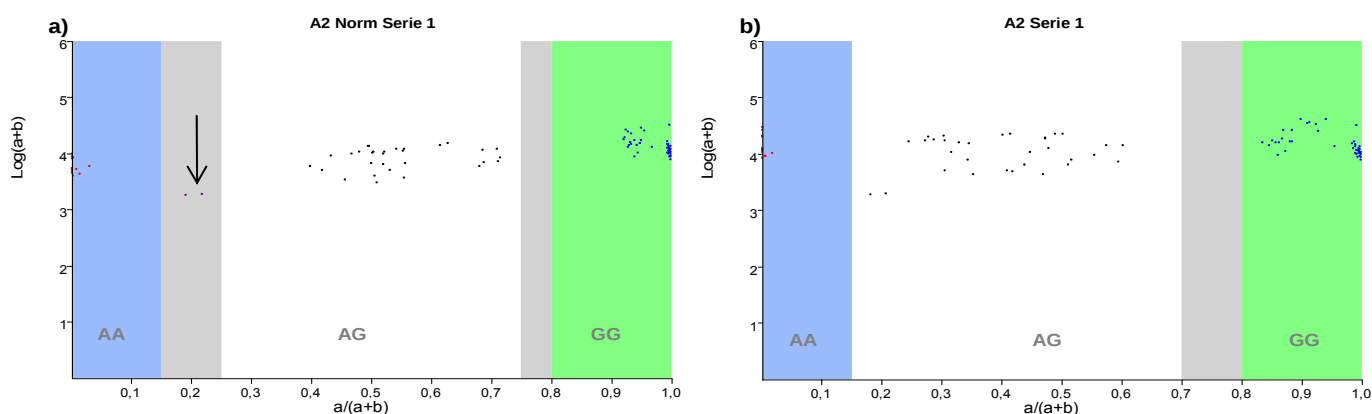


Abbildung 30: Abbildung von zwei Plots mit klar getrennten Clustern für den Marker A2. Innerhalb der grauen Flächen liegen die Punkte, deren Genotypen nicht eindeutig bestimmt werden konnten (L45b, Pfeil). Die Gruppierung erfolgte jedoch mit vergleichbarer Qualität sowohl mit als auch ohne eine Korrektur der Werte.

Eine ebensolche Clusterung wurde auch für den SNP A50 nachgewiesen (Abbildung 31 a und b). Die Bereiche der Genotypen waren klar getrennt und wiesen nur eine geringe Streuung auf. Nach Korrektur mit dem Intensitätsfaktor wurden die Werte gebündelt und grenzten sich in klaren Clustern ab (Abbildung 31 a). Die Analyse der Proben b zeigte ebenfalls eine Abgrenzung der Homozygoten für die Allele 1 (T) und 2 (C). Eine leichte Streuung war in der Analyse jedoch für die heterozygoten Werte nachweisbar. Die Probe

L4b fiel aus (Pfeile) und wurde aus der Gesamtwertung durch die schraffierte Fläche herausgenommen (vgl. Abbildung 31 b).

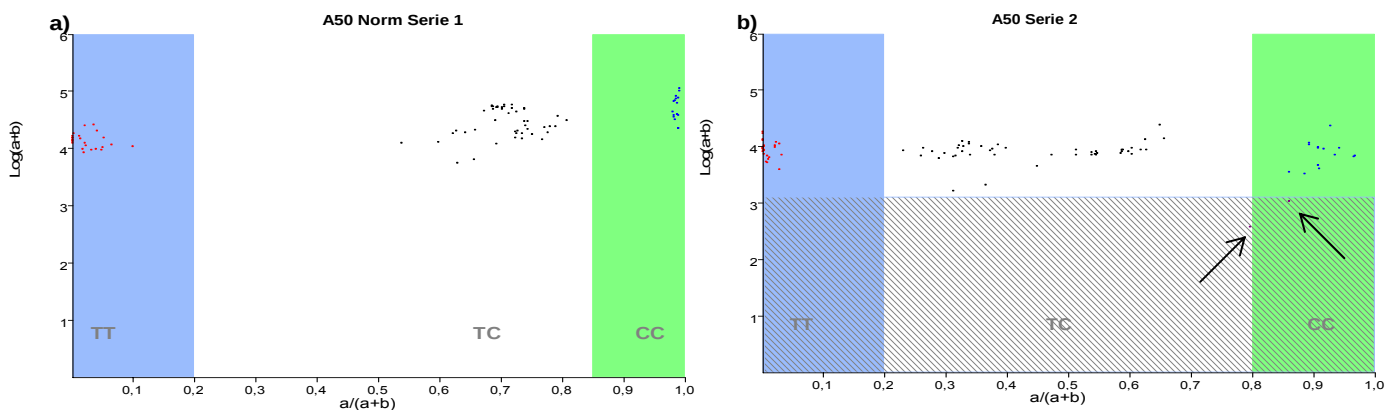


Abbildung 31: Abbildung von zwei Plots des Markers A50. a) Die Korrektur der Werte resultierte in einer sehr guten Gruppierung der Daten in den jeweiligen Clustern und in einer vollständigen Zuordnung der Genotypen. b) In der Typisierung der Probenserie 2 fiel Probe L4b aus (Pfeile unter der schraffierten Fläche). Diese wurde aus der weiteren Analyse entfernt.

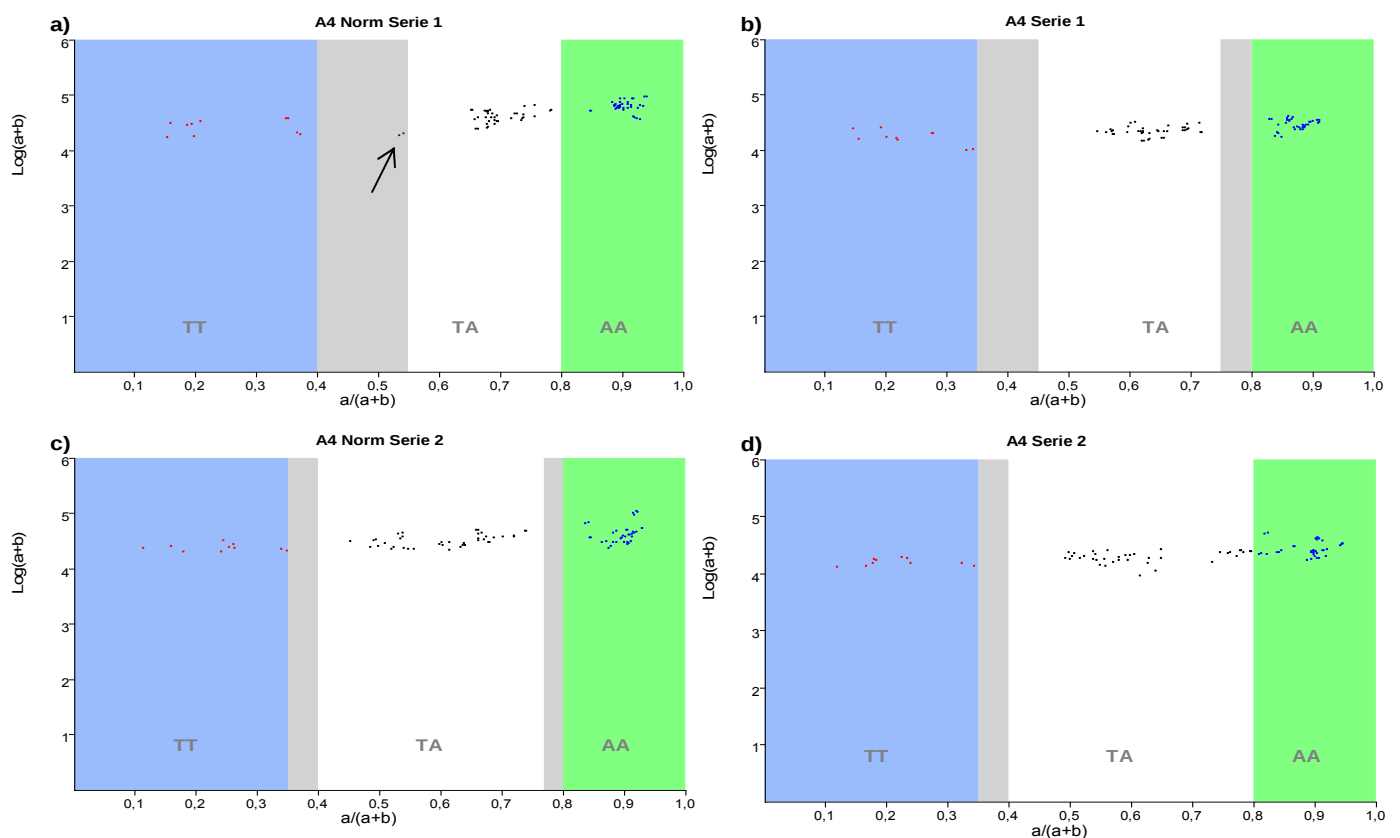


Abbildung 32: Abbildung der Gesamtanalyse des Markers A4. b) und d) Die analysierten Proben wurden eindeutig ihren Genotypen zugeordnet. a) und c) Nach der Korrektur der Werte erfolgte eine Streuung innerhalb der Cluster der Genotypen TT und TA. Probe L45a konnte nicht eindeutig einem Genotyp zugewiesen werden (Abb. a), Pfeil).

Der Marker A4 wurde in den Farbkanälen T und A typisiert. Er wies eine deutliche Bündelung der Proben innerhalb den jeweiligen Genotypen auf. Diese waren auf der Skala im

Gegensatz zum vorherigen SNP etwas verschoben, konnten jedoch in eindeutige Cluster abgetrennt werden und es erfolgte eine korrekte Genotypisierung. Unterschiede in der Streuung waren hierbei zwischen den DNA-Proben der Serie 1 und 2 zu erkennen (Abbildung 32 a – d). Die heterozygoten Proben der Serie 2 waren weniger gruppiert, lagen jedoch innerhalb gleicher Skalenwerte und konnten ebenfalls genau typisiert werden. In beiden Probensätzen kam es zu keinen Ausfällen einzelner Proben. Nach der Faktor-Multiplikation in der Probengruppe der Serie 1 erfolgte eine Abspaltung der Probe L45a aus der Heterozygotengruppe (Abbildung 32 b, Pfeil). Diese Probe wurde außerhalb des Clusters im grauen Feld platziert und daher nicht eindeutig zugeordnet. In der Probenserie 2 erfolgte eine vollständige Typisierung (Abbildung 32 c und d).

Wie in den vorab beschriebenen Plots wurde für den SNP A5 eine relativ genaue Zuordnung der drei möglichen Genotypen erreicht. Eine starke Streuung der Werte konnte lediglich in den homozygoten Proben (TT-Genotyp) in beiden Serien (Abbildung A-5 im Anhang) beobachtet werden. Die Korrektur der Werte führte zu einer stärkeren Gruppierung (Abbildung 33 a und b). Auffällig war dabei, dass sich die Werte innerhalb des TT-Clusters untereinander nochmals in ein Untercluster teilten. Besonders ausgeprägt war diese Unterteilung nach der Korrektur mit dem Intensitätsfaktor (Abbildung 33 a und b).

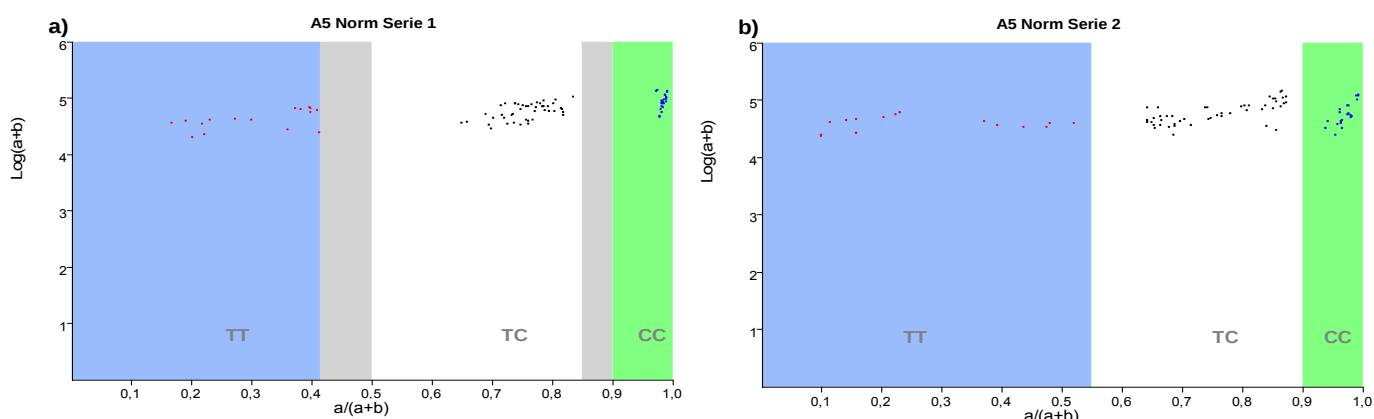


Abbildung 33: Abbildung der Plots der Probenserie 1 und 2 für den Marker A5 nach der Korrektur der Werte. Sie zeigen trotz einer guten Verteilung der Genotypen eine Streuung der Wertepunkte und ein gesondertes Cluster (Abb. b, blaue Fläche).

Die folgenden SNPs, die ebenfalls gute diskriminatorische Werte aufwiesen, werden hier nicht einzeln beschrieben. Sie wiesen eine ähnliche Verteilung innerhalb der Plots auf. Bis auf wenige Proben konnten beide Serien vollständig typisiert werden (siehe Abbildungen im Anhang).

Die größte Abweichung zwischen der Serie 1 und 2 trat beim SNP A15 auf. Die eindeutige Clusterung innerhalb der Serie 1, bis auf Probe 2a (Abbildung 34, Pfeil) konnte nicht in der Probenserie 2 reproduziert werden. Es kam dort zu vereinzelt Ausfällen der Marker (Abbildung A-9) sowie zu einer breiten Streuung der Proben entlang der x-Achse. Auffallend hier war zudem eine größere Menge an DNA-Proben, die nicht eindeutig typisiert werden konnten (Abbildung 34 b, grauer Bereich).

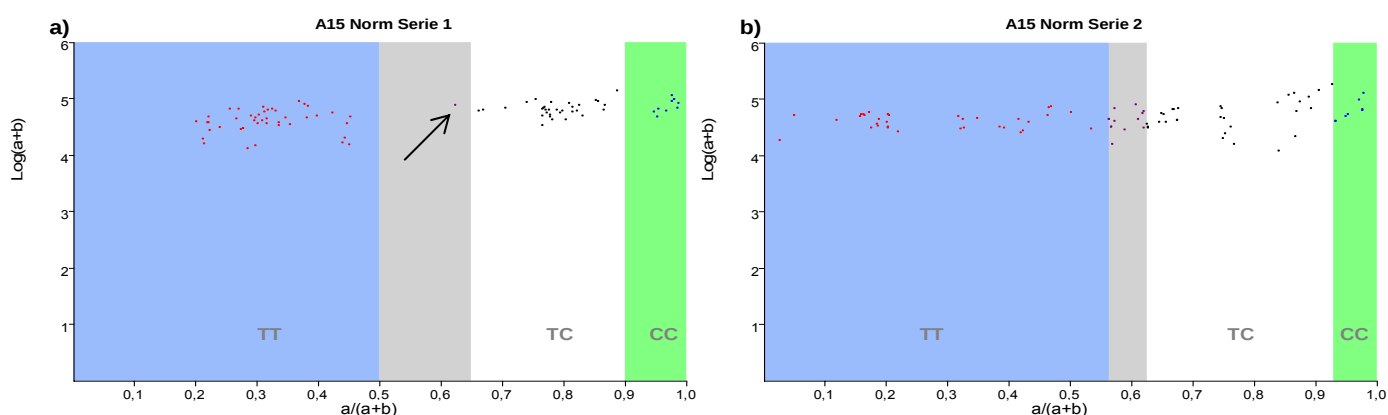


Abbildung 34: Abbildung der Plots der Probenserie 1 und 2 für den Marker A15 nach der Korrektur der Werte. a) Die Verteilung der Probenserie 1 erfolgte eindeutig, nur eine Probe konnte nicht zugeordnet werden (Pfeil). b) In der Probenserie 2 befanden sich durch eine Streuung der Wertepunkte eine Reihe von Proben innerhalb der grauen Fläche zwischen den Genotypen TT und TC.

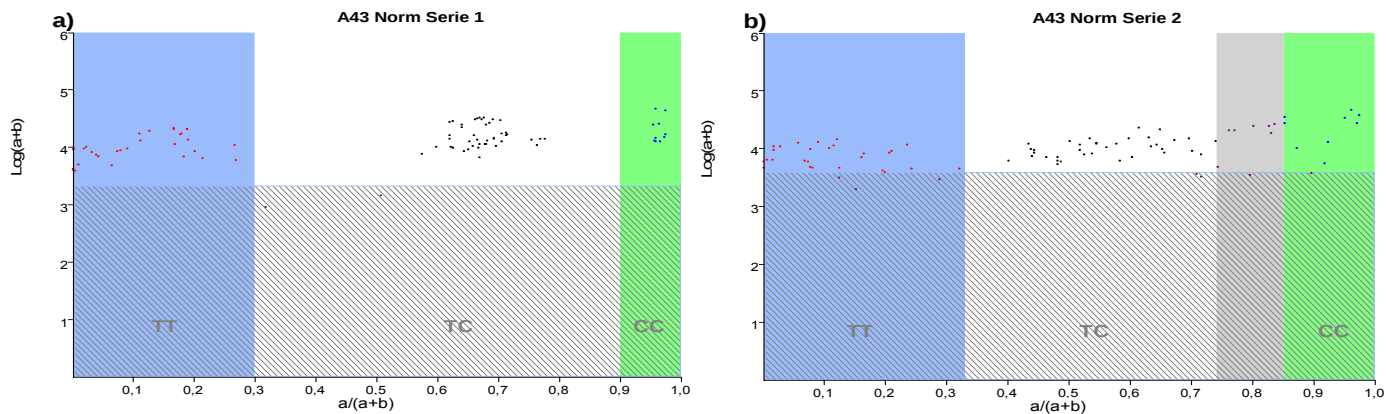


Abbildung 35: Abbildung der Plots der Probenserie 1 und 2 für den Marker A43 nach der Korrektur der Werte. a) Die Verteilung der Werte der Probenserie 1 erfolgte eindeutig, die Probe L14a fiel jedoch aus (schraffierte Fläche). b) Die Proben der Serie 2 konnten nicht eindeutig gruppiert werden (graue Fläche). Es kam zu einer breiten Streuung der Wertepunkte zwischen den Genotypen TC und CC. Sieben Proben wiesen zu geringe Fluoreszenzsignale auf und konnten nicht analysiert werden (schraffierte Fläche).

Die Problematik der unterschiedlichen Typisierungsqualität zwischen den Probenserien 1 und 2 wiederholte sich für den SNP A43 (Abbildung 35 a und b), wobei es hier zu sieben Ausfällen innerhalb der Probenserie 2 kam (10b, L13b, L4b, L23b, 1b, E9201b und E9225b; Abbildung 35 b - Punkte unterhalb der schraffierten Fläche). Im Gegensatz dazu wurde nur ein Ausfall innerhalb der Probenserie 1 beobachtet (L14a; Abbildung 35 a – Punkte unterhalb

der schraffierten Fläche). Durch die starke Streuung der Werte konnten die Genotypen einzelner Proben nach der Normalisierung nicht zugeordnet werden (Abbildung 35 b).

Betrachtete man alle Plots der Probenserie 2, so konnte festgestellt werden, dass die Werte im Gegensatz zur Analyse mit der Probenserie 1 im Allgemeinen eine breitere Streuung auf der x-Achse aufwiesen. Dadurch kam es verstärkt zu Überlagerungen der Genotypenbereiche. In dieser Gruppe kam es z. T. zu einer Unterteilung der Cluster, so dass für einen Genotyp ein fehlerhafter Eindruck eines vierten Clusters entstehen konnte (SNPs: A4, A5, A15, A30, A50 und A24). Korrekturen mit dem Intensitätsfaktor wirkten verstärkend auf die Streuung der Werte entlang der x-Achse. Abbildung 36 a) zeigt ein Beispiel einer Überlappung der Cluster im grauen Bereich. Abbildung 36 b) stellt ein Beispiel einer Gruppierung der Werte in vier Cluster dar.

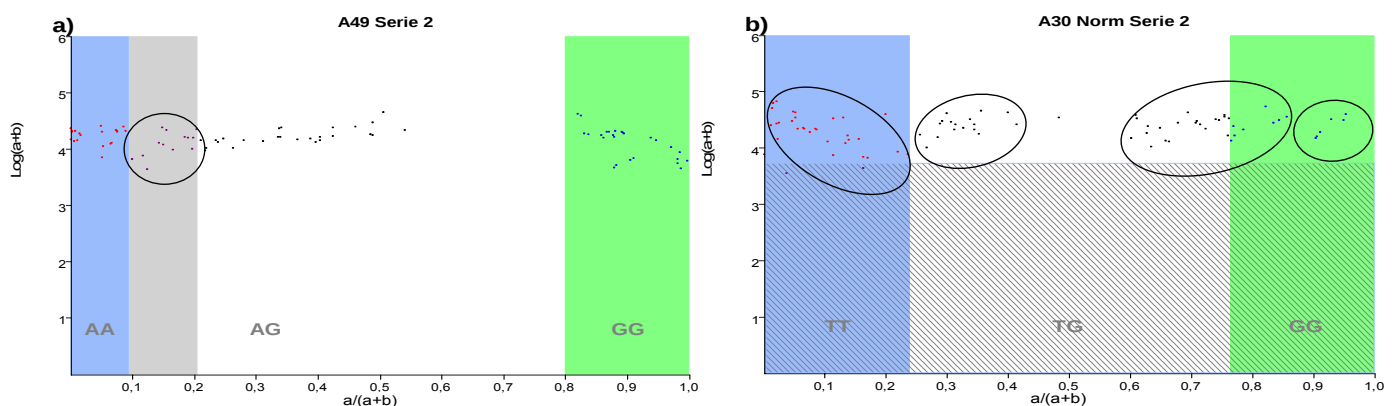


Abbildung 36: Abbildung der Plots der Probenserie 2 für den Marker A49 vor der Korrektur und für den Marker A30 nach der Korrektur der Werte. a) Beispiele für die schlechte Genotypisierungsqualität der Probenserie durch die Streuung und Überlappung der Wertepunkte unterschiedlicher Cluster. b) Ausfall der Proben (schraffierte Fläche) sowie die Bildung von Subclustern (eingekreisten Bereiche).

4.2.2.1.2 SNPs mit mittlerer Clusterqualität

Acht der ausgewählten SNPs wurden in eine Kategorie eingegliedert, aus welcher die Trennung der Cluster nicht eindeutig erkennbar war. Eine solche Datenansammlung bildete eine nicht gegliederte Ansammlung von Punkten im Plot und wurde in der Analyse genauer betrachtet und manuell ausgewertet. Dabei zeigte sich, dass die Werte durchaus eine Möglichkeit zur Interpretation boten, sie jedoch, wie am Beispiel der SNPs A1 und A8 gezeigt wird, nur unzureichend getrennte Cluster bildeten (Abbildung 37 a und b). Innerhalb beider Beispiele befanden sich die Datenpunkte aller Genotypen dicht in einer „Datenwolke“. Dennoch kam es zu keiner Überlappung der Grenzbereiche und durch Verschiebung der farbigen Flächen im Plot konnte eine tabellarische Typisierung erzielt werden. Nach einer durchgeführten Korrektur der Daten mit dem Intensitätsfaktor fielen vereinzelt Datenpunkte in den grau unterlegten Bereich und wurden nicht oder nur nach manueller Prüfung in die Genotypisierung einbezogen (Abbildung A-2 und Abbildung A-6 im Anhang).

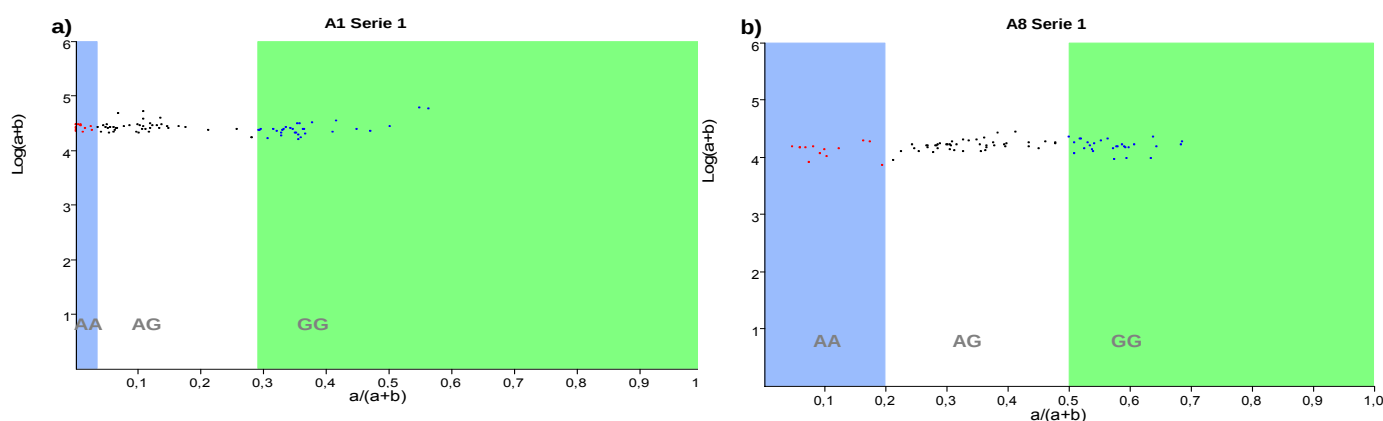


Abbildung 37: Abbildung der Plots der Probenserie 1 für die Marker A1 und A8 ohne Korrektur der Werte. Die Datenpunkte zeigen bei beiden Markern eine breite Streuung entlang der x-Achse auf, können jedoch eindeutig ihren Genotypen zugeordnet werden.

SNP A32 zeigte in der Analyse der Probenserie 1 eine sehr gute Clusterung der Datensätze mit einem Ausfall eines einzelnen Datenpunktes der Probe 2a (Abbildung 38 a, Pfeil). Die Datenpunkte waren hier bis auf den heterozygoten Bereich eng nebeneinander gruppiert. Die Situation änderte sich erheblich durch die Korrektur der Datensätze. Dabei fand eine Rechtsverschiebung der heterozygoten Proben statt. Die Mehrheit der Datenpunkte lag zwar enger beieinander, einzelne Datenpunkte blieben jedoch im vorderen Bereich der Skala. Weitere Proben hingegen konnten in ihrer Wertezuordnung nicht mehr von den Datensätzen des CC-Genotyps unterschieden werden (Abbildung 38 a, eingekreister Bereich).

Diese Vermischung der Datensätze fand sich auch in der Analyse der Probenserie 2 wieder. Zusätzlich konnte eine Streuung der Daten über den vollständigen Wertebereich zwischen 0 und 1 beobachtet werden, wobei sich die Datenwerte des TT-Genotyps untereinander in einem Subcluster gruppierten (Abbildung 38 d, umkreisten Bereiche). Vier der TT-homo-

zygoten Proben waren nach einer Korrektur auf der Skala in die Wertebereiche heterozygoter Marker verschoben (0,523/0,475 und 0,504/0,480) (Abbildung 38 c, eingekreister Bereich).

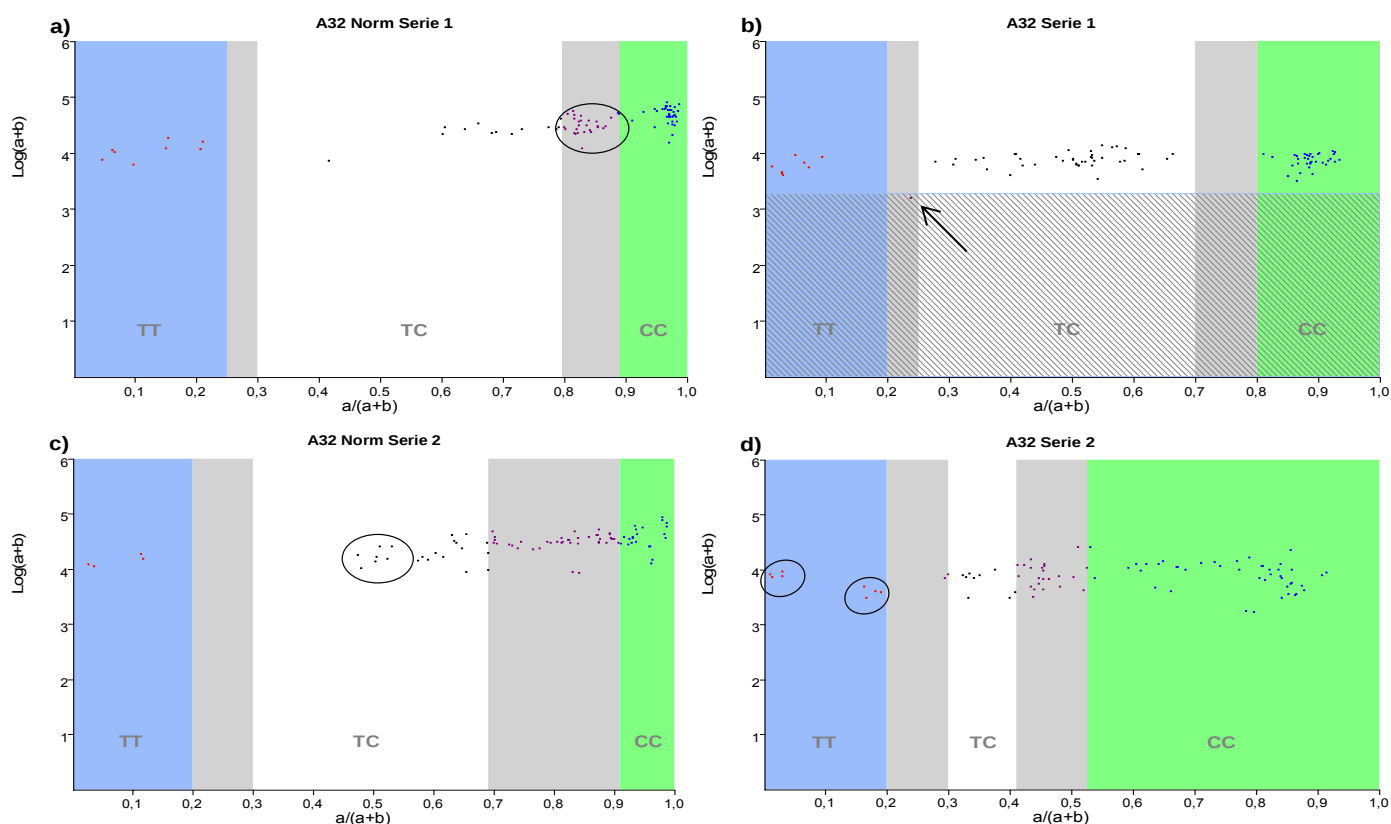


Abbildung 38: Abbildung der Plots der Gesamtanalyse für den Marker A32. b) Die Typisierung der Werte der Probenserie 1 zeigte bis auf einen Ausfall der Probe 2 (Pfeil) eine gute Gruppierung der Genotypen. a) Nach der Korrektur erfolgte eine Verschiebung der heterozygoten Datenpunkte, so dass keine eindeutige Zuordnung stattfinden konnte, (graue Fläche). c) und d) Die Probenserie 2 konnte aufgrund von Streuung und schlechter Gruppierung (eingekreister Bereiche) der Wertepunkte innerhalb der heterozygoten und CC-homozygoten Proben nur unzureichend typisiert werden.

Die Problematik der Subcluster wiederholte sich in dieser Gruppe bei den Markern A21, A33 und A37 (siehe Anhang, Abbildung A-13, Abbildung A-20 und Abbildung A-23). Zudem trat dort eine Überlappung der Werte in beiden Probenserien auf, die im grau unterlegten Bereich dargestellt ist. Für die Datenpunkte aus diesem Bereich war nur eine manuelle Nachanalyse der Genotypen möglich.

Der Ausfall mehrerer *spots* auf dem Array in der Probenserie 2 ist für den Marker A37 zu beobachten. Ausgefallen waren hier die Proben L30b, L1b, L38b, L4b, L67b und L22b (Abbildung 39, schraffierte Fläche). Zudem war in dieser Serie im Vergleich zu der gut gruppierten Serie 1 eine deutliche Streuung entlang beider Koordinatenachsen zu beobachten (Abbildung 39 und Abbildung A-23).

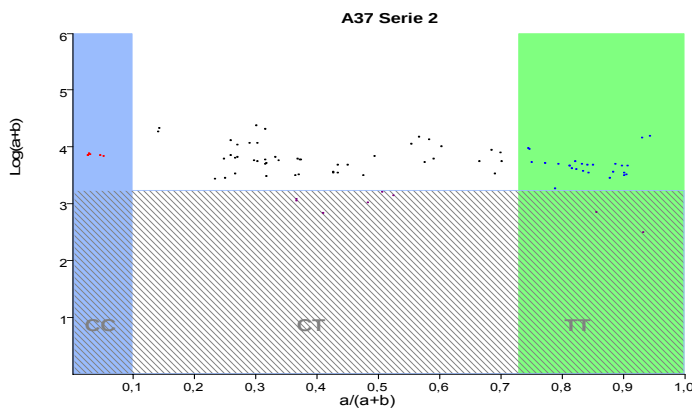


Abbildung 39: Abbildung eines Plots für den Marker A37 mit zahlreichen Probenausfällen innerhalb der Probenserie 2.

Eine breite Streuung nach der Korrektur der Werte ließ sich entlang der x-Achse für den SNP A33 beobachten. Die Genotypengruppierung der Probenserien 1 und 2 erfolgte in eindeutigen Clustern (Abbildung 40 a und Abbildung A-20), die Zuordnung des einzigen homozygoten CC-Genotyps konnte nach der Normalisierung der Werte nicht bestätigt werden (Abbildung 40 b, Pfeil). Dieser Genotyp konnte in der Analyse nur manuell von der Gruppe der heterozygoten Werte getrennt werden (Abbildung 40 a, Pfeil).

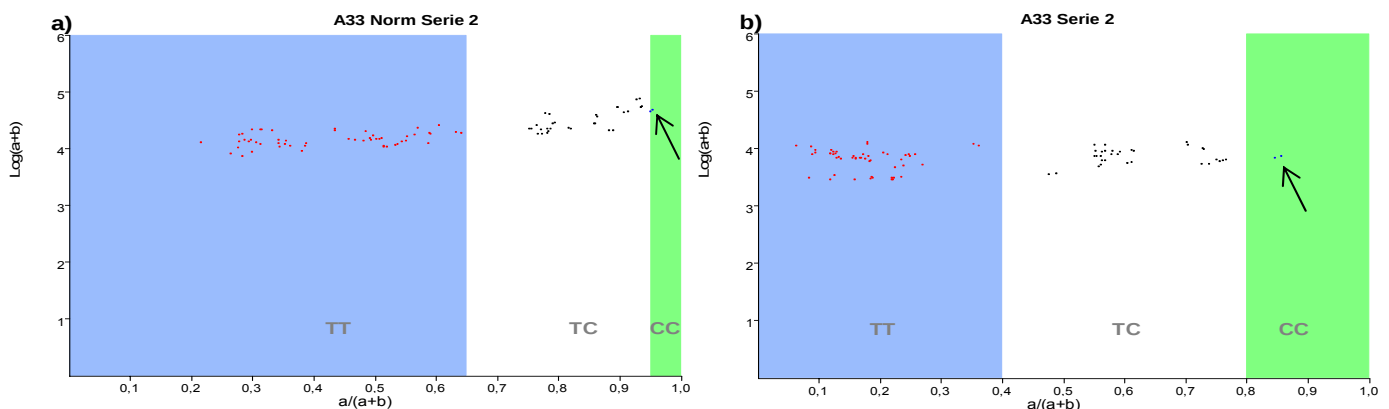


Abbildung 40: Abbildung der Plots der Probenserie 2 für den Marker A33 vor und nach der Korrektur der Werte. Beispiele für eine Probe mit nur einer CC-Heterozygoten (Pfeile). a) Diese konnte nach der Korrektur der Werte nur noch unzureichend von den heterozygoten Proben getrennt werden.

Eine weniger eindeutige Gruppierung der Genotypen in den Probenserien 1 und 2 wurde für den SNP A53 nachgewiesen (Anhang Abbildung A-33 und Abbildung 41). Diese blieb nach der Korrektur der Werte fast unverändert. Auffällig hier waren jedoch die beiden klar abgegrenzten homozygoten Proben im GG-Genotyp (Abbildung 41 a, Pfeile). In der Probenserie 2 konnte nur eine der Proben reproduziert werden (L3b). Die zweite Probe (L44b) lieferte keine Signale auf dem Array und fiel aus. Ein Einzelner der zugehörigen Datenpunkte für die ausgefallene Probe konnte jedoch gefunden werden. Dieser befand sich außerhalb des Plots, fast am Nullpunkt der x- und der y-Achse (Abbildung 41 b, Pfeil).

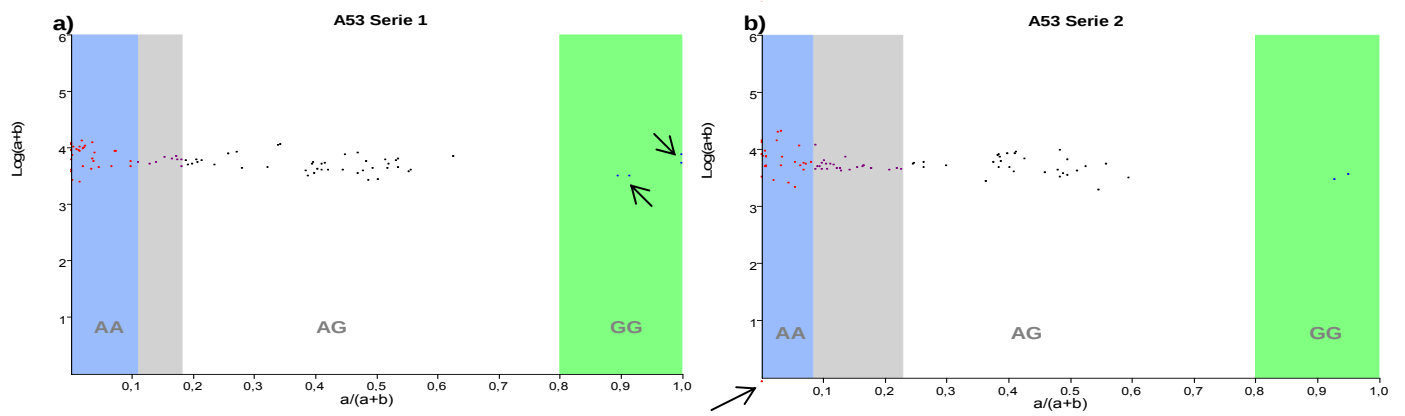


Abbildung 41: Abbildung der Plots der Probenreihen 1 und 2 für den Marker A53. Der GG-homozygote Genotyp konnte nur für zwei Proben nachgewiesen werden (Pfeile). b) Durch eine geringere Analysequalität wurde eine der homozygoten Proben in der Serie 2 nicht typisiert und konnte nur außerhalb des Plots detektiert werden (Pfeil).

4.2.2.1.3 Systeme mit schlechter Clusterqualität

Bei neun der 33 Marker konnte die Typisierung nur für eine geringe Anzahl von Proben erfolgreich durchgeführt werden. In allen Fällen war eine manuelle Korrektur der automatisierten Auswertung notwendig, da der überwiegende Teil der Proben nicht eindeutig zugeordnet werden konnte. In dieser Kategorie wurden SNPs zusammengefasst, die zu einem überwiegenden Teil Datenpunkte aufwiesen, die nicht von den Datenpunkten der benachbarten Genotypen zu trennen waren. Die Plots wiesen zu einem erheblichen Prozentsatz graue Flächen auf. Es wurden nur wenige (SNP A9, Abbildung 42 a), bzw. keine erkennbaren Heterozygoten (SNP A11, Abbildung 42 b) detektiert.

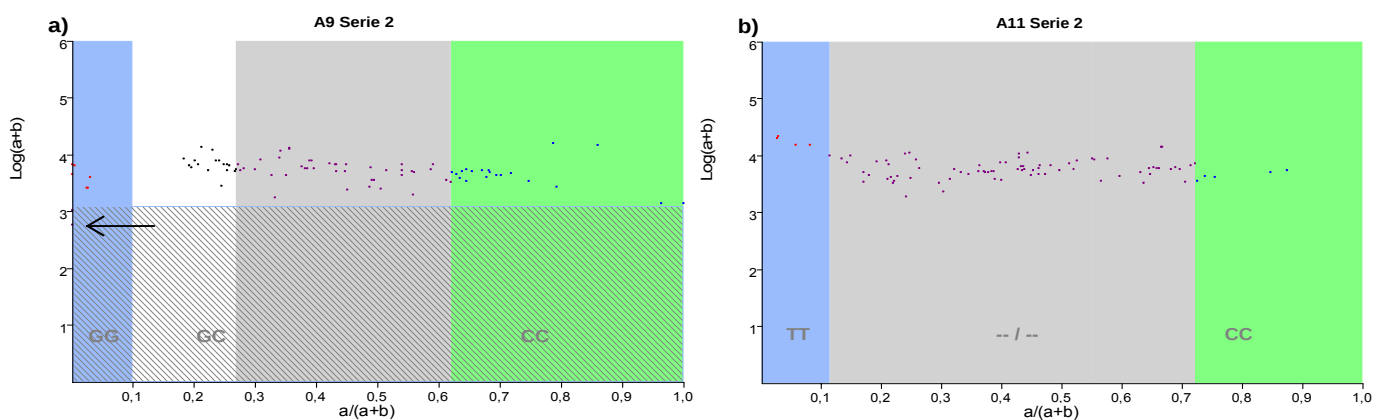


Abbildung 42: Abbildung von zwei Plotbeispielen mit schlechter Clusterqualität. a) Beim Marker A9 konnten nur wenige Polymorphismen korrekt zugeordnet werden (blaue, weiße und grüne Flächen). b) Beim Marker A11 kam es zu einer fast vollständigen Überlappung der Wertepunkte und es fand keine Zuordnung statt (graue Fläche).

Das Spektrum der Ausfälle reichte von einzelnen Proben wie für den SNP A9 (Probe E35b, Abbildung 42 a, Pfeil), für die SNPs A16 und A35 (Probe L44b, Abbildung A-10 und Abbildung A-26) und für die SNPs A25, A34 und A41 (Probe L4b, Abbildung A-16, Abbildung A-21 und Abbildung A-26). Im Fall des Markers A20 konnte bei der Mehrzahl der Proben der Genotyp aufgrund schwacher Signale nicht bestimmt werden. Die zugehörigen Punkte dieser Proben befanden sich im Plot unter der schraffierten Fläche und wurden nicht in die weitere Analyse einbezogen (Abbildung 43). Zusätzlich wurden in der Probenserie 1 Wertepunkte detektiert, die weit außerhalb des Plots lagen (Abbildung 43, Pfeil und Abbildung A-12 im Anhang).

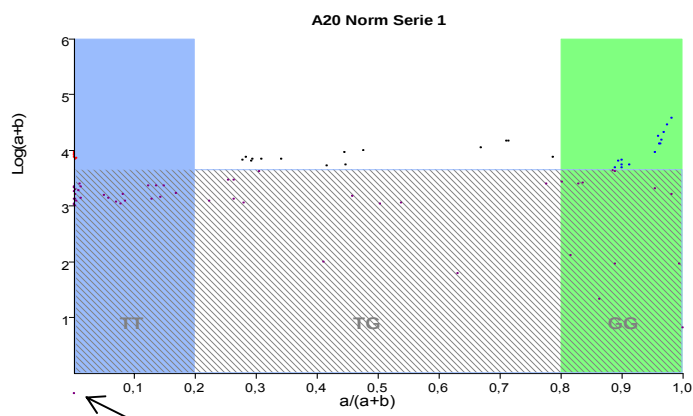


Abbildung 43: Plotdarstellung für den Marker A20. Die Mehrheit der Proben konnte für diesen Marker nicht analysiert werden (schraffierte Fläche). Zusätzlich wurde ein einzelner Datenpunkt unterhalb der x-Achse detektiert (Pfeil).

Mit besonderer Sorgfalt und unter kritischer Betrachtung fand die Auswertung der letzten Gruppe statt. Eine automatisierte Genotypisierung war nur in wenigen Fällen erfolgreich. Bereiche, die flächig grau unterlegt waren, wurden nicht nachtypisiert. Dieses wurde in der Tabelle als ein vollständiger Ausfall des jeweiligen Markers dargestellt (SNP A41, Abbildung 44).

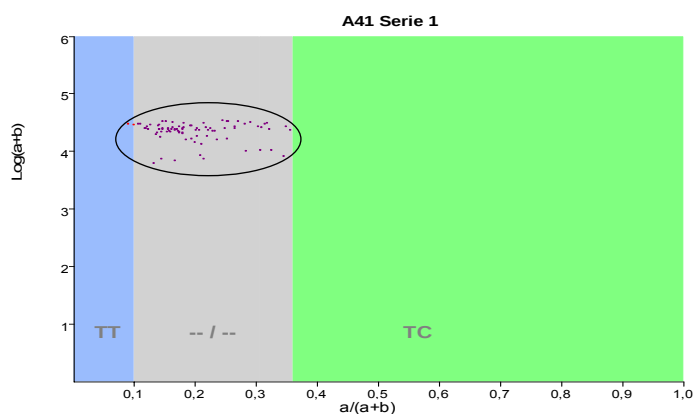


Abbildung 44: Abbildung des Plots für den Marker A41 bei dem keiner der Proben der korrekte Genotyp zugeordnet werden konnte. Alle Datenpunkte gruppieren sich im Bereich der grauen Fläche.

Gesondert muss der Marker A23 betrachtet werden. Dessen Genotypen konnten nach dem Auswerten der Probenserie 1 nur in wenigen Fällen zugeordnet werden (Abbildung 45 a und b, blaue und grüne Fläche). Es wurde vermutet, dass die Ursache dafür in einer fehlerhaften Primerkonzentration, bedingt durch einen Pipettierfehler begründet war. Für die nachfolgenden Analysen der Serie 2 wurden sowohl für die PCR-, als auch für die Minisequenzierungsreaktion frisch angesetzte Primer verwendet.

Als Ergebnis erfolgte eine erfolgreiche Analyse fast aller Proben. Es war eine deutliche Trennung der Genotypen zu erkennen. Eine automatisierte Typisierung konnte durchgeführt werden (Abbildung 45 c und d).

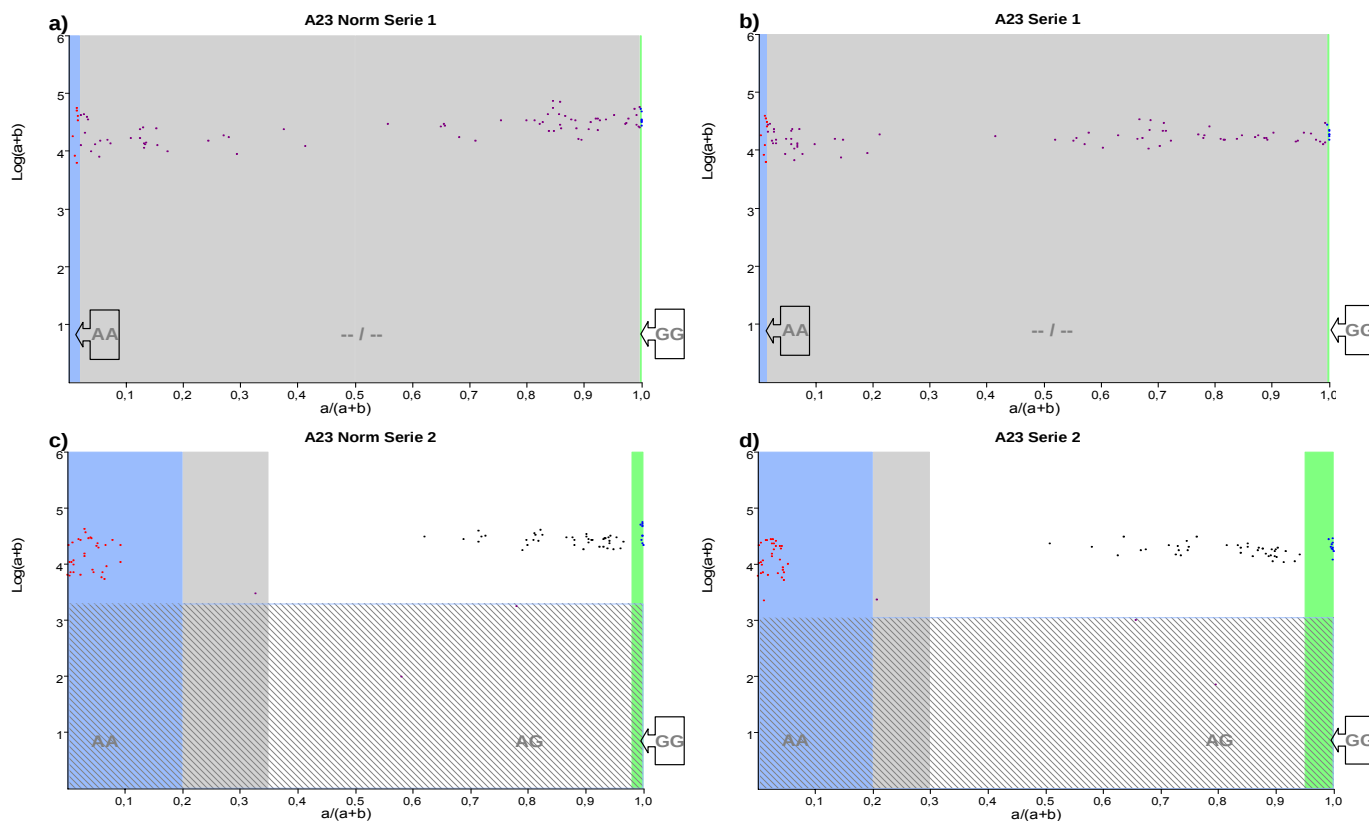


Abbildung 45: Darstellung der vollständigen Analyse der Datenpunkte für den SNP A23. a) und b) Hier konnte die Probenserie 1 aufgrund eines Fehlers nicht typisiert werden. c) und d) Die Probenserie 2 zeigte hingegen eine korrekte Typisierung der Proben und nur wenige Ausfälle (schraffierte Flächen).

In Einzelfällen in denen Datenpunkte manuell ausgewertet wurden, geschah dies ausschließlich nach genauer Betrachtung der Intensitätswerte und der Verhältnisse und wurde nur dann zugelassen, wenn die Werte für $\frac{a}{a+b}$, eine Distanz von 0,001-0,04 zur jeweiligen Typisierungsgrenze aufwiesen. Der Wert wurde rein deskriptiv festgelegt und konnte in den Fällen Anwendung finden, wenn die grau unterlegte Fläche eine leichte Vermischung der Genotypen aufwies. Das Einbauverhalten der Marker konnte innerhalb der Proben als einheitlich festgehalten werden, so dass eine Betrachtung der Verhältnisse $\frac{a}{b}$ eine solche Nachtypisierung zuließ.

4.2.3 Auswertung der Daten

4.2.3.1 Referenzmethoden

Die Evaluierung der Microarray-Analyse basierte auf der vergleichenden Betrachtung der Typisierungsergebnisse mit den bereits veröffentlichten Ergebnissen der Studie zur Evaluierung der *Genplex*-Methode (Phillips *et al.* 2007a). Ergebnisse aus den *SNaPshot*TM-Typisierungen sowie der Genotypisierung mittels des *Genplex*-Verfahrens wurden den Ergebnissen des Microarray-Verfahrens gegenübergestellt (Tabelle A-9, im Anhang).

Aus den unterschiedlich großen Marker-Pools der Referenzmethoden wurden die SNPs berücksichtigt, die mit der Markerauswahl für das Microarray-Verfahren konkordant waren. Für das Typisierungsverfahren mittels Microarray wurden 33 SNPs selektiert, die in einem Multiplexverfahren stabil hybridisierten. Mittels der *Genplex*-Methode können routinemäßig 48 Marker untersucht werden. Darauf Bezug nehmend wurden mittels *SNaPshot*TM parallel ebenfalls 48 SNPs typisiert. Aus dem Gesamtpool der validierten 52 forensischen Marker wurden somit vier nicht in die Evaluierungsstudie einbezogen. Es handelte sich dabei um die SNPs mit der Datenbankbezeichnung rs2016276, rs826472, rs2830795 und rs1028528. Diese Marker waren jedoch in die 33 SNP Typisierungsreaktion auf Glasobjektträgern integriert. Sie entsprachen der internen Kodierung A15, A23, A42 und A53. Die Marker konnten in die Arbeit mit einbezogen werden, da eine zusätzliche Typisierung der Proben mit der vollständigen Markerzahl erfolgte und die Daten vom Labor in Santiago de Compostela (C. Phillips, pers. Mitteilung) zur Verfügung gestellt wurden. Zusätzlich wurden in Mainz die einzelnen Marker mittels *SNaPshot*TM in Singleplexen nachtypisiert und abgeglichen.

In der Gesamtauswertung fiel auf, dass bei zwei Markern (rs873196 und rs1528460) in 19 bzw. 18 von 43 untersuchten Proben die Allelsignale der Evaluierung mittels *Genplex*-Methode (Phillips *et al.* 2007a) nicht mit den zugesandten Ergebnissen der 52plex Typisierung mittels *SNaPshot*TM übereinstimmten. Die Ergebnisse der Typisierung durch Microarray waren mit letzteren konkordant (Tabelle 17 a).

Tabelle 17: Ausschnitt aus der Typisierungstabelle für die Proben L38a und L38b (A-9, im Anhang). a) Zusammengeordnete Genotypen der Marker rs873196 und rs1528460 aus den Referenzdaten und Ergebnisse der Typisierung mittels Microarray. b) Überarbeitete Ergebnisse der Referenzdaten. GG-Genotypen aus den *SNaPshot*TM II und *Genplex*-Plattformen wurden für beide Marker nicht in die Gesamtauswertung einbezogen (NN).

a)

	rs873196	rs8037429	rs1528460
	A25	A50	A54
	SNP14_2	SNP15_2	SNP15_3
<i>SNaPshot</i> TM I	L38a AG	AG	AG
<i>SNaPshot</i> TM II	L38a GG	AG	GG
<i>Genplex</i>	L38a GG	AG	GG
Microarray-Plattform	L38a AG	TC	AG
<i>SNaPshot</i> TM I	L38b -	-	-
<i>SNaPshot</i> TM II	L38b -	-	-
<i>Genplex</i>	L38b GG	AG	GG
Microarray-Plattform	L38b AG	TC	AG

b)

rs873196	rs8037429	rs1528460
A25	A50	A54
SNP14_2	SNP15_2	SNP15_3
AG	AG	AG
NN	AG	NN
NN	AG	NN
AG	TC	AG
-	-	-
-	-	-
NN	AG	NN
AG	TC	AG

Bei der Betrachtung aller typisierten Proben für beide Marker fällt auf, dass in den Referenzmethoden *SNaPshot*TM II und *Genplex* keine heterozygoten Genotypen detektiert wurden. Anders in der *SNaPshot*TM I-Referenztypisierung und in den Ergebnissen der Microarray-Methode (Tabelle 17 b). Die Ergebnisse der Referenzmethoden wurden zunächst untereinander verglichen. Dabei wurde festgestellt, dass es sich ausschließlich um den Genotyp AG bzw. GG handelte. Fünf Proben wurden daraufhin sequenziert, um eine genaue Kenntnis über diesen Locus zu erhalten (Abbildung 46).

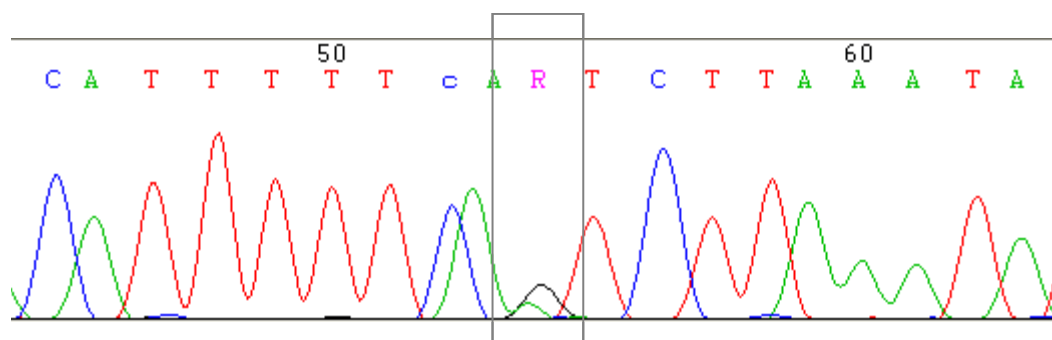


Abbildung 46: Ausschnitt der sequenzierten Probe E97a mit nachgewiesenem AG-Polymorphismus (Rahmen) für den Nachweis der falsch typisierten Genotypen in den Referenzen *SNaPshot*TM II und *Genplex*.

Für die folgende Bestimmung des Genotypisierungserfolges wurden falsche Homozygoten korrigiert und als Fehltypisierung in die Gesamtauswertung einbezogen (Tabelle 17 b, Tabelle A-9). Überprüfte man die Ergebnisse der Microarray-Typisierung, so waren die Verhältnisse der Farbstoffintensitäten eindeutig interpretierbar und ließen keinen Zweifel am heterozygoten Genotyp der jeweiligen Probe zu.

4.2.3.2 Typisierungserfolg

Die Summe der in die Studie einbezogenen Allele betrug 1419 für die vollständigen Profile der 33 SNPs und 1247 für die Referenzprofile mit 29 typisierten SNPs.

In der Gesamtauswertung des Microarray-Verfahrens wurden die Typisierungen der Proben mit und ohne Korrektur nebeneinander gestellt und verglichen.

Tabelle 18 gibt einen Überblick der erzielten Typisierungsergebnisse aller Verfahren wieder. Sie umfasst die Typisierungserfolge der jeweiligen Methode sowie die prozentuale Übereinstimmung der Ergebnisse aller Probenreplikate.

Aus der Anzahl der potentiellen Allele und der Gesamtheit der erfolgreich typisierten Genotypen, wobei in diese die Summe der editierten Genotypen einfließt, ergaben sich die prozentualen Angaben des Genotypisierungserfolges. Für die Microarray-Studie betrug der Typisierungserfolg 85,52% nach einer Korrektur mit den Intensitätswerten und 86,75% ohne die Korrektur. Die Typisierungsleistungen der Referenzstudien lagen zwischen 98,4% für die *SNaPshot*TM I-Studie und 99,92% für die *Genplex*-Studie.

Tabelle 18: Zusammengefassten Ergebnisse der Typisierung von 43 Proben mittels Microarray und der Referenzmethoden.

Vergleichsstudie	N	Σ typisierter SNPs	Σ typisierter Allele	Σ typisierter Genotypen	Σ ausgefallener Genotypen	Σ abweichender Genotypen	Σ editierter Ergebnisse	% Typisierungs- erfolg	% Übereinstim- mung mit Replikaten
Microarray – Plattform Mainz mit Korrektur	43	33	1419	2462	411	92	67	85,52	96,21
Microarray – Plattform Mainz ohne Korrektur	43	33	1419	2427	376	112	118	86,75	95,45
<i>SNaPshot</i> TM I Santiago	43	33	1419	1547	4	19	0	99,74	98,77
<i>SNaPshot</i> TM II Santiago	43	29	1247	1227	20	38	0	98,4	96,9
Genplex Gesamtstudie Santiago, London, Köln	43	29	1247	2405	2	64	1	99,92	97,34

Betrachtet man die Gesamtheit aller typisierten Genotypen und die Summe der von der Gesamtübereinstimmung abweichenden Genotypen, so zeigte sich eine prozentuale Konkordanz von 96,21% für den Microarray-Ansatz nach einer Korrektur und 95,45% für den Ansatz ohne Korrektur.

Die Prozentwerte der Referenzstudien zeigen eine Spanne der Übereinstimmung zwischen 96,9% für die *SNaPshot*TM II-Studie und 98,77% für die Ergebnisse der Probenotypisierung aus Santiago de Compostela, die ebenfalls mittels der *SNaPshot*TM-Methode (*SNaPshot*TM I) durchgeführt wurde.

Die Summe möglicher Genotypen, die aus dem Produkt der 33 untersuchten SNPs und den 43 Probenpaaren hervorgeht, lässt sich anschaulich in den beiden Kreisdiagrammen (Diagramm 5 a und b) in ihrer versuchsbedingten Zusammenstellung darstellen. Nach der

postexperimentellen Korrektur der Werte setzte sich die Summe (87,3%) der erfolgreich typisierten Polymorphismen aus 83,3% der detektierten Genotypen und aus 4% der in der manuellen Analyse nachtypisierten Ergebnisse zusammen. 12,7% der Genotypen fielen aus oder konnten in der Auswertung keinem Genotypcluster eindeutig zugeordnet werden (Diagramm 5 a).

Vergleichbare Ergebnisse ergaben sich in der Analyse der Werte ohne Korrektur (Diagramm 5 b). Die Differenz des Typisierungserfolges zum normalisierten Diagramm betrug 1,5%. Hierbei unterschieden sich die Prozentsätze der automatisch zugeordneten Genotypen um 0,2 Prozentpunkte zugunsten der nicht korrigierten Werte. Nach einer manuellen Korrektur konnten zusätzlich nur 2,3% der Datensätze ihrem Genotyp zugeordnet werden. In diesem Ansatz betrug der Prozentsatz der nicht typisierten Genotypen 14,1%.

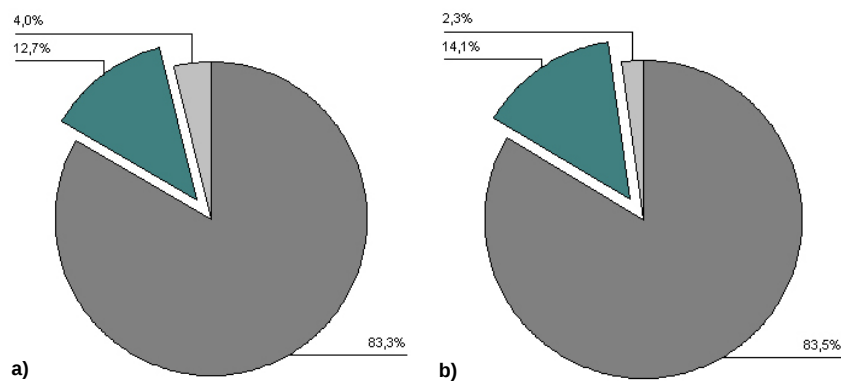


Diagramm 5 Typisierungserfolg (dunkelgraue Flächen) unter Berücksichtigung editierter Genotypen (hellgraue Flächen). Prozentsätze der nicht typisierten Genotypen sind aus dem Diagrammsegment herausgerückt und blau. a) Darstellung des Typisierungserfolges nach der Wertekorrektur. b) Diagramm mit Werten ohne Korrektur.

Um die Typisierungseffizienz der in der vorliegenden Arbeit untersuchten Methode und der Referenzmethoden vergleichend darstellen zu können, wurden die durchschnittlichen Erfolgsquoten der Studien berechnet. Zunächst wurde der Typisierungserfolg für jede Probe errechnet und dieser anschließend in Verhältnis zur Gesamtkonkordanz gesetzt. Die durchschnittliche Differenz aller typisierten Proben konnte so ermittelt werden und ergab eine Wertung der Typisierungseffizienz der einzelnen Plattformen.

Die Plattform mit der höchsten Effizienz im positiven Wertebereich zeichnet sich durch die beste Erfolgsquote bezüglich der Gesamttypisierung aus. Geringere Typisierungsquoten werden durch niedrigere Werte in der graphischen Darstellung verdeutlicht. Fallen diese unter den Durchschnitt, liegen sie im negativen Wertebereich (Diagramm 6).

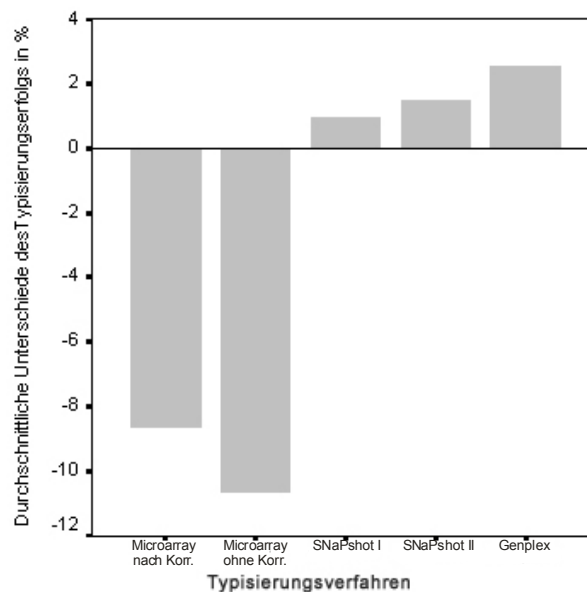


Diagramm 6: Vergleichende Darstellung der Typisierungseffizienz aller untersuchten Studien im Verhältnis zu ihrer Gesamtkonkordanz.

Die Erfassung der aus den Ergebnissen der unterschiedlichen Studien ermittelten Allelfrequenzen konnte gezielt dazu verwendet werden, die Qualität der Analysen zu visualisieren. Betrachtet man die Verteilung im Balkendiagramm kann man zwischen dem untersuchten Verfahren und den Referenzmethoden eine deutliche Übereinstimmung der Allelhäufigkeiten feststellen (Diagramm 7). Auffällig ist dabei der Marker A41 (Diagramm 7, roter Pfeil), der in beiden Typisierungsdurchläufen fast durchgängig ausgefallen war. Die Genotypen konnten nur für drei Proben ermittelt werden. Diese waren homozygot und ergaben somit die Allelfrequenz 1.

Kleine Differenzen der Häufigkeitsverteilung sind bei zwei weiteren Markern (A25 und A54, Diagramm 7, grüne Pfeile) zu beobachten. Die Unterschiede zeigen sich zwischen den Studien des *Genplex* und der *SNaPshot*TM II-Studie sowie zwischen *SNaPshot*TM I und der Microarrayanalyse. Es handelt sich hierbei um die Unterschiede in der Genotypisierung beider SNPs (siehe Kapitel 4.2.3.1) auf den unterschiedlichen Plattformen. Durch das veränderte Verhältnis des heterozygoten AG- und des homozygoten AA-Genotyps resultiert daraus eine Verschiebung der Frequenz zugunsten des A-Allels in beiden Markern der Microarray- und der *SNaPshot*TM I-Studie.

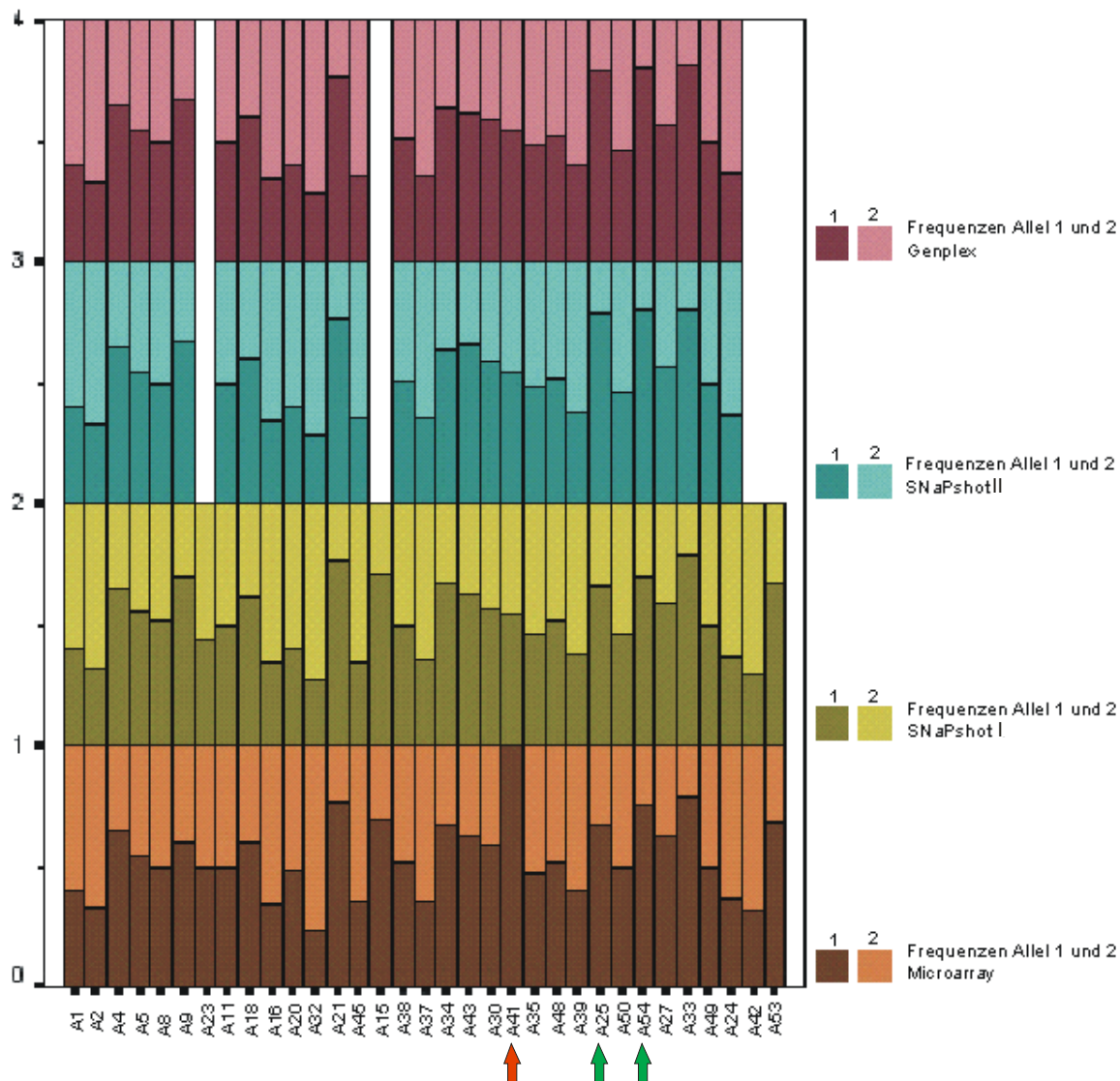


Diagramm 7: Graphische Darstellung der Allelfrequenzverteilung der vier Vergleichsstudien über die Gesamtanzahl der typisierten Marker. In den Analysen der *Genplex*- und *SNaPshot*TM II-Methoden fehlenden SNPs A23, A15, A42 und A53 sind durch weiße Flächen gekennzeichnet. Roter Pfeil kennzeichnet den Marker, welcher in der Studie nicht korrekt typisiert werden konnte. Grüne Pfeile deuten auf Marker mit einer Verschiebung der Allelfrequenzen aufgrund von Profilunterschieden.

Zusätzlich wurde die graphische Auswertung durch eine Überprüfung der Signifikanz der Varianzenverteilung untermauert. Die Überprüfung mit Hilfe des t-Tests für unabhängige Variablen wurde für die Ergebnisse der Microarray-Studie im Vergleich zu den Referenzmethoden durchgeführt. Die Ergebnisse zeigten bei einem Signifikanzniveau von $p < 0,05$ keine signifikanten Unterschiede zwischen den Vergleichspaaren (Tabelle 19).

Tabelle 19: Auflistung der p -Werte für den t-Test unabhängiger Variablen. Der Test wurde zweiseitig durchgeführt (Prüfung auf Gleichheit/Ungleichheit der Varianzen).

Werte	Sig. (2-seitig) $p=$	H1 $p<0,05$
Microarray vs. <i>SNaPshot</i> TM I	0,524/0,524	nicht signifikant
Microarray vs. <i>SNaPshot</i> TM II	0,865/0,864	nicht signifikant
Microarray vs. <i>Genplex</i>	0,844/0,843	nicht signifikant

Aufgrund des deutlich geringeren Typisierungserfolges der Microarray-Studie werden im weiteren Verlauf die potentiellen molekularbiologischen Faktoren besprochen, die auf die Resultate Einfluss nehmen können.

Sterische Inhibitionen beim Einbau von Farbstoffen sowie bei der Immobilisierung der Hybridisierungsprodukte auf dem Trägermaterial sind bekannt und nehmen auch in der vorliegenden Studie Einfluss auf ihre Durchführbarkeit.

Eine genauere Untersuchung dieser physikalischen Phänomene konnte in der vorliegenden Arbeit nicht durchgeführt werden. Es konnten jedoch sequenzabhängige Störungsmechanismen erfasst und analysiert werden.

Wie im Kapitel 4.2.2.1.1 - 4.2.2.1.3 beschrieben wurde bei der Auswertung eine Unterteilung der Ergebnisse in die „schlechten“, „mittleren“ und „guten“ Wertungsbereiche eingeführt. Zusammengefasst wurde zunächst, welche Polymorphismen in der Versuchsanordnung eine Rolle spielten. Alle sechs potentiellen Basenkombinationen wurden detektiert. Die Häufigkeit der untersuchten Polymorphismen war in der Versuchsanordnung jedoch nicht gleich verteilt. Der häufigste untersuchte Polymorphismus war C/T, der seltenste, nämlich nur einmal in seiner Kombination zu typisierende Polymorphismus, war A/C. Die Gesamtverteilung ist im Diagramm 8 graphisch dargestellt und tabellarisch aufgelistet. Auffallend ist, dass kein G/C-Polymorphismus in der „guten“ Kategorie detektiert wurde. Sie sind lediglich innerhalb der schlecht und mittelmäßig diskriminierenden Kategorien zu finden. Die beiden untersuchten T/A-Polymorphismen konnten dagegen alle vollständig typisiert werden.

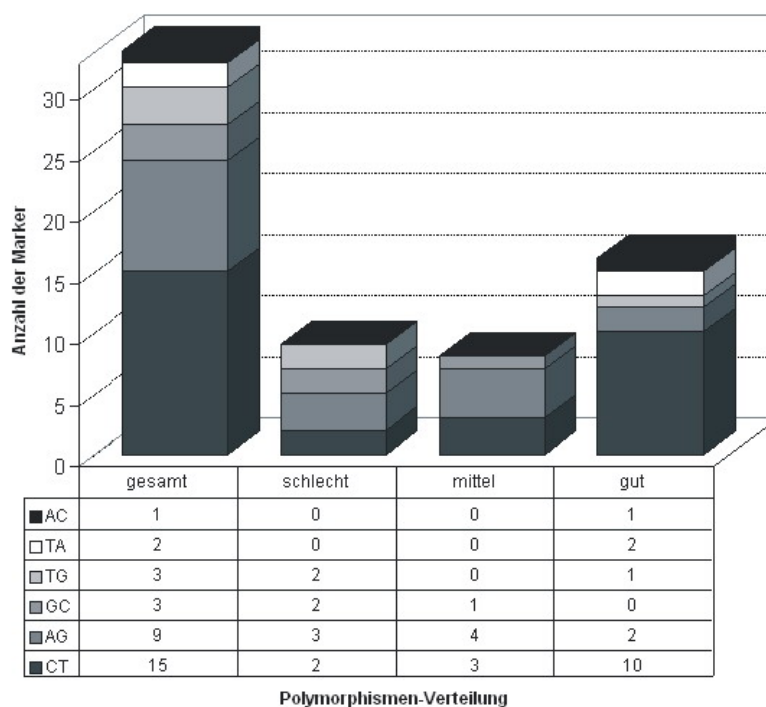


Diagramm 8: Häufigkeitsverteilung der untersuchten SNP-Kombinationen in Microarray-Experimenten. Der linke Balken umfasst alle untersuchten Polymorphismen, die drei folgenden Balken stellen die Verteilung der Polymorphismen hinsichtlich des Typisierungserfolges dar.

In Ergänzung zur Untersuchung der Einflussnahme des Polymorphismus auf den Genotypisierungserfolg wurden die Häufigkeiten der vier 5'-flankierenden Basen betrachtet, die als Bindungspartner für die jeweilig einzubauende Base der polymorphen Partner diente. Diese wurden in drei Kategorien des Typisierungserfolges eingeordnet (Tabelle 20).

Tabelle 20: Prozentualer Anteil der an den SNP 5'-grenzenden Basen beim Typisierungserfolg.

benachbarte Base	schlecht (%)	mittel (%)	gut (%)
A	22,2	25	31,3
T	11,1	12,5	37,5
C	44,4	25	18,8
G	22,2	37,5	12,5

Auf diese Weise konnte gezeigt werden, dass Cytosin als Nachbarnukleotid den größten prozentualen Anteil (44,44%) an fehlerhafter oder schwieriger Inkorporation darstellt. Die Nachbarsequenz mit einem 3'-terminalen A- oder T-Nukleotid, begünstigt hingegen den korrekten Einbau der an die Fluoreszenzfarbstoffe gekoppelten Base (31,25 und 37,5%).

4.2.4 Klassifizierung der Daten mittels Silhouette-Koeffizienten

Eine sehr gute Möglichkeit, einen Gesamtüberblick über die Qualität eines Assays zu erhalten, bot die Berechnung der *Silhouette*-Koeffizienten und die Darstellung der Ergebnisse mit Hilfe eines Verbundliniendiagramms (Diagramm 9). Sie diente als Kontrolle der bereits durchgeführten Clusteranalyse und erlaubte einen zusätzlichen wertenden Blick auf die Microarray-Studie.

Die *Silhouette*-Werte wurden im *ClusterA*-Programm für jeden Marker ausgerechnet. Es wurden beide DNA-Serien (1 und 2) sowie die Ergebnisse der analysierten Probenserien nach der durchgeführten Korrektur berücksichtigt. Die Qualität aller Ansätze lässt sich anhand der *score*-Werte auf der y-Achse ablesen. Der Wert 0,25 (rote Gitternetzlinie) wurde aus der Literatur entnommen (Liu *et al.* 2003, Lovmar *et al.* 2005) und beschreibt den niedrigsten *score*-Wert, anhand dessen die Clusterdistanzen noch zu einer zuverlässigen Genotypen-Diskriminierung führen kann.

Die größte Streubreite erzielte Probenserie 1 nach Korrektur der Werte (Diagramm 9, Serie 1_Norm). Die Werte erstrecken sich von 0,90 (rs10495407) bis -1,0 (rs737681). Etwas mehr als die Hälfte der Marker (19 von 33) liegen oberhalb des Wertes von 0,25. Betrachtet man dagegen die Ergebnisse der Analyse ohne eine zusätzliche Korrektur der Werte (Serie 1) befinden sich 21 von 33 Marker oberhalb dieses Wertes. Die Werte entlang der Verbundlinie haben eine geringe Streubreite. Diese reicht vom maximalen Wert von 0,84 (rs2831700) bis -0,37 (rs873196).

Die Analyse der Probenserie 2 erschien von geringerer Qualität. Die maximalen *score*-Werte lagen in den Bereichen von 0,73 (Serie 2_Norm, rs1493232) und 0,71 (Serie 2 – rs8037429), die Werte im negativen Bereich bei -0,51 (Serie 2_Norm, rs737681) und -1,0 (Serie 2, rs729172). Die Mittelwerte für alle vier Analysen waren 0,203 (Serie 1_Norm), 0,33 (Serie 1), 0,053 (Serie 2_Norm) und 0,16 (Serie 2).

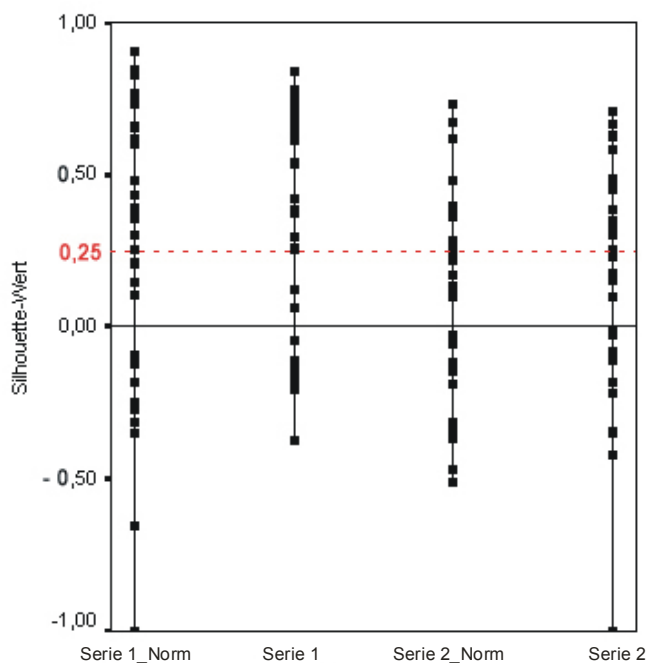


Diagramm 9: Darstellung der *Silhouette*-Werte im Verbundliniendiagramm. Beide Probenserien wurden nach und vor der Korrektur mit Hilfe der so genannten *score*-Werte dargestellt, um vergleichend die Qualität der Cluster innerhalb der Gesamtauswertung zu betrachten. Jedes schwarze Rechteck steht für den Wertepunkt einer Probe. Als Richtmaß für eine zuverlässige Genotypisierung dient die rote Linie am Wertepunkt 0,25.

4.2.5 Ergebnisse der Sensitivitätsstudie

Um zu untersuchen, welche Leistungen die Multiplexreaktion hinsichtlich des Typisierungsverhaltens bei geringen DNA-Mengen erzielt, wurden zwei Proben in unterschiedlichen DNA-Konzentrationen typisiert.

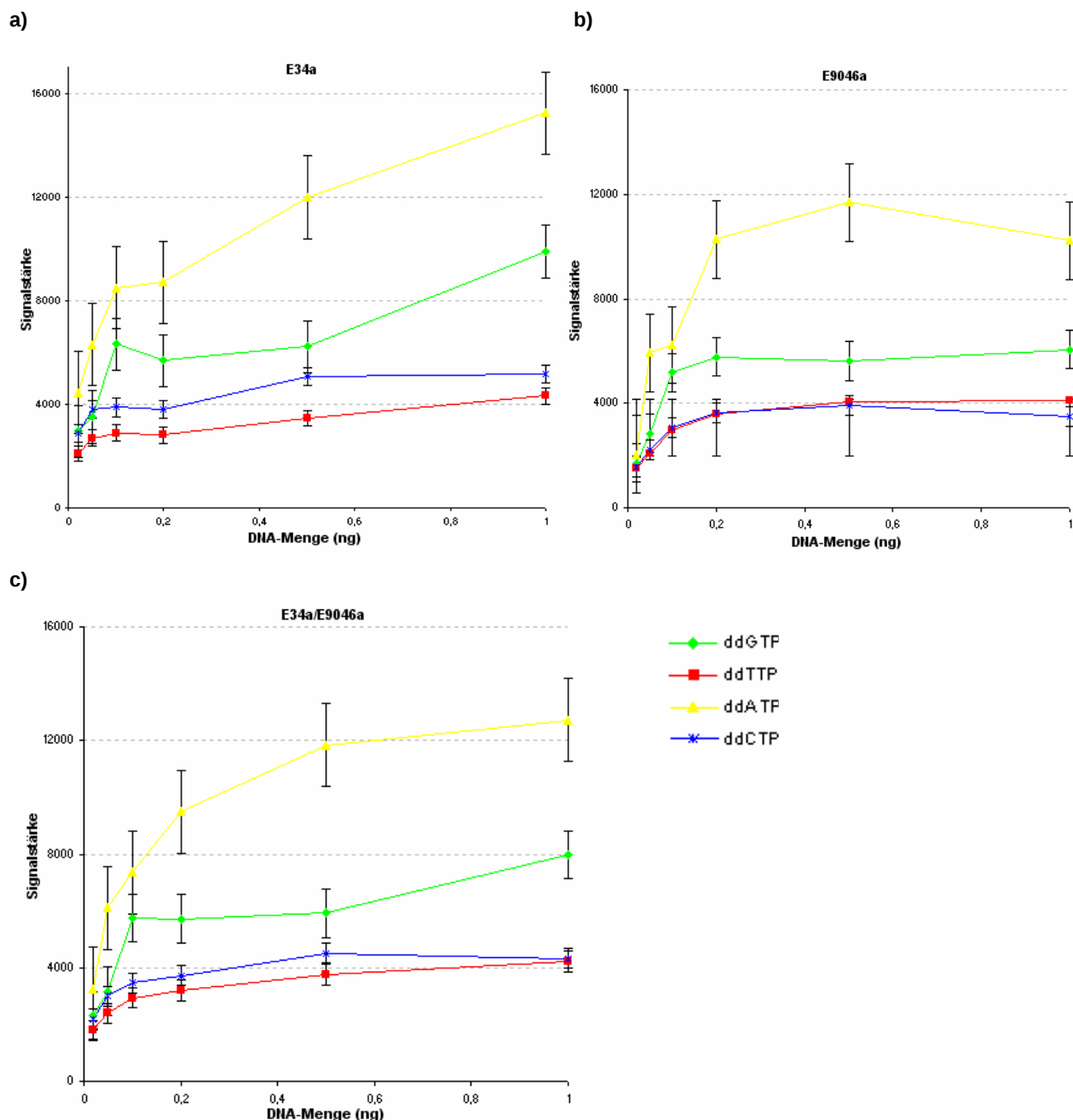


Diagramm 10: Darstellung der Verdünnungsreihe für die Proben a) E34a und b) E9046a. Die eingesetzte DNA-Menge in der PCR lag bei 0,02; 0,05; 0,1; 0,2; 0,5 und 1 ng. Die höchste gemessene Signalintensität betrug über 15 000 RFU in der Probe E34a für den an ddATP gekoppelten *Texas Red*-Farbstoff. In der Abbildung c) wurden die Mittelwerte beider Proben dargestellt.

Anhand von Fehlerbalkendiagrammen mit berechnetem Standardfehler wird die Tendenz der Signalstärkenverteilung bezüglich der eingesetzten DNA-Menge dargestellt. Vergleichbare Intensitätswerte konnten für die ddATP- und ddGTP-Nukleotide bei 0,5 ng

DNA gemessen werden (Diagramm 10 a und b). Bei beiden Proben weist ddATP (*Texas Red*-Farbstoff) die stärksten mittleren Intensitäten auf, gefolgt von ddGTP im mittleren Bereich und ddTTP und ddCTP im unteren Intensitätsbereich (Diagramm 10 c). Je niedriger die DNA-Konzentration, desto näher rücken die Wertepunkte der Farbstoffintensitäten zusammen, wobei auffällig ist, dass die mit *Texas Red* markierten Nukleotide den stärksten Intensitätsabfall durchlaufen (3,9-fach), während die restlichen Farbstoffe ein deutlich kleineres Gefälle aufweisen (2,2-1,8-fach). Speziell für die DNA-Probe E9046a ist sichtbar, dass die Intensitätswerte bei der eingesetzten DNA-Menge von 0,02 ng für alle Farbstoffe auf ein niedriges und fast identisches Niveau fallen können (Diagramm 10 b).

5 Diskussion

Das Interesse an der Nutzung populationsgenetischer Daten und ihre weiterführende Anwendung in der forensischen Genetik und in populationsgenetischen Analysen erfordert die stetige Weiterentwicklung der Untersuchungsmethoden. Im Hinblick auf die Kriterien, welche für die Wahl der untersuchten Marker und die zur Untersuchung herangezogenen Stichproben formuliert wurden, beschäftigt sich der erste Teil dieser Arbeit mit der Etablierung einer Multiplexmethode für die Untersuchung autosomaler, biallelischer Polymorphismen zur Identifizierung von Personen. Die Anforderungen an die Untersuchungsmethode setzten die Reproduzierbarkeit der Daten auf molekularer Ebene voraus, um diese in biostatistische Informationen über die genetische Variation umzuwandeln. Die statistische Auswertung der so gewonnenen populationsspezifischen Allelfrequenzen beschreibt die Verteilung der Diversität zwischen den Populationen und den untersuchten Loci. Sie basiert auf der Prüfung des Vorliegens des Hardy-Weinberg-Gleichgewichts und auf der Analyse der beobachteten und der erwarteten Heterozygotität mit dessen Hilfe Kenntnisse über mögliche Populationsstrukturen gewonnen werden können.

Die Analysen wurden in zwei aufeinander folgenden Abschnitten durchgeführt, wobei die Voruntersuchungen und die Auswertung der Daten eigenständig im Mainzer Institut für Rechtsmedizin durchgeführt wurden. Sie basieren auf der Anzahl von 23 Markern, welche im Partnerlabor in Santiago de Compostela ausgewählt wurden. Ihr folgte die Erweiterung der Multiplexanalyse auf insgesamt 52 SNPs. Die Ergebnisse aller am Projekt beteiligten Arbeitsgruppen und die zusammenfassende Diskussion werden im Kapitel 5.2 beschrieben. Die daraus hervorgegangene Veröffentlichung (Sanchez *et al.* 2006) findet sich im Anhang.

5.1 Diskussion zur Vorvalidierung der 23-SNP Multiplexreaktion

5.1.1 Probenumfang und Whole Genome Amplification (WGA)

Als Herausforderung erwies sich die Sicherung eines ausreichenden Probenumfangs der zu typisierenden Individuen. Bei populationsgenetischen Untersuchungen stellt sich die Frage, wie groß der Stichprobenumfang sein sollte, um den statistischen Fehler zu minimieren und eine valide Aussagekraft zu erhalten. Unabhängig von der Zahl der untersuchten Marker und des zu untersuchenden Effekts innerhalb der Populationen oder Subpopulationen ist der Stichprobenumfang oftmals durch die Möglichkeit beschränkt, eine größere Anzahl gewünschter Proben zu beschaffen.

Das Hauskollektiv des Mainzer Instituts umfasste Proben deutscher und ostasiatischer Populationen (Zhang *et al.* 1993, Waiyawuth *et al.* 1998, Schneider *et al.* 1998, Seo *et al.* 1998). Insbesondere von den Proben männlicher japanischer Individuen waren jedoch nur geringe DNA-Mengen vorhanden. Als eine viel versprechende Lösung zur Erhaltung des Kollektivs wurde die Methode der *Whole Genome Amplification* eingesetzt, welche im Besonderen für forensische aber auch für klinische Aufgabenstellungen einen enormen Nutzen darstellen kann (Lasken und Engholm, 2003).

Parallel zu den weiterführenden Sensitivitätsstudien, die mittels STRs durchgeführt worden sind (Schneider *et al.* 2004, Balogh *et al.* 2006) wurden die japanischen Proben mittels WGA für die Validierung der SNP-Multiplexmethoden aufbereitet.

In der Praxis konnten die bereits publizierten Vorzüge des Wirkungsprinzips und der Anwendung nur bezüglich der Steigerung der resultierenden Menge des genetischen Materials bestätigt werden (Cheung und Nelson, 1996, Dean *et al.* 2002, Lovmar *et al.* 2003). Sowohl in den oben genannten Sensitivitätsstudien an STRs als auch für die hier untersuchten SNPs konnte durch die WGA keine für forensische Analysen notwendige Qualität der DNA erzielt werden. Die präferentielle Amplifikation bestimmter Strangabschnitte, die durch unterschiedliche Affinität der Phi29-Polymerase zustande kommt, bedingt die beobachteten Allelverluste (*allelic drop outs*) sowie die Detektion von Zusatzallelen, was zur Typisierung falscher Heterozygoten führen kann. Besonders für kleine Mengen der Ausgangs-DNA und bei Mischspuren kommt die Methode an Grenzen und stellt ihre Anwendung im forensischen Bereich in Frage (Holbrook *et al.* 2005, Balogh *et al.* 2006, Thacker *et al.* 2006, Bergen *et al.* 2005, Barber und Foran 2006). Zur SNP-Detektion in Multiplexreaktionen werden mindestens 0,3 ng DNA benötigt, um zuverlässige Ergebnisse zu erhalten (Lovmar *et al.* 2003). Trotz der sehr hohen Prozessivität des bei dieser Amplifikationsmethode verwendeten Enzyms können nicht alle Abschnitte der DNA in gleicher Zuverlässigkeit kopiert werden. Die größten Defizite zeigen sich hier bei der Amplifikation von GC-reichen Regionen (Bredel *et al.* 2005). Es wird auch vermutet, dass nicht alle chromosomalen Bereiche mit gleicher Effizienz amplifiziert werden und dabei besonders Telomer-nahe Bereiche einen signifikant geringeren Amplifikationserfolg aufweisen können (Tzvetkov *et al.* 2005). Von Nutzen ist jedoch der Einsatz dieser Methode für die dauerhafte Probenerhaltung aus unterschiedlichen Geweben bei ausreichender Menge (ca. 50 ng) und Qualität des Ausgangsmaterials (Arriola *et al.* 2007). Nach einer technischen Lösung für die noch immer ungeklärten Beobachtungen innerhalb der viel versprechenden Anwendung wird stetig geforscht. Die Forscher um Pan *et al.* (2008) haben für die WGA-Amplifikation die Idee aufgegriffen, Trehalose, die als PCR *enhancer* gilt (Spiess *et al.* 2004), gezielt für diese Methode einzusetzen und haben erste Erfolge in der Verbesserung der *template*-Qualität erzielt.

5.1.2 Technischen Voraussetzungen

Die Sequenzqualität spielte ebenfalls bei der Selektion der Marker und der Herstellung der PCR- und Minisequenzierungsprimer eine entscheidende Rolle. Ersteres war ein entscheidender Faktor der Entwicklung eines Multiplexverfahrens. Über drei Millionen SNPs wurden Anfang des neuen Jahrtausends innerhalb des menschlichen Genoms vermutet (Stoneking, 2001) und weit über eine Million biallelische Polymorphismen wurden zur gleichen Zeit publiziert (Sachidanandam *et al.* 2001, *The International HapMap Consortium*, 2005, NCBI³¹). In folgenden Jahren erhöhte sich jedoch sowohl die Zahl der vermuteten als auch der validierten SNPs erheblich (Frazer *et al.* 2007). Bei der Auswahl eines Pools von Markern, die zusammen für forensische Zwecke eingesetzt werden konnten, mussten eine Reihe von Faktoren berücksichtigt werden. Ein ausreichender Abstand von mindestens 100 kb von bekannten Gensequenzen musste eingehalten werden. Die Entdeckung der gekoppelten Vererbung und des Mechanismus der Entkopplung durch „crossing-over“ durch Thomas Hunt Morgan macht eine überlegte Lokalisation der Marker zwingend. Die Berücksichtigung der Kopplungsmöglichkeit zwischen den Markern tritt bei der Erweiterung des Multiplexumfangs besonders in den Vordergrund. Die in der Spuren- und Abstammungsanalyse verwendeten STR-Standardkits sind in der Regel so konzipiert, dass keiner der Marker auf dem gleichen Chromosom liegt. Möchte man jedoch die notwendige höhere Anzahl der SNPs in einer Reaktion kombinieren, werden je Chromosom mehrere Marker gewählt. Eine mögliche Kopplung zwischen den benachbarten Markern muss auf jeden Fall ausgeschlossen werden, um eine korrekte biostatistische Berechnung durchführen zu können. In der Bearbeitung von Spurenfällen kann diese zwar vernachlässigt werden, für spezielle Anforderungen innerhalb der Verwandtschaftsanalyse (z.B. Defizienzfall), vor allem unter der Einbeziehung zusätzlicher Marker ist die Betrachtung unabhängig vererbter Merkmale eine Voraussetzung für die Berechnung richtiger Wahrscheinlichkeiten.

Die Vorgehensweise zur Vermeidung der Kopplung bei der Entwicklung der 52-SNP Multiplexmethode ist im Kapitel 5.2 aufgeführt. Die statistische Überprüfung des LD-Koeffizienten erfolgte wie im Kapitel 3.2.4.2 beschrieben.

Ein weiteres Kriterium bei der Auswahl der Loci für die Multiplexanalyse war die Selektion der Primer für die PCR-Reaktion. Hierbei waren die molekularen Interaktionen der Primer als besonders kritisch zu beachten. Die Bildung von Dimeren und somit eine potentielle Entstehung von Nebenprodukten durch unspezifische Verlängerung der Primer führt zur Abnahme der Reaktionssensitivität. Die Balance der Signalintensitäten wird beeinflusst und letztendlich kommt es zu einer erschwerten Typisierung durch unspezifische Signale in der Detektionsreaktion. Die Qualität der Primer muss für eine dauerhaft zuverlässige Funktion

³¹ www.ncbi.nlm.nih.gov/About/primer/snps.html

gewährleistet sein, so wie die Qualität der den Marker flankierender Sequenzen, eine zuverlässige Analyse mit weiteren Methoden zulassen sollte. Zusätzliche Herausforderungen, wie die Anwendung der Multiplexmethode bei der Untersuchung von degradierten Proben und Mischspuren, stellten an die Präzision des Oligodesigns zusätzliche Anforderungen.

Die erfolgreiche Analyse der 23 SNP Multiplexreaktion deutete bereits an, dass - unter Beachtung der auferlegten Kriterien - die weitere Entwicklung einer Multiplexreaktion mit über 50 Markern möglich sein würde. Dieses beinhaltete auch eine möglichst geringe Länge der Amplifikate. Diese konnte für die 23 SNP Multiplexreaktion mit den Längen zwischen 59 und 105 *bp* erfolgreich etabliert und experimentell bestätigt werden.

5.1.3 Diskussion zur Auswertung populationsgenetischer Daten

Der Vergleich der errechneten Allelfrequenzen der ancestralen Allele mit den in der Datenbank (Populationsdaten des *HapMap* Projekts in der dbSNP-Datenbank)³² gefundenen Häufigkeiten für die asiatischen Populationen wies eine maximale Abweichung von 0,12 für den Marker A9 (rs1015250) auf. Eine Abweichung um 0,188 innerhalb des gleichen Markers wurde beim Vergleich der Allelfrequenzen zwischen europäischen Proben aus der Datenbank und den deutschen Proben des Mainzer Labors gefunden. Eine etwa doppelt so hohe Abweichung der Häufigkeiten wurde für den Marker A13 (rs1886510) berechnet. Hier betrug die Differenz zwischen den errechneten Allelfrequenzen und den Frequenzen aus der Datenbank 0,356 Datenpunkte (Diagramm 1).

Die Abweichungen der Allelhäufigkeiten können auf Unterschiede zwischen den für die Datenbank untersuchten Populationen und den für die Studie getesteten Proben zurückgeführt werden. Wahrscheinlicher jedoch ist die Begründung, dass es sich um eine kleine Anzahl untersuchter Individuen in der SNP-Datenbank handelt und die Vergleichspopulationen nicht immer eine repräsentative Gruppe darstellten. Dennoch kann festgehalten werden, dass zwischen den Extremwerten beider Populationsgruppen keine Systematik zu erkennen war, d. h. die Sequenzen konnten richtig genotypisiert werden und gaben keine Anhaltspunkte auf eine Amplifikationsproblematik. Die aufgrund von fehlenden Datenbankfrequenzen für die Marker A15, A21 und A23 hinzugezogenen Frequenzen aus dem *SNPforID browser* unterstützen das Gesamtergebnis. Die Tendenz der Übereinstimmung deckt sich jedoch eher mit den aus der *HapMap* Datenbank errechneten Frequenzen, was auf die drei untersuchten asiatischen Stichproben zurückzuführen ist. Die Homogenität der Ergebnisse für die chinesischen Stichproben bestätigt den hierzu gewählten Stichprobenumfang und die Auswahl der Proben aus einer Bevölkerungsgruppe (Han Chinesen) für alle drei Datenquellen.

³² <http://www.ncbi.nlm.nih.gov/projects/SNP/>

Der Korrelationskoeffizient nach Pearson kann die errechneten Allelfrequenzen zusätzlich im Verhältnis zu den Datenbankfrequenzen darstellen. Der Koeffizient (r) ist sehr hoch für die chinesische Stichproben (0,965). Er weicht auch dann kaum ab, wenn man die Daten aller untersuchten Populationen poolt und diese mit den Datenbankwerten für die chinesische Population vergleicht (0,963; siehe Anhang Tabelle A-5 c). Dies veranschaulicht die gute Übereinstimmung der untersuchten Individuen mit den Vergleichsgruppen. Der Wert für die Gruppe der deutschen Proben im Vergleich zu den europäischen Proben aus der Datenbank weist eine deutlich niedrigere Korrelation auf. Möglicherweise deutet diese auf eine höhere Diversität innerhalb der 120 typisierten europäischen Proben aus der Datenbank hin. Die untersuchten Populationen waren dort nicht ausführlicher beschrieben und stellten eine Stichprobe aus der zentraleuropäischen Bevölkerung dar.

Der Einfluss der Unterteilung einer Population in Subpopulationen auf die resultierenden Genotypfrequenzen wird in der forensischen Genetik seit langem kritisch diskutiert (Lewontin und Hartl, 1991; Balding und Nichols, 1994; Budowle, 1995). Postulierte Unterschiede in den Frequenzen zwischen den Untergruppen einer Gesamtpopulationen (*major population*), die durch lokale, soziale oder linguale Grenzen entstehen, reduzieren die Unabhängigkeit der Allele und entstehen aus verminderten Heterozygotenpaarungen. Die kontroverse Fragestellung ergab sich für die Anwendung der publizierten Allelfrequenzen der Hauptpopulationen bei der Berechnung der *matching probability* in der Untersuchung von Spurentreffern oder Abstammungsuntersuchungen. Die Bedenken konnten in der forensischen Routine für STR-Daten weitestgehend ausgeräumt werden (Deka *et al.* 1995, Morton, 1995). Dennoch ist das Wissen um den Umfang und die Herkunft der Referenzpopulation von hoher Bedeutung. Die Korrelationen für die typisierten 146 Proben aus China, Japan und Thailand sprechen für die Verwendung der Datenbankdaten, die in diesem Fall 90 Individuen der chinesischen Han-Population beinhalten. Größere genetische Diversität zeigte sich innerhalb der europäischen Referenzstichprobe. Diese umfasste Individuen europäischer Herkunft aus Utah in den USA (Dausset *et al.* 1990) und konnte erst durch eine Studie von Lao *et al.* (2008) in die genetische Struktur innerhalb der europäischen Population eingeordnet werden. Die Daten sind weitestgehend irischen, niederländischen, dänischen und norwegischen Individuen zuzuordnen. Einzelne Daten stammen aus Norddeutschland und eine einzelne Probe ist einer süddeutschen Region zugeordnet.

Betrachtet man die Heterozygotenraten, liegen diese für die deutsche Stichprobe für alle Marker bei 0,375 (Median) und weichen um 0,045 Datenpunkte von den erwarteten Heterozygotenwerten ab. Der beobachtete Wert entsprach den Anforderungskriterien einer hohen Heterozygotenrate der Marker innerhalb der europäischen Bevölkerung und bestätigte die korrekte Selektion der SNPs sowie der Funktionalität der Primer. Abweichende beobachtete Heterozygotenwerte können Hinweise auf mögliche Amplifikationsprobleme,

besonders in der Untersuchung vom degradierten genetischen Material liefern. Bei kleinen Stichprobengrößen können diese Probleme in der Gesamtdarstellung stark ins Gewicht fallen. Erst ein größerer Probenumfang nivelliert die Werte und lässt einen Trend erkennen. Zusammenfassend konnten für jede untersuchte Stichprobe Heterozygotenraten von 0,367/Thailänder bis 0,425/Japaner festgestellt werden. Diese lieferten den entscheidenden Hinweis auf vorhandene genetische Variabilität.

Die Betrachtung der Genotypen nach Bonferroni-Korrektur der errechneten Genotypfrequenzen ergab keine signifikanten Abweichungen vom Hardy-Weinberg-Gleichgewicht. Vor einer Korrektur des α -Fehlers, d.h. der Multiplikation der p-Werte mit der Anzahl der Vergleiche, konnte man jedoch signifikante Abweichungen in drei der vier getesteten Populationsgruppen beobachten. Lediglich die japanische Population war im Gleichgewicht, was auf das Vorliegen einer homogenen Stichprobe hindeutet.

Die mögliche Kopplung wurde für die Marker A10 (rs735155) und A23 (rs826472) in Betracht gezogen, da beide auf Chromosom 10 liegen. Sie wurde jedoch für die beiden SNPs trotz ihrer Position auf dem p-Arm des Chromosoms nicht bestätigt.

Mit Hilfe der F-Statistiken wurde der Grad der genetischen Strukturierung der untersuchten Populationsstichproben definiert. Die Verteilung der F_{ST} -Werte nach Wright dient der Analyse der genetischen Differenzierung und gibt den relativen Anteil der Unterschiede zwischen den Subpopulationen bezogen auf die Gesamtpopulation an (Swofford und Selander, 1989). Sie wird durch den Anteil der Heterozygotie der Subpopulationen an der Heterozygotie der Gesamtpopulation bestimmt. In der vorliegenden Arbeit beziehen sich die F_{ST} -Werte auf die japanischen, thailändischen und chinesischen Populationen. Ihre Werte liegen in einem sehr niedrigen Bereich mit einem Maximum von 0,07 und einem Mittelwert von 0,008. Durch das Maß der Schiefe wird hier ebenfalls verdeutlicht, wie sich die Verteilung der F_{ST} -Werte hinsichtlich des Mittelwerts verhält. Die Verteilung ist bei den beschriebenen Daten stark rechtsschief (1,519), was das häufige Auftreten von Werten, die kleiner sind als der Mittelwert, aufzeigt (Tabelle 15, Tabelle A-7). Werte, die sehr nah an Null liegen, deuten auf einen hohen genetischen Fluss und auf eine geringe Differenzierung zwischen den Subpopulationen hin. Hierdurch wird auch deutlich, dass es keinen Hinweis auf eine funktionale Relevanz der SNPs gibt. Da für die europäische Population nur die untersuchten Proben deutscher Individuen vorlagen, wurden hier die F_{IS} -Werte berechnet. Bei dieser Beschreibung werden die einzelnen Individuen mit der dazugehörigen Subpopulation in Beziehung gesetzt. Da F_{IS} -Werte auch ein Maß der Abweichung vom Hardy-Weinberg-Gleichgewicht darstellen, überrascht es nicht, dass diese für die Marker A7, A11 und A15 (vgl. Tabelle 14) die höchste Abweichung von 0 aufzeigen (0,313, 0,388 und 0,340). Das gleiche Bild erhält man für die betroffenen Marker bei der Betrachtung der F_{IS} -Werte für die einzelnen asiatischen Subpopulationen (Anhang, Tabelle A-7).

Bleibt man bei der Annahme, dass größere Frequenzunterschiede zwischen den Subpopulationen zu Abweichungen in der Berechnung der Trefferwahrscheinlichkeit in der forensischen Biostatistik führen können, kann man den sog. Schätzwert für F_{ST} ($\theta - \text{Theta}$) als Korrekturfaktor für die o.g. Berechnung verwenden (Balding und Nichols, 1994). Dies wurde von der *Commission on DNA Forensic Science* (1996) vorgeschlagen und wird in einigen Ländern mit stark strukturierten Subpopulationen, wie z.B. den USA, berücksichtigt (Schneider, 2008).

Die genaue Bewertung der Varianzverhältnisse erfolgt über den Signifikanztest Φ . Wie schon die unter der Annahme des Hardy-Weinberg-Gleichgewichts berechneten F-Werte, stellen die Φ -Werte die Differenzierungsindizes auf der Basis der molekularen Varianz dar. Für die Berechnung wird im Vorfeld, unter Definition der einzelnen Populationen und deren Einteilung in Gruppen, die genetische Struktur festgelegt. Die aus den Varianzkomponenten errechneten Werte werden darauf hin gegen die Nullhypothese auf ihre Signifikanz getestet. Die Nullhypothese setzt dabei voraus, dass im Datensatz keine hierarchische Struktur vorhanden ist. Im vorliegenden Fall waren die genetischen Unterschiede innerhalb der definierten Gruppen der Asiaten und der deutschen Gruppe nicht signifikant. Deren Varianzanteil betrug 0,05% (Tabelle 16). Die höchste Variabilitätsrate wurde innerhalb der Populationen festgestellt (99,28%). Diese war hoch signifikant ($p=0,000$) auf dem Signifikanzlevel von 5%. Die Ergebnisse unterstützen die Auswahl der SNP-Marker unter dem Aspekt einer populationsübergreifenden Variabilität. Eine entgegengesetzte Varianzverteilung wäre bei einer Untersuchung von SNPs zu erwarten, die eine Extremverteilung der Häufigkeiten innerhalb der drei Hauptpopulationsgruppen aufweisen (AIMs – *ancestry-informative markers*, Rosenberg *et al.* 2003, Phillips *et al.* 2007b).

5.1.4 Fazit

Anhand der selektierten 23 Marker konnten die beiden Populationen eindeutig voneinander unterschieden werden.

Mit dem Ziel, eine ausreichend große Zahl von Markern analysieren zu können, wurde die Auswahl in Teilschritten durchgeführt. Die vorangegangene Diskussion bezog sich dabei auf die erste Gruppe von 23 SNPs. Die Ergebnisse zeigen eine positive Tendenz in Bezug auf die Möglichkeit, eine größere Markerzahl effizient und zuverlässig in einer Amplifikationsreaktion zu analysieren und die Methode auch statistisch abzusichern. Die nachfolgende Validierung der Multiplexreaktion mit insgesamt 52 SNPs wurde im Anschluss an die hier beschriebene Untersuchung durchgeführt. Diese wurde auf alle teilnehmenden Institute aufgeteilt und wird hier zusammengefasst ergänzend dargestellt. Die aus der Studie hervorgegangene Veröffentlichung befindet sich im Anhang dieser Arbeit (Sanchez *et al.* 2006).

5.2 Validierung der 52 SNP Multiplexreaktion – Ergebnisse und Diskussion der gemeinsamen Projektarbeit

Das primäre Ziel der Entwicklung einer Multiplexmethode mit einem Minimum von 50 Amplikons ist mit der Validierung einer 52 SNP Multiplexreaktion erreicht worden. Die technischen sowie statistischen Anforderungen an diese Untersuchungsmethode wurden vom *SNPforID Consortium* umfassend analysiert (Sanchez *et al.* 2006). Dabei standen die Qualität der gewählten Marker, die Bedienbarkeit der Anwendung sowie die biostatistische Effizienz im Vordergrund. Im Zuge der Entwicklung führten zwei Zusatznotationen in der dbSNP-Datenbank zur Entfernung von zwei Markern aus der ursprünglichen Selektion von 54 SNPs. Im ersten Fall handelte es sich um eine Duplikation des Locus rs2145294. Im zweiten Fall um eine korrigierte Sequenz, welche sich im Anschluss in unmittelbarer Position zum anderen ausgewählten Locus auf Chromosom 1 befand.

Die Robustheit konnte anhand einer Reihe zusätzlicher Methoden und Testsysteme in weiterführenden Studien genutzt und bestätigt werden (Balogh *et al.* 2006, Bender *et al.* 2006, Harrison *et al.* 2006, Sanchez *et al.* 2008, Børsting *et al.* 2008).

5.2.1 Technische Voraussetzungen und Typisierungserfolg innerhalb der Institute

Nach den Kriterien der SNP-Auswahl wurden im zweiten Teil der Multiplex-Validierung weitere 29 Marker selektiert und zusammen mit den bereits erfolgreich getesteten 23 SNPs untersucht. Bei der Auswahl der Detektionsplattform wurden am Mainzer Institut für Rechtsmedizin die Reaktionsdurchläufe zwischen dem ABI PRISM® 310 und dem ABI PRISM® 3100 *Genetic Analyzer* abgeglichen. Der Vorteil des ABI PRISM® 310 lag in der Möglichkeit einer direkt aufeinander folgenden Injektion beider Multiplexreaktionen im gleichen Lauf (vgl. Kapitel 3.3.3.2) sowie der Auswertung der Daten beider Multiplexe mit der *Genotyper 3.7 Software* in einer einzigen Analyse. Die Sensitivität der älteren Detektionsplattform lag jedoch um etwa 80% niedriger verglichen mit dem ABI 3100. Als einziges Partnerlaboratorium konnte das Kopenhagener Institut die Anwendung der „doppelten Injektion“ erfolgreich auf dem Applied Biosystems 3130xl *Genetic Analyser* Gerät nach der Norm ISO 17025 akkreditieren (Børsting *et al.* 2009).

In einer umfangreichen, gemeinschaftlichen Studie unter Teilnahme von insgesamt elf europäischen und US-amerikanischen Laboratorien wurde die Anwendbarkeit der Methode für forensische Zwecke unter vereinfachten Bedingungen überprüft (Sanchez *et al.* 2008). Nach der PCR-Amplifikation mit dem Primerset für alle 52 Marker sollten in der SBE-Reaktion 29 SNPs untersucht werden. Das Probenmaterial (zehn Blutproben und eine Negativkontrolle) wurde einheitlich hergestellt und an alle Laboratorien verschickt. Dabei

stellte eine degradierte Probe zusätzliche Herausforderung dar. Die zu 98% übereinstimmenden Ergebnisse der EDNAP (*The European DNA Profiling Group*) Studie haben aufgezeigt, dass die Anwendung, basierend auf der Typisierung von SNPs mittels *SNaPshot*TM, durchaus eine Ergänzung der forensischen DNA-Analyse darstellen kann. Zudem bot die Ausweitung der Untersuchungen die Möglichkeit, über die Schwierigkeiten der automatisierten Anwendung laborübergreifend zu diskutieren. Diese resultierten aus der technischen Ausstattung sowie aus der unterschiedlichen Erfahrung der Institute in der SNP-Analyse. Schwankungen in den relativen Signalstärken waren von der Sensitivität des Analysegeräts sowie der Trennleistung des eingesetzten Polymers abhängig (Abbildung 47).

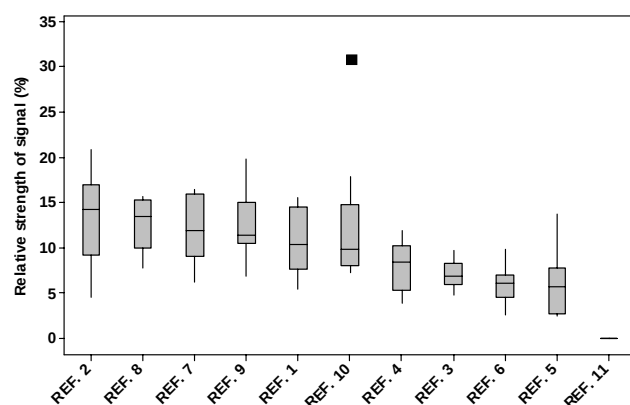


Abbildung 47: Boxplot-Darstellung der Signalstärken aller teilnehmenden Labore aus der EDNAP-29plex Studie. Bestimmt wurden die relativen Signalstärken für jede Probe (REF. 1-11) durch den Quotienten der Summe der Fluoreszenzeinheiten (RFUs) für alle Allele in der Probe und der Summe der RFUs aller beobachteten Allele der Proben. Die Signalstärken schwankten um maximal ca. 7% für die Probe 2 innerhalb der beiden Quartile und um ca. 12% im gesamten Wertebereich des Datensatzes für die gleiche Probe. Einen Ausreißer gab es in Probe 10 (schwarzes Quadrat). Probe 11 enthielt eine Negativkontrolle und ergab keine Signale (aus Sanchez *et al.* 2008).

Die laborübergreifenden Versuche bestätigten auch die Tendenz der Adenin-Nukleotide, im Detektionsbereich des grünen Fluoreszenzfarbstoffs, starke Hintergrundsignale im Elektropherogramm zu erzeugen. Eine matrizenunabhängige Bindung von basen-gekoppelten Fluoreszenzfarbstoffen, speziell von ddATP, an 3'-Enden der kürzeren Sonden führte zum Auftreten der unspezifischen Signale. Diese verhielten sich immer konstant und konnten so durch Einengung der Detektionsbereiche der *GenoTyper 3.7 Software* aus dem Allelmarkierungsbereich herausgefiltert werden. Ebenso verhielt sich die Auswertung im Mainzer Labor, mittels des Softwarepakets *GeneMapper ID* (Kapitel 3.3.3.2). Hier war es durch die Definition der sog. *bins* möglich, die Bereiche der Zusatzsignale zu umgehen.

Ammoniumionen als Pufferbestandteil haben eine stabilisierende Wirkung und werden bei der Analyse großer Multiplexansätze gerne eingesetzt (Bríón *et al.* 2005). Nicht zu vernachlässigen sind jedoch die einfachsten Vorkehrungen, die die Qualität der SBE-Primer sicherstellen bzw. erhalten können. Dazu gehören die Aufreinigung der langen Moleküle nach der Synthese, um unvollständige Syntheseprodukte abzutrennen sowie eine

schonende Aufbewahrung, die das vorzeitige Degradieren der Primer und damit ein inhomogenes Verhalten in der Reaktion verhindert.

5.2.2 Analyse der 52-SNP Multiplexreaktion

Nach der Testphase, in welcher die zehn auf FTA-Papier gespotteten Blutproben in allen teilnehmenden Laboratorien untersucht wurden, begann die umfangreiche Typisierungsphase der Populationsproben.

Die bereits anhand von Verdünnungsreihen durchgeführte Analyse der Funktionsfähigkeit der Multiplexreaktion bei unterschiedlichen DNA-Mengen ergab ein vollständiges Profil der Reaktion bereits ab 500 pg DNA.

Um die Sensitivität der 52-SNP Multiplexreaktion zu testen und sie mit der von STRs zu vergleichen, wurde eine Untersuchung mit degradierter DNA durchgeführt. Das Teilexperiment ergab vollständige Profile mit dem Einsatz der DNA-Menge zwischen 200 und 500 pg sowie bei der Verwendung von degradiertem Material. Dem gegenüber wurden keine vollständigen Profile bei der STR-Typisierung der gleichen DNA-Proben mit dem *AmpF/STR SGM Plus Kit* erzielt. Hier wurden Signale ausschließlich für kurze Amplifikate detektiert (Siehe Anhang, Abb. 5, Sanchez *et al.* 2006).

Eine umfangreiche Validierung der 52-SNP Multiplexmethode für forensische Anwendungen erfolgte durch Musgrave-Brown *et al.* (2007). Sie beinhaltete die Untersuchungen von Spurenmaterialien, stark degradierten Proben sowie Mischspuren. Die Vergleichsmethode war das *Powerplex[®] 16* STR-Analysekit (Promega). Die vergleichende Beurteilung verfolgte das Ziel, die Grenzbereiche beider Methoden im Hinblick auf schwieriges Untersuchungsmaterial herauszuarbeiten. Die Ergebnisse zeigten, dass die Methode der SNP-Typisierung in Hinblick auf die vollständige Identifikation der Allele bei degradiertem DNA-Material, welches in geringer Menge vorhanden ist, zuverlässig funktioniert. Erfolgreiche Anwendungen der Methode in der Fallanalyse unterstützen das Ziel, mit Hilfe der SNPs forensische Fragestellungen aufklären zu können (Fondevila *et al.* 2008). Unerwartete Unterschiede in der Sensitivität der Methode traten jedoch im Vergleich zu der bei Sanchez *et al.* (2006) durchgeführter Studie auf. Die 52-SNP Multiplexmethode erzielte bei einer Konzentration von weniger als 1 ng DNA kein vollständiges Profil. Die Untersuchung von Mischspuren, die in der Routine der forensischen Analyse eine besondere Herausforderung darstellen, konnte eine weitere Hürde aufzeigen. Aus den Signalstärken konnten keine Hinweise auf die Mischungsverhältnisse der Probe abgeleitet werden. Eine genauere Diskussion dieser Ergebnisse findet sich in der Dissertation von Balogh (2009).

Im Rahmen eines Workshops im Kopenhagener Labor zur bereits beschriebenen Methode der „doppelten Injektion“ (Kapitel 3.3.3.2) wurden 46 Primaten-Proben zur Bestimmung des

anzestralen Wildtyp-Allels eingesetzt. 46 Loci aus der 52-SNP Multiplex konnten aus der Primaten-DNA typisiert werden. Die SNP-Anzahl pro Primatenart ist in Tabelle 21 zusammengefasst.

Tabelle 21: Unterarten der Primaten und Anzahl der typisierten SNPs.

Primatengruppe	Anzahl typisierter SNPs (von 52)
Schimpanse	44
Orang-Utan	37
Gorilla	36
Pavian	23
Rhesus	22
Cynomolgus	21

Fast alle Primaten-Stichproben wiesen einen homozygoten Status auf. Ausnahmen bildeten zwei der Schimpansen, die für den Locus rs2056277 heterozygot waren. Mit der Ausnahme von fünf Loci wurden in allen untersuchten Primatenarten gleiche Allele nachgewiesen, die in der dbSNP Datenbank als ancestrale Allele definiert sind. In den detektierten Ausnahmefällen handelte es sich um das jeweils dem bekannten humanen Polymorphismus zugeordnete zweite Allel (rs717302 (A/G), rs873196 (C/T), rs2056277 (C/T), rs1413212 (A/G), rs2040411 (A/G), siehe Tabelle S3 im Anhang).

Es stellt sich die Frage, ob es sich bei den hier detektierten Polymorphismen um Loci handelt, die eine höhere Mutationsrate aufweisen als die durchschnittlich errechnete Mutationsrate von $2,5 \times 10^{-8}$ (Nachman und Crowell, 2000). Es kann sich hierbei auch um Polymorphismen handeln, die bereits vor der Auseinanderentwicklung der phylogenetischen Linien von Mensch und Primat entstanden sind (*"identical by descent"*). Aufgrund der Unabhängigkeit der Loci und ihrer Position in nichtcodierenden, autosomalen Chromosomenabschnitten, kann keine Begründung für eine höhere Mutationsrate angeführt werden. Bekannte Abweichungen beruhen auf strukturellen Sequenzunterschieden (z.B.: CpG-reiche Regionen) oder deuten auf eine Funktionalität hin (Chen und Li, 2001). Hacia *et al.* (1999) definieren aus ihrer Studie zur Bestimmung des ursprünglichen Allels die Folgerung, dass die meisten gemeinsamen Polymorphismen unabhängig voneinander entstanden sind und lediglich einige davon aufgrund niedrigerer Mutationsraten Primaten und Menschen gemeinsam sind (*"identical by state"*).

Das Vorkommen des ancestralen Allels mit einer leicht höheren Frequenz in der somalischen Population (0,53) im Vergleich zu den beiden übrigen Populationsgruppen der Europäer und Asiaten (beide 0,48, siehe Tabelle im Anhang) kann aufgrund der geringen Stichprobengröße nur in Ansätzen diskutiert werden. Weiterführende Untersuchungen an molekularen Markern, die das Vorkommen des Ursprungsallels mit einer höheren Frequenz

(zusammengefasst bei Takahata *et al.* 2001, Excoffier, 2002) sowie eine höhere genetische Diversität in afrikanischen Populationen beschreiben (Yu *et al.* 2002), können das vorliegende Ergebnis ergänzen. Die Ergebnisse bleiben jedoch nicht unumstritten. Gerade hinsichtlich der spannenden Frage nach dem regionalen Ursprung des *Homo sapiens* und mit Hilfe hochdurchsatzfähiger Untersuchungsmethoden konnten Wissenschaftler um Rogers *et al.* (2007) nach der Auswertung von codierenden und nichtcodierenden Markersets sowie von *Alu*-Elementen aufzeigen, dass die oben beschriebene Tendenz des Ursprungsallels nicht für alle Marker zutrifft.

5.2.2.1 Allelverteilung in neun Populationen

Die biostatistische Aussagekraft aller forensischer Marker wird durch den sog. Diskriminationsindex [DI] bestimmt. Dieser beschreibt die Wahrscheinlichkeit der Unterscheidung von zwei beliebigen, nicht verwandten Individuen und wird durch die Anzahl der Allele und deren Frequenz bestimmt. Allelfrequenzen des selteneren Allels (*minor allele*) geben hier die Leistungsfähigkeit der Differenzierung an. Sie beträgt bei einer Allelfrequenz von 0,5 = 0,62. Sinkt die Häufigkeit des seltenen Allels, sinkt auch der Diskriminationsindex. Betrachtet man die Kurve in Diagramm 11, so zeigt sich ab der Frequenz 0,3 mit dem Indexwert 0,58 ein deutlicher Abfall. Die Wahrscheinlichkeit einer zufälligen Übereinstimmung von zwei nicht verwandten Individuen steigt. Auf der Basis des DI wurden die Marker innerhalb der Variabilitätsskala von 0,62 und 0,58 Indexpunkten in den drei Hauptpopulationen ausgewählt (Phillips *et al.* 2004b). Diese Indexspanne sollte in mindestens zwei der drei Vergleichspopulationsgruppen gewährleistet sein.

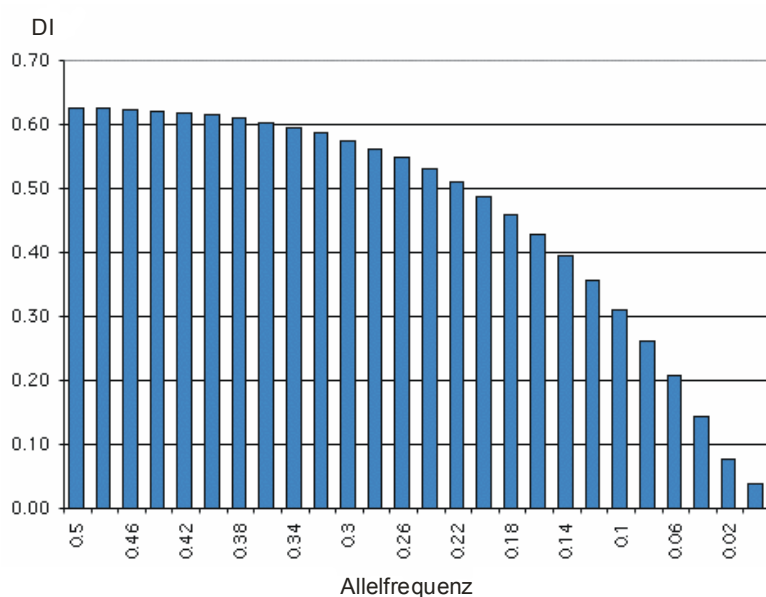


Diagramm 11: Werte der Diskriminationsindizes [DI] für unterschiedliche „minor allele frequencies“ (Abbildung zu Phillips *et al.* 2004b).

Abweichungen im Hardy-Weinberg-Gleichgewicht wurden für einen SNP festgestellt. Die Genotypverteilung des Markers A38 (rs907100) wich in sechs von neun Populationen von den erwarteten Heterozygotenwerten ab. Bei einem Signifikanzniveau von <5% lag der χ^2 -Wert bei 12,04. Nach der Bonferroni-Korrektur waren die Abweichungen in drei Populationen weiterhin signifikant. Zum Zeitpunkt des Studienabschlusses wurde ein neuer SNP (rs11689319) veröffentlicht³³. Dieser A/G-Polymorphismus lag neun Nukleotide *upstream* vom 3'-Ende des *reverse* Primers für den SNP A38. In einer umfassenden Studie an 124 Mutter-Vater-Kind-Trios (Børsting *et al.* 2008) wurde eine Kopplung des A-Allels mit dem G-Allel des Markers A38 bestätigt. Dadurch lag eine geringere Amplifikationsfähigkeit des gewählten Abschnitts vor, die mit der Beeinträchtigung der Primerbindung durch den neu detektierten Polymorphismus erklärt werden konnte. Die benachbarte polymorphe Stelle bedingte die hohe Homozygotenrate in der Validierung der 52-SNP Multiplexreaktion und gab den Hinweis, dass weitere umfassende Studien notwendig sind, um mögliche „stille“ Allele (*silent alleles*) zu charakterisieren.

Die Berechnungen der Fixationsindizes und der molekularen Varianz bestätigten die Auswahl der SNPs. Die Populationen waren gut voneinander getrennt und die F_{ST} -Werte deuteten auf eine geringe Differenzierung zwischen den Subpopulationen hin. Die genetische Variabilität innerhalb der Populationsgruppen lag bei 93%. Möglicherweise führte die Integration der grönländischen Proben in die asiatische Populationsgruppe zu einem leicht niedrigeren Wert gegenüber der vorangegangenen Analyse der 23 Marker.

³³ http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?searchType=adhoc_search&type=rs&rs=rs11689319

5.2.2.2 Clusteranalyse

Zusätzlich zu der Allelverteilung wurde eine Klassifizierung der Populationen mittels der *Structure*-Software vom J. Pritchard durchgeführt (Pritchard *et al.* 2000). Sie führt mittels eines Algorithmus eine geografisch-genetische Strukturierung von Individuen zu einer oder mehreren Populationen durch, deren Anzahl frei gewählt werden kann. Die Software stützt ihre Analyse auf einer modellbasierten Statistik der Clusterzuteilung, in der angenommen wird, dass die Daten aus einer Anzahl K unabhängiger Populationen entstammen, deren Gruppenzuordnung nicht bekannt ist.

Die Unterteilung erfolgte zwischen $K=2$ und $K=7$, wobei bereits bei $K=4$ die Abtrennung der Proben aus Grönland von den ostasiatischen Proben erfolgte (Abb. 3 in Sanchez *et al.* 2006). Eine Steigerung der K -Werte auf $K>4$ bewirkte keine weitere Zugehörigkeitsunterteilung in Subpopulationen. Es folgte jedoch eine genauere Charakterisierung innerhalb der Cluster aufgrund des angenommenen „*admixture*“ Modells. Hierdurch wurden die Verhältnisgrößen der genetischen Anteile der Vorfahren für jedes Individuum dargestellt. Sichtbar wurden dadurch z.B. Einwirkungen von europäischen Allelen innerhalb der sehr homogenen somalischen Population (Abbildung 48). Das vorliegende Phänomen kann auf die geringe Anzahl der Marker zurückgeführt werden. Speziell für die hier untersuchte somalische Population kann die Begründung auch darin liegen, dass die somalischen Proben von in Dänemark ansässigen Individuen stammten (Phillips *et al.* 2007a).

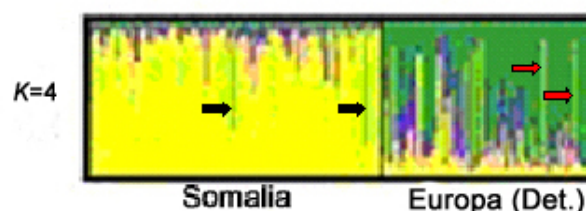


Abbildung 48: Vergrößerter Ausschnitt der Clusterdarstellung mit den Einschlüssen innerhalb der Cluster. Die schwarzen Pfeile deuten auf Individuen europäischer Herkunft innerhalb der somalischen Populationsgruppe hin, roten Pfeile auf den umgekehrten Fall innerhalb des europäischen Clusters (Ausschnitt aus Abb. 3, Sanchez *et al.* 2006 im Anhang).

Vor einigen Jahren veröffentlichte die Forschergruppe um Rosenberg eine Studie, in der 1046 Individuen aus 52 Populationen anhand 377 Mikrosatellitenmarker untersucht wurden (Rosenberg *et al.* 2002). Die Clusterstruktur der *Structure*-Analyse in der vorliegenden Arbeit wies eine vergleichbare Populationsgruppierung auf wie die der erheblich umfangreicheren Rosenberg-Studie. Bei den für das Projekt ausgesuchten Markern handelte es sich nicht um SNPs, die unter dem strikten Aspekt der Populationsspezifität ausgesucht worden sind. Sie zeigten trotzdem Übereinstimmung und bestätigten die korrekte Zuordnung der Wahlpopulationen und deren Herkunft sowie die Clusterseparation. Diese Schlussfolgerung

muss trotzdem kritisch betrachtet werden, denn wegen der niedrigen Proben- und Markeranzahl sowie der geringen Aussagekraft der gewählten Marker im Hinblick auf die Bestimmung der Ursprungspopulation, hätten diese Muster nicht intrinsisch entstehen müssen. Multiplexreaktionen bestehend aus einer relativ kleinen Anzahl von Markern können nur dann erfolgreich Ursprungspopulationen bestimmen, wenn die SNPs maximale Allelfrequenzunterschiede zwischen den Populationsgruppen aufweisen. Eine solche Multiplexreaktion mit populationsspezifischen Markern (AIMs, siehe auch Kapitel 5.1.3) wurde in der weiterführenden Studie des *SNPforID Consortium* entwickelt (Phillips *et al.* 2007b). Sie berücksichtigt die so genannten *fixed difference* Marker³⁴. Marker, deren Allelfrequenzen über die Populationsgruppen ungleich verteilt sind sowie SNPs mit drei möglichen Allelen (triallelische Marker). Genetische Marker mit starken populations-spezifischen Ausprägungen ihrer Frequenzen haben im evolutiven Verlauf eine positive Selektion erfahren. Es handelt sich dabei um funktionale Marker, die z.B. einen Hinweis auf Pigmentierung oder bestimmte populationsspezifische Resistenzen liefern können. Eine solche Auswahl der SNPs könnte bei forensischen Untersuchungen speziell in Fällen von Katastrophenopfern hilfreich sein, kann aber auch im Rahmen der polizeilichen Fahndungsanalyse unterstützende Ergebnisse liefern (Kayser und Schneider, 2009).

Eine weitere Vorgehensweise zur Identifizierung von Personen wurde von Kidd *et al.* (2006) vorgestellt. Diese bedient sich einer Gruppe von SNPs, die sich durch eine hohe Heterozygotie und einen niedrigen Inzuchtkoeffizienten (F_{ST} -Wert) auszeichnet. Letzteres deutet auf eine geringe Unterteilung in Subpopulationen hin und sichert einen maximalen Diskriminationsindex (siehe auch Kapitel 5.2.2.1). In mehreren Schritten wurden durch die Forscher zunächst 19 unabhängige SNP-Marker selektiert, die in 40 Vergleichspopulationen eine *matching probability* von ca. 10^{-7} erreichten. Nach weiteren Untersuchungen (Pakstis *et al.* 2007) konnte schließlich eine Anzahl von insgesamt 92 Markern mit den oben definierten Kriterien publiziert werden (Pakstis *et al.* 2010). Ein daraus zusammengestelltes Panel von 45 unabhängigen SNPs erreicht selbst in kleinen isolierten Populationen *matching probability*-Werte, die mit VNTRs vergleichbar sind (Chakraborty und Kidd, 1991).

5.2.2.3 Forensische Biostatistik

Zur Vervollständigung der Validierungsstudie wurden die biostatistischen Kriterien berechnet, die als Grundlage zur Beurteilung von Vaterschaften, zur Berechnung der Spurenlegerwahrscheinlichkeiten oder zur Ermittlung von Verwandtschaftsanalysen dienen. Sie basieren auf der Errechnung von Ausschluss- oder Einschluss-Chancen sowie der Ermittlung des *Likelihood*-Quotienten zum Hypothesenvergleich.

³⁴ Jeweils eines der beiden Allele ist spezifisch in einer der drei Hauptpopulationsgruppen vorhanden.

Die errechnete diskriminatorische Kraft des SNP-Sets erreicht ihre Grenzen in Fällen mit einer Mischspur. Um diese zu steigern und damit an die Beweiskraft der Mikrosatellitenanalyse zu kommen, muss die Anzahl der Polymorphismen erhöht werden. Hierfür sollten auch Marker mit einer geringen Frequenz des seltenen Allels in die Analyse einbezogen werden (Gill, 2001), da es hier in einzelnen Fällen zum Auftreten seltener Genotypen in der Mischung kommen kann.

Die Anwendbarkeit der SNP-Analyse in Verwandtschaftsfällen profitiert von der Allelfrequenzverteilung von annähernd 0,5. Die in der vorliegenden Validierung errechneten Vaterschaftsindizes erreichen ihre größte Aussagekraft in der europäischen Populationsgruppe (PI=549 000), da hier eine gleichmäßigere Verteilung der Allele vorlag. Dieses Ergebnis wurde auch in einer umfangreichen Studie mit der hier beschriebenen 52 SNP Multiplexreaktion an 124 dänischen und somalischen Familientrios und Vater-Kind-Duos bestätigt (Børsting *et al.* 2008). Die ermittelten hohen Vaterschaftsindizes für die dänischen und somalischen Probanden (durchschnittlich $1,32 \times 10^6$ / $9,77 \times 10^3$ und $4,04 \times 10^5$ / $4,31 \times 10^3$ jeweils für Trio/Duo) lagen dennoch deutlich unter den Vaterschaftsindizes von STRs oder VNTRs. Zudem müssen in Vaterschaftsfällen, in denen die Mutter nicht in die Untersuchung hinzugezogen werden kann, 70-80 SNPs untersucht werden, um eine Nichtübereinstimmung in 99,5% der Fälle zu detektieren (Ayres, 2005). Die Vorteile gegenüber der VNTRs liegen bei der SNP-Analyse in ihrer einfachen Handhabung und gegenüber den Mikrosatelliten in der erheblich geringeren Mutationsrate. Im Gegensatz zu insgesamt sechs Mutationen, die in echten Vaterschaften durch VNTRs und STRs detektiert wurden, konnte erwartungsgemäß keine Mutation der biallelischen Marker nachgewiesen werden.

5.3 Diskussion zur Microarray-Analyse der 33 SNP Multiplexreaktion

Durch die erfolgreiche Etablierung und Validierung der im Vorfeld beschriebenen Multiplexreaktionen zur Detektion von SNPs mittels der *SNaPshot*TM-Methode stand eine große Auswahl von Markern zur Verfügung, die auf weiteren Plattformen etabliert werden konnte (Bríón *et al.* 2005, Sanchez *et al.* 2006). Das Methodenspektrum erstreckte sich von herkömmlichen Microarrays bis zur Nanochip-Technologie (Dissertation Balogh, 2009), der Pyrosequenzierung-Methode (Bender *et al.* 2006) und der Detektion mittels MALDI-TOF Massenspektrometrie (Mengel-Jorgensen *et al.* 2004). Die vorliegende Arbeit beschreibt die Etablierung der SNP-Detektionsmethode basierend auf dem Prinzip der Sequenzelongation in Reaktion und der anschließenden Immobilisierung auf Glasobjektträgern.

5.3.1 Technische Voraussetzungen

5.3.1.1 Versuchskonzeption und Optimierung

Die variable Methode der Immobilisierung von Sequenzen auf diversen Substratträgern entstand aus der auf dem Prinzip der Hybridisierung basierten *Southern blot* Methode (Southern, 1975). Die Technik entwickelte sich im Laufe der Jahre zu einer effizienten hochdurchsatzfähigen Analyse mit einem hohen Parallelisierungsgrad, die vor allem in der Genexpressionsdetektion ihre Anwendung findet (Schena *et al.* 1995, Johnston, 1998). Die Möglichkeit des simultanen Einsatzes einer großen Sondenzahl bietet ebenfalls Einsatzmöglichkeiten für den Bereich der Mutationsforschung (Gerry *et al.* 1999, Hacia *et al.* 1999). Eine Methode mit dem Ausblick einer Anwendung im Spurenbereich sollte die Eigenschaften der hohen Spezifität sowie die eines herausragenden Detektionserfolges mit sich bringen. In der vorliegenden Arbeit sollten die Vor- und Nachteile der Technologie in Bezug auf eine mögliche Anwendung in der forensischen Genetik getestet und kritisch diskutiert werden. Der Schwerpunkt lag in der Multiplexfähigkeit und der Standardisierung der Technologie, deren Anwendung aufgrund der Problematik des nur in geringen Mengen verfügbaren Spurenmaterials enge Grenzen gesetzt sind.

Die technische Voraussetzung für die Anwendung der Microarray-Technologie ist die Verfügbarkeit eines Laserscanners, welcher mit mindestens drei Lasern ausgestattet ist. In der vorliegenden Arbeit wurde gezeigt, dass eine Verteilung der optischen Detektionsspektren mittels einer Kombination von Bandpassfiltern in einem Anregungsbereich zwischen 532 nm und 635 nm, wie in Zwei-Laser-Scannern üblich, nicht möglich ist. Der Einsatz von drei oder vier Lasern hingegen erhöht die Flexibilität und erlaubt eine Detektion aller potentiellen Polymorphismen in einer Reaktion (Hirschhorn *et al.* 2000),

unabhängig von der Amplifikationsrichtung der Minisequenzierungsprimer (Pastinen *et al.* 1997).

5.3.1.2 „Array of arrays“-Prinzip

Ein großer Schritt in Richtung Hochdurchsatzanalyse und Multiplexfähigkeit erfolgte mit der Etablierung des „array of array“ Prinzips in der Microarray-Technologie (Pastinen *et al.* 2000). Ein geringerer Materialverbrauch und eine höhere Effizienz der Entwicklungsphase der Experimente waren die Folge. Die Methodenerweiterung bietet Vorteile beim Vergleich von Reaktionsansätzen, die unter gleichen Konditionen hybridisiert werden und so miteinander verglichen werden können. Zu dieser Entwicklung gehörte ein nach Maß hergestellter Silikonpad als Abtrennung der Proben und Reduktion des Kammervolumens.

Die hier beschriebene Form der Probenotypisierung erforderte jedoch eine zusätzliche Entwicklung, um eine ausreichende Feuchtigkeit in der Hybridisierungskammer zu gewährleisten. Die mechanische Ursache eines möglichen Probenausfalls könnte nämlich das Austrocknen der Array-Flächen während des eigentlichen Hybridisierungsschrittes sein. Es ist deswegen besonders wichtig, auf ein geschlossenes System mit genügend Feuchtigkeit während der Reaktion zu achten.

5.3.1.3 Kontamination

Eine Kontaminationsgefahr durch das Ineinanderlaufen der Proben zwischen zwei Arrays unter dem Silikonpad wird durch den konstanten, mechanischen Druck des Hybridisierungskammerdeckels verhindert. Ein Kapillareffekt, der zu einem Übertrag über die obere Öffnung führen könnte, wurde nicht festgestellt. Die eingesetzte Flüssigkeitsmenge von 22 µl ist sehr gering und verteilt sich ausschließlich über die Trägersoberfläche. Durch vorsichtiges Pipettieren der Reaktion in die Öffnungen der Hybridisierungskammer konnte gewährleistet werden, dass keine übermäßige Schaumbildung entstand. Diese kann sich durch das Hinzufügen von SDS in den Hybridisierungspuffer entwickeln und über die Ränder der Kammer treten.

Multiplexfähige Hybridisierungskammern mit vorgefertigten Trennmasken für Microarrays werden zwischenzeitlich von einigen kommerziellen Herstellern angeboten (*ArrayIt*[®] Microarray Technology, TeleChem International, Inc.).

Kontamination spielte jedoch vermutlich eine Rolle beim Vorbereiten der *ctag*-Oligos vor dem Spotten. Das Ansetzen frischer Stammlösungen vor dem Bspotten neuer Glasobjektträger erwies sich als notwendig. Es konnte nicht gewährleistet werden, dass es beim Versiegeln und wieder Öffnen der Microtiterplatte zu keinem Übertrag der Moleküle in die Nachbarvertiefungen kam. Eine auf diese Weise erzielte Spotverunreinigung konnte nur

mittels Singleplexreaktionen nachgewiesen werden. Eine derartige *tag*-Verschleppung ist irreversibel.

5.3.2 Signaldetektion

5.3.2.1 Signalüberlappung

Eine andere Form der Signalüberlappung war die reproduzierbar in allen Versuchen auftretende Detektion markerfremder Signale. Der Scanvorgang und das Aufsplitten der Scanbilder in vier Einzelbilder der Farbkanäle zeigen, dass falsch-positive Signale in den allermeisten Spots vorhanden waren (Abbildung 26). Es konnte nicht quantitativ bewiesen werden, dass die falsch positiven Signale ausschließlich in den markerfremden Detektionsbereichen auftraten. Gleichwohl waren nur diese als solche für den Experimentator sichtbar. Die Spots, die zu Genotypisierungszwecken quantifiziert wurden, beinhalteten jedoch mit großer Wahrscheinlichkeit auch additive Signale, die ihren Ursprung nicht in einer primerspezifischen Elongationsreaktion hatten. Sie hatten zur Folge, dass die Diskriminierung der Genotypen nicht immer eindeutig war, bzw., dass sich die Verhältnisse der Signalintensitäten zueinander zu Ungunsten der Clusterdistanzen veränderten. Mehrere Erklärungen physikalischer und struktureller Art können für dieses Phänomen von Bedeutung sein und werden im Folgenden erläutert.

Vorab muss jedoch erwähnt werden, dass die Markerauswahl nach vorgegebenen Kriterien der 52-SNP Multiplexreaktion erfolgte. Diese Auswahl erlaubte das Generieren der Minisequenzierungsprimer nach den vorgegebenen thermodynamischen Kriterien aus zwei potentiellen Richtungen. Dieses ergab ≤ 104 Möglichkeiten der Primerkonstruktion. Beachtet man jedoch den Einfluss der thermodynamischen und sequenzabhängigen Faktoren, welche zusätzlich auf die Signaldetektion Auswirkung haben, so muss die Schlussfolgerung gezogen werden, dass ein begrenzter Primerpool, den der Experimentator vorab festlegt, nicht allen Methoden mit der gleichen Effizienz zur Verfügung stehen kann.

Folglich konnten nicht alle negativen Einflüsse durch ein „*redesign*“ umgangen oder eliminiert werden. So wurden auch Primer zugelassen, die mit einer freien Enthalpie, $\Delta G < 0$ auch an Fremdsequenzen annealen konnten (Kaderali *et al.* 2003).

5.3.2.2 Untersuchungen des Bindeverhaltens der *tag/ctag*-Sequenzen während der Hybridisierung

Vermutet wurde ein SNP-unabhängiger Einbau des fluoreszenzmarkierten Dideoxynukleotids, was zur Detektion falscher Genotypen führte (4.2.1.5). Dieses Phänomen trat auch in kleineren Reaktionen (z.B. Single- bis Quadruplexen) auf und widersprach der Hypothese einer Kreuzreaktion (Abbildung 27). Die Annahme, dass mit einer reduzierten

Anzahl von Matrizen der vorwiegende Anteil der falsch positiven Signale zurückgegangen sein würde, bestätigte sich nicht und die Problematik konnte nicht endgültig geklärt werden.

Nachdem ein unspezifisches Bindungsverhalten der *tag/ctag*-Sequenzen durch nicht-homologe *tags* ausgeschlossen werden konnte, verbleiben noch die Möglichkeiten der Dimerbildung der *tag*-Primer-Sequenzen sowie der Ausbildung von Haarnadelstrukturen der Primeroligos. Besonders im letzten Fall ist ein Zusammenfallen des Primers mit dem 3'-Bereich kritisch. In Reaktion kann es zu einer unspezifischen Basenverlängerung kommen, die nach einer korrekten Hybridisierung zu den beschriebenen, unspezifischen Signalen innerhalb der Spots führen kann.

Eine weitere Erklärungsmöglichkeit ist der unspezifische Einbau von markierten Dideoxynukleotiden nach spezifischer Hybridisierung des Primers. Dieser konnte durch die Optimierung der Zyklenzahl in der Minisequenzierungsreaktion adaptiert (Lovmar und Syvänen, 2005), jedoch nicht vollständig verhindert werden. Hierbei kann auch die Menge der *target*-Sequenz für den Falscheinbau der Fluoreszenzmarker ausschlaggebend sein (Pastinen *et al.* 1997). Dieser Fakt deutet stark auf eine zusätzliche Abhängigkeit der Reaktionseffektivität vom PCR-Produkt hin. Dieser Aspekt wird in der Diskussion der Sensitivitätsstudie nochmals aufgegriffen.

5.3.2.3 Limitierende Faktoren der Minisequenzierungsreaktion

Hauptaugenmerk der Optimierungsphase lag jedoch nicht auf der Hybridisierungs- sondern auf der Minisequenzierungsreaktion. Sie stellt die kritische Größe dar, welche für die Ausgewogenheit der Signalintensitäten, den daraus resultierenden Quotienten zur Determinierung der Genotypen sowie für die Sensitivität und Präzision der Gesamtanalyse zuständig ist.

Vergleichsstudien zwischen den Minisequenzierungsmethoden und den auf der Hybridisierung auf Glasobjektträgern basierten Microarray-Methoden mit allelspezifischen Sonden (ASO) zeigen, dass das Prinzip der hier untersuchten Methode in einer zehnfachen Steigerung der diskriminatorischen Stärke resultiert (Pastinen *et al.* 1997). Schon aus diesem Grund und durch den flexiblen Einsatz der universellen *tag*-Sequenzen, statt des Einsatzes allelspezifischer Oligos, wurde diese Methode für die Applikation bevorzugt. Sie eignet sich zur Genotypisierung, ist im höchsten Maß multiplexfähig und kann im Niedrigdurchsatz- wie auch im Hochdurchsatz-Verfahren eingesetzt werden (Lindroos *et al.* 2002).

Es zeigten sich jedoch eine Reihe weiterer limitierender Faktoren auf, die auf den Gesamterfolg der Reaktion Einfluss nahmen. Diese umfassten überwiegend die signalliefernden Bestandteile der SBE-Reaktion. Zusätzlich lagen die Schwierigkeiten auch in den Versuchsvorgaben. Diese werden im folgenden Abschnitt dargestellt.

5.3.2.3.1 *Signalstärke und Einbauverhalten*

Die Signalstärke der Spots und die Balance des Signaleinbaus verringerten sich mit der Erhöhung der Markerzahl in der Reaktion. Entgegenwirkend wurden die Marker zunächst in kleinen Reaktionsansätzen betrachtet. SNPs mit der größten Signalstärke wurden in die Reaktion reintegriert. Die nächste Herausforderung sollte die Ausbalancierung der Signalstärken sein. Eine Maßnahme zur Erhöhung der Signalausbeute der Multiplexreaktion stellte die Verringerung des Markerumfangs dar. Die endgültige Analysegröße umfasste 33 von ursprünglich 52 SNPs. Der Einfluss des Reaktionsumfangs fiel in diesem Ansatz stark ins Gewicht, denn die Anzahl der Spots auf dem Array war begrenzt und das Umdisponieren innerhalb der Positionen erforderte zusätzlich viel Material und Zeit. Der am stärksten limitierende Faktor jedoch war die bereits vorgegebene Anzahl der Marker. Der Multiplexansatz zur Detektion von 52 SNPs konnte nicht vollständig auf die Microarray-Plattform übertragen werden, könnte jedoch mit der Hinzunahme zusätzlicher Marker auf die statistisch erforderlicher Markerzahl ausgeweitet werden.

Ein weiterer Aspekt, der großen Einfluss auf die Effizienz der Reaktion nahm, war die unterschiedliche Einbaurate der Dideoxynukleotide in der Minisequenzierungsreaktion. Diese begründet sich sowohl in der uneinheitlichen Quantenausbeute und den molekularen Eigenschaften der Fluorophore, als auch in ihrer physikalischen Affinität zu den Nukleotiden, an die diese markierten Moleküle gekoppelt werden. Diese werden in der Reaktion den Polymeraseenzymen präsentiert, welche ihrerseits Differenzen in den Einbauraten bergen.

Die Beobachtung einer erhöhten Einbaurate der *TAMRA*-ddGTP-Nukleotide wurde bereits in der Literatur vermerkt (Jain *et al.* 2003). Im Experiment führte dies zu einem erhöhten Hintergrundrauschen für diesen Farbstoff sowie zu falsch positiven Ergebnissen. Letzteres wurde durch einen stark herabgesenkten Einbau von *Texas Red*-ddATP und *Fluorescein*-12-ddCTP verstärkt. Die differente Präferenz des Einbaus gekoppelter Nukleotide konnte bei Lindroos *et al.* (2002) nachgewiesen werden. Drei Parameter bezüglich der relativen Signalintensität konnten auf diese Weise verglichen werden: a) die Einbaurate des Farbstoffs durch das Sequenzierenzym, b) das Emissionsvermögen des Fluoreszenzfarbstoffs und c) das Emissionsverhalten nach der Kopplung an das Zielmolekül.

5.3.2.3.2 *Einbauverhalten und Fehleinbauten*

Durch die beobachtete Divergenz im Einbauverhalten der ddNTPs zielten die Optimierungsversuche zunächst auf das Sequenzierenzym. Statt der empfohlenen *ThermoSequenase*-Polymerase (Hirschhorn *et al.* 2000, Lindroos *et al.* 2002) wurde das Enzym *Therminator* DNA Polymerase verwendet. Entgegen der Erwartungen verbesserte sich das Ergebnis nicht. Die Präferenz im Einbau des Cy5-ddUTP, welches im Vorversuch stabile Einbauraten

zeigte, ging deutlich zurück, während *Texas Red*-ddATP wieder vollständig detektiert werden konnte. Die Verwendung beider Enzyme (*ThermoSequenase*-Polymerase und *Therminator* DNA Polymerase) in der SBE-Reaktion, zeigte hingegen den gewünschten Effekt des zeitgleichen und balancierten Einbaus der ddNTPs (Abbildung 28).

Aus der Literatur sind wenige Studien bekannt, in der Sequenzierenzyme hinsichtlich ihrer Einbauqualitäten getestet wurden (Gardner und Jack, 2002, Di Giusto und King, 2003, Lovmar und Syvänen, 2005). Die Resultate der Untersuchungen unterstützen die hier ermittelte Problematik in ihrer Hauptaussage, erlauben jedoch nur einen vorsichtigen Rückschluss auf die Notwendigkeit der Verwendung beider Enzyme. Vergleichend zu Lovmar und Kollegen, die lediglich die Qualität der Assays vergleichen, zielt die Studie von Gardner und Jack spezieller auf den Vergleich der Einbaueffizienz verschiedener Enzymfamilien hinsichtlich Dideoxy- und azyklischen terminalen Nukleotiden ab. Daraus geht hervor, dass das Zusammenwirken mindestens dreier Faktoren in additiver Weise die Inkorporation der terminalen Nukleotide beeinflussen kann: die Polymerase-Eigenschaft, Fluoreszenzfarbstoffe mit erhöhter Einbauleistung sowie die Eigenschaften der Kopplungsnukleotide. Zu diesen limitierenden Faktoren kommen kinetische Eigenschaften des Polymerase-Primer-*template*-Komplexes als Einflussgrößen in der SBE-Reaktion hinzu und bedingen den Fehleinbau, wobei hier eine Rückkopplung mit den verschiedenen Farbstofffamilien der Fluorophore sowie den Bindungsunterschieden dieser zwischen den Nukleotiden (Purine vs. Pyrimidine) stattfindet (Brandis, 1999).

5.3.3 Typisierungserfolg und vergleichende Analyse der Referenzmethoden

Der Genotypisierungserfolg lag innerhalb der Referenzmethoden mit 96,4-99,92 % deutlich höher als bei der hier untersuchten microarraybasierten Methode (85,52 und 86,75 %). Hintergründe dafür sind im vorigen Kapitel diskutierte strukturelle und physikalische Probleme.

Vergleicht man jedoch die Ergebnisse der beiden Marker rs873196 und rs1528460 in der evaluierten Studie bei Phillips *et al.* (2007a), fällt in der Gesamttabelle (Tabelle A-9 im Anhang) zunächst auf, dass der Marker rs873196 als unzuverlässig deklariert wurde. Dieser wurde mit allen Methoden detektiert und lieferte bei der Typisierung mit dem *MassARRAY*[®] *iPLEX* System (SEQUENOM) falsche Ergebnisse. Für Marker rs1528460 wurden die Genotypen nur mit *SNaPshot*[™] und *Genplex* bestimmt. Der Marker war in die *iPLEX*-Typisierung nicht integriert. Betrachtete man die beiden SNPs, so fiel zunächst auf, dass für beide keine heterozygoten Genotypen detektiert wurden. Mit dem Hintergrund einer erforderlichen Variabilität der untersuchten SNPs (Phillips *et al.* 2004c) und der aus Datenbanken entnommenen Heterozygotenrate von mind. 0,22 für rs873196³⁵ und 0,51

³⁵ http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=rs873196

(außer bei afrikanischen Populationen) für den SNP rs1528460³⁶, wurden beide Marker noch einmal genauer betrachtet. Nach einem Vergleich mit den Ergebnissen der *SNaPshot*TM I-Typisierung und der Typisierung mittels Microarray, die eindeutige Heterozygote aufwiesen, wurden die Ergebnisse der Referenzplattformen angezweifelt und mittels einer nachträglichen, stichprobenartigen Sequenzierung wurde das Vorliegen einer Heterozygotie bestätigt. Die Ergebnisse wurden zugunsten der Microarray-Typisierung korrigiert.

Hansen *et al.* (2007) diskutieren die Problematik der partiellen Fehltypisierung einzelner Polymorphismen durch das *SNaPshot*TM-Verfahren. Selbst eindeutig homozygote Ergebnisse konnten den entgegengesetzten Genotyp aufweisen. In der Gesamtstudie wurden für vier Marker Fehleinbauraten zwischen 6 und 15% detektiert. Die Detektionshäufigkeit von G/G statt G/A, so wie es die beiden hier untersuchten Marker betrifft, lag bei 4,5%. Auch hier lag die Vermutung nahe, dass neben einigen strukturellen Problemen des genetischen Ausgangsmaterials eine unbalancierte Allel amplifikation dem Phänomen zu Grunde liegt.

Von einer bekannten Problematik der Heterozygotentypisierung sprechen auch Simsek *et al.* (2001) und verweisen auf Überlegungen von Leren *et al.* (1993), die Schwierigkeiten in der Typisierung von G/A-Heterozygoten in Sequenzierungsreaktionen auf ungleichmäßige Signalstärken zurückführen. Sie fügen an, dass sowohl die Polymerase und ihre Wirksamkeit des Nukleotideinbaus als auch der Sequenzkontext die Ursachen der Signalunterschiede darstellen. Dieses erkannten auch Tabor und Richardson (1989, 1990). Sie beobachteten den ungleichen Einbau der terminalen Nukleotide und führten diesen auf einen Seitenkettenunterschied innerhalb der Aminosäuren der Polymerasen zurück. Sie veränderten die Polymerase, indem sie Thyrosin statt Phenylalanin einbauten und erhielten damit einen einheitlicheren Einbau der Abbruchmoleküle (Tabor und Richardson, 1995). Die beiden Forscher waren auch die Ersten, die Versuche durchführten bei denen statt Magnesium, Mangan in der Sequenzierungsreaktion eingesetzt wurde, was zu einer weiteren Optimierung des Nukleotideinbaus führte (Tabor und Richardson, 1987).

Korch und Drabkin (1999a und 1999b) griffen diese Idee auf und setzten Mangancitrat in der Sequenzierung ein, was speziell dazu führte, dass der G/A-Genotyp einen deutlich verbesserten Einbau beider Basen aufwies und eindeutig als heterozygot erkannt werden konnte.

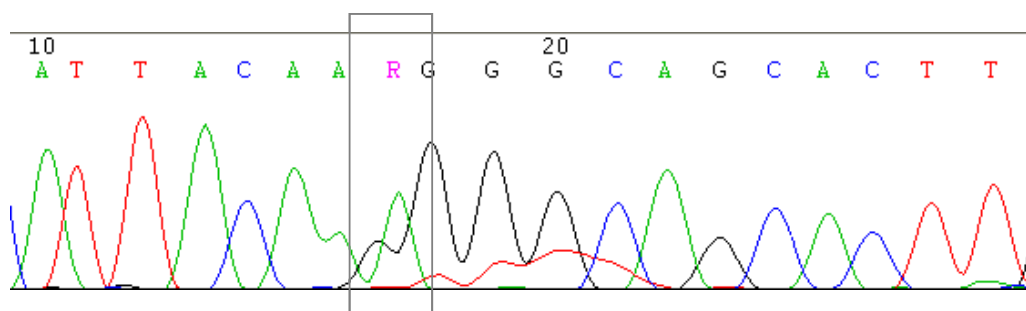
Im Gebrauch von Rhodamin-Farbstoffen als Fluoreszenzmarker terminaler Nukleotide erhärtete sich der Verdacht, dass die markierten G-Basen eine Suppression der Signalstärke erfahren, wenn es sich bei der vorangegangenen Base um Adenin handelt. Zusätzlich kommt es dabei zu einer Erhöhung der Signalstärke des nachfolgenden Nukleotids, was ebenfalls eine Unterdrückung der G-Base zufolge hatte (Parker *et al.* 1995, 1996, Zakeri *et al.* 1998).

³⁶ http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=rs1528460

Mit neuen Farbstoffen (dRhodaminen und den *Big Dyes*³⁷) verzeichneten die Forscher zwar eine deutlich stärkere Uniformität der Signalstärken sowie eine geringere Supression der G-Moleküle, dennoch tritt die Problematik der Signalunterdrückung in der Sequenzierung weiterhin vereinzelt auf.

Zu vergleichbaren Eigenschaften der hier untersuchten Marker gehörten der G/A-Polymorphismus sowie ein in der Sequenz 3' vorangehendes A-Nukleotid (Abbildung 46, Abbildung 49).

a)



b)

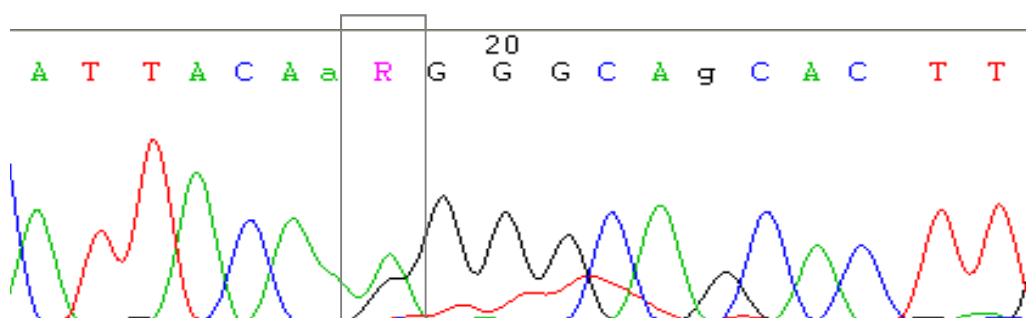


Abbildung 49: Elektropherogramme der sequenzierten Marker. a) SNP rs873196 Probe E34a, b) SNP rs873196 Probe E73a. Die SNP-Bezeichnung nach IUB-Code ist mit R gekennzeichnet. Der heterozygote Marker wird durch das unterdrückte G-Signal nicht erkannt und als ein homozygoter SNP detektiert.

Die niedrige Signalstärke beider SNPs sowie die unterdrückten Signale des G-Nukleotids (Abbildung 49), deuten auf das oben beschriebene Phänomen hin.

Ausgehend von dem Fakt, dass es sich auch bei den im *SNaPshot*TM-Kit verwendeten Farbstoffen um Rhodaminderivate handelt und die zur polymorphen Base benachbarte Sequenz eine Rolle beim Einbau der farbstoffgekoppelten ddNTPs spielt, ist dies als eine potentielle Fehlerquelle zu erkennen. Zudem lagen von 22 A/G-Polymorphismen in der gesamten 52 Marker Multiplexreaktion drei der SNPs in der Nachbarschaft zu einem A-Nukleotid. Bei zwei Markern handelte es sich um die bereits beschriebenen problematischen Marker. Der dritte SNP (rs733164, interne Bezeichnung: A22) war nicht Teil der microarraybasierten Methode und wurde nicht verglichen. In den jeweiligen

³⁷ Energie-Transfer-Varianten der dRhodamin-Farbstoffe. Fluorophore im *BigDye Terminator Kit* (Applied Biosystems)

Referenzmethoden mit denen der Marker untersucht wurde, zeigte dieser keine Detektionsprobleme auf, musste aber in einigen Fällen nach einer Prüfung korrigiert werden.

5.3.4 Qualitätsprüfung mittels *Silhouette*-Koeffizienten

Der Verlauf der Methodenevaluierung beinhaltete die Visualisierung der Daten sowie ihre objektive Auswertung mit kalkulatorischen Mitteln. Diese umfasste die Darstellung der selektierten Marker sowie die vergleichende Prüfung ihrer Genotypisierungserfolge im Assay.

Die mathematische Beurteilung der Assay-Qualität wurde im Anschluss an die Analysen mittels des *Silhouette*-Koeffizienten durchgeführt. Der hierzu benutzte Algorithmus im *ClusterA*-Programm erlaubt eine interne Distanzprüfung im und innerhalb der Cluster im Plot, durch welche man auf die Robustizität des Versuchsansatzes schließen kann. Eine hohe Robustizität eines Experiments definiert sich durch eine möglichst kurze Distanz der Signalpunkte innerhalb eines Genotypclusters sowie lange Distanzen zwischen den drei möglichen Genotypclustern. Diese Art der Qualitätskontrolle ist besonders hilfreich in der Optimierungsphase eines Experiments, bei dem mehrere Versuchsansätze unter differenten Bedingungen getestet werden (Lovmar *et al.* 2005).

Um aufgrund der *Silhouette scores* keinen vorzeitigen Ausschluss der Marker zu erzielen, wurde die Analyse im Anschluss an die bereits beschriebenen Experimente durchgeführt. Sie sollte nach der eigentlichen Analyse die Beurteilung in der Diskussion unterstützen. Diese Darstellung bündelt zusätzlich die Informationen der Assays und stellt neben der statistischen Betrachtung eine qualitative Darstellung dar. Der Einfluss des alternativ angewendeten Normalisierungsschrittes konnte auf diese Weise präziser visualisiert werden.

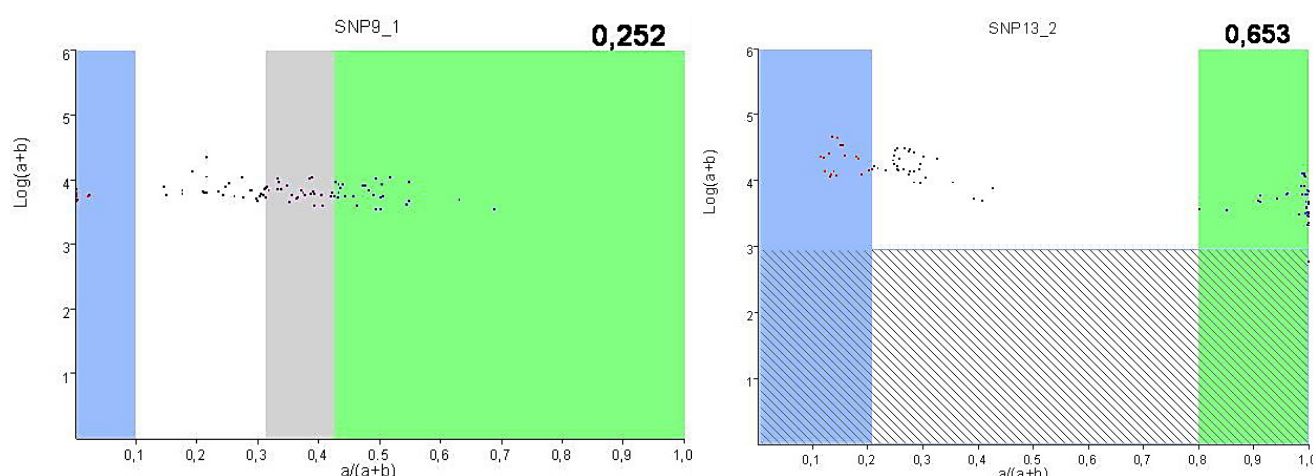


Abbildung 50: Beispiele von Clusterplots mit errechneten *Silhouette scores* entsprechend der Literaturvorgaben (siehe Text).

Generalisierende Kriterien eines Ausschlusses sind *scores*, die niedriger ausfallen als 0,25 (Liu *et al.* 2003, Lovmar *et al.* 2005, Abbildung 50 a). Erst ab einem *score* von über 0,65 können die Cluster als uneingeschränkt getrennt betrachtet werden (Abbildung 50 b).

Die Ergebnisse der *Silhouette*-Koeffizientenauswertung untermauern die bisherigen Resultate, die nach einer visuellen Einschätzung der Clusterqualität und einer Betrachtung der Werte innerhalb der Plots ermittelt wurden. Es ist auf diese Weise möglich, eine Auswertung der Daten zu erzielen, ohne ein objektives Mittel zur Unterteilung der Qualität zu besitzen. Die Ermittlung der Werte mittels des *Silhouette*-Koeffizienten kann die Auswertung erleichtern. Sie ersetzt jedoch nicht die langwierige Umwandlung der Daten zur Ermittlung der Genotypen. Abbildung 50 zeigt stark abweichende Werte der *Silhouette*-Koeffizienten für zwei Marker, die eine vergleichbare Zuordnung ohne die Koeffizienten-Ermittlung erfahren haben.

Anhand des Diagramms 12 kann zusammenfassend die Qualität der Auswertung dargestellt und beurteilt werden. Es verdeutlicht noch einmal die Unterschiede der beiden Probenserien 1 und 2, die eine unterschiedliche DNA-Qualität besaßen. Die Unterschiede können jedoch auch auf die Chargenunterschiede der Materialien zurückzuführen sein. Über die gesamte Versuchsreihe hindurch wurden mehrere Komponenten wie Glasobjektträger, Polymerasen und fluoreszenzgekoppelten ddNTPs nachbestellt. Die Qualitätsänderungen nach Austausch der Minisequenzierungschemikalien waren weniger stark. In der Qualität der gespotteten Glasobjektträger bestanden jedoch große Unterschiede, die sich auf die Gesamtauswertung ausgewirkt haben.

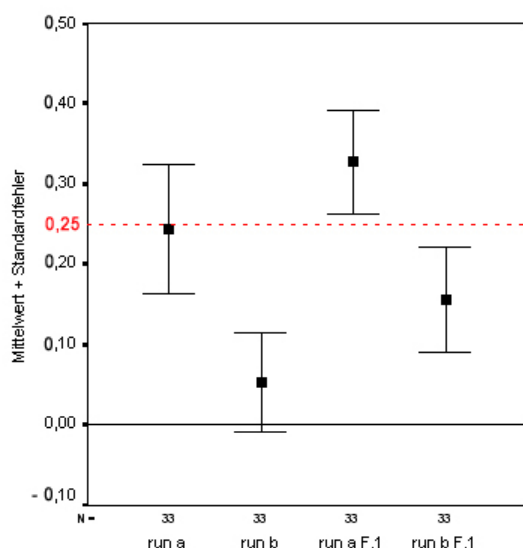


Diagramm 12: Zusammenfassende Plotdarstellung der Mittelwerte der errechneten *Silhouette scores*. Zu Qualitätsbeurteilung dient der Wert von 0,25 (rote gestrichelte Linie).

5.3.5 Sensitivität der Microarray-Multiplexreaktion

Von besonderer forensischer Relevanz waren die Ergebnisse der Adressierung von Proben mit stark reduzierter DNA-Konzentration nach der Applikation unterschiedlich stark verdünnter DNA-Mengen in der PCR-Reaktion. Die Tendenz hinsichtlich der Sensitivität der Microarray-Methode konnte in einem kleinen Versuchsansatz untersucht werden.

Die erzielten Ergebnisse unterstützen die bisherigen Vermutungen der Abhängigkeit der Signalintensität sowie der Spezifität von der Amplikonmenge, welche aus der PCR hervorgeht. Es kam hier zu Signalverlusten sowie zu einer stark unbalancierten Amplifikation und dem daraus resultierenden stochastischen Fehler im weiteren Verlauf der Detektion. Die bereits angedeuteten Signalverluste traten vermehrt bei DNA-Mengen von weniger als 500 pg auf. Eine genauere Nachweisgrenze ließ sich jedoch nur mit einem größeren Pool von Proben und einer geringeren Verdünnungsskala unterhalb dieser Grenze bestimmen. Das Auftreten unspezifischer Signale konnte nicht, wie erwartet, mit der Reduktion der PCR-Produktkonzentration verhindert werden. Im Gegenteil dazu zeigten einige der Marker einen systematischen Rückgang der Signalintensitäten auf, nicht jedoch der Fehlsignale in den signalfremden Farbkanälen (SNPs A7 und A5).

Zur Veranschaulichung des geringen Rückgangs der Falschsignale in der Verdünnungsreihe wurden die Signalintensitäten aus nicht gescannten Kanälen gegen die DNA-Menge aufgetragen. Sie lagen, verglichen mit den tatsächlichen Signalen, überwiegend im geringeren Intensitätsbereich, bargen aber die Gefahr der sterischen Konkurrenz auf dem Spot. Zudem stieg das Verhältnis der Fehlsignale zu den tatsächlichen Signalen, bei geringerer Menge DNA in der Reaktion (vergleiche Diagramm 10 b, 0,02 ng sowie Diagramm 13).

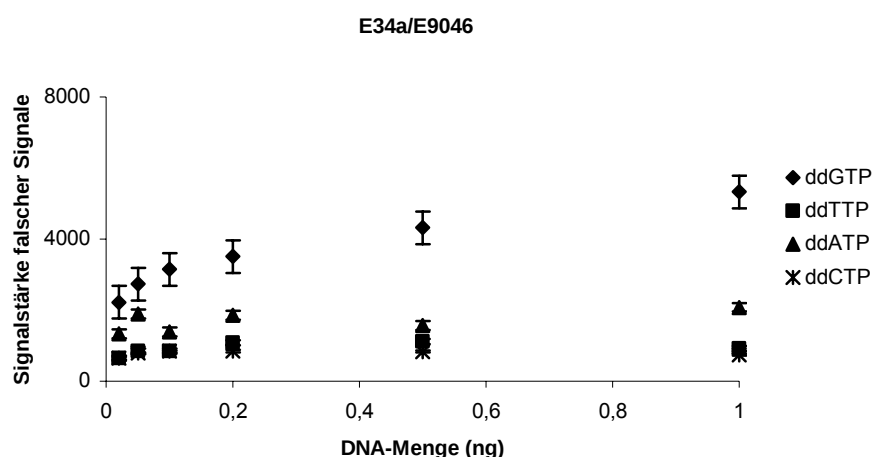


Diagramm 13: Darstellung der Signalstärken der in den vier Farbkanälen fehlerhaft detektierten Signale. Die stärksten Intensitäten wurden für den an ddGTP gekoppelten TAMRA-Farbstoff nachgewiesen. Diese führten im Experiment zu fehlerhaften Ergebnissen durch überlagerte Signalintensitätswerte.

Abschließend kann man feststellen, dass mit einem geringen Rückgang der Falschsignale auf den benachbarten Farbkanälen als Folge unspezifischer Hybridisierung, sich die Gefahr einer geringen Balance der detektierten Loci beim Einsetzen geringer DNA-Mengen verstärkt. Es folgt so ein unkontrollierter Anstieg falsch-positiver Signale, der vor einer Auswertung der Proben eine zusätzliche Normalisierung erfordert.

5.3.6 Abschließende Bewertung der Methode

Die Ursachen der ungleichmäßigen Verteilung der Werte sind bei diesem Assay jedoch so vielfältig, dass sie zusammenfassend in der Abfolge vom Resultat bis zur Versuchskonzipierung beschrieben werden. Den Umfang der Wirkungsweise der unterschiedlichen Komponenten kann man mit der erforderlichen Optimierung des experimentellen Ansatzes in Korrelation stellen. Nachfolgend sind die bisher diskutierten Parameter aufgelistet.

Normalisierung der Daten

Die mathematische Korrektur der detektierten Werte sollte im Versuchsansatz durch das Einbringen einer zusätzlichen Variablen erreicht werden. Sie resultierte aus der internen Reaktionskontrolle eines jeden Arrays. Obwohl die Reaktionskontrolle stabile Parameter beinhaltete, konnte die Hybridisierungsleistung nicht konstant gehalten werden, was der Vermutung entsprach, dass der nicht kontrollierbare Faktor der Sondendichte im Spot zusätzlich zu den Inkorporationsfaktoren wie Form und Größe der Fluoreszenzfarbstoffe eine Rolle spielte.

Heterozygotenkontrollen für jeden potentiellen Polymorphismus könnten eine Weiterentwicklung der Normalisierung darstellen.

Fluoreszenz-Scan

Das Einscannen der vorbereiteten Objektträger ist eine fast vollständig automatisierte Tätigkeit, die wenige Fehlerquellen in der Versuchsdurchführung bedingt. Die Möglichkeiten der Spektralfarben-Darstellung boten die notwendige Unterstützung bei der Einstellung des Photomultipliers. Die durch das Gerät fixierte Einstellung der Laserstärke erlaubte keine interne Anpassung an die Eigenschaften der Fluoreszenzfarbstoffe.

Hybridisierungskonditionen

Dieser Aspekt hat einen starken Einfluss auf den experimentellen Erfolg. Obgleich die vielen Schritte der Hybridisierungsreaktion im Versuchskontext relativ robust erscheinen, müssen die Bedingungen die höchstmögliche Konstanz erreichen. Chemische Lösungen, die manuell angesetzt werden, müssen in ihrer Zusammensetzung akribisch dem Protokoll entsprechen. Auch die Reaktionsvolumina sollten immer erreicht sein, um mögliches Austrocknen der Arrays zu vermeiden.

Einer der kritischsten Faktoren ist die Reaktions- bzw. Waschtemperatur. Durch die Erhöhung der Waschtemperatur wird eine deutliche Herabsetzung des Hintergrundrauschens erreicht. Die Erhöhung der Hybridisierungstemperatur um 1°C führte zu einer dauerhaften Steigerung der Reaktionsstringenz. Nur ein sehr geringer Anteil der Signale wurde als Hintergrund

detektiert und von den tatsächlichen Signalen subtrahiert. Die Spotreinheit wurde so optimiert.

Minisequenzierungsreaktion

Hier verbergen sich mit Abstand die sensitivsten Aspekte der Methode. Die Reaktion besteht aus einer Vielzahl von Komponenten, die im Einzelnen einen Einfluss auf den Reaktionserfolg nehmen und nachfolgend noch mal zusammenfassend aufgeführt werden.

Sequenzierenzyme

Es gibt eine Reihe von Sequenzierenzymen, die in ihrer Funktion gleich, in der Einbaupräferenz jedoch große Unterschiede aufweisen. Diesem Verhalten beim Nukleotideinbau kann nicht durch eine Konzentrationsänderung der Enzyme entgegengewirkt werden. Daher muss die Verwendung eines Enzyms den Versuchsbedingungen angepasst werden. Sie ist erheblich in der Auswirkung auf die Stringenz und Sensitivität der Untersuchungen. Die Anpassung erfolgte hier durch Verwendung von zwei unterschiedlichen Sequenzierenzymen.

Fluoreszenzmarkierte ddNTPs

Bei den fluoreszenzmarkierten Nukleotiden handelt es sich um Komponenten, bei deren Auswahl die meisten Faktoren berücksichtigt werden müssen. Die Eigenschaften der Fluorophore im gekoppelten und ungekoppelten Zustand unterscheiden sich erheblich. Sie korrelieren jedoch nicht miteinander und die Ursachen dafür sind nur in geringem Maße erforscht. Sterische Effekte, chemische Umgebung sowie die Kopplungs- und Nachbarmoleküle führen zu veränderter Quantenausbeute. Im Experiment erfordert die Konzentrationseinstellung der gekoppelten Farbstoffe ein besonderes Augenmerk. Die Anzahl der gekoppelten Fluoreszenzfarbstoffe ist begrenzt. Die Farbstoffe selbst sind patentrechtlich geschützt und unterliegen den Patentrechten eines Unternehmens (hier Perkin Elmer). Da die Farbstoffe auch maßgeblich für die Kosten der Reaktion sind, muss die Wahl gut überlegt sein und an die technischen Eigenschaften des Laserscanners angepasst werden. Spektralüberlappungen führen zur Notwendigkeit zusätzlicher mathematischer Korrektur.

Minisequenzierungsprimer

Die in der Reaktion eingesetzten Oligonukleotide sind die Komponenten, auf die der Experimentator den größten Einfluss ausüben kann. Physikalische Eigenschaften können mit diversen Programmen überprüft und eingestellt werden. Die Länge der Minisequenzierungsprimer kann variieren. Sie muss den energetischen Faktoren angepasst werden. Homo- und Heterodimerbildung der Primer und Bildung einer Sekundärstruktur mit der *tag*-Sequenz müssen verhindert werden. Nicht unerheblich für die Detektion der Polymorphismen kann die *annealing*-Richtung des Primers sein. Oft kann die Verlagerung des Primers auf den komplementären DNA-Strang eine Optimierung des Versuchsansatzes darstellen. Das Herabsetzen der Primerkonzentration in Reaktion kann zu einer Erhöhung der Spezifität führen. Sie vermindert das Konkurrieren der Moleküle im Reaktionsvolumen.

Zyklenzahl

Eine kritische, jedoch leicht einstellbare Größe in der Reaktion. Die Zyklenzahl kann nach Bedarf sogar bis auf 99 gesteigert werden (Lovmar und Syvänen, 2005). Die notwendige Sensitivität der Reaktion kann bei 30 Zyklen erreicht werden.

Multiplex-PCR-Reaktion

Die PCR-Reaktion für eine große Anzahl von Markern einzustellen, erfordert einige Übung. Anzahl der eingesetzten Marker sowie die Menge der Amplifikationsprodukte sind für den weiteren Erfolg der Reaktion ausschlaggebend und können zu erheblichen Signalverlusten

	sowie zu einer stark unausbalancierten Reaktion führen. Das Poolen kleinerer Reaktionen kann eine Alternative darstellen, erfordert jedoch eine gezielte Einstellung der Konzentration. Eine Alkohol-Fällung bzw. Dialyse, kann zu einer Erhöhung der Amplifikate führen und dadurch zu einer Verbesserung der Reaktion.
PCR-Primer	Die Anforderungen an das Primerdesign steigen mit dem Vorhaben einer Multiplexreaktion. Sekundärstrukturen verringern erheblich die Produktausbeute und müssen für alle Oligonukleotidsequenzen gegeneinander geprüft werden.
Blocken der Glasobjektträger	Ein notwendiger Schritt zur Erhöhung der Spezifität der Spots.
Spotten der Arrays	Die Konzentration der <i>tag</i> -Sequenzen im Spottingpuffer ist ausschlaggebend für die spätere Moleküldichte im Spot (Guo <i>et al.</i> 1994). Diese muss so gewählt sein, dass sie nicht zu gering ist, um die notwendige Signalstärke erreichen zu können. Eine zu hohe Moleküldichte kann ebenfalls in geringer Signalstärke resultieren. Sie ist die Folge sterischer Interferenz zwischen den immobilisierten Molekülen.
Microarray-Glasobjektträger	Die Wahl der Oberflächenbeschichtung der Glasobjektträger kann variieren. <i>CodeLink™</i>-Objektträger werden durch ihre Oberflächenbeschichtung für Oligo-Arrays empfohlen (Lindroos <i>et al.</i> 2001). Die Polymerbeschichtung der Glasobjektträger sorgt für eine vertikale Immobilisierung der Moleküle und erfordert in der Theorie keinen zusätzlichen <i>spacer</i>, dieser wird jedoch bei kurzen Oligos trotzdem integriert, um die <i>tag</i>-Sequenzen den Minisequenzierungsprimern in der Hybridisierung zugänglicher zu machen. Es hat sich gezeigt, dass ein <i>poly-dT(15)-spacer</i> das Hybridisierungssignal deutlich erhöht (Guo <i>et al.</i> 1994). Die kovalente Bindungsform der Oligonukleotide erlaubt stringente Waschschrte und eine Möglichkeit der Wiederverwendung der Objektträger nach mehreren <i>stripping</i>-Vorgängen (Dolan <i>et al.</i> 2001, Beier und Hoheisel, 1999).

5.4 Vergleichende Betrachtung anderer Typisierungsmethoden und Ausblick

Entgegen der anfänglichen Skepsis, eine statistisch erforderliche Anzahl von über 50 SNP Markern in einer Multiplexreaktion typisieren zu können, haben die am EU-Projekt teilnehmenden Laboratorien eine solche Methode mit der 52 SNP Multiplex erfolgreich etabliert (Sanchez *et al.* 2006). Die für forensische Zwecke entwickelte und validierte Multiplexreaktion wurde mit Hilfe der *SNaPshot™*-Minisequenzierung entwickelt. Durch ihre Robustheit konnte diese auf weiteren Analyseplattformen getestet und weiterentwickelt werden.

Unentbehrlich für die erfolgreiche Etablierung der Methode waren die vorformulierten Kriterien für die SNP-Selektion. Eine Unabhängigkeit aller Marker zu benachbarten SNPs und funktionalen Genen, eine gute Sequenzqualität sowie einer biostatistische Effizienz durch die Auswahl der Marker aus Frequenzbereichen zwischen 0,18 und 0,28, machten es

schließlich möglich, eine Reaktion mit dieser Anzahl von Markern durchzuführen. Zusätzlich konnten zwei weitere Reaktionen, die in einem PCR-Ansatz durchgeführt wurden, auf der Basis der Minisequenzierung etabliert werden: eine 34 SNP Reaktion, die aufgrund der *ancestral informative markers* Hinweise auf die Ursprungspopulation liefern konnte (Phillips *et al.* 2007b) und die Y-chromosomale 29 SNP Multiplexreaktion (Brion *et al.* 2005). Diverse molekularbiologische Verfahren konnten auf der Basis der beschriebenen Methode auf ihre Effektivität, Sensitivität und Genauigkeit ausgetestet werden. Nicht unwesentlich war dabei auch die Kosteneffizienz der Reaktionen. In der vorliegenden Arbeit stand im Vordergrund, die Multiplexreaktion auf einer hybridisierungsbasierten Microarray-Plattform zu testen. Dieses besonders robuste Verfahren war aufgrund der niedrigen Herstellungskosten und der einfachen Handhabung für das Projekt von Interesse.

Umso enttäuschender wurde das Ergebnis zur Kenntnis genommen, dass ohne Einbußen in der Sensitivität, der Reproduzierbarkeit und der Genauigkeit keine Ergebnisse erzielt werden konnten. Die Verluste jener Eigenschaften machten die Methode nicht zuverlässig genug für forensische Untersuchungen.

Eine aufwendige Umgestaltung des Versuchsansatzes hin zur Genotypisierung jeder der vier Basen in vier getrennten Farbkanälen hatte eine Steigerung der Reproduzierbarkeit zur Folge. Dies konnte nur im geringen Maße die Fehleranfälligkeit des Ansatzes aufwiegen. Die Ursachen der Problematik konnten nur in Ansätzen mit den zur Verfügung stehenden Mitteln behoben werden. Die Anpassungen wurden zum Nachteil der Effizienz vorgenommen. Sie erforderten eine Reduzierung der Markeranzahl, um sterische und molekulare Interaktionen der Fluoreszenzmoleküle herabzusetzen. Weiterhin konnte die aufwendige Auswertung der Signale nur in begrenztem Maße mit Hilfe von Makros automatisiert werden. Die große Datenmenge, die durch die Typisierung der Marker in vier Farbkanälen anfiel und eine Umwandlung der Daten in vielen Einzelschritten erforderte, war mit einem hohen Zeitaufwand verbunden und beinhaltete eine große Anzahl möglicher Fehlerquellen.

Ein anderer Ansatz einer Multiplexreaktion auf Glasobjektträgern ist im Vorfeld im Partnerlabor in Santiago de Compostela entwickelt worden (Sobrinho *et al.* 2004). Die für 29 Y-chromosomale Marker etablierte Reaktion wurde dabei in vier Multiplexen durchgeführt. In jeder der Multiplexreaktionen wurden die PCR-Produkte mit jeweils einer fluoreszenzgekoppelten Base markiert. Die Genotypen konnten so aus der Reihe der vier den Basen zugeordneten Spots ermittelt werden (Abbildung 51).

Durch die Verwendung eines einzigen Fluoreszenzfarbstoffs (hier JOE, Anregung: 520 nm; Emission: 548 nm) konnte die Problematik der unterschiedlichen Einbaueigenschaften umgangen werden, wenn auch hier die Kopplung an unterschiedliche Nukleotide zu Verlusten der Signalintensität führen kann. Die Reaktion ist jedoch in der Gesamtheit stabiler. Nachteilig ist hier jedoch die Tatsache, dass die Reaktion nach der PCR in vier

unterschiedliche Multiplexe aufgeteilt werden muss. Dies führt neben einem erhöhten Verbrauch an terminalen Nukleotiden auch zu einem Mehraufwand an Arbeits- und Auswertungsschritten. Der methodische Ansatz scheint dennoch eine höhere Reproduzierbarkeit aufzuweisen.

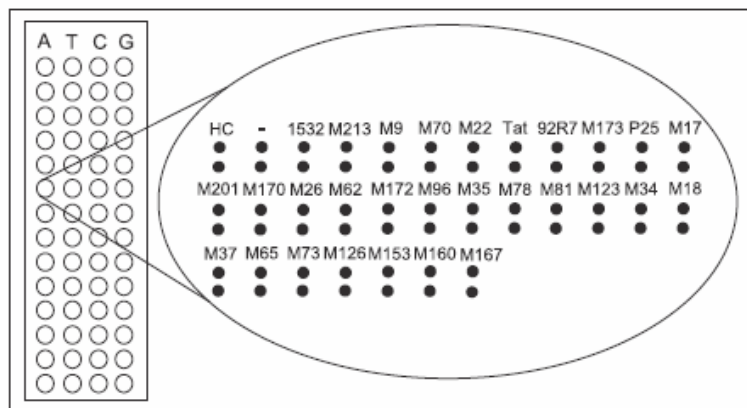


Abbildung 51: Mit insgesamt 4 x 14 Spots belegter Glasobjektträger für die 29 SNP Multiplexreaktion der Y-chromosomalen SNPs. Das Format wurde durch das „array of arrays“ Prinzip erzielt. In der Minisequenzierungsreaktion mit einem Fluoreszenzfarbstoff, gekoppelt an jeweils eine der vier Basen, wurden die Amplikons verlängert (aus Sobrino *et al.* 2004).

Ausgehend vom APEX-Prinzip (Kapitel 2.2.4.2) und speziell für forensische Zwecke haben die Forscher um Krjutškov die weiterführende APEX-2-Methode entwickelt (Krjutškov *et al.* 2008), um mit ihr 124 biallelische Marker mittels Microarray zu typisieren (Krjutškov *et al.* 2009). Das Prinzip basiert auf einer zwei-Phasen-Reaktion, bei welcher zunächst in der PCR spezifische Primer in beide Richtungen unmittelbar an die polymorphe Stelle anlagern. Sie werden verlängert und im zweiten Schritt hybridisiert eine universelle Sequenz am 3'-Ende des vorab synthetisierten Abschnitts. Dieser wird nun ihrerseits verlängert. Die APEX-Primer, die ebenfalls mit einer universellen Sequenz versehen sind, können an die Glasobjektträger gekoppelt werden. Die im zweiten Schritt generierte Sequenz kann nun an den immobilisierten Primer anlagern und die SBE-Reaktion mit fluoreszenzmarkierten Nukleotiden findet auf dem Objektträger statt. Diese Methode erwies sich trotz ihres Durchsatzes als sehr exakt mit einer geringen Zahl falscher Signale. Die Grenze ihrer Genauigkeit lag jedoch bei einer Ausgangsmenge von 25 ng DNA, durch welche dieser Ansatz mehr für Untersuchungen zu genealogischen Zwecken nützlich ist, zumal neben autosomaler auch mitochondriale und gonosomale DNA parallel untersucht werden kann. Die unterschiedlichen Ansätze, eine durchsatzfähige und dennoch sensitive Methode für die SNP-Microarray-Detektion auf Glasobjektträgern zu etablieren, gestalteten sich aufgrund der beschriebenen Hürden als schwierig. Im Besonderen die Detektion unterschiedlicher Fluoreszenzfarbstoffe erforderte eine aufwendige Laborausrüstung und unkonventionelle Programmlösungen bei der Auswertung.

Eine Weiterentwicklung der konventionellen Microarray-Methode auf Glasobjektträgern stellt die *Nanogen*-Plattform dar. Es handelt sich dabei um elektronisch adressierte Microarrays auf einem so genannten *NanoChip*TM. Diese Technologie wurde zunächst im Partnerlabor in Kopenhagen angewendet (Børsting *et al.* 2004). Im Institut für Rechtsmedizin in Mainz wurden weiterführende Experimente an der *NanoChip*TM *Molecular Biology Workstation* (MBW, Nanogen) und am *NanoChip*TM 400 System (NC400, Nanogen) durchgeführt. Die Entwicklung mehrerer Hybridisierungsassays und die Etablierung unterschiedlicher Multiplexanalysen auf diesen Plattformen sind in der Dissertationsarbeit von Balogh (2009) ausführlich dargestellt. Das Prinzip der Analyse basiert auf der elektronischen Adressierung einzelsträngiger DNA-Sonden, ihrer Immobilisierung über reversible Biotin-Streptavidin-Bindungen auf dem mit einer Hydrogelschicht beschichtetem Mikrochip und der anschließenden Hybridisierung der Zielsequenzen, gefolgt von der thermischen Diskriminierung und der fluoreszenzbasierten Detektion. Nach der vorab durchgeführten PCR erfolgen alle Schritte vollautomatisch, verteilt entweder auf zwei computergesteuerte Instrumente (*loader* und *reader* bei der *NanoChip*TM *Molecular Biology Workstation*) oder über eine einzige Geräteeinheit wie bei der NC400-Plattform. Der herausragende Fortschritt der elektronischen Microarray-Technologie besteht zudem in der vollen Individualisierbarkeit der Versuchsansätze. Diese wird dadurch gewährleistet, dass kein vorab festgelegtes Spotten der Chipoberfläche erfolgt und dem Anwender dadurch keine Grenzen gesetzt werden.

Die *NanoChip*TM-Technologie wurde als erste elektronische Microarray-Technologie im Rahmen des Projekts für forensische Analysen angewendet. Vor allem die Methode des so genannten *capture down* etablierte sich als besonders multiplexfähig und konnte für eine Analyse von 32 SNPs validiert werden. Mit einer Effizienz von 94 % bei der Genotypisierung wurde zwar noch nicht die erforderliche Genauigkeit für den Einsatz in der forensischen Routine erreicht, die Methode überzeugt jedoch in der Handhabung und der Schnelligkeit sowie ihrer Flexibilität. Sie kann bei einer entsprechend größeren Anzahl von Markern auf 98 erweitert werden. Die Nachteile für die Anwendung liegen in der begrenzten Verfügbarkeit des Geräts und der die Microchips beinhaltenden *cartridges*. Letztere können jedoch mehrmals verwendet werden und durch die Etablierung von universellen Reportern entfallen weitere Kosten für einzelne fluoreszenzmarkierte Sonden.

Neben den microarraybasierten Anwendungen konnte die massenspektrometrische Methode des MALDI-TOF gute Ergebnisse hinsichtlich der parallelen Typisierung einer großen Anzahl von Markern erzielen. Bei dieser Methode werden die einzelnen Nukleotide nach ihrem spezifischen Molekulargewicht detektiert. Die Detektion durchläuft zunächst die Schritte der PCR und der Minisequenzierung. Nach einer Aufreinigung der Sequenzabschnitte mit der terminalen allelspezifischen Base werden die auf eine Matrix überführten Produkte durch

Laserenergie ionisiert und im elektrischen Feld beschleunigt. Die Messung ihrer Molekülmasse erfolgt in der Abhängigkeit zur Flugzeit. Hierbei bereitete zunächst die Ungenauigkeit in der Trennung einzelner Polymorphismen große Schwierigkeiten. Diese wurden durch die Kopplung von ddUTP an einen Aminolinker und der so herbeigeführten Erhöhung der spezifischen Masse gelöst (Mengel-Jorgensen *et al.* 2004). Weitere Modifikationen in der Versuchsanordnung, wie die Verwendung von biotinkoppelten ddNTPs in der SBE-Reaktion führten zu einer gezielten Separation der markierten Oligonukleotide. Im folgenden Aufreinigungsschritt ließen sich diese durch die Bindung an Streptavidin-*beads* gezielt von unmarkierten Oligos trennen, was eine höhere Effizienz der Reaktion zur Folge hatte. Mit dem Ziel den Messbereich zu erweitern, konnte die so genannte MARNA-Methode etabliert werden. Durch die Spaltung einer RNA-Sequenz im SBE-Primer mittels RNase konnten besonders Bereiche mit einem geringen Molekulargewicht für die Detektion gewonnen werden (Mengel-Jorgensen *et al.* 2005). Aufgrund der beschriebenen Modifikationen konnten schließlich in dem Bereich zwischen 3000 und 10000 Da, 50 der validierten 52 SNPs erfolgreich mittels MALDI-TOF typisiert werden (Wächter *et al.* 2008). Trotz der hohen Multiplexfähigkeit und der vergleichbar niedrigen Kosten der Reaktion sowie der schnellen Detektion auf dem Massenspektrometer ist diese in der Anwendung der beschriebenen Modifikationen jedoch sehr arbeits- und zeitaufwendig.

Eine weitere sequenzbasierte Methode, die im Rahmen des Projekts untersucht wurde, behilft sich der Detektion von Lichtsignalen. Die so genannte Pyrosequenzierung beinhaltet eine PCR-Amplifikation mit 5'-biotinylierten Primern, die anschließend durch eine Streptavidin-Sepharose-Bindung von den übrigen Produkten getrennt werden. In der folgenden Minisequenzierungsreaktion erfolgt eine gezielte und sukzessive Zugabe von dNTPs. Erfolgt der Einbau eines komplementären Nukleotids, wird durch die Freisetzung eines Pyrophosphat-Moleküls (PPi) eine enzymunterstützte Signalkaskade in Gang gesetzt, die in der Freisetzung eines Lichtsignals resultiert. Die Intensität des Lichtsignals steigt dabei proportional zu der Menge der eingebauten Nukleotide (Ahmadian *et al.* 2006, Nyren, 2007). Diese Methode wird nicht in die vergleichende Diskussion einbezogen. Eine Validierung konnte nicht abgeschlossen werden, da der ursprüngliche Hersteller (Pyrosequencing) im Verlauf des Projekts den Geschäftsbereich (Sequenzierung) und den Namen wechselte (Biotage, 2003 und anschließend Qiagen, 2008) und für weitere Etablierungen für forensische Bereiche nicht weiter zur Verfügung stand. Bis zu diesem Zeitpunkt konnte eine Multiplexreaktion etabliert werden, die 22 SNPs in acht Triplexen erfolgreich typisieren konnte (Bender *et al.* 2006). Versuche im Londoner Laboratorium konnten für eine 5 SNPs Multiplexreaktion die Nachweisgrenze von 30 pg aufzeigen (Harrison *et al.* 2006).

Aus den Anwendungsmöglichkeiten der Kapillarelektrophorese in einem forensischen Labor für die Routineuntersuchungen resultieren die Vorteile der Genotypisierung mittels *SNaPshot*TM-Referenzmethode. Das Gerät befindet sich in fast jedem Routinelabor und kann ohne großen Aufwand nach Bedarf auf die Detektion von SNPs umgestellt werden. Die Schwierigkeiten zeigen sich jedoch in den Bereichen der Etablierung und in der Auswertung der Multiplexdaten. Aus den laborübergreifenden Validierungsstudien ging hervor, dass nicht alle Geräte das gleiche Potential für die Methode aufweisen. Typabhängige Abweichungen sowie Unterschiede in der Sensitivität der Geräte mussten für den weiteren Verlauf der Studie einkalkuliert werden. Es folgte eine Anpassung der Protokolle, so dass in den meisten Laboratorien die 52 SNP Multiplex auf zwei getrennte Reaktionen aufgeteilt wurde. In der Validierungsstudie mussten in vielen Laboratorien bis zu 20 % der Proben mindestens ein drittes Mal nachtypisiert werden, bis die Ergebnisse übereinstimmten. Eine manuell durchgeführte Nachkorrektur der Ergebnisse war ebenfalls nicht selten erforderlich. Hier zeigte sich, dass die Auswertung von SNPs ein geschultes Auge erfordert und von in der SNP-Analyse erfahrenen Personen durchgeführt werden sollte.

Einschränkungen der *SNaPshot*TM-Methode in ihrer Multiplexfähigkeit waren durch die Fragmentlängentrennung vorgegeben. Laufabhängig weist jedes Molekül eine gewisse Unschärfe des Detektionsbereichs auf, so dass ein Abstand von mindestens vier Basen zwischen zwei Produkten eingehalten werden musste. Im Gegensatz zur STR-Analyse oder der Sequenzierung konnte für die SNP-Detektion nur ein relativ kurzer Bereich von ca. 130 bp für eine effiziente Genotypisierung genutzt werden. Längere Minisequenzierungsprimer sind in der Herstellung sehr kostenintensiv und können erhebliche Qualitätseinbußen durch die Bildung von Dimeren aufweisen. Somit ist mit der Typisierung der 52 SNP Multiplex bereits ein Limit gesetzt. Ein größerer Multiplexansatz könnte nach einer zusätzlichen Aufteilung der Reaktion typisiert werden, dies würde jedoch die Kosten und den Zeitaufwand erheblich steigern. Die Hochdurchsatzfähigkeit ist zudem durch die limitierte parallele Typisierung der Individuen gegeben. Zwar gibt es Geräte mit bis zu 96 Kapillaren, diese sind jedoch selten in den forensischen Laborbereichen. Die routinemäßige Typisierung in einem solchen Labor erfolgt gewöhnlich mit 4 bis 16 Kapillaren.

Dennoch ist die *SNaPshot*TM-Methode trotz ähnlichen Zeitaufwands bei der Reaktionsdurchführung im Hinblick auf die nachfolgende Detektion und Auswertung der Daten erheblich benutzerfreundlicher als die hybridisierungsbasierte Microarray-Methode, die hier lediglich durch ihre Parallelität einen Vorteil aufzeigt.

Zuletzt sollte die zweite *Genplex*-Referenzmethode näher betrachtet werden. Diese Technologie ist auf die Maximalmenge von 48 SNPs, die gleichzeitig typisiert werden können, beschränkt. Vergleichend zu der Microarray-Plattform liegt sie dabei im oberen Bereich der etablierten Multiplexe und durch ihre Sensitivität konnten DNA-Mengen von 1 ng

DNA erfolgreich typisiert werden (Phillips *et al.* 2007a, Musgrave-Brown *et al.* 2008). Die ligationsbasierte Methode zeigt ihre Vorteile in der einfachen und schnellen Handhabung sowie ihrer Kosteneffizienz. Gegenüber den beschriebenen Plattformen ist sie jedoch eine der unflexibelsten. Die Entwicklung des gesamten Assays erfolgt durch den Hersteller (Applied Biosystems) nach der Vorgabe des Experimentators, kann jedoch nicht durch ihn beliebig variiert werden.

Für die beiden Referenzmethoden konnte jedoch eine essentielle Schwachstelle aufgedeckt werden. Wie im Kapitel 5.3.3 ausführlich beschrieben, zeigte sich erst im Vergleich die Problematik der Detektion von G/A-Polymorphismen, die möglicherweise durch eine Unterdrückung des G-Signals zustande kommt, wenn diesem in der Sequenzabfolge ein A vorangegangen ist (Parker *et al.* 1995, 1996, Zakeri *et al.* 1998). Diese Problematik wurde bisher für beide Plattformen nicht weiter verfolgt, kann jedoch beim entsprechenden Nachbarnukleotid eine Limitierung der SNP-Auswahl zur Folge haben.

Im vorangegangenen Abschnitt wurden die zahlreichen Arbeitsplattformen für die Genotypisierung von SNPs diskutiert. Den SNPs selbst gebührt jedoch auch ein kurzer Ausblick. Über die letzten Jahre hinweg haben sich die Einzelbasenpolymorphismen zu erfolgreichen Werkzeugen in vielen Bereichen der biomedizinischen Forschung und der medizinischen Diagnostik entwickelt. Die Analysemöglichkeit der SNPs, als Veränderungen der Sequenz an einer definierten Stelle, statt der Detektion von Längenpolymorphismen erlaubt mit den für diagnostische Bereiche konzipierten Hochdurchsatzmethoden in Kopplungsanalysen und genomweiten Assoziationsstudien, Gene zu entdecken und zu kartieren, die für eine große Zahl von Krankheiten verantwortlich sind.

Der Umfang der weltweiten Datenbanken, die Informationen über die Polymorphismen und ihre flankierenden Sequenzbereiche enthalten, wurde auch im Rahmen des SNPforID-Projekts erfolgreich genutzt, um u.a. eine Auswahl von biostatistisch relevanten Markern zu treffen, die eine sichere Identifizierung von Individuen erlauben. Unter Wissenschaftlern herrscht Konsens darüber, dass die Etablierung einer SNP-Multiplexmethode nicht dazu dienen sollte auf lange Sicht die STR-Datenbanken zu ersetzen (Gill *et al.* 2001). Durch die Vorteile der kurzen Fragmente und der geringen Mutationsrate unterstützen SNPs vielmehr die Fälle, die mittels Längenpolymorphismen nicht gelöst werden können.

Aus diesem Grund wurden mit der Zeit weitere genetische Bereiche interessant und so liegt bereits ein Schwerpunkt in der Erforschung abstammungsrelevanter Gene, die Hinweise auf phänotypische Merkmale geben können. Dabei kann es sich um Merkmale wie die Pigmentierung der Haut (Shriver *et al.* 2003), der Haare (Flanagan *et al.* 2000) oder der Iris (Frudakis *et al.* 2003) handeln. Diese weiterführenden und nicht unumstrittenen Untersuchungen solcher Merkmale im Hinblick auf die Eingrenzung oder Identifizierung unbekannter Personen im Verbrechensfall werden in der Gesellschaft kontrovers diskutiert,

werden jedoch in den Vereinigten Staaten und einigen europäischen Ländern bereits angewendet (Kayser und Schneider, 2009).

In Deutschland bleiben SNPs vorerst unterstützende Werkzeuge, die in Fällen von Untersuchungen mit degradiertem DNA oder bei der Erhöhung der Diskriminierungskraft in Verwandtschaftsanalysen Verwendung finden.

Die Bestrebungen zur SNP-Analyse können, wie in der vorliegenden Arbeit dargestellt, nicht immer mit dem gewünschten Erfolg abgeschlossen werden. Durch die Untersuchung der Marker an mehreren Analyseplattformen wurde deutlich, dass jede Methode ihre Vor- und Nachteile hinsichtlich der geforderten Kriterien birgt. Es gibt jedoch keine Methode zur SNP-Analyse, die alle gewünschten Eigenschaften in sich vereinigt.

6 Literaturverzeichnis

- Ahmadian A, Ehn M, Hober S (2006) Pyrosequencing: history, biochemistry and future. *Clin Chim Acta* 363:83-94.
- Allard MW, Miller K, Wilson M, Monson K, Budowle B (2002) Characterization of the Caucasian haplogroups present in the SWGDAM forensic mtDNA dataset for 1771 human control region sequences. Scientific Working Group on DNA Analysis Methods. *J Forensic Sci* 47:1215-1223.
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457-465.
- Arriola E, Lambros MB, Jones C, Dexter T, Mackay A, Tan DS, Tamber N, Fenwick K, Ashworth A, Dowsett M, Reis-Filho JS (2007) Evaluation of Phi29-based whole-genome amplification for microarray-based comparative genomic hybridisation. *Lab Invest* 87:75-83.
- Ayres KL (2005) The expected performance of single nucleotide polymorphism loci in paternity testing. *Forensic Sci Int* 154:167-172.
- Balding DJ, Nichols RA (1994) DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Sci Int* 64:125-140.
- Balogh MK, C. Børsting, P. Sanchez Diz, C. Thacker, D. Syndercombe-Court, A. Carracedo, N. Morling, P.M. Schneider (2006) Application of whole genome amplification for forensic analysis. Elsevier, International Congress Series 1288: 725– 727.
- Balogh MK (2009) Forensisch relevante SNP-Genotypisierung mittels elektronischer Microarray-Technologie. Dissertation, 2009.
- Barber AL, Foran DR (2006) The utility of whole genome amplification for typing compromised forensic samples. *J Forensic Sci* 51:1344-1349.
- Beier M, Hoheisel JD (1999) Versatile derivatisation of solid support media for covalent bonding on DNA-microchips. *Nucleic Acids Res* 27:1970-1977.
- Bender K, C. Nehlich, C. Harrison, E. Musgrave-Brown, D. Syndercombe-Court, P.M. Schneider, the SNPforID Consortium (2006) A multiplex SNP typing approach for the DNA pyrosequencing technology International Congress Series 1288, pp 73-75.
- Bergen AW, Haque KA, Qi Y, Beerman MB, Garcia-Closas M, Rothman N, Chanock SJ (2005) Comparison of yield and genotyping performance of multiple displacement amplification and OmniPlex whole genome amplified DNA generated from multiple DNA sources. *Hum Mutat* 26:262-270.
- Blouin MS, Yowell CA, Courtney CH, Dame JB (1995) Host movement and the genetic structure of populations of parasitic nematodes. *Genetics* 141:1007-1014.
- Bonferroni CE (1935) Il calcolo delle assicurazioni su gruppi di teste., Rome, Italy.
- Borsting C, Sanchez JJ, Morling N (2004) Multiplex PCR, amplicon size and hybridization efficiency on the NanoChip electronic microarray. *Int J Legal Med* 118:75-82.
- Borsting C, Sanchez JJ, Morling N (2005) SNP typing on the NanoChip electronic microarray. *Methods Mol Biol* 297:155-168.
- Borsting C, Sanchez JJ, Hansen HE, Hansen AJ, Bruun HQ, Morling N (2008) Performance of the SNPforID 52 SNP-plex assay in paternity testing. *Forensic Sci Int Genet* 2:292-300.
- Borsting C, Rockenbauer E, Morling N (2009) Validation of a single nucleotide polymorphism (SNP) typing assay with 49 SNPs for forensic genetic testing in a laboratory accredited according to the ISO 17025 standard. *Forensic Sci Int Genet* 4:34-42.
- Brandis JW (1999) Dye structure affects Taq DNA polymerase terminator selectivity. *Nucleic Acids Res* 27:1912-1918.
- Brandstatter A, Klein R, Duftner N, Wiegand P, Parson W (2006) Application of a quasi-median network analysis for the visualization of character conflicts to a population sample of mitochondrial DNA control region sequences from southern Germany (Ulm). *Int J Legal Med* 120:310-314.

- Bredel M, Bredel C, Juric D, Kim Y, Vogel H, Harsh GR, Recht LD, Pollack JR, Sikic BI (2005) Amplification of whole tumor genomes and gene-by-gene mapping of genomic aberrations from limited sources of fresh-frozen and paraffin-embedded DNA. *J Mol Diagn* 7:171-182.
- Brenner C, Morris, J.W. (1989) Paternity Index calculations in single locus hypervariable DNA probes: Validation and other studies. Paper presented at Proceedings for The International Symposium on Human Identification 1989, 1990.
- Brenner CH (1993) A note on paternity computation in cases lacking a mother. *Transfusion* 33:51-54.
- Brinkmann B (1998) Overview of PCR-based systems in identity testing. *Methods Mol Biol* 98:105-119.
- Brion M (2005) Y chromosome SNP analysis using the single-base extension: a hierarchical multiplex design. *Methods Mol Biol* 297:229-242.
- Brion M, Sanchez JJ, Balogh K, Thacker C, Blanco-Verea A, Borsting C, Stradmann-Bellinghausen B, Bogus M, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N (2005) Introduction of an single nucleotide polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages. *Electrophoresis* 26:4411-4420.
- Budowle B (1995) The effects of inbreeding on DNA profile frequency estimates using PCR-based loci. *Genetica* 96:21-25.
- Carracedo A, Bar W, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Budowle B, Brinkmann B, Gill P, Holland M, Tully G, Wilson M (2000) DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. *Forensic Sci Int* 110:79-85.
- Chakraborty R, Kidd KK (1991) The utility of DNA typing in forensic work. *Science* 254:1735-1739.
- Chen FC, Li WH (2001) Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am J Hum Genet* 68:444-456.
- Chen X, Livak KJ, Kwok PY (1998) A homogeneous, ligase-mediated DNA diagnostic test. *Genome Res* 8:549-556.
- Cheung VG, Nelson SF (1996) Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA. *Proc Natl Acad Sci U S A* 93:14676-14679.
- Consortium TIH (2005) A haplotype map of the human genome. *Nature* 437:1299-1320.
- Dausset J (1954) Présence des antigènes A et B dans le leucocytes, décléé par des épreuves d'agglutination. *C R Soc Biol* 148: 1607-1608.
- Dausset J, Cann H, Cohen D, Lathrop M, Lalouel JM, White R (1990) Centre d'etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* 6:575-577.
- Dean FB, Nelson JR, Giesler TL, Lasken RS (2001) Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res* 11:1095-1099.
- Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M, Lasken RS (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A* 99:5261-5266.
- Deka R, Shriver MD, Yu LM, Ferrell RE, Chakraborty R (1995) Intra- and inter-population diversity at short tandem repeat loci in diverse populations of the world. *Electrophoresis* 16:1659-1664.
- Di Giusto D, King GC (2003) Single base extension (SBE) with proofreading polymerases and phosphorothioate primers: improved fidelity in single-substrate assays. *Nucleic Acids Res* 31:e7.
- Dolan PL, Wu Y, Ista LK, Metzenberg RL, Nelson MA, Lopez GP (2001) Robust and efficient synthetic method for forming DNA microarrays. *Nucleic Acids Res* 29:E107-107.
- Erich HA, Higuchi R, Lichtenwalter K, Reynolds R, Sensabaugh G (1990) Reliability of the HLA-DQ alpha PCR-based oligonucleotide typing system. *J Forensic Sci* 35:1017-1019.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479-491.
- Excoffier L (2002) Human demographic history: refining the recent African origin model. *Curr Opin Genet Dev* 12:675-682.

- Fan JB, Chen X, Halushka MK, Berno A, Huang X, Ryder T, Lipshutz RJ, Lockhart DJ, Chakravarti A (2000) Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res* 10:853-860.
- Flanagan N, Healy E, Ray A, Phillips S, Todd C, Jackson IJ, Birch-Machin MA, Rees JL (2000) Pleiotropic effects of the melanocortin 1 receptor (MC1R) gene on human pigmentation. *Hum Mol Genet* 9:2531-2537.
- Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D (1991) Light-directed, spatially addressable parallel chemical synthesis. *Science* 251:767-773.
- Fondevila M, Phillips C, Naveran N, Fernandez L, Cerezo M, Salas A, Carracedo A, Lareu MV (2008) Case report: identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur. *Forensic Sci Int Genet* 2:212-218.
- Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851-861.
- Frudakis T, Thomas M, Gaskin Z, Venkateswarlu K, Chandra KS, Ginjupalli S, Gunturi S, Natrajan S, Ponnuswamy VK, Ponnuswamy KN (2003) Sequences associated with human iris pigmentation. *Genetics* 165:2071-2083.
- Gardner AF, Jack WE (2002) Acyclic and dideoxy terminator preferences denote divergent sugar recognition by archaeon and Taq DNA polymerases. *Nucleic Acids Res* 30:605-613.
- Gerry NP, Witowski NE, Day J, Hammer RP, Barany G, Barany F (1999) Universal DNA microarray method for multiplex detection of low abundance point mutations. *J Mol Biol* 292:251-262.
- Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, Sullivan K (1994) Identification of the remains of the Romanov family by DNA analysis. *Nat Genet* 6:130-135.
- Gill P (2001) An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med* 114:204-210.
- Gill P, Brenner C, Brinkmann B, Budowle B, Carracedo A, Jobling MA, de Knijff P, Kayser M, Krawczak M, Mayr WR, Morling N, Olaisen B, Pascali V, Prinz M, Roewer L, Schneider PM, Sajantila A, Tyler-Smith C (2001) DNA commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. *Int J Legal Med* 114:305-309.
- Goudet J, Perrin N, Waser P (2002) Tests for sex-biased dispersal using bi-parentally inherited genetic markers. *Mol Ecol* 11:1103-1114.
- Gunderson KL, Huang XC, Morris MS, Lipshutz RJ, Lockhart DJ, Chee MS (1998) Mutation detection by ligation to complete n-mer DNA arrays. *Genome Res* 8:1142-1153.
- Guo SW, Thompson EA (1992) A Monte Carlo method for combined segregation and linkage analysis. *Am J Hum Genet* 51:1111-1126.
- Guo Z, Guilfoyle RA, Thiel AJ, Wang R, Smith LM (1994) Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res* 22:5456-5465.
- Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsie L, Robbins CM, Brody LC, Wang D, Lander ES, Lipshutz R, Fodor SP, Collins FS (1999) Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. *Nat Genet* 22:164-167.
- Haff LA, Smirnov IP (1997) Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. *Genome Res* 7:378-388.
- Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet* 22:239-247.
- Hansen LL, Madsen BE, Pedersen K, Wiuf C (2007) Conflicting results in SNP genotype assessment. *Biotechniques* 43:756, 758, 760 passim.
- Harrison C, E. Musgrave-Brown, K. Bender, A. Carracedo, N. Morling, P. Schneider, D. Syndercombe-Court, The SNPforID Consortium (2006) A sensitive issue: Pyrosequencing as a valuable forensic SNP typing platform. In: Elsevier (ed) International Congress Series 1288, pp 52-54.

- Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum Mol Genet* 6:799-803.
- 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* 97:12164-12169.
- Holbrook JF, Stabley D, Sol-Church K (2005) Exploring whole genome amplification as a DNA recovery tool for molecular genetic studies. *J Biomol Tech* 16:125-133.
- Holland M, Parsons TJ (1999) Mitochondrial DNA Sequence Analysis—Validation and Use for Forensic Casework. *Forensic Science Review* 11:22-50.
- Hopkinson DA SN, Harris H (1963) Red cell acid phosphatase variants: a new human polymorphism. *Nature* 199:969-971.
- Hopkinson DA SN, Harris H (1964) Genetic studies on human red cell acid phosphatase. *Am J Hum Genet* 16:141-154.
- Hosono S, Faruqi AF, Dean FB, Du Y, Sun Z, Wu X, Du J, Kingsmore SF, Egholm M, Lasken RS (2003) Unbiased whole-genome amplification directly from clinical samples. *Genome Res* 13:954-964.
- Jain M, Thorstenson YR, Faulkner DM, Pourmand N, Jones T, Au M, Oefner PJ, White KP, Davis RW (2003) Genotyping African haplotypes in ATM using a co-spotted single-base extension assay. *Hum Mutat* 22:214-221.
- Jansen RP (2000) Origin and persistence of the mitochondrial genome. *Hum Reprod* 15 Suppl 2:1-10.
- Jeffreys AJ, Wilson V, Wong Z, Royle N, Patel I, Kelly R, Clarkson R (1987) Highly variable minisatellites and DNA fingerprints. *Biochem Soc Symp* 53:165-180.
- Jobling MA, Heyer E, Dieltjes P, de Knijff P (1999) Y-chromosome-specific microsatellite mutation rates re-examined using a minisatellite, MSY1. *Hum Mol Genet* 8:2117-2120.
- Johnston M (1998) Gene chips: array of hope for understanding gene regulation. *Curr Biol* 8:R171-174.
- Jones DA (1972) Blood samples: probability of discrimination. *J Forensic Sci Soc* 12:355-359.
- Kaderali L, Deshpande A, Nolan JP, White PS (2003) Primer-design for multiplexed genotyping. *Nucleic Acids Res* 31:1796-1802.
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18:830-838.
- Kayser M, Schneider PM (2009) DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. *Forensic Sci Int Genet* 3:154-161.
- Khrapko KR, Khorlin AA, Ivanov IB, Chernov BK, Lysov Iu P, Vasilenko SK, Florent'ev VL, Mirzabekov AD (1991) [Hybridization of DNA with oligonucleotides immobilized in a gel: a convenient method for recording single base replacements]. *Mol Biol (Mosk)* 25:718-730.
- Kidd KK, Pakstis AJ, Speed WC, Grigorenko EL, Kajuna SL, Karoma NJ, Kungulilo S, Kim JJ, Lu RB, Odunsi A, Okonofua F, Parnas J, Schulz LO, Zhukova OV, Kidd JR (2006) Developing a SNP panel for forensic identification of individuals. *Forensic Sci Int* 164:20-32.
- Korch C, Drabkin H (1999a) Improved DNA sequencing accuracy and detection of heterozygous alleles using manganese citrate and different fluorescent dye terminators. *Genome Res* 9:588-595.
- Korch C, Drabkin H (1999b) Manganese citrate improves base-calling accuracy in DNA sequencing reactions using rhodamine-based fluorescent dye-terminators. *Nucleic Acids Res* 27:1405-1407.
- Krjutskov K, Andreson R, Magi 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 Res* 36:e75.
- Krjutskov 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 Sci Int Genet* 4:43-48.
- Kwok PY (2001) Methods for genotyping single nucleotide polymorphisms. *Annu Rev Genomics Hum Genet* 2:235-258.

- Landegren U, Kaiser R, Sanders J, Hood L (1988) A ligase-mediated gene detection technique. *Science* 241:1077-1080.
- Landsteiner (1901) Ueber Agglutinationserscheinungen normalen menschlichen Blutes. *Wien Klin Wchnschr.* Vol. 14, Wien, pp 1132 – 1134.
- Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A, Balascakova M, et al. (2008) Correlation between genetic and geographic structure in Europe. *Curr Biol* 18:1241-1248.
- Lasken RS, Egholm M (2003) Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens. *Trends Biotechnol* 21:531-535.
- Leren TP, Rodningen OK, Rosby O, Solberg K, Berg K (1993) Screening for point mutations by semi-automated DNA sequencing using sequenase and magnetic beads. *Biotechniques* 14:618-623.
- Lewis PO, Zaykin D (2002) Genetic data analysis: Computer program for the analysis of allelic data, v. 1.1.
- Lewontin RC, Hartl DL (1991) Population genetics in forensic DNA typing. *Science* 254:1745-1750.
- Li N, Stephens M (2003) Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 165:2213-2233.
- Lightowlers RN, Chinnery PF, Turnbull DM, Howell N (1997) Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. *Trends Genet* 13:450-455.
- Liljedahl U, Fredriksson M, Dahlgren A, Syvanen AC (2004) Detecting imbalanced expression of SNP alleles by minisequencing on microarrays. *BMC Biotechnol* 4:24.
- Lindroos K, Liljedahl U, Raitio M, Syvanen AC (2001) Minisequencing on oligonucleotide microarrays: comparison of immobilisation chemistries. *Nucleic Acids Res* 29:E69-69.
- Lindroos K, Sigurdsson S, Johansson K, Ronnblom L, Syvanen AC (2002) Multiplex SNP genotyping in pooled DNA samples by a four-colour microarray system. *Nucleic Acids Res* 30:e70.
- Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ (1999) High density synthetic oligonucleotide arrays. *Nat Genet* 21:20-24.
- Liu WM, Di X, Yang G, Matsuzaki H, Huang J, Mei R, Ryder TB, Webster TA, Dong S, Liu G, Jones KW, Kennedy GC, Kulp D (2003) Algorithms for large-scale genotyping microarrays. *Bioinformatics* 19:2397-2403.
- Lovmar L, Fredriksson M, Liljedahl U, Sigurdsson S, Syvanen AC (2003) Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA. *Nucleic Acids Res* 31:e129.
- Lovmar L, Ahlford A, Jonsson M, Syvanen AC (2005) Silhouette scores for assessment of SNP genotype clusters. *BMC Genomics* 6:35.
- Lovmar L, Syvanen AC (2005) Genotyping single-nucleotide polymorphisms by minisequencing using tag arrays. *Methods Mol Med* 114:79-92.
- Mack M, Bender K, Schneider PM (2004) Detection of retroviral antisense transcripts and promoter activity of the HERV-K(C4) insertion in the MHC class III region. *Immunogenetics* 56:321-332.
- Macmillan C, Lach B, Shoubbridge EA (1993) Variable distribution of mutant mitochondrial DNAs (tRNA(Leu[3243])) in tissues of symptomatic relatives with MELAS: the role of mitotic segregation. *Neurology* 43:1586-1590.
- Maier E, Meier-Ewert S, Ahmadi AR, Curtis J, Lehrach H (1994) Application of robotic technology to automated sequence fingerprint analysis by oligonucleotide hybridisation. *J Biotechnol* 35:191-203.
- Mengel-Jorgensen J, Sanchez JJ, Borsting C, Kirpekar F, Morling N (2004) MALDI-TOF mass spectrometric detection of multiplex single base extended primers. A study of 17 y-chromosome single-nucleotide polymorphisms. *Anal Chem* 76:6039-6045.
- Mengel-Jorgensen J, Sanchez JJ, Borsting C, Kirpekar F, Morling N (2005) Typing of multiple single-nucleotide polymorphisms using ribonuclease cleavage of DNA/RNA chimeric single-base extension primers and detection by MALDI-TOF mass spectrometry. *Anal Chem* 77:5229-5235.
- Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215.
- Mir KU, Southern EM (1999) Determining the influence of structure on hybridization using oligonucleotide arrays. *Nat Biotechnol* 17:788-792.

- Morton NE (1995) DNA forensic science 1995. *Eur J Hum Genet* 3:139-144.
- Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 51 Pt 1:263-273.
- Musgrave-Brown E, N. Anwar, K. Elliott, C. Phillips, D. Syndercombe Court, A. Carracedo, N. Morling, P. Schneider, B. McKeown (2006) Mixture analysis using SWaP SNPs and non-biallelic SNPs. In: Elsevier (ed) *International Congress Series* 1288, pp 34-36.
- Musgrave-Brown E, Ballard D, Balogh K, Bender K, Berger B, Bogus M, Borsting C, Brion M, Fondevila M, Harrison C, Oguzturun C, Parson W, Phillips C, Proff C, Ramos-Luis E, Sanchez JJ, Sanchez Diz P, Sobrino Rey B, Stradmann-Bellinghausen B, Thacker C, Carracedo A, Morling N, Scheithauer R, Schneider PM, Syndercombe Court D (2007) Forensic validation of the SNPforID 52-plex assay. *Forensic Sci Int Genet* 1:186-190.
- Musgrave-Brown E, Ballard D, Fondevila Álvarez M, Fang R, Harrison C, Phillips C, Prasad Y, Sobrino Rey B, Thacker C, Wiluhn J, Carracedo A, Schneider PM, Syndercombe Court D and The SNPforID Consortium (2008) Forensic validation of the Genplex SNP typing system—Results of an inter-laboratory study. *Forensic Science International: Genetics Supplement Series* Vol 1, Is 1: 389-393.
- Nachman MW, Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics* 156:297-304.
- Nur I, Elkaim R, Herzberg M (1990) HybriComb. A novel diagnostic tool for DNA probing. *Ann Biol Clin (Paris)* 48:557-559.
- Nyren P (2007) The history of pyrosequencing. *Methods Mol Biol* 373:1-14.
- Pakstis AJ, Speed WC, Kidd JR, Kidd KK (2007) Candidate SNPs for a universal individual identification panel. *Hum Genet* 121:305-317.
- Pakstis AJ, Speed WC, Fang R, Hyland FC, Furtado MR, Kidd JR, Kidd KK (2010) SNPs for a universal individual identification panel. *Hum Genet* 127:315-324.
- Pan X, Urban AE, Palejev D, Schulz V, Grubert F, Hu Y, Snyder M, Weissman SM (2008) A procedure for highly specific, sensitive, and unbiased whole-genome amplification. *Proc Natl Acad Sci USA* 105:15499-15504.
- Parker LT, Deng Q, Zakeri H, Carlson C, Nickerson DA, Kwok PY (1995) Peak height variations in automated sequencing of PCR products using Taq dye-terminator chemistry. *Biotechniques* 19:116-121.
- Parker LT, Zakeri H, Deng Q, Spurgeon S, Kwok PY, Nickerson DA (1996) AmpliTaq DNA polymerase, FS dye-terminator sequencing: analysis of peak height patterns. *Biotechniques* 21:694-699.
- Pastinen T, Partanen J, Syvanen AC (1996) Multiplex, fluorescent, solid-phase minisequencing for efficient screening of DNA sequence variation. *Clin Chem* 42:1391-1397.
- 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* 7:606-614.
- Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, Syvanen AC (2000) A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res* 10:1031-1042.
- Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP (1994) Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc Natl Acad Sci U S A* 91:5022-5026.
- Phillips C, Lareu V, Salas A, Carracedo A (2004a) Nonbinary single-nucleotide polymorphism markers. *International Congress Series* 1261 27– 29. Vol. 1261 27– 29.
- Phillips C, Lareu M, Salas A, Fondevila M, Berniell Lee G, Carracedo A, Morling N, Schneider P, Syndercombe Court D (2004b) Population specific single nucleotide polymorphisms. *International Congress Series* 1261 233–235.
- Phillips C, Lareu M, Sanchez JJ, Brion M, Sobrino B, Morling N, Schneider PM, Syndercombe-Court, D CA (2004c) Selecting single nucleotide polymorphisms for forensic applications. *International Congress Series* 1261: 18-20.

- Phillips C, Fang R, Ballard D, Fondevila M, Harrison C, Hyland F, Musgrave-Brown E, Proff C, Ramos-Luis E, Sobrino B, Carracedo A, Furtado MR, Syndercombe Court D, Schneider PM (2007a) Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel. *Forensic Sci Int Genet* 1:180-185.
- Phillips C, Salas A, Sanchez JJ, Fondevila M, Gomez-Tato A, Alvarez-Dios J, Calaza M, de Cal MC, Ballard D, Lareu MV, Carracedo A (2007b) Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Sci Int Genet* 1:273-280.
- Prak ET, Kazazian HH, Jr. (2000) Mobile elements and the human genome. *Nat Rev Genet* 1:134-144.
- Prinz M, Boll K, Baum H, Shaler B (1997) Multiplexing of Y chromosome specific STRs and performance for mixed samples. *Forensic Sci Int* 85:209-218.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Queller DC (1993) Worker control of sex ratios and selection for extreme multiple mating by queens. *Am Nat* 142:346-351.
- Rappold GA (1993) The pseudoautosomal regions of the human sex chromosomes. *Hum Genet* 92:315-324.
- Roewer L (2001) Die Haplotypisierung des Y-Chromosoms - Grundlagen und Anwendungen einer neuen molekulargenetischen Identifizierungsmethode. Habilitationsschrift, 2001.
- Roewer L, Epplen JT (1992) Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work. *Forensic Sci Int* 53:163-171.
- Rogers AR, Wooding S, Huff CD, Batzer MA, Jorde LB (2007) Ancestral alleles and population origins: inferences depend on mutation rate. *Mol Biol Evol* 24:990-997.
- Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. *Science* 298:2381-2385.
- Rosenberg NA, Li LM, Ward R, Pritchard JK (2003) Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet* 73:1402-1422.
- Rousseeuw P (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20:53 – 65.
- Rousset F, Raymond M (1995) Testing heterozygote excess and deficiency. *Genetics* 140:1413-1419.
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928-933.
- Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350-1354.
- 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* 27:1713-1724.
- 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, Schmitter H, Schneider PM, Vallone PM, Morling N (2008) Forensic typing of autosomal SNPs with a 29 SNP-multiplex--results of a collaborative EDNAP exercise. *Forensic Sci Int Genet* 2:176-183.
- Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467-470.
- Schena M, Heller RA, Thériault TP, Konrad K, Lachenmeier E, Davis RW (1998) Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol* 16:301-306.
- Schneider PM, Meuser S, Waiyawuth W, Seo Y, Rittner C (1998) Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. *Forensic Sci Int* 97:61-70.

- Schneider PM, K. Balogh, N. Naveran, M. Bogus, K. Bender, M. Lareu, A. Carracedo (2004) Whole genome amplification—the solution for a common problem in forensic casework? Elsevier, International Congress Series 1261 (2004) 24–26.
- Schneider PM (2008) Populationsgenetik autosomaler Polymorphismen. medgen, Springer Medizin Verlag 20:298–301.
- Schneider S, D. Roessli and L.Excoffier (2000) Arlequin: a software for population genetics data analysis. User manual ver 2.000.
- Seo Y, Stradmann-Bellinghausen B, Rittner C, Takahama K, Schneider PM (1998) Sequence polymorphism of mitochondrial DNA control region in Japanese. Forensic Sci Int 97:155-164.
- Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, Pfaff C, Jones C, Massac A, Cameron N, Baron A, Jackson T, Argyropoulos G, Jin L, Hoggart CJ, McKeigue PM, Kittles RA (2003) Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet 112:387-399.
- Shumaker JM, Metspalu A, Caskey CT (1996) Mutation detection by solid phase primer extension. Hum Mutat 7:346-354.
- Simsek M, Tanira MO, Al-Baloushi KA, Al-Barwani HS, Lawatia KM, Bayoumi RA (2001) A precaution in the detection of heterozygotes by sequencing: comparison of automated DNA sequencing and PCR-restriction fragment length polymorphism methods. Clin Chem 47:134-137.
- Smit AF, Toth G, Riggs AD, Jurka J (1995) Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. J Mol Biol 246:401-417.
- Smit AF, Riggs AD (1996) Tiggers and DNA transposon fossils in the human genome. Proc Natl Acad Sci U S A 93:1443-1448.
- Smithies O (1955) Zone electrophoresis in starch gels: group variations in the serum proteins of normal human adults. The Biochemical Journal. 61, 629-641.
- Sobrinho B, Lareu M, Brion M, Carracedo A (2004) SNP genotyping with single base extension-tag microarrays. International Congress Series 1261, 331– 333.
- Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503-517.
- Spencer N, Hopkinson, D. A., Harris, H. (1964) Phosphoglucomutase polymorphism in man. Nature 204:742.
- Spiess AN, Mueller N, Ivell R (2004) Trehalose is a potent PCR enhancer: lowering of DNA melting temperature and thermal stabilization of taq polymerase by the disaccharide trehalose. Clin Chem 50:1256-1259.
- Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am J Hum Genet 48:370-382.
- Stoneking M (2001) Single nucleotide polymorphisms. From the evolutionary past. Nature 409:821-822.
- Strachan TR, Andrew P. (2005) Molekulare Humangenetik. Vol 3. Auflage. Spektrum Akademischer Verlag.
- Swofford DL, and R. B. Selander (1989) Biosys-1: a computer program for analysis of allelic variation in population genetics and biochemical statistics. Illinois Natural History Survey, Champaign, IL.
- Syvanen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet 2:930-942.
- Tabor S, Richardson CC (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc Natl Acad Sci U S A 84:4767-4771.
- Tabor S, Richardson CC (1989) Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and Escherichia coli DNA polymerase I. Proc Natl Acad Sci U S A 86:4076-4080.
- Tabor S, Richardson CC (1990) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Effect of pyrophosphorolysis and metal ions. J Biol Chem 265:8322-8328.
- Tabor S, Richardson CC (1995) A single residue in DNA polymerases of the Escherichia coli DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. Proc Natl Acad Sci U S A 92:6339-6343.

- Takahata N, Lee SH, Satta Y (2001) Testing multiregionality of modern human origins. *Mol Biol Evol* 18:172-183.
- Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res* 17:6463-6471.
- Tautz D, Renz M (1984a) Simple DNA sequences of *Drosophila virilis* isolated by screening with RNA. *J Mol Biol* 172:229-235.
- Tautz D, Renz M (1984b) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res* 12:4127-4138.
- Thacker CR, M.K. Balogh, C. Børsting, E. Ramos, P. Sanchez Diz, A. Carracedo, N. Morling, P. Schneider, D. Syndercombe Court, the SNPforID Consortium (2006) The effect of whole genome amplification on samples originating from more than one donor. Elsevier, International Congress Series 1288 (2006) 722–724.
- Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW (2000) Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc Natl Acad Sci U S A* 97:7360-7365.
- Tzvetkov MV, Becker C, Kulle B, Nurnberg P, Brockmoller J, Wojnowski L (2005) Genome-wide single-nucleotide polymorphism arrays demonstrate high fidelity of multiple displacement-based whole-genome amplification. *Electrophoresis* 26:710-715.
- Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 7:996-1005.
- Vallone PM, Just RS, Coble MD, Butler JM, Parsons TJ (2004) A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. *Int J Legal Med* 118:147-157.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, et al. (2001) The sequence of the human genome. *Science* 291:1304-1351.
- Vergnaud G, Mariat D, Apiou F, Aurias A, Lathrop M, Lauthier V (1991) The use of synthetic tandem repeats to isolate new VNTR loci: cloning of a human hypermutable sequence. *Genomics* 11:135-144.
- Wächter A, Mengel-From J, Børsting, Morling N. (2008) A 50 SNP-multiplex mass spectrometry assay for human identification. *Forensic Science International: Genetics Supplement Series Vol 1, Is 1, Pages 487-489.*
- Waiyawuth W, Zhang L, Rittner C, Schneider PM (1998) Genetic analysis of the short tandem repeat system D12S391 in the German and three Asian populations. *Forensic Sci Int* 94:25-31.
- Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T, Itakura K (1979) Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Res* 6:3543-3557.
- Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, et al. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280:1077-1082.
- Watterson GA, Guess HA (1977) Is the most frequent allele the oldest? *Theor Popul Biol* 11:141-160.
- Weir BS (1979) Inferences about linkage disequilibrium. *Biometrics* 35:235-254.
- Wright S (1951) The genetical structure of populations. *Annals of Eugenics*:323-354.
- Yu N, Chen FC, Ota S, Jorde LB, Pamilo P, Patthy L, Ramsay M, Jenkins T, Shyue SK, Li WH (2002) Larger genetic differences within africans than between Africans and Eurasians. *Genetics* 161:269-274.
- Zakeri H, Amparo G, Chen SM, Spurgeon S, Kwok PY (1998) Peak height pattern in dichlororhodamine and energy transfer dye terminator sequencing. *Biotechniques* 25:406-410, 412-404.
- Zhang L, Stradmann-Bellinghausen B, Schneider PM, Rittner C (1993) Genetic polymorphisms of the A and B subunits of coagulation factor XIII in the Chinese population. *Exp Clin Immunogenet* 10:137-140.

ANHANG

I. Veröffentlichung

Electrophoresis 2006, 27, 1713–1724

1713

Juan J. Sanchez¹
Chris Phillips²
Claus Børsting¹
Kinga Balogh³
Magdalena Bogus³
Manuel Fondevila²
Cheryl D. Harrison⁴
Esther Musgrave-Brown⁴
Antonio Salas²
Denise Syndercombe-Court⁴
Peter M. Schneider³
Angel Carracedo²
Niels Morling¹

¹Department of Forensic Genetics,
Institute of Forensic Medicine,
University of Copenhagen,
Copenhagen, Denmark

²Institute of Legal Medicine,
University of Santiago de
Compostela,
Santiago de Compostela, Spain

³Institute of Legal Medicine,
Johannes Gutenberg University,
Mainz, Germany

⁴Centre for Haematology,
ICMS, Barts and The London,
Queen Mary's School of Medicine
and Dentistry,
London, UK

Research Article

A multiplex assay with 52 single nucleotide polymorphisms for human identification

A total of 52 SNPs reported to be polymorphic in European, Asian and African populations were selected. Of these, 42 were from the distal regions of each autosome (except chromosome 19). Nearly all selected SNPs were located at least 100 kb distant from known genes and commonly used STRs. We established a highly sensitive and reproducible SNP-typing method with amplification of all 52 DNA fragments in one PCR reaction followed by detection of the SNPs with two single base extension reactions analysed using CE. The amplicons ranged from 59 to 115 bp in length. Complete SNP profiles were obtained from 500 pg DNA. The 52 loci were efficiently amplified from degraded samples where previously only partial STR profiles had been obtained. A total of 700 individuals from Denmark, Greenland, Somalia, Turkey, China, Germany, Taiwan, Thailand and Japan were typed, and the allele frequencies estimated. All 52 SNPs were polymorphic in the three major population groups. The mean match probability was at least 5.0×10^{-19} in the populations studied. Typical paternity indices ranged from 336 000 in Asians to 549 000 in Europeans. Details of the 52 SNP loci and population data generated in this work are freely available at <http://www.snpsforid.org>.

Keywords: Autosomes / Human identification / Multiplex PCR / Single base extension / Single nucleotide polymorphism
DOI 10.1002/elps.200500671

Received September 10, 2005

Revised October 15, 2005

Accepted October 16, 2005



1 Introduction

SNPs have a number of characteristics that make them ideal markers for human identification. First, they have lower mutation rates than the STR and VNTR (variable number tandem repeat) loci typically used for relationship analysis in paternity and immigration testing. Second, SNPs can be analysed after PCR amplification of very short DNA-regions surrounding the substitution site, making SNPs preferable for anthropological and crime case investigations where the DNA is often degraded. Third, SNPs can be genotyped with a growing range of high-throughput tech-

nologies; an important factor in the implementation of large criminal DNA databases [1, 2]. Finally, SNPs, as binary polymorphisms, are comparatively easy to validate, because precise allele frequency estimates, required for the accurate interpretation of forensic genotyping data, can be obtained by analysing fewer samples compared to those needed for allele frequencies estimates of STRs and VNTRs. Seeking to match the discriminatory power of the 10–15 multiple allele STRs routinely used in forensic investigations, a set of about 50 polymorphic SNP markers are predicted to be required [3, 4]. Furthermore, it has been suggested that 50 unlinked SNP loci with high overall heterozygosity should be sufficient to adjust for population stratification in population-based associations studies [5]. SNPs that are polymorphic in one population may be almost or completely monomorphic in another population [6, 7], while others are known to be polymorphic in all major population groups. Thus, it should be possible to select SNPs that are useful for human identification purposes in the majority of populations, and to supplement these with SNPs

Correspondence: Dr. Juan J. Sanchez, Department of Forensic Genetics, Institute of Forensic Medicine, University of Copenhagen, 11 Frederik V's Vej, DK-2100 Copenhagen, Denmark
E-mail: juan.sanchez@forensic.ku.dk
Fax: +45-35-32-61-20

Abbreviations: RFU, relative fluorescence unit; SBE, single base extension

showing highly contrasting allele frequency distributions in particular populations. These latter SNPs can provide valuable information for population admixture detection, in addition to the estimation of biogeographical ancestry.

The SNPforID group (<http://www.snpforid.org>) is a consortium supported by the EU GROWTH programme with the following objectives: (i) selection of at least 50 autosomal SNPs suitable for the identification of persons of unknown population origin and determination of allele frequencies in the major population groups; (ii) development of a highly efficient DNA amplification strategy for the simultaneous analysis of up to 50 independent SNPs in a single assay; (iii) assessment of automated, high-throughput DNA-typing platforms for reliable and accurate multiplex SNP typing; (iv) assessment of the forensic application of the high-throughput SNP-typing methods developed.

If SNP typing is to be a realistic alternative to STR typing in forensic analyses and other fields of investigation where the target material is scarce and often of poor quality, the relevant SNP loci must be amplified efficiently in a single multiplexed PCR reaction and preferably analysed with a method that is well established and robust in routine use.

In this work we present five stages in the development of an SNP-based human identification genotyping assay: (i) a set of 52 unlinked autosomal SNPs that are highly polymorphic in European, Asian and African populations; (ii) a multiplex PCR amplification strategy that allows the simultaneous amplification of 52 fragments in a single reaction from as little as 500 pg DNA; (iii) an SNP detection system based on two sets of multiplexed single base extension (SBE) reactions with 23 and 29 SBE primers, respectively, that can be analysed in one CE run and automated computer based allele detection; (iv) validation of assay reproducibility, sensitivity and robustness; (v) determination of allele frequency distributions of each SNP in 700 individuals from 9 European, Asian and African populations as well as 9 animal species, including 46 samples from 6 primate species.

2 Materials and methods

2.1 Samples and DNA purification

A total of 700 samples (numbers in parentheses) from Denmark (156), Greenland (149), Somalia (104), Turkey (96), China (63), Germany (49), Taiwan (43), Thailand (33) and Japan (7) were typed in duplicate. In addition, samples from unrelated chimpanzees (29), gorillas (3), orangutan (1), baboons (2), rhesus macaques (5) and Cynomolgus monkeys (6) were tested together with dog (1), cat (1) and horses (2).

We used blood on FTA® cards (Whatman) and DNA purified by phenol/chloroform extraction or the QIAamp DNA blood mini kit (Qiagen). DNA concentrations were determined by real-time PCR using the Quantifiler™ Human DNA Quantification Kit (Applied Biosystems) with the ABI 7300 real-time PCR system (Applied Biosystems) or by using SYBR Green I (Roche) with the LightCycler system (Roche). The performance of the 52 SNP-plex assay was also tested on DNA purified from paraffin-embedded tissues and DNA from bones and muscle tissue samples taken from seven human cadavers found under various environmental conditions in crime cases.

All protocols were approved by the Danish ethical committee (KF-01-037/03).

2.2 Criteria for selection of SNP loci

The following SNP selection criteria were used when choosing suitable candidate loci: (i) the size of the amplicon generated from optimum primer designs less than 120 bp; (ii) reported minimum 30% heterozygosity (0.28 minor allele frequency) in at least one population, and minimum 20% heterozygosity (0.17 minor allele frequency) in all three populations; (iii) a freely assorting marker set using SNPs from the distal parts of the *p* and *q* arms of each autosome; (iv) a minimum distance of 100 kb between candidate SNPs and neighbouring genes; (v) no likely association with the STR loci most commonly used in forensic analysis; and (vi) flanking DNA sequence reliably reported and free from interfering polymorphisms, such as nucleotide substitutions in potential primer binding sites.

2.3 Marker selection *in silico*

Suitable regions were chosen for scrutiny using NCBI Map Viewer (<http://www.ncbi.nlm.nih.gov/mapview/maps.cgi/>) with alignments of the gene and variation maps following the general guidelines described by Phillips [8] (2004). Only loci marked on the gene map as 'confirmed gene models based on mRNA alignments' were used to define gene-free regions. In addition, SNP Browser™ (<http://www.allsnps.com/snpbrowser>) and HapMap (<http://www.hapmap.org/>) genome browsers were used as these became publicly available during the search phase. Three searches were performed using dbSNP builds 112 (*p*-arm loci), 115 (*q*-arm loci) and 118 (supplementary loci) corresponding to genome builds 28, 33 and 34, respectively. From these searches, sets of 46, 67 and 25 SNPs were selected giving a median two loci *per p*-arm and four loci *per q*-arm from each autosome. We primarily selected SNPs genotyped by The SNP Consortium, although these

were supplemented by a small number of SNPs validated by the Perlegen and HapMap genotyping initiatives. SNPs from The SNP Consortium had been validated in most cases using genotyped individuals from the three major populations with an average sample size of 38 from each group. Approximately 60% of SNPs scrutinized during the selection process had insufficient variability in one or more population groups using the comparative allele frequency criterion previously described [8] (see also Section 2.2). In the case of the third SNP selection of 25 loci, several markers with limited variability in one of the three major population groups were included to add predictive power for population of origin. The final screening of SNPs before assimilation into candidate pools involved examination of the flanking sequence to ensure that the region available for primer design (approximately 100 bp on each side of the SNP) was free from clustering SNPs and low-complexity sequence. In several cases, a single clustering SNP was permitted and subsequently circumvented during primer design. Simple sequence quality checks were performed using the Fasta sequence report in dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>). Candidate SNPs with a proportion of low-complexity flanking sequence greater than 20% were rejected to ensure primer specificity before starting the multiplex design process. Table S1 lists the selected SNPs and the flanking sequences.

2.4 Selection of PCR amplification primers

PCR primers were designed to give amplicon lengths in the range from 59 to 115 bps (Table S1). The aim was to obtain a theoretical melting temperature of $60 \pm 2^\circ\text{C}$ at a salt concentration of 180 mM and a purine:pyrimidine content close to 1:1. All primer candidates were analysed for primer-dimer formation, hairpin structures, homology and complementarity to other primers in the multiplex using Primer 3.2 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). Primer characteristics were chosen to ensure equal PCR amplification efficiency for all DNA fragments as previously described [9]. The primers were HPLC-purified and checked for homogeneity by MALDI-TOF MS (DNA Technology A/S, Denmark). Table S1 shows the sequences and the concentrations of the amplification primers in the final multiplex PCR.

2.5 PCR conditions and purification of PCR products

PCR amplification conditions were: 1–10 ng DNA in a 25 μL reaction volume containing $1 \times$ PCR buffer, 8 mM MgCl_2 , 700 μM of each dNTP, 0.01–0.17 μM of each primer and 2 U AmpliTaq Gold DNA polymerase (Applied Biosystems). For the PCR inhibition study, haematin

(Sigma-Aldrich) was diluted to 15 mM in 0.1 N NaOH and added to the mixture. Cycling was performed in a GeneAmp 9600 (Perkin Elmer) or Eppendorf Mastercycler gradient (Eppendorf) thermal cycler with the following cycle programme: denaturation at 94°C for 5 min followed by 35 cycles of 95°C for 30 s, 60°C for 30 s and 65°C for 30 s, followed by 7 min at 65°C .

Excess primers and dNTPs were removed by using Min Elute PCR purification spin columns (Qiagen). The PCR products were eluted in 20 μL of Milli-Q water. Alternatively, 1 μL ExoSAP-IT kit (Amersham Pharmacia Biotech) or 0.75 μL (1 U/ μL) shrimp alkaline phosphatase (Amersham Pharmacia Biotech) and 0.023 μL (10 U/ μL) Exonuclease I (Amersham Pharmacia Biotech) were added to 2.5 μL PCR product and incubated at 37°C for 15 min, 80°C for 15 min or 37°C for 1 h, and 75°C for 15 min, respectively.

2.6 Design of SBE primers

Two SBE multiplexes with 23 and 29 SBE primers, respectively, were developed according to the guidelines described by Sanchez *et al.* [9] (2004). The lengths of the SBE primers were between 16 and 92 nucleotides. Size intervals of five to six nucleotides were used for primers shorter than 35 nucleotides and size intervals of four nucleotides were used for primers longer than 35 nucleotides (Table S2). The primers were checked for primer-dimer formation using the AutoDimer program (<http://www.cstl.nist.gov/biotech/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm>). The lengths of the SBE primers were increased with tails of nonhuman sequence and/or poly-dNTP stretches (see Table S2). Opposite allele combinations (e.g. A/G with C/T) were analysed in the same size interval whenever possible.

2.7 SBE reaction and purification of the SBE products

SBE reactions were performed in 8 μL with 1 μL of purified PCR product, 4 μL of SNaPshot reaction mix (Applied Biosystems), 1 μL of SBE primer mix (0.01–0.27 μM , Table S2) and 2 μL of Milli-Q water. The SBE primer mix was diluted in 160 mM ammonium sulfate (Sigma-Aldrich) to minimize primer-dimer artefacts. The SBE reaction was performed in a GeneAmp 9600 or Eppendorf Mastercycler gradient instruments with 30 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 30 s. Excess nucleotides were removed by addition of 1 μL (1 U/ μL) shrimp alkaline phosphatase to the SBE mix and incubation at 37°C for 45 min followed by 75°C for 15 min.

2.8 Detection and analysis of the SBE products

Two microlitres of SBE product was mixed with 20 μ L of Hi-Di formamide (Applied Biosystems) and analyzed by CE using ABI Prism 310, ABI Prism 3100 or ABI Prism 3100-Avant Genetic Analyzers (Applied Biosystems) with 36 cm capillary arrays and POP-4 polymer (Applied Biosystems). The following modifications were made to the SNP36-POP-4 default module of the GeneScan analysis 3.7 software: Three extra washing steps were added (i) after injection of the polymer, (ii) after the prerun and (iii) after the injection of the samples using the spare water reservoir, SIT Water2. The time that the capillary tips were immersed in the wash buffers was increased to 60 s to minimize carry-over. The run time was decreased to 300 s to allow sufficient spacing between the two injections. Another GeneScan method file was created to allow injection of the second SBE multiplex without a capillary polymer fill stage by removing the steps that fill the syringe and the capillary array along with the prerun. The run time of the second injection was 1000 s. The two GeneScan methods were run sequentially by saving the methods with number sequence default module names in the ABI 3100 Data Collection 1.1 software using the Module Editor. We used 22 s injections at 3000 and 2000 V in the first and second injections, respectively. Analysis was made using GeneScan Analysis 3.7 with peak thresholds set to a minimum of 120 relative fluorescence units (RFUs) (blue color), 60 RFUs (green color) and 30 RFUs (yellow, red and orange color). GeneScan-120 Liz internal size standard (Applied Biosystems) was used in both injections but the reference sizes were modified in the second injection to the standard size plus 200 (Fig. 1).

Automated allele calls were made using Genotyper 3.7 (Applied Biosystems) macros. The macros used are available on the SNPforID website (www.snpforid.org). All peaks in the size standard had to be detected and the peak height of the largest peak in the electropherogram had to be a minimum of

1000 RFUs before analysis could proceed. Peaks detected in predefined allele windows with peak heights larger than 10% (blue color), 7% (green color) or 5% (yellow and red colour) of the maximum peak height in the respective color were labelled with allele names, peak heights and sizes. The predefined windows were determined from prior analyses of 96 samples. The widths of the predefined windows varied from 1 to 2 bp depending on the lengths of the extended SBE primers. The ratios of the fluorophore emissions of the dR110 (blue), dR6G (green), TAMRA™ (yellow) and dROX™ (red) dyes were approximately 4:2:1:1. A few windows had slightly different ratios due to locus-specific variations and the peak heights were normalized accordingly. A maximum peak height ratio of 3:1 after normalization was accepted as a heterozygote. A minimum peak height of 400 RFUs (blue), 200 RFUs (green) or 100 RFUs (yellow and red) and a peak height ratio of minimum 5:1 after normalization were accepted as a homozygote.

2.9 Reproducibility and sensitivity

Validation of the 52 SNP-plex assay was, in principle, conducted according to the revised guidelines of DNA analysis of the 'Scientific Working Group on DNA Analysis Methods' (SWGDM, <http://www.fbi.gov/hq/lab/fsc/backissu/july2004/index.htm>). Multiplex PCR performance was assessed by analyses of dilution series of genomic DNA (0.07, 0.14, 0.27, 0.55, 1.09, 2.19, 4.37, 8.75, 17.50, 35 and 70 ng) from two individuals. The RFUs of each of the four dyes were collated and normalised by dividing homozygote allele values by two.

2.10 Population studies

SNP characterization including allele frequencies, Hardy-Weinberg equilibrium and linkage disequilibrium tests was carried out using the SNP Assistant Program v. 1.0.9.

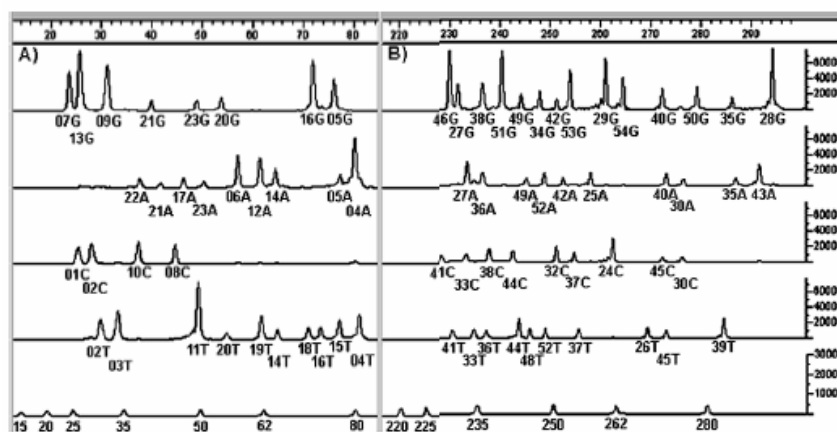


Figure 1. Analysis of the 52 SNP-plex assay. (A) Results of the first injection (23 SNPs). (B) Results of the second injection (29 SNPs).

The Genetic Data Analysis software (<http://hybridictyon.eeb.uconn.edu/people/plewis/software.php>) was used to estimate F_{ST} values [10] and exact tests for associations between alleles at different loci. The software package Arlequin v. 2.000 (<http://anthro.unige.ch/arlequin>) was used to determine the molecular variance. Pearson coefficients (r) were calculated using Excel spreadsheets. Cluster analysis was performed by assigning individuals to a predefined number (K) of 'inferred populations' according to their genotypes using the Bayesian approach implemented in Structure v. 2.0 [11]. A model of admixture was used that assumes individuals may have mixed ancestry. All Structure v. 2.0 runs used 200 000 Markov Chain Monte Carlo steps after a burn-in of length 200 000. Five independent replicates were performed for each value of K , and these gave consistent results. Posterior probabilities of K were calculated using the values of $\ln P(X > K)$ where X denotes the genotypes of the sampled individuals.

2.11 Autosomal STR typing and sequencing

STR typing was performed using the AmpF/STR Identifier or the AmpF/STR SGM Plus amplification kit (Applied Biosystems). The PCR was performed as recommended by the manufacturer.

Fourteen PCR amplicons carrying the SNPs rs1028528, rs1029047, rs1979255, rs2016276, rs2046361, rs2056277, rs2107612, rs2831700, rs354439, rs733164, rs735155, rs901398, rs907100 and rs938283 were sequenced in at least three individuals using the BigDye Terminator Kit (Applied Biosystems) as recommended by the manufacturer and analysed on an ABI Prism 377 Genetic Analyzer (Applied Biosystems).

2.12 Forensic statistical analysis

The power of discrimination was calculated as described by Jones [12] (1972) using PowerStats (Promega). The matching probability, mean exclusion probability and the typical paternity index were calculated as described by Brenner and Morris [13] (1989) using DNAVIEW™ 27.19.

3 Results

3.1 Selection of SNP loci and primers for PCR and SBE

Three groups of candidate SNP loci fulfilling the criteria described in Section 2 were collected from the NCBI SNP database, AB SNP Browser and the HapMap genome browser. The first group of 46 SNPs comprised 2–3 SNPs

from the distal region of the p -arm of each autosome (with the exception of the gene dense chromosome 19), the second group of 67 SNPs comprised 3–4 SNPs from the distal region of the equivalent q -arm and the third group contained 25 SNPs selected mostly from the large autosomes. The final selection of SNP loci for the multiplex was determined during the phase of PCR primer design. The SNP from each chromosome end that gave the best chance to be amplified as a short fragment in a large multiplex under the conditions set for primer design was selected from the first two candidate groups. No suitable loci on the distal regions of chromosome 19 were found in this selection, a finding not completely unexpected, since chromosome 19 has a very high gene density [14]. Another 12 loci, including one SNP from chromosome 19, were selected from the third group of loci bringing the total number of loci to 54. All loci were successfully amplified in singleplex PCR and then divided into two large multiplexes with 23 and 31 fragments, respectively. The two PCR multiplexes were optimized and finally combined into a single multiplex with 54 loci. At no point during this process was it necessary to replace any of the originally designed PCR primers, emphasizing the importance of strict adherence to clearly defined guidelines for primer design when developing large multiplexes. Similarly, only 5 of the 54 SBE primer designs were altered during the development of the SBE multiplexes. Alternative primers on the opposite DNA strand were needed in three instances because the SBE reactions for these primers were so efficient that it had a deleterious effect on the balance and stability of the multiplex extension reaction. In two cases, the lengths of the SBE primers were altered to allow more efficient analysis of the electropherograms.

During the multiplex development phase, the genomic positions or the sequences surrounding the selected SNPs underwent revision in no less than 35 of the 54 loci. This arose from one major and several minor updates of the reference sequence builds on which the dbSNP database is founded, principally correcting numerous segments of inverted sequences. In most cases, this had no impact on the PCR or SBE primer designs. However, the new information reported in dbSNP disqualified 2 of the 54 SNP loci selected for the multiplex. In one case, SNP rs2145294 was reported as duplicated, and in the other case, a revised position was given for SNP rs1360674 only 4.24 kb from another selected locus on chromosome 1. This reduced the final combination to 52 SNPs.

3.2 Analysis of the SBE reaction

We chose to use the ABI Prism SNaPshot™ Multiplex System for generation of the SBE products because SBE primers extended with a fluorescently labelled ddNTP can

be detected with the CE equipment available in most forensic genetic laboratories, including each of the consortium laboratories. However, it was immediately clear that all 52 SNPs could not be detected in one SBE reaction in a five-colour electropherogram as it was impossible to construct suitably spaced size windows for 104 possible alleles when four to six nucleotide spacing between size windows in the same colour was needed and the maximum size of extended SBE primers could not realistically exceed 90 nucleotides [15, 9]. Therefore, we decided to develop two SBE multiplexes with 23 and 29 SBE primers, respectively, and analyse the 52 SNPs in two different electrophoretic runs.

On the ABI Prism 310 Genetic Analyser, it was possible to inject the two SBE multiplexes shortly after each other in the same capillary and thus analyse all 52 SNPs in one data file. This was not possible with the ABI Prism 3100 Genetic Analysers because there is a default capillary polymer fill stage between each injection. To address this problem, we created a new GeneScan method file for the ABI 3100 that allowed sequential injection into the same capillary (see Section 2). During this work it also became clear that the same SBE product analysed on an ABI 310 and on an ABI 3100 gave very different signal strengths. On average, the sensitivity of the ABI 310 was only 20% of the sensitivity of the ABI 3100, and most of the work presented here (see below) was analysed using ABI 3100 or ABI 3100-Avant Genetic Analysers.

In the early phase of the optimisation of the SBE multiplexes, we observed some small, distinct peaks that did not have the appearance of background noise. Addition of ammonium sulfate to the SBE primer mix significantly reduced the number of these nonspecific peaks, indicating that several of the peaks originated from extension of primer-dimers [16]. However, even after extensive optimisation of the sample purification and the SBE reaction, certain peaks were observed occasionally, mostly in the green dye electropherograms. We assume that some of these peaks originate from nontemplate addition of fluorescent ddATP to the 3'-end of the shortest amplicons. Since the electrophoretic mobilities of these amplicons are constant, we designed the GenoTyper 3.7 allele calling macros to label only the peaks in the predefined allele windows and ignore peaks outside the prescribed size windows.

The largest concern during analysis of the electropherograms was the difference in fluorophore emission from the four fluorophores dR110 (blue), dR6G (green), TAMRA (yellow) and dROX (red) used in the SNaPshot Multiplex System. On average, the ratio between the signal strengths of the colours was 4:2:1:1, and the peak thresholds in GeneScan 3.7 and the rules for heterozygous and homozygous allele calls had to be adjusted accordingly (see

Section 2). For the large majority of the samples, this was sufficient to ensure robust and reproducible analysis of the electropherogram. However, for some weakly amplified samples, the differences in signal strengths of the four colours gave rise to 'no calls' for certain SNPs. All 700 samples analysed in the population validation work were typed twice and 80% of the samples gave results for all 52 SNPs in the first and second runs. In the other 20% of samples, between one and five of the 52 SNPs were designated no call in one of the typing runs, and a third analysis using increased amount of DNA in the PCR was necessary to make the correct allele call. No call results were equally likely in all SNPs and not confined to particular loci.

3.3 Validation of the 52 SNP-plex assay

PCR and SBE primers were distributed to the four consortium laboratories along with ten blood samples spotted on FTA cards. The ten samples were typed in all laboratories with identical results and this test was used to validate the successful implementation of the 52 SNP-plex assay in each laboratory before more extensive typing of samples began. In general, genomic DNA purified from blood by phenol/chloroform or column based methods were used in the PCR reaction, but Chelex-purified DNA also gave satisfactory amplification of all 52 SNP loci.

In order to analyse the performance of the multiplex PCR and SBE reactions, sequential dilutions of DNA from two different samples purified by the QIAamp DNA blood mini kit were made and different amounts of DNA from 68 pg to 70 ng were used in the PCR reaction (Fig. 2). As expected, allele drop-outs and unusual peak height ratios for heterozygote SNPs due to stochastic phenomena in the PCR reaction [17] were frequently observed when less than 200 pg of DNA was used. However, a complete SNP profile was obtained for both samples from only 500 pg DNA. The electropherograms with the best peak balance were obtained with 1–17 ng DNA, but acceptable results were also obtained with up to 70 ng DNA.

The influence of PCR inhibitors on the multiplex PCR was tested by addition of porcine haematin, a haeme derivative that is known to inhibit *Taq* polymerase [18]. Three samples with 10 ng template DNA were treated with haematin to final concentrations of 5, 10, 15, 20, 25, 30, 35 and 40 μ M. Reduced PCR amplification products were observed with 5 μ M haematin and complete inhibition of the PCR with 10 μ M haematin (data not shown).

The DNA-regions surrounding 14 SNPs were sequenced to verify the allele calling of the SBE reactions. For each SNP, sequences for at least three individuals were obtained. All sequences were in agreement with the SNP-typing results.

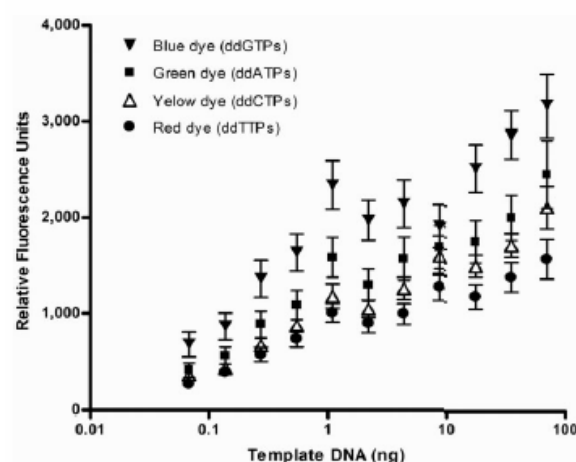


Figure 2. Sensitivity of the 52 SNP-plex assay. Different amounts of genomic DNA (0.07, 0.14, 0.27, 0.55, 1.09, 2.19, 4.37, 8.75, 17.50, 35 and 70 ng) from two individuals were used. RFUs of each of the four dyes were collated and normalized by dividing the RFU values from homozygote alleles by two. Normalised average RFUs are shown as a function of the amounts of template DNA in the multiplex PCR on a $\log_{10}$ scale. Error bars indicate the S.E.Ms.

The 52 SNP-plex assay was tested with DNA from dog, cat and horse, as well as 46 individual samples representing 6 different primate species. In total, 46 of the 52 SBE primers were extended in the SBE multiplexes in one or more of the primate species. In chimpanzees, 44 loci were typed, 37 loci were typed in orangutans, 36 in gorillas, 23 in baboons, 22 in rhesus monkeys and 21 in Cynomolgus monkeys. No amplification was detected from the dog, cat and horse samples. Table S3 shows the results obtained from the 46 primate samples, each sample being homozygous for all loci with the exception of 2 out of 29 chimpanzees that were heterozygous for SNP rs2056277 (#28). For the loci where SBE products were detected in more than one primate species, the same allele was observed in most cases. However, in five loci both the known human alleles were detected among the primates tested. For the other 41 loci, the frequencies of the alleles observed in primates indicated that the proposed ancestral alleles (alleles found in the last common ancestor of humans and chimpanzees) were marginally more frequent in Somalis (average: 0.53) than in Europeans and Asians (average for both: 0.48).

3.4 SNP allele distributions in nine different populations

A total of 700 samples from 9 different populations were typed with the 52 SNP-plex assay and the allele frequencies in each population determined (Table S4). The

populations were divided in three groups: Somali, Asian (Chinese, Taiwanese, Thais, Greenlanders, Japanese) and European (Danes, Germans, Turks), and the combined allele frequencies compared to the allele frequencies in three major population groups; African-American, East Asian and European reported by The SNP Consortium, Celera and the RealSNP databases (Table S4, March 2005 release). The observed allele frequencies deviated by an average of 0.08 (156 comparisons) from the database allele frequencies, with a range from 0.00 to 0.31. Pearson correlation coefficients (r) were calculated to measure correlations between the allele frequency estimates in our grouped populations and the database allele frequencies. The correlation coefficients were $r = 0.80$ (Danes, Germans, Turks vs. European), $r = 0.72$ (Somalis vs. African-American) and $r = 0.83$ (Chinese, Taiwanese, Thais, Greenlanders and Japanese vs. East Asian). Similarly, correlation coefficients between allele frequencies from combinations of our nine populations and those of three of the four listed populations in the HapMap database were calculated. Not all of the 52 SNPs have been genotyped by HapMap, but where allele frequency estimates were listed, appropriate population groupings were compared: $r = 0.51$ (39 SNPs for Somalis vs. Yoruba in Ibadan, Nigeria), $r = 0.94$ (41 SNPs for combined Danes and Germans vs. Utah residents with ancestry from northern and Western Europe) and $r = 0.96$ (36 SNPs for combined Taiwanese and Chinese vs. Han Chinese in Beijing, China).

Cluster analysis using the model-based approach implemented in Structure v.2.0 was performed with the 700 individuals. The splitting order illustrated in Fig. 3, was as follows: at $K = 2$, one cluster contained Somalis, Turkish, Danes and Germans, whereas Asians and Greenlanders were grouped together in another cluster. Each increase in K split one of the clusters obtained with the previous value. Thus, at $K = 3$, Somalis were separated from Turkish, Danes and Germans without altering the Asian and Greenland group. At $K = 4$, Greenlanders clearly split from the East Asian group. Higher values of K produced further equitable membership coefficients within the European and Middle East individuals without modifying the patterns of the other clusters (the proportion of the sample assigned to each population is roughly symmetric indicating lack of further substructure within these populations). The posterior probability was almost equal to one for $K = 4$ and virtually zero for other K values. Several individuals had high membership coefficients in more than one cluster. For instance, at $K = 4$, Somalis included two individuals with clear ancestry coefficients from the main European group, while the opposite pattern was seen in individuals in the European cluster (Fig. 3).

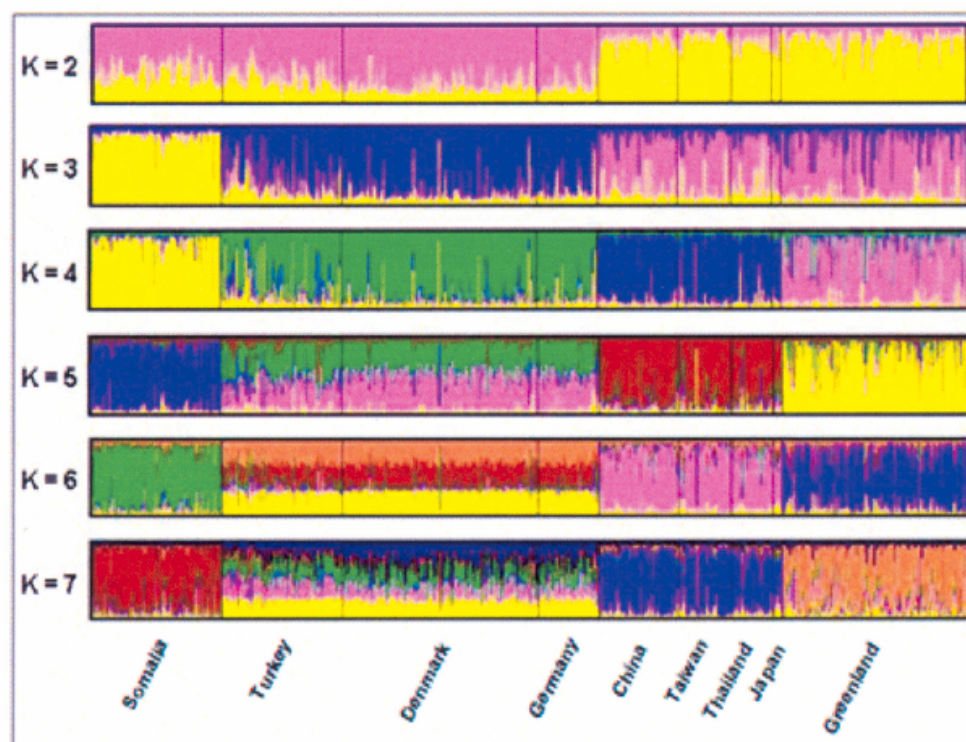


Figure 3. Cluster analysis. Each bar represents a single individual, and colours correspond to the coefficients of ancestry.

The genotype distributions of SNP rs907100 (#38) differed from expectations based on Hardy–Weinberg equilibrium in six of the nine populations due to an excess of individuals typed as homozygotes ($\chi^2 = 12.04$, $p < 0.05$) and the deviations were still significant in three populations after Bonferroni correction [19]. In response to this, we are currently testing a new reverse PCR primer in 100 mother–father–child trios from Denmark and Somalia in order to determine whether the observed disequilibrium is related to a recently described SNP in position 9 from the 3' end of the reverse primer. The remaining SNP genotype distributions all matched expectations. Average heterozygosities were 0.44, 0.41 and 0.38 in Europeans, Somalis and Asians, respectively. The lowest heterozygosity was found in the Greenland and Taiwanese groups (both 0.37).

Association of alleles across loci (linkage disequilibrium – LD) was estimated using the χ^2 -test, the r^2 coefficient [20] and the Exact test [21]. LD tests for pairs of SNPs on the same chromosome demonstrated no significant deviation from the expectations ($p > 0.05$ after sequential Bonferroni correction).

Table S5 shows analyses of variances for all 52 markers in European (F_{ST}), Somali (F_{IS}) and Asian populations (F_{ST} including Greenlanders). The F_{ST} values ranged from 0.003 to 0.217 with a mean of 0.07. The distribution of the F_{ST} values is shown in Fig. 4. The tendency to bimodality was mainly due to the contribution of the European populations.

The allele distribution pattern showed uniformity in the studied populations with approximately 93% genetic variability within populations. To determine the distribution of residual genetic variance, we grouped the populations into Europeans, Somalis and Asians (including Greenlanders) and calculated the Φ statistics. Low values of both inter- and intragroup variability ($\Phi_{CT} = 0.054$; $\Phi_{SC} = 0.035$) were observed in each case.

3.5 Forensic statistics

The combined mean match probability using the 52 SNPs was between 5.0×10^{-19} (Asian) and 5.0×10^{-21} (European), corresponding to a combined power of

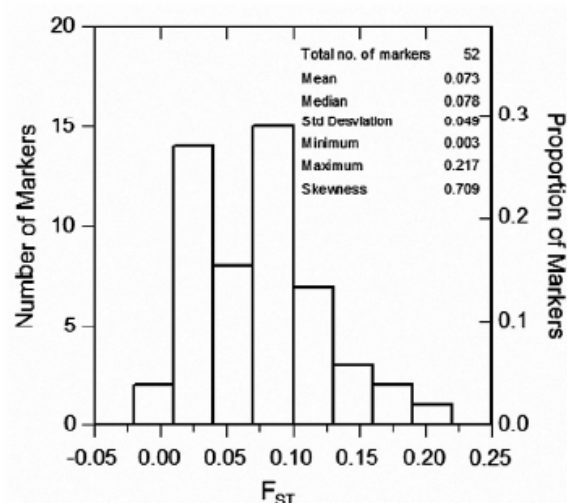


Figure 4. F_{ST} distribution of the 52 autosomal SNPs. Values were calculated for a global population including 700 individuals from Denmark (156), Greenland (149), Somalia (104), Turkey (96), China (63), Germany (49), Taiwan (43), Thailand (33) and Japan (7).

discrimination of 99.9999 and 99.99999%, respectively (Table 1). We performed pair-wise comparisons of the 700 samples typed in the validation study. The average number of identical SNP loci between two unrelated individuals was 21 with a minimum of 7 and a maximum of 41 matching loci (Fig. 6).

The typical paternity indices [13] obtained with the 52 SNPs ranged from 336 000 in Asians to more than 549 000 in Europeans, corresponding to a mean exclusion probability of 99.91 and 99.98%, respectively. In motherless cases, the typical paternity indices ranged from 2880 (Asian) to 4640 (European) (Table 1).

3.6 Typing of partially degraded DNA from crime case samples

Purified DNA from bone, muscle or other tissues collected from seven cadavers found under different environmental conditions and in various stages of decomposition, obtained previously as part of crime case investigations,

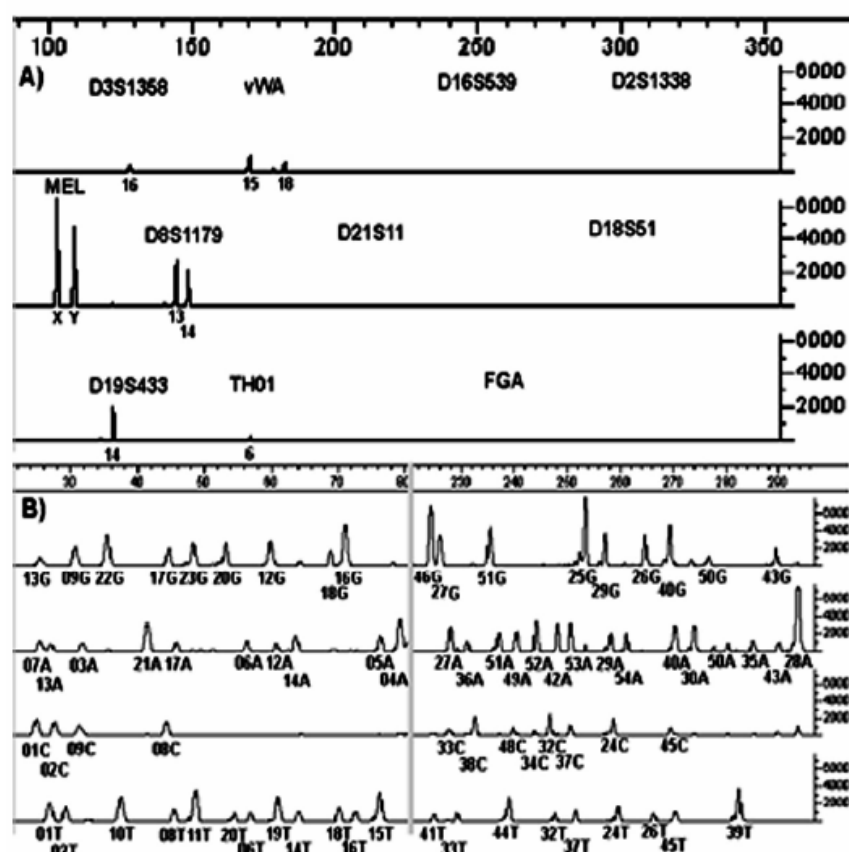
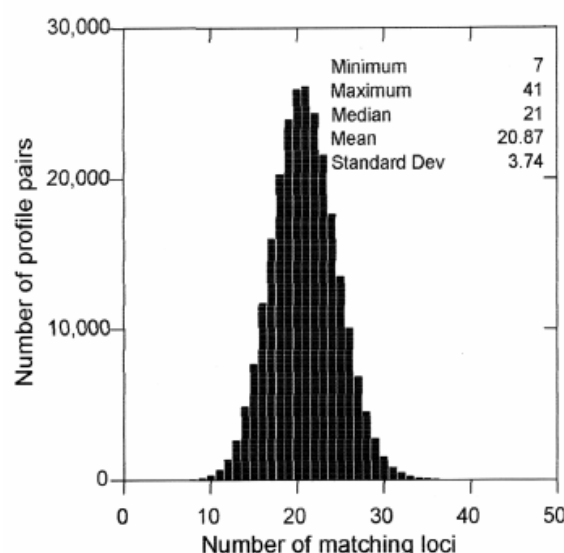


Figure 5. Representative electropherograms of (A) AmpF/STR SGM Plus and (B) 52 SNP-plex assay typing of a sample with partly degraded DNA. A total of 2.6 μ L of 0.4 ng/ μ L DNA was used in the PCR amplifications.

Table 1. Forensic statistical parameters

	European	Somali	Asian
Mean match probability	5.0×10^{-21}	1.1×10^{-19}	5.0×10^{-19}
Combined power of discrimination	>99.99999%	>99.9999%	>99.9999%
Mean exclusion probability	99.98%	99.95%	99.91%
'Typical' paternity index (trios)	549 000	337 000	336 000
Typical paternity index (motherless)	4 640	3 160	2 880

**Figure 6.** Distribution of the number of identical SNP loci found in unrelated individuals. Values were calculated for 700 individuals belonging to nine different populations. Total number of comparisons was 244 650.

were analyzed for ten STR loci using the AmpFLSTR SGM Plus PCR Amplification kit. The short-PCR fragments in the STR kit were amplified efficiently whereas the longer fragments (more than 200 bp) were very poorly amplified or not amplified at all (example shown in Fig. 5A), suggesting that the DNA was partially degraded. A series of dilutions of DNA from the seven samples were made and different amounts of DNA were used in the 52-plex PCR. When using concentrations identical to those used in the AmpF/STR SGM plus PCR (approximately 0.5 and 1 ng), all 52 SNPs were successfully typed (example shown in Fig. 5B). At lower concentrations (approximately 0.225 ng), the 52 SNP-plex assay also gave full SNP profiles for six of the seven samples (data not shown). Five of the SNPs (rs717302, rs729172, rs1015250, rs1528460 and rs1886510) consistently gave weak signals suggesting that they perform poorly under these conditions, but this did not correlate with the sizes of the amplicons; the sizes of the amplicons for these loci varied from 60 to

115 bp. Similarly, no relationship was observed between the amplicon sizes and successful SNP allele calls when genotyping the 700 blood samples (see above).

4 Discussion

This study demonstrates that it is possible to construct large, sensitive multiplex PCR assays for the detection of more than 50 SNPs for forensic applications. As a result, a set of SNP markers carefully tailored for use in human identification and readily genotyped with established technology is now available to the forensic community. STR typing is the most commonly used typing method for forensic genetic investigations in crime casework, but this technique may not always give conclusive results in cases with heavily degraded DNA [22]. In such instances, the size of the amplicon to be investigated is critical. With the widely used STR markers, the lengths of the amplicons range from approximately 100 to 400 bp. Efforts are being made to reduce the amplicon lengths of STR systems currently in routine use [23, 24]. In addition, new shorter STR loci are being identified and tested for practical forensic use [25]. SNP typing has the advantage that amplicons less than 60 bp in length may be investigated because the length of the amplicon is restricted only by the compositions of the flanking sequences and the lengths of the PCR primers. Therefore, in the 52 SNP-plex assay outlined here, amplicons have a maximum size of only 115 bp.

The most important factors allowing the construction of the present 52 SNP-plex assay were (i) a careful selection of primers to avoid intra- and interactions between the primers; (ii) high-quality primers that are pure and homogeneous; and (iii) careful balancing of the PCR multiplex and the SBE multiplex reactions [9]. Until now, only a few large multiplexes have been reported [26], but larger multiplexes that are constructed based on the same principles as the present 52-plex are emerging, e.g. packages with Y chromosome SNPs [15, 27] or autosomal SNPs with contrasting allele frequency distributions in different populations useful for the estimation of the population of origin (Phillips *et al.*, in preparation). In order to speed up the typing process and keep costs low, we developed a double injection method

that has the advantage that both the 23-plex and the 29-plex SBE products can be investigated in the same electrophoresis run and computer analysis. In addition, automatic allele calling macros have been developed and are available for use with this SNP set. These macros reduce both the analysis time involved and the risk of allele calling errors. Overall, the 52 SNP-plex assay performed well in four different laboratories when reaction mixtures were sent to the participating laboratories on dry ice. Preliminary results demonstrated that the 52-plex also performed well when applied to other SNP-typing platforms. Currently, 48 of 52 SNPs have been successfully incorporated into the Sequenom MassArray™ MALDI-TOF-based genotyping assay (Phillips *et al.*, unpublished), and 39 of the 52 SNPs have been implemented successfully onto the Nanochip™ Molecular Biology Workstation (Balogh *et al.*, unpublished). Furthermore, we are also implementing the 52-plex into alternative MALDI-TOF-based assays [28], conventional microarrays and oligonucleotide ligation-based assays.

The lack of polymorphism in various primates allowed the recognition of the ancestral allele in 41 of the 52 SNPs analysed. For five of the SNPs, both human alleles were detected in the primates studied. It is not possible to conclude whether this was caused by a higher mutation rate at these loci compared to the average SNP mutation rate, or if the observation reflects that the substitutions appeared before the genetic divergence of the primates studied and present-day humans.

The observed allele frequencies deviated on average 0.08 (maximum 0.31) from those reported by The SNP Consortium, Perlegen and HapMap databases when our data were pooled into three major groups. In the few cases where large frequency estimate differences were observed, it is possible that these resulted from divergence between the listed populations and those we analysed or that the sample sizes used in the SNP database validation were small or unrepresentative.

Although we had selected SNPs with allele distributions as close to 0.5/0.5 as possible in Europeans, Asian and African populations, using Cluster analysis (Fig. 3), we observed a broadly comparable population-grouping pattern to that found in a previous study in which a considerably higher number of polymorphic markers were used [29]. Although the interpretation of the patterns in our case requires caution due to our limited range of sampled populations and the very limited number of SNPs as well as the low power of the majority of the SNPs for prediction of the origin of population of individuals, grouping individuals into appropriate principal regions gave consistent results that provided a high correlation between predefined populations, geography and the inferred clusters. However, if one wants to use SNPs for prediction of the origin of

population, the SNPs should be selected for this specific purpose. The observation that several individuals exhibited high membership coefficients in multiple clusters indicates a degree of coancestry in some of the populations sampled, suggesting some population admixture. This is consistent with the hypothesis that human populations are not discrete groups, since admixture commonly occurs between neighbouring populations [30]. However, the use of a limited number of SNPs in our analysis could also have contributed to the extent to which multiple group membership was observed.

For the European and East Asian population groups, we observed high-correlation (*r*) values when comparing the allele frequencies of the present study with those of HapMap. In contrast, the coefficient of correlation between the Hapmap Yoruba Nigerian population and our Somali populations was much lower, and this mirrors the high genetic diversity of the African continent [31, 32]. Clearly, further studies and a broader range of population samples from this continent will be needed to refine our knowledge of the allele frequency patterns for the 52 SNPs amongst Africans.

The genotypes of all loci studied except SNP rs907100 (#38) were distributed as expected based on the assumption of Hardy–Weinberg equilibrium. Recently, a new SNP (rs11689319) was reported in position 9 from the 3' end of the reverse PCR primer used to amplify rs907100. A new reverse PCR primer with a degenerate base in position 9 from the 3' end is currently being tested in order to determine whether the unexpectedly high number of homozygotes is caused by this substitution.

The heterozygosity values of the 52 SNPs in the tested populations were close to those expected with the highest average of 0.44 in Europeans. This is not surprising considering one of the primary SNP selection criteria was a priority for maximum heterozygosity in European populations with less emphasis on the same variability in other populations for most SNPs chosen. The low average F_{ST} value of 0.07 (maximum of 0.22 for the total group) underlines the fact that none of these markers has any known or likely functional relevance, another key factor in locus selection being sufficient distance between SNPs and neighbouring genes.

Due to the limited amount of DNA in many crime cases, it is important that typing of a sufficient number of SNPs can be performed reliably on small amounts of DNA from traces of evidential material. STR typing can often work on as little as 200 pg of target DNA. In our study partially degraded, purified DNA from seven different cadavers were typed for the full set of 52 SNP markers using 200–500 pg DNA (Fig. 5). In contrast, only the short-PCR fragments from the

AmpF/STR SGM Plus STR amplification kit were amplified efficiently in these samples. These results are very promising for practical crime casework in which only very small amounts of partially degraded DNA are available.

The mean power of discrimination (5.0×10^{-19} to 5.0×10^{-21}) would be satisfactory in crime cases with a full profile from only one contributor. However, if a mixture of DNA from two or more individuals is found, additional SNPs, including SNPs with unequal allelic distribution (i.e. much lower average minor allele frequencies), will be necessary in order to obtain a weight of evidence equivalent to that obtained by STR typing [3]. For paternity testing, the 52 SNP-plex assay will be most valuable in European populations (typical paternity index: 549 000) while the unequal distribution of the alleles in Asians and Somalis resulted in lower paternity indices (typical paternity indices: 336 000 and 337 000, respectively). This demonstrates the importance of selecting SNPs with allele frequencies close to 0.5 for relationship testing applications.

In conclusion, we have identified a set of 52 polymorphic SNPs that can be typed by standard methods and we have developed an assay that allows multiplex PCR amplification of very limited amounts of DNA. The SNP-typing procedure uses standard CE equipment available to most modern forensic genetic laboratories. Furthermore, the 52-plex SNP set can be readily adapted to a range of other genotyping methods offering the possibility of high-throughput solutions in the future. The SNPs presented are freely available and can be part of a future core set of SNPs for forensic genetic investigations.

We thank Ms. Annette Holbo Birk for technical assistance, Jock Nielsen, PhD, for help in programming Access macros and Bo Simonsen, PhD for helpful discussions. The work was supported by grants to J. J. Sanchez from Ellen and Aage Andersen's Foundation, the European Commission (GROWTH programme, SNPforID project, contract G6RD-CT-2002-00844) and a grant from the 'Ministerio de Ciencia y Tecnología' (BMC2003-09822) to C. Phillips, M. Fondevila, A. Salas and A. Carracedo.

5 References

- [1] Schneider, P. M., Martin, P. D., *Forensic Sci. Int.* 2001, 119, 232–238.
- [2] Martin, P. D., Schmitter, H., Schneider, P. M., *Forensic Sci. Int.* 2001, 119, 225–231.
- [3] Gill, P., *Int. J. Legal Med.* 2001, 114, 204–210.
- [4] Amorim, A., Pereira, L., *Forensic Sci. Int.* 2005, 150, 17–21.
- [5] Hao, K., Li, C., Rosenow, C., Wong, W. H., *Eur. J. Hum. Genet.* 2004, 12, 1001–1006.
- [6] Pfaff, C. L., Barnholtz-Sloan, J., Wagner, J. K., Long, J. C., *Genet. Epidemiol.* 2004, 26, 305–315.
- [7] Shriver, M., Kennedy, G., Parra, E., Lawson, H. et al., *Hum. Genomics* 2004, 1, 274–286.
- [8] Phillips, C., in: Carracedo, A. (Ed.), *Forensic DNA Typing Protocols. Series: Methods in Molecular Biology*, Humana Press, Totowa, NJ 2004, pp. 83–106.
- [9] Sanchez, J. J., Børsting, C., Morling, N., in: Carracedo, A. (Ed.), *Forensic DNA Typing Protocols. Series: Methods in Molecular Biology*, Humana Press, Totowa, NJ 2004, pp. 209–228.
- [10] Weir, B. S., Hill, W. G., *Annu. Rev. Genet.* 2002, 36, 721–750.
- [11] Pritchard, J. K., Stephens, M., Donnelly, P., *Genetics* 2000, 155, 945–959.
- [12] Jones, D. A., *J. Forensic Sci.* 1972, 12, 355–359.
- [13] Brenner, C. H., Morris, J. W., *Proceedings for The International Symposium on Human Identification* 1989, Promega Corporation, Madison, WI pp. 21–53.
- [14] Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M. et al., *Nature* 2001, 409, 928–933.
- [15] Sanchez, J. J., Børsting, C., Hallenberg, C., Buchard, A. et al., *Forensic Sci. Int.* 2003, 137, 74–84.
- [16] Doi, Y., Yamamoto, Y., Inagaki, S., Shigeta, Y. et al., *Leg. Med. (Tokyo)* 2004, 6, 213–223.
- [17] Krenke, B. E., Tereba, A., Anderson, S. J., Buel, F. S. et al., *J. Forensic Sci.* 2002, 47, 773–785.
- [18] Wallin, J. M., Buonocristiani, M. R., Lazaruk, K. D., Fildes, N. et al., *J. Forensic Sci.* 1998, 43, 854–870.
- [19] Rice, W. R., *Evolution* 1989, 43, 223–225.
- [20] Weir, B. S., *Biometrics* 1979, 35, 235–254.
- [21] Zaykin, D., Zhivotovsky, L., Weir, B. S., *Genetica* 1995, 96, 169–178.
- [22] Schneider, P. M., Bender, K., Mayr, W. R., Parson, W. et al., *Forensic Sci. Int.* 2004, 139, 123–134.
- [23] Hellmann, A., Rohleder, U., Schmitter, H., Wittig, M., *Int. J. Legal Med.* 2001, 114, 269–273.
- [24] Butler, J. M., Shen, Y., McCord, B. R., *J. Forensic Sci.* 2003, 48, 1054–1064.
- [25] Coble, M. D., Butler, J. M., *J. Forensic Sci.* 2005, 50, 43–53.
- [26] Dixon, L. A., Murray, C. M., Archer, E. J., Dobbins, A. E. et al., *Forensic Sci. Int.* 2005, 154, 62–77.
- [27] Brion, M., Sanchez, J. J., Balogh, K., Thacker, C. et al., *Electrophoresis* 2005, 26, 4411–4420.
- [28] Mengel-Jørgensen, J., Sanchez, J. J., Børsting, C., Kirpekar, F., Morling, N., *Anal. Chem.* 2005, 77, 5229–5235.
- [29] Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M. et al., *Science* 2002, 298, 2381–2385.
- [30] Serre, D., Paabo, S., *Genome Res.* 2004, 14, 1679–1685.
- [31] Salas, A., Richards, M., De la Fe, T., Lareu, M. V. et al., *Am. J. Hum. Genet.* 2002, 71, 1082–1111.
- [32] Sanchez, J. J., Hallenberg, C., Børsting, C., Hernandez, A., Morling, N., *Eur. J. Hum. Genet.* 2005, 13, 856–866.

II. Tabellen und Abbildungen

Table S1: Information and primer sequences for PCR amplification of 52 autosomal SNPs									
Marker code	Chro	NCBI human moso chromosome me position	NCBI rs No.	TSC No.	PCR forward primers (5'-->3')	PCR reverse primers (5'-->3')	Final primer concentration (microm)	Amplicon size (bp)	TSC allele reported
1	1	4037521	rs1490413	TSC0724193	GTGTGGACTGGGTGATGT	TTCTCACTAGTGTCCCTGCTCTG	0,06	68	A/G
2	2	104974	rs876724	TSC0208870	GCAGGCTCCATTTTATACACT	GAATATCTATGAGCAGGAGTTAGC	0,02	83	C/T
3	3	936782	rs1357617	TSC0496080	AGCTGATGCAGACCATTAC	GGATAGCTGATAAGAAAACATGACCA	0,02	90	A/T
4	4	10719942	rs2046361	TSC1065282	CCTATTCTGATCTATCTATTGTCTATGAACG	GTCATTGTTGACACTTCACCTTCTA	0,03	79	A/T
5	5	2932133	rs717302	TSC0039610	CTTTAGAAAGGCATATCGTATTAACGTGTG	AACACAGAAAGAGGTTCCATATGTTGG	0,04	86	A/G
6	6	1080939	rs1029047	TSC0253802	CATAACGTGGATTGTGCAGCA	GGAAATAAAGCTGAAGGCTAAAGAAAAG	0,06	100	A/T
7	7	4201341	rs917118	TSC0229630	GCCCTTTAGGGTCGGTTC	GTAAGAGATGACTGAGGTCACGAG	0,03	87	C/T
8	8	1363017	rs763869	TSC0065968	ATCAAGTCTTTCTGTGTGACATTTG	GGCTACTCCCTCATTAATGTAATGC	0,09	100	C/T
9	9	1813774	rs1015250	TSC0097236	AAGTGATGGAGTTAGGAAAAGAAC	AAGACATTAGGTGGATTTCATAGCTG	0,03	95	C/G
10	10	3328178	rs735155	TSC0027519	GGAGAAAACCGGAGAGCTG	GAGTGTACCGAAATTCACAG	0,05	100	A/G
11	11	11060530	rs901398	TSC0177752	CTGGGTGCAAACTAGCTGAATATC	CTGGAATGTAAGTAGGCAAGAACTAA	0,02	70	C/T
12	12	741262	rs2107612	TSC1108144	GAGCATTCCTCTCTGTGTTAAATTCG	TGAGTACATTATTCAACTGTTTGGAG	0,05	93	A/G
13	13	20172700	rs1886510	TSC0904551	GTCTTGTCAATCTTTCTACCAGAG	GGATTTTCAACAACAACACTTGC	0,01	86	C/T
14	14	23840960	rs1454361	TSC0684657	AGGAAATACACCCCTGAGCTG	AGCTGTCCATCATCAGTAAGACAC	0,02	73	A/T
15	15	22119157	rs2016276	TSC0326920	TGCATCCAGCCCTCCACT	ATTGTACCTTGGCCACTTTTGTGTG	0,17	90	A/G
16	16	5606490	rs729172	TSC0028090	CATTAAATAGCAAGGCTCCTCT	ACATTTCCCTCTTTCGGTTAC	0,02	60	A/C
17	17	5907188	rs740910	TSC0105771	GTATAACAGTTTGTCTAAGTAAGTGAGTG	AGATAGGTTTCGAGTTTTCGGCTTTA	0,08	87	A/G
18	18	1117986	rs1493232	TSC0729796	CTATTCTCTCTTTTGGGTGCTAGG	CAAACTGTTTATTGTGTAGGCCCTGT	0,07	59	C/A
19	19	33155177	rs719366	TSC0044147	CCACAGCATCTTTTAACTCTTTTATATACC	GTAAGGACTTATAGTGTAGTAAGGACAGG	0,06	105	C/T
20	20	4442483	rs1031825	TSC0334834	CTTATCTTTTCCACATTTATGTCCT	AAGATATAATCACTGCTTTCAGTATGC	0,03	98	A/C
21	21	15607469	rs722098	TSC0050288	GGAAGTACACATCTGTTGACAGTAATGA	GGGTAAGAAAATATTCAGCACATCC	0,12	81	A/G
22	22	26141338	rs733164	TSC0023085	AGCTTTCAGCCCCAGTTC	CGGCTCAGGAATGTCAGG	0,02	68	A/G
23	10	2360631	rs826472	TSC0557086	TGAATTTTGTCTCTGTTATATTAGTCACC	TGTAATTGAAATTTGTAGGCAATAGAC	0,04	85	C/T

24	21	28601558	rs2831700	NA	GGCTAAACTATTGCCGGAGA	TTCCCTAGAACCAACAATTATCTGTCTC	0,04	62	G/A
25	14	96835572	rs873196	TSC0202434	GCATTCAAATCCCAAGTGCT	GCAGGAGTTGGAGTCAATCAG	0,03	63	C/T
26	16	79885888	rs1382387	TSC0544547	ACGAAGGAGAAACACTGAACCT	TGGAGTACTTAAATAAGACGCTGCAT	0,02	69	G/T
27	12	104830721	rs2111980	TSC1113476	AGCATCTTTGGCAGCATCC	AGCAAGATCTTTGGCCAGTGAGT	0,01	72	A/G
28	8	139370172	rs2056277	TSC1082757	CCAAACTGGGTGTTAGGGAGAC	TCATTATCTCGTCATACACTCCCTGT	0,03	73	C/T
29	18	73559363	rs1024116	TSC0247167	CCATGTGTTCTAAATAAAGAGATTGC	TGGGAAGTGAGCAAAAGTAAATACA	0,03	76	A/G
30	6	164954622	rs727811	TSC0062764	GTGTTTCTTTTTTCTTACCGAAC	GTGAATGAAATCATGAGATTGCTG	0,03	78	A/C
32	1	239753521	rs1413212	TSC0607362	AACCTCCTTTTGGAAACACTGAC	CAACATTCCATTATCCAGGAGAC	0,05	84	A/G
33	17	78065617	rs938283	TSC0357388	CATTGAAAGTCCTAACCCCTAGTACG	GGATGAGGCCCCACCCATA	0,03	85	C/T
34	4	191013970	rs1979255	TSC0925231	TCAGAGACTATGGATGGTATTTAGGTC	CATGGAACTGTTGGAACTCTTTG	0,03	86	C/G
35	9	122257493	rs1463729	TSC0377760	ACTATCAGTCTCTGCCCTTATTCTG	CACATGTGCATGCTTTTGG	0,04	87	A/G
36	11	134205198	rs2076848	TSC0022275	GCCTCACACCACAGAAATCAG	GACATCAGAAAACCTCCCATGAAACT	0,04	89	A/T
37	3	192127021	rs1355366	TSC0491662	CCATGATTTTCTTGTGGTGAGA	CACATGTGCTTAGGCCACAAC	0,05	90	A/G
38	2	239850329	rs907100	TSC0186810	GGAGTTTCTTGATAACGATTCTGAAG	ACAGAAAAAGAGCCAGTTTGA*	0,04	91	C/G
39	13	104636412	rs354439	TSC0700528	GGCTTCTCTTTTCCCTTATGTATCTC	CAGGTTCCGATAGAAAAACAGTGAAT	0,05	93	A/T
40	22	46048653	rs2040411	TSC1056845	TCTGGAATGCCAGTTCTTTTGT	CAGAAACGCTATGAAAAACCACT	0,03	94	A/G
41	7	154850085	rs737681	TSC0033074	ACATGTGAGGCCATCTCCAC	CCTTACTGTGTAGTAGGCACCTGTTT	0,03	96	C/T
42	21	27530034	rs2830795	NA	CACTTCTATAGACATAGGACACACCAT	ATCTAGGCTCTGAATCAGGATGAG	0,02	97	A/G
43	5	174759601	rs251934	TSC0220872	AGAGGGCAGTGAGGCTTTTAAGTAG	TGCTAGAAATCCAGACTTAACTACCAG	0,01	98	C/T
44	21	41336325	rs914165	TSC0197658	AGCAGCAGAGCCTGGATG	AGACCACTCACCTCTTTTGCACT	0,07	100	A/G
45	1	235480457	rs10495407	NA	AGATCTCCACTTCTCTCTTGGTTG	CTCCCAAATTTACATTGCCACT	0,05	102	G/A
46	9	124344108	rs1360288	TSC0501229	AGACTCTCTGTGTGTGGCTTTG	GAGGGGGCATCTGTTGAG	0,04	103	C/T
48	10	132172819	rs964681	TSC0270699	GTACCTGGAGTGATTTCTGTGAG	GTTATGGAGGATTTGGTAAGAACAG	0,03	106	C/T
49	20	40172539	rs1005533	TSC0082071	GGTTTGTGTGTGAGTGTTTCAGAT	CCTTATGCCTCCCTGGAAC	0,03	107	A/G
50	15	51332965	rs8037429	NA	TTCACTTTGTGTACACCTCCATAGTA	TGCTACGTAAGAGGTCATTGCTATC	0,03	108	C/T
51	1	236923075	rs891700	TSC0162577	TTTTTCAGAGGTGGTATTTCTAGCTG	GCTATGACACTCCTTTAGAACTATGCAA	0,04	109	A/G
52	13	18699724	rs1335873	TSC0829150	GTGGATGATATGTTTCTCAAGG	TTCAACAAACCGTGTGATGCTCT	0,03	110	A/T
53	22	46574531	rs1028528	TSC0253071	ACAGCTGATGCCCTCCCTGA	GAGGATGAAGTTTAGAGCCAGAC	0,02	113	A/G
54	15	52926761	rs1528460	TSC0798410	TCCTGGAGATCAATATTAGCCCTTA	GGGTGACCAGTAGTTCTATGAGC	0,06	115	C/T

NA = Not Available.

*Bold lower case define the base replaced with a pyrimidine derivate P base analoge.

Table S2: Single base extension primer sequences of 52 autosomal SNP markers

Marker code	NCBI rs No.	Poly (dNTP) (5'-->3')	Neutral sequence (5'-->3')	SBE target specific sequence (microm)	Final primer concentration	SBE primer size	SNP detected	Detection orientation ^a
First 23 Autosomal SNP SBE set								
1	rs1490413	None	A	TCTGAGGCCAGCCAGTT	0,01	18	T/C	Reverse
2	rs876724	None	CAA	CACTGCACCTGAAGTATAAGTA	0,04	24	C/T	Forward
3	rs1357617	None	TCTGACAA	TGATAAGAAAACATGACCAAGC	0,02	29	T/A	Reverse
4	rs2046361	17 (dC)	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	TTGTTGACACTTCACTTCTA	0,27	78	T/A	Reverse
5	rs717302	13 (dC)	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	GGCATAATCGTATTAACTGTGT	0,02	74	A/G	Forward
6	rs1029047	None	AAACTAGGTGCCACGTCGTGAAAGTCTGACAA	GTAAGAATTCAAGATGGTATTT	0,04	54	A/T	Forward
7	rs917118	None	None	TGACTGAGGTCAACGAGC	0,02	18	G/A	Reverse
8	rs763869	None	CGTGAAAGTCTGACAA	GTTTGTATTATATTATTTCTAACTCAA	0,03	42	C/T	Forward
9	rs1015250	None	GTCTGACAA	ATGAAAAATCACTGGGACAG	0,02	29	G/C	Reverse
10	rs735155	None	GTGAAAGTCTGACAA	CACCGAATTCAACGGGAAG	0,02	34	T/C	Reverse
11	rs901398	None	AGGTGCCACGTCGTGAAAGTCTGACAA	ACTAGCTGAATATCAGCCC	0,04	46	C/T	Forward
12	rs2107612	None	TAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	GTGTTTTTCTTAAATCATATGTCT	0,02	58	A/G	Forward
13	rs1886510	None	ACAA	GAATTTTACAAACACATTGC	0,05	25	G/A	Reverse
14	rs1454361	4 (dC)	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	AAATACACCTGAGCTGC	0,02	62	A/T	Forward
15	rs2016276	14 (dC)	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	TGAGAGAGAGAAAAATGTGC	0,02	74	T/C	Reverse
16	rs729172	10 (dC)	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	TCTTGCGGTTACATACATT	0,02	70	T/G	Reverse
17	rs740910	None	CGTCGTGAAAGTCTGACAA	TAAAGTGAAGTGGTATAATCATAT	0,01	42	A/G	Forward
18	rs1493232	5 (dC)	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	TTGTGAGGCCTGTTTATTTTG	0,01	66	T/G	Reverse
19	rs719366	None	ACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	TTTTCTCTCCCATTTCTAG	0,04	58	C/T	Forward
20	rs1031825	None	AAACTAGGTGCCACGTCGTGAAAGTCTGACAA	TTTTCAAGTATGCTCGGGG	0,02	50	T/G	Reverse
21	rs722098	None	CGTGAAAGTCTGACAA	TGACAGTAATGAAATATCCTTG	0,01	38	A/G	Forward
22	rs733164	None	CGTGAAAGTCTGACAA	GGCCATCCCACTTGAAA	0,01	34	A/G	Forward
23	rs826472	None	TGCCACGTCGTGAAAGTCTGACAA	GACTAAATTAATCTGCTATCCTG	0,01	46	G/A	Reverse

24	rs2831700	dT + 16 (dCdT)	None	TA	TCTGTCTTTGGCCTTCTAACT	0, 03	56	C/T	Reverse
25	rs873196	dT + 14 (dCdT)	None	AA	TACAGGCATGTTACATTACAA	0, 03	52	A/G	Reverse
26	rs1382387	dT + 21 (dCdT)	None	G	GAGAAACACCTGAACTTTCAA	0, 03	64	G/T	Forward
27	rs2111980	2 (dCdT)	None	CA	CTTTGGCAGCATCCTTC	0, 03	23	A/G	Forward
28	rs2056277	dT + 37 (dCdT)	None	CC	CTGTCTTTGAACTCCC	0, 02	92	A/G	Reverse
29	rs1024116	dT + 15 (dCdT)	None	TG	TTCTAATAAAAAAGGATTGCTCAT	0, 03	56	A/G	Forward
30	rs727811	dT + 26 (dCdT)	None	CC	GAACTTCAACGACTTA	0, 04	72	A/C	Forward
32	rs1413212	9 (dCdT)	None	AG	ACATTTGTTTCATATAAGTCAAAATG	0, 05	44	C/T	Reverse
33	rs938283	dT	None	CT	AGTACGTTAGATGTGACCG	0, 05	22	C/T	Forward
34	rs1979255	dT + 10 (dCdT)	None	GA	AGTTGGAACCTCTTGCT	0, 05	40	G/C	Reverse
35	rs1463729	32 (dCdT)	None	C	AGATACACTCATAGCCAC	0, 05	84	A/G	Forward
36	rs2076848	5 (dCdT)	None	CC	ACAGAAATCAGGGC	0, 05	27	A/T	Forward
37	rs1355366	dT + 15 (dCdT)	None	CT	GGCTTTGGCAAGTCC	0, 03	48	C/T	Reverse
38	rs907100	dT + 4 (dCdT)	None	CC	AGTTGAGGCTTCTTT	0, 03	27	G/C	Reverse
39	rs354439	dT + 28 (dCdT)	None	CT	CTCTCATGTATCACATTCT	0, 04	80	A/T	Forward
40	rs2040411	21 (dCdT) + dC	None	CT	CTGTATTTTCTTACTCTAAGTGC	0, 05	68	A/G	Forward
41	rs737681	None	None	TG	AGGCCATCTCCACC	0, 03	16	C/T	Forward
42	rs2830795	dT + 9 (dCdT)	None	GG	ACACACCAATTTTATTGTCTAAAG	0, 03	44	A/G	Forward
43	rs251934	dT + 33 (dCdT)	None	TA	GTAGATATCTGGCTGTCCC	0, 03	88	A/G	Reverse
44	rs914165	dT + 7 (dCdT)	None	GA	GCTGCTGAGGTGT	0, 05	32	C/T	Reverse
45	rs10495407	22 (dCdT)	None	G	AAAAACAACAGAAATGGAATAGGA	0, 04	68	C/T	Reverse
46	rs1360288	None	None	AG	GGGATGCAGCCAGT	0, 03	16	A/G	Reverse
48	rs964681	dT + 9 (dCdT)	None	TG	TAGGAACGTCGACA	0, 03	36	C/T	Forward
49	rs1005533	dT + 9 (dCdT)	None	G	CAAGAGCCGTGGAATT	0, 03	36	A/G	Forward
50	rs8037429	dT + 26 (dCdT)	None	T	CAGAAATAACTCTTTCGCTAC	0, 03	76	A/G	Reverse
51	rs891700	dT + 3 (dCdT)	None	TC	CAATCTTTTTTTTTTGAAGCCT	0, 02	32	A/G	Forward
52	rs1335873	dT + 7 (dCdT)	None	AG	TACCTAGCTATGTACTCAGTAT	0, 04	40	T/A	Reverse
53	rs1028528	dT + 13 (dCdT)	None	G	AAAGGTCTTACTCGACATC	0, 03	48	A/G	Forward
54	rs1528460	dT + 17 (dCdT)	None	AA	AACTTTTACTGTAACTTTTCTCA	0, 03	60	A/G	Reverse

NA = Not available.

^aThe orientation is defined as described by the SNP Consortium (TSC).

Table S3: 46 autosomal SNPs detected in primates.

rs No.	Marker code	Primates						Human populations				P-values		
		Chimpanzee (29)	Orang-Utan (1)	Gorilla (3)	Baboon (2)	Rhesus (5)	Cynomolgous (6)	Ancestral alleles	Europeans (301)	Somalis (104)	Asian (295)	Europeans vs Somalis	Somalis vs Asian	Europeans vs Asian
rs1490413	1	G	G	G	G	G	G	G	0,58	0,49	0,43	-	-	-
rs876724	2	C	C	C	C	C	C	C	0,70	0,90	0,60	-	-	-
rs1357617	3	A	A	A	A	A	A	A	0,71	0,85	0,83	-	-	-
rs2046361	4			A				A	0,64	0,57	0,58	-	-	-
rs717302	5	G	G	G	A			G	0,44	0,51	0,18	-	-	P < 0.05
rs917118	7	T	T	T				T	0,30	0,54	0,33	-	-	-
rs763869	8	T	T	T	T	T	T	T	0,54	0,42	0,72	-	-	-
rs1015250	9	C	C	C				C	0,21	0,51	0,47	-	-	P < 0.05
rs735155	10	G	G	G				G	0,51	0,63	0,16	-	P < 0.001	P < 0.001
rs901398	11	T	T	T	T	T	T	T	0,64	0,63	0,73	-	-	-
rs2107612	12	A	A	A				A	0,69	0,74	0,86	-	-	-
rs1886510	13	T	T	T	T	T	T	T	0,47	0,09	0,17	P < 0.05	-	P < 0.01
rs1454361	14	A	A	A	A	A	A	A	0,51	0,60	0,51	-	-	-
rs740910	17	A	A		A	A	A	A	0,70	1,00	0,88	-	-	-
rs1493232	18	G	G	G	G	G	G	G	0,38	0,31	0,58	-	-	-
rs719366	19	T		T				T	0,61	0,81	0,77	-	-	-
rs1031825	20	C						C	0,72	0,73	0,57	-	-	-
rs722098	21	G	G					G	0,19	0,67	0,54	P < 0.01	-	P < 0.001
rs733164	22	G	G	G	G	G	G	G	0,70	0,52	0,76	-	-	-
rs826472	23	C	C	C				C	0,60	0,72	0,79	-	-	-
rs873196	25	C	C	C	T	C	C	C	0,41	0,33	0,13	-	-	P < 0.01
rs1382387	26	G	G	G	G	G	G	G	0,27	0,33	0,35	-	-	-
rs2111980	27	A	A	A	A	A	A	A	0,42	0,62	0,58	-	-	-
rs2056277	28	C/C, T C			T	T	T	T	0,24	0,25	0,18	-	-	-
rs1024116	29	C	C	C				C	0,44	0,61	0,81	-	-	P < 0.001
rs727811	30	C		C				C	0,46	0,65	0,33	-	-	-
rs1413212	32		A	G				A	0,30	0,21	0,41	-	-	-
rs938283	33	C	C	C	C	C	C	C	0,15	0,17	0,18	-	-	-
rs1979255	34	G	G	G				G	0,34	0,38	0,50	-	-	-
rs1463729	35	A	A	A	A			A	0,52	0,40	0,43	-	-	-
rs2076848	36	T	T	T	T	T	T	T	0,56	0,61	0,68	-	-	-
rs1355366	37	G	G	G				G	0,38	0,55	0,18	-	P < 0.05	-
rs354439	39	A	A	A	A	A	A	A	0,43	0,62	0,35	-	-	-
rs2040411	40	A		A	G			A	0,58	0,60	0,33	-	-	P < 0.05
rs737681	41	T	T	T	T	T	T	T	0,43	0,37	0,20	-	-	-
rs2830795	42	A	A			A		A	0,74	0,69	0,57	-	-	-
rs251934	43	C	C					C	0,34	0,29	0,10	-	-	P < 0.05
rs914165	44	G	G		G			G	0,59	0,31	0,68	-	P < 0.05	-
rs10495407	45	G	G		G	G	G	G	0,67	0,81	0,72	-	-	-
rs1360288	46	C	C		C	C	C	C	0,66	0,65	0,66	-	-	-
rs964681	48	C		C				C	0,37	0,20	0,27	-	-	-
rs1005533	49	A						A	0,55	0,40	0,33	-	-	-
rs8037429	50	C		C				C	0,52	0,50	0,59	-	-	-
rs891700	51	G	G	G				G	0,49	0,65	0,54	-	-	-
rs1335873	52	A	A	A	A	A	A	A	0,31	0,63	0,29	-	-	-
rs1028528	53	G		G				G	0,24	0,52	0,35	-	-	-
No. of loci amplified		44	37	36	23	22	21	Mean	0,48	0,53	0,48			

-: P > 0.05

Table S4: Frequencies of 52 autosomal SNP markers in various populations.

Marker code	Chromos- ome	NCBI rs No.	TSC alleles reported ^a	TSC No.	TSC											
					African- American ^b	Asian ^b	European ^b	Danish (156)	German (49)	Turkish (96)	Somali (104)	Greenland (149)	Thai (33)	Taiwanese (43)	Chinese (63)	Japanese (7)
1	1	rs1490413	A/G	TSC0724193	0,59/0,41	0,46/0,54	0,47/0,53	0,49/0,51	0,43/0,57	0,35/0,65	0,51/0,49	0,58/0,42	0,68/0,32	0,56/0,44	0,55/0,45	0,50/0,50
2	2	rs876724	C/T	TSC0208870	0,72/0,28	0,46/0,54	0,69/0,31	0,69/0,31	0,66/0,34	0,74/0,26	0,90/0,10	0,87/0,13	0,62/0,38	0,55/0,45	0,56/0,44	0,36/0,64
3	3	rs1357617	A/T	TSC0496080	0,85/0,15	0,80/0,20	0,65/0,35	0,73/0,27	0,70/0,30	0,71/0,29*	0,85/0,15	0,93/0,07	0,88/0,12	0,86/0,14	0,78/0,22	0,71/0,29
4	4	rs2046361	A/T	TSC1065282	0,65/0,35	0,70/0,30	0,80/0,20	0,71/0,29	0,65/0,35	0,55/0,45	0,57/0,43	0,67/0,33	0,50/0,50	0,53/0,47	0,56/0,44	0,64/0,36
5	5	rs717302	A/G	TSC0039610	0,92/0,08	0,83/0,17	0,45/0,55	0,53/0,47	0,52/0,48	0,64/0,36	0,49/0,51	0,91/0,09	0,80/0,20	0,80/0,20	0,80/0,20	0,79/0,21
6	6	rs1029047	A/T	TSC0253802	0,61/0,39	0,27/0,73	0,37/0,63	0,36/0,64	0,44/0,56	0,33/0,67	0,64/0,36	0,21/0,79	0,39/0,61	0,30/0,70	0,39/0,61	0,14/0,86
7	7	rs917118	C/T	TSC0229630	0,10/0,90	0,50/0,50	0,50/0,51	0,75/0,25	0,69/0,31	0,66/0,34	0,46/0,54	0,70/0,30	0,71/0,29	0,64/0,36	0,75/0,25	0,57/0,43
8	8	rs763869	C/T	TSC0065968	0,48/0,52	NA	0,56/0,44	0,49/0,51	0,55/0,45	0,34/0,66	0,58/0,42	0,15/0,85	0,44/0,56	0,37/0,63	0,37/0,63	0,07/0,93
9	9	rs1015250	C/G	TSC0097236	0,55/0,45	0,40/0,60	0,15/0,85	0,18/0,82	0,18/0,82	0,27/0,73	0,51/0,49	0,46/0,54	0,38/0,62	0,49/0,51	0,52/0,48	0,50/0,50
10	10	rs735155	A/G	TSC0027519	0,38/0,62	0,82/0,18	0,41/0,59	0,42/0,58	0,51/0,49	0,54/0,46	0,37/0,63	0,85/0,15	0,91/0,09	0,93/0,07	0,82/0,18	0,71/0,29
11	11	rs901398	C/T	TSC0177752	0,39/0,61	0,29/0,71	0,34/0,66	0,36/0,64	0,40/0,60	0,31/0,69	0,37/0,63	0,32/0,68	0,21/0,79	0,26/0,74	0,19/0,81	0,36/0,64
12	12	rs2107612	A/G	TSC1108144	0,64/0,36	0,75/0,25	0,70/0,30	0,76/0,24	0,65/0,35	0,67/0,33	0,74/0,26	0,72/0,28	0,91/0,09	0,84/0,16	0,90/0,10	0,93/0,07
13	13	rs1886510	C/T	TSC0904551	0,85/0,15	NA	0,50/0,50	0,42/0,58	0,53/0,47	0,64/0,36	0,91/0,09	0,73/0,27	0,77/0,23	0,87/0,13	0,91/0,09	0,86/0,14
14	14	rs1454361	A/T	TSC0684657	0,57/0,43	0,62/0,38	0,47/0,53	0,52/0,48	0,57/0,43	0,44/0,56	0,60/0,40	0,60/0,40	0,52/0,48	0,44/0,56	0,56/0,44	0,43/0,57
15	15	rs2016276	A/G	TSC0326920	0,91/0,09	0,59/0,41	0,79/0,21	0,77/0,23	0,69/0,31	0,78/0,22	0,84/0,16	0,52/0,48	0,58/0,42	0,64/0,36	0,62/0,38	0,43/0,57
16	16	rs729172	A/C	TSC0028090	0,23/0,77	0,06/0,94	0,38/0,62	0,46/0,54	0,34/0,66	0,38/0,62	0,32/0,68	0,18/0,82	0,14/0,86	0,06/0,94	0,11/0,89	0,07/0,93
17	17	rs740910	A/G	TSC0105771	0,92/0,08	0,92/0,09	0,57/0,43	0,67/0,33	0,72/0,28	0,69/0,31	1,00/0,00**	0,66/0,34	0,91/0,09	0,88/0,12	0,93/0,07	1,00/0,00**
18	18	rs1493232	A/C	TSC0729796	0,44/0,56	0,39/0,61	0,56/0,44	0,61/0,39	0,62/0,38	0,63/0,37	0,69/0,31	0,77/0,23	0,33/0,67	0,28/0,72	0,39/0,61	0,36/0,64
19	19	rs719366	C/T	TSC0044147	0,24/0,76	0,15/0,85	0,38/0,62	0,46/0,54	0,37/0,63	0,35/0,65	0,19/0,81	0,27/0,73	0,26/0,74	0,19/0,81	0,23/0,77	0,21/0,79
20	20	rs1031825	A/C	TSC0334834	0,60/0,40	0,41/0,59	0,30/0,70	0,24/0,76	0,28/0,72	0,32/0,68	0,27/0,73	0,31/0,69	0,41/0,59	0,45/0,55	0,40/0,60	0,57/0,43
21	21	rs722098	A/G	TSC0050288	0,17/0,83	0,50/0,50	0,86/0,14	0,81/0,19	0,85/0,15	0,78/0,22	0,33/0,67	0,40/0,60	0,50/0,50	0,50/0,50	0,46/0,54	0,43/0,57
22	22	rs733164	A/G	TSC0023085	0,26/0,74	0,16/0,84	0,34/0,66	0,32/0,68	0,34/0,66	0,24/0,76	0,48/0,52	0,47/0,53	0,18/0,82	0,16/0,84	0,12/0,88	0,29/0,71
23	10	rs826472	C/T	TSC0557086	0,70/0,30	0,66/0,34	0,44/0,56	0,57/0,43	0,61/0,39	0,62/0,38	0,72/0,28	0,85/0,15	0,85/0,15	0,87/0,13	0,87/0,13	0,50/0,50
24	21	rs2831700	A/G	NA	0,42/0,58	0,50/0,50	0,62/0,38	0,61/0,39	0,57/0,43	0,56/0,44	0,39/0,61	0,34/0,66	0,59/0,41	0,43/0,57	0,44/0,56	0,50/0,50
25	14	rs873196	C/T	TSC0202434	0,30/0,70	0,19/0,81	0,37/0,63	0,37/0,63	0,47/0,53	0,39/0,61	0,33/0,67	0,07/0,93	0,14/0,86	0,09/0,91	0,12/0,88	0,21/0,79
26	16	rs1382387	G/T	TSC0544547	0,55/0,45	0,40/0,60	0,25/0,75	0,32/0,68	0,26/0,74	0,24/0,76	0,33/0,67	0,36/0,64	0,27/0,73	0,34/0,66	0,34/0,66	0,43/0,57
27	12	rs2111980	A/G	TSC1113476	0,69/0,31	0,60/0,40	0,56/0,44	0,40/0,60	0,38/0,62	0,47/0,53	0,62/0,38	0,59/0,41	0,67/0,33	0,44/0,56	0,65/0,35	0,57/0,43
28	8	rs2056277	C/T	TSC1082757	0,70/0,30	0,75/0,25	0,55/0,45	0,74/0,26	0,69/0,31	0,84/0,16	0,75/0,25	0,55/0,45	0,89/0,11	0,88/0,12	0,91/0,09	0,86/0,14
29	18	rs1024116	A/G	TSC0247167	0,37/0,63	0,12/0,88	0,50/0,50	0,54/0,46	0,63/0,37	0,52/0,48	0,39/0,61	0,54/0,46	0,11/0,89	0,08/0,92	0,08/0,92	0,14/0,86
30	6	rs727811	A/C	TSC0062764	0,30/0,70	0,66/0,34	0,61/0,39	0,65/0,35	0,53/0,47	0,43/0,57	0,35/0,65	0,33/0,67	0,59/0,41	0,74/0,26	0,70/0,30	1,00/0,00**
32	1	rs1413212	A/G	TSC0607362	0,61/0,39	0,47/0,53	0,41/0,59	0,26/0,74	0,27/0,73	0,36/0,64	0,21/0,79	0,40/0,60	0,35/0,65	0,52/0,48	0,44/0,56	0,50/0,50
33	17	rs938283	C/T	TSC0357388	0,09/0,91	0,17/0,83	0,30/0,70	0,15/0,85	0,14/0,86	0,16/0,84	0,17/0,83	0,04/0,96**	0,38/0,62	0,14/0,86	0,16/0,84	0,21/0,79

34	4	rs1979255	TSC0925231	C/G	0,38/0,62	0,33/0,67	0,35/0,65	0,66/0,34	0,65/0,35	0,67/0,33	0,62/0,38	0,27/0,73	0,58/0,42	0,59/0,41	0,51/0,49	0,57/0,43
35	9	rs1463729	TSC0377760	A/G	0,79/0,21	0,48/0,52	0,56/0,44	0,54/0,46	0,52/0,48	0,51/0,49	0,40/0,60	0,56/0,44	0,48/0,52	0,35/0,65	0,45/0,55	0,29/0,71
36	11	rs2076848	TSC0022275	A/T	0,38/0,62	0,58/0,42	0,49/0,51	0,36/0,64	0,52/0,48	0,45/0,55	0,39/0,61	0,09/0,91	0,27/0,73	0,43/0,57	0,37/0,63	0,43/0,57
37	3	rs135366	TSC0491662	A/G	0,40/0,60	0,95/0,05	0,70/0,30	0,60/0,40	0,62/0,38	0,64/0,36	0,45/0,55	0,77/0,23	0,85/0,15	0,92/0,08	0,85/0,15	0,71/0,29
38	2	rs907100	TSC0186810	C/G	0,74/0,26	NA	0,48/0,52	0,29/0,71*	0,37/0,63	0,42/0,58*	0,71/0,29	0,57/0,43*	0,50/0,50	0,56/0,44	0,43/0,57	0,36/0,64
39	13	rs354439	TSC0700528	A/T	0,50/0,50	0,58/0,42	0,64/0,36	0,34/0,66	0,36/0,64	0,60/0,40	0,62/0,38	0,12/0,88	0,45/0,55	0,43/0,57	0,41/0,59	0,36/0,64
40	22	rs2040411	TSC11056845	A/G	0,75/0,25	0,08/0,92	0,71/0,29	0,64/0,36	0,57/0,43	0,54/0,46	0,60/0,40	0,41/0,59	0,35/0,65	0,24/0,76	0,22/0,78	0,43/0,57
41	7	rs737681	TSC0033074	C/T	0,56/0,44	0,80/0,20	0,58/0,42	0,60/0,40	0,51/0,49	0,61/0,39	0,63/0,37	0,71/0,29	0,83/0,17	0,92/0,08	0,86/0,14	0,71/0,29
42	21	rs2830795	NA	A/G	0,88/0,12	0,45/0,55	0,67/0,33	0,79/0,21	0,74/0,26	0,70/0,30	0,69/0,31	0,63/0,37	0,52/0,48	0,49/0,51	0,41/0,59	0,79/0,21
43	5	rs251934	TSC0220872	C/T	0,26/0,74	0,16/0,84	0,42/0,58	0,35/0,65	0,36/0,64	0,32/0,68	0,29/0,71	0,11/0,89	0,11/0,89	0,08/0,92	0,05/0,95	0,14/0,86
44	21	rs914165	TSC0197658	A/G	0,66/0,34	0,27/0,73	0,49/0,51	0,37/0,63	0,44/0,56	0,41/0,59	0,69/0,31	0,52/0,48	0,26/0,74	0,22/0,78	0,30/0,70	0,29/0,71
45	1	rs10495407	NA	A/G	0,11/0,89	0,42/0,58	0,29/0,71	0,32/0,68	0,31/0,69	0,35/0,65	0,19/0,81	0,25/0,75	0,36/0,64	0,34/0,66	0,30/0,70	0,14/0,86
46	9	rs1360288	TSC0501229	C/T	0,70/0,30	0,50/0,50	0,50/0,50	0,70/0,30	0,64/0,36	0,65/0,35	0,65/0,35	0,78/0,22	0,62/0,38	0,58/0,42	0,70/0,30	0,64/0,36
48	10	rs964681	TSC0270699	C/T	0,37/0,63	0,23/0,77	0,41/0,59	0,44/0,56	0,32/0,68	0,36/0,64	0,20/0,80	0,27/0,73	0,29/0,71	0,36/0,64	0,28/0,72	0,14/0,86
49	20	rs1005533	TSC0082071	A/G	0,55/0,45	0,60/0,40	0,45/0,55	0,55/0,45	0,62/0,38	0,48/0,52	0,40/0,60	0,49/0,51	0,29/0,71	0,29/0,71	0,35/0,65	0,21/0,79
50	15	rs8037429	NA	C/T	0,58/0,42	0,42/0,58	0,62/0,38	0,50/0,50	0,51/0,49	0,54/0,46	0,50/0,50	0,49/0,51	0,65/0,35	0,59/0,41	0,61/0,39	0,57/0,43
51	1	rs891700	TSC0162577	A/G	0,53/0,47	0,53/0,47	0,57/0,43	0,54/0,46	0,59/0,41	0,39/0,61	0,35/0,65	0,49/0,51	0,50/0,50	0,44/0,56	0,52/0,48	0,36/0,64
52	13	rs1335873	TSC0829150	A/T	0,91/0,09	0,28/0,72	0,36/0,64	0,26/0,74	0,27/0,73	0,40/0,60	0,63/0,37	0,14/0,86	0,36/0,64	0,40/0,60	0,35/0,65	0,21/0,79
53	22	rs1028528	TSC0253071	A/G	0,35/0,65	0,65/0,35	0,74/0,26	0,78/0,22	0,73/0,27	0,78/0,22	0,48/0,52	0,72/0,28	0,61/0,39	0,66/0,34	0,69/0,31	0,57/0,43
54	15	rs1528460	TSC0798410	C/T	0,65/0,35	0,36/0,64	0,29/0,71	0,32/0,68	0,34/0,66	0,28/0,72	0,72/0,28	0,29/0,71	0,32/0,68	0,41/0,59	0,40/0,60	0,29/0,71

NA = Not available,

aLetters in bold indicate the ancestral alleles based on primate typing results.

bFrequencies reported in The SNP Consortium (TSC), Celera or RealSNP databases using at least 24 individuals. The bold numbers are frequencies only available in TSC.

*p < 0,05 for Hardy-Weinberg equilibrium tested with Chi-square test or Fisher's exact test in case of small dataset and sequential Bonferroni correction (Selvin, 1998).

**Markers with very low minor allele frequency.

Table S5: Analysis of variance (F-statistics) for the 52 SNPs in the populations studied.

Locus	European (Fst)	Somali (Fis)	Asian (Fst)	All (Fst)	Statistic analysis	
					European	
rs1490413	0,019	0,023	-0,001	0,024	No, of markers	52
rs876724	0,001	-0,102	0,167	0,092	Mean	0,01
rs1357617	-0,005	0,045	0,051	0,049	Median	0,005
rs2046361	0,032	-0,036	0,018	0,019	Std, deviation	0,018
rs717302	0,011	0,004	0,031	0,136	Minimum	-0,001
rs1029047	0,003	0,025	0,038	0,081	Maximum	0,090
rs917118	0,008	0,038	-0,001	0,038	Skewness (G1)	2,479
rs763869	0,038	-0,003	0,112	0,113	Somali	
rs1015250	0,009	0,254	-0,001	0,090	No, of markers	52
rs735155	0,015	-0,073	0,014	0,217	Mean	0,028
rs901398	0,001	0,118	0,018	0,014	Median	0,025
rs2107612	0,015	-0,009	0,056	0,029	Std, deviation	0,085
rs1886510	0,052	0,020	0,047	0,171	Minimum	-0,002
rs1454361	0,007	0,103	0,011	0,011	Maximum	0,254
rs2016276	0,003	-0,120	0,009	0,075	Skewness (G1)	0,263
rs729172	0,011	-0,074	0,014	0,099	Asian	
rs740910	-0,004	0,000	0,136	0,111	No, of markers	52
rs1493232	-0,005	-0,079	0,258	0,100	Mean	0,057
rs719366	0,011	-0,037	-0,003	0,048	Median	0,028
rs1031825	0,005	-0,083	0,016	0,018	Std, deviation	0,074
rs722098	0,001	0,072	0,001	0,186	Minimum	-0,001
rs733164	0,006	-0,017	0,158	0,071	Maximum	0,306
rs826472	-0,001	-0,095	0,026	0,084	Skewness (G1)	1,602
rs2831700	-0,001	-0,002	0,034	0,046	All populations combined	
rs873196	0,002	0,174	0,005	0,120	No, of markers	52
rs1382387	0,004	0,029	-0,003	0,003	Mean	0,073
rs2111980	0,005	0,070	0,019	0,041	Median	0,078
rs2056277	0,023	0,082	0,200	0,084	Std, deviation	0,049
rs1024116	0,005	0,079	0,306	0,152	Minimum	0,003
rs727811	0,059	0,028	0,201	0,110	Maximum	0,217
rs1413212	0,013	-0,033	0,004	0,038	Skewness (G1)	0,709
rs938283	-0,005	0,077	0,147	0,046		
rs1979255	-0,005	0,086	0,110	0,105		
rs1463729	-0,005	0,086	0,030	0,015		
rs2076848	0,020	-0,002	0,153	0,088		
rs1355366	-0,004	0,013	0,022	0,097		
rs907100	0,015	0,239	0,009	0,081		
rs354439	0,090	0,086	0,163	0,146		
rs2040411	0,009	0,063	0,035	0,090		
rs737681	0,003	-0,073	0,057	0,057		
rs2830795	0,007	-0,047	0,047	0,058		
rs251934	-0,003	0,021	0,001	0,084		
rs914165	-0,001	0,147	0,091	0,081		
rs10495407	-0,003	-0,172	0,005	0,010		
rs1360288	-0,001	0,070	0,034	0,012		
rs964681	0,009	0,063	0,001	0,030		
rs1005533	0,010	0,006	0,042	0,036		
rs8037429	-0,003	0,024	0,010	0,004		
rs891700	0,035	0,139	-0,005	0,020		
rs1335873	0,024	0,077	0,089	0,121		
rs1028528	-0,003	0,042	0,001	0,048		
rs1528460	-0,001	0,096	0,006	0,099		
Overall	0,011	0,033	0,059	0,074		

Tabelle A-1: Richtlinien zur manuellen Auswertung der 52 SNP Multiplexreaktion unterteilt in zwei SNP Sets (23plex und 29plex).

First 23 Autosomal SNP SBE set				
AUTO 1:	Code:	Peak heights should be:		
rs1490413	A1	C	=	T
rs876724	A2	C	=	T
rs1357617	A3	1/2-1/3 A	=	T
rs2046361	A4	1/2-2/3 A	=	T
rs717302	A5	1/2-1/3 G	=	A
rs1029047	A6	1/3-2/3 A	=	T
rs917118	A7	2/3 G	=	A
rs763869	A8	C	=	T
rs1015250	A9	1/2 G	=	C
rs735155	A10	2/3 C	=	T
rs901398	A11	C	=	T
rs2107612	A12	1/2 G	=	A
rs1886510	A13	1/2 G	=	A
rs1454361	A14	1/2-2/3 A	=	T
rs2016276	A15	C	=	T
rs729172	A16	1/4 G	=	T
rs740910	A17	1/2-1/3 G	=	A
rs1493232	A18	1/2 G	=	T
rs719366	A19	C	=	T
rs1031825	A20	1/4 G	=	T
rs722098	A21	1/2 G	=	A
rs733164	A22	1/3-1/2 G	=	A
rs826472	A23	T/A	=	A
Second 29 Autosomal SNP SBE set				
AUTO 2:	Code:	Peak heights should be:		
rs2831700	A24	C	=	T
rs873196	A25	1/2-1/3 G	=	A
rs1382387	A26	1/4 G	=	T
rs2111980	A27	G	=	A
rs2056277	A28	2/3 G	=	A
rs1024116	A29	1/2-2/3 G	=	A
rs727811	A30	1/2 A	=	C
rs1413212	A32	C	=	T
rs938283	A33	C	=	2/3-1/1 T
rs1979255	A34	1/3-1/4 G	=	C
rs1463729	A35	2/3 G	=	A
rs2076848	A36	1/2 A	=	T
rs1355366	A37	C	=	T
rs907100	A38	1/2-1/3 G	=	C
rs354439	A39	1/2 A	=	T
rs2040411	A40	1/2-2/3 G	=	A
rs737681	A41	C	=	T
rs2830795	A42	2/3-1/1 G	=	A
rs251934	A43	1/2 G	=	A
rs914165	A44	C	=	2/3 T
rs10495407	A45	C	=	T
rs1360288	A46	2/3 G	=	A
rs964681	A48	C	=	1/2-2/3 T
rs1005533	A49	1/3-1/2 G	=	A
rs8037429	A50	2/3 G	=	A
rs891700	A51	1/2 G	=	A
rs1335873	A52	A	=	T
rs1028528	A53	A	=	G
rs1528460	A54	1/4-1/3 G	=	A

Observations in general	
Alleles	Ratio
C/T	1:1
A/T	2:1
G/A	2:1 or 3:1
G/T	4:1

Tabelle A-2: In der 23plex Reaktion detektierten Genotypen aller Populationsstichproben aus dem Hauskollektiv des Instituts für Rechtsmedizin der Johannes Gutenberg Universität, Mainz. Chinesen (B): n = 57, Japaner (JPN): n = 40, Thailänder (T): n = 49, Deutsche (Dt): n = 64.

SNP Ind.	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23
B1	CT	CC	TT	TT	AA	AT	GG	CT	GC	TT	CT	AA	GG	AT	CT	GG	AA	TG	TT	GG	GG	AG	GG
B2	CT	CT	TT	AT	AA	AT	AA	TT	GC	TT	TT	AA	GG	TT	CT	TG	AA	TG	TT	TG	GG	GG	GG
B3	CT	CC	AT	AA	AG	AT	AG	TT	GC	CT	CC	AG	GG	AT	CT	TG	AA	GG	TT	TG	AG	GG	GG
B4	CT	CT	TT	TT	AA	TT	GG	CT	GC	TT	CT	AA	GG	AT	CT	TG	AA	GG	TT	GG	AG	GG	GG
B6	CT	CT	TT	TT	AA	AA	AG	TT	GC	TT	CT	AA	GG	AT	TT	GG	AA	TG	CT	GG	GG	GG	AG
B7	CT	TT	TT	AT	AA	AT	AG	CT	GC	TT	TT	AA	GG	AA	TT	GG	AA	GG	CT	GG	AG	GG	GG
B8	CT	CT	TT	TT	GG	TT	GG	CC	GC	CT	TT	AG	GG	AA	CT	GG	AA	TG	TT	TG	AA	AG	GG
B9	CT	CT	TT	AT	AA	AA	AG	TT	GG	TT	TT	AA	GG	AA	CT	GG	AA	TG	CT	GG	GG	AG	GG
B10	CC	CT	AT	AT	AG	AT	GG	TT	GG	CT	TT	AA	GG	AT	CT	TG	AG	TT	TT	GG	AG	AG	AG
B11	TT	CC	AT	AT	AA	AT	GG	CT	GC	TT	CT	AG	GG	AT	CT	GG	AA	GG	TT	TG	GG	GG	GG
B13	CC	CT	AT	AA	AA	AT	AG	TT	CC	TT	TT	AG	AG	AA	CT	TG	AA	TT	CT	TG	AA	GG	AA
B14	TT	CT	AT	TT	AA	AT	GG	CT	GG	CT	CT	AA	AG	AT	CT	GG	AA	TG	TT	TT	AA	GG	GG
B15	TT	CC	AT	AA	AA	AT	AG	CC	GC	TT	TT	AA	GG	AT	TT	GG	AA	TG	TT	TT	AA	AG	GG
B16	CC	TT	TT	AT	AA	AT	AA	CT	GG	TT	CT	AA	GG	AA	CT	GG	AG	TT	TT	TG	AA	GG	AG
B17	CT	CC	AA	AT	GG	TT	AG	TT	CC	TT	TT	AA	GG	AT	CT	GG	AA	GG	CT	TT	GG	GG	GG
B18	CC	CC	TT	AT	AG	TT	GG	CT	GG	TT	TT	AA	GG	TT	TT	GG	GG	TG	CT	TT	AG	GG	GG
B21	CT	CC	AT	TT	AA	AT	GG	TT	CC	CT	TT	AA	GG	AT	CT	GG	AA	TG	CT	GG	AG	GG	AG
B22	CT	TT	AT	AT	AA	AT	AG	CT	GC	TT	TT	AG	GG	TT	TT	TG	AA	GG	TT	GG	AA	AG	GG
B23	CT	TT	AT	AT	AA	AT	AG	CT	GC	TT	TT	AG	GG	TT	TT	TG	AA	GG	TT	GG	AA	AG	GG
B26	TT	CC	AT	TT	AA	TT	GG	CT	GG	TT	CT	AA	GG	TT	TT	GG	AA	GG	TT	GG	GG	GG	GG
B27	TT	CT	AT	AT	AA	AA	GG	CT	GG	CT	CT	AA	GG	AA	TT	GG	AA	TT	CT	TG	AG	AG	GG
B29	CC	TT	TT	TT	AA	AT	AG	TT	GC	CT	CT	AA	GG	AT	TT	GG	AG	TG	TT	GG	AG	GG	GG
B30	CC	CC	TT	AA	AA	AT	GG	CT	GC	CT	TT	AA	AG	TT	TT	GG	AA	TG	CT	TT	GG	AG	GG
B31	CT	TT	TT	AT	AG	AT	GG	CC	CC	CT	TT	AA	AA	AA	CT	TG	AA	TG	TT	TG	AG	GG	GG
B34	CT	CT	AT	TT	AA	TT	GG	TT	GC	TT	TT	AA	GG	AT	CC	GG	AA	TG	TT	TG	AG	GG	AG
B37	TT	CT	TT	TT	AA	TT	AG	CC	CC	TT	TT	AG	GG	AA	TT	GG	AA	TG	TT	TG	AG	AG	GG
B38	TT	CC	TT	AA	AG	AT	AG	CT	CC	CT	TT	AA	AG	AT	CC	GG	AA	GG	CT	TG	AG	GG	GG
B39	CC	CC	TT	AA	AA	AT	AG	TT	GG	TT	TT	AA	GG	AA	TT	GG	AA	TG	TT	GG	GG	AG	GG
B40	TT	CT	AT	AT	AA	TT	GG	TT	CC	CC	TT	AA	GG	AT	CT	GG	AA	TG	TT	TT	AG	GG	GG
B41	CT	CC	AT	AA	AA	TT	GG	CT	GC	CT	TT	AA	GG	AA	CT	GG	AA	TG	TT	TG	AA	GG	GG
B42	CT	TT	TT	AT	GG	TT	AG	TT	GC	TT	CT	AA	GG	AT	TT	TG	AA	GG	CT	TG	GG	AG	GG
B44	CC	CT	AT	AT	AG	AT	GG	TT	GC	TT	CT	AA	GG	AT	CC	GG	AG	TG	CT	TG	AG	GG	GG
B45	CT	CT	TT	AT	AG	AA	GG	TT	GC	CT	CT	AA	GG	AT	CC	GG	AA	TG	CT	GG	AA	GG	GG
B47	CT	CT	AT	AT	AA	TT	AA	TT	GG	TT	TT	AA	GG	TT	TT	GG	AA	GG	CT	TG	AG	GG	GG
B48	CT	CC	TT	AT	AA	AA	GG	TT	GC	TT	TT	AA	GG	AA	CC	GG	AA	TG	TT	GG	AG	GG	GG
B49	TT	CT	AT	AT	AA	TT	AG	TT	GG	TT	CT	AA	AG	TT	TT	GG	AA	GG	CT	TG	GG	GG	AG
B50	CT	CT	TT	AT	AA	TT	GG	CC	GG	TT	TT	AA	GG	AT	TT	TG	AA	TT	CT	GG	AG	GG	GG
B52	TT	TT	AT	AT	AA	AT	AA	TT	GC	CT	CT	AA	GG	AT	CT	GG	AA	TG	CT	TG	AG	GG	GG
B56	CT	CT	TT	AA	AA	TT	GG	CC	GC	CT	TT	AA	GG	AA	CT	TT	AA	GG	CT	TG	AA	GG	GG
B58	TT	CC	TT	AT	AA	AA	GG	TT	GC	TT	TT	AA	GG	AA	CC	TG	AA	GG	CT	TG	AG	GG	GG
B59	TT	CT	TT	TT	AA	AA	GG	CC	CC	TT	CT	AG	GG	AT	TT	GG	AA	GG	CT	GG	GG	GG	GG
B61	CT	CT	AT	AT	GG	AT	GG	CC	GC	TT	TT	AA	GG	TT	CC	GG	AA	TT	CT	TG	AG	GG	GG
B62	CC	CC	AA	TT	AA	TT	GG	CT	GC	TT	CT	AA	AG	AT	CT	TG	AA	GG	TT	TG	AA	AG	GG
B63	TT	CT	AT	TT	AA	TT	GG	CT	GC	CT	TT	AG	GG	AT	CC	GG	AA	GG	TT	TG	AG	GG	GG
B66	CC	CT	TT	AA	GG	AT	GG	CT	GC	CT	TT	AA	GG	AA	CT	GG	AA	TG	TT	TG	AA	GG	AG
B69	CC	CC	TT	AT	AG	TT	GG	TT	GC	TT	TT	AA	GG	AA	TT	GG	AA	TG	TT	TT	AG	GG	GG
B71	TT	CT	AT	AA	AA	TT	AA	CT	GG	TT	CT	AG	AA	TT	TT	GG	AA	GG	CC	TG	AG	AG	AG
B73	CT	CT	AT	AT	AA	AT	GG	CT	GC	CC	TT	AA	AG	AT	CT	GG	AA	TG	CT	TG	GG	AG	AG
B75	CC	CC	TT	AT	AA	AA	GG	TT	GC	TT	TT	AA	GG	AT	CC	GG	AA	TG	CT	TG	GG	AG	GG
B79	TT	CT	TT	TT	AA	TT	GG	TT	GG	TT	TT	AA	GG	AT	CC	TG	AA	TG	TT	GG	AA	GG	AG
B80	CT	CC	AT	AT	AA	TT	GG	TT	GC	CT	CC	AA	GG	AT	TT	GG	AG	TG	TT	TG	AG	GG	GG
B81	TT	CC	TT	AT	AA	AA	GG	CT	GC	CT	TT	AA	GG	AT	CT	GG	AA	GG	TT	GG	GG	GG	GG
B83	CC	CT	AT	TT	AG	AT	GG	CC	GC	TT	TT	AG	GG	AT	TT	GG	AA	TG	CT	TT	AG	GG	GG

SNP Ind.	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23
B84	CT	TT	TT	AT	AA	TT	AG	TT	CC	TT	TT	AA	GG	AT	TT	TG	AA	TG	TT	TT	AG	GG	GG
B86	CC	TT	TT	AA	AA	AT	AA	TT	GG	CT	TT	AA	GG	AT	TT	TG	AA	TG	TT	GG	AG	GG	GG
B87	TT	TT	AT	TT	AG	AT	GG	CT	GC	CT	CC	AA	GG	AA	TT	GG	AA	GG	TT	TG	AG	GG	AA
B89	CT	CT	TT	AT	AG	AT	AG	CT	CC	TT	TT	AA	GG	TT	CT	GG	AA	TG	TT	GG	GG	GG	GG
B90	TT	TT	AT	AT	AA	AT	GG	CC	CC	TT	TT	AG	GG	AT	CT	TG	AA	GG	CT	TG	GG	GG	GG
B91	CC	CT	TT	AT	AA	AA	GG	CT	GG	TT	TT	AA	GG	AT	CC	TG	AG	TT	TT	TG	AG	GG	GG
B92	CT	TT	AT	TT	AA	TT	AG	TT	GG	TT	TT	AA	GG	AT	CT	GG	AA	GG	CT	TG	AG	GG	GG
B95	CT	CC	TT	AA	AA	AA	GG	CT	GC	TT	TT	AA	GG	AA	TT	GG	AG	GG	TT	TT	AG	GG	GG
B96	CT	CC	TT	AA	AG	TT	AA	TT	GC	CT	TT	AA	GG	AT	CT	GG	AG	TT	TT	GG	AG	GG	AG
B97	CT	CC	AT	TT	AA	AT	GG	CC	GC	CT	TT	AG	GG	AT	CT	GG	AA	TG	CT	GG	GG	GG	GG
B98	TT	CT	TT	AT	AA	AT	AG	CT	GG	TT	CT	AA	GG	AT	TT	GG	AA	TG	TT	TG	AG	AG	AG
B99	CT	CC	AT	TT	AA	TT	GG	TT	CC	TT	CT	GG	GG	AT	CT	GG	AA	GG	TT	TG	AG	GG	AG
JPN1	TC	TT	TT	TT	AA	AT	AG	CC	CC	TC	TT	AA	AG	AT	TC	GG	AA	TT	TT	GG	GG	AA	AG
JPN2	CC	TC	TT	AT	AG	TT	GG	TT	CC	TT	TC	AA	GG	AT	TC	TG	AA	TT	TC	TG	GG	GG	AG
JPN3	CC	TC	AA	TT	AG	TT	GG	TT	GG	TT	TC	AA	GG	AT	TT	GG	AG	GG	TT	TT	AA	AG	GG
JPN4	TC	CC	TT	AT	AA	AT	GG	TC	CC	TT	TC	AA	GG	AA	TT	TG	GG	TG	TT	TG	AA	GG	GG
JPN5	CC	TT	TT	TT	AA	TT	AG	TT	CG	TT	TT	AA	GG	AA	CC	GG	AA	GG	TT	GG	AG	GG	AG
JPN6		CC	TT	TT	AA	TT	AA	TT	GG	TT	TC	AG	AA	AT	CC	TG	AG	GG	TT	TT	AA	GG	AG
JPN7	TT	TT	TT	AT	AG	TT	AG	TC	GG	TC	TC	AA	AG	AT	CC	GG	AA	GG	TC	TG	AG	GG	AG
JPN8	TT	CC	TT	AT	AA	AT	GG	TC	CG	TT	TC	GG	AG	AT	TC	GG	AA	TG	TC	TG	AA	GG	AA
JPN9	TC	TC	TT	AT	AA	TT	AG	TT	GG	TT	TT	AA	GG	AT	TC	TG	AG	GG	TC	TG	AG	GG	GG
JPN10	CC	TC	AT	TT	AA	AT	AG	TC	CG	TC	TT	GG	GG	AA	TT	GG	AA	TG	TT	GG	AG	AG	GG
JPN11	TC	TC	TT	AT	AA	AT	AG	TT	CG	TT	TC	AG	GG	AT	TT	GG	AA	GG	TT	TG	AG	AG	AG
JPN12	TT	CC	AT	AT	AA	TT	AG	TT	CG	TT	CC	AA	GG	TT	TC	GG	AA	GG	TT	TG	GG	AG	GG
JPN13	TC	TC	TT	TT	AA	TT	GG	TT	CG	TT	TT	AA	AG	AT	TT	GG	AA	TT	CC	TG	AG	GG	AG
JPN14	TT	TC	TT	TT	AA	TT	AA	TT	CG	TT	TT	AG	AG	AT	TC	TG	AA	TG	TT	TG	AA	GG	AA
JPN15	TC	TT	TT	AT	AA	AT	GG	TT	GG	CC	TC	AA	AG	AT	TT	TG	AA	TG	TT	TG	AA	GG	GG
JPN16	TC	CC	TT	AT	AA	TT	AG	TC	CC	TC	TT	AA	GG	TT	TC	GG	AA	TG	TC	TG	AA	GG	AG
JPN17	TC	TT	TT	AA	AA	AA	GG	TC	GG	TC	TT	AA	AG	AA	TC	GG	AA	GG	TT	GG	GG	GG	GG
JPN18	TC	TC	AT	AT	AA	TT	GG	CC	CG	TT	TT	AA	AG	AT	CC	GG	GG	TG	TT	TG	GG	AG	AA
JPN19	CC	TC	AT	TT	AA	AA	AG	CC	CC	TT	TT	AA	GG	AA	TC	GG	AA	TT	CC	TG	GG	GG	AG
JPN20	TT	TC	AT	AT	AA	TT	AG	TC	CG	TC	TT	AA	AG	AT	CC	GG	AA	GG	TT	TG	GG	AG	GG
JPN21	TT	TC	TT	AA	AG	TT	GG	TT	CC	TC	TC	AG	GG	AA	TT	GG	GG	TG	TT	TT	AG	AG	GG
JPN22	TC	TT	TT	AT	AA	TT	GG	TT	CC	TC	TC	AA	GG	AT	TC	GG	AA	GG	TT	TT	GG	GG	GG
JPN23	CC	TC	AA	AA	AG	TT	AG	TT	CC	TT	TC	AA	AG	AT	TC	GG	AA	TG	TC	GG	AA	GG	AG
JPN24	TC	CC	TT	AT	AA	AT	AA	TC	CG	TC	TC	AA	GG	AT	TC	GG	AA	TG	TC	GG	AG	GG	AG
JPN25	TT	CC	AT	AT	AG	TT	GG	TT	CG	TC	TC	AA	AG	AA	TT	TG	AA	TG	TC	TT	AG	GG	AG
JPN26	TT	TC	TT	AT	AA	TT	AG	TT	CG	TT	TC	AA	GG	AT	TC	GG	AA	GG	TT	TG	AG	GG	AG
JPN27	CC	TT	AT	TT	AA	TT	AG	TT	CG	TC	TC	AA	GG	TT	TT	GG	AA	GG	TT	TG	GG	AG	GG
JPN28	TC	TT	AT	AT	AA	AT	AG	TC	CG	TC	TC	AG	GG	AT	TC	TG	AG	TG	TT	GG	AG	AG	AG
JPN29	TT	TC	TT	AT	AA	TT	GG	TT	GG	TT	TC	AG	GG	AT	TC	TG	AA	TT	TC	GG	AG	AG	GG
JPN30	CC	TC	AT	TT	AA	AT	GG	TT	CG	TT	TT	AA	AG	AT	TC	TG	AG	GG	TT	GG	AG	GG	AG
JPN31	CC	TC	TT	AT	AA	AT	AG	TT	CG	TT	TC	AA	AG	AT	TC	GG	AA	TG	TC	TG	AA	GG	GG
JPN32	TT	TC	AT	AT	AG	AT	GG	TC	CG	TT	TT	AA	AA	AT	TT	GG	AA	TG	TC	TG	AG	GG	GG
JPN33	TC	CC	AT	TT	AA	AT	AG	CC	CG	TT	TC	AA	GG	TT	TC	TG	AG	TG	TC	TG	AG	AG	GG
JPN34	TC	TT	TT	AT	AA	TT	GG	TC	GG	TT	TT	AA	AG	AT	CC	GG	AA	TT	TC	TT	AG	GG	GG
JPN35	TC	TC	AT	AT	AA	AT	AG	TT	CG	TT	CC	AA	AG	AT	TC	GG	AG	GG	TC	TG	AG	AG	GG
JPN36	TC	TT	TT	AA	AG	TT	GG	TC	GG	TT	TT	AA	AG	AT	TT	GG	AA	TG	TC	TG	AA	AG	GG
JPN37	TC	TC	TT	TT	AA	AA	GG	TT	CC	TC	TC	AG	GG	AA	TC	GG	AG	TG	TC	TG	AG	AG	GG
JPN38	TC	TT	AT	TT	AG	AT	AG	TC	CG	TT	TC	AA	GG	AA	CC	TG	AA	GG	TT	TG	AA	GG	AG
JPN39	CC	TC	TT	TT	AA	AT	GG	TC	CC	TC	TT	AA	GG	TT	TT	GG	AG	TG	TT	GG	AG	GG	AG
JPN40	TC	TT	TT	AA	AA	TT	AA	TT	GG	TT	TT	AG	GG	AT	CC	GG	AG	GG	CC	GG	AG	GG	AA
T1	TT	CT	TT	AT	AG	TT	AG	TT	CC	TT	TT	AG	GG	AA	CT	TG	AA	TT	TT	GG	AG	AG	GG
T3	CC	CT	AT	AT	AG	AT	AG	CT	GG	CT	TT	AG	GG	AA	CT	TG	AA	TG	CT	TG	AG	GG	AG
T10	CC	CT	AT	AT	AG	AT	AG	CT	GG	CT	TT	AG	GG	AA	CT	TG	AA	TG	CT	TG	AG	GG	AG
T11	CC	CC	TT	AT	AG	TT	AA	TT	GC	CT	TT	AA	GG	AA	TT	GG	AA	TG	CC	TG	GG	GG	GG
T12	CC	CC	TT	AT	AG	TT	GG	TT	CC	TT	TT	AA	GG	AT	CT	GG	AA	TG	CC	TG	GG	AG	AG

SNP Ind.	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23
T13	TT	CT	AT	TT	AA	AT	GG	TT	CC	CC	CT	AA	GG	TT	CT	TG	AA	TG	CT	GG	GG	GG	AG
T15	CC	CC	AT	AT	GG	AT	GG	TT	GC	TT	CT	AG	GG	AT	CC	GG	AA		TT	TG	GG	AG	GG
T16	TT	CC	TT	TT	AA	TT	GG	CC	CC	TT	TT	AA	AA	AA	CT	GG	AA	TG	TT	GG	AA	AG	AG
T20	TT	CT	AT	AA	GG	AA	AA	CC	GG	TT	CT	AA	GG	AA	CT	GG	AA	TT	CT	TT	AG	GG	GG
T21	CT	CT	TT	TT	AA	TT	GG	CT	CC	TT	CT	AG	GG	AT		GG	AA		CT	TG	AG	GG	AA
T22	TT	CT	AT	AT	AA	AT	GG	CC	CC	TT	TT	AA	GG	TT	CT	TG	AA	TG	CT	TG	AG	GG	GG
T23	CT	CC	TT	AT	AA	AT	GG	CT	GC	TT	TT	AA	GG	AT	CC	GG	AA	TG	TT	TG	GG	GG	AG
T24	TT	CT	TT	AT	AG	AT	GG	CT	GC	TT	CT	AA	GG	AA	CC	TG	AA	TG	TT	TT	AG	GG	GG
T26	CT	CC	AT	AT	AA	AT	GG	CT	GC	TT	TT	AG	GG	AA	TT	GG	AA	TG	CT	GG	AG	GG	GG
T27	CC	CT	AT	TT	AA	TT	AG	CC	GC	TT	CT	AA	GG	AA	CC	TG	AA	GG	CC	TG	AA	GG	GG
T31	CT	CT	AT	AT	AG	TT	GG	CC	GC	TT	TT	AA	GG	AT	CT	GG	AG	TT	TT	GG	AG	GG	GG
T34	CC	CT	AT	TT	AG	TT	GG	TT	GC	TT	TT	AG	GG	AT	TT	TG	AA	TG	CT	TG	GG	GG	AG
T71	TT	CC	AT	AT	AA	TT	AG	TT	GC	TT	TT	AA	AG	AT	CT	GG	AG	TG	TT	TG	AA	GG	AA
T85	CT	TT	TT	AT	AA	AT	AG	CC	CC	TT	TT	AA	GG	AT	CC	TG	AA	GG	TT	GG	AG	GG	GG
T86	TT	CT	TT	AT	AA	TT	GG	CT	CC	TT	TT	AG	GG	TT	TT	GG	AA	TG	TT	TT	GG	AA	AA
T87	TT	CC	AA	AT	AA	AA	AA	CT	GG	TT	TT	AG	GG	AT	TT	GG	GG	TG	CT	TG	AG	GG	GG
T88	CT	CC	AT	TT	AA	TT	GG	CT	CC	TT	CT	AA	GG	AA	CT	GG	AA	GG	CT	TG	GG	AG	GG
T89	CT	TT	AT	AT	AA	AA	AG	CC	GG	CT	CT	AA	GG	AT	CT	GG	AA	TG	TT	TG	AG	GG	AG
T92	TT	CT	TT	AT	AA	AT	AG	TT	CC	CT	CT	AA	GG	AT	CT	GG	AA	TG	CC	GG	AG	GG	AG
T93	TT	CT	TT	AT	AA	AT	GG	TT	CC	TT	TT	AA	GG	AA	CT	GG	AA	TG	CT	TG	AA	AG	GG
T95	TT	CC	TT	AT	AG	TT	AG	CT	CC	TT	TT	AA	AG	AT	CT	GG	AA	TG	CT	TG	AA	GG	GG
T100	TT	CT	TT	TT	AA	AA	GG	TT	CC	TT	CT	AA	GG	AT	CT	GG	AA	TG	CT	TG	GG	GG	GG
T108	TT	CT	AT	AT	AG	TT	GG	CT	CC	CT	TT	AA	GG	AT	TT	GG	AA	TG	CT	GG	AG	AG	GG
T109	CC	CC	AT	AA	AG	AT	AG	CC	GC	TT	CT	AA	AG	TT	TT	TG	AA	TG	TT	TG	AG	AG	AG
T112	TT	CC	AT	AT	AG	AA	AG	TT	GC	TT	TT	AA	GG	AT	CC	TG	AA	GG	TT	GG	AG	GG	GG
T121	CT	CT	AT	AT	AG	TT	GG	CT	GG	TT	TT	AA	GG	TT	TT	GG	AA	?G	TT	TG	AG	GG	GG
T124	TT	CC	TT	AT	AA	AT	GG	CC	CC	TT	TT	AA	AG	AT	CT	GG	AA	TG	CC	TG	AG	GG	AG
T128	CT	CT	TT	AT	AA	AA	AG	CC	GG	TT	CT	AA	GG	AT	CT	GG	AA	TG	CT	TG	AA	AG	AG
T136	TT	TT	TT	AT	GG	AT	AA	CC	GC	TT	TT	AG	AG	AT	CT	GG	AA	TG	CT	TG	AG	GG	GG
T138	CT	CT	TT	AA	AG	TT	AG	TT	GC	CT	CT	AA	GG	AA	TT	GG	AA	GG	TT	TG	AA	AG	AG
T139	TT	TT	TT	AT	AG	AT	GG	TT	GC	TT	TT	AA	GG	AA	CT	TG	AA	TG	TT	GG	AG	AG	GG
T142	CC	CC	TT	TT	AA	TT	AG	CC	GC	CT	CT	AA	GG	AA	CT	GG	AA	TT	CT	TT	AG	AA	AA
T147	TT	TT	AT	AT	AA	TT	GG	TT	GG	CT	TT	AA	GG	AA	CT	TG	AA	GG	CT	TG	AG	GG	GG
T152	TT	CT	TT	AA	AA	AT	AA	CC	CC	TT	TT	AA	GG	AA	CT	TT	AA	GG	TT	TG	AA	GG	GG
T153	TT	CT	AT	TT	AA	TT	GG	CT	GC	TT	CT	AA	GG	AA	TT	GG	AG	TG	CT	GG	AG	GG	GG
T170	CC	TT	TT	AA	AG	AT	AG	CT	GC	CT	CT	AA	GG	AT	CC	TG	AA	TG	CT	TG	AG	AG	AG
T175	CT	CC	TT	AA	AA	TT	AA	CT	GC	TT	CT	AA	GG	AA	TT	TG	AG	GG	TT	TG	GG	AG	GG
T177	CT	CC	TT	AA	AA	TT	AG	TT	GG	CT	CT	AG	GG	AT	TT	GG	AA	TG	TT	TG	GG	AG	GG
T179	CT	CC	AT	AT	AG	TT	AG	CT	GC	CT	TT	AG	GG	AT	CC	GG	AA	TT	CT	TG	AA	AG	GG
T182	TT	CC	TT	TT	AA	TT	GG	CT	GG	TT	TT	AA	AG	TT	CT	TG	AA	TG	TT	TG	GG	GG	GG
T184	TT	CT	TT	AT	AA	AT	GG	CC	GG	TT	TT	AA	GG	AT	CT	GG	AA	TG	CT	GG	AG	GG	GG
T190	CT	CT	TT	TT	AA	TT	AA	CT	GG	TT	TT	AA	GG	AA	CT	GG	AA	GG	CT	GG	GG	GG	AG
T195	TT	CT	AT	TT	AA	TT	GG	TT	CC	CT	TT	AG	GG	AT	CT	GG	AA	TG	TT	TG	GG	AG	GG
T200	TT	CT	AT	TT	AA	TT	AG	TT	GC	TT	TT	AG	AG	AT	CT	GG	AA	TG	CT	TG	AA	GG	AG
Dt1	CC	CC	AT	TT	AG	AA	GG	CT	CC	TT	TT	AG	AG	AA	CT	GG	AG	TG	TT	GG	AA	AG	AA
Dt2	CT	CT	TT	AT	GG	TT	GG	TT	CC	CC	CC	AA	AG	AA	TT	GG	AA	TT	CC	GG	AA	GG	GG
Dt3	TT	CC	AT	TT	AA	AT	AA	TT	GC	CT	TT	AG	AA	AA	CT	TG	AA	TT	TT	GG	AA	GG	GG
Dt4	CC	CC	TT	AA	AG	AT	GG	CC	CC	CT	CT	AA	AG	AT	TT	TT	AA	TT	TT	GG	AG	GG	AG
Dt5	CC	CC	TT	TT	GG	AT	AG	CC	CC	TT	TT	AG	AG	AA	TT	GG	AA	TT	CT	GG	AA	GG	GG
Dt6	TT	TT	TT	TT	AG	AA	AG	CC	CC	TT	CT	AG	GG	TT	CC	GG	AA	TT	CT	GG	AA	AA	GG
Dt7	CT	TT	TT	TT	AA	TT	GG	CT	GG	CT	CC	AA	AA	AT	TT	GG	AA	TT	TT	GG	AA	GG	GG
Dt8	CT	CT	TT	TT	AG	AA	AG	CT	CC	CC	CC	AG	GG	AA	TT	TT	AA	TT	CT	TG	AA	AG	AG
Dt9	CC	CT	TT	AT	AA	TT	AA	CT	CC	CT	TT	AA	AG	TT	CT	GG	GG	TG	CT	GG	AA	AG	AG
Dt10	CC	CT	AT	TT	AG	AA	GG	CC	CC	CC	CT	AG	AG	AT	TT	TG	AA	TG	CC	GG	AA	AG	AG
Dt11	CT	CT	TT	AT	AG	AT	GG	CT	CC	CT	CT	AG	GG	TT	TT	TG	AG	TT	CT	GG	AG	AG	AG
Dt12	CT	TT	AT	TT	AG	TT	AG	CT	CC	CC	CC	AG	AG	TT	CC	GG	AG	TG	TT	GG	AA	GG	GG
Dt13	CT	CC	AT	TT	AA	TT	GG	CC	GC	TT	CT	AA	AG	AT	TT	TG	GG	TT	CC	GG	AG	AG	AG

SNP Ind.	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23
Dt14	CT	CT	TT	TT	AA	TT	GG	CC	CC	TT	CT	AA	GG	AT	TT	TG	AG	TG	CT	GG	AA	AA	GG
Dt15	CC	CC	TT	AA	AA	AA	GG	CT	CC	TT	CC	GG	GG	TT	TT	TT	AA	TG	TT	TG	AG	GG	AG
Dt16	CT	CT	AT	TT	AG	AT	GG	TT	GC	CT	CC	AA	AA	AA	TT	TG	AG	TG	CC	GG	AA	AG	AA
Dt17	CT	CT	AT	TT	AG	AT	AG	TT	GC	TT	TT	AA	AG	AA	CC	TG	AA	GG	TT	TT	AG	AG	AG
Dt18	CT	CT	AT	AA	GG	TT	GG	CT	CC	CC	TT	AA	AA	AT	CC	GG	AA	TG	TT	GG	AG	AG	AG
Dt19	TT	CC	AT	AT	GG	TT	AG	CT	CC	CT	TT	AA	AG	TT	CT	GG	AA	TG	CT	TT	AG	GG	GG
Dt20	CT	CT	TT	AA	AG	AT	AG	CT	GC	CT	CC	AA	AG	AT	CT	GG	AA	GG	CT	TG	AA	GG	AG
Dt21	CC	CC	AT	TT	AG	AT	GG	TT	GC	CC	CC	GG	AG	AA	TT	TT	AA	GG	CT	TG	AA	AG	AG
Dt22	CC	CC	AT	AA	AG	AT	GG	CT	CC	CT	TT	AG	AG	AT	CT	TG	AG	TG	TT	GG	AA	GG	AG
Dt23	TT	CC	AA	AA	AG	AT	AA	CT	CC	CT	TT	AG	AG	AT	TT	TG	AG	TG	TT	TT	AG	GG	AG
Dt24	CT	CT	AT	AT	AG	AA	AG	TT	CC	TT	TT	AA	GG	TT	TT	TG	AG	TG	TT	GG	AG	GG	GG
Dt25	TT	CT	TT	TT	AA	TT	AG	CC	CC	TT	TT	AG	GG	AA	TT	GG	AA	TG	TT	TG	AG	AG	GG
Dt26	TT	CC	TT	AA	AG	AT	AG	CT	CC	CT	TT	AA	AG	AT	CC	GG	AA	GG	CT	TG	AG	GG	GG
Dt27	CC	CC	TT	AA	AA	AT	AG	TT	GG	TT	TT	AA	GG	AA	TT	GG	AA	TG	TT	GG	GG	AG	GG
Dt28	TT	CT	AT	AT	AA	TT	GG	TT	CC	CC	TT	AA	GG	AT	CT	GG	AA	TG	TT	TT	AG	GG	GG
Dt29	CT	CC	AT	AA	AA	TT	GG	CT	GC	CT	TT	AA	GG	AA	CT	GG	AA	TG	TT	TG	AA	GG	GG
Dt30	CT	TT	TT	AT	GG	TT	AG	TT	GC	TT	CT	AA	GG	AT	TT	TG	AA	GG	CT	TG	GG	AG	GG
Dt31	CC	CT	AT	AT	AG	AT	GG	TT	GC	TT	CT	AA	GG	AT	CC	GG	AG	TG	CT	TG	AG	GG	GG
Dt32	CT	CT	TT	AT	AG	AA	GG	TT	GC	CT	CT	AA	GG	AT	CC	GG	AA	TG	CT	GG	AA	GG	GG
Dt33	CT	CT	AT	AT	AA	TT	AA	TT	GG	TT	TT	AA	GG	TT	TT	GG	AA	GG	CT	TG	AG	GG	GG
Dt34	CT	CC	TT	AT	AA	AA	GG	TT	GC	TT	TT	AA	GG	AA	CC	GG	AA	TG	TT	GG	AG	GG	GG
Dt35	TT	CT	AT	AT	AA	TT	AG	TT	GG	TT	CT	AA	AG	TT	TT	GG	AA	GG	CT	TG	GG	GG	AG
Dt36	CT	CT	TT	AT	AA	TT	GG	CC	GG	TT	TT	AA	GG	AT	TT	TG	AA	TT	CT	GG	AG	GG	GG
Dt37	TT	TT	AT	AT	AA	AT	AA	TT	GC	CT	CT	AA	GG	AT	CT	GG	AA	TG	CT	TG	AG	GG	GG
Dt38	CT	CT	TT	AA	AA	TT	GG	CC	GC	CT	TT	AA	GG	AA	CT	TT	AA	GG	CT	TG	AA	GG	GG
Dt39	TT	CC	TT	AT	AA	AA	GG	TT	GC	TT	TT	AA	GG	AA	CC	TG	AA	GG	CT	TG	AG	GG	GG
Dt40	TT	CT	TT	TT	AA	AA	GG	CC	CC	TT	CT	AG	GG	AT	TT	GG	AA	GG	CT	GG	GG	GG	GG
Dt41	CT	CT	AT	AT	GG	AT	GG	CC	GC	TT	TT	AA	GG	TT	CC	GG	AA	TT	CT	TG	AG	GG	GG
Dt42	CC	CC	AA	TT	AA	TT	GG	CT	GC	TT	CT	AA	AG	AT	CT	TG	AA	GG	TT	TG	AA	AG	GG
Dt43	TT	CT	AT	TT	AA	TT	GG	CT	GC	CT	TT	AG	GG	AT	CC	GG	AA	GG	TT	TG	AG	GG	GG
Dt44	CC	CT	TT	AA	GG	AT	GG	CT	GC	CT	TT	AA	GG	AA	CT	GG	AA	TG	TT	TG	AA	GG	AG
Dt45	CC	CC	TT	AT	AG	TT	GG	TT	GC	TT	TT	AA	GG	AA	TT	GG	AA	TG	TT	TT	AG	GG	GG
Dt46	TT	CT	AT	AA	AA	TT	AA	CT	GG	TT	CT	AG	AA	TT	TT	GG	AA	GG	CC	TG	AG	AG	AG
Dt47	CT	CT	AT	AT	AA	AT	GG	CT	GC	CC	TT	AA	AG	AT	CT	GG	AA	TG	CT	TG	GG	AG	AG
Dt48	CC	CC	TT	AT	AA	AA	GG	TT	GC	TT	TT	AA	GG	AT	CC	GG	AA	TG	CT	TG	GG	AG	GG
Dt49	TT	CT	TT	TT	AA	TT	GG	TT	GG	TT	TT	AA	GG	AT	CC	TG	AA	TG	TT	GG	AA	GG	AG
Dt50	CT	CC	AT	AT	AA	TT	GG	TT	GC	CT	CC	AA	GG	AT	TT	GG	AG	TG	TT	TG	AG	GG	GG
Dt51	TT	CC	TT	AT	AA	AA	GG	CT	GC	CT	TT	AA	GG	AT	CT	GG	AA	GG	TT	GG	GG	GG	GG
Dt52	CC	CT	AT	TT	AG	AT	GG	CC	GC	TT	TT	AG	GG	AT	TT	GG	AA	TG	CT	TT	AG	GG	GG
Dt53	CT	TT	TT	AT	AA	TT	AG	TT	CC	TT	TT	AA	GG	AT	TT	TG	AA	TG	TT	TT	AG	GG	GG
Dt54	CC	TT	TT	AA	AA	AT	AA	TT	GG	CT	TT	AA	GG	AT	TT	TG	AA	TG	TT	GG	AG	GG	GG
Dt55	TT	TT	AT	TT	AG	AT	GG	CT	GC	CT	CC	AA	GG	AA	TT	GG	AA	GG	TT	TG	AG	GG	AA
Dt56	CT	CT	TT	AT	AG	AT	AG	CT	CC	TT	TT	AA	GG	TT	CT	GG	AA	TG	TT	GG	GG	GG	GG
Dt57	TT	TT	AT	AT	AA	AT	GG	CC	CC	TT	TT	AG	GG	AT	CT	TG	AA	GG	CT	TG	GG	GG	GG
Dt58	CC	CT	TT	AT	AA	AA	GG	CT	GG	TT	TT	AA	GG	AT	CC	TG	AG	TT	TT	TG	AG	GG	GG
Dt59	CT	TT	AT	TT	AA	TT	AG	TT	GG	TT	TT	AA	GG	AT	CT	GG	AA	GG	CT	TG	AG	GG	GG
Dt60	CT	CC	TT	AA	AA	AA	GG	CT	GC	TT	TT	AA	GG	AA	TT	GG	AG	GG	TT	TT	AG	GG	GG
Dt61	CT	CC	TT	AA	AG	TT	AA	TT	GC	CT	TT	AA	GG	AT	CT	GG	AG	TT	TT	GG	AG	GG	AG
Dt62	CT	CC	AT	TT	AA	AT	GG	CC	GC	CT	TT	AG	GG	AT	CT	GG	AA	TG	CT	GG	GG	GG	GG
Dt63	TT	CT	TT	AT	AA	AT	AG	CT	GG	TT	CT	AA	GG	AT	TT	GG	AA	TG	TT	TG	AG	AG	AG
Dt64	CT	CC	AT	TT	AA	TT	GG	TT	CC	TT	CT	GG	GG	AT	CT	GG	AA	GG	TT	TG	AG	GG	AG

Tabelle A-3: Allelfrequenzen der 23 SNPs für die vier untersuchten Populationen. Die Tabelle beinhaltet neben den errechneten Frequenzen, die in der HapMap-Datenbank und im SNPforID browser notierten Häufigkeiten für die jeweiligen SNP-Marker.

rs Nr.	Ancestr. Al		Ref. Al.	Durchschn.									
	SNP	lel		Polym.	Europäer (HapMap)	Europäer (52-plex)	Deutsche	Japaner (40)	Thais (49)	Chinesen (57)	Chinesen (HapMap)	Chinesen (52-plex)	Asiaten (52-plex)
rs1490413	A1	G	A/G	0,575	0,568	0,492	0,500	0,357	0,483	0,422	0,364	0,364	0,447
rs876724	A2	C	C/T	0,638	0,674	0,609	0,450	0,633	0,561	0,567	0,570	0,514	0,548
rs1357617	A3	A	A/T	0,717	0,706	0,742	0,788	0,765	0,763	0,844	0,744	0,795	0,772
rs2046361	A4	A	A/T	0,333	0,344	0,438	0,388	0,459	0,447	0,344	0,402	0,475	0,431
rs717302	A5	G	A/G	0,543	0,508	0,281	0,113	0,235	0,175	0,111	0,174	0,181	0,174
rs1029047	A6	T	A/T	0,700	0,584	0,594	0,738	0,694	0,588	0,689	0,679	0,641	0,673
rs917118	A7	T	C/T	0,233	0,276	0,258	0,325	0,347	0,263	0,318	0,352	0,338	0,312
rs763869	A8	T	C/T	0,483	0,502	0,586	0,250	0,541	0,658	0,689	0,651	0,662	0,483
rs1015250	A9	C	C/G	0,203	0,212	0,391	0,500	0,459	0,553	0,433	0,523	0,462	0,504
rs735155	A10	G	A/G	0,475	0,534	0,305	0,200	0,153	0,193	0,156	0,131	0,152	0,182
rs901398	A11	T	C/T	0,658	0,676	0,719	0,688	0,827	0,816	0,733	0,682	0,703	0,777
rs2107612	A12	A	A/G	0,692	0,660	0,820	0,850	0,847	0,912	0,911	0,932	0,910	0,870
rs1886510	A13	T	C/T	0,575	0,469	0,219	0,250	0,092	0,070	0,133	0,057	0,138	0,137
rs1454361	A14	A	A/T	0,542	0,551	0,547	0,550	0,622	0,570	0,511	0,570	0,538	0,581
rs2016276	A15	-	A/G	0,267	0,230	0,375	0,450	0,479	0,395	-	0,330	0,377	0,441
rs729172	A16	C	A/C	0,583	0,596	0,773	0,850	0,816	0,860	0,833	0,837	0,838	0,842
rs740910	A17	A	A/G	0,661	0,700	0,867	0,800	0,949	0,912	0,922	0,898	0,917	0,887
rs1493232	A18	C	A/C	0,375	0,347	0,547	0,625	0,553	0,597	0,670	0,381	0,490	0,592
rs719366	A19	T	C/T	0,625	0,625	0,711	0,725	0,663	0,754	0,778	0,784	0,782	0,714
rs1031825	A20	C	A/C	0,667	0,739	0,656	0,563	0,582	0,605	0,489	0,512	0,569	0,583
rs722098	A21	G	A/G	-	0,228	0,414	0,475	0,541	0,544	-	0,440	0,485	0,520
rs733164	A22	G	A/G	0,733	0,702	0,828	0,800	0,796	0,877	0,956	0,920	0,864	0,824
rs826472	A23	C	C/T	-	0,611	0,789	0,688	0,755	0,860	-	0,895	0,838	0,767

Tabelle A-4: Heterozygotenraten der 210 Proben der vier untersuchten Populationen. Berechnet wurden die erwartete und die beobachtete Heterozygotenrate für jede Population einzeln und für die asiatischen Populationen zusammengefasst.

SNP	Heterozygotenrate Expected				Heterozygotenrate observed				Asian ex- pected	Asian Ob- served
	JPN	THAI	CHN	GER	JPN	THAI	CHN	GER		
A1	0,5	0,459	0,499	0,5	0,487	0,306	0,474	0,453	0,486	0,422
A7	0,439	0,453	0,379	0,383	0,450	0,367	0,292	0,266	0,297	0,370
A2	0,495	0,465	0,492	0,476	0,500	0,490	0,491	0,469	0,484	0,494
A13	0,375	0,167	0,131	0,342	0,400	0,143	0,105	0,281	0,224	0,216
A3	0,335	0,359	0,361	0,383	0,325	0,429	0,404	0,453	0,352	0,386
A9	0,5	0,497	0,494	0,476	0,500	0,388	0,544	0,438	0,497	0,477
A22	0,32	0,325	0,215	0,285	0,350	0,327	0,246	0,281	0,287	0,307
A10	0,32	0,259	0,311	0,424	0,350	0,265	0,316	0,359	0,297	0,310
A21	0,499	0,497	0,496	0,485	0,500	0,510	0,526	0,516	0,497	0,512
A17	0,32	0,0968	0,16	0,23	0,250	0,061	0,140	0,203	0,192	0,151
A8	0,399	0,497	0,45	0,485	0,350	0,347	0,368	0,391	0,449	0,355
A11	0,43	0,287	0,301	0,404	0,525	0,347	0,298	0,250	0,339	0,390
A23	0,43	0,37	0,241	0,333	0,425	0,327	0,211	0,328	0,347	0,321
A20	0,492	0,487	0,478	0,451	0,575	0,673	0,509	0,438	0,486	0,586
A6	0,387	0,425	0,485	0,482	0,375	0,367	0,439	0,375	0,432	0,394
A19	0,399	0,447	0,371	0,411	0,400	0,469	0,456	0,422	0,406	0,442
A12	0,255	0,259	0,16	0,295	0,200	0,306	0,175	0,266	0,225	0,227
A14	0,495	0,463	0,49	0,496	0,650	0,521	0,579	0,531	0,483	0,583
A18	0,469	0,494	0,481	0,496	0,450	0,674	0,526	0,500	0,481	0,550
A16	0,255	0,3	0,241	0,35	0,300	0,327	0,246	0,297	0,265	0,291
A15	0,495	0,499	0,478	0,469	0,500	0,583	0,439	0,313	0,491	0,507
A5	0,2	0,359	0,281	0,404	0,225	0,347	0,185	0,344	0,186	0,252
A4	0,475	0,497	0,494	0,492	0,525	0,592	0,544	0,406	0,489	0,554
				$\emptyset = 0,415$			$\emptyset = 0,373$		$\emptyset = 0,378$	$\emptyset = 0,396$

Tabelle A-5: Berechnung der Korrelationskoeffizienten nach Pearson anhand der **a)** Allelfrequenzen für die deutsche Stichprobe vs. der in der *HapMap*-Datenbank gefundenen Frequenzen, **b)** untersuchten chinesischen Population vs. der *HapMap*-Datenbankwerte für Han Chinesen und ergänzend **c)** die Korrelationswerte für die untersuchten asiatischen Stichproben vs. der *HapMap*-Datenbankwerte für Han Chinesen.

a)

SNP	Deutsche	Europäer (HapMap)	Korrelationen		Deutsche	Europäer (HapMap)
A1	0,492	0,575	Deutsche	Korrelation nach Pearson	1	0,682
A2	0,609	0,638		Signifikanz (2-seitig)		0,001
A3	0,742	0,717		Quadratsummen und Kreuzprodukte	0,881	0,452
A4	0,438	0,333		Kovarianz	0,040	0,023
A5	0,281	0,543		N	23	21 **
A6	0,594	0,700	Europäer (HapMap)	Korrelation nach Pearson	0,682 **	1
A7	0,258	0,233		Signifikanz (2-seitig)	0,001	
A8	0,586	0,483		Quadratsummen und Kreuzprodukte	0,452	0,543
A9	0,391	0,203		Kovarianz	0,023	0,027
A10	0,305	0,475		N	21	21
A11	0,719	0,658				
A12	0,820	0,692				
A13	0,219	0,575				
A14	0,547	0,542				
A15	0,375	0,267				
A16	0,773	0,583				
A17	0,867	0,661				
A18	0,547	0,375				
A19	0,711	0,625				
A20	0,656	0,667				
A21	0,414	-				
A22	0,828	0,733				
A23	0,789	-				

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

b)

snp	Chinesen	Han Chinesen (HapMap)	Korrelationen		Chinesen	Han Chinesen (HapMap)
A1	0,483	0,422	Chinesen	Korrelation nach Pearson	1	0,965 **
A2	0,561	0,567		Signifikanz (2-seitig)		0,000
A3	0,763	0,844		Quadratsummen und Kreuzprodukte	1,329	1,250
A4	0,447	0,344		Kovarianz	0,060	0,066
A5	0,175	0,111		N	23	20
A6	0,588	0,689	Han Chinesen (HapMap)	Korrelation nach Pearson	0,965 **	1
A7	0,263	0,318		Signifikanz (2-seitig)	0,000	
A8	0,658	0,689		Quadratsummen und Kreuzprodukte	1,250	1,379
A9	0,553	0,433		Kovarianz	0,066	0,073
A10	0,193	0,156		N	20	20
A11	0,816	0,733				
A12	0,912	0,911				
A13	0,070	0,133				
A14	0,570	0,511				
A15	0,395	-				
A16	0,860	0,833				
A17	0,912	0,922				
A18	0,597	0,670				
A19	0,754	0,778				
A20	0,605	0,489				
A21	0,544	-				
A22	0,877	0,956				
A23	0,860	-				

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

c)

snp	Durchschn. Frequenzwerte		Korrelationen		Chinesen (HapMap)	Durchschn. Frequenzwerte Asiaten
	Chinesen (HapMap)	Asiaten				
A1	0,422	0,447	Chinesen (HapMap)	Korrelation nach Pearson	1	0,963 **
A2	0,567	0,548		Signifikanz (2-seitig)		0,000
A3	0,844	0,772		Quadratsummen und Kreuzprodukte	1,379	1,160
A4	0,344	0,431		Kovarianz	0,073	0,061
A5	0,111	0,174		N	20	20
A6	0,689	0,673	Durchschn. Frequenzwerte Asiaten	Korrelation nach Pearson	0,963 **	1
A7	0,318	0,312		Signifikanz (2-seitig)	0,000	
A8	0,689	0,483		Quadratsummen und Kreuzprodukte	1,160	1,111
A9	0,433	0,504		Kovarianz	0,061	0,050
A10	0,156	0,182		N	20	23
A11	0,733	0,777		** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.		
A12	0,911	0,870				
A13	0,133	0,137				
A14	0,511	0,581				
A15		0,441				
A16	0,833	0,842				
A17	0,922	0,887				
A18	0,670	0,592				
A19	0,778	0,714				
A20	0,489	0,583				
A21		0,520				
A22	0,956	0,824				
A23		0,767				

Locus	SNP	P-values							
		JPN	JPN	THAI	THAI	CHN	CHN	GER	GER
rs1490413	A1	0,8759	-	0,02025	P < 0.05	0,69803	-	0.39050	-
rs917118	A7	0,85505	-	0,16285	-	0.06830	-	0.01300	P < 0.05
rs876724	A2	0,87255	-	0,88475	-	0,89533	-	0.90195	-
rs1886510	A13	0,83685	-	0,18585	-	0,1243	-	0.10130	-
rs1357617	A3	0,8271	-	0,18205	-	0,60163	-	0.14745	-
rs1015250	A9	0,8759	-	0,1241	-	0,5123	-	0.52285	-
rs733164	A22	0,81535	-	0,83315	-	0,3843	-	0.84070	-
rs735155	A10	0,81585	-	0,79345	-	0,83243	-	0.19420	-
rs722098	A21	0,8746	-	0,885	-	0,6891	-	0.70950	-
rs740910	A17	0,09261	-	0,05355	-	0,2006	-	0.17765	-
rs763869	A8	0,32785	-	0,03075	P < 0.05	0,1967	-	0.09755	-
rs901398	A11	0,2094	-	0,22775	-	0,8316	-	0.00225	P < 0.05
rs826472	A23	0,85785	-	0,34745	-	0,17647	-	0.85680	-
rs1031825	A20	0,27455	-	0,01345	P < 0.05	0,67887	-	0.68300	-
rs1029047	A6	0,84735	-	0,24505	-	0,50177	-	0.05172	-
rs719366	A19	0,8456	-	0,87775	-	0,11003	-	0.88265	-
rs2107612	A12	0,10385	-	0,4308	-	0,6762	-	0.30025	-
rs1454361	A14	0,08565	-	0,44345	-	0,2349	-	0.53660	-
rs1493232	A18	0,62005	-	0,01325	P < 0.05	0,4969	-	0.90155	-
rs729172	A16	0,3797	-	0,8277	-	0,7864	-	0.22085	-
rs2016276	A15	0,8726	-	0,3157	-	0,50283	-	0.00575	P < 0.05
rs717302	A5	0,70001	-	0,5669	-	0.00640	P < 0.05	0.17425	-
rs2046361	A4	0,62245	-	0,20075	-	0,51123	-	0.16264	-
Overall									

Tabelle A-7: Werte der Varianzanalyse nach Wright für 23 Loci in der deutschen Population und in den drei untersuchten asiatischen Populationsstichproben.

SNP	Locus	Deutsche (Fis)	Asiaten(Fst)	Alle (Fst)
A1	rs1490413	0,101	0,025	0,025
A7	rs917118	0,313	-0,002	-0,002
A2	rs876724	0,023	0,021	0,021
A13	rs1886510	0,185	0,067	0,067
A3	rs1357617	-0,176	-0,009	-0,009
A9	rs1015250	0,089	-0,001	-0,001
A22	rs733164	0,020	0,006	0,006
A10	rs735155	0,160	-0,006	-0,006
A21	rs722098	-0,055	-0,005	-0,005
A17	rs740910	0,126	0,044	0,044
A8	rs763869	0,203	0,023	0,023
A11	rs901398	0,388	0,023	0,023
A23	rs826472	0,022	0,032	0,032
A20	rs1031825	0,038	-0,007	-0,007
A6	rs1029047	0,230	0,016	0,016
A19	rs719366	-0,019	0,002	0,002
A12	rs2107612	0,107	0,003	0,003
A14	rs1454361	-0,064	-0,003	-0,003
A18	rs1493232	-0,001	-0,004	-0,004
A16	rs729172	0,161	-0,006	-0,006
A15	rs2016276	0,340	-0,002	-0,002
A5	rs717302	0,157	0,011	0,011
A4	rs2046361	0,182	0,001	0,001
Overall		0,110	0,008	0,008

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
A	E34a	E34a	E95a	E95a	E8802a	E8802a	E9225a	E9225a	L11a	L11a	L31a	L31a	E64b	E64b	E99b	E99b	E8889b	E8889b	L2b	L2b	L14b	L14b	L33b	L33b	A
B	E34a	E34a	E95a	E95a	E8802a	E8802a	E9225a	E9225a	L11a	L11a	L31a	L31a	E64b	E64b	E99b	E99b	E8889b	E8889b	L2b	L2b	L14b	L14b	L33b	L33b	B
C	E35a	E35a	E97a	E97a	E8966a	E8966a	L1a	L1a	L13a	L13a	L33a	L33a	E67b	E67b	E115b	E115b	E9046b	E9046b	L3b	L3b	L17b	L17b	L38b	L38b	C
D	E35a	E35a	E97a	E97a	E8966a	E8966a	L1a	L1a	L13a	L13a	L33a	L33a	E67b	E67b	E115b	E115b	E9046b	E9046b	L3b	L3b	L17b	L17b	L38b	L38b	D
E	E64a	E64a	E99a	E99a	E8989a	E8989a	L2a	L2a	L14a	L14a	L38a	L38a	E73b	E73b	E2125b	E2125b	E9142b	E9142b	L4b	L4b	L22b	L22b	L44b	L44b	E
F	E64a	E64a	E99a	E99a	E8989a	E8989a	L2a	L2a	L14a	L14a	L38a	L38a	E73b	E73b	E2125b	E2125b	E9142b	E9142b	L4b	L4b	L22b	L22b	L44b	L44b	F
G	E67a	E67a	E115a	E115a	E9046a	E9046a	L3a	L3a	L17a	L17a	L44a	L44a	E82b	E82b	E2152b	E2152b	E9175b	E9175b	L6b	L6b	L23b	L23b	L45b	L45b	G
H	E67a	E67a	E115a	E115a	E9046a	E9046a	L3a	L3a	L17a	L17a	L44a	L44a	E82b	E82b	E2152b	E2152b	E9175b	E9175b	L6b	L6b	L23b	L23b	L45b	L45b	H
I	E73a	E73a	E2125a	E2125a	E9142a	E9142a	L4a	L4a	L22a	L22a	L45a	L45a	E86b	E86b	E2174b	E2174b	E9201b	E9201b	L8b	L8b	L30b	L30b	L347-2d	L347-2d	I
J	E73a	E73a	E2125a	E2125a	E9142a	E9142a	L4a	L4a	L22a	L22a	L45a	L45a	E86b	E86b	E2174b	E2174b	E9201b	E9201b	L8b	L8b	L30b	L30b	L347-2d	L347-2d	J
K	E82a	E82a	E2152a	E2152a	E9175a	E9175a	L6a	L6a	L23a	L23a	E34b	E34b	E95b	E95b	E8802b	E8802b	E9225b	E9225b	L11b	L11b	L31b	L31b	HTC	HTC	K
L	E82a	E82a	E2152a	E2152a	E9175a	E9175a	L6a	L6a	L23a	L23a	E34b	E34b	E95b	E95b	E8802b	E8802b	E9225b	E9225b	L11b	L11b	L31b	L31b	HTC	HTC	L
M	E86a	E86a	E2174a	E2174a	E9201a	E9201a	L8a	L8a	L30a	L30a	E35b	E35b	E97b	E97b	E8966b	E8966b	L1b	L1b	L13b	L13b	L347-40c	L347-40c	L347-2e	L347-2e	M
II	E86a	E86a	E2174a	E2174a	E9201a	E9201a	L8a	L8a	L30a	L30a	E35b	E35b	E97b	E97b	E8966b	E8966b	L1b	L1b	L13b	L13b	L347-40c	L347-40c	L347-2e	L347-2e	II
O	L347-1a	L347-1a	L347-1b	L347-1b	L347-2a	L347-2a	L347-2b	L347-2b	L347-40a	L347-40a	L347-40b	L347-40b	L347-45a	L347-45a	L347-45b	L347-45b	L347-4c	L347-4c	L347-2c	L347-2c	L347-45c	L347-45c	Ladder	Ladder	O
P	L347-1a	L347-1a	L347-1b	L347-1b	L347-2a	L347-2a	L347-2b	L347-2b	L347-40a	L347-40a	L347-40b	L347-40b	L347-45a	L347-45a	L347-45b	L347-45b	L347-4c	L347-4c	L347-2c	L347-2c	L347-45c	L347-45c	Ladder	Ladder	P

Abbildung A-1: Anordnung der Referenzproben aus dem *SNiPlex™* panel. Die Abbildung entspricht der Applizierung der DNA-Proben in einer 384er Mikrotiterplatte, wie sie von der Firma *Applied Biosystems* bereitgestellt wurde. Die Proben waren in Gruppen je vier *wells* pro DNA-Probe (Bsp.: E34a – A1-A2 und B1-B2) vorpipettiert und wurden vom Institut für Rechtsmedizin in Köln zur Verfügung gestellt. Der Inhalt von jeweils einem *well* der Platte wurde in eine 96er Mikrotiterplatte überführt. Die Probe L33b war nicht mehr verfügbar und wurde nicht typisiert.

Tabelle A-8: Erweiterte Auswahl von *tag*-Sequenzen und Abfolge der komplementären *tags* zum Spotten auf den Glasobjektträgern.

N°	Name	<i>tag</i> -Sequenz		<i>ctag</i> 5' 3'	Modifikation 3'
1	ProbeSet00122	ACGCGGTCACTCAGCATATA	cProbeSet00122	TATATGCTGAGTGACCGGTTTTTTTTTTTTTTT	C7 Amino
2	ProbeSet00132	CTTCGTGGCTAGTCTGTGAC	cProbeSet00132	GTCACAGACTAGCCACGAAGTTTTTTTTTTTTTT	C7 Amino
3	ProbeSet00138	CGTACCAATGGATGCGGTCT	cProbeSet00138	AGACCGCATCCATTGGTACGTTTTTTTTTTTTTT	C7 Amino
4	ProbeSet00142	CCACGAGCTGTAATCCGGTA	cProbeSet00142	TACCGGATTACAGCTCGTGGTTTTTTTTTTTTTT	C7 Amino
5	ProbeSet00147	ACGAGACGTGACGACTGCAT	cProbeSet00147	ATGCAGTCGTACGCTCTGTTTTTTTTTTTTTTT	C7 Amino
6	ProbeSet00152	TCGCCGTTGGTCTGTATGCA	cProbeSet00152	TGCATACAGACCAACGGCGATTTTTTTTTTTTTTT	C7 Amino
7	ProbeSet00153	GTGCGGTATGGTTCCTCTCA	cProbeSet00153	TGAGAGGAACCATACCGCACTTTTTTTTTTTTTTT	C7 Amino
8	ProbeSet00154	GTCACGTATGGTTCGCTGCT	cProbeSet00154	AGCAGCGAACCATACGTGACTTTTTTTTTTTTTTT	C7 Amino
9	ProbeSet00155	GTCCGTCATGTGTTAGATGC	cProbeSet00155	GCATCTAACACATGACGGACTTTTTTTTTTTTTTT	C7 Amino
10	ProbeSet00156	ATGCAGCGTAGGTATCGACT	cProbeSet00156	AGTCGATACCTACGCTGCATTTTTTTTTTTTTTT	C7 Amino
11	ProbeSet00161	GCGTGTGTGGACTCTCTCTA	cProbeSet00161	TAGAGAGAGTCCACACACGCTTTTTTTTTTTTTTT	C7 Amino
12	ProbeSet00164	CTGCGGTGTCAGTGATCTCT	cProbeSet00164	AGAGATCACTGACACCGCAGTTTTTTTTTTTTTTT	C7 Amino
13	ProbeSet00171	GCGCCGAATGTGTCTGTAAT	cProbeSet00171	ATTACAGACACATTGGCGCTTTTTTTTTTTTTTTT	C7 Amino
14	ProbeSet00175	GGCGGCTATTGACGAAGTCT	cProbeSet00175	AGAGTTCGTCAATAGCCGCCTTTTTTTTTTTTTTT	C7 Amino
15	ProbeSet00177	GCAGGGAATTGCCGACCATA	cProbeSet00177	TATGGTCGGCAATCCCTGCTTTTTTTTTTTTTTTT	C7 Amino
16	ProbeSet00178	TATGGGTCTTGCTGATACGC	cProbeSet00178	GCGTATCAGCAAGACCCATATTTTTTTTTTTTTTT	C7 Amino
17	ProbeSet00182	CGCGGCGTGTCTCAGAATAT	cProbeSet00182	ATATTCTGAGACACGCCGCTTTTTTTTTTTTTTTT	C7 Amino
18	ProbeSet00189	ATCCGGTCTCATCGCTGAAT	cProbeSet00189	ATTACGCGATGAGACCGGATTTTTTTTTTTTTTTT	C7 Amino
19	ProbeSet00202	AATGCTCACATCGCAGGTAC	cProbeSet00202	GTACCTGCGATGTGAGCATTTTTTTTTTTTTTTT	C7 Amino
20	ProbeSet00207	ACGCTAATGACGGCAGTGCA	cProbeSet00207	TGCACTGCCGTCATTAGCGTTTTTTTTTTTTTTT	C7 Amino
21	ProbeSet00210	ATGTGTCCGAACGTCGAGCT	cProbeSet00210	AGCTCGACGTTCCGACACATTTTTTTTTTTTTTTT	C7 Amino
22	ProbeSet00213	GCCGTCGGTTCAGGTCATAT	cProbeSet00213	ATATGACCTGAACCGACGGCTTTTTTTTTTTTTTTT	C7 Amino
23	ProbeSet00214	GTCGCGGGTCTGCACATAT	cProbeSet00214	ATATGTGCAGAACCCGCGACTTTTTTTTTTTTTTTT	C7 Amino
24	ProbeSet00222	TCCGTCTGTTGAGTTAGGCC	cProbeSet00222	GGCCTAACTCAACAGACGGATTTTTTTTTTTTTTTT	C7 Amino
25	ProbeSet00223	TCCTCTCGTTGGATGTGAGC	cProbeSet00223	GCTCACATCCAACGAGAGGATTTTTTTTTTTTTTTT	C7 Amino
26	ProbeSet00224	TCGGTTGCTTGGATGTACCC	cProbeSet00224	GGGTACATCCAAGCAACCGATTTTTTTTTTTTTTTT	C7 Amino
27	ProbeSet00225	GGCGTTCGTGCAGCTTACTT	cProbeSet00225	AAGTAAGCTGCACGAACGCCTTTTTTTTTTTTTTTT	C7 Amino
28	ProbeSet00226	CGCGTACCTGCTGTTTGCAT	cProbeSet00226	ATGCAACAGCAGGTACGCGTTTTTTTTTTTTTTT	C7 Amino
29	ProbeSet00243	CAATACCTGTGACGAGCTGC	cProbeSet00243	GCAGCTCGTCACAGGTATTGTTTTTTTTTTTTTTT	C7 Amino
30	ProbeSet00253	CGAGTGCTCCGTGCGAAATA	cProbeSet00253	TATTCGCACGGAGCACTCGTTTTTTTTTTTTTTT	C7 Amino

Abbildung A-2: Plotdarstellung der Genotypcluster für den SNP A1.

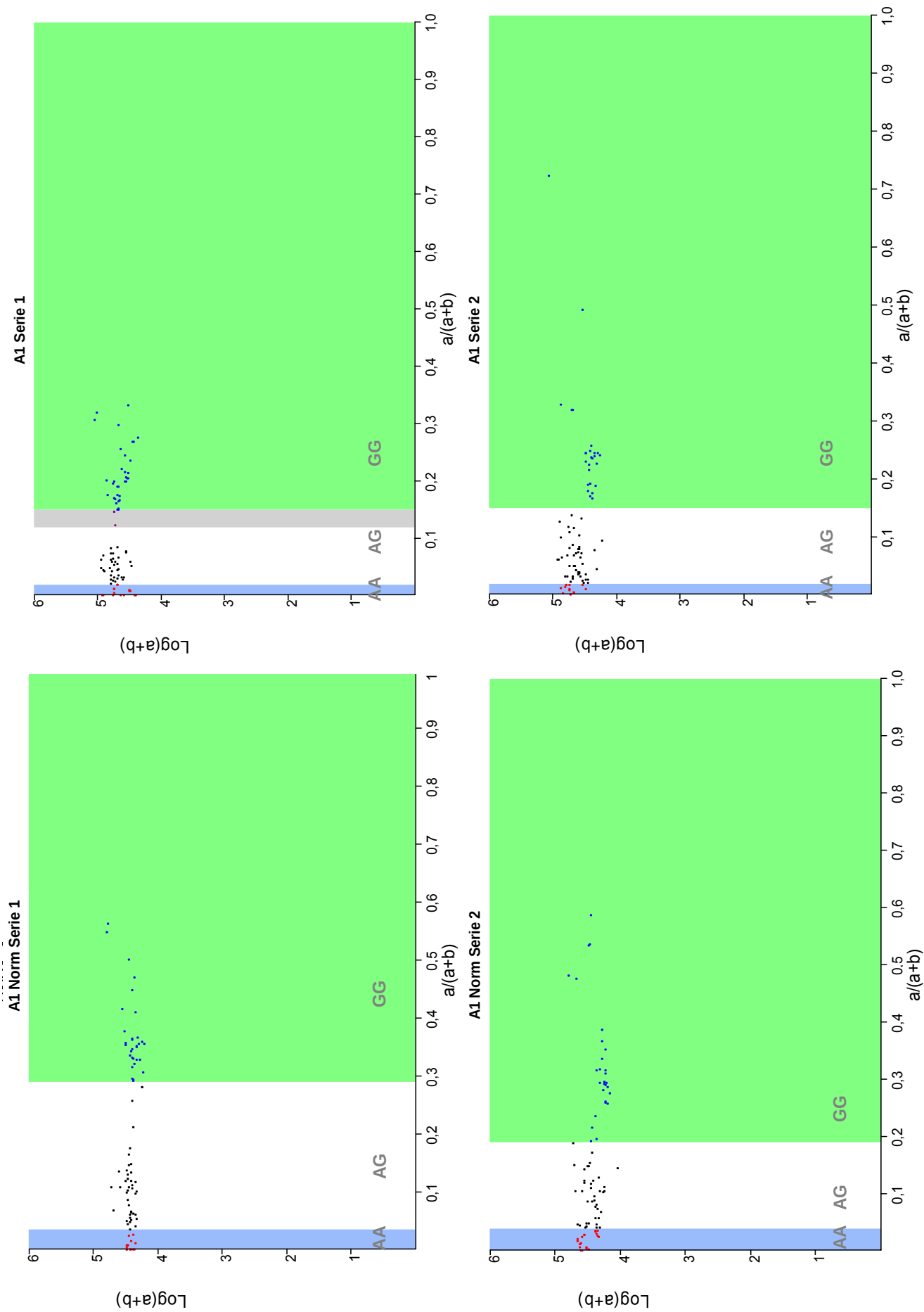


Abbildung A-3: Plotdarstellung der Genotypcluster für den SNP A2.

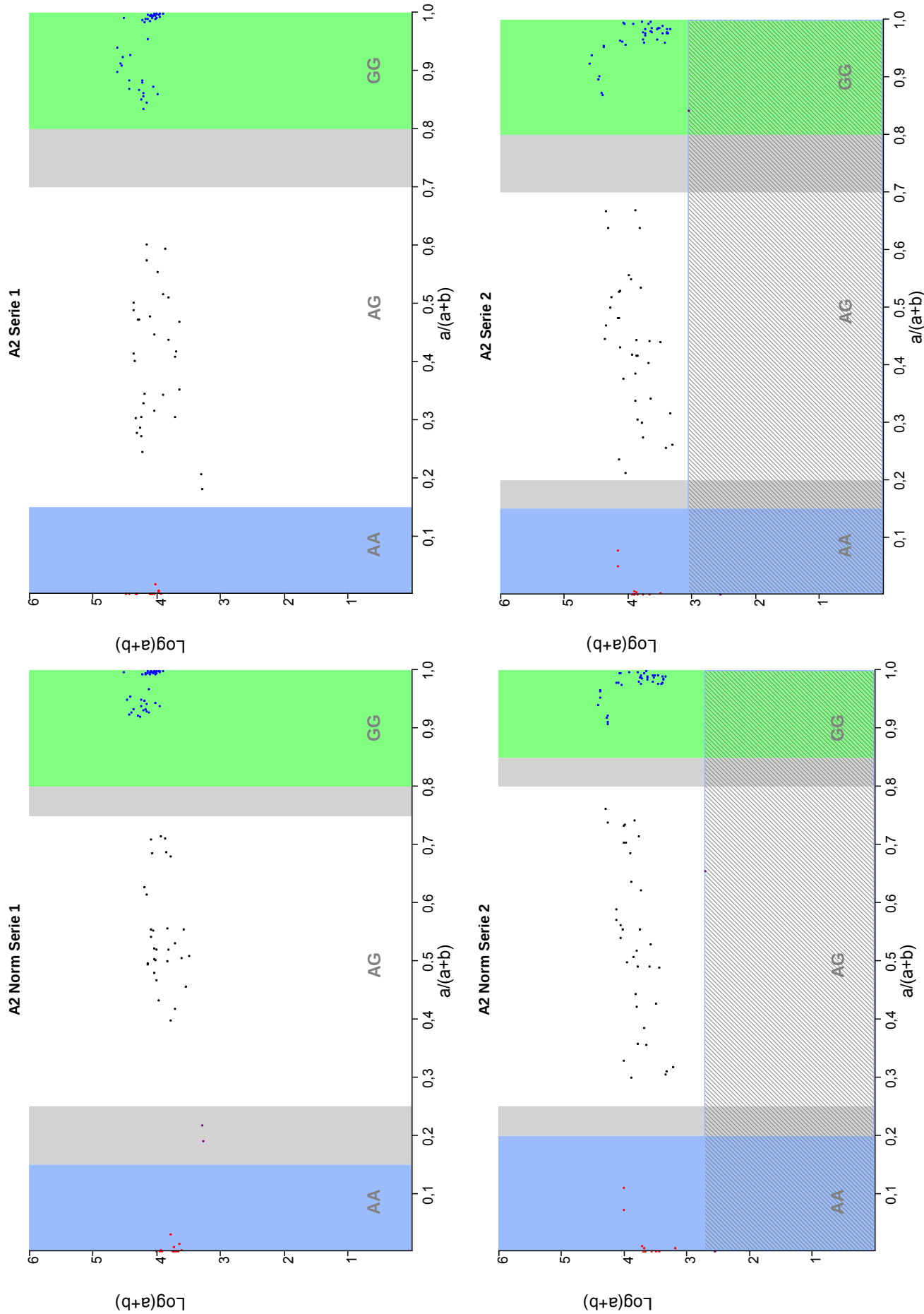


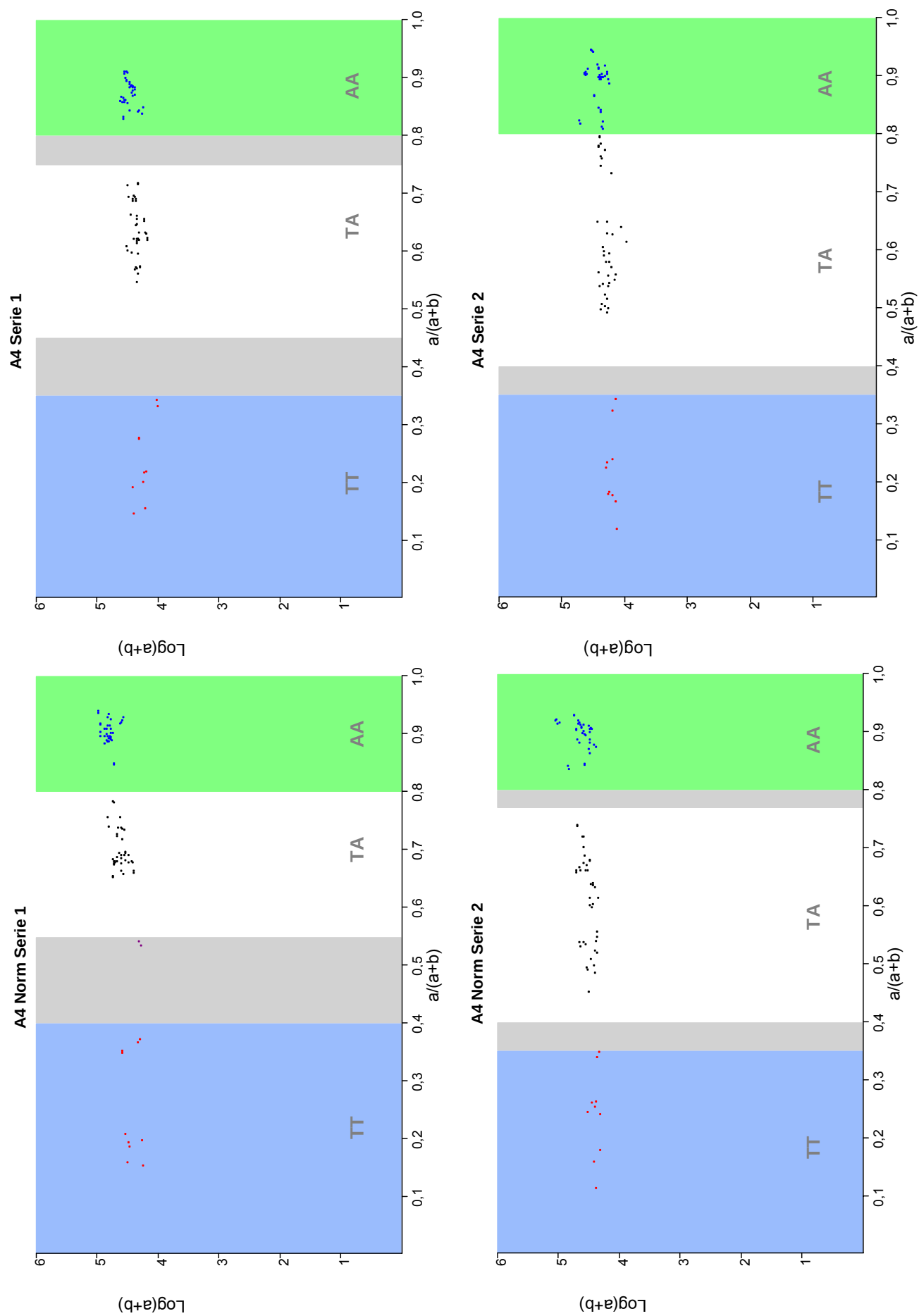
Abbildung A-4: Plotdarstellung der Genotypcluster für den SNP A4.

Abbildung A-5: Plotdarstellung der Genotypcluster für den SNP A5.

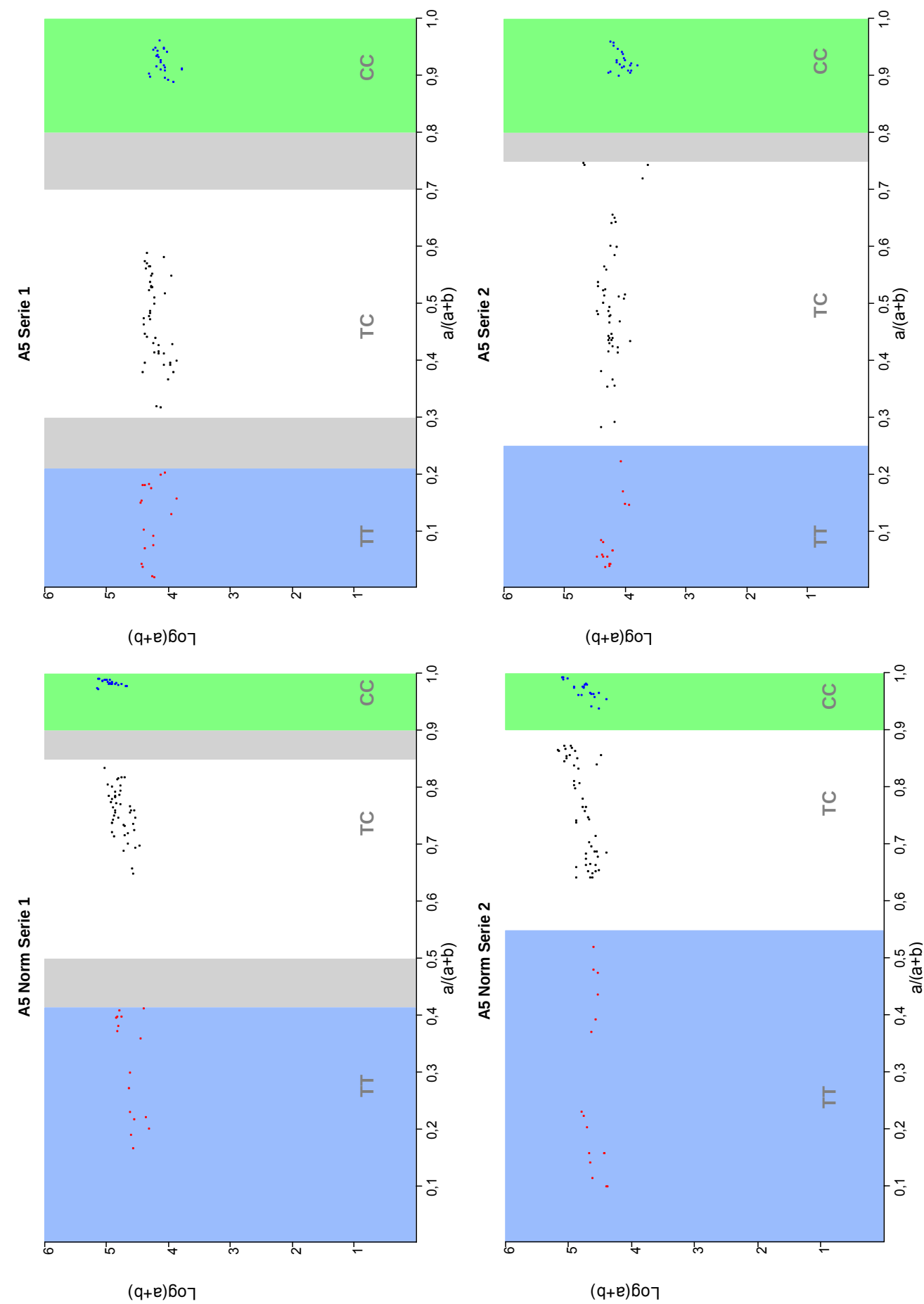


Abbildung A-6: Plotdarstellung der Genotypcluster für den SNP A8.

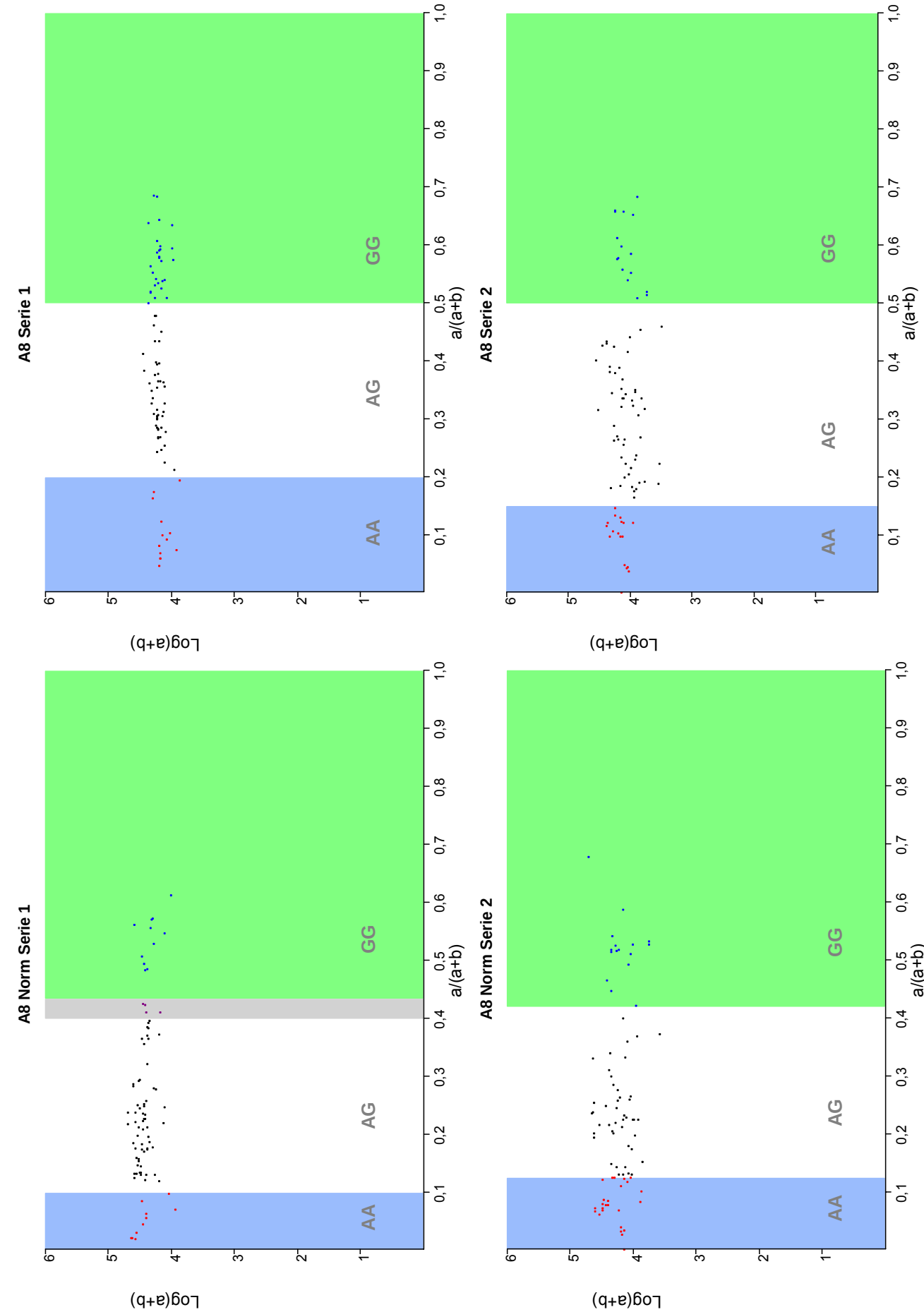


Abbildung A-7: Plotdarstellung der Genotypcluster für den SNP A9.

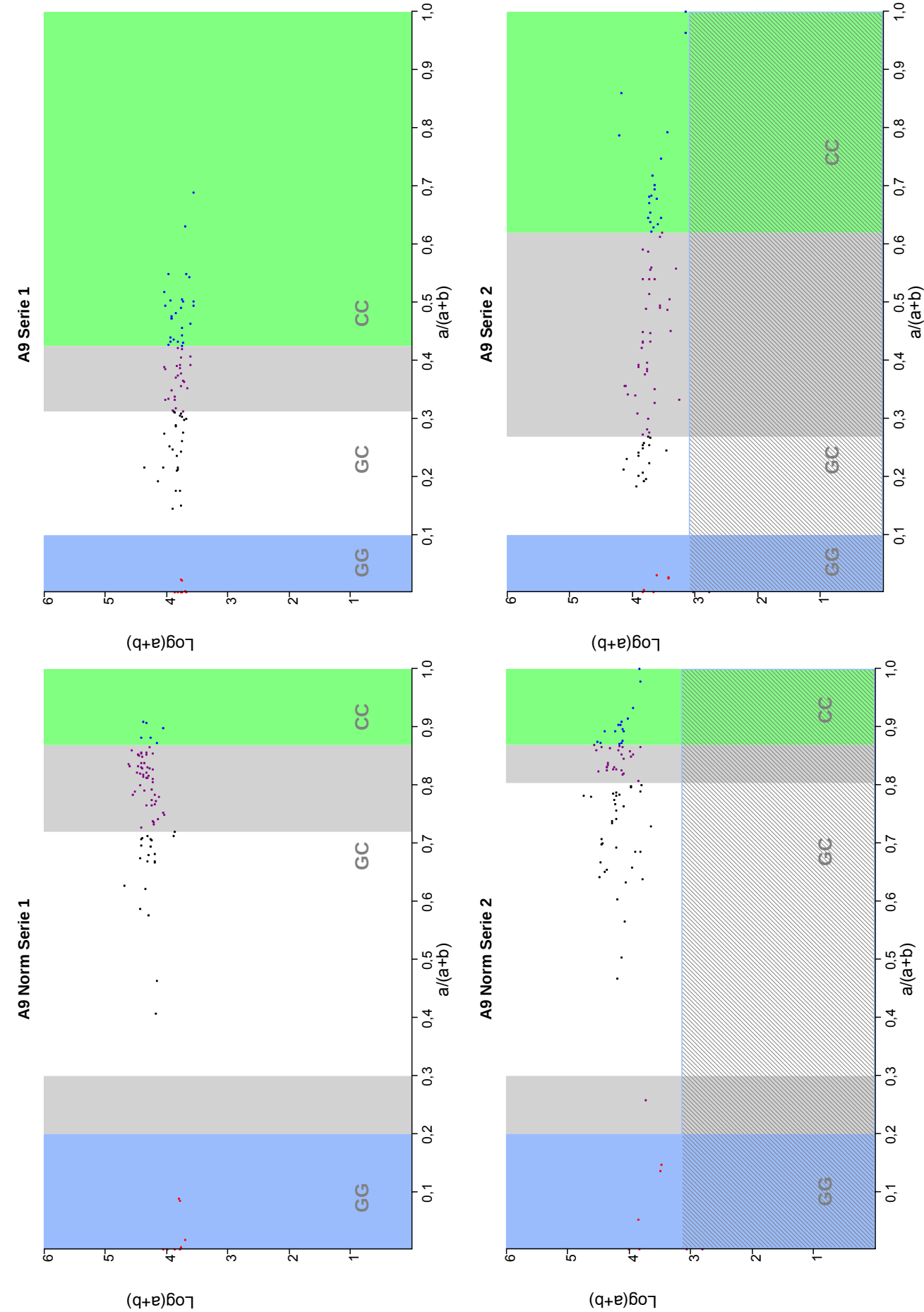


Abbildung A-8: Plotdarstellung der Genotypcluster für den SNP A11.

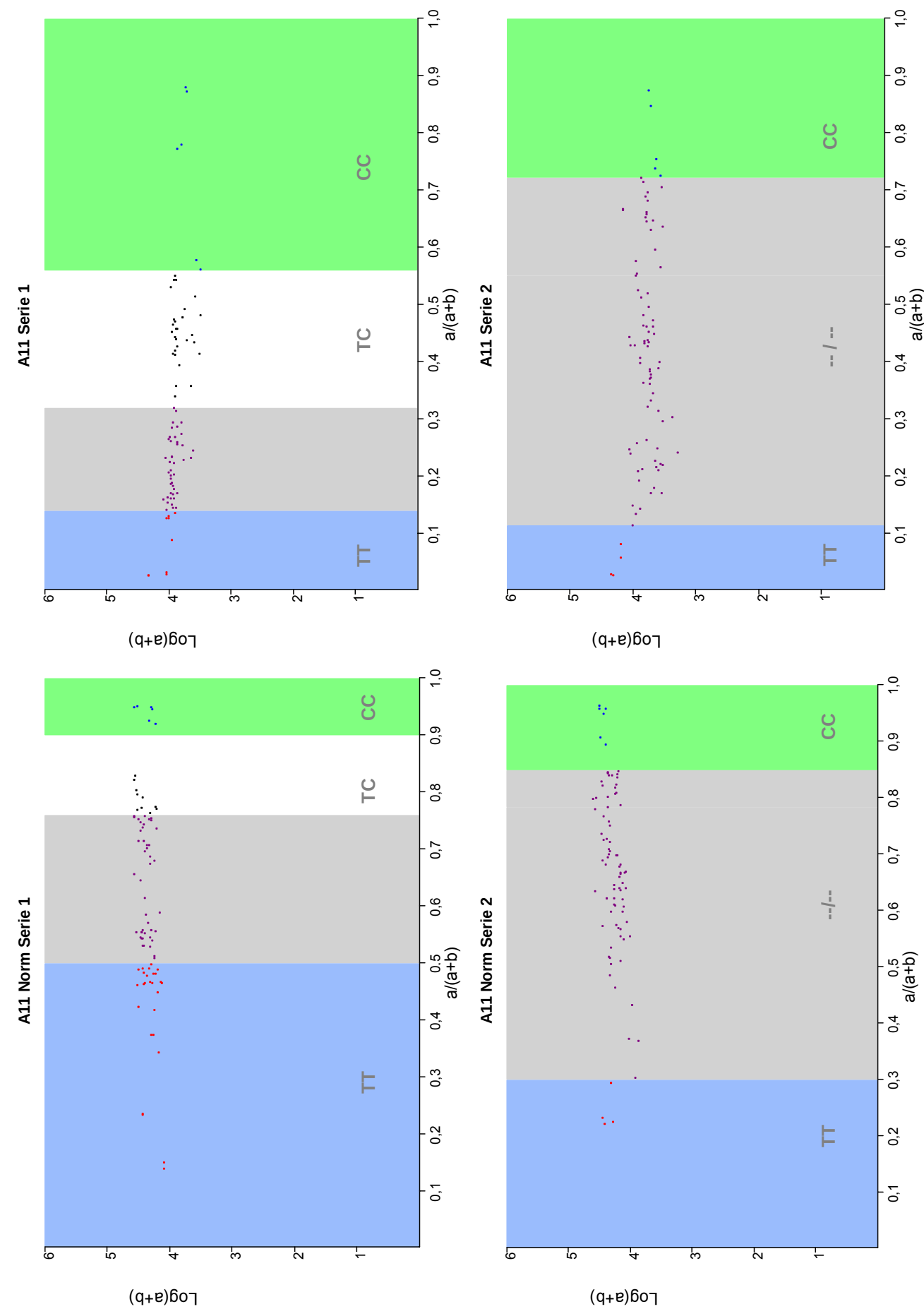


Abbildung A-9: Plotdarstellung der Genotypcluster für den SNP A15.

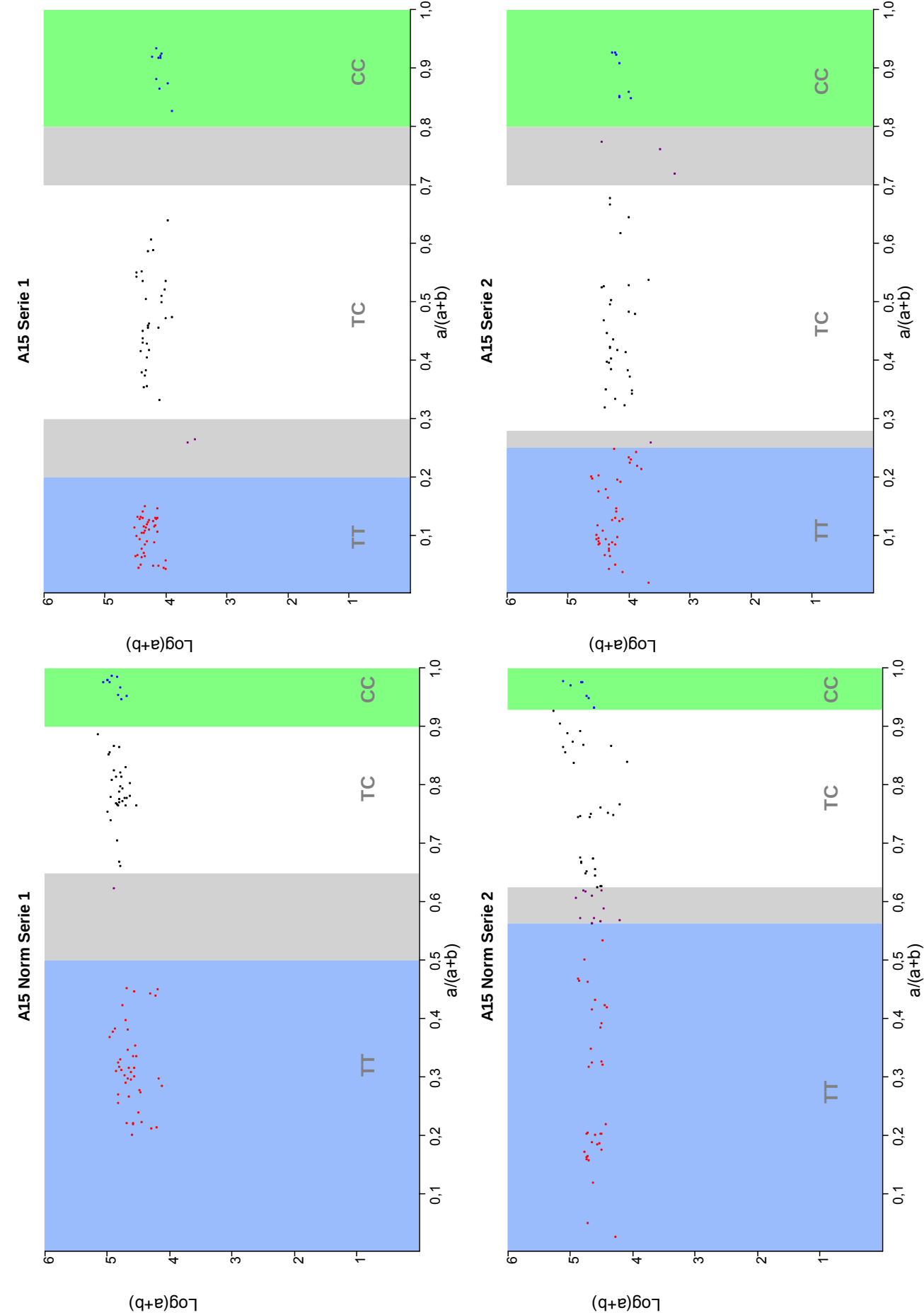


Abbildung A-10: Plotdarstellung der Genotypcluster für den SNP A16.

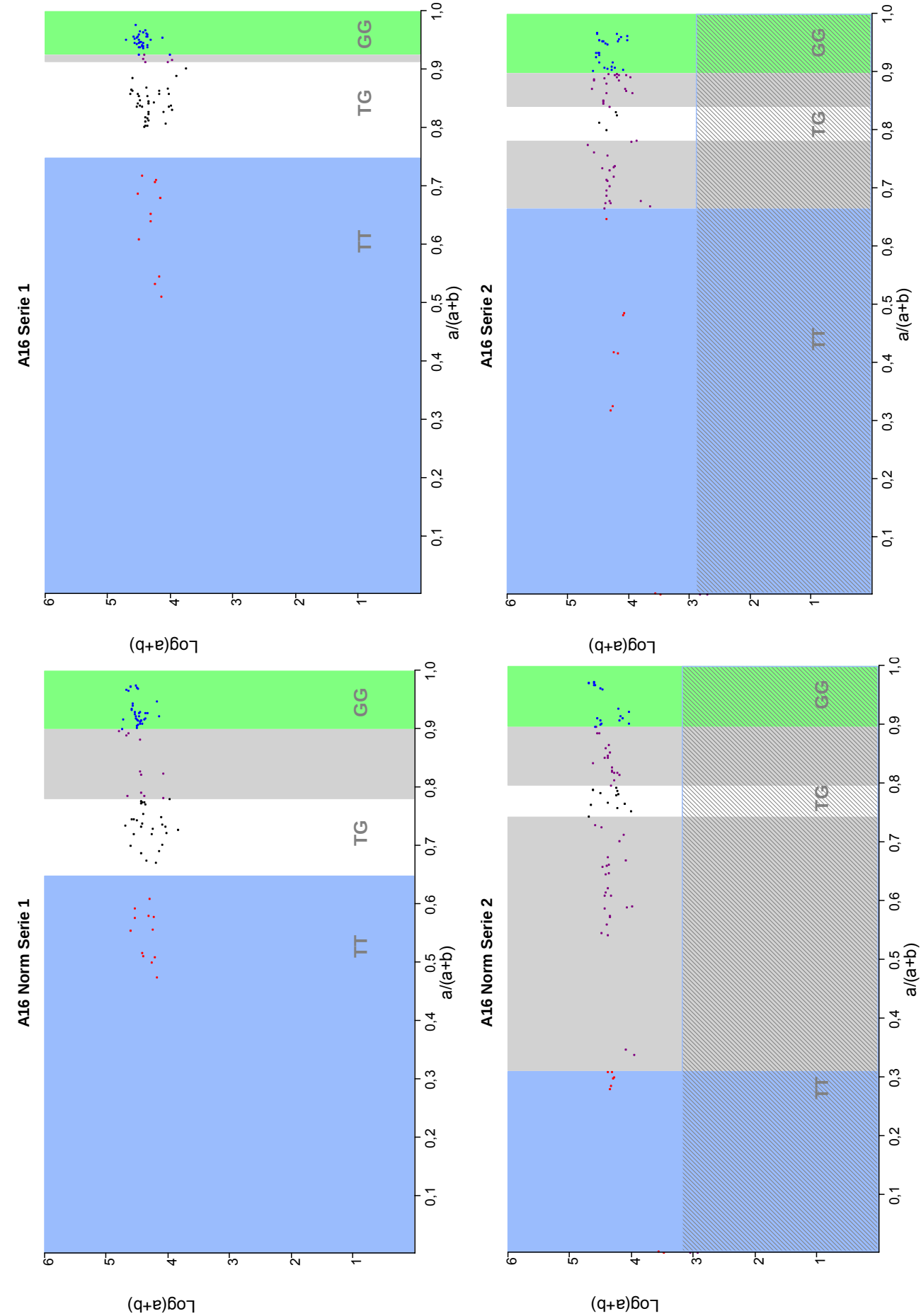


Abbildung A-11: Plotdarstellung der Genotypcluster für den SNP A18.

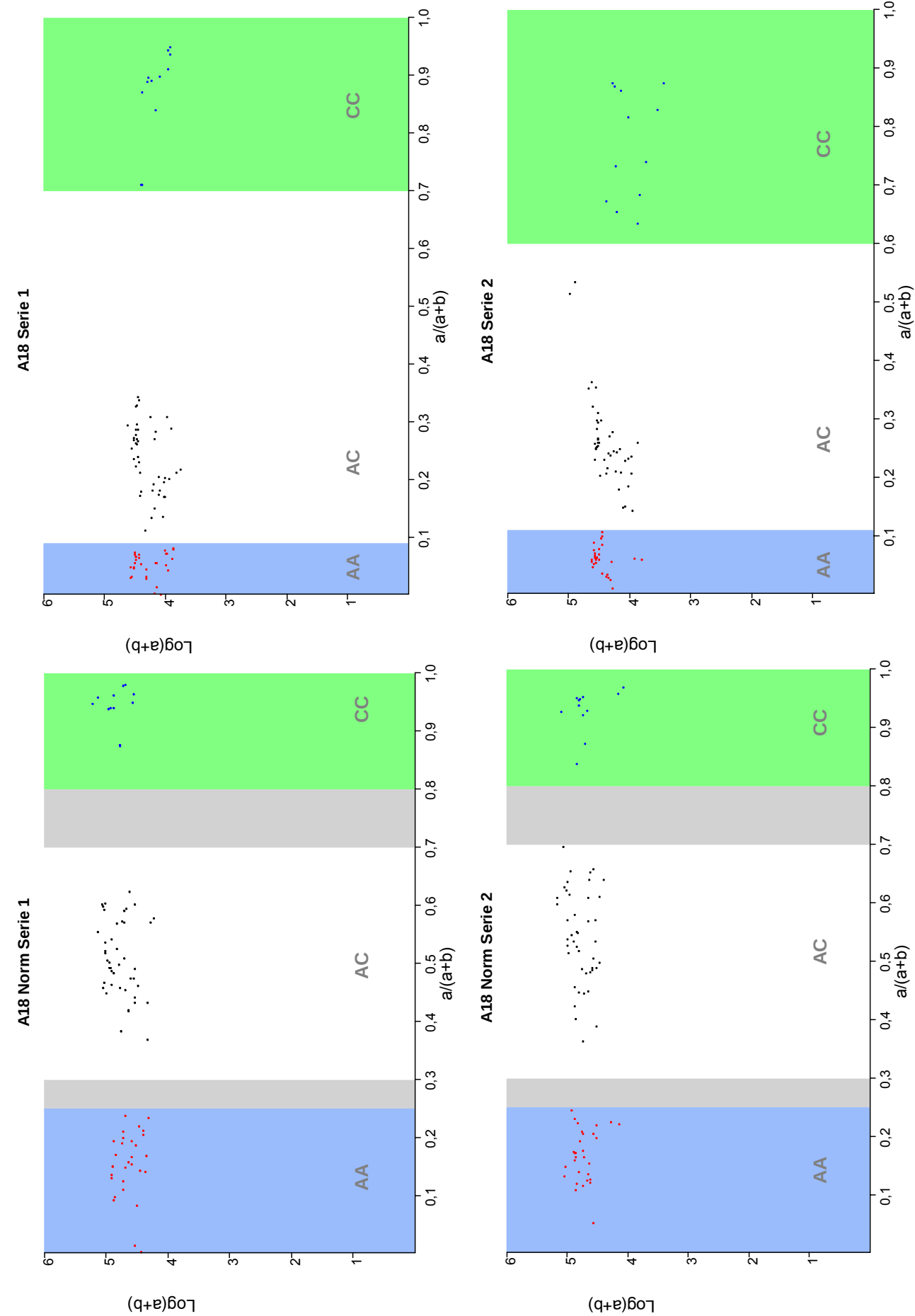


Abbildung A-12: Plotdarstellung der Genotypcluster für den SNP A20.

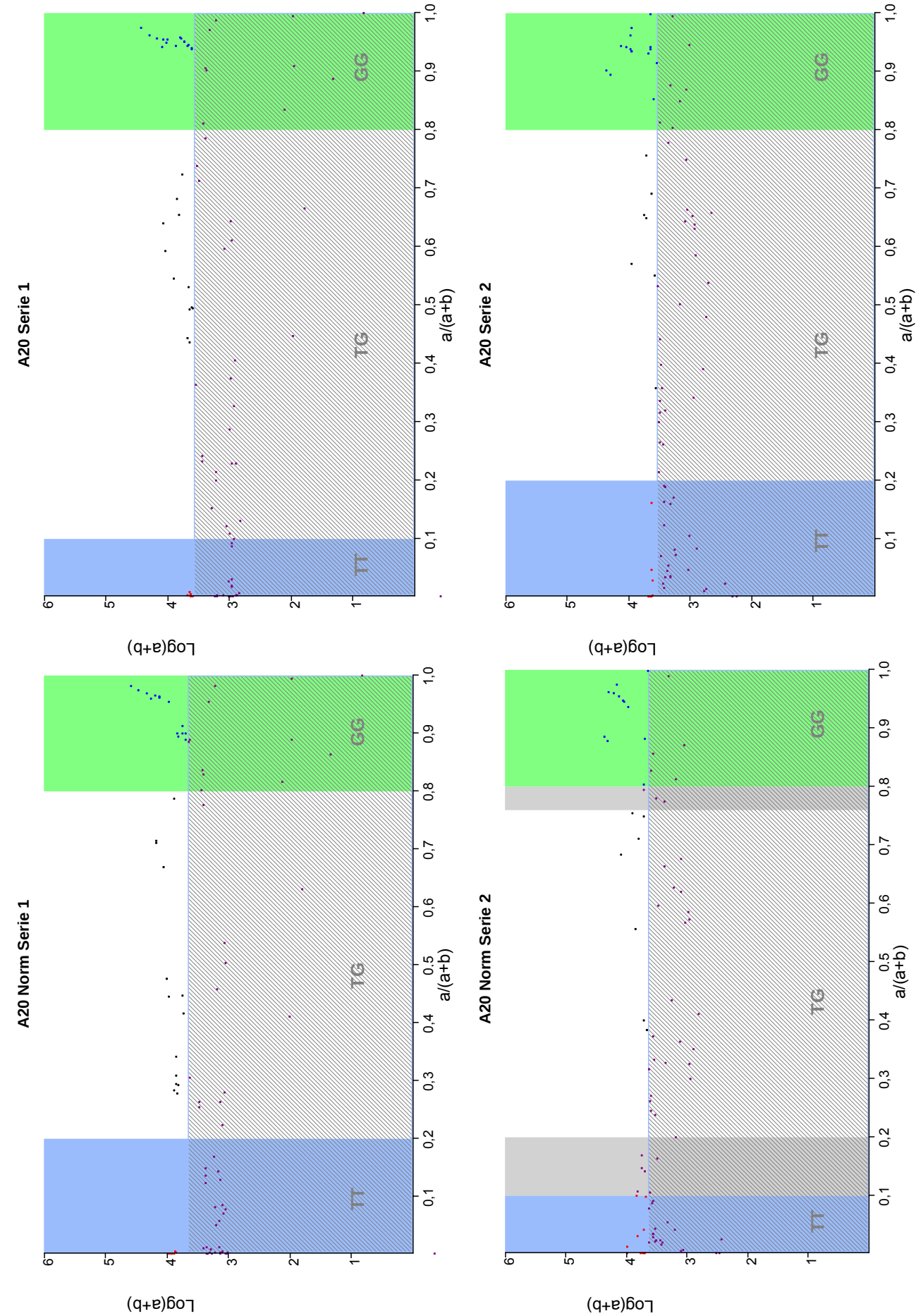


Abbildung A-13: Plotdarstellung der Genotypcluster für den SNP A21.

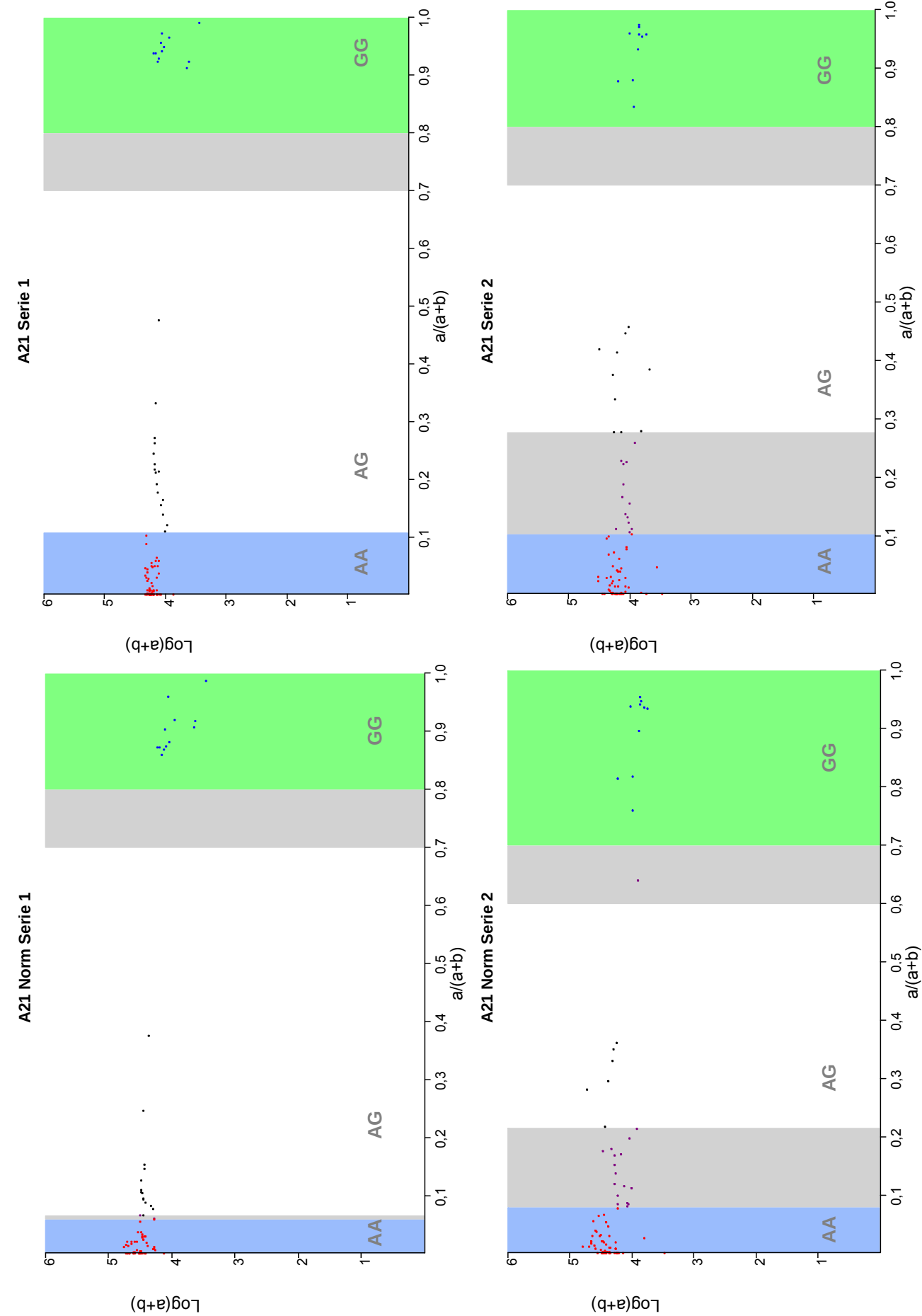


Abbildung A-14: Plotdarstellung der Genotypcluster für den SNP A23.

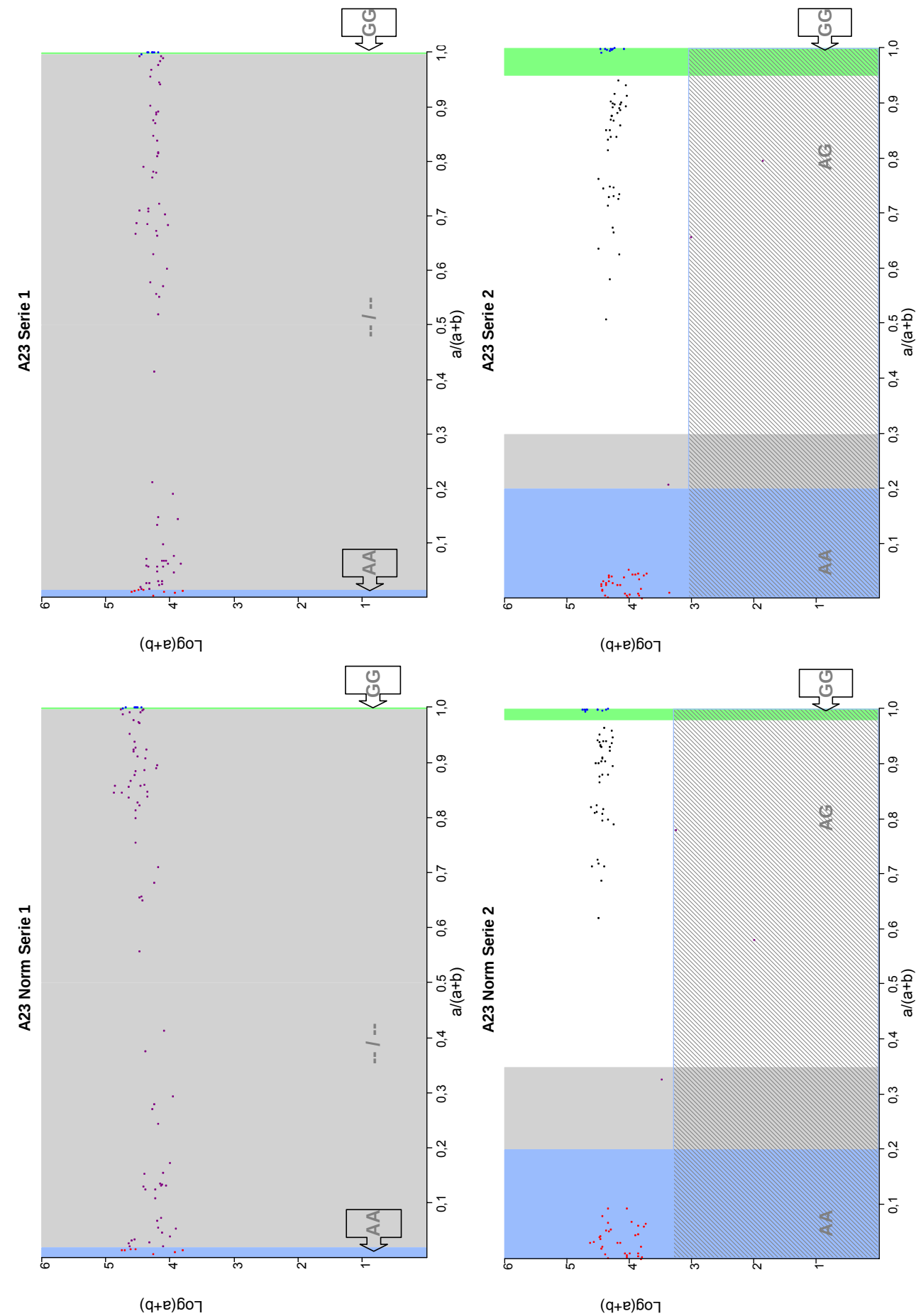


Abbildung A-15: Plotdarstellung der Genotypcluster für den SNP A24.

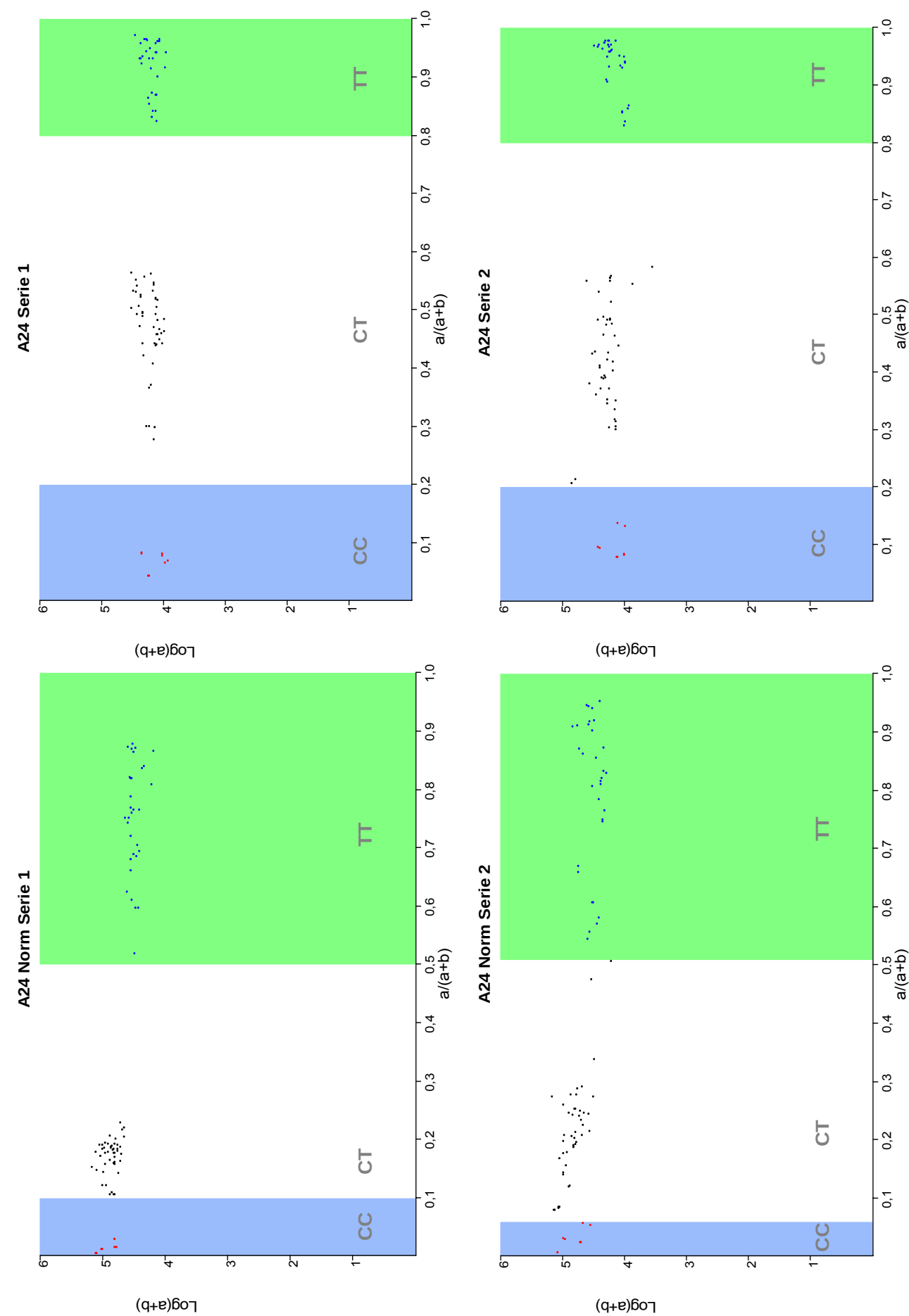


Abbildung A-16: Plotdarstellung der Genotypcluster für den SNP A25.

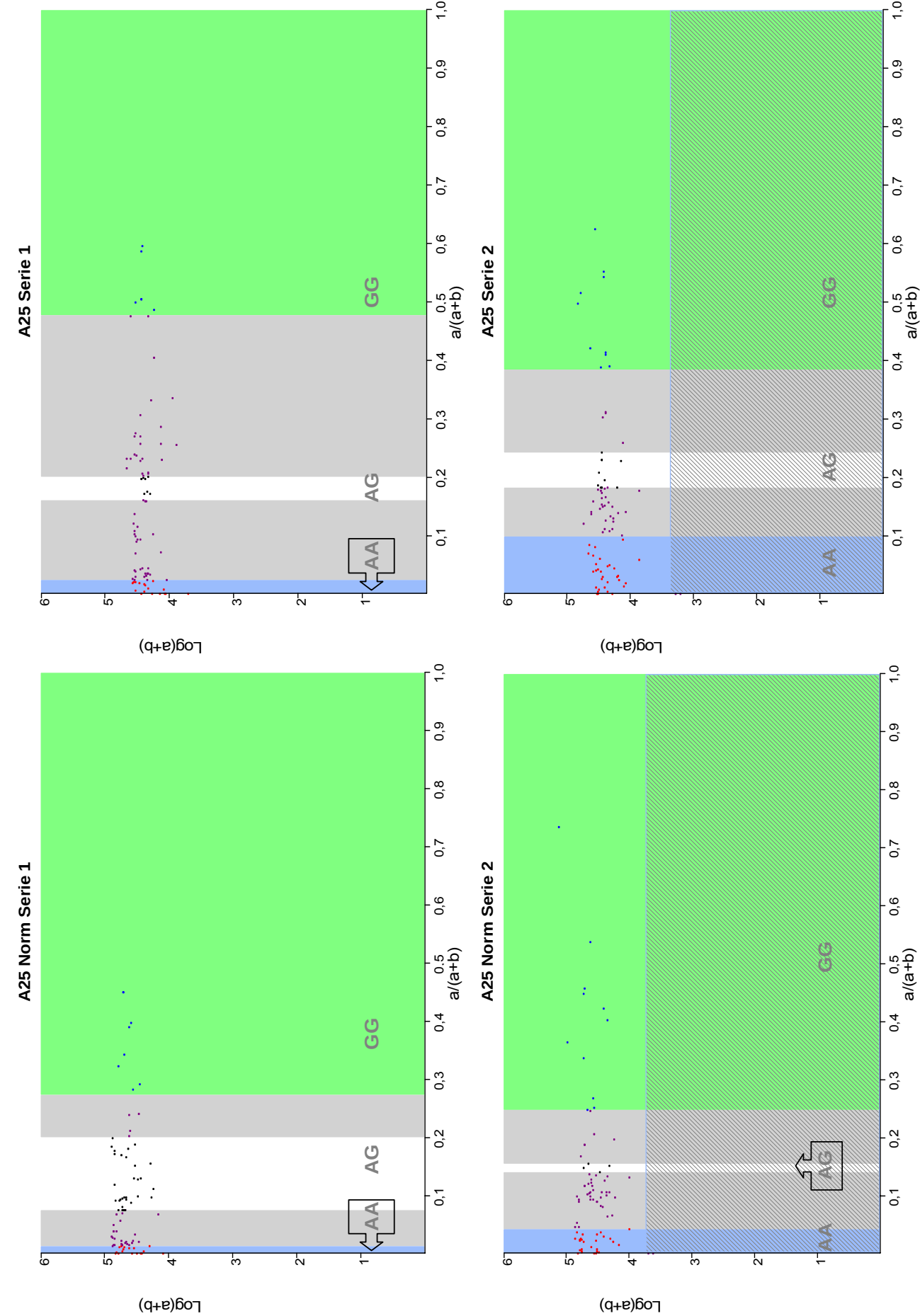


Abbildung A-17: Plotdarstellung der Genotypcluster für den SNP A27.

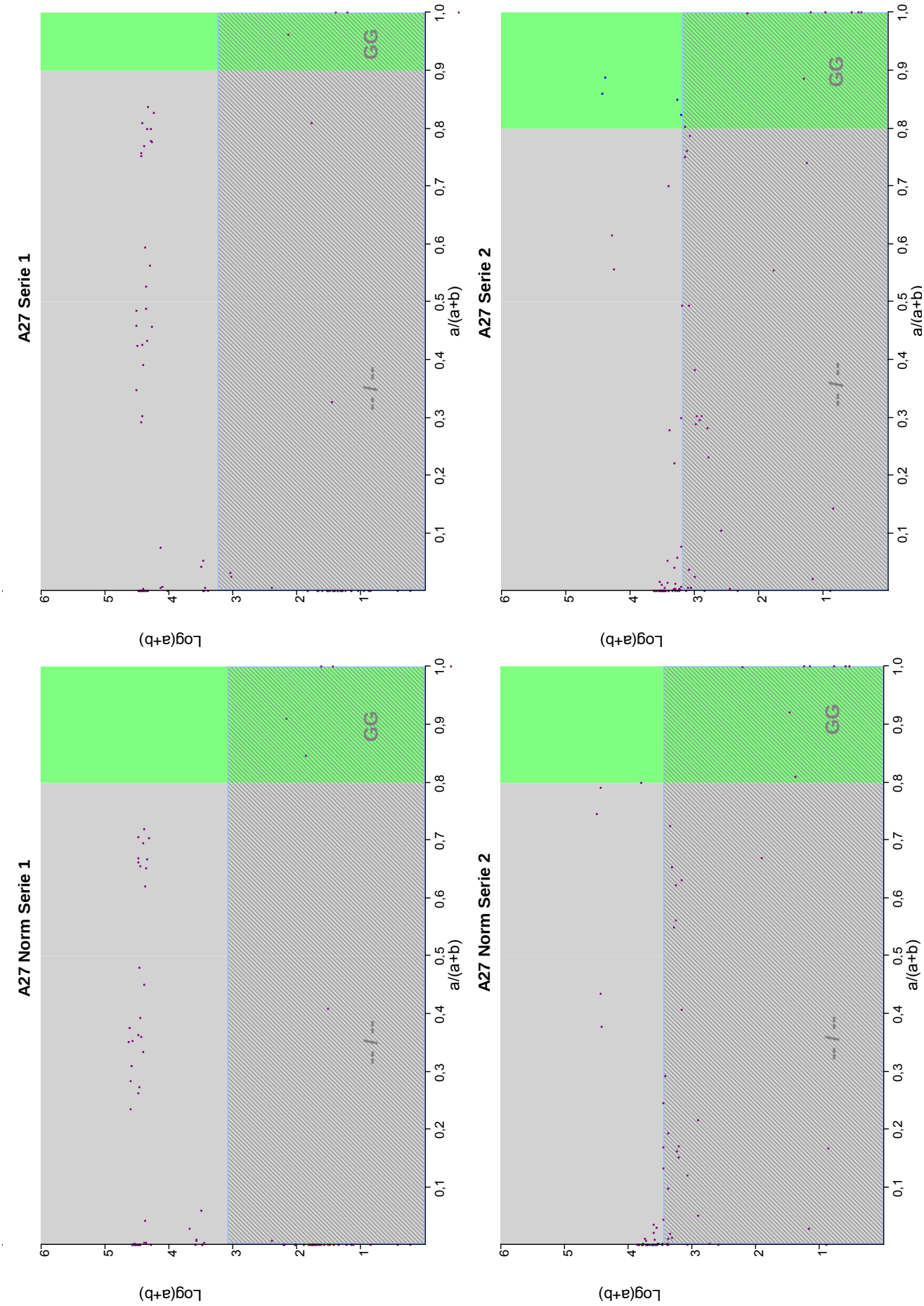


Abbildung A-18: Plotdarstellung der Genotypcluster für den SNP A30.

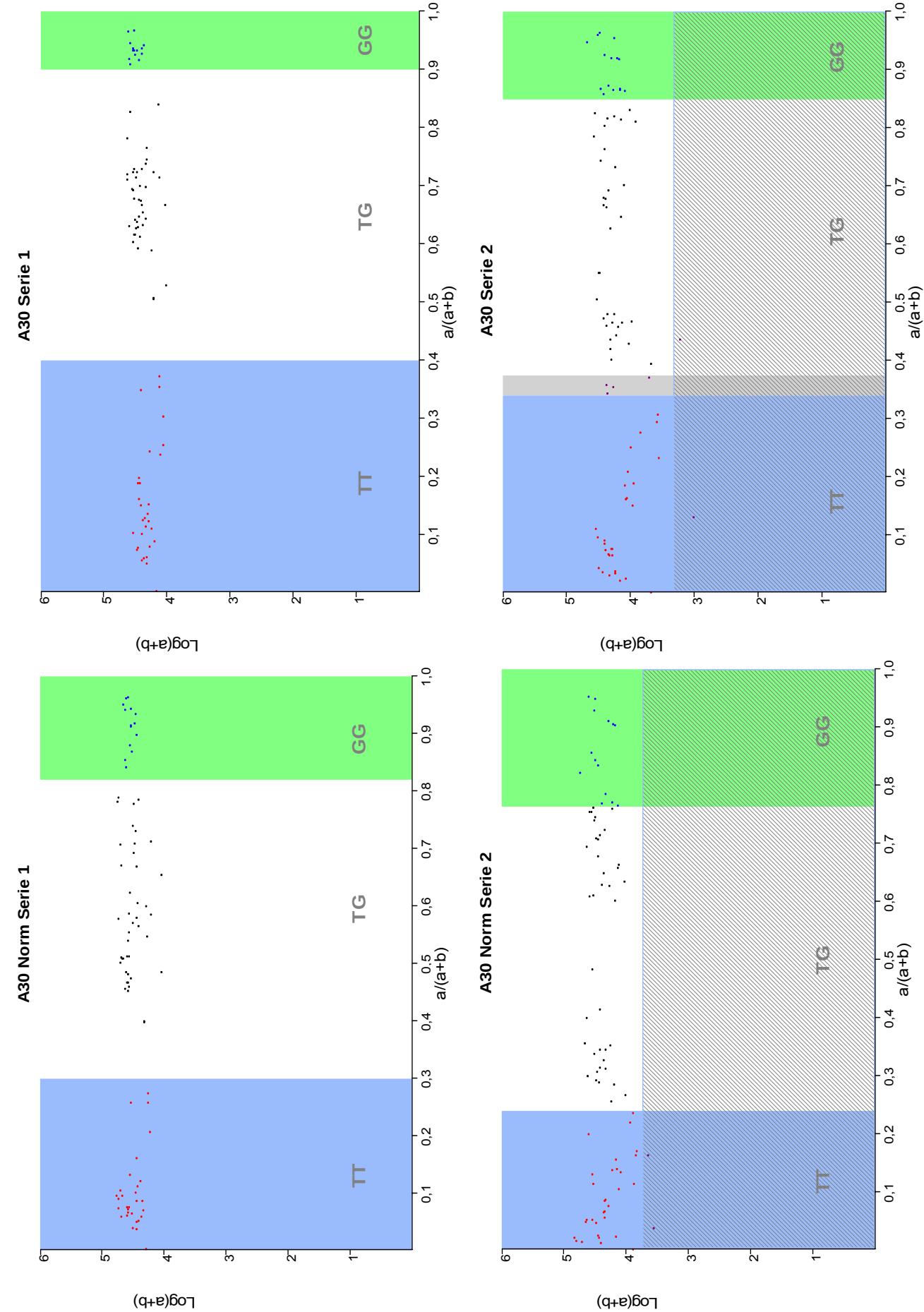


Abbildung A-19: Plotdarstellung der Genotypcluster für den SNP A32.

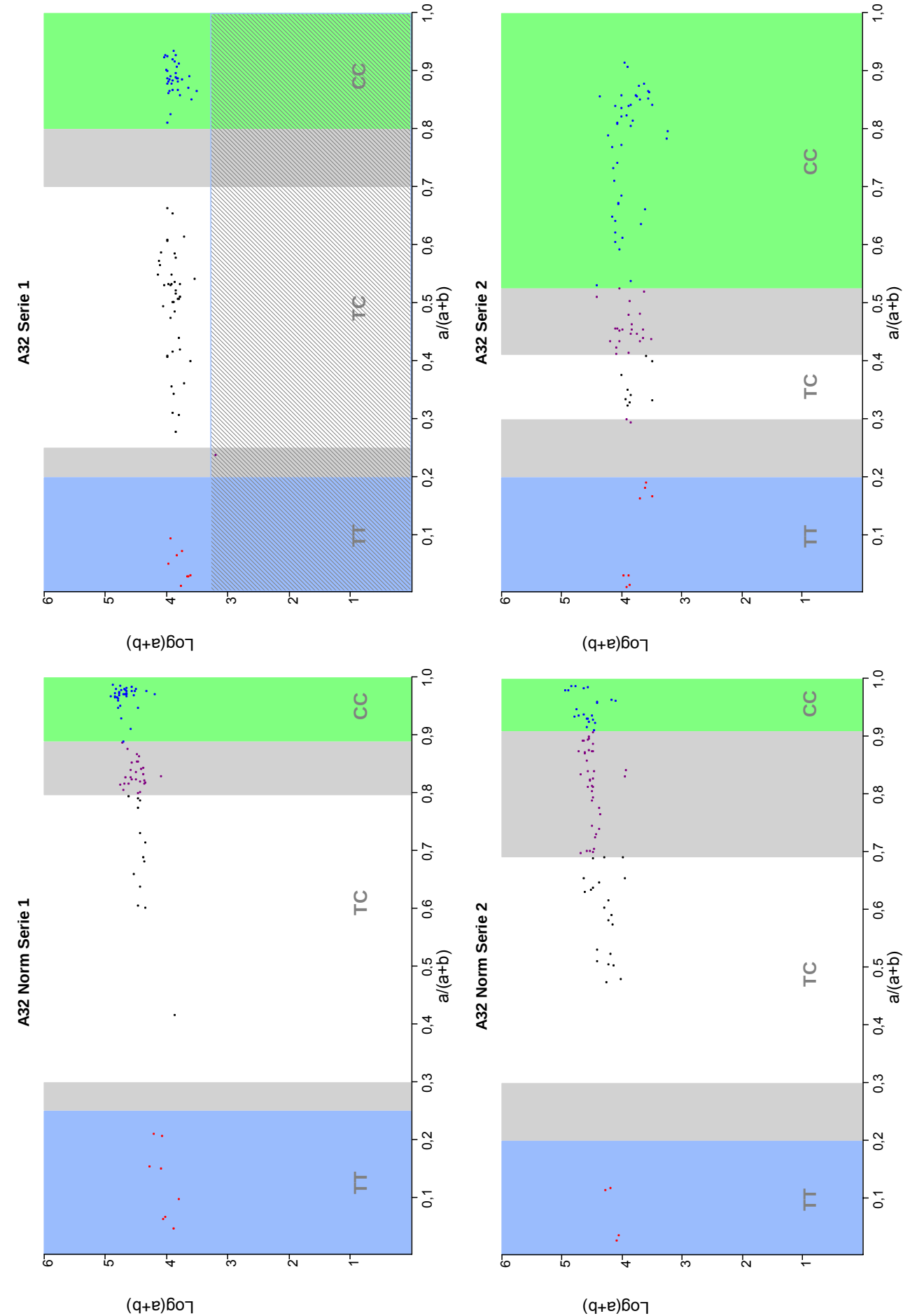


Abbildung A-20: Plotdarstellung der Genotypcluster für den SNP A33.

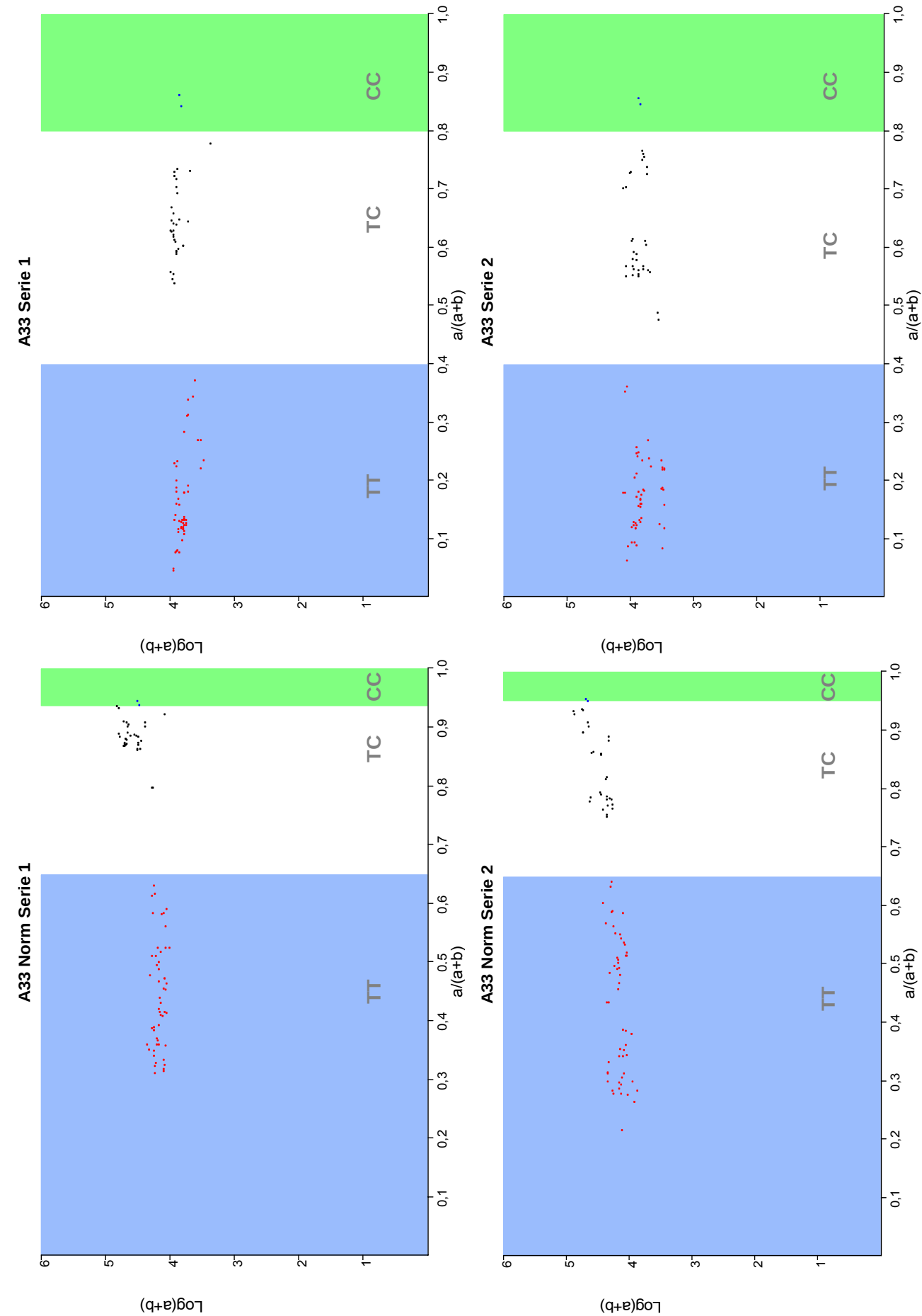


Abbildung A-21: Plotdarstellung der Genotypcluster für den SNP A34.

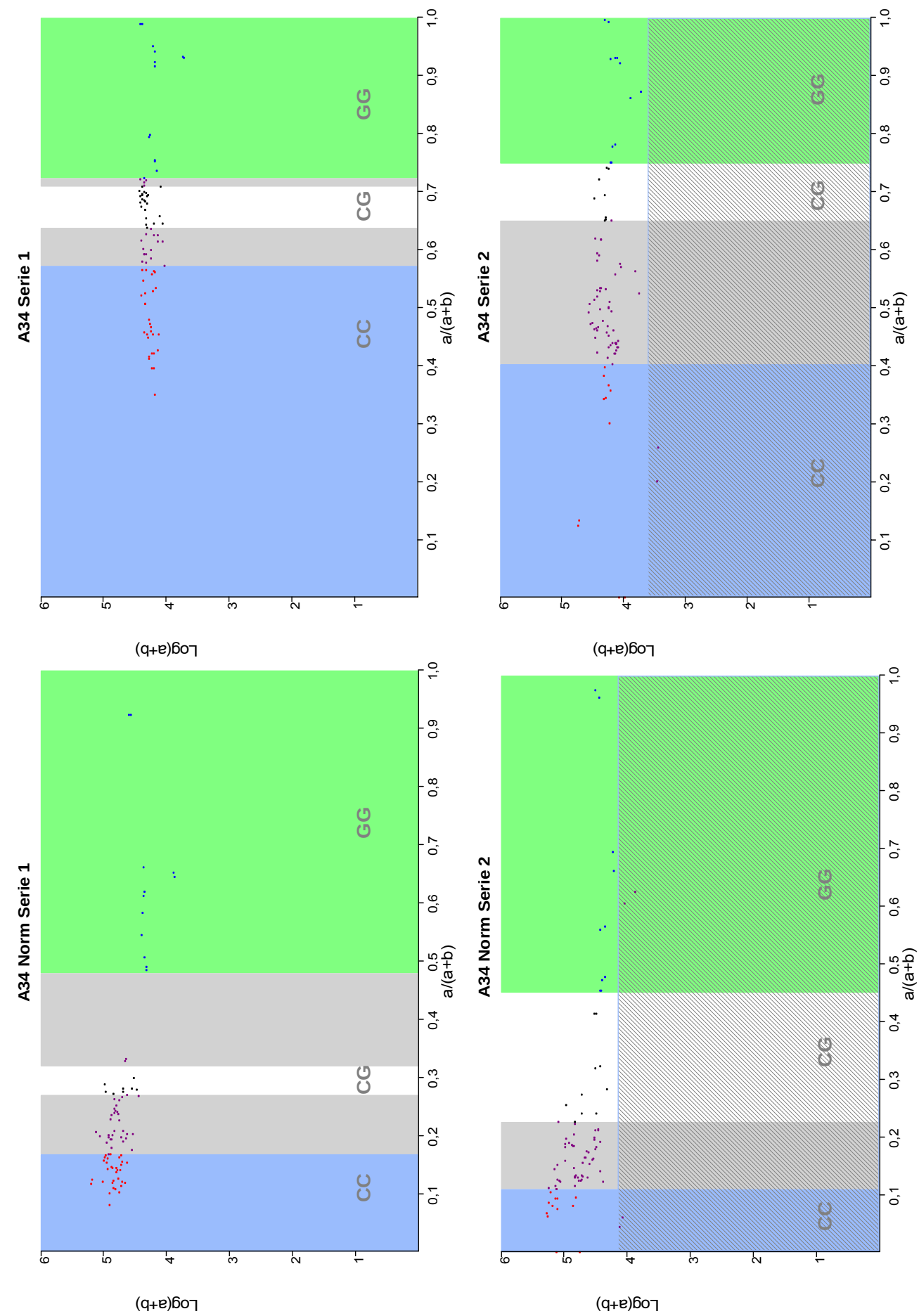


Abbildung A-22: Plotdarstellung der Genotypcluster für den SNP A35.

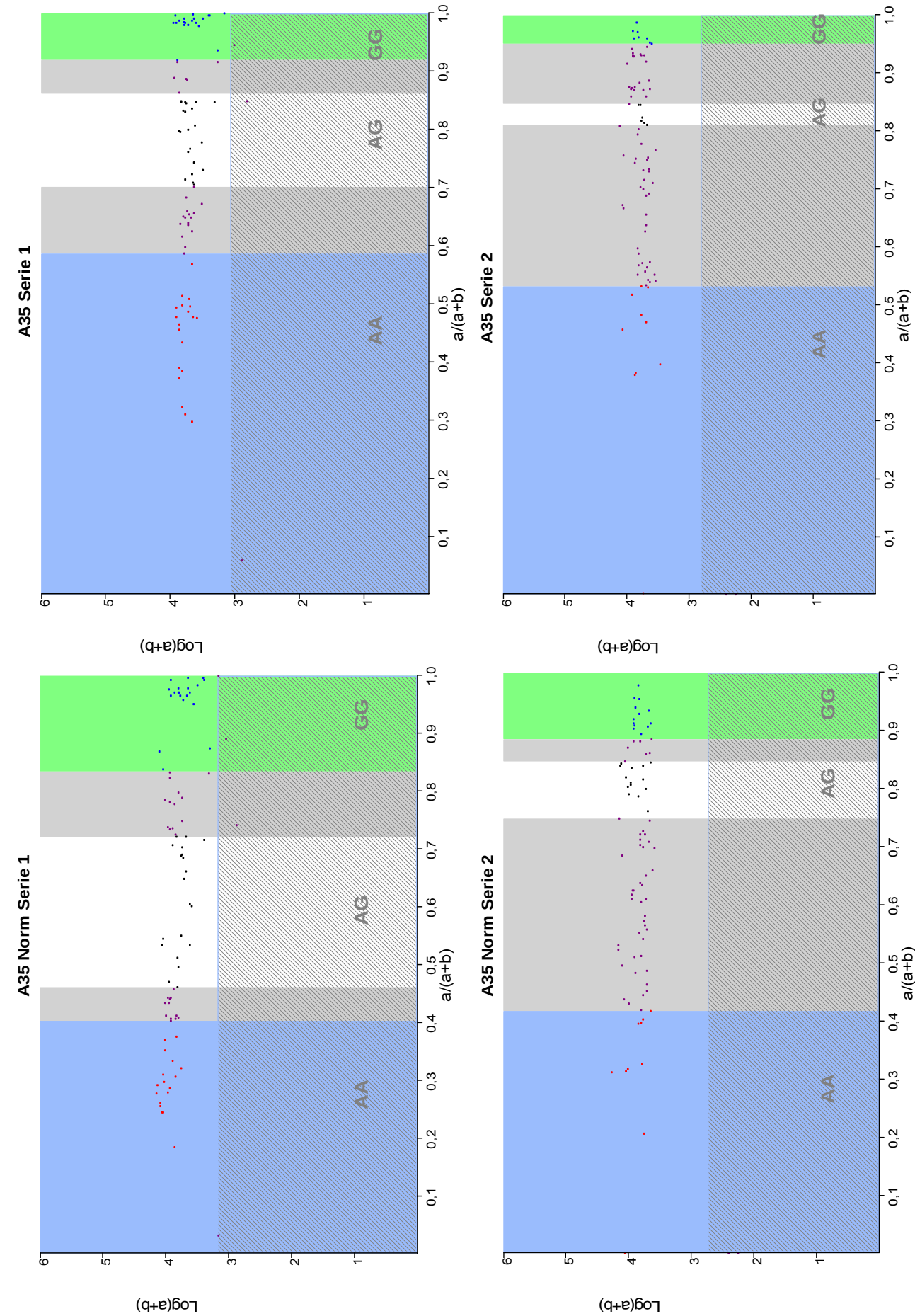


Abbildung A-23: Plotdarstellung der Genotypcluster für den SNP A37.

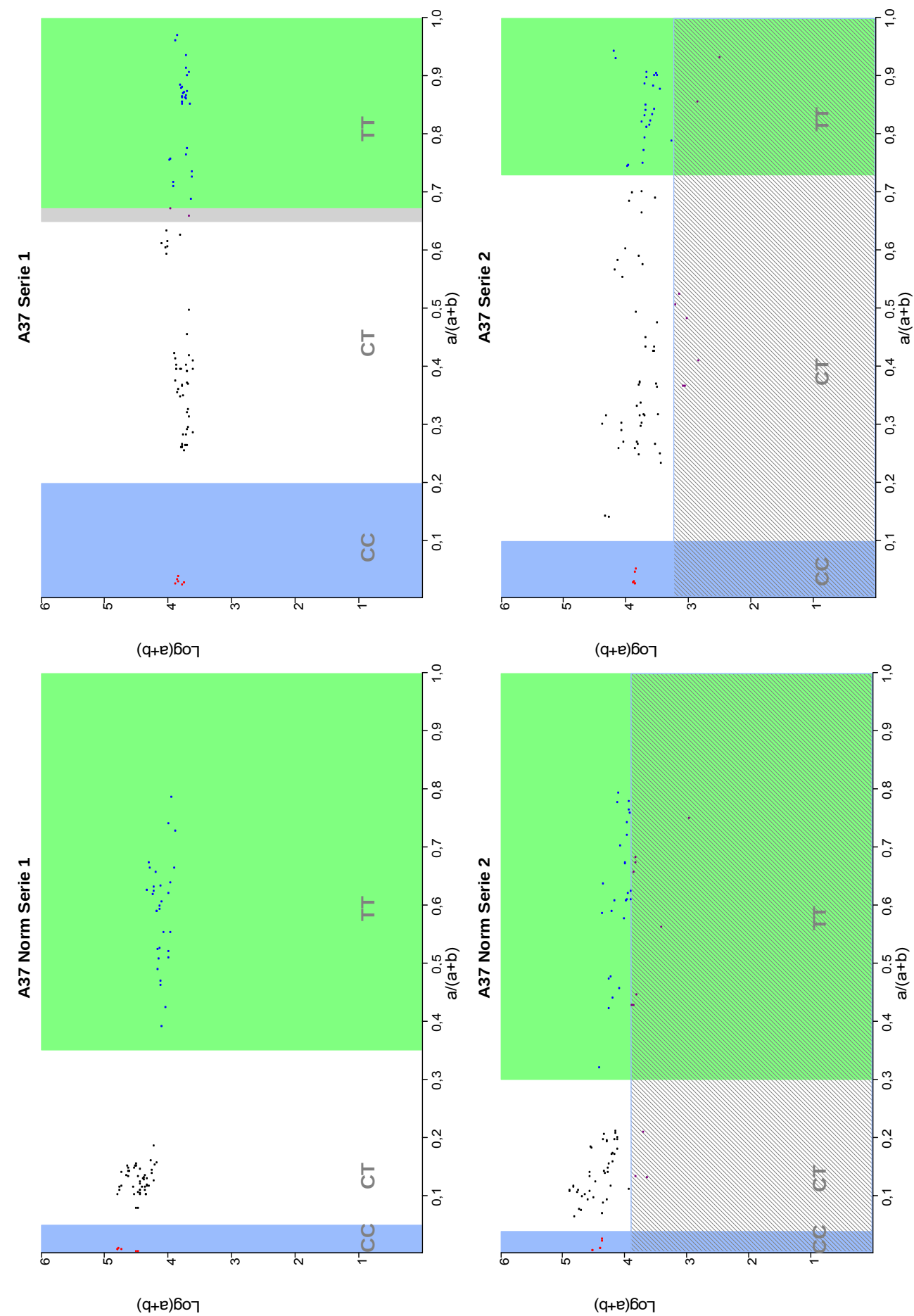


Abbildung A-24: Plotdarstellung der Genotypcluster für den SNP A38.

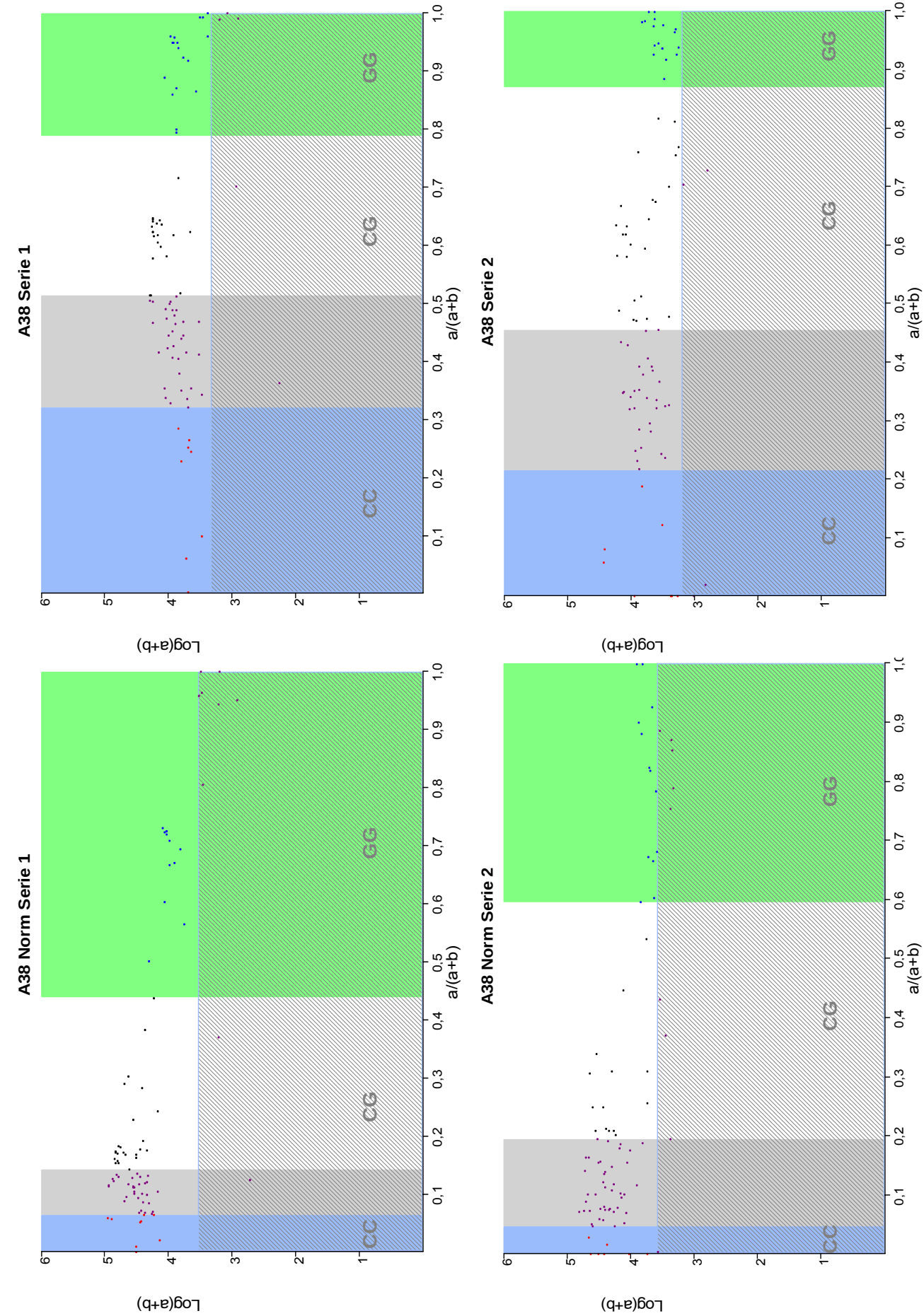


Abbildung A-25: Plotdarstellung der Genotypcluster für den SNP A39.

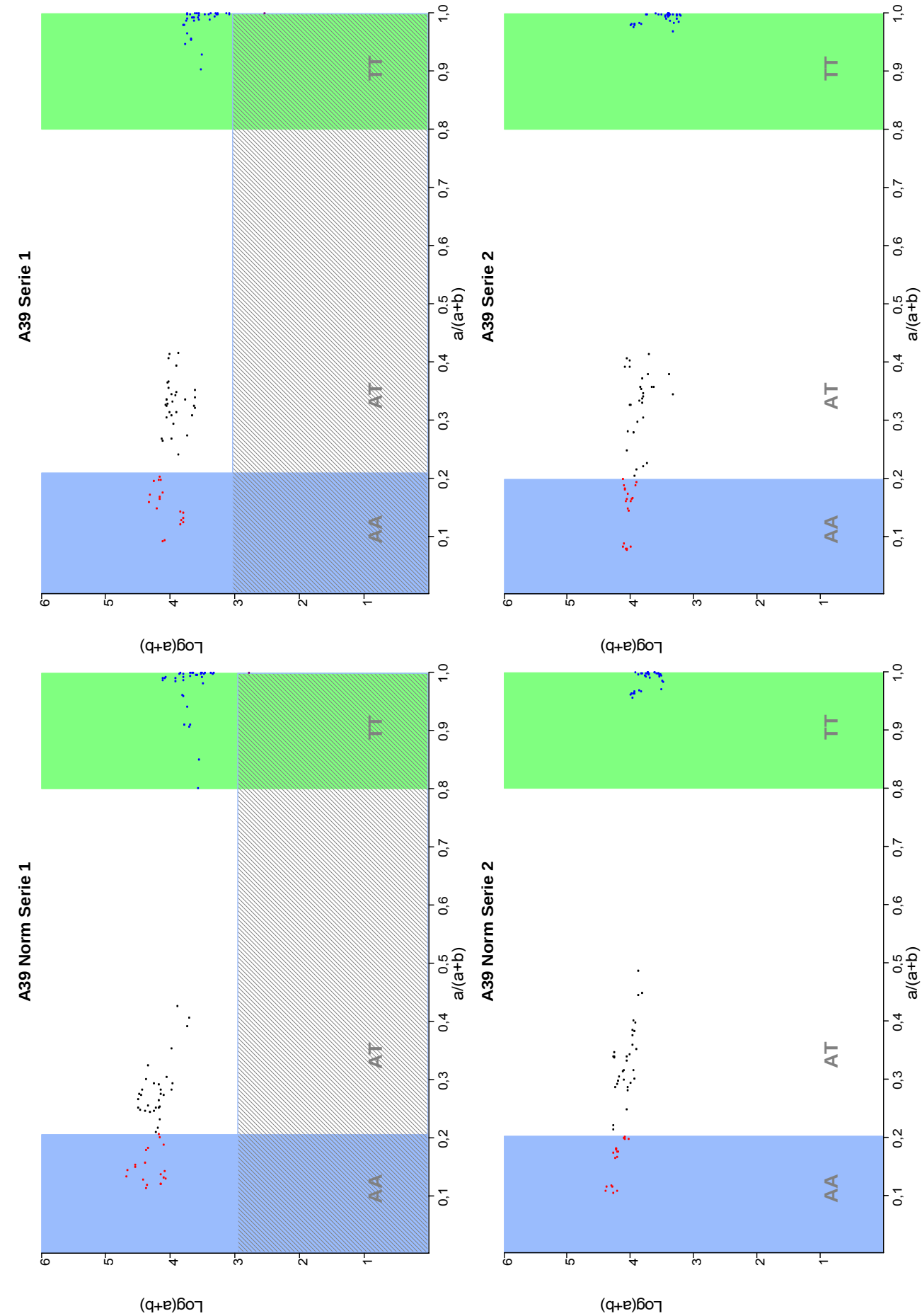


Abbildung A-26: Plotdarstellung der Genotypcluster für den SNP A41.

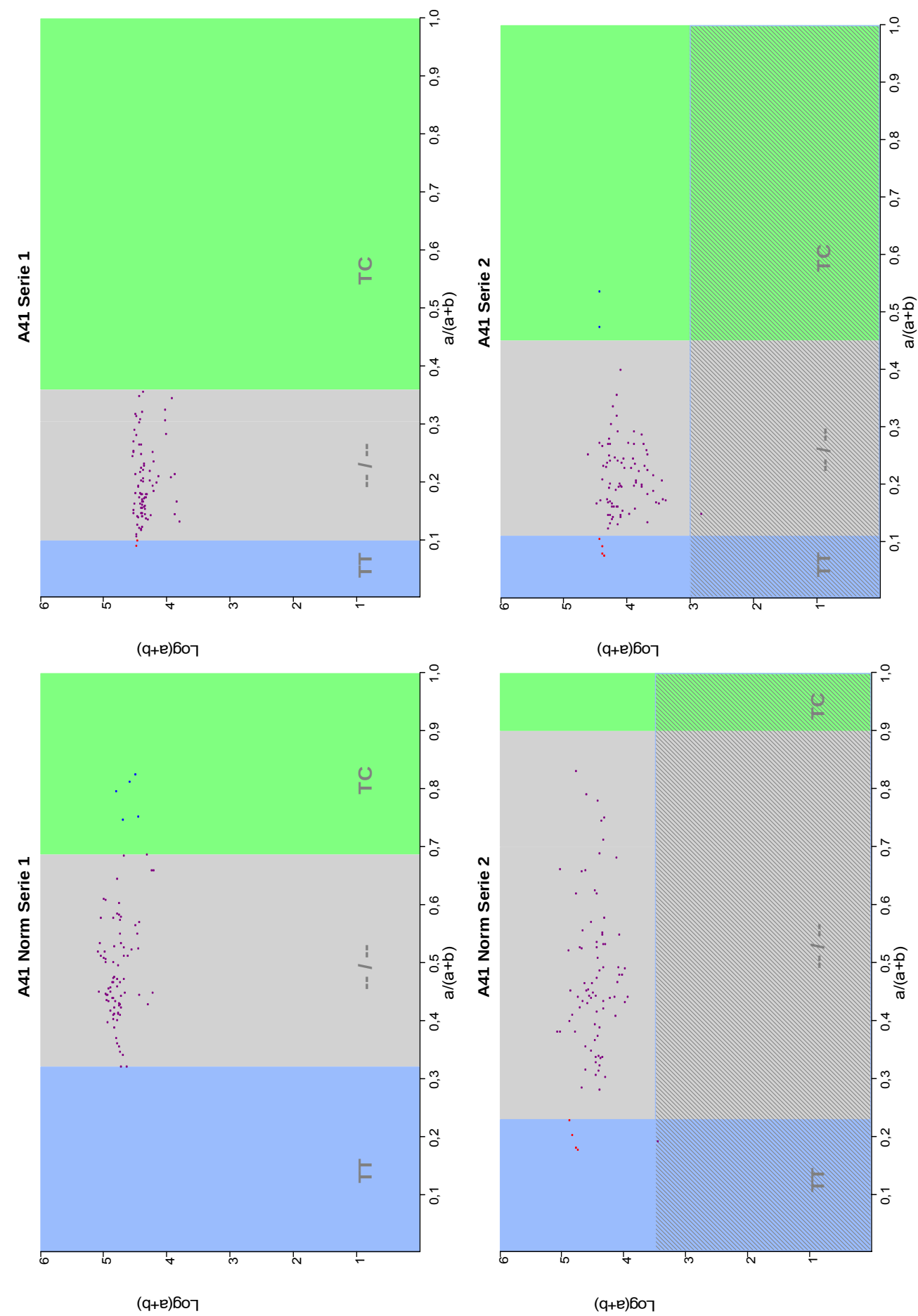


Abbildung A-27: Plotdarstellung der Genotypcluster für den SNP A42.

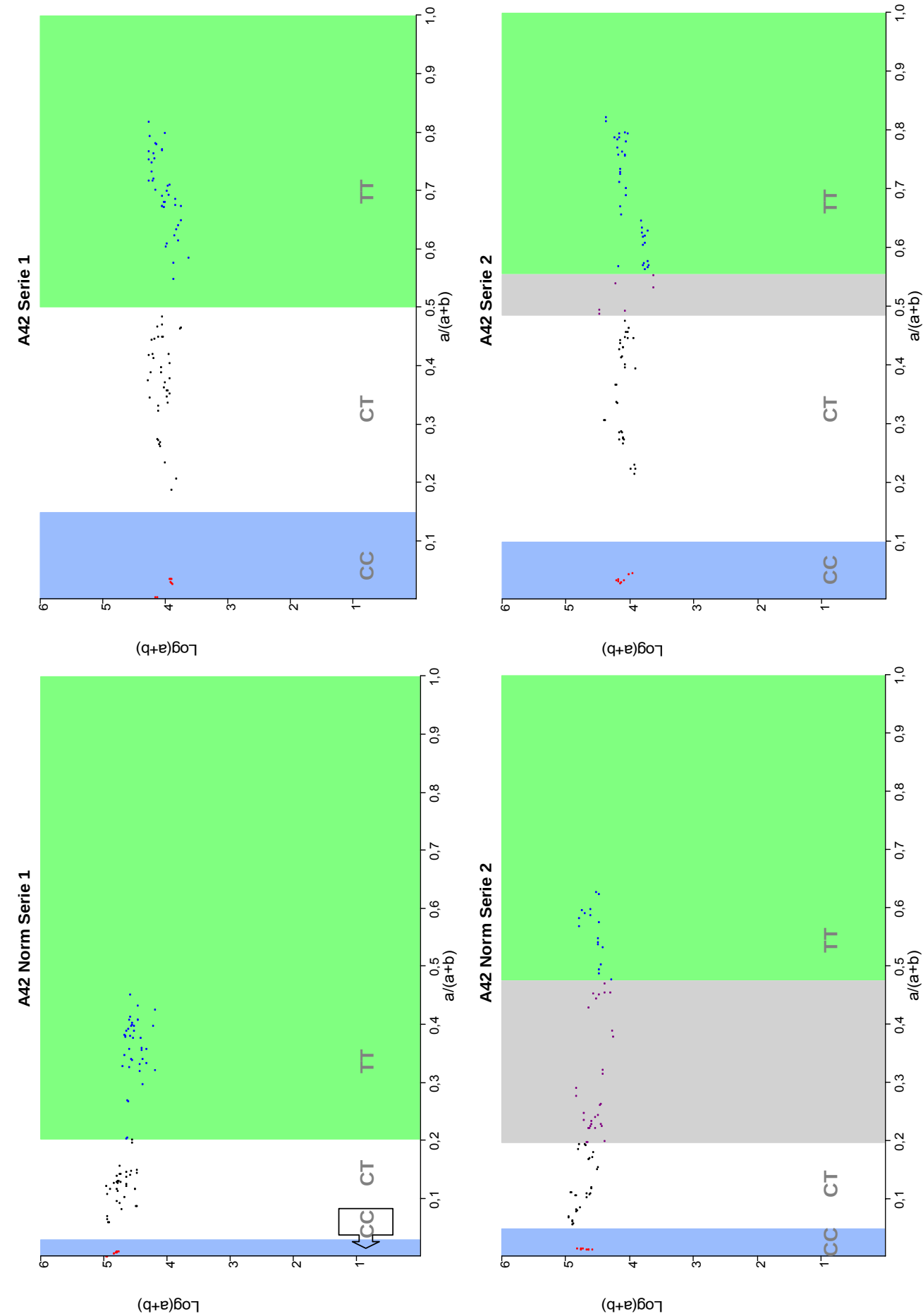


Abbildung A-28: Plotdarstellung der Genotypcluster für den SNP A43.

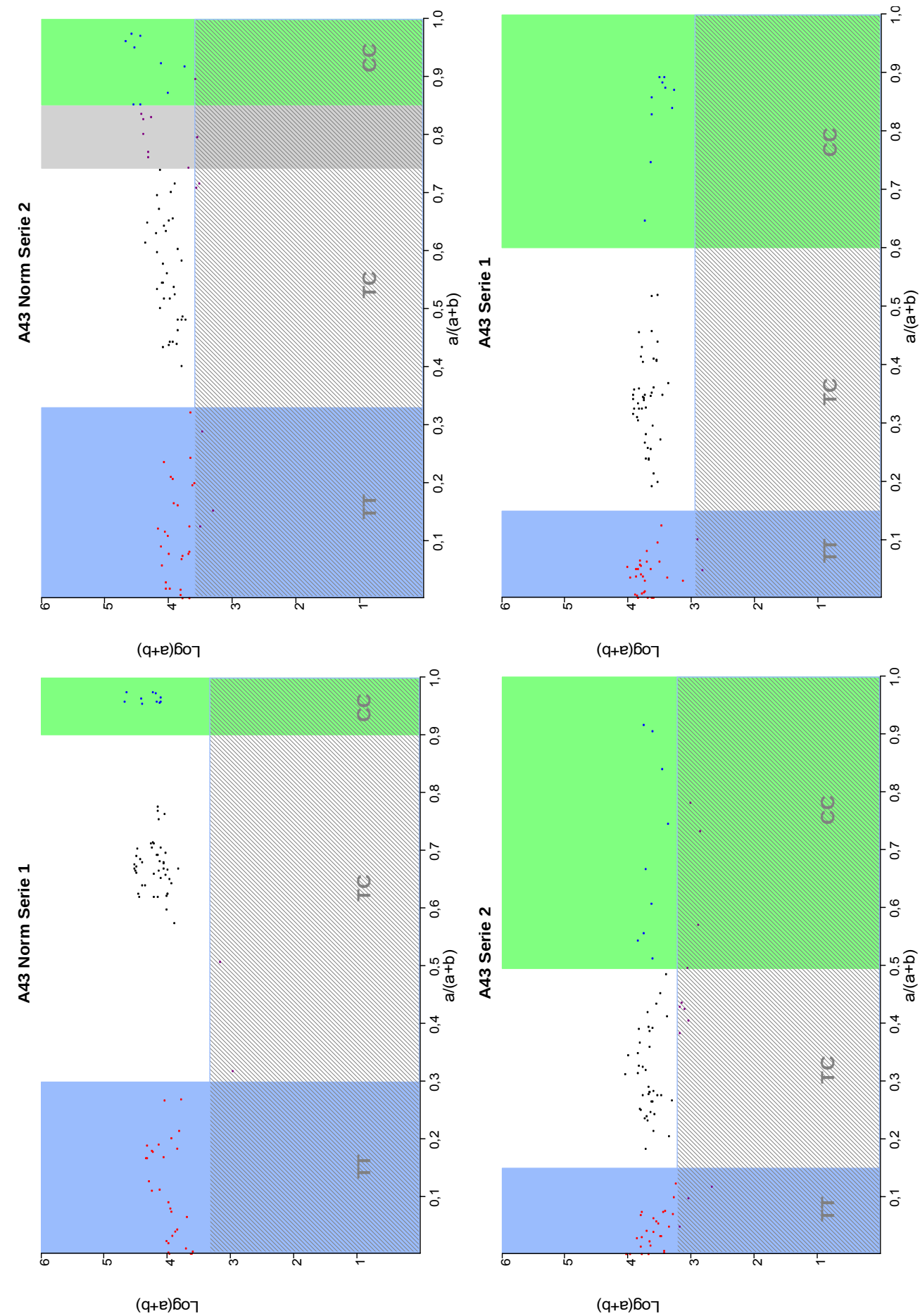


Abbildung A- 29: Plotdarstellung der Genotypcluster für den SNP A45.

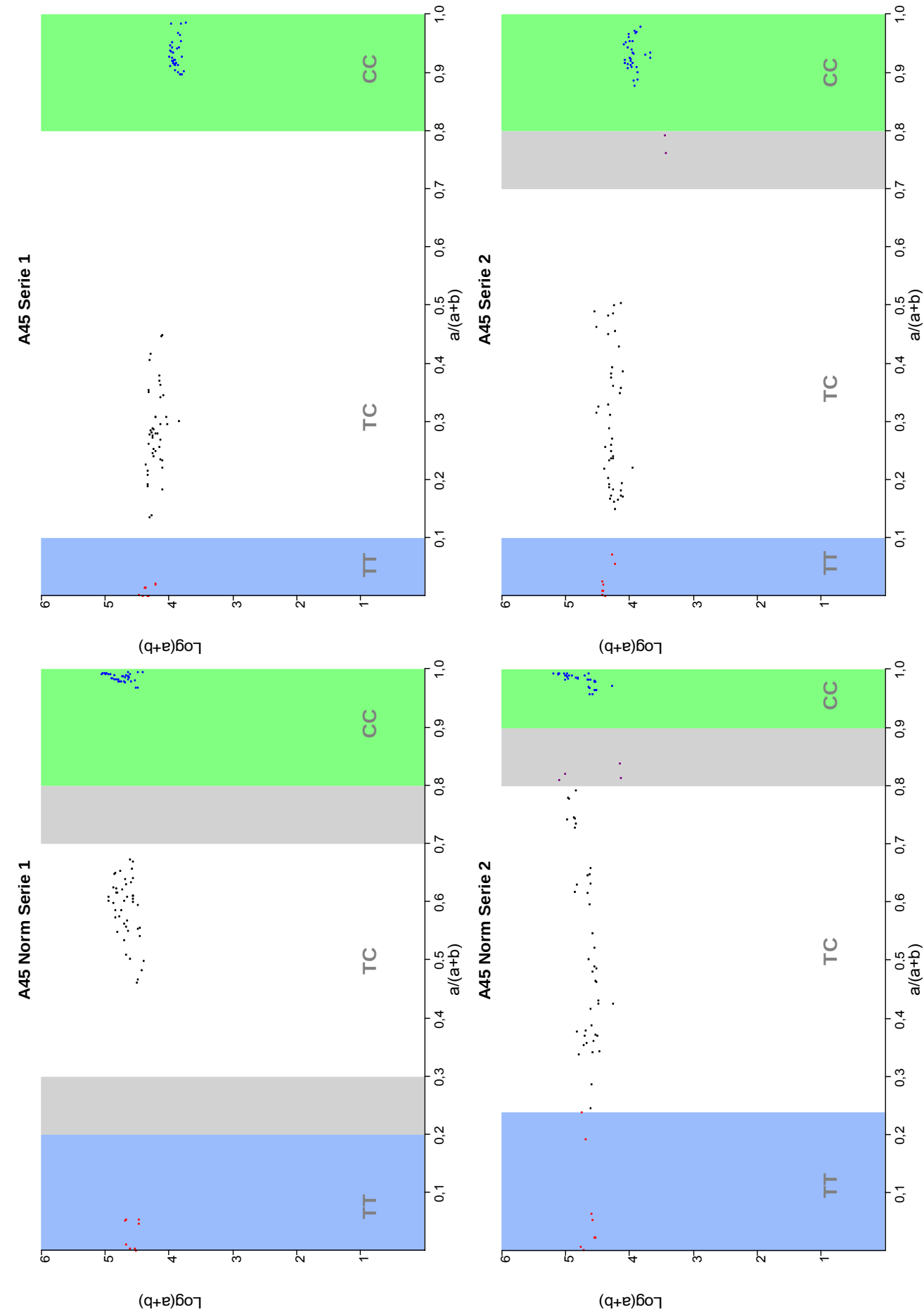


Abbildung A-30: Plotdarstellung der Genotypcluster für den SNP A48.

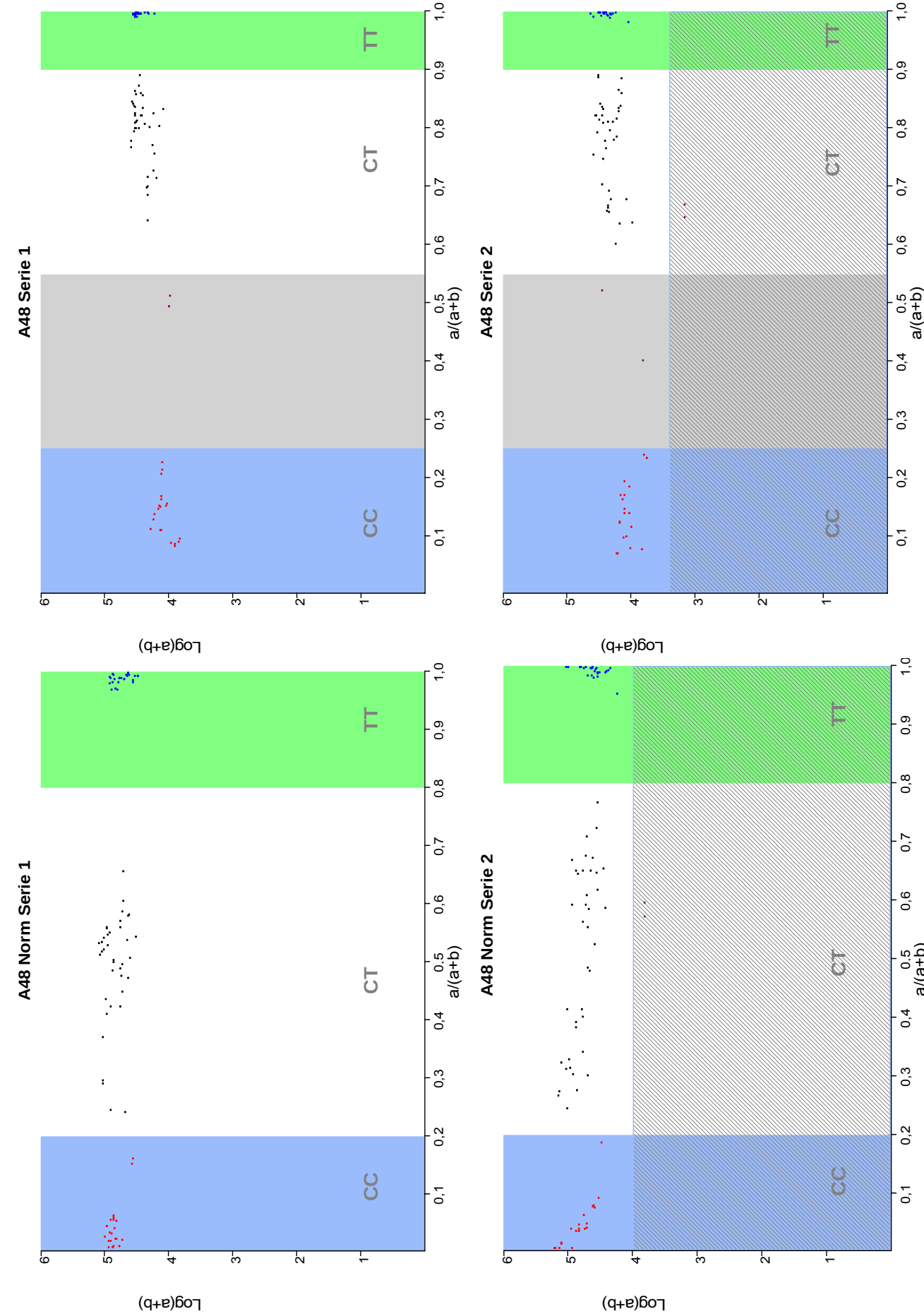


Abbildung A-31: Plotdarstellung der Genotypcluster für den SNP A49.

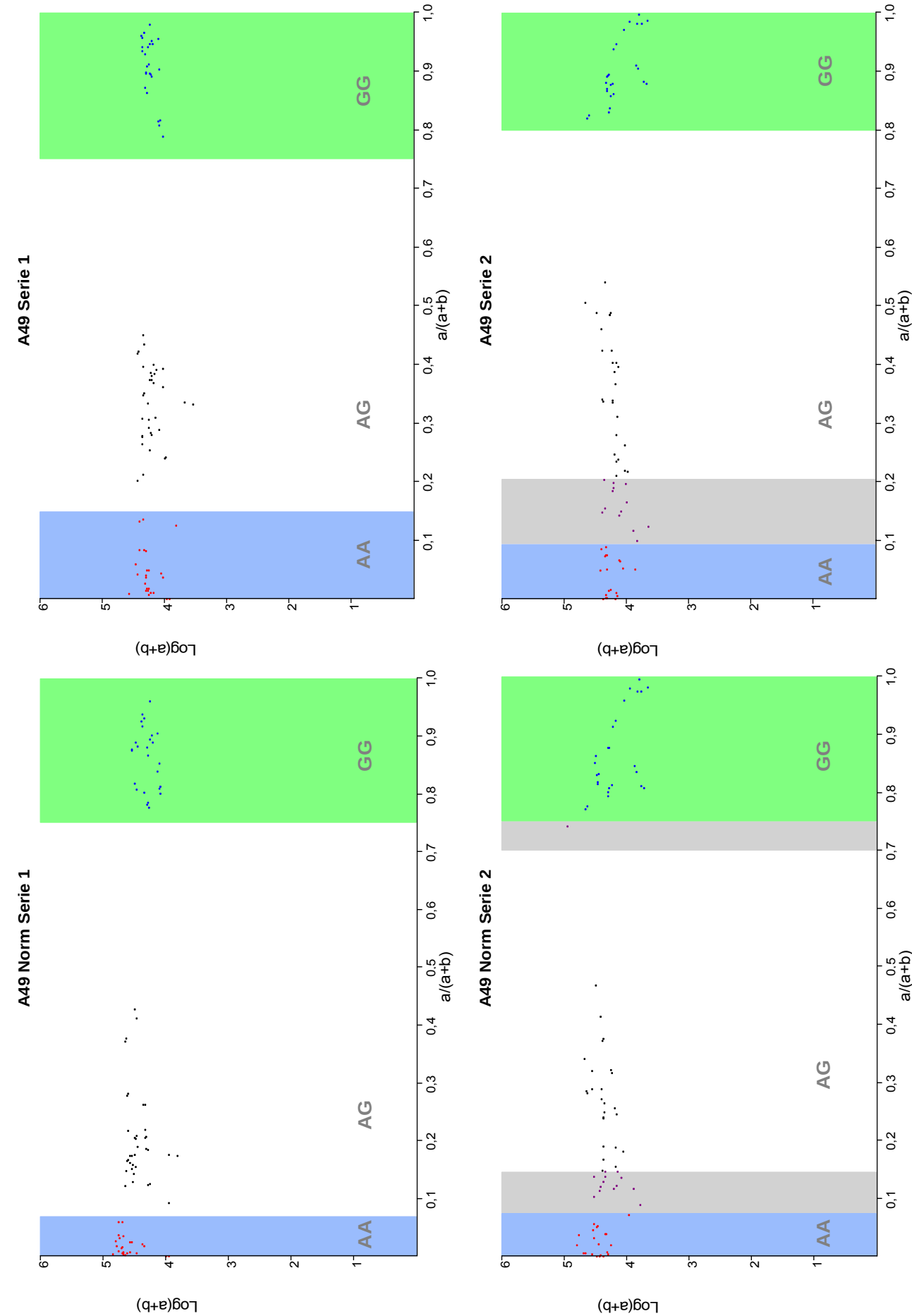


Abbildung A-32: Plotdarstellung der Genotypcluster für den SNP A50.

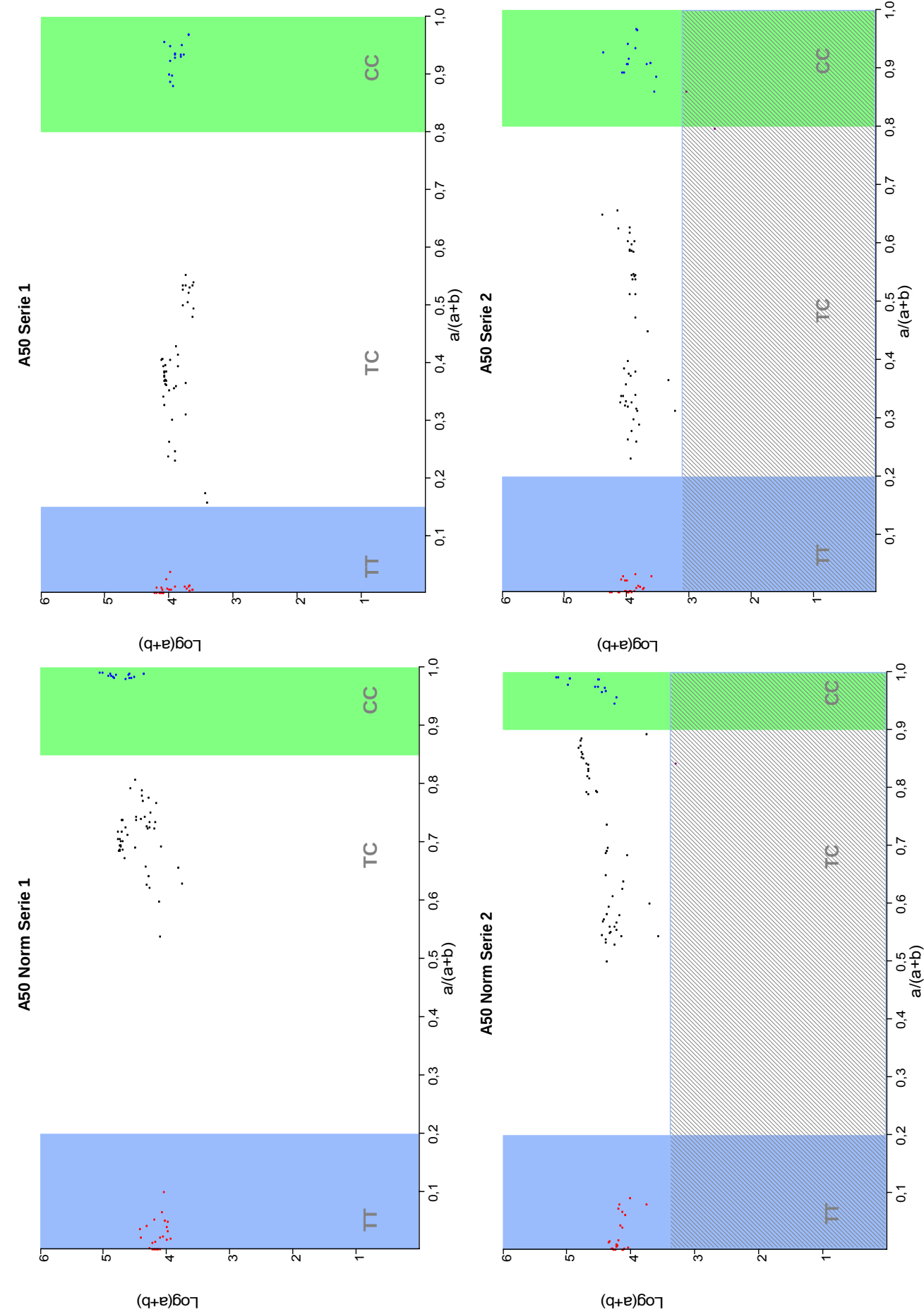


Abbildung A-33: Plotdarstellung der Genotypcluster für den SNP A53.

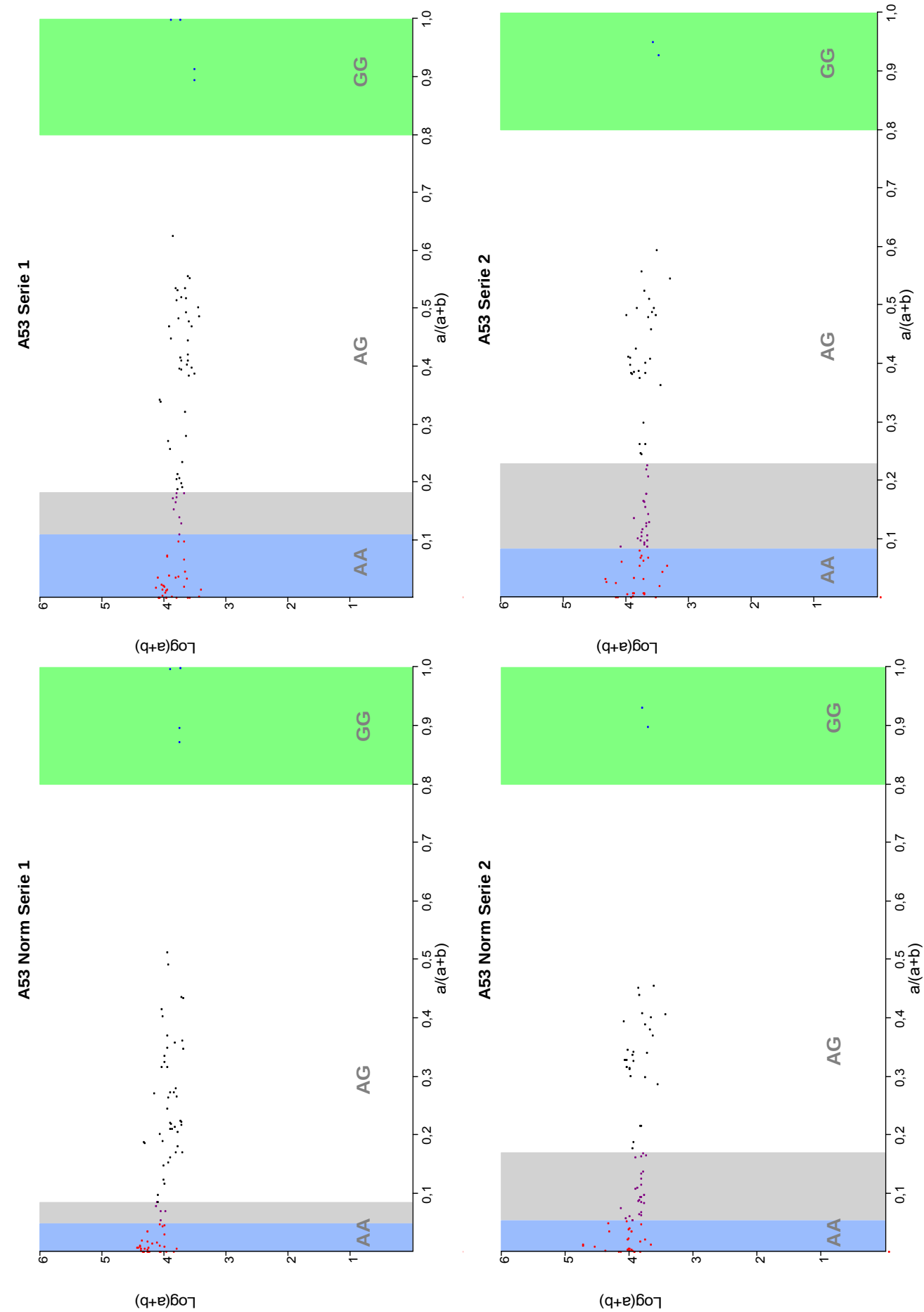


Abbildung A-34: Plotdarstellung der Genotypcluster für den SNP A54.

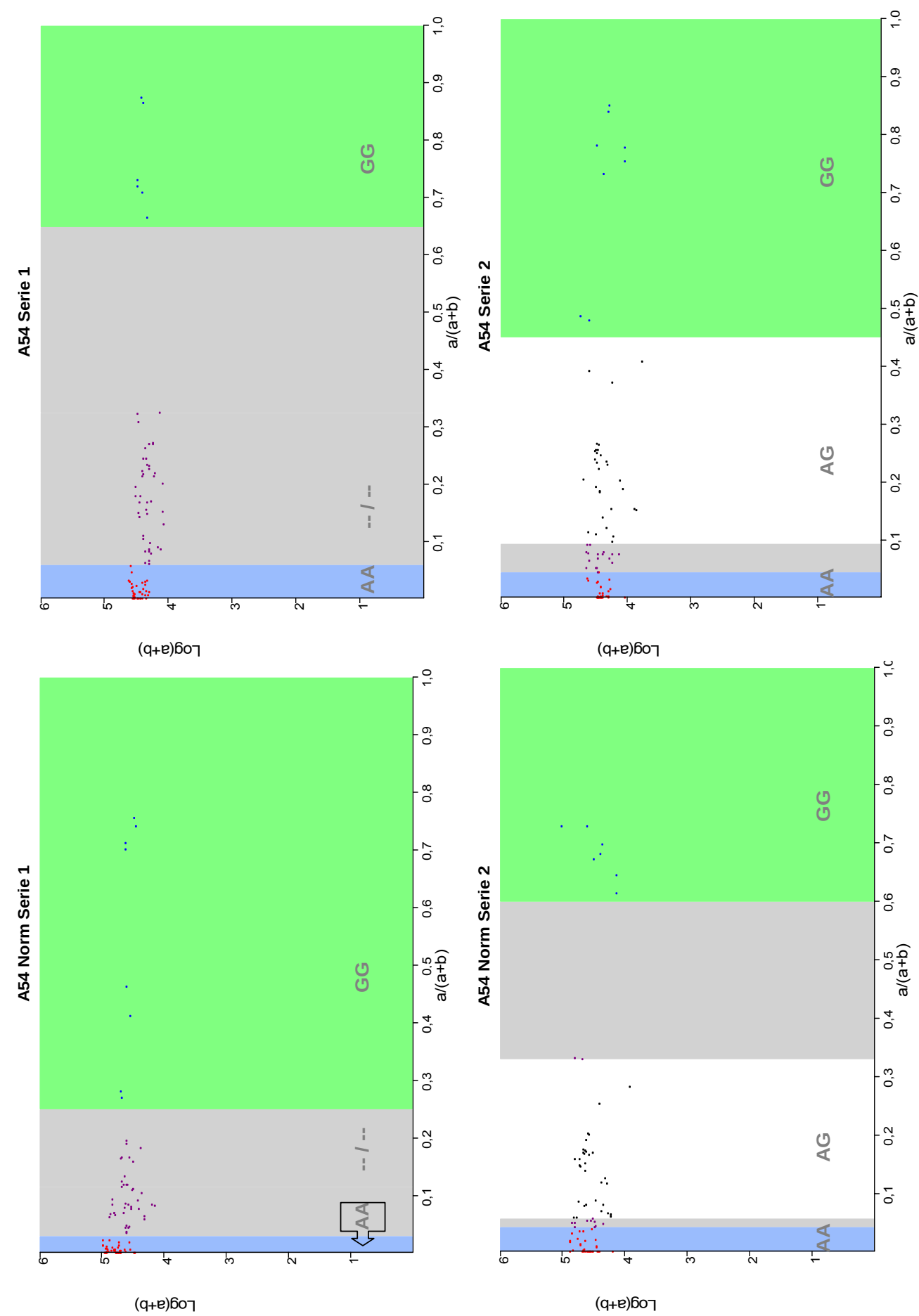


Tabelle A-9: Die zur Evaluierung der Microarray-Analyse hinzugezogene Ergebnistabelle aller mit der 33 SNP Multiplexreaktion doppelt typisierten Proben (Felder ohne farbigen Hintergrund) und der aus den Referenzmethoden ermittelten Genotypen (Genplex-Methode in hellgrün, *SNaPshot*TM-Methode, orange und grau unterlegt). Gelb, rot und blau hinterlegte Felder = keine Ergebnisse in der jeweiligen Methode. Grüne Felder: nacheditierte Ergebnisse aus der Microarray-Analyse.

	SNP	A1	A2	A4	A5	A8	A9	A23	A11	A18	A16	A20	A32	A21	A45	A15	A38	A37	A34	A43	A30	A41	A35	A48	A39	A25	A50	A54	A27	A33	A49	A24	A42	A53		
Sample 1	E34a	CT	CC	AT	GG	CT	GG	GG	TT	GT	GT	CC	GG	CT	AA	CC	GG	CT	GG	AA	AC	GG	GG	GG	GG	CT	AG	AA	NN	GG	CT	AG	NN	AG	AG	
	E34a	CT	CC	AT	GG	CT	GG	GG	TT	GT	GT	CC	GG	CT	AA	CC	GG	CT	GG	AA	AC	GG	GG	GG	GG	CT	AG	AA	NN	GG	CT	AG	NN	AG	AG	
	E34a	CT	CC	AT	GG	CT	GG	GG	TT	GT	GT	CC	GG	CT	AA	CC	GG	CT	GG	AA	AC	GG	GG	GG	GG	CT	AG	AA	NN	GG	CT	AG	NN	AG	AG	
	E34a	NN	GG	TA	CC	AG	-	GG	TT	AC	NN	GG	CT	AA	CC	TT	CC	CT	GG	TT	TG	CC	GG	CC	TT	AG	TT	NN	CC	TC	AG	TT	CT	AG	AG	
	E34b	CT	CC	AT	GG	CT	CC	GG	TT	GT	GT	CC	GG	CT	AA	CC	GG	CT	CC	AA	AC	GG	GG	GG	CT	AG	AA	NN	GG	CT	AG	TT	CT	AG	AG	
Sample 2	E34b	CT	CC	AT	GG	CT	GG	TT	GT	GT	CC	CC	AG	AA	CC	GG	CC	GG	AA	AC	GG	GG	GG	CC	TT	NN	AA	AA	NN	GG	CT	AG	TT	CT	AG	AG
	E34b	AG	GG	TA	CC	AG	NN	GG	TT	AC	NN	GG	CT	AA	CC	TT	CC	CT	GG	TT	TG	-	NN	CC	TT	AG	TT	AG	CC	TC	AG	TT	CT	AG	AG	
	E35a	TT	CC	NN	CC	GG	AG	CT	GT	GT	GT	CC	GG	AA	CC	CC	CG	CT	CG	AA	AC	CC	AA	CC	TT	AA	GG	NN	GG	TT	AG	CT	AA	AG	AG	
	E35a	TT	CC	AT	AG	CC	CC	CT	GT	GT	GT	CC	GG	AA	CC	CC	CG	CT	CG	AA	AC	GG	AA	CC	TT	AA	GG	NN	GG	TT	AG	CT	AA	AG	AG	
	E35a	TT	CC	AT	AG	CC	CC	CT	GT	GT	GT	CC	GG	AA	CC	CC	CG	CT	CG	AA	AC	GG	AA	CC	TT	AA	GG	NN	GG	TT	AG	CT	AA	AG	AG	
Sample 3	E35b	TT	CC	TA	TC	GG	GG	GG	CT	GT	GT	GG	CC	AA	CC	CC	CG	CT	CG	TT	TG	CC	AA	CC	TT	AA	GG	AG	GG	TT	AG	CT	AA	AG	AG	
	E35b	TT	CC	AT	AG	CC	CC	CC	CT	GT	GT	CC	GG	AA	CC	CC	CG	CT	CG	AA	AC	GG	AA	CC	TT	AA	GG	GG	GG	TT	AG	CT	AA	AG	AG	
	E35b	AA	NN	TA	TC	NN	-	-	-	AC	NN	CC	TT	CC	AA	CC	NN	CT	GG	TT	TG	-	AA	CC	TT	AA	CC	NN	CC	TT	AG	CT	NN	AG	AG	
	E64a	CT	CT	TT	NN	CT	GC	GG	CT	TT	GT	GT	CC	AA	CC	CC	CG	TT	GG	AA	AA	CC	AA	TT	TT	AA	AG	AA	AA	GG	CT	AA	CT	AG	AA	
	E64a	CT	CT	AA	AA	CT	CG	CT	CT	TT	GT	AC	GG	AA	CC	CC	CG	TT	CC	AA	AA	GG	AA	TT	TT	AA	AG	AA	AA	GG	CT	AA	CT	AG	AA	
Sample 4	E64a	AG	AG	AA	TT	AG	GC	-	-	AA	TG	TG	CC	AA	CC	TT	CG	TT	CC	TT	TT	-	AA	TT	TT	AA	AG	TC	AA	CC	TC	AA	CT	CT	AG	AA
	E64b	CT	CT	AA	AA	CT	CG	GG	CT	TT	GT	AC	GG	AA	CC	CC	CG	TT	CC	AA	AA	GG	AA	TT	TT	AA	AG	AA	AA	GG	CT	AA	CT	AG	AA	
	E64b	AG	AG	AA	TT	AG	GC	GG	-	AA	TG	-	-	AA	CC	TT	-	TT	-	TT	TT	-	AA	TT	TT	AA	TC	AA	CC	TC	AA	CT	CT	AG	AA	
	E67a	CC	CT	TT	AG	CT	GC	AA	TT	GT	GT	GG	CC	AA	CT	TT	CG	TT	CC	AA	AC	CT	AG	CT	TT	AA	AG	GG	GG	TT	AA	TT	AA	AA	AA	
	E67a	CC	CT	AA	AG	CT	CG	TT	GT	GT	GT	CC	GG	AA	CT	CC	CG	TT	GG	AA	AC	AG	CT	TT	AA	AG	GG	GG	GG	TT	AA	TT	AA	AA	AA	
Sample 5	E67a	CC	CT	AA	AG	TC	AG	-	-	AC	TG	GG	CC	AA	TC	TT	CG	TT	GG	TT	TG	-	AG	CT	TT	AA	AG	TC	AA	CC	TT	AA	TT	TT	AA	AA
	E67b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	E67b	CC	CT	AA	AG	CT	CG	GG	TT	GT	GT	CC	GG	AA	CT	CC	CG	TT	GG	AA	AC	AG	AG	CT	TT	AA	AG	GG	GG	TT	AA	TT	AA	AA	AA	
	E67b	GG	AG	AA	TC	AG	-	AA	-	AC	TG	GG	CC	AA	TC	TT	-	TT	GG	TT	TG	-	CT	CT	TT	AA	TC	GG	CC	TT	AA	TT	-	AA	AA	
	E73a	CC	CC	AT	GG	CT	GC	AG	TT	TT	GT	TT	CC	AG	CT	TT	CG	TT	GC	AA	AC	CT	AA	TT	AT	AG	AG	AA	AA	AA	TT	AG	CT	AA	AA	
Sample 6	E73a	CC	CC	AT	GG	CT	CG	TT	TT	TT	GT	AA	GG	AG	CT	CC	CG	TT	GC	AA	AC	AG	AA	TT	AT	NN	AG	NN	AA	AA	TT	AG	CT	AG	AG	
	E73a	CC	CC	AT	GG	CT	CG	TT	TT	TT	GT	AA	GG	AG	CT	CC	CG	TT	GC	AA	AC	AG	AA	TT	AT	NN	AG	NN	AA	AA	TT	AG	CT	AG	AG	
	E73a	GG	GG	TA	CC	AG	GC	-	TT	AA	TG	TT	CC	AG	TC	TT	CG	TT	GG	TT	TG	-	AA	TT	AT	AG	TC	-	TT	TT	AG	CT	TT	AA	AA	
	E73b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	E73b	CC	CC	AT	GG	CT	CG	GG	TT	TT	GT	AA	GG	AG	CT	CC	CG	TT	GC	AA	AC	AG	AA	TT	AT	NN	AG	NN	AA	AA	TT	AG	CT	AG	AG	
Sample 7	E73b	GG	GG	TA	CC	AG	GC	GA	-	AA	-	TT	-	AG	CT	TT	CG	TT	GC	AA	TG	-	CT	CT	AT	AG	TC	AG	CC	AG	TT	NN	CT	TT	AA	AA
	E82a	CC	CC	AT	AG	CT	CC	AG	CT	TT	GT	GT	CC	AG	CC	CC	CG	CT	CC	AG	AC	CT	AG	CT	AT	AG	AA	AA	AA	AG	CT	AA	CT	AG	AG	
	E82a	CC	CC	AT	AG	CT	GG	CC	CT	TT	GT	AC	AG	AG	CC	CC	CG	CT	GG	AG	AC	AG	AG	CT	AT	NN	AG	AA	AG	CT	AA	CT	AA	CT	AG	
	E82a	CC	CC	AT	AG	CT	GG	CC	CT	TT	GT	AC	AG	AG	CC	CC	CG	CT	GG	AG	AC	AG	AG	CT	AT	NN	AG	AA	AG	CT	AA	CT	AA	CT	AG	
	E82a	GG	GG	TA	TC	AG	GC	-	TC	AA	TG	TG	CT	AG	CC	TC	GG	CT	GG	TC	TG	-	-	CT	AT	AG	TC	AA	CT	TC	AA	CT	CT	AG	AG	
Sample 8	E82b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	E82b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	E82b	CC	CC	AT	AG	CT	GG	AG	CT	TT	GT	AC	AG	AG	CC	CC	CG	CT	GG	AG	AC	AG	AG	CT	AT	NN	AG	AA	AG	CT	AA	CT	CT	AG	AG	
	E82b	GG	GG	TA	TC	AG	GC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	E82b	GG	GG	TA	TC	AG	GC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

SNP	A1	A2	A4	A5	A8	A9	A23	A11	A18	A16	A20	A32	A21	A45	A15	A38	A37	A34	A43	A30	A41	A35	A48	A39	A25	A50	A54	A27	A33	A49	A24	A42	A53	
7	E86a	CT	CT	AT	AG	CT	CC	AA	TT	GT	TT	CC	AG	CC	CC	NN	TT	GC	AG	AC	TT	AA	CT	TT	AA	AG	AA	AG	TT	AA	CC	AG	AG	
	E86a	CT	CT	AT	AG	CT	GG	TT	GT	TT	AA	GG	AG	CC	CC	CC	TT	CG	AG	AC	AA	AA	CT	TT	AA	AG	AA	AG	TT	AA	CC	AG		
	E86a	CT	CT	AT	AG	CT	GG	TT	GT	TT	AA	GG	AG	CC	CC	CC	TT	CG	AG	AC	AA	AA	CT	TT	AA	AG	AA	AG	TT	AA	CC	AG		
	E86a	AG	AG	TA	TC	AG	-	TT	AC	TT	TT	CC	AG	CC	CC	CC	GG	TT	GG	TC	TG	-	AA	CT	TT	AA	TC	AA	CT	TT	AA	CC	CT	AG
	E86b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	E86b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	E86b	CT	CT	AT	AG	CT	GG	TT	GT	TT	AA	GG	AG	CC	CC	CC	CC	TT	CG	AG	AC	AA	AA	CT	TT	AA	AG	AA	AG	TT	AA	CC	AG	
8	E86b	AG	AG	TA	TC	AG	CC	AA	-	AC	TT	CC	-	CC	CC	-	TT	-	TC	TG	-	AA	CT	TT	AA	TC	AA	CT	TT	AA	CC	CT	AG	
	E86b	CC	CC	AT	GG	CC	CC	AA	TT	GT	GG	CT	AA	CC	CC	CC	CG	CT	CG	AG	CC	TT	AG	TT	AT	AA	AA	AA	AG	TT	AG	CT	AA	
	1a	CC	CC	AT	GG	CC	GG	TT	TT	GT	CC	AG	AA	CC	CC	CC	CG	CT	CG	AG	CC	AA	AG	TT	AT	AA	AA	AA	AG	TT	AG	CT	AA	
	1a	CC	CC	AT	GG	CC	GG	TT	TT	GT	CC	AG	AA	CC	CC	CC	CG	CT	CG	AG	CC	AA	AG	TT	AT	AA	AA	AA	AG	TT	AG	CT	AA	
	1a	GG	GG	TA	CC	NN	CC	-	TT	AA	TG	GG	TC	AA	CC	CC	CG	CT	NN	TC	GG	-	AG	TT	AT	AA	TT	AA	CT	TT	AG	CT	TT	AA
	1b	CC	CC	AT	GG	CC	CC	AG	TT	TT	GT	GG	CT	AA	CC	CC	CC	CG	CT	CG	AG	CC	TT	AG	TT	AT	AA	AA	AG	TT	AG	CT	AA	
	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
1c	CC	CC	AT	GG	CC	GG	TT	TT	GT	CC	AG	AA	CC	CC	NN	CG	CT	CG	AG	CC	AA	AG	TT	AT	AA	AA	AA	AG	TT	AG	CT	AG		
1b	NN	GG	TA	NN	GG	-	NN	-	AA	TG	GG	TC	AA	CC	CC	CG	CT	-	TC	GG	-	TT	AT	AA	TT	AA	CT	TT	AG	CT	-	AA		
9	E95a	CC	CT	TT	AA	CC	CC	GG	TT	GT	GG	CC	AA	TT	CC	CC	CT	CG	AG	AC	CT	AG	TT	TT	AG	AA	AA	AA	GG	TT	GG	CT	AG	
	E95a	CC	CT	AA	CC	CC	GG	TT	GT	GG	CC	GG	AA	TT	CC	CC	CG	CT	CG	AG	AC	AG	TT	TT	NN	AA	AA	AA	GG	TT	GG	CT	AG	
	E95a	CC	CT	AA	AA	CC	GG	TT	GT	GG	CC	GG	AA	TT	CC	CC	GG	CT	CG	AG	AC	AG	TT	TT	NN	AA	AA	AA	GG	TT	GG	CT	AG	
	E95a	GG	AG	AA	TT	NN	CC	-	CT	AC	GG	CC	AA	TT	CC	-	CT	CG	TC	TG	-	AG	TT	TT	AG	TT	AA	CC	CT	GG	CT	CC	AG	
	E95b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	E95b	CC	CT	AA	AA	CC	GG	TT	GT	GG	CC	GG	AA	TT	CC	CC	GG	CT	CG	AG	AC	AG	AG	TT	TT	NN	AA	AA	GG	TT	GG	CT	AG	
	E95b	GG	AG	AA	TT	GG	CC	GG	-	AC	GG	-	AA	TT	CC	-	CT	CG	TC	TG	-	-	TT	TT	AG	TT	AA	CC	CT	GG	CT	CC	-	
10	E97a	CT	TT	TT	AG	TT	CC	AG	CT	TT	GG	TT	CC	AA	CT	CT	CG	CG	AG	AA	CT	GG	TT	AT	AG	AG	AG	AG	CT	AG	CT	AG	AG	
	E97a	CT	TT	AA	AG	TT	GG	CT	TT	GG	AA	GG	AA	CT	CT	CG	CC	CG	AG	AA	AG	GG	TT	AT	NN	AG	NN	AG	CT	AG	CT	AG	AG	
	E97a	CT	TT	AA	AG	TT	GG	CT	TT	GG	AA	GG	AA	CT	CT	CG	CC	CG	AG	AA	AG	GG	TT	AT	NN	AG	NN	AG	CT	AG	CT	AG	AG	
	E97a	AG	AA	AA	TC	AA	CC	-	CT	AA	NN	TT	NN	AA	TC	TC	CG	CC	-	TC	TT	-	TT	AT	AG	TC	-	CT	TC	AG	CT	CT	AG	
	E97b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	E97b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	E97b	AG	AA	AA	TC	AA	CC	NN	-	AA	-	TT	CC	AA	TC	TC	-	CG	-	TC	TT	-	NN	TT	AT	-	TC	AG	CT	TC	AG	CT	AG	
11	E99a	CT	CT	TT	NN	CC	CC	GG	TT	GT	GG	CC	AG	CT	CT	CC	CT	CG	AG	CC	CT	AG	CC	AA	GG	AG	AA	AG	CT	AG	CT	GG	AA	
	E99a	CT	CT	AA	AG	CC	GG	TT	TT	GT	CC	GG	AG	CT	CT	CG	CC	CG	AG	CC	AG	CC	AA	GG	AG	AA	AG	AG	CT	AG	CT	GG	AA	
	E99a	CT	CT	AA	AG	CC	GG	TT	TT	GT	CC	GG	AG	CT	CT	CG	CC	CG	AG	CC	AG	AG	CC	AA	GG	AG	AA	AG	CT	AG	CT	GG	AA	
	E99a	AG	AG	AA	TC	NN	CC	-	TT	AA	TG	GG	CC	AG	TC	TC	-	CT	CG	TC	GG	-	AG	CC	AA	GG	TC	AA	CT	TC	AG	CT	CC	AA
	E99b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	E99b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	E99b	AG	AG	AA	TC	GG	CC	GG	-	AA	-	-	CC	-	TC	-	CT	CG	TC	GG	-	AG	CC	AA	GG	TC	AA	CT	TC	AG	CT	CC	AA	
12	E115a	CT	CC	TT	NN	TT	CG	GG	TT	GT	GG	CC	AA	CT	TT	GG	TT	CG	AG	AA	CC	AA	TT	AA	AG	GG	AA	AG	TT	AA	CT	GG	AG	
	E115a	CT	CC	AA	AG	TT	CG	TT	GT	GT	CC	GG	AA	CT	CC	CT	CG	AG	AA	GG	AA	TT	AA	NN	GG	AA	AG	TT	AA	CT	AA	CT	GG	
	E115a	CT	CC	AA	AG	TT	CG	TT	GT	GT	CC	GG	AA	CT	CC	CT	CG	AG	AA	GG	AA	TT	AA	NN	GG	AA	AG	TT	AA	CT	AA	CT	GG	
	E115a	AG	GG	AA	TC	AA	-	-	AC	TG	GG	CC	AA	TC	TT	GG	TT	-	TC	TT	-	AA	TT	AA	AG	CC	AA	CT	TT	AA	CT	CC	AG	
	E115b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	E115b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	E115b	CT	CC	AA	AG	TT	CG	TT	GT	GT	CC	GG	AA	CT	CC	CC	CG	CT	CG	AG	AA	GG	AA	TT	NN	GG	AA	AG	TT	AA	CT	CC	AG	
13	E115b	AG	GG	AA	TC	AA	-	GG	-	AC	-	CC	AA	TC	TT	GG	TT	-	TC	TT	-	TT	AA	-	CC	AA	CT	TT	AA	CT	CC	-	AG	
	E2125a	CC	CC	AT	AG	CT	CC	AG	CT	GT	GT	CC	AA	CT	TT	GG	CC	GG	AG	AC	CT	AA	CT	AT	AG	AG	AG	AG	TT	AG	TT	AG	AG	
	E2125a	CC	CC	AT	AG	CT	GG	CT	GT	GT	AC	GG	AA	CT	CC	CC	CC	CC	AG	AC	AG	AA	CT	AT	NN	AG	NN	AG	TT	AG	TT	AG	AG	
	E2125a	CC	CC	AT	AG	CT	GG	CT	GT	GT	AC	GG	AA	CT	CC	CC	CC	CC	AG	AC	AG	AA	CT	AT	NN	AG	NN	AG	TT	AG	TT	AG	AG	
	E2125a	GG	GG	TA	TC	AG	-	-	CT	AC	TG	CC	AA	TC	TT	CC	CC	CC	TC	TG	-	AA	CT	AT	AG	TC	-	CT	TT	AG	TT	TT	AG	
	E2125b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	E2125b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
E2125b	CC	CC	AT	AG	CT	GG	CT	GT	GT	AC	GG	AA	CT	TT	GG	CC	CC	AG	AC	AG	AA	CT	AT	NN	AG	NN	AG	TT	AG	TT	TT	-		
E2125b	GG	GG	TA	TC	AG	-	GA	-	AC	-	-	CC	AA	TC	TT	GG	CC	-	TC	TG	-	AA	CT	AT	-	TC	NN	CT	TT	AG	TT	TT	-	

SNP	A1	A2	A4	A5	A8	A9	A23	A11	A18	A16	A20	A32	A21	A45	A15	A38	A37	A34	A43	A30	A41	A35	A48	A39	A25	A50	A54	A27	A33	A49	A24	A42	A53		
14	E2152a	CC	TT	AT	GG	CT	CC	AG	CT	TT	GT	GT	AA	CC	TT	CG	TT	CG	AA	AC	CT	AG	CT	AT	AA	AG	AA	AG	AA	GG	TT	GG	TT	AG	
	E2152a	CC	TT	AT	GG	CT	GG	CT	TT	GT	AC	AG	AA	CC	AG	CG	TT	CG	AA	AC	AG	CT	AT	AA	AG	AA	AG	AA	GG	TT	GG	TT	AG		
	E2152a	CC	TT	AT	GG	CT	GG	CT	TT	GT	AC	AG	AA	CC	CC	CG	TT	CG	AA	AC	AG	AG	CT	AT	AA	AG	AA	GG	TT	GG	TT	AG			
	E2152a	GG	AA	TA	CC	AG	CC	-	TC	AA	TG	CT	AA	CC	TT	CG	TT	-	TT	TG	-	AG	CT	AT	AA	TC	AA	CC	TT	GG	TT	CT	AG		
	E2152b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
15	E2152b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	E2152b	CC	TT	AT	GG	CT	GG	GA	-	AA	-	NN	TC	AA	CC	TT	-	CG	AA	AC	AG	AG	CT	AT	AA	AG	AA	GG	TT	GG	TT	CT	-		
	E2152b	GG	AA	TA	CC	AG	CC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	E2174a	CC	CC	AA	AG	CT	GG	TT	GT	GT	AC	AG	AA	CT	CC	GG	CT	CC	AA	AA	GG	AG	TT	AT	NN	AG	NN	GG	TT	AG	CT	AA	AA		
	E2174a	CC	CC	AA	AG	CT	GG	TT	GT	GT	AC	AG	AA	CT	CC	GG	CT	CC	AA	AA	GG	AG	TT	AT	NN	AG	NN	GG	TT	AG	CT	AA			
16	E2174a	GG	GG	AA	TC	AG	-	GG	TT	AC	TG	-	CT	AA	TC	TT	CC	CT	TT	TT	-	AG	TT	AT	NN	TC	NN	CC	TT	AG	CT	TT	AA		
	E2174b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	E2174b	CC	CC	AA	AG	CT	GG	GG	TT	GT	GT	AC	AG	AA	CT	GG	CT	CC	AA	AA	GG	AG	TT	AT	NN	AG	GG	GG	TT	AG	CT	TT	AA		
	E2174b	GG	GG	AA	TC	AG	NN	GG	-	AC	-	-	TC	AA	TC	TT	CC	GG	TC	TT	-	-	TT	AT	AG	TC	-	CC	TT	AG	CT	TT	AA		
	E8802a	CC	CC	AT	AG	CT	CC	AA	TT	GT	TT	GT	CC	AA	CC	CT	CG	CT	GG	AG	CC	CT	AG	CT	AA	GG	AA	AA	AG	TT	AG	CT	AA		
17	E8802a	CC	CC	AT	AG	CT	GG	TT	GT	TT	AC	GG	AA	CC	CC	CG	CT	CC	AG	AA	AG	AG	CT	AA	NN	AA	AA	AA	AG	TT	AG	CT	AA		
	E8802a	CC	CC	AT	AG	CT	GG	TT	GT	TT	AC	GG	AA	CC	CC	CG	CT	CC	AG	AA	AG	AG	CT	AA	NN	AA	AA	AG	TT	AG	CT	AA			
	E8802a	CC	CC	AT	AG	CT	GG	TT	GT	TT	AC	GG	AA	CC	CC	CG	CT	CC	AG	AA	AG	AG	CT	AA	NN	AA	AA	AG	TT	AG	CT	AA			
	E8802a	GG	GG	TA	TC	AG	-	AA	TT	AC	TT	-	CC	AA	CC	TC	CC	CC	TC	TT	-	AG	CT	AA	AG	TT	AA	CT	TT	AG	CT	TT	AA		
	E8802b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
18	E8802b	CC	CC	AT	AG	CT	GG	TT	GT	TT	AC	GG	AA	CC	CC	CG	CT	CC	AG	AA	AG	AG	CT	AA	GG	AA	AG	AA	AG	TT	AG	CT	TT	AA	
	E8802b	GG	GG	TA	TC	AG	CC	AA	-	AC	TT	-	AA	CC	TC	-	CT	CC	TC	TT	-	-	CT	AA	-	TT	AA	CT	CT	TT	AG	CT	TT	AA	
	E8966a	CT	CT	TT	GG	CT	CG	AG	CC	TT	GG	CC	AG	CC	CC	CG	TT	CG	AG	NN	CC	AA	CC	AA	GG	AG	AG	AG	TT	AA	CT	AA	AG		
	E8966a	CT	CT	AA	GG	CT	CG	CC	TT	GG	CC	GG	AG	CC	CC	CC	CC	TT	CG	AG	AA	GG	AA	CC	AA	GG	AG	NN	AG	TT	AA	CT	AA		
	E8966a	NN	AG	AA	CC	AG	GC	-	CC	AA	GG	-	CC	AG	CC	-	TT	-	TC	TT	-	AA	CC	AA	GG	TC	-	CT	TT	AA	CT	TT	AG	AG	
19	E8966b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	E8966b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	E8966b	CT	CT	AA	GG	CT	CG	CC	TT	GG	CC	GG	AG	CC	CC	CG	TT	CG	AG	AA	GG	AA	CC	AA	GG	AG	NN	AG	TT	AA	CT	TT	AG		
	E8966b	AG	AG	AA	CC	AG	CC	GA	CC	AA	-	GG	CC	AG	CC	CC	GG	TT	CG	TC	TT	-	CC	AA	GG	TC	AG	NN	AG	CT	TT	-	CT	TT	AG
	E8989a	TT	CT	AA	AA	TT	CC	AG	TT	TT	GG	GT	TT	AA	CC	TT	CC	CT	CG	AA	AC	CC	AG	CC	TT	GG	AA	AG	AG	TT	AG	TT	AA	AG	
20	E8989a	TT	CT	TT	AA	TT	GG	TT	TT	GG	AC	AA	AA	CC	CC	GG	CT	CG	AA	AC	GG	AG	CC	TT	GG	AA	NN	AG	TT	AG	TT	AA	AG	AG	
	E8989a	TT	CT	TT	AA	TT	GG	TT	TT	GG	AC	AA	AA	CC	CC	GG	CT	CG	AA	AC	GG	AG	CC	TT	GG	AA	NN	AG	TT	AG	TT	AA	AG		
	E8989a	AA	AG	TT	TT	NN	CC	-	TT	AA	GG	-	TT	AA	CC	TT	NN	CT	CG	TT	TG	-	CC	TT	-	NN	TT	-	CT	TT	AG	TT	AG		
	E8989b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	E8989b	TT	CT	TT	AA	TT	GG	TT	TT	GG	AC	AA	AA	CC	CC	GG	CT	CG	AA	AC	GG	AG	CC	TT	GG	AA	NN	AG	TT	AG	TT	AG	TT	AG	
21	E9046a	CT	CC	AT	AG	TT	CG	AA	CT	GT	GG	TT	CC	AA	CT	TT	CC	AA	AA	AA	TT	AA	CT	AT	AG	AG	AG	AG	CT	AG	CT	AA	AA		
	E9046a	CT	CC	AT	AG	TT	CG	AA	CT	GT	GG	AA	GG	AA	CT	GG	TT	GG	AA	AA	AA	AA	CT	AT	NN	AG	NN	AG	CT	AG	CT	AA	AA		
	E9046a	CT	CC	AT	AG	TT	CG	AA	CT	GT	GG	AA	GG	AA	CT	GG	TT	GG	AA	AA	AA	AA	CT	AT	NN	AG	NN	AG	CT	AG	CT	AA	AA		
	E9046a	AG	GG	TA	TC	AA	GC	AA	TC	AC	GG	TT	CC	AA	TC	TT	CC	TT	GG	TT	TT	-	AA	CT	AT	AG	TC	AG	CT	TC	AG	CT	TT	AA	
	E9046b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
22	E9046b	CT	CC	AT	AG	TT	CG	TT	GT	GG	AA	GG	AA	CT	GG	TT	GG	AA	AA	AA	AA	AA	CT	AT	NN	AG	NN	AG	CT	AG	CT	TT	AA		
	E9046b	AG	GG	TA	TC	AA	GC	AA	-	AC	-	TT	CC	AA	TC	TT	CC	TT	GG	TT	TT	-	CT	AT	-	TC	AG	CT	TC	AG	CT	TT	AA		
	E9142a	CT	CC	TT	GG	CT	CG	AG	TT	TT	GT	CT	GG	CT	CT	GG	TT	GG	AG	AC	TT	GG	CC	TT	AA	AA	AA	AG	TT	GG	TT	AA	AG		
	E9142a	CT	CC	AA	GG	CT	CG	TT	TT	GT	AC	AG	GG	CT	CC	CC	CT	CC	AG	AA	AA	GG	CC	TT	AA	AA	AA	AG	TT	GG	TT	AA	AG		
	E9142a	AG	GG	AA	CC	AG	GC	-	TT	AA	TG	-	NN	CT	GG	TC	CC	CC	TC	TG	-	GG	CC	TT	AA	AA	AA	CT	TT	GG	TT	TT	AG		
23	E9142b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	E9142b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	E9142b	CT	CC	AA	GG	CT	CG	GA	TT	TT	GT	AC	AG	GG	CT	CC	TT	CC	AG	AC	AA	GG	CC	TT	AA	AA	AA	AG	TT	GG	TT	TT	AG		
	E9142b	AG	GG	AA	CC	AG	GC	GA	-	AA	-	-	TC	GG	TC	-	CC	CC	TC	TG	TT	GG	CC	TT	AA	TT	AA	CT	CT	TT	GG	TT	AG		
	E9142b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		

	SNP	A1	A2	A4	A5	A8	A9	A23	A11	A18	A16	A20	A32	A21	A45	A15	A38	A37	A34	A43	A30	A41	A35	A48	A39	A25	A50	A54	A27	A33	A49	A24	A42	A53		
21	E9175a	NN	CC	TT	NN	NN	NN	AG	CT	NN	GG	GT	CT	AA	TT	CT	CG	CT	GG	AG	AA	CT	AG	CT	AT	AA	AG	AA	AG	CT	GG	CT	AA	AG		
	E9175a	CT	CC	AA	GG	CT	GG	CT	GG	AC	AG	AA	TT	TT	CG	CT	CC	AG	AA	AG	CT	AG	CT	AT	AA	AG	AA	AG	CT	GG	CT	AA	AG			
	E9175a	CT	CC	AA	GG	CT	GG	CT	GG	AC	AG	AA	TT	TT	CG	CT	CC	AG	AA	AG	CT	AG	CT	AT	AA	AG	AA	AG	CT	GG	CT	AA	AG			
	E9175a	AG	GG	AA	CC	AG	-	-	TC	AC	GG	-	CT	AA	TT	TC	NN	CT	CC	TC	TT	-	AG	CT	AT	AA	TC	AA	CC	TC	GG	CT	TT	AG		
	E9175b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
22	E9175b	CT	CC	AA	GG	CT	GG	CT	GT	GG	AC	AG	AA	TT	TC	GG	CT	CC	AG	AA	AG	AG	CT	AT	AA	AG	AA	AG	CT	GG	CT	AA	AG			
	E9175b	AG	GG	AA	CC	AG	NN	NN	AC	GG	-	-	AA	TT	TC	GG	CT	-	TC	TT	-	-	CT	AT	AA	TC	AA	CC	TC	GG	CT	AA	AG			
	E9201a	NN	CC	TT	NN	NN	NN	GG	CT	NN	GG	GT	CC	AA	TT	CT	CG	TT	GG	GG	AA	CT	GG	CC	AT	GG	GG	GG	AA	CT	AG	CT	AA	AG		
	E9201a	CC	CC	AA	AG	CC	CC	CT	TT	GG	AC	GG	AA	TT	CG	TT	CG	TT	CC	GG	AA	AG	GG	CC	AT	GG	GG	GG	AA	CT	AG	CT	AA	AG		
	E9201a	CC	CC	AA	AG	CC	CC	CT	TT	GG	AC	GG	AA	TT	CG	TT	CG	TT	CC	GG	AA	AG	GG	CC	AT	GG	GG	GG	AA	CT	AG	CT	AA	AG		
23	E9201a	GG	GG	AA	TC	GG	GG	-	TC	AA	GG	NN	CC	AA	TT	TC	GG	TT	CC	CC	TT	-	GG	CC	AT	GG	CC	GG	CC	GG	TT	TC	AG	CT	AA	
	E9201b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
	E9201b	CC	CC	AA	AG	CC	CC	CT	TT	GG	AC	GG	AA	TT	TC	GG	TT	CC	GG	AA	AG	GG	CC	AT	GG	GG	GG	GG	AA	CT	AG	CT	AA	AG		
	E9201b	GG	GG	AA	TC	GG	GG	NN	-	AA	NN	NN	CC	AA	TT	TC	-	TT	CC	-	TT	-	GG	CC	AT	GG	CC	GG	CC	GG	TT	TC	AG	CT	AA	
	2a	CT	CT	AA	AG	CC	CG	CT	GT	GT	GT	GT	CT	AA	CT	TT	GG	CT	CC	AA	AC	CC	AG	TT	AA	AA	AA	AA	AA	AA	TT	GG	CT	AA	AG	
24	2a	CT	CT	TT	AG	CC	CG	CT	GT	GT	AC	AG	AA	CT	-	CC	CT	GG	AA	AC	GG	AG	TT	AA	AA	AA	AA	AA	NN	AA	TT	GG	CT	AA	AG	
	2a	CT	CT	TT	AG	CC	CG	-	CT	GT	GT	AC	AG	AA	CT	-	CC	CT	GG	AA	AC	GG	AG	TT	AA	AA	AA	AA	NN	AA	TT	GG	CT	AA	AG	
	2a	AG	AG	TT	TC	GG	GG	CC	-	AC	TG	TG	TC	AA	TC	TT	NN	CT	-	TT	TG	-	AG	TT	AA	NN	TT	AG	TT	TT	GG	CT	CT	AG	AG	
	2b	CT	CT	AA	AG	CC	CG	AG	CT	GT	GT	GT	CT	AA	CT	TT	GG	CT	CC	AA	AC	CC	AG	TT	AA	AA	AA	AA	AA	TT	GG	CT	AG	AG	AG	
	2b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
25	2b	NN	AG	TT	TC	GG	GC	GA	-	AC	TG	-	TC	AA	TC	TT	-	CT	-	NN	TT	TG	-	TT	AA	AA	TT	AG	NN	TT	GG	CT	CT	AG	AG	
	E9225a	CC	CT	AT	AA	CT	CC	AG	TT	GT	GG	GG	CT	AG	CT	CT	CG	CT	GG	CC	CC	AG	CT	TT	AG	AG	AA	AA	GG	TT	GG	CT	AG	AG	AG	
	E9225a	CC	CT	AT	AA	CT	GG	TT	GT	GG	CC	AG	AG	CT	CT	GG	CT	CC	GG	CC	GG	AG	CT	TT	NN	AG	AA	AA	GG	TT	GG	TT	AG	AG		
	E9225a	GG	AG	TA	TT	AG	CC	-	TT	AC	GG	-	CT	AG	TC	TC	-	CT	CC	CC	GG	-	AG	CT	TT	AG	TC	AA	CC	CT	GG	TT	NN	NN	AG	
	E9225b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
26	E9225b	CC	CT	AT	AA	CT	GG	-	TT	GT	GG	CC	AG	AG	CT	TC	GG	CT	CC	GG	CC	GG	AG	CT	TT	NN	AG	AA	GG	TT	GG	TT	AG	AG	AG	
	E9225b	GG	AG	TA	TT	AG	-	GA	-	AC	-	TC	-	TC	TC	-	CT	-	CC	CC	GG	-	AG	CT	TT	AG	TC	AA	CC	CT	GG	TT	NN	AG	AG	
	L1a	CT	TT	TT	GG	CT	CC	GG	CT	TT	GG	GT	CT	AA	CT	CT	CG	CT	CG	AA	CC	CT	AG	CT	TT	AG	GG	AA	GG	CT	AG	CT	AG	AA	AG	
	L1a	CT	TT	AA	GG	CT	GG	CT	CT	TT	GG	AC	AG	AA	CT	CT	CG	CT	CG	-	CC	AG	CT	TT	NN	GG	AA	AA	GG	CT	AG	CT	AG	AA	AG	
	L1a	AG	AA	AA	CC	AG	-	CT	AA	GG	-	CT	AA	TC	TC	-	CT	CG	TT	GG	-	NN	CT	TT	AG	CC	AA	CC	CC	TC	AG	CT	CT	AA	AG	
27	L1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	L1b	CT	TT	AA	GG	CT	GG	CT	TT	GG	AC	AG	AA	CT	TC	GG	CT	CG	AA	CC	AG	AG	CT	TT	NN	GG	AA	AA	GG	CT	AG	CT	CT	AA	AG	
	L1b	AG	AA	AA	CC	AG	CC	NN	-	AA	-	TC	AA	TC	TC	-	CT	GG	TT	GG	-	AG	CT	TT	-	CC	AA	CC	CC	TC	AG	CT	CT	AA	AG	
	L2a	CT	CC	AA	AA	TT	CG	CT	GT	GG	GT	CT	AG	CC	TT	GG	CT	CC	CT	CG	AG	CT	TT	AA	GG	AA	AA	AG	CT	GG	TT	AG	AA	AG		
	L2a	CT	CC	AA	AA	TT	CG	CT	GT	GG	AC	AG	AG	CC	CG	GG	CT	CG	-	AG	AA	GG	CT	TT	AA	GG	AA	AA	AG	CT	GG	TT	AG	AA	AG	
28	L2a	CT	CC	AA	AA	TT	CG	-	CT	GT	GG	AC	AG	AG	CC	CG	GG	CT	CG	-	AG	AA	GG	CT	TT	AA	GG	AA	AA	AG	CT	GG	TT	AG	AA	AG
	L2a	NN	GG	AA	TT	AA	GC	-	CT	AC	GG	TG	CT	AG	CC	TT	CC	CT	TC	TG	-	-	CT	TT	AA	GG	AA	AG	CT	GG	TT	CT	CT	AA	AG	
	L2b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	L2b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	L2b	CT	CC	AA	AA	TT	CG	CT	GT	GG	AC	AG	AG	CC	CG	GG	CT	GG	CT	AG	AC	AA	GG	CT	TT	AA	GG	AA	AA	AG	CT	GG	TT	AG	AA	AG
29	L2b	AG	GG	AA	TT	AA	GC	GA	-	AC	GG	TG	-	AG	CC	TT	CC	CT	GG	AG	CT	NN	CT	TT	AA	CC	AA	CC	CC	CT	GG	TT	CT	AA	AG	
	L3a	CT	CT	AA	GG	CC	CG	AG	CT	TT	GG	GT	CC	AA	CC	CT	CG	CT	GG	CC	TT	AG	CT	AT	GG	AA	AG	GG	TT	GG	CC	AA	AG	AG	AG	
	L3a	CT	CT	TT	GG	CC	CG	CT	TT	GG	AC	GG	AA	CC	CG	GG	CT	CC	-	CC	AA	AG	CT	AT	GG	AA	NN	NN	GG	TT	GG	CC	AA	AG	AG	
	L3a	CT	CT	TT	GG	CC	CG	CT	TT	GG	AC	GG	AA	CC	CG	GG	CT	CG	CC	GG	CC	AA	AG	CT	AT	GG	AA	NN	NN	GG	TT	GG	CC	AA	AG	
	L3a	AG	AG	TT	CC	GG	GC	-	-	AA	GG	TG	CC	AA	CC	TC	CG	CT	CC	CC	GG	-	AG	CT	AT	GG	TT	AG	CC	CT	GG	CC	TT	GG	AG	AG
30	L3b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	L3b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	L3b	CT	CT	TT	GG	CC	CG	CT	TT	GG	AC	GG	AA	CC	CG	TC	CG	CT	CC	GG	CC	AA	AG	CT	AT	GG	AA	NN	NN	GG	TT	GG	CC	AA	AG	
	L3b	AG	AG	TT	CC	GG	GC	NN	-	AA	GG	TG	CC	AA	CC	TC	CG	CT	CC	CC	GG	-	CT	AT	GG	TT	AG	CC	CT	GG	CC	-	GG	AG	AG	
	L3b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

	SNP	A1	A2	A4	A5	A8	A9	A23	A11	A18	A16	A20	A32	A21	A45	A15	A38	A37	A34	A43	A30	A41	A35	A48	A39	A25	A50	A54	A27	A33	A49	A24	A42	A53	
28	L4a	CT	TT	AT	GG	CT	CG	AA	TT	GT	GT	GG	CC	AA	CC	TT	CC	CT	GG	AG	AC	CC	AG	CC	AT	AA	GG	GG	CT	AA	CT	AG	AA		
	L4a	CT	TT	AT	GG	CT	CG	AA	TT	GT	GT	CC	GG	AA	CC	GG	CT	CC	-	AC	GG	AG	CC	AT	AA	GG	NN	GG	CT	AA	CT	AA			
	L4a	CT	TT	AT	GG	CT	CG	AA	TT	GT	GT	CC	GG	AA	CC	GG	CT	CC	AG	AC	AG	GG	AG	CC	AT	AA	GG	NN	GG	CT	AA	CT	AA		
	L4a	AG	AA	TA	CC	AG	GC	AA	TT	AC	TG	-	CC	AA	CC	TT	CC	CT	CC	TC	TG	-	AG	CC	AT	AA	CC	NN	CC	TC	AA	CT	AA		
	L4b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	L4b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
29	L4b	CT	TT	AT	GG	CT	CG	AA	TT	GT	GT	CC	GG	AA	CC	GG	CT	CC	AG	AC	AG	GG	AG	CC	AT	AA	GG	NN	GG	CT	AA	CT	NN	AA	
	L4b	AG	AA	TA	CC	NN	GC	AA	-	AC	-	-	-	AA	-	TT	CC	-	-	-	-	AG	-	AT	-	NN	AG	CC	NN	AA	CT	AG	AG		
	L6a	TT	CC	AA	AG	CT	CC	AG	CT	GT	TT	GG	CT	AA	CT	TT	CG	CT	GG	AG	AC	CT	GG	CT	TT	AA	GG	AG	GG	CT	GG	CT	AG		
	L6a	TT	CC	AA	AG	CT	GG	CT	GT	TT	CC	AG	AA	CT	TT	CG	GT	CC	-	AC	AG	GG	CT	TT	AA	GG	NN	GG	CT	GG	CT	AG	AG		
	L6a	TT	CC	AA	AG	CT	GG	-	TC	GT	TT	CC	AG	AA	CT	TT	CG	CT	CC	AG	AC	AG	GG	CT	TT	AA	GG	NN	GG	TT	GG	CT	NN		
	L6b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
30	L6b	TT	CC	AA	AG	CT	GG	CT	GT	TT	CC	AG	AA	CT	TT	GG	CT	CC	CC	AG	AC	AG	GG	CT	TT	AA	GG	NN	GG	TT	GG	CT	-		
	L6b	AA	GG	AA	TC	AG	NN	NN	-	AC	NN	GG	-	AA	TC	TT	GG	CT	CC	TG	-	GG	CT	TT	AA	CC	NN	CC	AG	CC	TT	GG	CT	AG	
	L8a	CC	CC	AT	AA	CT	CG	AG	TT	GG	GG	GT	TT	GG	CT	TT	CG	TT	CG	AA	CC	CC	GG	TT	TT	AA	AG	AA	AA	CT	GG	CC	AG	AG	
	L8a	CC	CC	AT	AA	CT	CG	AG	TT	GG	GG	AC	AA	GG	CT	TT	CG	TT	CG	AA	AA	GG	GG	TT	TT	AA	AG	AA	AA	CT	GG	CC	AG	AG	
	L8a	GG	GG	TA	TT	AG	GC	-	TT	CC	GG	-	TT	GG	TC	TT	GG	TT	CG	TT	TT	-	GG	TT	TT	AA	TC	AA	-	TC	GG	CC	CT	AG	
	L8b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
31	L8b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	L8b	CC	CC	AT	AA	CT	CG	AG	TT	GG	GG	AC	AA	GG	CT	TT	GG	TT	CG	AA	AA	GG	GG	CT	TT	AA	AG	AA	AA	CT	GG	CC	AG	AG	
	L8b	GG	GG	TA	TT	AG	GC	GA	-	CC	GG	-	TT	GG	TC	TT	GG	-	CG	TT	TT	-	GG	TT	TT	AA	TC	NN	TT	TC	GG	CC	AG	AG	
	L11a	CC	CC	TT	GG	CT	CG	AG	TT	GG	TT	GT	CC	AA	CT	TT	CG	CT	GG	GG	AC	CT	GG	CT	TT	AG	AA	AG	AA	CT	AA	TT	AA	AA	
	L11a	CC	CC	AA	GG	CT	CG	AG	TT	GG	TT	AC	GG	AA	CT	TT	CG	CT	CC	-	AC	AG	GG	CT	TT	NN	AA	NN	AG	CT	AA	TT	AA	AA	
	L11a	CC	CC	AA	GG	CT	CG	AG	TT	GG	TT	AC	GG	AA	CT	TT	GG	CT	CC	GG	AC	AG	GG	CT	TT	NN	AA	NN	AG	CT	AA	TT	AA	AA	
32	L11a	GG	GG	AA	CC	AG	-	-	TT	CC	TT	-	CC	AA	TC	TT	-	CT	CC	CC	TG	-	GG	CT	TT	AG	TT	AG	CT	TC	AA	TT	TT	AA	AA
	L11b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	L11b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	L11b	CC	CC	AA	GG	CT	CG	AG	TT	GG	TT	AC	GG	AA	CT	TT	GG	CT	CC	GG	AC	AG	GG	CT	TT	NN	AA	NN	AG	CT	AA	TT	AA	AA	
	L11b	GG	GG	AA	CC	AG	-	GA	-	CC	TT	-	CC	AA	TC	TT	-	CT	CC	CC	TG	-	GG	CT	TT	AG	TT	AG	CT	TC	AA	TT	TT	AA	
	L13a	CT	CT	TT	AA	TT	CG	AA	TT	GT	GG	GT	CC	GG	CT	TT	GG	CT	GG	AA	AA	CT	GG	CC	AA	AA	AA	AA	AA	GG	TT	AA	TT	AA	
33	L13a	CT	CT	AA	AA	TT	CG	TT	GT	GG	AC	GG	GG	CT	TT	CC	CT	CC	-	AA	AG	CT	GG	CC	NN	AA	AA	AA	AA	GG	TT	AA	TT	AA	
	L13a	CT	CT	AA	AA	TT	CG	TT	GT	GG	AC	GG	GG	CT	TT	CC	CT	CC	-	AA	AG	GG	CC	NN	AA	AA	AA	AA	AA	GG	TT	AA	TT	AA	
	L13a	CT	CT	AA	AA	TT	CG	TT	GT	GG	AC	GG	GG	CT	TT	CC	CT	CC	AA	AG	AG	GG	CC	AA	AA	AA	AA	AA	AA	GG	TT	AA	TT	AA	
	L13a	AG	AG	AA	TT	NN	GC	AA	TT	AC	GG	-	CC	GG	TC	TT	GG	CT	CC	TT	TT	-	GG	CC	AA	AA	AA	AA	CC	TT	AA	TT	TT	AA	
	L13b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	L13b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
34	L13b	CT	CT	AA	AA	TT	CG	AG	TT	GT	GG	AC	GG	GG	CT	NN	GG	CT	CC	AA	AA	AG	GG	CC	AA	AA	AA	AA	GG	TT	AA	TT	-	AA	
	L13b	AG	AG	AA	TT	AA	GC	NN	-	AC	GG	-	CC	GG	TC	NN	GG	CT	CC	NN	TT	-	GG	CC	AA	AA	AA	TT	CC	TT	-	TT	-	AA	
	L14a	CT	CC	AA	AG	TT	CC	CC	TT	GG	CC	GG	AA	CC	CC	TT	CC	CC	AG	AC	AG	GG	CC	TT	NN	AG	AA	AA	AA	TT	GG	CT	AG	AG	
	L14a	CT	CC	AA	AG	TT	CC	CC	TT	GG	CC	GG	AA	CC	CC	TT	CC	CC	AG	AC	AG	GG	CC	TT	NN	AG	AA	AA	AA	TT	GG	CT	AG	AG	
	L14a	NN	GG	AA	TC	NN	GG	-	CC	AA	GG	GG	CC	AA	CC	TT	-	TT	CC	-	TG	NN	-	CC	TT	AG	TC	-	TT	TT	GG	CT	CT	AG	
	L14b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
35	L14b	CT	CC	AA	AG	TT	CC(*)	CC	TT	GG	CC	GG	AA	CC	CC	TT	GG	CT	CC	AG	AC	AG	GG	CC	TT	NN	AG	AA	AA	TT	GG	CT	CT	AG	
	L14b	AG	GG	AA	TC	AA	GG	GA	CC	AA	GG	GG	CC	AA	CC	TT	GG	TT	CC	-	TG	-	NN	CC	TT	AG	TC	AA	TT	TT	GG	CT	AG	AG	
	L17a	TT	CT	AT	GG	CC	CG	GG	TT	GT	GG	GG	CC	AA	CT	TT	GG	CC	GG	AG	AA	CC	AA	CC	AT	AG	GG	AG	CT	AG	TT	AG	AG	AG	
	L17a	TT	CT	AT	GG	CC	CG	GG	TT	GT	GG	CC	GG	AA	CT	TT	GG	CC	CC	-	AA	GG	AA	CC	AT	NN	AG	GG	AG	CT	AG	TT	AG	AG	
	L17a	TT	CT	AT	GG	CC	CG	GG	TT	GT	GG	CC	GG	AA	CT	TT	GG	CC	CC	AG	AA	GG	AA	CC	AT	NN	AG	GG	AG	CT	AG	TT	AG	AG	
	L17a	AA	AG	TA	CC	GG	GC	GG	TT	AC	GG	GG	CC	AA	TC	TT	GG	CC	NN	TC	TT	-	AA	CC	AT	AG	TC	GG	CT	TC	AG	TT	CT	AG	
36	L17b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	L17b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	L17b	TT	CT	AT	GG	CC	CG	GG	TT	GT	GG	CC	GG	AA	CT	TT	GG	CC	CC	AG	AA	GG	AA	CC	AT	NN	AG	GG	AG	CT	AG	TT	CT	AG	
	L17b	AA	AG	TA	CC	GG	GC	GG	TT	AC	GG	GG	CC	AA	TC	TT	GG	CC	NN	TC	TT	-	AA	CC	AT	NN	TC	GG	CT	TC	AG	TT	CT	AG	

	SNP	A1	A2	A4	A5	A8	A9	A23	A11	A18	A16	A20	A32	A21	A45	A15	A38	A37	A34	A43	A30	A41	A35	A48	A39	A25	A50	A54	A27	A33	A49	A24	A42	A53	
35	L22a	CC	CC	AA	AG	CT	GG	AG	TT	GG	GG	GT	TT	GG	CT	TT	GG	CT	GG	AA	AA	CT	GG	CT	TT	AA	AG	AA	AA	CT	AG	CT	AG		
	L22a	CC	CC	CT	AG	CT	CC	TT	GG	GG	AC	AA	GG	CT	CC	GG	CT	CC	-	AA	AG	GG	CT	TT	AA	AG	AA	AA	CT	AG	CT	AG			
	L22a	CC	CC	TT	AG	CT	CC	TT	GG	GG	AC	AA	GG	CT	CC	GG	CT	CC	AA	AA	AG	GG	CT	TT	AA	AG	AA	AA	CT	AG	CT	AG			
	L22a	GG	GG	TT	TC	AG	GG	-	TT	GG	GG	-	TT	GG	TC	TT	GG	CT	CC	TT	TT	-	GG	CT	TT	AA	TC	AA	-	TC	AG	CT	AG		
	L22b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	L22b	CC	CC	TT	AG	CT	CC	TT	GG	GG	AC	AA	GG	CT	AA	GG	CT	TT	CC	AA	AA	AG	GG	CT	TT	AA	AG	AA	AA	CT	AG	CT	AG		
36	L22b	NN	GG	TT	TC	AG	GG	NN	-	CC	GG	-	NN	GG	TC	TT	GG	CT	CC	TT	TT	-	GG	CT	TT	AA	TC	AA	NN	TC	AG	CT	AG		
	L23a	CT	TT	AT	AG	CT	CG	-	CT	GT	TT	GT	CT	GG	AA	TT	CC	CT	CG	AG	AC	TT	GG	CT	AA	AG	AG	AA	AA	TT	AG	TT	AA		
	L23a	CT	TT	AT	AG	CT	CG	CT	GT	TT	AC	GG	AA	TT	GG	CT	GG	CT	CG	-	AC	AA	GG	CT	AA	NN	AG	AA	AA	TT	AG	TT	AG		
	L23a	CT	TT	AT	AG	CT	CG	CT	GT	TT	AC	GG	AA	TT	GG	CT	GG	CT	CG	AG	AC	AA	GG	CT	AA	NN	AG	AA	AA	TT	AG	TT	AG		
	L23a	AG	AA	TA	TC	AG	GC	-	CT	AC	TT	-	CC	AA	TT	TT	GG	CT	CG	TC	TG	-	GG	CT	AA	AG	TC	AA	TT	TT	AG	TT	AG		
	L23b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
37	L23b	CT	TT	AT	AG	CT	CG	GA	CT	GT	TT	AC	GG	AA	TT	GG	CT	GG	CT	CG	AG	AC	AA	GG	CT	AA	GG	AG	AA	AA	TT	AG	TT	AG	
	L23b	AG	NN	TA	TC	AG	GC	-	AC	-	TG	-	CC	AA	TT	TT	CC	CT	-	TC	TG	-	GG	CT	AA	-	TC	AG	AA	NN	TT	AG	TT	AG	
	L30a	TT	CT	TT	AG	CT	CC	AG	TT	GT	GG	TT	TT	AA	CT	CT	GG	CT	GG	AG	AA	TT	AA	CT	AT	AG	AG	AA	AA	TT	AG	TT	AA		
	L30a	TT	CT	AA	AG	CT	GG	TT	GT	GG	AA	AA	AA	CT	CC	CT	CG	CT	CC	-	AA	AA	CT	AT	NN	AG	NN	AG	NN	AG	TT	AG	TT	AG	
	L30a	AA	AG	AA	TC	AG	CC	-	TT	AC	GG	TT	TT	AA	TC	TC	CG	CT	CC	AG	AA	AA	CT	AT	NN	AG	NN	AG	NN	AG	TT	AG	TT	AG	
	L30b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
38	L30b	TT	CT	AA	AG	CT	GG	AA	TT	GT	GG	AA	AA	AA	CT	TC	GG	CT	CC	AG	AA	AA	AA	CT	AT	NN	AG	NN	AG	TT	AG	TT	AG	TT	AG
	L30b	AA	AG	AA	TC	AG	CC	GA	-	AC	GG	-	TC	-	TC	GG	CT	CC	NN	TT	-	AA	CT	AT	AG	TC	AG	TC	AG	CT	TT	AG	TT	-	AG
	10a	CC	CT	AT	AG	CC	CG	AG	TT	GT	GT	GT	CT	AA	CT	CT	GG	TT	CC	AG	CC	CT	AG	TT	AT	AA	AA	AA	AA	TT	GG	CT	AG	AG	
	10a	CC	CT	AT	AG	CC	CG	AG	TT	GT	GT	AC	AG	AA	CT	-	CC	TT	GG	AG	CC	AG	AG	TT	AT	-	AA	AA	AA	TT	GG	CT	AG	AG	
	10a	NN	AG	TA	TC	GG	GC	-	TT	AC	-	GT	TC	AA	TC	TC	GG	TT	GG	TC	GG	-	AG	TT	AT	AA	TT	AA	-	TT	GG	CT	AG	AG	
	10b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
39	10b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	10b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	10b	NN	AG	TA	TC	GG	GC	NN	-	AC	TG	-	TC	AA	TC	TC	NN	TT	GG	-	GG	-	TT	AT	AA	TT	AA	-	TT	GG	CT	AG	CT	-	AG
	L31a	CT	CC	AT	AA	TT	CG	AG	TT	GG	GT	TT	CT	AG	CC	TT	CG	CT	CG	AG	AC	CT	AG	CT	AA	AG	AG	AG	AG	CT	AG	CT	AA	AA	
	L31a	CT	CC	AT	AA	TT	CG	AG	TT	GG	GT	AA	AG	CC	CC	CT	CG	CT	CG	-	AC	AG	CT	AA	AG	AA	AG	NN	AG	NN	AG	CT	AG	CT	AA
	L31a	AG	GG	TA	TT	-	NN	-	-	TT	CC	TG	TT	TC	AG	CC	TT	-	CT	-	TC	TG	-	AG	CT	AA	AA	TC	AG	CT	TT	AG	CT	TT	AA
40	L31b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	L31b	CT	CC	AT	AA	TT	CG	AG	TT	GG	GT	AA	AG	CC	CC	GG	CT	CG	AG	AC	AG	AG	CT	AA	AG	AA	AG	NN	AG	CT	AG	CT	AA	AA	
	L31b	AG	GG	TA	TT	AA	-	NN	-	CC	TG	TT	-	AG	CC	-	CT	-	-	TG	-	AG	CT	AA	AG	AA	TC	AG	CT	TT	AG	CT	-	AA	
	L38a	CT	TT	AT	AG	CT	CG	GG	CT	GT	GT	TT	CC	AA	CT	CC	GG	CT	CG	GG	CC	CT	TT	TT	AG	AG	AG	GG	CT	AA	TT	AA	AA		
	L38a	CT	TT	AT	AG	CT	CG	GG	CT	GT	GT	AA	GG	AA	CT	CC	CT	CG	-	CC	AG	AG	TT	TT	NN	AG	NN	NN	GG	-	AA	TT	AA	AA	
	L38a	AG	AA	TA	TC	AG	GC	-	-	AC	TG	TT	CC	AA	TC	CC	GG	CT	CG	GG	CC	AG	AG	TT	TT	NN	AG	NN	GG	-	AA	TT	TT	AA	
41	L38b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	L38b	CT	TT	AT	AG	CT	CG	GG	CT	GT	GT	AA	GG	AA	CT	CC	GG	CT	CG	CC	AG	AG	TT	TT	NN	AG	NN	NN	GG	-	AA	TT	-	AA	
	L38b	AG	AA	TA	TC	AG	GC	NN	-	AC	TG	-	-	TC	CC	GG	CT	CG	CC	GG	-	AG	TT	TT	GG	TC	AG	TC	AG	NN	AA	TT	-	AA	
	L44a	CT	CC	AT	AG	CT	CG	GG	TT	GG	GG	GG	CT	GG	CC	CT	CC	TT	GG	AG	AA	CT	GG	CT	TT	AA	GG	GG	AG	CT	AA	TT	AA	GG	
	L44a	CT	CC	AT	AG	CT	CG	GG	TT	GG	GG	CC	AG	GG	CC	CT	CC	TT	CC	-	AA	AG	CT	TT	AA	GG	GG	GG	AG	CT	AA	TT	AA	GG	
	L44a	AG	GG	TA	TC	AG	GC	-	TT	CC	GG	GG	TC	GG	CC	TC	CC	TT	CC	AG	AA	AG	GG	CT	TT	AA	GG	GG	AG	CT	AA	TT	TT	GG	
42	L44b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	L44b	CT	CC	AT	AG	CT	CG	GG	TT	GG	GG	CC	AG	GG	CC	CC	GG	TT	CC	AG	AA	AG	GG	CT	TT	AA	GG	GG	AG	CT	AA	TT	-	NN	
	L44b	NN	-	TA	TC	AG	NN	GG	-	CC	-	-	TC	NN	CC	TC	CC	TT	CC	TC	TT	-	-	CT	TT	AA	CC	GG	CT	TC	AA	TT	-	NN	
	L44b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

SNP	A1	A2	A4	A5	A8	A9	A23	A11	A18	A16	A20	A32	A21	A45	A15	A38	A37	A34	A43	A30	A41	A35	A48	A39	A25	A50	A54	A27	A33	A49	A24	A42	A53	
42	L45a	TT	CT	AT	AG	TT	CC	GG	CC	TT	GT	GG	CC	GG	CC	CT	GG	CT	GG	AA	AC	TT	AG	TT	AT	AG	AG	AG	CT	AA	TT	AG	AG	
	L45a	TT	CT	AT	AG	TT	GG	CC	TT	GT	CC	GG	GG	CC	CC	CT	CC	CC	CC	-	AC	AA	AG	TT	AT	NN	AG	NN	AG	CT	AA	TT	AG	
	L45a	TT	CT	AT	AG	TT	GG	CC	TT	GT	CC	GG	GG	CC	CC	CT	CC	CC	AA	AC	AA	AG	TT	AT	NN	AG	NN	AG	CT	AA	TT	AG		
	L45a	AA	AG	TA	TC	AA	CC	GG	CC	AA	NN	-	NN	GG	CC	TC	-	CC	TT	TG	-	AG	TT	AT	AG	TC	-	CT	TC	AA	TT	CT	AG	
L45b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
L45b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
L45b	TT	CT	AT	AG	TT	GG	CC	GG	CC	TT	GT	CC	GG	CC	CC	CC	CT	CC	AA	AC	AA	AG	TT	AT	NN	AG	NN	AG	CT	AA	TT	CT	AG	
L45b	AA	AG	TA	TC	AA	CC	GG	CC	AA	TG	GG	NN	GG	CC	TC	GG	CT	CC	TT	TG	-	AG	TT	AT	AG	TC	AG	CT	TC	AA	TT	CT	AG	
43	15a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	15a	CT	CT	TT	AG	CT	GG	CC	GG	GG	CC	AG	AA	CT	CC	GG	TT	GG	AG	AC	GG	AA	TT	AA	AA	AA	AA	AA	AA	TT	GG	CT	AG	
	15a	CT	CT	TT	AG	CT	GG	CC	GG	GG	CC	AG	AA	CT	CC	GG	TT	GG	AG	AC	GG	AA	TT	AA	AA	AA	AA	AA	AA	TT	GG	CT	AG	
	15a	NN	AG	TT	TC	AG	-	NN	CC	GG	NN	CC	AA	TC	TC	CC	GG	TT	GG	TC	TG	-	AA	TT	AA	AA	AA	TT	-	TT	GG	CT	CT	AG
	15c	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	15c	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
15c	CT	CT	TT	AG	CT	GG	CC	GG	GG	CC	AG	AA	CT	CC	GG	TT	GG	AG	AC	GG	AA	TT	AA	AA	AA	AA	AA	AA	TT	GG	CT	CT	AG	
15b	NN	AG	TT	TC	AG	-	GA	-	CC	GG	-	TC	AA	TC	CC	TT	GG	TC	TG	-	AA	TT	AA	AA	AA	AA	TT	-	TT	GG	CT	CT	AG	

III. Abkürzungsverzeichnis

%	Prozent	f	<i>forward</i> (bezüglich der Orientierung einer DNA-Sequenz)
.tiff	<i>tagged image file format</i>	f	Frequenz
.txt	Dateinamenserweiterung reiner Textdateien	F	<i>Fixation index</i>
©	<i>Copyright</i>	FP-TDI	<i>fluorescence polarization template-directed dye incorporation</i>
®	Eingetragene Marke	FRET	Fluoreszenz-Resonanz-Energietransfer
°C	Grad Celsius	g	Gramm
µg	Mikrogramm	G	Guanin
µl	Mikroliter	ggf.	gegebenenfalls
µM	Mikromolar	h	Stunde(n)
µm	Mikrometer	H ₂ O	Wasser: chemische Verbindung aus den Elementen Sauerstoff (O) und Wasserstoff (H)
A	Adenin	HC	<i>hybridisation control</i> = Hybridisierungskontrolle
A.dest	<i>Aqua bidestillata</i>	H _{exp/obs}	erwarteter/beobachteter Heterozygotengrad
AB0	AB0-System der Blutgruppen	HLA	<i>human leucocyte antigene</i>
Abb.	Abbildung	HPLC	<i>high pressure liquid chromatography</i> = Hochdruck-Flüssigkeitschromatographie
AG	Aktiengesellschaft	HV1/HV2	<i>hypervariable region</i> 1 und 2
AIM	<i>ancestry-informative markers</i>	HWG	Hardy-Weinberg-Gleichgewicht
Alu	ursprünglich aus <i>Arthrobacter luteus</i> isolierte Endonuklease	Inc.	<i>incorporation</i>
AMOVA	<i>Analysis of Molecular Variance</i>	ISO	<i>International Organization for Standardization</i>
APEX	<i>arrayed primer extension</i>	IUB	<i>International Union of Biochemistry</i>
ASBER	<i>allele-specific base extention reaction</i>	K	<i>cluster (Structure-Software)</i>
ASO	<i>Allele-Specific Oligonucleotide Probes</i> = allelspezifische Sonden	kb	Kilobasenpaare = 1000 bp
ATP	Adenosintriphosphat	KFB	Kompetenzzentrum für Fluoreszente Bioanalytik
aut.	autosomal	LD	<i>linkage disequilibrium</i>
bp	<i>base pair(s)</i> = Basenpaar(e)	LINE	<i>long interspersed nuclear element</i>
BSA	<i>bovine serum albumin</i>	ln	natürlicher Logarithmus
bzw.	beziehungsweise	log	Logarithmus
C	Cytosin	LR	<i>likelihood ratio</i>
ca.	circa	LTR	<i>long terminal (direct) repeat</i>
CCD	<i>Charge-coupled Device</i>	M	Molar
cDNA	<i>copy DNA</i> = DNA-Kopie	MALDI-TOF	<i>Matrix-Assisted Laser Desorption/Ionisation – Time of Flight</i>
cEXP	kombinierte Ausschlusswahrscheinlichkeit	max.	Maximum
cm	Zentimeter	mg	Milligramm
cov	Kovarianz	MgCl ₂	Magnesiumchlorid
ctag	<i>complementary tag</i>	Min.	Minute(n)
Cy3	<i>Cyanine 3</i>	min.	Minimum
Cy5	<i>Cyanine 5</i>	ml	Milliliter
d.h.	das heißt	mM	Millimolar
Da	Dalton	mt	<i>mitochondrial</i>
db	<i>database</i>	mW	Milliwatt
ddATP	Didesoxyadenosintriphosphat	N	Normal
ddCTP	Didesoxycytosintriphosphat	n/N	Stichprobengröße
ddGTP	Didesoxyguanosintriphosphat	NaOH	Natronlauge
ddNTP	Dideoxynukleotidtriphosphat	NCBI	<i>National Center for Biotechnology Information</i>
ddUTP	Didesoxyuridintriphosphat	ng	Nanogramm
DI	Diskriminierungsindex	NH ₂	Aminogruppe
div.	diverse	nm	Nanometer
D-Loop	<i>displacement loop</i>	nt	<i>nucleotide</i>
DNA	deoxyribonucleic acid = Desoxyribonukleinsäure	o.g.	oben genannte
dNTP	Deoxynukleotidtriphosphat	OLA	<i>oligonucleotide ligation assay</i>
EDNAP	European DNA Profiling Group	Oligo(dT)	Oligodesoxythymidinspacer
EDTA	Ethylendiamintetraessigsäure	p	Wahrscheinlichkeitswert
ENFSI	<i>European Network of Forensic Science Institutes</i>	p	<i>petit (smaller chromosome arm)</i>
engl.	englisch		
et al.	<i>et alii</i>		
EtOH	Ethanol		
EU	Europäische Union		
Exol	Exonuklease I		
EXP	Ausschlusswahrscheinlichkeit		

PAR	pseudoautosomale Region	V	Volt
PBS	<i>phosphat buffered saline</i> = Phosphat-gepufferte Salzlösung	v.	Version
PCR	<i>Polymerase Chain Reaction</i> = Polymerasekettenreaktion	V/mm	Volt pro Millimeter
pg	Picogramm	ver.	Version
pH	<i>pondus Hydrogeni</i> = Protonenaktivitätsexponent	vgl.	vergleiche
PI	<i>paternity index</i>	VNTR	<i>Variable Number Tandem Repeats</i>
PM	<i>matching probability</i>	vs.	<i>versus</i>
pmol	Picomol	WGA	<i>whole genome amplification</i>
PMT	<i>photomultiplier tube</i>	x	mal
POP	<i>performance optimized polymer</i>	x g	Zentrifugalbeschleunigung
PPi	Pyrophosphat-Molekül	YCC	<i>Y-Chromosome Consortium</i>
q	<i>queue (longer chromosome arm - at the chromosome's tail)</i>	z.B.	zum Beispiel
r	<i>reverse</i> (bezüglich der Orientierung einer DNA-Sequenz)	α	Alpha
r	Korrelationskoeffizient	Δ	Delta
R110	<i>Rhodamine 110</i>	θ	Theta
R6G	<i>Rhodamine 6G</i>	λ	Lambda
RC	<i>reaction control</i> = Reaktionskontrolle	σ	Sigma
REF.	Referenzprobe	Σ	Sigma (Summenzeichen)
RFLP	<i>restriction fragment length polymorphism</i> = Restriktions-Fragment-Längen-Polymorphismus"	χ	Chi
RFU	<i>relative fluorescence unit</i>	Φ	Phi
ROX	<i>x-Rhodamine</i>		
rpm	<i>rounds per minute</i>		
RT	Raumtemperatur		
s	Sekunde(n)		
S _{0/1/2}	Singulett-Zustand		
s.a.	siehe auch		
SABER	<i>single-allele base extension reaction</i>		
SAP	<i>shrimp alkaline phosphatase</i>		
SBE	<i>Single Base Extension</i> = Einzelbasen-Verlängerung		
SD	standard deviation = Standardabweichung		
SDS	Sodiumdodecylsulfat		
Sek.	Sekunde		
SINE	<i>short interspersed nuclear element</i>		
SNBE	<i>single nucleotide base extension</i>		
SNP	<i>Single Nucleotide Polymorphism</i>		
SNPforID	<i>Single Nucleotide polymorphisms for the Forensic Identification of Persons</i>		
SNuPE	<i>single nucleotide primer extension</i>		
SSC	Standard Salz Citrat		
Std.	Stunde(n)		
STR	<i>Short Tandem Repeat</i>		
SWaP	<i>S/W amid Palindrome</i>		
T	Thymin		
TAE	Tris-Acetat-EDTA		
TAMRA	<i>carboxytetramethylrhodamine</i>		
TBE	Tris-Borat-EDTA-Puffer		
T _m	<i>melting temperature</i> = Schmelztemperatur		
Tris	Tris(Hydroxymethyl)aminomethan		
TSC	<i>The SNP consortium</i>		
™	<i>trade mark</i>		
U	<i>Unit</i> (Einheit Enzymaktivität)		
u.a.	unter anderem, unter anderen		
u.ä.	und ähnlichen		
U/min	Umdrehungen pro Minute		

IV. Material- und Herstellerverzeichnis

0-9

00112_RC_2_tag, metabion international AG, Martinsried, Deutschland
 100 bp Leiter (Längenstandard), Invitrogen, Karlsruhe
 2-Mercaptoethanol, Sigma-Aldrich, Steinheim, Deutschland
 2-Propanol, ROTISOLV® ≥99,9 %, UV/IR-Grade, Carl Roth, Karlsruhe, Deutschland

A

ABI PRISM® 310 DNA Genetic Analyzer, Applied Biosystems Inc., Foster City, CA, USA
 ABI PRISM® 3100 DNA Genetic Analyzer, Applied Biosystems Inc., Foster City, CA, USA
 ABI PRISM® 3130 DNA Genetic Analyzer, Applied Biosystems Inc., Foster City, CA, USA
 ABI PRISM® Big Dye™ Terminator Cycle Sequencing Kit, Applied Biosystems Inc., Foster City, CA, USA
 ABI PRISM® SNaPshot™ Multiplex Kit, Applied Biosystems Inc., Foster City, CA, USA
 Agar, Sigma-Aldrich, Steinheim, Deutschland
 High Resolution Agarose Gel Electrophoresis, Invitrogen, Karlsruhe, Deutschland
 Ammoniumacetat (CH₃COONH₄), Merck, Darmstadt, Deutschland
 Ammoniumsulfat (H₈N₂O₄S), ≥99,5 %, p.a., ACS, ISO Carl Roth, Karlsruhe, Deutschland
 AmpliTaq Gold® DNA Polymerase, 250 U, Applied Biosystems Inc., Foster City, CA, USA
 ARLEQUIN v. 3.11, Software for Population Genetics Data Analysis, (c) Laurent Excoffier, Institute of Ecology and Evolution, University of Bern

B

BioPhotometer® 6131, Eppendorf, Hamburg, Deutschland
 Blue Dextran D5751 aus Leuconostoc mesenteroides Sigma-Aldrich, Steinheim, Deutschland
 Bromphenolblau, Merck, Darmstadt, Deutschland
 Bovine Serum Albumin (BSA), Fermentas, St. Leon-Rot, Deutschland

C

ClusterA Software tool, The research group of molecular medicine, Uppsala Universitet
 CodeLink™ Activated Slides, GE Healthcare, Chalfont St. Giles, UK (formerly Amersham Biosciences)
 cProbe Sets, metabion international AG, Martinsried, Deutschland
 ctag Oligos, MWG Biotech AG (jetzt Eurofins MWG Operon), Ebersberg, Deutschland
 Cuvette, trUVView™, Bio-Rad Laboratories, München, Deutschland

D

Data Collection Software 3.1, Applied Biosystems Inc., Foster City, CA, USA
 Desoxynukleotid-Triphosphate (einzelne dNTPs), PEQLAB Biotechnologie, Erlangen, Deutschland
 Didesoxynukleotid-Triphosphate (einzelne fluoreszenzmarkierte ddNTPs): Perkin Elmar, Waltham, MA, USA
 DNAVIEW™ Software v. 27.17, software package for DNA identification, Charles Brenner, Forensic Mathematics, Oakland, USA

E

EDTA, (((HOOCCH₂)₂NCH₂)₂) ≥99 %, p.a., ACS (Ethyldiamintetraessigsäure), Carl Roth, Karlsruhe, Deutschland
 Elastosil® RT 625A und B, Wacker-Chemie, München, Deutschland
 Emissionsfilter, PhotoMed GmbH, Seefeld, Deutschland
 Ethanol ROTIPURAN® (C₂H₅OH) ≥99,8 %, p.a., Carl Roth, Karlsruhe, Deutschland
 Ethanol vergällt, ≥99,8 %, mit ca. 1 % MEK, Carl Roth, Karlsruhe, Deutschland
 Ethanolamin ROTIPURAN®, ≥99 %, p.a., ACS, Carl Roth, Karlsruhe, Deutschland
 Ethidiumbromid (EtBr), Sigma-Aldrich, Steinheim, Deutschland
 Exonuklease I (Exol) aus E.coli, GE Healthcare, Chalfont St. Giles, UK (formerly Amersham Biosciences)

F

Formaldehyd 37 % (HCHO), Carl Roth, Karlsruhe, Deutschland
 Formamid, ≥99,5 % (HCONH₂), Sigma-Aldrich, Steinheim, Deutschland
 FSTAT v. 2.9.3.2 Phylogenetic Analysis, Jerome Goudet, Department of Ecology & Evolution, Lausanne, Switzerland
 FTA®-Papier, Whatman International Ltd, Maidstone, Kent, UK

G

GDA Genetic Data Analysis software, Program by Paul O. Lewis and Dmitri Zaykin, Paul Lewis Phylogenetics, Department of Ecology and Evolutionary Biology at the University of Connecticut
 GeneMapper® ID 3.0 Software, Human Identification Analysis, Applied Biosystems Inc., Foster City, CA, USA
 GENEPOP v. 4.0, Population Genetics Software, Michel Raymond und Francois Rousset, Laboratoire de Genetique et Environment, Montpellier, France
 GeneScan® Analysis 3.7 Software, Applied Biosystems Inc., Foster City, CA, USA
 GeneScan™, 120 / 500 LIZ® Size Standard, Applied Biosystems Inc., Foster City, CA, USA
 GenomiPhi™ DNA Amplification Kit, GE Healthcare, Chalfont St. Giles, UK (formerly Amersham Biosciences)
 Genotyper 3.7 Software, Applied Biosystems Inc., Foster City, CA, USA
 GenPlex, Applied Biosystems Inc., Foster City, CA, USA

H

H33258 Hoechst, Bisbenzimid, AppliChem, Darmstadt, Deutschland
 Handschuhe, PFE Safeskin, Fisher Scientific GmbH, Schwerte, Deutschland
 Hi-Di™ Formamide, Applied Biosystems Inc., Foster City, CA, USA

HPLC-H₂O ROTISOLV® HPLC Gradient Grade, Carl Roth, Karlsruhe, Deutschland
 Hybridisierungskammer, Hersteller unbekannt, Schenkung Zytometrielabor, Universität Mainz
 Hybridisierungsöfen BE 500, Bachhofer, Reutlingen

I

IDT SciTools OligoAnalyzer 3.1, Integrated DNA Technologies, Coralville, IA, USA
 ImaGene™ 5.5 Standard Microarray Analysis Software, BioDiscovery Inc., El Segundo, CA, USA

Inkubations-Schüttler CERTOMAT®IS, B.Braun Biotech, Melsungen, Deutschland

Isoamylalkohol ROTIPURAN® (C₅H₁₂O), ≥98,5 %, p.a., ACS, Carl Roth, Karlsruhe, Deutschland

J

K

genomische DNA (gDNA), European Collection of Cell Cultures, Applied Biosystems Inc., Foster City, USA

L

Laserscanner Affymetrix® 428™ Array Laser-Scanner, Affymetrix, Santa Clara, CA, USA, jetzt Eurofins MWG Operon
Laserscanner GeneTAC™ LSIV Microarray Scanner, Genomic Solutions Inc., Ann Arbor, MI, USA

M

Magnesiumchlorid (MgCl₂) 25mM, Applied Biosystems Inc., Foster City, CA, USA

MassARRAY® iPLEX®, Sequenom, San Diego, CA, USA
mfold-Programm, Michael Zuker, Rensselaer Polytechnic Institute, Troy, NY, USA

Microsoft® Office Professional, Microsoft Corporation, USA

Microsoft® Windows 2000 / Windows XP, Microsoft Corporation, USA

N

N-hydroxysuccinimidester (NHS), Sigma-Aldrich, Steinheim, Deutschland

N-Laurylsacrosinat, Sigma-Aldrich, Steinheim, Deutschland

Natriumacetat, (CH₃COONa), 99,99 Suprapur®, Merck, Darmstadt, Deutschland

Natriumchlorid (NaCl), ≥99,5 %, p.a., ACS, Carl Roth, Karlsruhe, Deutschland

O

OmniGrid spotter, GeneMachines, San Carlos, CA, USA

P

PCR-Cooler Rack für 0,2 ml und 0,5 ml Reaktionsgefäße, Eppendorf, Hamburg, Deutschland

PCR-Reaktionspuffer Gold 10x, Applied Biosystems Inc., Foster City, CA, USA

PCR-Reaktionsgefäß, 0,5 ml, 1,5 ml, 2 ml Safe-Lock, Eppendorf, Hamburg, Deutschland

PCR-Reaktionsgefäß, 0,2 ml, (Domed Cap/Flat Cap), PEQLAB Biotechnologie, Erlangen, Deutschland

PEQLAB E.Z.N.A.® Blood DNA Kit II, PEQLAB Biotechnologie, Erlangen, Deutschland

PicoGreen® dsDNA Quantitation Kit, Molecular Probes, Eugene, Oregon, USA

Pipettenspitzen epT.I.P.S., 0,1-10 µl, 0,5-20 µl, 2-200 µl, 50-1000 µl, mit und ohne Filter, Eppendorf, Hamburg, Deutschland
POP-4™, POP-6™ Performance Optimized Polymers Applied Biosystems Inc., Foster City, CA, USA

PowerPlex® 16 System, Promega GmbH, Mannheim, Deutschland

PowerStats Software, Promega Corporation, Madison, WI, USA

Primer, biomers.net GmbH, Ulm, Deutschland

Q

QIAamp DNA Blood Mini Kit, Qiagen, Hilden, Deutschland

QIAquick PCR Purification Kit, Qiagen, Hilden, Deutschland

QPCR Tube Strips, 0,2 ml mit Einzeldeckel, PEQLAB

Biotechnologie, Erlangen, Deutschland

R

Reagenz Röhrchen 12 ml, steril, Greiner-Bio-One, Frickenhausen, Deutschland

S

Shrimp Alkaline Phosphatase (SAP) GE Healthcare, Chalfont St. Giles, UK (formerly Amersham Biosciences)

Sequencing Analysis Software 3.1 Applied Biosystems Inc., Foster City, CA, USA

Slides MWG Biotech AG (jetzt Eurofins MWG Operon), Ebersberg, Deutschland

SNP Assistant™ v. 1.0.9 Programm, BioData Ltd., Tartu, Estonia

SNPSnapper 3.88b software developed by: Juha Saharinen, National Public Health Institute, Helsinki, Finland

Sodiumdodecylsulfat (SDS) Carl Roth, Karlsruhe, Deutschland

Sorvall RC-5B Refrigerated Superspeed Centrifuge Du Ponts Instruments, Thermo Fisher Scientific

SPSS-Software Version 11.5 SPSS Inc., an IBM Company, Chicago, IL, USA

Sterilbank Gelaire BSB4A®. Gelaire Pty Ltd., Sydney, NSW, Australia

Structure Software, developed by Pritchard, Stephens & Donnelly, The Division of Biological Sciences, The University of Chicago

T

TEMED 99 %, p.a., für die Elektrophorese N,N,N',N'-Tetramethylethylendiamin, 1,2-Bis(dimethylamino)-ethan Carl Roth, Karlsruhe, Deutschland

Therminator™ DNA Polymerase, New England Biolabs Inc., Ipswich, MA, USA

Thermocycler PE 9600 und 2400 Perkin Elmer Cetus, Norwalk, USA

Thermo Sequenase™ GE Healthcare, Chalfont St. Giles, UK (formerly Amersham Biosciences)

tri-Natriumcitrat-Dihydrat, Merck, Darmstadt, Deutschland

Tris PUFFERAN® ≥99,3 %, Buffer Grade Tris-(hydroxymethyl)-aminomethan Carl Roth, Karlsruhe, Deutschland

Triton X-100 Sigma-Aldrich, Steinheim, Deutschland

U

V

VersaFluor™ Fluorometer, Bio-Rad Laboratories, München, Deutschland

W

WGA, siehe GenomiPhi™ DNA Amplification Kit

X

Y

Z

ERKLÄRUNG

(gemäß § 6 (2), h, i der Promotionsordnung vom 26. Juli 2000, zuletzt geändert am 15. August 2005)

Hiermit erkläre ich, Magdalena Bogus, dass ich die eingereichte Dissertation selbständig, ohne fremde Hilfe verfasst und mit keinen anderen als den darin angegebenen Hilfsmitteln angefertigt habe, dass die wörtlichen oder dem Inhalt nach aus fremden Arbeiten entnommenen Stellen, Zeichnungen, Skizzen, bildliche Darstellungen und dergleichen als solche genau kenntlich gemacht sind.

Die Arbeit wurde in gleicher oder abgewandelter Form noch keiner anderen Stelle als Prüfungsleistung vorgelegt.

Magdalena Bogus

