

### 3. Ergebnisse

#### 3.1 Klonierung und Sequenzierung der Region um *tbpAB* und *opaA*

Die Region zwischen dem *tbpAB*-Operon und dem *opaA*-Gen wurde ausgewählt, um die Variation von Virulenzgenen mit "house keeping"-Genen vergleichen zu können. Um diesen ca. 30 kb lange Sequenzabschnitt zu analysieren, wurden zunächst Lambda-Genbanken erstellt. Dazu wurde chromosomale DNA des Serogruppe A, Subgruppe IV-1 Stammes Z2491 mit dem Restriktionsenzym *Sau3AI* partiell verdaut und Fragmente mit Größen zwischen 12 bis 23 kb nach Vorschrift in das Lambda-Phagen-System DASH II (Stratagene) kloniert. Die erhaltenen rekombinanten Phagen wurden zusammen mit *E. coli*-Wirtsbakterien ausplattiert und nach der Phagenplaquebildung über Nacht in 96-Well-Mikrotiterplatten einzeln gepickt. Bei einer durchschnittlichen Insert-Größe der Lambda-Phagen von 18 kb ergibt sich bei der Genomgröße von *N. meningitidis* mit 2,2 Mb rechnerisch eine Mindestanzahl von 122 Phagen, um das Genom theoretisch einmal abzudecken. Um eine ca. 10fache Redundanz der klonierten Fragmente zu erhalten, wurden je 12 96er Mikrotiterplatten (entspricht 1152 einzelnen Phagen pro Bank) isoliert. Es wurden vier verschiedene Lambda-Bänke mit Insertgrößen von 12-18 kb, 18-23 kb, 12-15 kb und 15-23 kb erstellt. Diese Phagenbanken wurden mit Hilfe eines 96-Pin-Replikators (Genomed) in je 12 viereckige Petrischalen (20 x 20 cm) ausplattiert, in denen die Wirtsbakterien vorgelegt waren. Die nach der Inkubation über Nacht entstandenen Phagenplaques im 96er Raster wurden durch einen Kontaktabzug auf Nylonmembranen übertragen, um dann mit Digoxigenin-markierten Sonden gegen die an den Membranen haftende Phagen-DNA hybridisieren zu können. Es wurden hierzu das komplette *tbpB*-Gen mit den Primern 3'Met2/P1 und 5'inter1/P2, das *opaA*-Gen mit den Primern O80 und O87b (Tabelle 2.7) und das *tbpA*-Gen mit den Primern *tbp1\_1* und *tbp1\_A* (Tabelle 2.8) als DIG-markierte PCR-Produkte amplifiziert und als Sonden verwendet. Mehrere mit den Sonden reagierende Phagen wurden identifiziert und drei Phagen (C7E1, C12E3 und B10F3) für die Sequenzierung ausgewählt (Siehe Abbildung 3.1).

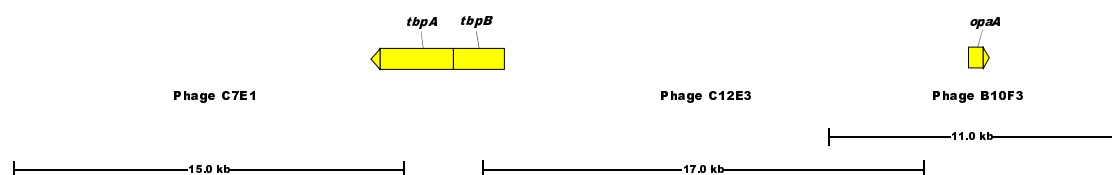


Abbildung 3.1: Überblick über die für die Sequenzierung ausgewählten Lambdaphagen.

Die Gene, aus denen die für die Identifizierung der Phagen verwendeten Sonden amplifiziert wurden, sind als gelbe Blöcke eingezeichnet. Die ermittelte Länge der Phageninserts ist angegeben.

Die zweimal gereinigten und überprüften Einzelplaque-Lysaten dienten nun zur Herstellung von Lysaten mit hohen Titern ( $>10^{12}$  pfu/ml). Aus diesen Phagenkulturen wurden nun die klonierten Abschnitte mittels der *long-range*-PCR (*lr*PCR) unter Verwendung der Primer DASH II T3long und DASH II T7long (siehe Tabelle 2.2) amplifiziert. Die so erhaltenen PCR-Produkte mit einer Länge von 15-18 kb waren nach der Aufreinigung über MICROCON-100 Filter dann Ausgangspunkt für eine sogenannte "*shot gun*"-Klonierung. Dies bedeutet, daß viele kleine DNA-Fragmente eines großen Ausgangsproduktes rein zufällig in einen Vektor kloniert werden und diese Fragmente nach der Sequenzierung von hier ca. 200 Inserts pro *lr*PCR-Fragment wieder mittels des Computers (*in silico*) zu der Sequenz des Ausgangsprodukt zusammengesetzt werden. Dazu wurden die in Puffer gelösten *lr*PCR-Fragmente mechanisch durch das Einleiten von Stickstoff in die Lösung geschert. Über die Variation von Druck und Zeit kann man die Größe der Fragment regulieren, hier wurden Längen von 0,5 bis 1 kb verwendet. Bei dieser sog. "*Nebulisation*" entstehen überwiegend gerade Strangabbrüche, sicherheitshalber wurde aber noch zusätzlich eine Auffüllreaktion der evtl. vorhandenen einzelsträngigen Enden mittels Taq-Polymerase durchgeführt. Die so erhaltenen kurzen Fragmente wurden dann in pUC19 oder in den Vektor pCR2.1 des *TA-cloning-Kits* kloniert. Aus den gewonnenen 200 Klonen pro *lr*PCR-Produkt wurde direkt mittels Kolonie-PCR das klonierte Fragment mit den Primern pUC/M13Forward und pUC/M13Reverse (siehe Tabelle 2.3) amplifiziert, das PCR-Produkt mit dem automatischen Rosys-System aufgereinigt und eine Sequenzierung der so erhaltenen PCR-Fragmente mit dem für beide Vektoren spezifischen Primer M13Rev (siehe Tabelle 2.3) durchgeführt. Nach der Bearbeitung der so erhaltenen Sequenzläufe jedes Phageninserts mit Hilfe des ASP-Programms wurden diese in das im STADEN-Paket enthaltene Programm GAP4 eingelesen und dort zu einer Gesamtsequenz des Inserts zusammengesetzt. Nach dem Bereinigen von Unklarheiten und Fehlstellen in den einzelnen Läufen durch manuelle Editierung konnten die Konsensus-Sequenzen der verschiedenen Inserts der überlappenden Phagen vereinigt werden. Es wurde ein ca. 35 kb langer Sequenzabschnitt erhalten, wobei für die folgende Analyse eine Beschränkung auf einen ca. 24,5 kb langen Bereich (5 kb abwärts von *tbpA* und bis an *opaA* heran) erfolgte. Die Einschränkung auf den kurzen Bereich neben *tbpA* mußte erfolgen, da der Phage C7E1 eine "Chimäre" war, d.h. im Genom nicht benachbarte Sequenzen beinhaltete. Es stellte sich bei einem Vergleich der erhaltenen Sequenzen mit dem Sanger-Genom heraus, daß ein ca. 5 kb langes, den N-terminaler Abschnitt des *tbpA* Gens enthaltendes DNA-Stück zusammen mit einem genomisch völlig anders lokalisiertem DNA-Abschnitt in dem Phagen vereinigt waren. Dieser DNA-Abschnitt wurde dann verworfen. Der Bereich um das *opaA*-Gen wurde zeitgleich von G. Morelli untersucht und die Ergebnisse für die nachfolgenden Analysen mit der Genehmigung von G. Morelli auszugsweise verwendet.

Bei der hier verwendeten Klonierungs- und Sequenzierungsstrategien konnte es zu einer Reihe von Fehlern und diversen unerwünschten Effekten kommen, wie z.B. der bereit erwähnten Chimärenbildung (dem Zusammenfinden zweier nicht benachbarter DNA-Fragmente bei einer Ligationsreaktion) oder einem Einbau falscher Nukleotide durch Taq-Polymerasefehler bei den aufeinanderfolgenden PCR-Reaktionen der Insert-Amplifikation. Darum sollte die aus den Phagen erhaltene Sequenz durch eine Sequenzierung direkt aus dem Genom amplifizierter DNA-Fragmente überprüft und bestätigt werden. Dazu wurden nun mit Hilfe der aus den Phagen gewonnenen Sequenz PCR-Primer für die Amplifikation von vier ca. 5-7 kb langen DNA-Fragmenten ausgewählt (siehe Tabelle 2.4, Seite 34).

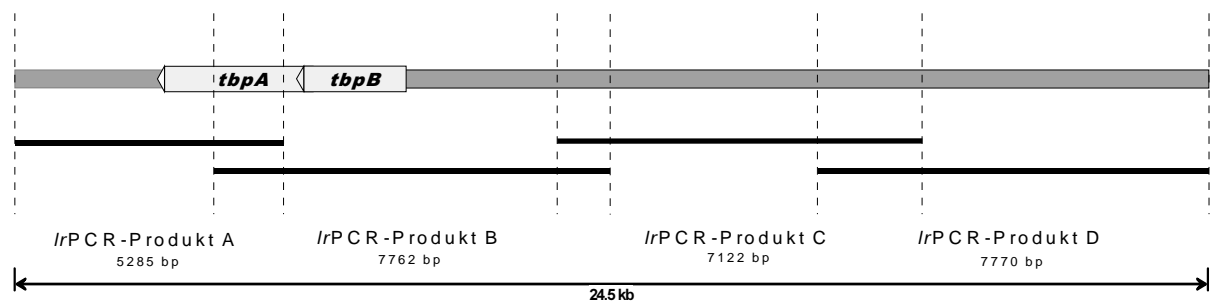


Abbildung 3.2: **Darstellung der IrPCR-Produkte für die Nachsequenzierung aus dem Genom.**

Das *tbpAB*-Operon ist nur als Orientierungshilfe eingezeichnet, das Produkt A wurde mit den Primern *potF\_lr\_forw* und *lr\_rev3*, das Produkt B mit den Primern *forw4* und *rev4*, das Produkt C mit den Primern *lr\_forw5* und *lr\_rev5* und das Produkt D mit den Primern *forw6* und *rev6.2* amplifiziert. (siehe **Error! Reference source not found.**)

Um diese damit erhaltenen Fragmente nun sequenzieren zu können, wurden spezifische Sequenzierungs-Primer in einem Abstand von ca. 400 bp (bei aufeinanderfolgenden Primern) und von ca. 100 bp (bei gegenläufigen Primern) ausgewählt (Tabelle 2.5, Seite 35). Besonders kritisch mußte die Auswahl dieser Primer in den Bereichen mit repetitiven DNA-Elementen (*Repeats*) erfolgen, da selbst speziell für die Aufgabe der Primerauswahl entwickelte Programme wie PRIMERPREMIER4 hier versagten. Hier wurden Primer "per Hand" ausgewählt und nur kleinere interne Abschnitte der *Repeats* amplifiziert und sequenziert, um Mehrfachbindungen der Primer an anderen Stellen zu vermeiden. Zusätzlich wurde mit Hilfe längerer Sequenzierungsläufe (bis ca. 800 bp, sonst nur 500 bis 550 bp) die Konsistenz der erhaltenene Sequenzabschnitte in den *Repeats* überprüft und bestätigt.

Mit diesen Sequenzierungs-Primern wurde nun nach Amplifikation und Aufreinigung der 4 IrPCR-Fragmente (A, B, C und D) die gesamte Region um das *tbpAB*-Operon bis hin zu dem Beginn der *Repeats* vor *opaA132* direkt vom Chromosom bestätigt. Die Sequenz und Struktur der *Repeats* um das in Z2491 vorhandene *opaA*-Gen wird an anderer Stelle veröffentlicht (Morelli *et al.*, 1998), so daß auf diesen Abschnitt hier verzichtet wurde. In

der Abbildung 3.3 auf Seite 66 ist die gesamte ermittelte Sequenz mit der auf ihr gefundenen offenen Leserahmen (*open reading frames*, kurz ORFs) dargestellt.

Die erhaltene Nukleotidsequenz wurde mit dem Programm CODONUSE 3.5.5 (MacIntosh) unter Verwendung der für *N. meningitidis* typischen Verwendungshäufigkeit von Kodons (*codon usage bias*) und des alternativen Startcodons GTG (gewöhnlich ATG) übersetzt. Die dabei erhaltenen Proteinsequenzen der offenen Leserahmen wurden mittels der Internet-Suchprogramme BLAST (Altschul *et al.*, 1990) und PSI-BLAST (Altschul *et al.*, 1997) gegen bereits bekannte Proteinsequenzen in den Internet-Datenbanken (alle nicht-redundanten GenBank CDS Übersetzungen, PDB, SwissProt, PIR und PRF) geprüft.

### 3.2 Analyse der kodierenden Bereiche

Der untersuchte Bereich wird von dem *opaA*-Gen und dem ca. 3 kb abwärts von *tbpA* liegenden partiellen offenen Leserahmen mit einer Homologie zu dem Gen *potF* begrenzt. Das *tbpAB*-Operon und das *opaA*-Gen sind von ca. 2 kb langen Sequenzbereichen mit repetitiven DNA-Elementen, die in kurzen Abständen aufeinander folgen (*Repeats*), umgeben. Bei allen anderen Genen in diesem Abschnitt sind keine dieser flankierenden *Repeats* gefunden worden.

Es wurden folgende ORFs identifiziert und deren Protein-Übersetzungen analysiert (siehe auch Tabelle 3.1):

#### **potF:**

Offener Leserahmen mit einer Identität von 60% und einer Homologie von 90% zu dem 370 AA langen periplasmatischen Putrescin-Transportprotein PotF aus *E. coli* mit der Zugriffsnummer A45313. Der ORF ist nur der C-Terminus des reifen Proteins, das Gen ist bei der Analyse nicht vollständig sequenziert worden. Das Protein ist für die Bindung und den Transport von Putrescin (einem Polyamin) im Periplasma verantwortlich. In *E. coli* ist *potF* mit den Genen *potG*, *potH* und *potJ* zu einem Operon zusammengefaßt.

#### **tbpA:**

Das unprozessierte TbpA-Protein (mit 24 AS langem Signalpeptid) aus *N. meningitidis* Z2491

#### **tbpB:**

Das unprozessierte TbpB-Protein (mit 20 AS langem Signalpeptid) aus *N. meningitidis* Z2491

## Ergebnisse

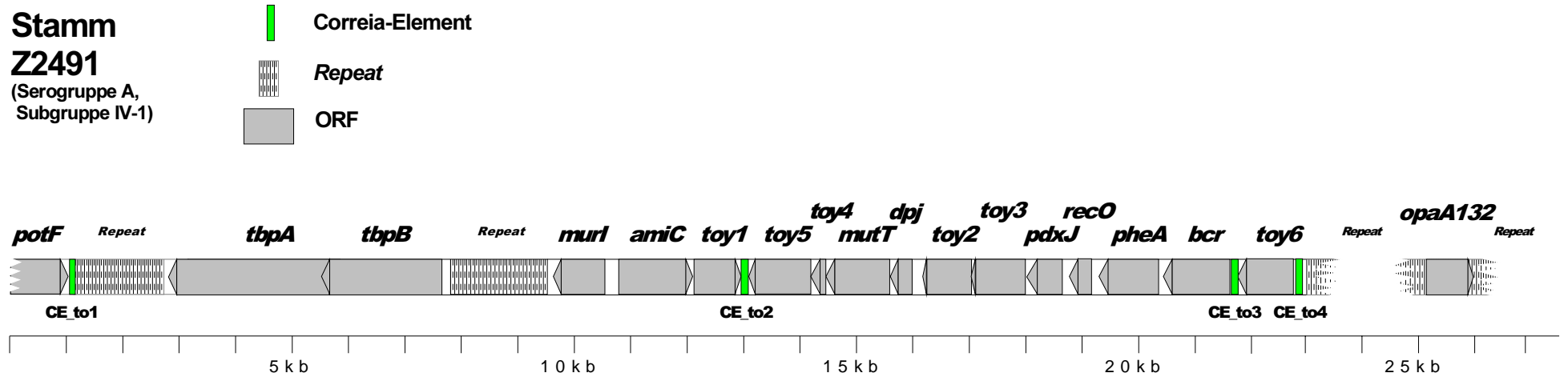


Abbildung 3.3: **Darstellung der gefundenen offenen Leserrahmen und Repeats im Stamm Z2491.**

Alle gefundenen offenen Leserahmen (*open reading frames*, ORF) werden als grauer Kasten mit einem Pfeil in der vermuteten Transkriptionsrichtung dargestellt. Die Genbezeichnungen sind über den ORFs angegeben. Repetitive DNA-Elemente sind als schwarzweiß gestreifter Kasten dargestellt, in dieser Arbeit nicht analysierte Bereiche um das in Z2491 vorhandene *opaA132*-Allel sind durch Abbruchkanten angedeutet. Correia-Elemente der *tbp-opaA*-Region (CE\_toX) sind durchnummeriert und als hellgrüne Blöcke dargestellt.

Ergebnisse

Tabelle 3.1: Gefundene ORFs in dem *tbpAB-opaA* Genabschnitt in Z2491.

Name des ORFs	<i>potF</i>	<i>tbpA</i>	<i>tbpB</i>	<i>murI</i>	<i>amiC</i>	<i>toy1</i>	<i>toy5</i>	<i>toy4</i>	<i>mutT</i>
Länge der ORFs in Aminosäuren	249	910	698	270	458	272	380	140	269
Wahrscheinlichkeit (P) eines zufälligen Treffers	1e <sup>-38</sup>	0	0	1e <sup>-52</sup>	1e <sup>-100</sup>	-	2e <sup>-68</sup>	-	5e <sup>-18</sup>
Name des Genbankeintrags	<i>potF</i>	<i>tbpA</i>	<i>tbpB</i>	<i>murI</i>	<i>amiC</i>	-	YCBY	-	<i>mutT</i>
Organismus	<i>E. coli</i>	<i>N. men</i>	<i>N. men</i>	<i>B. subt</i>	<i>E. coli</i>	-	<i>E. coli</i>	-	<i>H. inf</i>
Länge der Genbankeinträge	370	910	698	272	417	-	702	-	234
Zugriffsnummer	M93239	AF058689	AF058689	P94556	Q46929	-	P75864	-	I64101
Name des ORFs	<i>dpj</i>	<i>toy2</i>	<i>toy3</i>	<i>pdxJ</i>	<i>recO</i>	<i>pheA</i>	<i>bcr</i>	<i>toy6</i>	<i>opaA</i>
Länge der ORFs in Aminosäuren	125	255	259	242	264	375	410	255	265
Wahrscheinlichkeit (P) eines zufälligen Treffers	2e <sup>-16</sup>	-	-	1e <sup>-79</sup>	1e <sup>-107</sup>	0	4e <sup>-49</sup>	2e <sup>-37</sup>	0
Name des Genbankeintrags	<i>dpj</i>	-	-	<i>pdxJ</i>	<i>recO</i>	<i>pheA</i>	<i>bcr</i>	HI0976	<i>opa132</i>
Organismus	<i>E. coli</i>	-	-	<i>E. coli</i>	<i>N. gon</i>	<i>N. gon</i>	<i>H. inf</i>	<i>H. inf</i>	<i>N. men</i>
Länge der Genbankeinträge	126	-	-	243	247	375	398	170	240
Zugriffsnummer	P24224	-	-	P24223	AF047375	AF047375	P45123	U32778	AF001180

*N. men*: *N. meningitidis*; *B. subt*: *B. subtilis*; *H. inf*: *H. influenzae*; *N. gon*: *N. gonorrhoeae*

**murI:**

Offener Leserahmen mit einer Identität von 41% und einer Homologie von 58% zu einer Glutamat-Racemase aus *Bacillus subtilis*. In *E. coli* ist MurI als Glutamat-Racemase an der Umwandlung von L-Glutamin in D-Glutamin und umgekehrt beteiligt und ist in der Murein-Synthese involviert.

**amiC:**

ORF mit einer Identität von 55% und einer Homologie von 75% zu dem N-Acetylmuramoyl-L-Alanin Amidase-Vorläufer AmiC aus *E. coli*. Dieses periplasmatische Protein ist für die Hydrolyse der Bindungen zwischen N-Acetylmuroamyl und L-Alanin aus Glykopeptiden verantwortlich und wahrscheinlich an der Zellwand-Hydrolyse, Septierung der Zellen und beim Wiederverwenden von Zellwandbestandteilen beteiligt.

**toy1:**

offener Leserahmen ohne signifikante Homologie zu bekannten Proteinen in den Datenbanken. *toy* steht für (*tbpAB-opaA*-Region,  $\gamma$ : unbekanntes Protein). Die höchste Homologie zu *toy1* zeigt ein cDNA EST aus *Caenorhabditis elegans* (EMBL:C11278) mit  $P=1e^{-11}$ .

**toy5:**

ORF mit einer Identität von 38% und einer Homologie von 55% mit einem hypothetischen 78.9 KDa großen Protein in der *pyrD-pqiA* intergenischen Region in *E. coli* mit der Bezeichnung YCBY.

**toy4:**

ORF ohne signifikante Homologie zu bekannten Proteinen in den Datenbanken.

**mutT:**

ORF mit einer Identität von 35% und einer Homologie von 55% mit dem Mutator-Protein MutT aus *H. influenzae*. Das Protein MutT verhindert AT-GC Transversionen. Bei einer Mutante erhöht sich die Rate der T:A  $\rightarrow$  G:C Transversionen, da MutT 8-Oxo-7,8-Dihydroguaninriphosphat (8-oxoGTP) abbaut und somit eine Inkorporation dieser mutagenen Base verhindert. Ist 8-oxoGTP (ein Produkt einer Sauerstoffradikal-Reaktion von Guanin mit oxidierenden Agenzien) in einem DNA-Strang vorhanden, wird bevorzugt Adenin in den komplementären Strang eingebaut, was zu der Transversion führt.

**dpj:**

ORF mit einer Identität von 28% und einer Homologie von 46% zu dem Dpj-Protein aus *E. coli*. Dieses magnesiumabhängige Enzym liegt als Dimer vor und transferiert im Zytoplasma die 4'-Phosphopantethein-Gruppe des Coenzym A auf das Serin 36 des Acyl-Transporter-Proteins. In *E. coli* bildet *dpj* ein Operon mit *pdxJ*, das hier 2 ORFs weiter vorliegt.

**toy2:**

ORF ohne signifikante Homologie zu bekannten Proteinen in den Datenbanken.

**toy3:**

ORF ohne signifikante Homologie zu bekannten Proteinen in den Datenbanken.

**pdxj:**

ORF mit einer Identität von 58% und einer Homologie von 71% zu dem Pyridoxalphosphat Biosynthese Protein PdxJ (Vitamin B-6 Biosynthese) aus *E. coli*. Bei der Biosynthese des Pyridoxal-Phosphats werden die Enzyme aus den Genen *pdxA*, *pdxB*, *pdxJ* und *serC* für den Aufbau des Pyridoxin-Skeletts aus Grundbausteinen benötigt, *pdxH* wandelt dann Pyridoxol in Pyridoxal um. In *E. coli* bildet *pdxJ* ein Operon mit *dpj*.

**recO:**

ORF mit einer Identität von 80% und einer Homologie von 85% zu dem DNA-Reparatur-Protein RecO (*recombination protein O*) aus *N. gonorrhoeae*. In *E. coli* ist RecO ein DNA-Einzelstrang-bindendes Protein, das an der DNA-Reparatur beteiligt ist. Seine genaue Funktion und seine Partner sind jedoch noch nicht bestimmt.

**pheA:**

ORF mit einer Identität von 98% und einer Homologie von 98% zu der Chorismat Mutase/ Prephenat Dehydratase PheA aus *N. gonorrhoeae*. Dieses Enzym ist in *E. coli* an der Biosynthese der aromatischen Aminosäuren beteiligt, im besonderen bei der Phenylalanin-Erzeugung. Dabei wird erst Chorismat zu Prephenat umgewandelt und dann unter Abspaltung von CO<sub>2</sub> und H<sub>2</sub>O Phenolpyruvat erzeugt.

**bcr:**

ORF mit einer Identität von 29% und einer Homologie von 50% zu dem Bicyclomycin-Resistenz-Protein Homolog (Bcr) aus *H. influenza*. In *E. coli* ist Bcr ein membranständiges Protein, das Antibiotika wieder über die Zellwand ausschleust, um es aus der Zelle zu entsorgen. Es ist an der Sulfathiazol/Sulfonamid-Resistenz beteiligt.



**toy6:**

ORF mit einer Identität von 50% bzw. 48% und einer Homologie von 65% bzw. 69% zu den C- und N-terminalen Einträgen für das konservierte hypothetische Protein HI0976 aus *H. influenzae* mit der Zugriffsnummer U32778.

**opaA132:**

Das unprozessierte OpaA132-Protein aus *N. meningitidis* Z2491. Die Aminosäuren 1-24 entsprechen dem Signalpeptid, das reife Protein kodiert ab Aminosäure 25.

### **3.3 Analyse der nichtkodierenden Bereiche**

#### **3.3.1 Repetitive DNA-Elemente um das *tbpAB*-Operon (NIMEs)**

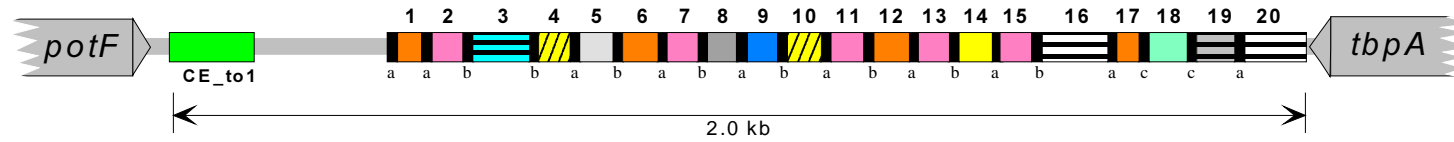
Die in Abbildung 3.3 mit *Repeat* bezeichneten Sequenzabschnitte flankieren die Gene *tbpAB* und *opaA132* und sind je ca. 2 kb lang. Die generelle Organisation der *Repeats* um *tbpAB* ist in Abbildung 3.4 dargestellt. Die Analyse der *Repeats* um *tbpAB* umfaßt die Bereiche vom Stop-Kodon des *potF*-Gens bis zum Stop-Kodon von *tbpA* und vom Start-Kodon des *tbpB*-Gens bis zum Stop-Kodon des *murI*-Gens. Diese beiden Bereiche werden als *tbpA-Repeat* und als *tbpB-Repeat* bezeichnet. Die hier zum Vergleich hinzugezogenen *Repeats* um das *opaA132*-Gen stammen aus einer bisher unveröffentlichten Analyse von G.Morelli und werden mit Erlaubnis hier teilweise verwendet.

Die *Repeats* um das *tbpAB*-Operon konnten im Gegensatz zu den das *opaA*-Gen flankierenden *Repeats* ohne Subklonierungen erfolgreich durchsequenziert werden, da zwischen den sich wiederholenden Elementen einzelne Abschnitte vorhanden waren, die für spezifische Sequenzierungs-Primer genutzt werden konnten (siehe Tabelle 2.5)

Ergebnisse

**Serogruppe A,  
Subgruppe IV-1,  
Z2491:**

***tbpA* - Repeat**



***tbpB* - Repeat**



dRS3-Elemente:

dRS3a,b,c: a,b,c

rs-Element-Familien:

RS14:	RS26:
RS15:	RS27:
RS22:	RS28:
RS23:	RS29:
RS24:	RS30:
RS25:	

Abbildung 3.4: **Darstellung der Repeats um das *tbpAB*-Operon.**

Alle repetitiven Elemente sind als Kästen dargestellt, nichtrepetitive Sequenzabschnitte als graue Linie. Der Typ der dRS3-Elemente ist unter den schwarzen Balken als a, b oder c angegeben (Vergleiche Tabelle 3.2). Die rs-Elemente wurden zu Familien zusammengefaßt; ihre Familienzugehörigkeit ist aus der Farbkodierung der Legende abzulesen. Die Numerierung bezieht sich auf die Abfolge der rs-Elemente in Z2491 (jeweils von 1-20 im *tbpA*-Repeat und von 21-47 im *tbpB*-Repeat). Zwischen zwei rs-Elementen aus einer Familie können bis zu 20% Polymorphismen auf Nukleotidebene auftreten. Die Gene *potF*, *tbpAB* und *murI* sind nicht maßstabsgerecht gezeichnet. Das Correia-Element CE\_to1 ist extra bezeichnet.

Es wurden 3 verschiedene dRS3-Elemente in den *Repeats* um *tbpAB* und um *opaA* gefunden. Die dRS3s wurden nach ihrer 8 bp langen internen Sequenz klassifiziert, um sie unterscheiden zu können, es können aber durchaus noch andere Variationen im Genom vorkommen. (Tabelle 3.2)

dRS3a	<b>ATTCCy</b> ACGAAAGy <b>GGAAT</b>
dRS3b	<b>ATTCCs</b> rmGyAGGC <b>GGAAT</b>
dRS3c	<b>ATTCCC</b> GCrAAAGC <b>GGAAT</b>

Tabelle 3.2: **Gefundene dRS3 Elemente.**

(nur ein Strang angegeben, können auch als revers-komplementäre Sequenz vorkommen)

y = C/T, s = G/C, r = G/A, m = C/A

In der Tabelle 3.3 sind die dRS3-Elemente nach ihrem Auftreten in den *Repeats* aufgelistet. Nur im *tbpA-Repeat* kommen dRS3c-Elemente vor, diese sind im Vergleich mit den dRS3a- und dRS3b-Elementen unterrepräsentiert. Das häufigste Element ist das dRS3b mit insgesamt 26 Einheiten in beiden *Repeats* (siehe auch Abbildung 3.4).

	dRS3a	dRS3b	dRS3c
<i>tbpA-Repeat</i>	7	11	2
<i>tbpB-Repeat</i>	12	15	-

Tabelle 3.3: **Anzahl der dRS3-Elemente in den *Repeats* um *tbpAB*.**

In der Tabelle 3.4 ist die Anzahl der rs-Elemente in den einzelnen *Repeats* aufgelistet. Die rs-Es mit der Bezeichnung RS28, RS29 und RS30 sind nur jeweils im *tbpA-Repeat* vorhanden, während alle anderen rs-Es in beiden *Repeats* auftreten. RS14 und RS15 wurden auch in den *opaA-Repeats* gefunden.

	RS14	RS15	RS22	RS23	RS24	RS25	RS26	RS27	RS28	RS29	RS30
<i>tbpA-Repeat</i>	4	1	2	1	1	1	5	1	2	1	1
<i>tbpB-Repeat</i>	6	1	1	3	2	2	10	2	-	-	-

Tabelle 3.4: **Anzahl der RS-Elemente in den *Repeats* um *tbpAB*.**

Die Konsensus-Sequenzen aller in den *tbpAB-Repeats* gefundenen rs-Elemente sind in Abbildung 3.5 gezeigt. Einzelne rs-Elemente zeigen Längenvariationen, während die meisten rs-Es eine definierte Länge besitzen. Es können aber auch Insertionen einzelner Nukleotide auftreten (z.B. in RS26). Die *Repeats* sind nun aus diversen *NIMEs* (dRS3-rs-E-dRS3) zusammengesetzt, wobei auch "halbe" *NIMEs* (nur ein dRS3 Element zusammen mit einem rs-E) möglich sind. Die Abfolge und Kombinationen der *NIMEs* ist unterschiedlich in den *Repeats* auf- und abwärts von *tbpAB* und *opaA*, so daß hier von einer Orientierung des gesamten Abschnittes *Repeat-Gen-Repeat* gesprochen werden kann. Auch sind die Sequenzen der meisten rs-Es in den *Repeats* um das *opaA*-Gen signifikant

unterschiedlich zu denen um das *thpAB*-Operon. Nur zwei der rs-Es (RS14 und RS15, siehe Abbildung 3.5) treten in beiden *Repeat*-loci auf.

**RS14 (139bp) :**

GACGAATTCGAGATTGCGGTGTTGTGCGGGAATGACGRATCCATCCRYACGGAAACCTGCAYCMCGTCATTCCYACGAACCACATCCCGTCATTCCWCARRACAGAAAACCA  
AAAWCAGAAAACCTAAAATCCCGTC

**RS15 (61bp) :**

CYRRDYCKKTCVGYRCRGRAACTTATCGGRWAAAAMGGTTTCTTSAGATTTTRCRTCTRG

**RS22 (54bp) :**

CCAGTGCCTTGAGYTTGAGCTATTTAGAATAAATTTTGAAACTCTAATCGCGTC

**RS23 (52bp) :**


CYRGYTCGTTSGGTTTCGCTTGTTTTAAGTTTCGGGTAACCTCCACTTCGTC

**RS24 (55bp) :**

GACGGGATTTGAGATTGCGGYATTTATCGGGAGCAACAGAAGCCGCTCTGCCGTC

**RS25 (max. 102bp) :**

GACGGCGGACGGTTTCTGTTTKYKYGCGTAAATACCCACAAGCTAAAATCCTGTTATTTTCACAAAAACAGAAAACCAAAACAGCAACCTGAAATTCGTC **Position**  
3, 36  
41




**RS26 (56-57bp) :**

<sup>T</sup>  
CYAGTTTTTTGAGTTTCAGTCATTYYCRAKAAATTRCCTTARYATTGAATGYCTRG

**RS27 (max. 100bp) :**

GACGGGATTTKAGGTTTCTGTTWHKYKHBKMYMTGTTTTTCGTAGGAATGACGGTTTAGAGTTACCCGAAACCTCAAAAAAACCGAAACCGAACGGACTGG **19**  
**45**  
**47**



**RS28 (120 bp) :**

GACGGGWTTYRAGATTACRGTKTRTCGGGRAKATGGGAAACGGTGGGAATTGTGTAAAAAATGCCGCTCTGAAGCCTTTCAGACGGCATTTCGGGCATTYGGACRTTTAAAMYYCMWTY

Abbildung 3.5: **Konsensus-Sequenzen der rs-Element-Familien.**

Y = C/T, R = A/G, K = G/T, S = G/C, die grauen Balken geben mögliche Deletionen in einzelnen rs-Elementen aus dieser Familie an. Diese sind mit ihren Positionen aus Abbildung 3.4 angegeben. Die Elemente RS14 und RS15 kommen auch in den *Repeats* um *opa132* vor.

### 3.3.2 Weitere repetitive DNA-Elemente

In der analysierten Sequenz sind auch 4 Strukturen enthalten, die nach ihrem Entdecker als Correia-Elemente beschrieben werden (Correia *et al.*, 1988). Sie liegen unmittelbar hinter dem Stop-Kodon von *potF* (CE\_to1), zwischen *toy1* und *toy5* (CE\_to2) und vor und hinter *toy6* (CE\_to3 und CE\_to4). Die hier vorkommenden Correia-Elemente sind 151 bis 155 bp lang und liegen als direkte oder invertierte repetitive Elemente zueinander vor (siehe Tabelle 3.5). Ihre Funktion ist bis jetzt unklar, es besteht aber die Möglichkeit, daß Correia-Elemente Bindungsstellen für eine DNA-Gyrase besitzen ( Morelli *et al.*, 1998).

Name des Correia-Elements	CE_to1	CE_to2	CE_to3	CE_to4
Position des Correia-Elements <sup>1</sup>	830	13620	22610	23890
Länge in bp	155	151	154	154
Orientierung <sup>2</sup>	⇒	⇐	⇐	⇒

Tabelle 3.5: **Correia-Elemente in Z2491.**

<sup>1)</sup> Position relativ zum ersten Nukleotid der Sequenz von *potF* bis *opaA*-Repeats in Z2491 (beginnend mit dem Primer *potF\_lr\_forw*), <sup>2)</sup> Orientierung in Bezug auf CE\_to1.

### 3.4 Vergleich des IV-1 Stammes Z2491 mit dem ET-37 Komplex Stamm Z4400

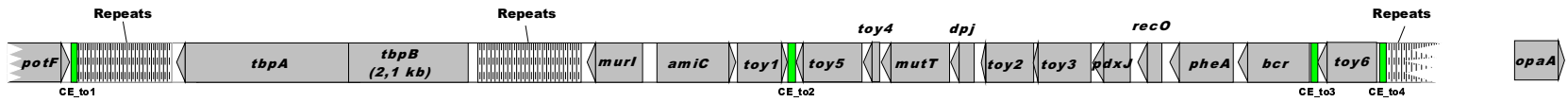
Um die Sequenzvariabilität zwischen den genetisch relativ weit auseinander liegenden (vgl. Abbildung 1.1) Serogruppen A und C untersuchen zu können, wurde der gleiche Sequenzabschnitt wie bei der Serogruppe A Subgruppe IV-1 auch aus einem Stamm der Serogruppe C ET-37 Komplex mit der Bezeichnung Z4400 amplifiziert und sequenziert.

Wie in Abbildung 3.6 dargestellt, zeigt die gewonnene Sequenz die gleiche Anordnung der offenen Leserahmen wie im IV-1 Stamm. Ausgenommen von diesem Vergleich sind die *Repeat*-Regionen um *tbpAB* und *opaA*, da schon in den ersten 100 Nukleotiden deutliche Abweichungen zu den Sequenzen aus dem IV-1 Stamm auftraten. Damit war eine direkte Sequenzierung dieser ca. 2 kb langen Bereiche mit den aus dem IV-1-Stamm abgeleiteten Primern nicht möglich. Da es hier in erster Linie um den Vergleich der beiden Sequenzabschnitte in den kodierenden Bereichen und dem möglichen Austausch von DNA zwischen den beiden Subgruppen ging, wurde auf eine detaillierte Analyse dieser repetitiven Bereiche des ET-37 Komplex Stammes verzichtet. Der Vergleich beschränkt sich somit auf das *potF*-Gen, das *tbpAB*-Operon, den Bereich zwischen *murI* und *toy6* und das *opaA*-Gen. Bekannt war, daß im Vergleich zu den IV-1 Stämmen jeweils eine andere Variante des *tbpAB*-Operons und des *opaA*-Gens in den ET-37 Komplex Stämmen vorliegen. Es zeigte sich aber, daß sich alle ORFs von denen der Subgruppe IV-1 in der Nukleotid- und Aminosäuresequenz unterschieden. Die Abweichungen auf Proteinebene und Nukleotidebene sind in der Tabelle 3.6 angegeben.

Ergebnisse

**Stamm  
Z2491**

(Serogruppe A,  
Subgruppe IV-1)



**Stamm  
Z4400**

(Serogruppe C,  
ET37-Komplex)

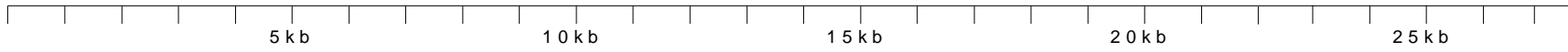
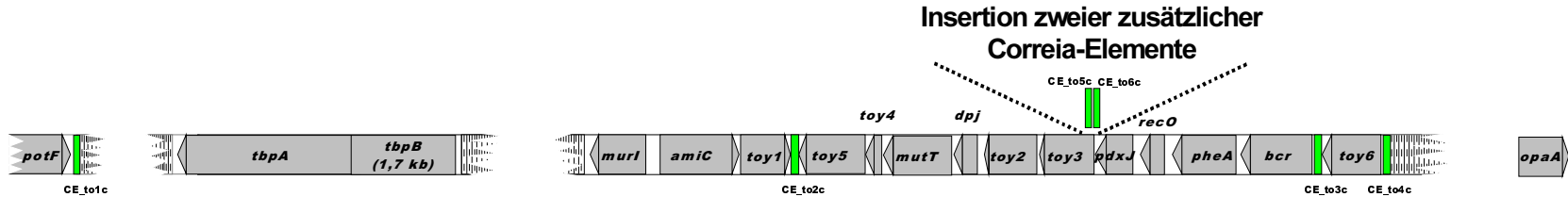


Abbildung 3.6: Vergleich der Sequenzen des Subgruppe IV-1 Stammes mit dem ET-37 Komplex-Stamm.

	<i>potF</i>	<i>tbpA</i>	<i>tbpB</i>	<i>murI</i>	<i>amiC</i>	<i>toy1</i>	<i>toy5</i>	<i>toy4</i>	<i>mutT</i>	<i>dpj</i>	<i>toy2</i>	<i>toy3*</i>	<i>pdxJ</i>	<i>recO</i>	<i>pheA</i>	<i>bcr</i>	<i>toy6</i>	<i>opaA</i>
Protein-Ähnlichkeit in %:	99,2	80,9	57,6	99,3	98,3	98,2	98,4	100	96,6	96,8	98,4	97,6	98,3	89,0	98,7	99,5	98,8	85,5
Protein-Identität in %:	99,2	77,1	48,8	99,3	98,0	98,2	97,6	100	96,3	96,0	98,0	96,9	96,7	88,6	98,1	99,5	98,8	83,8
DNA-Identität in %:	97,2	75,6	64,2	96,1	97,7	99,0	96,2	99,7	96,5	94,4	98,8	98,6	95,3	91,7	98,8	97,7	99,0	89,2

Tabelle 3.6: Vergleich der Sequenz der offenen Leserahmen des IV-1-Stammes mit denen des ET-37 Komplex Stammes.

Im ET-37 Komplex Stamm enthält *toy3* ein integriertes 155 bp langes Correia-Element (CE\_to5c) und ein verfrühtes Stop-Kodon. Durch die Integration des Correia-Elements ändern sich die Protein- und DNA-Identitäten und Ähnlichkeiten in *toy3* nicht, da das Element bei der Berechnung nicht berücksichtigt wird. Das 107 bp lange Correia-Element (CE\_to6c) liegt zwischen *toy3* und *pdxJ*.

In dem Serogruppe C, ET-37 Komplex Stamm Z4400 liegen ausserdem zwei zusätzliche Correia-Elemente in dem Bereich um *toy3*. Auffällig ist der geringe Abstand von nur ca. 200 bp zwischen den beiden Elementen. Sie unterscheiden sich auch in ihrer Länge voneinander. In Z2491 kommen in dem Abschnitt zwischen *potF* und *opaA* nur Correia-Elemente mit einer Länge von 151-155 bp vor, in Z4400 ist jedoch ein Element mit nur 107 bp gefunden worden (siehe Tabelle 3.7). Dieses 107 bp langes Correia-Element mit der Bezeichnung CE\_to6c ist so in *toy3* inseriert, daß das letzte Nukleotid des Elements gleichzeitig das erste Nukleotid des putativen *toy3*-Startkodons bildet. Das 155 bp lange CE\_to5c-Element liegt innerhalb des *toy3*-ORFs und führt zu einer Verschiebung des Leserahmens mit mehreren Stop-Kodons. Jedoch befindet sich in dem ET-37 Komplex Stamm schon ein Stop-Kodon vor der Integrationsstelle des Correia-Elements, so daß eine Expression von *toy3* in Z4400 unwahrscheinlich ist oder die Expression diese ORFs in beiden Stämmen generell erst weiter stromabwärts beginnt, wo sich ein weiteres mögliches ATG Start-Kodon in der Sequenz befindet.

Name des Correia-Elements	CE_to1c	CE_to2c	CE_to5c	CE_to6c	CE_to3c	CE_to4c
Pos. des Correia-Elements <sup>1</sup>	830	13000	18300	18500	22500	23500
Länge	155	155	155	107	155	155
Orientierung <sup>2</sup>	⇒	⇐	⇐	⇐	⇐	⇒

Tabelle 3.7: **Correia-Elemente in Z4400.**

<sup>1</sup>) Position relativ zum ersten Nukleotid der Sequenz von *potF* bis *opaA*-Repeats in Z2491 (beginnend mit dem Primer *potF\_lr\_forw*), <sup>2</sup>) Orientierung in Bezug auf das Correia-Element CE\_to1c. Die im Vergleich mit Z2491 zusätzlich gefundenen Correia-Elemente sind grau unterlegt.

### 3.5 Sequenzvariationen in der untersuchten Region

Es war bereits bekannt, daß um das *tbpAB*-Operon häufig Sequenzvariationen auftreten können. Um nun einen Überblick über die Art und Frequenz dieser Ereignisse innerhalb mehrerer Jahrzehnte und in verschiedenen Ländern zu erhalten, wurden verschiedene Gruppen von Stämmen untersucht. Um gleichzeitig die Größe dieser rekombinanten Bereiche zu untersuchen, wurden aus *tbpB* selbst, den Genen *potF* und *amiC* (ca. 5 kb auf- und abwärts von *tbpB*) und *opaA* (ca. 20 kb aufwärts von *tbpB*) jeweils ca. 550 bp große Bereiche ausgewählt, die aus allen Stämmen amplifiziert und sequenziert wurden. Mit diesen Genabschnitten wurde somit ein ca. 28 kb großer Bereich abgedeckt, in dem Rekombinationsereignisse detektierbar waren.

Untersucht wurden:

- frühe Isolate der Serogruppe A, Subgruppen III, IV-1 und IV-2
- Stämme der Subgruppe IV-1 aus Mali im Vergleich zu frühen Isolaten der Gambia-Epidemie (1982)
- 98 Stämme der Serogruppe A, Subgruppe IV-1 aus Gambia in Vergleich zu 96 Stämmen der Serogruppe C, ET-37 Komplex

Den bei dieser Analyse gefundenen Allelen der einzelnen Gene wurden willkürliche Nummern (z.B. *tbpB5* oder *opaA132*) zugeordnet, um sie unterscheiden zu können.

### 3.5.1 Variationen in älteren Isolaten

Die u.a. durch die MLEE (Wang *et al.*, 1992) definierten Subgruppen III, IV-1 und IV-2 der Serogruppe A zeigen ein stark unterschiedliches epidemiologisches Verhalten. Die Subgruppe IV-1 ist größtenteils auf Afrika beschränkt und verursacht dort seit den 60er Jahren wiederholt Epidemien (Olyhoek *et al.*, 1987). Die Subgruppe IV-2 war für große Epidemien vor dem 2ten Weltkrieg verantwortlich und ist seither rar geworden. Die Subgruppe III verursachte zwei pandemische Wellen. Die erste Welle begann in den 60ern und endete in den 70ern, die zweite Welle begann Anfang der 80er Jahre und dauert bis heute an. Bis auf wenige Ausnahmen waren die ältesten Isolate dieser drei Subgruppen identisch in vielen *loci* wie z.B. den *opa-loci* und dem *iga-locus* und sind wahrscheinlich vor ca. 100 Jahren aus einem gemeinsamen Vorfahren entstanden (Morelli *et al.*, 1997). Diese die ältesten vorhandenen Isolate der Serogruppe A, Subgruppen III (1966), IV-1 (1966) und IV-2 (1917) wurden nun hier untersucht. Bei dem Vergleich fällt auf, daß die Subgruppen IV-1 und IV-2 in allen hier untersuchten Allelen fast identisch sind (die *tbpB*-Allele 1 und 5 unterscheiden sich nur in einem Nukleotid) und nur die Subgruppe III in den *potF*- und *amiC*-Allelen abweicht (Tabelle 3.8). Dies ist umso mehr erstaunlich, da die Isolate aus drei verschiedenen Kontinenten stammen (Amerika, Asien und Afrika) und die IV-1 und IV-2-Isolate zeitlich weit auseinander liegen. Die drei Subgruppen scheinen von einem gemeinsamen Vorfahren abzustammen und in der ersten Hälfte dieses Jahrhunderts noch nicht besonders viele Rekombinationen oder Mutationen erfahren zu haben. Bemerkenswert ist die Übereinstimmung der *tbpB*-Allele, da ja gerade dieses Gen sehr variabel ist.



Subgruppe (Isol.-Jahr) Land	Stamm	<i>potF</i> Allel	<i>tbpB</i> Allel	<i>amiC</i> Allel	<i>opaA</i> Allel
IV-2 (1917) U.S.A.	Z1293	1	5	1	132
III (1966) China	Z3906	3	1	2	132
IV-1 (1966) Kamerun	Z1362	1	1	1	132

Tabelle 3.8: Auftreten der untersuchten Allele in frühen Isolaten.

Identische Allele sind hellgrau unterlegt. Die Allele 1 und 5 unterscheiden sich nur in einem einzelnen Nukleotid, das keine Aminosäureänderung bewirkt und werden somit als sehr eng miteinander verwandt angesehen.

### 3.5.2 Variation der Subgruppe IV-1 in verschiedenen Ländern

Es wurden verschiedene Subgruppe IV-1 Isolate vom Beginn der Epidemie 1982 in Gambia mit dem Referenzstamm Z2491 (1983, Gambia) und endemischen Isolaten aus den 90igern Jahren aus Mali verglichen, um die Entwicklung in der Subgruppe IV-1 zu untersuchen. Hier zeigte sich, daß das *potF*-Allel 1 aus den frühen IV-1 und IV-2-Isolaten auch in den IV-1-Stämmen in Mali (1990-1992) wiederzufinden ist (Tabelle 3.9), wohingegen in den Stämmen aus Gambia eine neue Variante (hier Allel 2) vorherrscht. Vier von fünf der Isolate aus der Gambia-Epidemie sind identisch, was für die Ausbreitung eines einzelnen Klons spricht. Nur in einem einzelstem Stamm aus Gambia (G786) ist das ursprüngliche *potF*-Allel 1 noch zu finden, dieser Stamm weicht aber in drei der vier Allele von diesem Klon ab und wurde nicht noch ein zweites Mal isoliert. Es scheint mit Beginn der Epidemie in Gambia 1982 zu einer Fixierung dieses neuen *potF*-Allels 2 gekommen zu sein.

In den untersuchten Subgruppe IV-1 Stämmen aus Mali hingegen finden sich ein Isolat (Z4421), das noch den frühen Alleltyp (*potF1/tbpB1/amiC1/opaA132*) trägt. Somit kann man davon ausgehen, daß dieser ursprüngliche Alleltyp in Mali weiter besteht. Die anderen untersuchten Isolate zeigen aber eine große Heterogenität in ihren *tbpB*-Allelen. In Mali scheint das *potF1*-Allel vorherrschend zu sein, da nur ein Isolat das neuere *potF2*-Allel trägt.

Subgruppe (Isol.-Jahr) Land	Stamm	<i>potF</i> Allel	<i>tbpB</i> Allel	<i>amiC</i> Allel	<i>opaA</i> Allel
IV-1 (1982) Gambia	G1	2	1	1	132
	G371	2	1	1	132
	G372	2	1	1	132
	G378	2	1	1	132
	G786	1	4	3	132
IV-1-Standard (1983) Gambia	Z2491	2	1	1	132
IV-1 (1990-92) Mali	Z4409	1	24	4	132
	Z4421	1	1	1	132
	Z4860	2	25	1	132
	Z6132	1	26	1	132
	Z6138	1	26	1	132

Tabelle 3.9: **Gefundene Allele in der Subgruppe IV-1 in verschiedenen Ländern.**  
Identische Allele sind hellgrau bzw. dunkelgrau unterlegt, um Gruppen kenntlich zu machen

### 3.5.3 Rekombinationshäufigkeiten in klonalen Gruppierungen

Um zu klären, wie häufig nun Rekombinationsereignisse in der Natur stattfinden, wurden 2 größere Gruppen von verschiedenen klonalen Gruppierungen untersucht: Dazu wurden 98 Serogruppe A, Subgruppe IV-1 Stämme aus der Epidemie in Gambia ab 1982 und 96 endemische Serogruppe C, ET-37 Komplex Stämme aus Mali, die zwischen 1990 und 1992 isoliert wurden, ausgewählt.

In den 96 endemischen ET-37 Komplex Stämmen wurden bis auf eine nichtsynonyme Punktmutation im *opaA*-Gen keinerlei weitere Varianten in den anderen Genen gefunden. (siehe Tabelle 3.10).

Bei den 98 epidemischen Subgruppe IV-1 Stämmen gab es hingegen vielfältige Abweichungen vom Referenzstamm Z2491. Es wurden 1 Allelvariante für *potF*, 20 Allelvarianten für *tbpB*, 8 Allelvarianten für *amiC* und 3 Allelvarianten für das *opaA*-Gen gefunden (siehe Tabelle 3.10). Aus dem Vergleich der beiden klonalen Gruppierungen kann man ableiten, daß hier ein völlig unterschiedlichen Rekombinationsverhalten in den

Ergebnisse

Serogruppe A, Subgruppe IV-1 und Serogruppe C, ET-37 Komplex vorliegt. Wodurch dies bedingt ist, ist noch nicht bekannt.

Isolationsort		Mali	Gambia
Serogruppe		C	A
Subgruppe		ET-37 Komplex	IV-1
Anzahl der identischen Stämme: ( <i>potF</i> , <i>tbpB</i> , <i>amiC</i> und <i>opaA</i> )		<b>95/96</b>	<b>54/98</b>
Anzahl der Varianten:		<b>1/96</b>	<b>44/98</b>
Allelvarianten	<i>potF</i> :	0	1
	<i>tbpB</i> :	0	20
	<i>amiC</i> :	0	8
	<i>opaA</i> :	1	3

Tabelle 3.10: Variationen in den untersuchten Gruppen aus Mali und Gambia.

In der Tabelle 3.11 sind alle Allele der bisher untersuchten Stämme aufgelistet, um einen Überblick über die gefundenen Varianten zu ermöglichen.

	<i>potF</i>	<i>tbpB</i>	<i>amiC</i>	<i>opaA</i>
Z2491 (IV-1 Std.)	2	1	1	132
<b>frühe Isolate:</b>				
Z1293 (IV-2)	1	5	1	132
Z1362 (IV-1)	1	1	1	132
Z3906 (III)	3	1	2	132
<b>Mali (IV-1):</b>				
Z4409	2	24	4	132
Z4421	1	1	1	132
Z4860	2	25	1	132
Z6132	1	26	1	132
Z6138	1	26	1	132
<b>Gambia (IV-1):</b>				
G1*	2	1	1	132
G4**	2	2	1	132
G11	2	—	5	132
G126	2	6	1	132
G135	2	1	1	137
G223	2	3	1	132
G361	2	9	1	132
G372	2	4	1	132
G396	2	6	1	132
G765	2	8	1	132

<b>noch Gambia (IV-1):</b>				
G786	<b>1</b>	<b>12</b>	<b>3</b>	132
G849	2	1	1	<b>141</b>
G852	2	<b>16</b>	<b>6</b>	132
G862	2	<b>19</b>	1	132
G867	2	<b>12</b>	1	132
G872	2	<b>18</b>	1	132
G882	2	<b>14</b>	1	132
G907	2	<b>2</b>	1	<b>136</b>
G1401	2	<b>13</b>	1	132
G1405	2	<b>8</b>	<b>7</b>	132
G1407	2	<b>7</b>	1	132
G1409	2	<b>20</b>	1	132
G1411	2	<b>21</b>	<b>8</b>	132
G1413	2	<b>22</b>	1	132
G1417	2	<b>11</b>	<b>9</b>	132
G1420	2	<b>15</b>	1	132
G1423	2	<b>10</b>	1	132
<b>Mali (ET37-Komplex):</b>				
Z4400 (Std.)	100	100	100	5200
Z4806***	100	100	100	<b>5203</b>

Tabelle 3.11: Überblick über alle gefundenen Allele.

\*: dem IV-1 Standardstamm Z2491 entsprechen die 54 Stämme G1, G7, G29, G43, G136, G137, G139, G152, G160, G169, G178, G206, G211, G216, G218, G219, G221, G222, G226, G230, G231, G265, G268, G277, G279, G359, G371, G378, G526, G561, G577, G623, G708, G719, G760, G767, G774, G790, G804, G806, G842, G857, G889, G892, G902, G912, G1217, G1219, G1221, G1223, G1226, G1403, G1415 und G1425. Abweichungen von den Standardallelen sind **fett** dargestellt.

\*\* : eine Gruppe von 19 IV-1 Stämmen mit der gleichen Allelkombination (G4, G204, G264, G322, G328, G333, G768, G776, G789, G793, G805, G881, G1210, G1211, G1212, G1213, G1214, G1215 und G1216).

\*\*\*: einzige Variante der 96 untersuchten ET37-Komplex Stämme.

Std.: Standard, -: keine Amplifikation möglich.

### 3.6 Analyse der gefundenen Allel-Varianten

Die in den IV-1 Stämmen gefundene *potF*-Variante *potF1* fand sich bereits in den frühen IV-1 und IV-2-Stämmen ebenso wie in Stämmen aus Mali von 1990-1992 wieder und entspricht damit dem ursprünglichen *potF*-Allel. Es wurden 20 neue unterschiedliche *tbpB*-Allele, 8 neue *amiC* und 4 (3 in der Subgruppe IV-1; 1 im ET-37 Komplex) neue *opaA*-Allele isoliert. Im Gegensatz zu teilweise mehrfach gefundenen *tbpB*-Allelen traten die *amiC*- und *opaA*-Allelvarianten jeweils nur einmal auf. Alle Varianten unterschieden sich in mehreren Nukleotiden voneinander. In der Abbildung 3.7 sind alle gefundenen Allele der Gene *potF*, *amiC* und *opaA* als HAPLOT dargestellt. Dabei wurden Abweichungen von der Referenzsequenz des Stammes Z2491 als senkrechter Strich dargestellt, weiße Blöcke zeigen Identität an.

## Ergebnisse

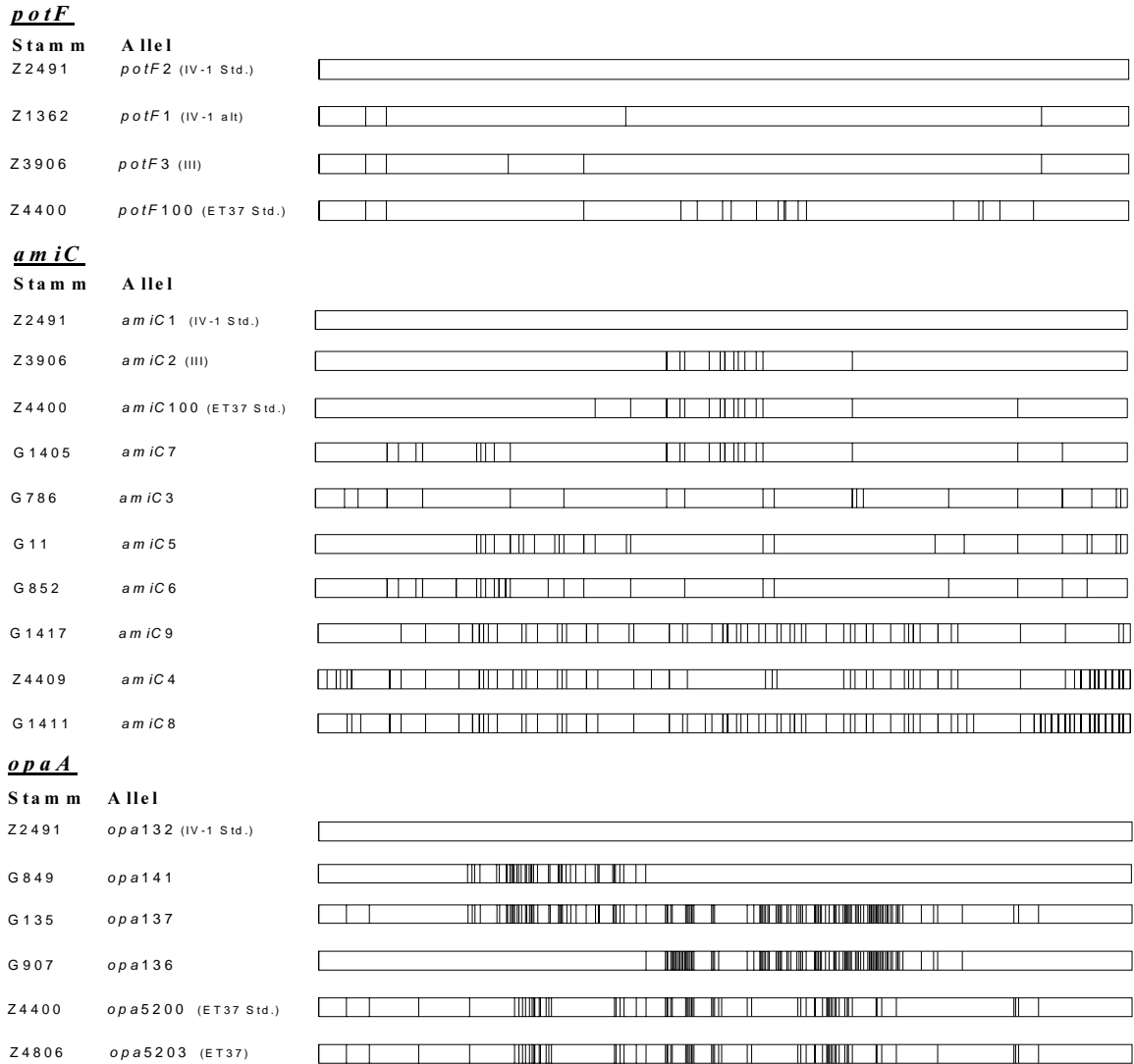


Abbildung 3.7: **HAPLOT-Darstellung aller Allelvarianten von *potF*, *amiC* und *opaA*.**

Abweichungen von Z2491 sind als senkrechter Strich dargestellt, die Stamm- und Allelbezeichnungen sind jeweils links angegeben. (Std.: Standard, ET-37: ET-37 Komplex, III: Subgruppe III).

Eine Darstellung der *thpB*-Allele als HAPLOT war nicht möglich, da hier so viele Polymorphismen auftraten, daß aus der Abbildung keinerlei Informationen mehr entnommen werden konnten. Die *thpB*-Allele werden später analysiert. Alle IV-1 Allele unterscheiden sich von denen der ET-37 Komplex Stämme. Um nun die möglichen evolutionären Verwandtschaften zwischen allen Allel-Varianten besser darstellen zu können, wurde das Program SPLITSTREE verwendet (Huson, 1998). SPLITSTREE verwendet die von H.J. Bandelt und A.W.M. Dress entwickelte Methode der "*splits decomposition*" und stellt im Gegensatz zu den "*maximum parsimony*" und "*maximum likelihood*"-Berechnungen die Ergebnisse als ein Netzwerk von möglichen verwandtschaftlichen Beziehungen dar (Bandelt and Dress, 1992).

Bei evolutionär idealen Daten ergibt sich wie bei den anderen Methoden ein klarer Baum, jedoch wird bei nicht idealen Daten ein Netzwerk als ein Hinweis auf eine unklare und nicht eindeutige Phylogenie dargestellt. Da SPLITSTREE diese Daten nicht in einen evolutionären Baum "zwingt", ist die Darstellung ein guter Indikator für die phylogenetischen Beziehungen der Ausgangsdaten. Bei einer ausgeprägten Netzdarstellung sind Rekombinationsereignisse wesentlich wahrscheinlicher als sukzessive Mutationen, die einen Baum erzeugen würden.

In den folgenden Darstellungen sind die Stämme Z2491 und Z4400 jeweils als IV-1 Standard und als ET-37 Komplex Standard farblich dargestellt, um die Abweichungen von ihnen besser kenntlich zu machen.

In der Abbildung 3.8 sind die *potF*-Allele dargestellt.



Abbildung 3.8: **SPLITSTREE-Darstellung der *potF*-Allele.**

Die genetische Distanz ist unter der Graphik angegeben. Die Allele sind zusätzlich mit Stammnummern bezeichnet, um einen Vergleich mit den anderen Abbildungen zu ermöglichen.

Hier ist die hohe Verwandtschaft der Stämme Z2491, G786 und Z3906 zu sehen, wobei der phylogenetische Entstehungsweg dieser Allele nicht eindeutig ist. Das ET-37 Komplex Allel ist wesentlich weiter von den Subgruppe IV-1 Stämmen entfernt als die IV-1 Varianten voneinander.

In der Abbildung 3.9 sind die Beziehungen der 10 gefundenen *amiC*-Allele dargestellt. Die *amiC*-Allele der IV-1 und ET-37 Komplex Stämme mit den Varianten Z3906 und G1405 können durch Rekombinationen untereinander entstanden sein, während die Allele G1411 und G1417 von einem gemeinsamen Vorgänger abstammen.

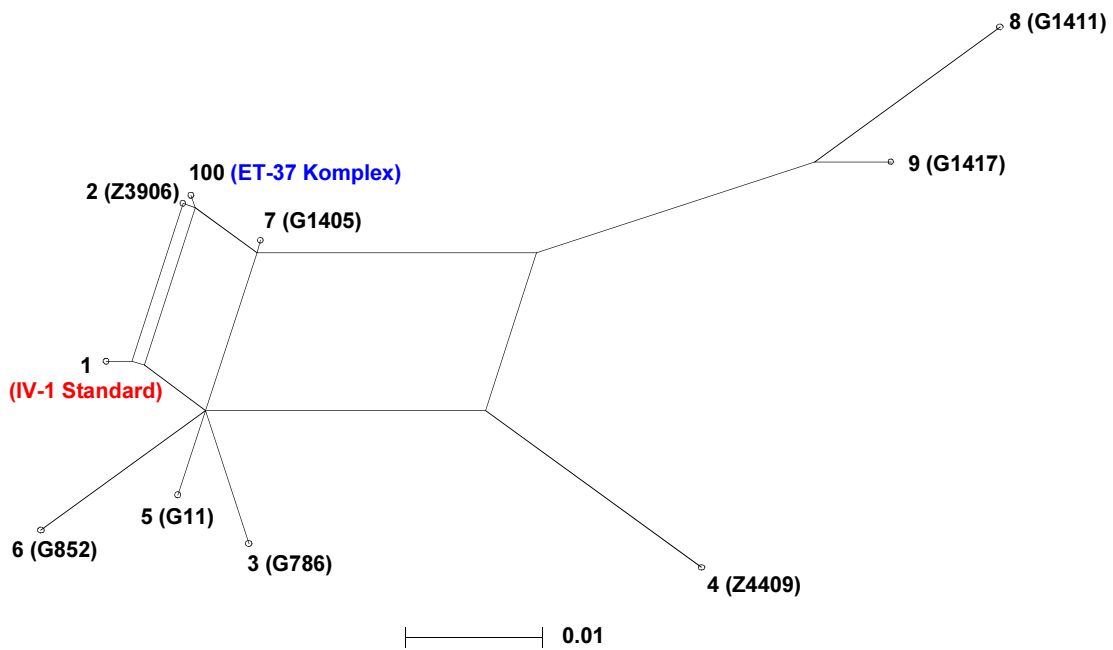


Abbildung 3.9: SPLITSTREE-Darstellung der gefundenen *amiC*-Allele.

Die genetische Distanz ist unter der Graphik angegeben. Die Allele sind zusätzlich mit Stammnummern bezeichnet, um einen Vergleich mit den anderen Abbildungen zu ermöglichen.

Nach einer Analyse von Hobbs *et al.* wurden in der Subgruppe IV-1 drei verschiedene *opa* Loci identifiziert, die als *opaA*, *opaB* und *opaD* bezeichnet wurden. (Hobbs *et al.*, 1994). Ein vierter Locus (*opaJ*) wurde in der Subgruppe III entdeckt (Morelli *et al.*, 1997). Bei den hier untersuchten *opaA*-Allelen war in 2 Stämmen eine Translokation von an anderer Stelle im Genom der Subgruppe IV-1 auftretender *opa*-Gene (in G907 *opa136* aus dem *opaD*-Locus und in G135 *opa137* aus dem *opaB*-Locus) in den *opaA*-Locus zu beobachten. Bei dem Stamm G849 hat die HV1-Region eines *opa137*-Gens die HV1-Region des *opa132*-Allels bei einer Rekombination ersetzt und so ein neues *opa*-Allel (*opa141*) gebildet. Die Variante des ET-37 Komplex *opaA*-Allels (Z4806) unterscheidet sich nur in einer Punktmutation von dem ursprünglichen Allel. Aus der Abbildung 3.10 ist eindeutig abzulesen, daß die *opaA*-Varianten aus verschiedenen Rekombinationsereignissen entstanden sind, da auch hier keinerlei Baumstruktur auftritt.

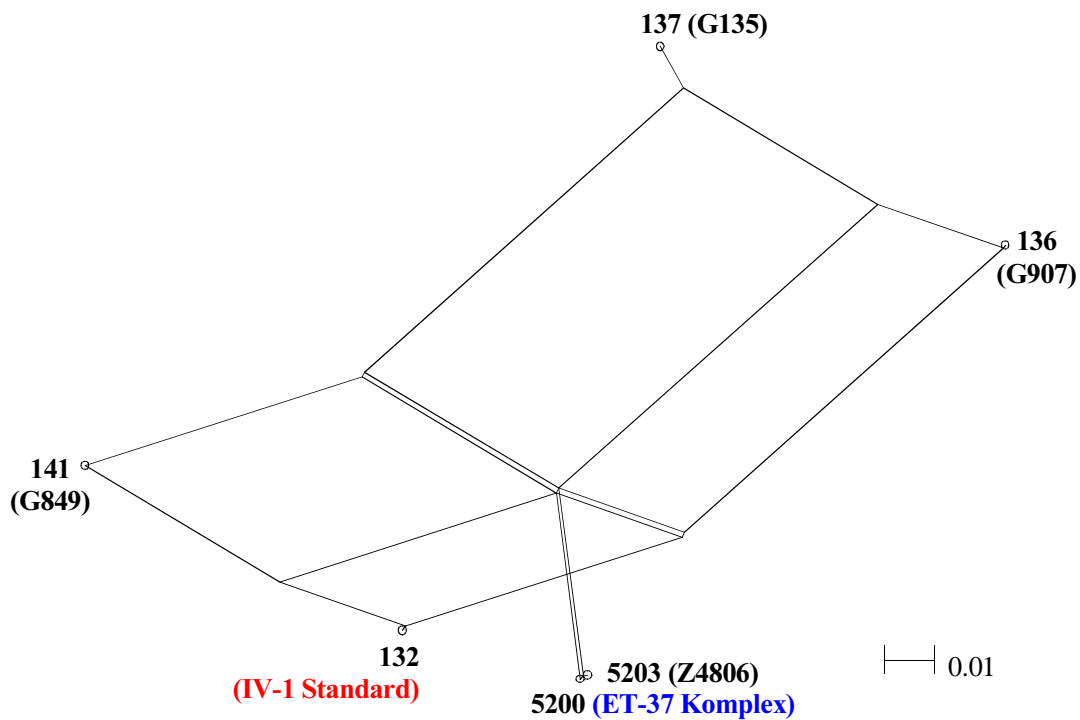


Abbildung 3.10: **SPLITSTREE-Darstellung der *opaA*-Allele.**

Die genetische Distanz ist unter der Graphik angegeben. Die Allele sind zusätzlich mit Stammnummern bezeichnet, um einen Vergleich mit den anderen Abbildungen zu ermöglichen.

Aufgrund ihrer hohen Sequenz-Diversität konnten die *tbpB*-Allele nur in einer SPLITSTREE-Graphik (Abbildung 3.11) dargestellt werden. Bei der Analyse der 98 Subgruppe IV-1 Stämme waren 20 verschiedene Allele isoliert worden. Eine Gruppe von 20 Stämmen zeigte eine gemeinsame *tbpB*-Allelvariante, die nur mit einem Nukleotid von der Referenzsequenz abwich (vgl. Tabelle 3.11). Der Stamm G4 dient als Referenzstamm für diese Gruppe. Zwei weitere Stämme, G223 und G372, unterschieden sich ebenfalls nur um ein Nukleotid an jeweils verschiedenen Positionen von dem Standardstamm Z2491. Die übrigen neuen Allele besaßen mehrere Polymorphismen innerhalb des untersuchten Bereiches.



## Ergebnisse

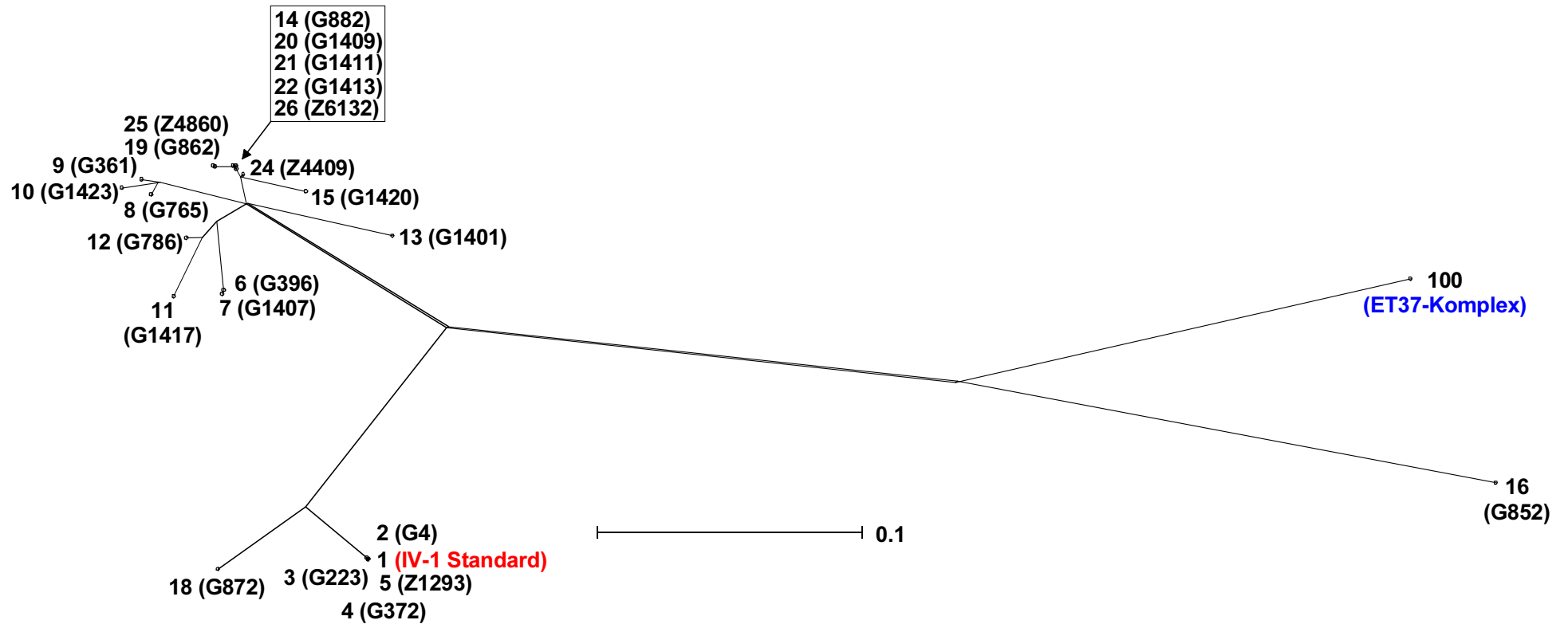


Abbildung 3.11: **SPLITSTREE-Darstellung aller gefundenen *thpB*-Allele.**

Die Allele sind zusätzlich mit den Stammnummern angegeben, um einen Bezug mit den anderen Abbildungen zu ermöglichen. Die Skala für die genetische Distanz ist unter dem Baumdiagramm dargestellt, wobei der Maßstab um den Faktor 10 größer als bei den anderen Graphiken ist.

In dem ermittelten Baum sind klar drei Gruppen von *tbpB*-Allelen auszumachen, die alle einen relativ großen genetischen Abstand voneinander aufweisen. Dies sind einmal die Isolate Z1293, G4, G223, G372 und G872, die höchstwahrscheinlich durch Mutationen des Ursprungsallels Z2491 entstanden sind, da sie nur mit wenigen Nukleotiden von dieser Sequenz abweichen. Die zweite Gruppe bilden der IV-1 Stamm G852 und der ET-37-Komplex Standard. Jedoch ist der genetische Abstand der beiden Stämme zu groß, um durch einen Import eines ET37-Komplex *tbpB*-Allels in den IV-1 Stamm G852 erklärt werden zu können. Diese beiden Stämme liegen genetisch weiter auseinander als die erste und dritte Gruppe der restlichen Stämme zueinander. Diese dritte Gruppe scheint aus unverwandten Bakterien in die *N. meningitidis* importiert worden zu sein und zeigt eine hohen Sequenzvariabilität. Die Phylogenie dieser dritten *tbpB*-Allelgruppe ist separat in einem weiteren SPLITSTREE-Diagramm dargestellt.

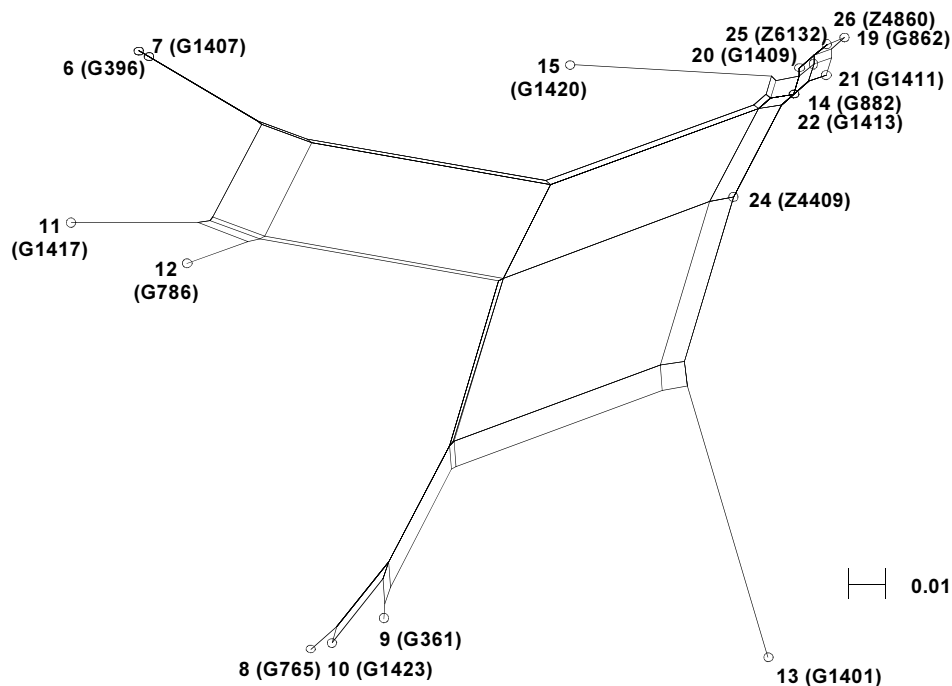


Abbildung 3.12: **SPLITSTREE-Darstellung der dritten Gruppe von *tbpB*-Allelen.**

Die Allele sind zusätzlich mit den Stammnummern angegeben, um einen Bezug mit den anderen Abbildungen zu ermöglichen. Die Skala für die genetische Distanz ist unter dem Baumdiagramm dargestellt.

In der Abbildung 3.12 zeigen sich komplexe Verbindungen zwischen den einzelnen Allelen. Auch hier ist keine Baumstruktur zu erkennen, was nur durch vielfältige Rekombinationen der Allele miteinander erklärbar ist. Ob alle gefundenen Allelvarianten einzeln aus fremden Bakterien importiert wurden oder erst in *N. meningitidis* entstanden sind, ist ungeklärt.

### 3.7 Restriktionsanalyse der gefundenen *tbpB*-Varianten in der Subgruppe IV-1

Es wurden 20 der gefundenen Allelvarianten ausgewählt, um die Größen der importierten Fragmente zu bestimmen. Bei der Sequenzanalyse wurden 3 *tbpB*-Allel-Paare (G126/G396, G786/G867 und G765/G1405) gefunden; 2 Paare besaßen aber ein unterschiedliches *amiC*-Allel (G786/G867 und G765/G1405). Um die gefundenen Varianten zu analysieren, wurden von den Stämmen sowie von der Referenz Z2491 vier *lrPCR*-Produkte (wie für die Sequenzierung) amplifiziert und aufgereinigt (siehe Abbildung 3.1, Seite 62). Der Stamm G11 wurde zusätzlich für die Restriktionsanalyse ausgewählt, da aus ihm zuvor kein *tbpB* amplifiziert werden konnte. Es folgte ein Verdau mit je drei geeigneten Restriktionsenzymen pro *lrPCR*-Fragment, die es ermöglichten, einen Größenvergleich mit Restriktionsfragmenten aus dem Referenzstamm Z2491 durchzuführen (Abbildung 3.13). Die Enzyme wurden mit Hilfe des Programmes DNAMAN4.0 ausgewählt, Dieses Programm bietet die Möglichkeit, *in silico* Test-Restriktionen auf symbolisierten Gelen darzustellen. Damit können geeignete Enzyme ausgewählt werden, um gut voneinander getrennt laufende Banden zu erhalten.

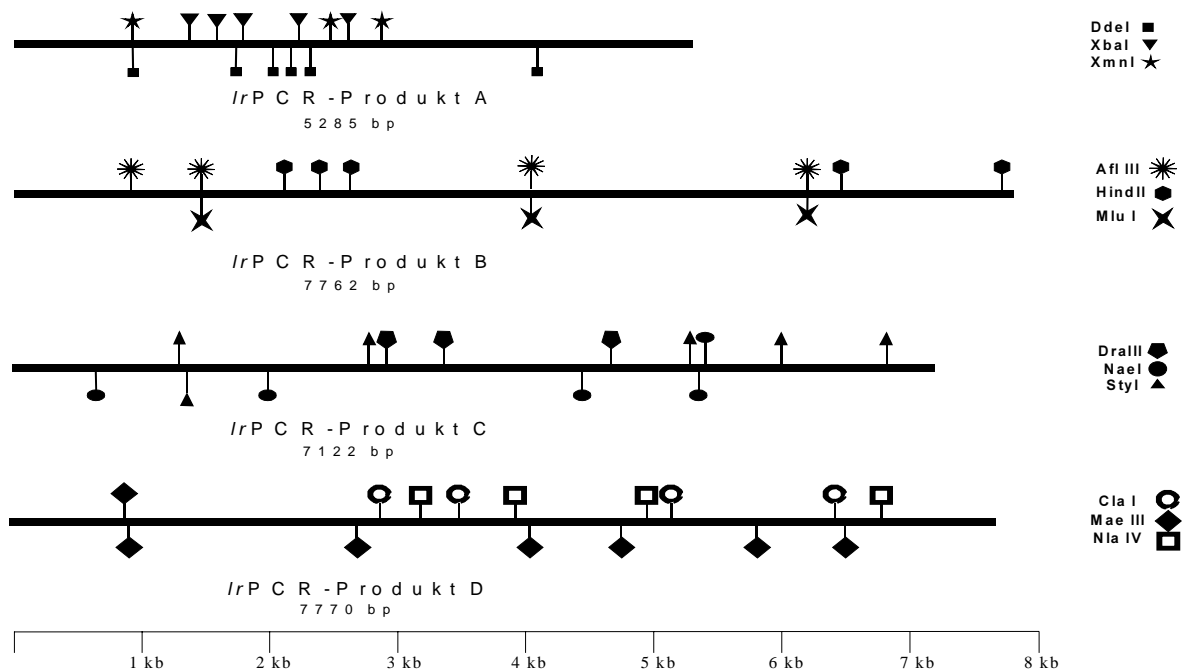


Abbildung 3.13: Enzymschnittstellen auf den *lrPCR*-Fragmenten für die Restriktionsanalyse. Angezeigt sind die Enzym-Schnittstellen in der Sequenz des Referenzstammes Z2491.

Nach dem Verdau der *lrPCR*-Produkte und der Auftrennung auf 1,2%-Agarose-Gelen wurden die Fragmente mit abweichenden Größen zum Standardstamm bestimmt (Abbildung 3.14).

## Ergebnisse

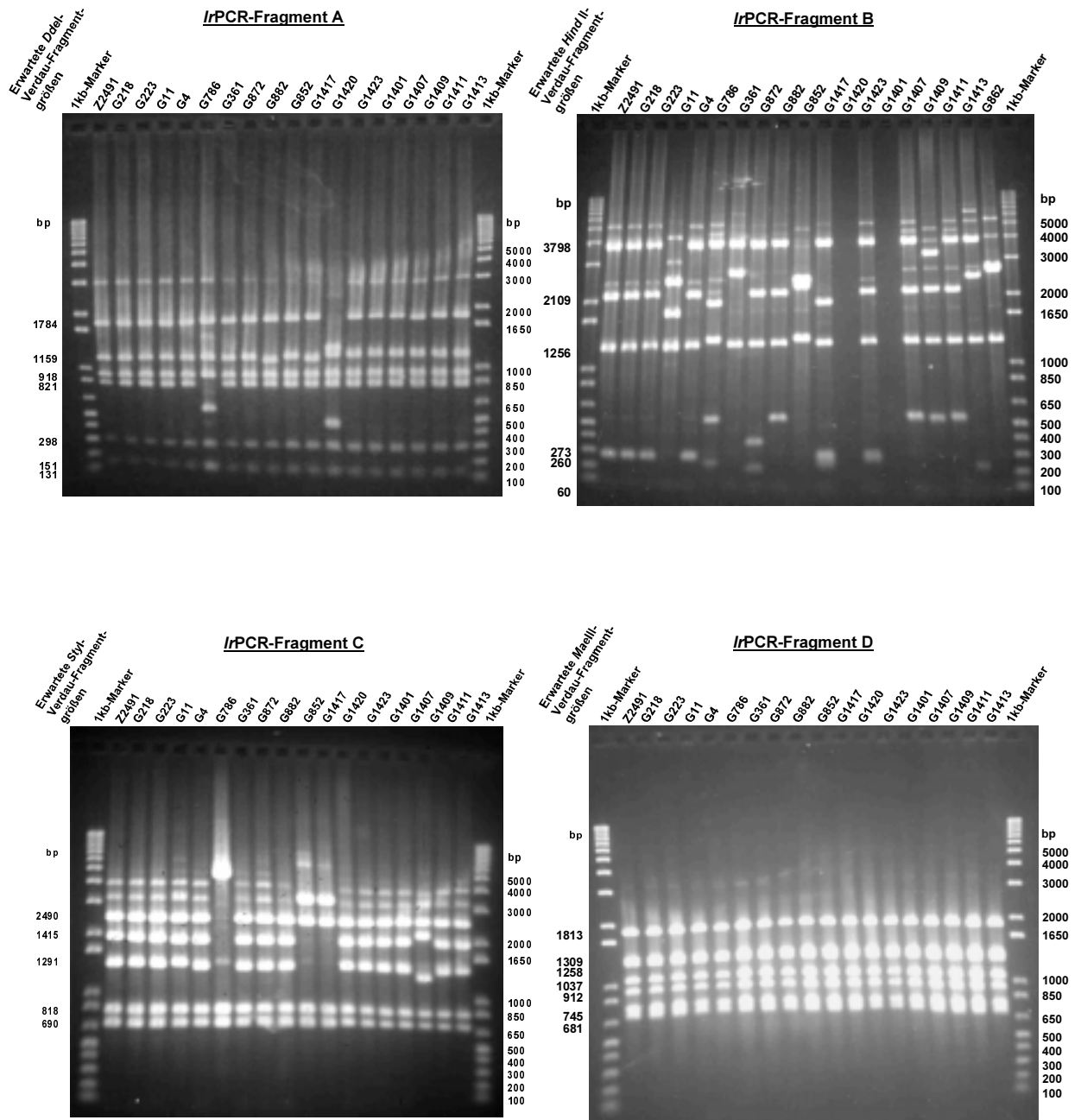


Abbildung 3.14: Verdau der lrPCR-Fragmente (je ein Enzym pro Fragment als Beispiel).

Die aus dem Verdau von Z2491 erwarteten Fragmentgrößen sind zusammen mit dem verwendeten Enzym jeweils links vom Gelbild angegeben (siehe Abbildung 3.13), auf der rechten Seite ist die 1 kb-Leiter als Größenstandard dargestellt. Bei dem lrPCR-Produkt B haben die Stämme G1420 und G1401 Sequenzvariationen in den Primerbindungsstellen, so daß hier keine PCR-Produkte amplifiziert werden konnten. Die Größe der Rekombinationsereignisse in diesen Stämmen wurde über die benachbarten lrPCR-Produkte A und C bestimmt.

Bei dieser Art der Analyse sind nur

- a) Sequenzvariationen in den Schnittstellenbereichen der drei Enzyme und/oder
- b) Längenvariationen zwischen zwei gleichen Schnittstellen detektierbar.

Diese Analyse wurde für alle Varianten mit den vier *lr*PCR-Fragmenten durchgeführt. Da die Positionen der Enzym-Schnittstellen im Referenzstamm bekannt waren, konnte nun nach der Zuordnung der Fragmente zu der bekannten Sequenz die Größe der Rekombinationsereignisse in den Varianten abgeschätzt werden. In der Abbildung 3.15 wurde das Ende des letzten abweichenden Fragments (Enzymschnittstelle) als Grenze der Rekombination angenommen.

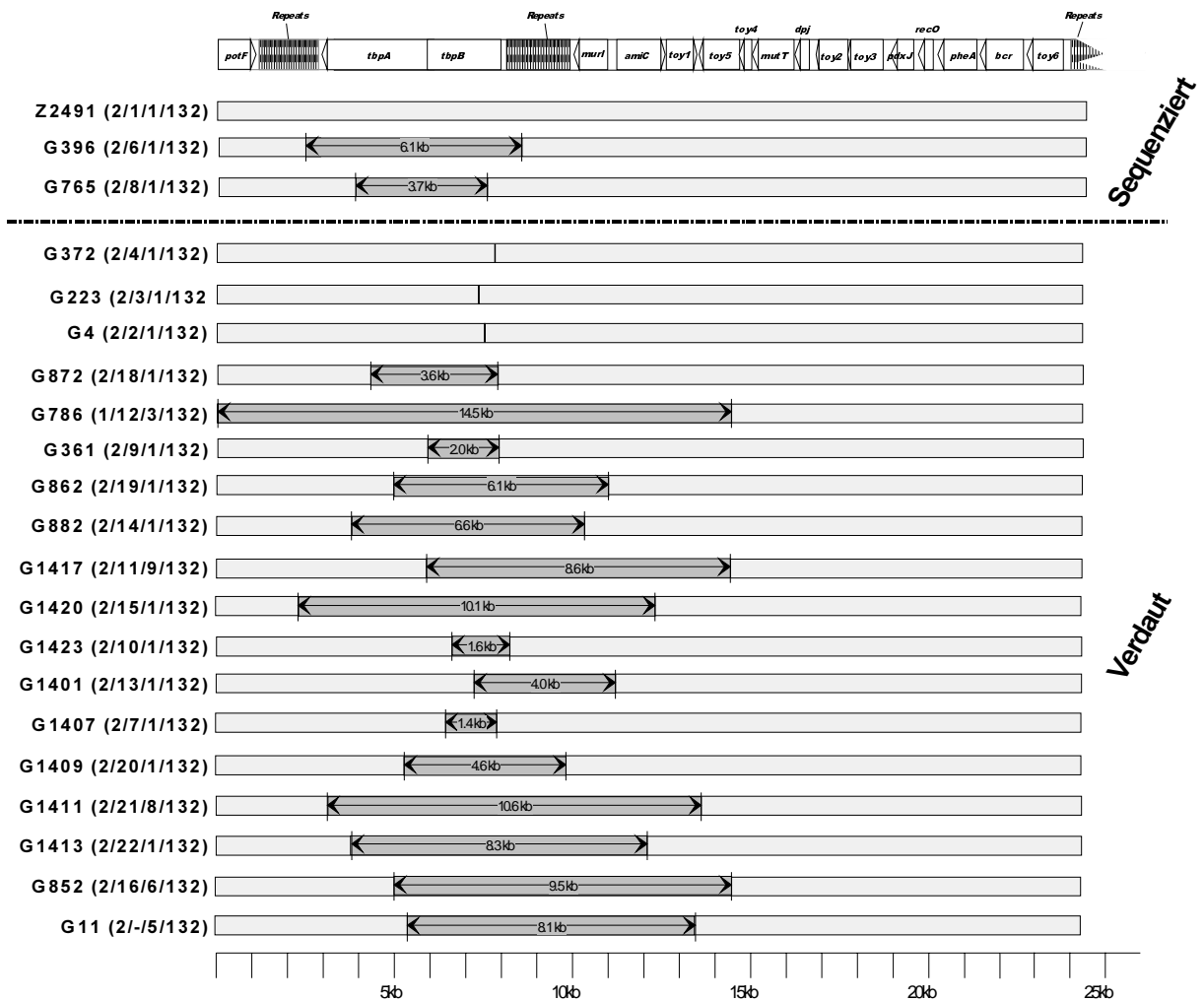


Abbildung 3.15: **Größe der importierten Fragmente.**

Die Längen der Rekombinationsereignisse sind angegeben. Einzelne Nukleotid-Abweichungen sind als senkrechte Striche dargestellt, die Stämme mit ihren Allelkombinationen (*potF/ttpB/amnC/opaA*) sind links bezeichnet. Die komplett sequenzierte Stämme (siehe Kapitel 3.8) sind von denen durch die Restriktionsanalyse untersuchten Stämme getrennt dargestellt.

Die importierten DNA-Fragmente besitzen bei einer gefundenen minimalen Länge von ca. 1,5 kb (z.B. G1407) bis maximal 14,5 kb (z.B. G786) eine mittlere Größe von 6,5 kb. Es ist eine starke Häufung der Rekombinationsereignisse um das *tbpAB*-Operon zu beobachten. Auffällig ist, daß kaum DNA-Import/Mutationen aufwärts des *toy5*-ORFs gefunden wurden. Entweder ist die Sequenzvariabilität in diesem Genomabschnitt so gering, daß Rekombinationen aufgrund fehlender informativer Nukleotide nicht sichtbar sind oder es gibt für den Bereich um *tbpAB* einen Mechanismus, der eine erhöhte Rekombinationsrate unterstützt. Somit scheint die Häufung der Rekombinationen im *tbpAB*-Operon auf ein funktionelles Prinzip hinzuweisen (siehe Diskussion).

### 3.8 Vergleichende Sequenzierung des Bereichs *potF-toy6* von 2 Varianten

Es wurden aus den gefundenen Allelvarianten 2 Stämme ausgewählt, die einer vergleichenden Sequenzierung in dem kompletten Bereich von *potF* bis *opaA* unterzogen wurden, um die genaue Rekombinationsendpunkte der importierten DNA-Fragmente bestimmen zu können (siehe Abbildung 3.15). Es handelte sich um die Serogruppen A, Subgruppe IV-1 Stämme G396 und G765. Es zeigte sich, daß diese beiden Varianten fast identisch mit dem Standardstamm waren und nur im Bereich um *tbpAB* Abweichungen zeigten. Die Endpunkte dieser Rekombinationsereignisse wurden auf das letzte von dem Referenzstamm abweichende Nukleotid gelegt. Wie aus der Abbildung 3.15 hervorgeht, sind die Rekombinationsgrenzen der sequenzierten *tbpAB*-Varianten an unterschiedlichen Stellen zu finden. So ist bei dem Stamm G765 ein nur 3,7 kb großer DNA-Block innerhalb des *tbpAB*-Operons ausgetauscht, während die Rekombinationendpunkte im Stamm G396 in den angrenzenden *Repeats* liegen. Hier wurde ein 6,1 kb großer Block mit den Rekombinationsenden in beiden *Repeats* integriert. Es wurde eine vergleichende Analyse der *Repeat*-Regionen um *tbpAB* durchgeführt, um Variationen in diesen Elementen aufzuzeigen und mögliche Funktionen bei Rekombinationen zu verstehen. Bei dem direkten Vergleich der vollständigen *Repeats* der Stämme Z2491, G396 und G765 wurden alle Abweichungen berücksichtigt, nur 100% identische Bereiche zu Z2491 sind in den Rekombinaten als weiße Blöcke dargestellt (siehe Abbildung 3.16). Bei der Rekombination der neuen *tbpAB*-Allele in den Stamm G396 wurde auch ein Abschnitt eines *tbpB-Repeat*s mit übertragen, was zu einem Austausch der sechs *NIMEs* Nr. 21-26 mit acht neuen *NIMEs* und somit einer Verlängerung des *tbpB-Repeat*s um 200 bp führte. Im *tbpA-Repeat* sind von der Rekombination nur die beiden unmittelbar neben *tbpA* liegenden *NIMEs* Nr. 19 und 20 betroffen worden, ohne daß es zu gravierenden Insertionen gekommen ist.

Ergebnisse

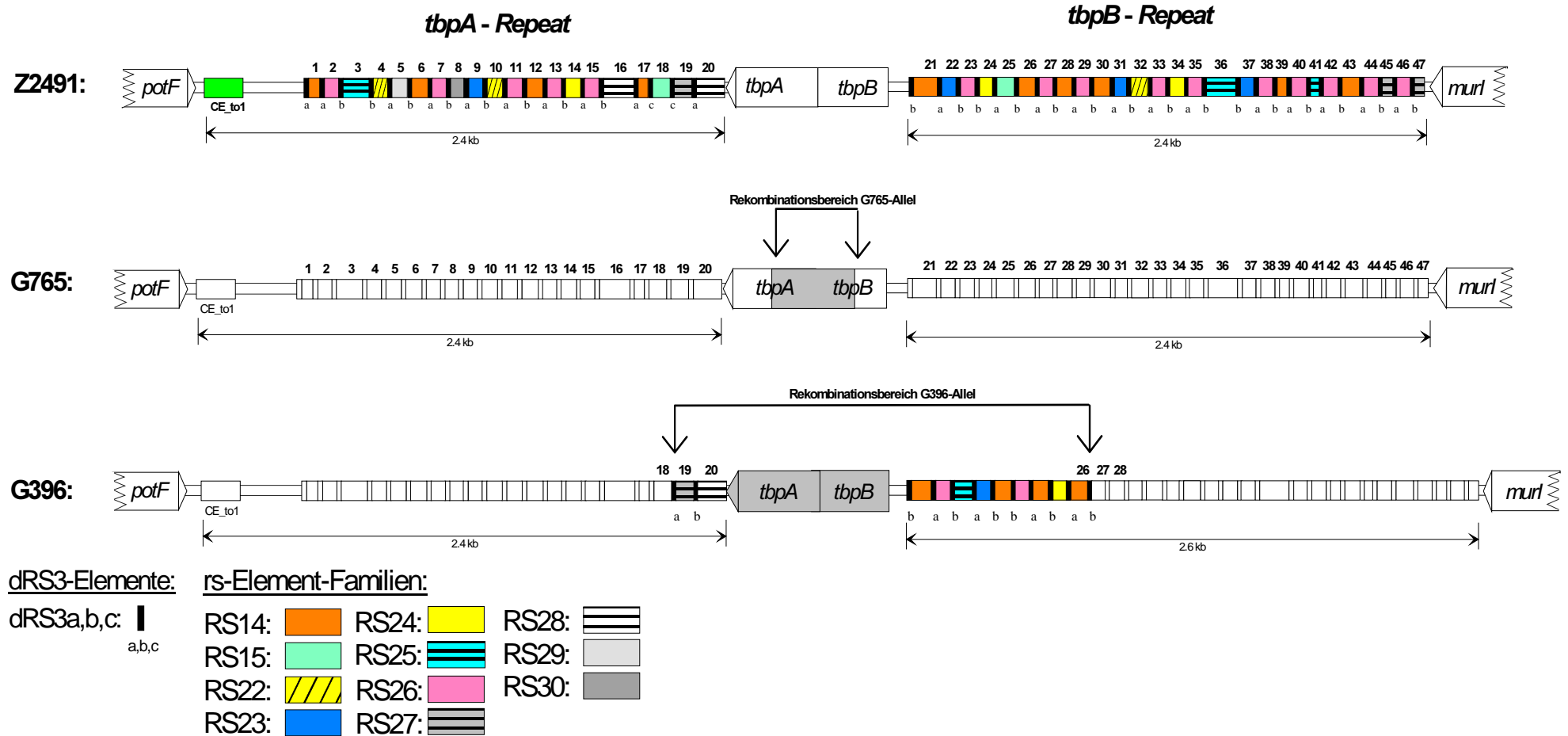


Abbildung 3.16: Aufbau der Repeats um *tbpAB* in den rekombinanten Stämmen G765 und G396.

Alle repetitiven Elemente sind als Kästen dargestellt, nichtrepetitive Sequenzabschnitte als graue Linie. Die abweichenden rs-Elemente sind als farbige Blöcke eingezeichnet, die Numerierung bezieht sich auf die Abfolge der rs-Elemente in Z2491 (jeweils von 1-20 im *tbpA*-Repeat und von 21-47 im *tbpB*-Repeat), um Bezugspunkte für einen Vergleich zu haben. Die Familienzugehörigkeit der rs-Elemente ist aus der Farbkodierung der Legende abzulesen. Zwischen rs-Elementen gleicher Bezeichnung können Polymorphismen von bis zu 20% auf Nukleotidebene auftreten. Die Gene *potF*, *tbpAB* und *murI* sind nicht maßstabsgerecht gezeichnet. Das Correia-Element ist extra bezeichnet. Alle Abweichungen zu der Z2491-Sequenz sind als farbige (rs-Elemente) oder dunkelgraue (Gene) Blöcke dargestellt. Identische Bereiche zu Z2491 sind als farblose Blöcke gezeichnet.

In den beiden *NIMEs* sind diverse Sequenzpolymorphismen gefunden worden, jedoch sind die beiden Elemente noch den RS27 und RS28 Familien zuzuordnen. Die Rekombinationsendpunkte des Stammes G396 wurden somit in das RS-Element mit der Nr. 19 auf der *tbpA*-Seite und in das RS-Element mit der Nr.26 auf der *tbpB*-Seite gelegt.

### **3.9 Suche nach potentiellen Donoren der neuen Allele in der Subgruppe IV-1**

Nachdem in den 100 Stämmen aus Gambia 20 verschiedene neue *tbpB*-Allele gefunden wurden, stellte sich die Frage, woher diese Allele kommen könnten. Es wurden 11 *N. meningitidis* Stämme der Serogruppen B, C, W-135, X, und Y, 9 nicht gruppierbare *N. meningitidis* Stämme, 8 *N. lactamica* und ein *N. sicca* Stamm ausgewählt, die alle in Gambia während der Epidemie isoliert wurden. Aus allen Stämmen wurden die *potF*, *tbpB*, *amiC* und *opaA* Allele bestimmt (siehe Tabelle 2.13).

In diesen 29 untersuchten Stämmen zeigten alle Allele starke Abweichungen zu den bisher gefundenen Sequenzen, so daß keiner dieser zusätzlich untersuchten Stämme als Donor für die gefundenen neuen Allele in der Subgruppe IV-1 in Frage kam (Daten nicht gezeigt).