

Aus dem Universitätsklinikum Münster

Institut für Hygiene

– Direktor: Univ.-Prof. Dr. rer. nat. Dr. h. c. Helge Karch –

Single Nucleotide Polymorphism (SNP)-Analyse im Kerngenom der HUSEC-
Kollektion – Ein phylogenetisches Porträt

INAUGURAL - DISSERTATION

zur

Erlangung des doctor medicinae

der Medizinischen Fakultät

der Westfälischen Wilhelms-Universität Münster

vorgelegt von: Meyer, Gregor Jodokus

aus Gütersloh

2013

Gedruckt mit freundlicher Genehmigung der Medizinischen Fakultät der
Westfälischen Wilhelms-Universität Münster.

Dekan: Univ.-Prof. Dr. med. Dr. h. c. Wilhelm Schmitz

1. Berichterstatter: Priv.-Doz. Dr. med. Alexander Mellmann

2. Berichterstatter: Univ.-Prof. Dr. med. Dag Harmsen

Tag der mündlichen Prüfung: 20.11.2013

Aus dem Universitätsklinikum Münster

Institut für Hygiene

– Direktor: Univ.-Prof. Dr. rer. nat. Dr. h. c. H. Karch –

Referent: Priv.-Doz. Dr. med. Alexander Mellmann

Koreferent: Univ.-Prof. Dr. med. Dag Harmsen

ZUSAMMENFASSUNG

Single Nucleotide Polymorphism (SNP)-Analyse im Kerngenom der HUSEC-Kollektion – Ein phylogenetisches Porträt

Gregor Jodokus Meyer

In der vorliegenden Arbeit wurde die phylogenetische Entwicklung und genetische Struktur der 42 Stämme umfassenden HUSEC-Kollektion anhand von 90 Zielgenen untersucht, die zufällig über das Genom verteilt sind und aus dem Kerngenom von EHEC O157:H7/H⁻ stammen. Hierbei wurden jeweils interne Genabschnitt mit einer Länge von 492 bis 618 bp untersucht. Interessanterweise konnte nur 70 der 90 Gene in allen Stämmen nachgewiesen werden, was schon einen ersten deutlichen Hinweis auf die Diversität der HUSEC-Kollektion gab. In den ubiquitär nachweisbaren ORFs konnten an 3034 Positionen SNPs detektiert werden, was ebenfalls die hohe genomische Diversität unterstreicht. Es wurden in allen Zielgenen Varianten detektiert, eine signifikant höhere Mutationsrate als im arithmetischen Mittel wurde allerdings in keinem ORF festgestellt. Für die wichtigsten Serogruppen (O157, O26 und O103) konnten serogruppenspezifische Varianten nachgewiesen werden. Zur weiteren Analyse wurden auf Basis der SNPs Phylogramme nach neighbour-joining- sowie minimum-spanning-Verfahren erstellt und mit den etablierten Methoden MLST und Serotypisierung verglichen. Mittels MLST können 30 unterschiedliche Sequenztypen bestimmt werden, der SNP-basierte Ansatz differenziert 5 weitere, also insgesamt 35 Stämme. Auch können einige Aufzweigungen der phylogenetischen Bäume mit höheren „bootrap-consensus“-Werten untermauert bzw. korrigiert werden. Es zeigt sich, dass in beiden Phylogrammen die Anordnung der Stämme weitestgehend ähnlich ist, also auch mit dem deutlich kleineren Ansatz der MLST (7 Gene) bereits eine gute Abbildung der Verwandtschaftsverhältnisse erreicht wurde. Lediglich HUSEC032 gruppiert wesentlich anders.

Tag der mündlichen Prüfung: 20.11.2013

ERKLÄRUNG

Ich gebe hiermit die Erklärung ab, dass ich die Dissertation mit dem Titel:

Single Nucleotide Polymorphism (SNP)-Analyse im Kerngenom der HUSEC-Kollektion –
Ein phylogenetisches Porträt

in der/im (Klinik, Institut, Krankenanstalt):

Institut für Hygiene

unter der Anleitung von:

PD Dr. med. Alexander Mellmann

1. selbständig angefertigt,
2. nur unter Benutzung der im Literaturverzeichnis angegebenen Arbeiten angefertigt und sonst kein anderes gedrucktes oder ungedrucktes Material verwendet,
3. keine unerlaubte fremde Hilfe in Anspruch genommen,
4. sie weder in der gegenwärtigen noch in einer anderen Fassung einer in- oder ausländischen Fakultät als Dissertation, Semesterarbeit, Prüfungsarbeit, oder zur Erlangung eines akademischen Grades, vorgelegt habe.

Rietberg, den 19.12.2013

Name: Gregor Jodokus Meyer

Unterschrift

1	Einleitung.....	1
1.1	Enterohämorrhagische <i>Escherichia coli</i>	1
1.2	Klinik und Epidemiologie von EHEC.....	1
1.3	Die HUSEC-Kollektion	2
1.4	Phylogenie von EHEC.....	3
1.5	Zielsetzung.....	4
2	Material und Methoden.....	5
2.1	Untersuchungsgut und Materialien	5
2.1.1	Chemikalien und Enzyme	5
2.1.2	Bakterienstämme	5
2.2	Methoden.....	7
2.2.1	Polymerase-Chain-Reaction – Amplifikation.....	7
2.2.2	PCR-Produktaufreinigung mit EXOSAP	7
2.2.3	Sequenzierungs-PCR.....	8
2.2.4	Aufreinigung des Sequenzierungsproduktes	8
2.2.5	Sequenzauswertung	9
2.2.6	Statistische Analyse.....	9
2.2.7	Software	10
3	Ergebnisse	11
4	Diskussion.....	19
5	Literaturverzeichnis.....	24
6	Lebenslauf	30
7	Danksagung	31
8	Anhang	I

1 Einleitung

1.1 Enterohämorrhagische *Escherichia coli*

Das zur Gattung der *Enterobacteriaceae* gehörende, gramnegative und peritrich begeißelte Bakterium *E. coli* wurde von Theodor Escherich 1885 als erstes spezifisches Darmbakterium beschrieben. *E. coli* ist als apathogener Kommensale regelmäßig im Darm von Warmblütern und Vögeln zu finden (4, 5, 28, 29, 42, 44). Neben diesen harmlosen Stämmen existieren aber auch pathogene Subgruppen, die sich im Verlauf der Evolution durch akquirieren verschiedener genetischer Elemente zu obligat pathogenen *E. coli* weiterentwickelt haben. Das durch diese Keime hervorgerufene Krankheitsspektrum umfasst u.a. Sepsen, Meningitiden, Harnwegsinfektionen und Durchfälle (3, 8, 21, 28, 47). Sie lassen sich an Hand des von ihnen hervorgerufenen klinischen Bildes, von Pathomechanismen und Pathogenitätsfaktoren in sieben Untergruppen unterteilen: Enteropathogene *E. coli* (EPEC), enteroinvasive *E. coli* (EIEC), enterotoxigene *E. coli* (ETEC), enteroaggregative *E. coli* (EAggEC), Diarrhö-assozierte hämolytische *E. coli* (DHEC) und enterohämorrhagische *E. coli* (EHEC) (7). Letztere sind als Subgruppe der Shiga Toxin (Stx)-produzierenden *E. coli* (STEC) definiert und sind in der Lage, beim Menschen schwere Krankheitsbilder wie z. B. blutige Diarrhö oder das hämolytisch-urämische Syndrom (HUS) hervorzurufen (29, 30, 44). Ihr natürliches Reservoir sind vornehmlich landwirtschaftliche Nutztiere, aber auch Wildtiere. Beide sind meist symptomlose Träger der Erreger (30).

1.2 Klinik und Epidemiologie von EHEC

Die Übertragung auf Menschen erfolgt entweder über direkten Kontakt - sowohl von Tier zu Mensch als auch von Mensch zu Mensch - oder indirekt durch kontaminierte Lebensmittel und Wasser (25, 26, 43, 77). Nachdem die säuretoleranten Erreger den Darmtrakt erreicht haben, gelangen die von Ihnen produzierten Stx in den Kreislauf und schädigen vaskuläre Endothelzellen. Des Weiteren werden auch non-Stx-Virulenzfaktoren, wie LEE-kodierende Proteine, EHEC-Hämolysin, cytolethal – distending – protein (CDT), Eisenaufnahmeproteine und andere sezerniert (30).

Eine Infektion kann asymptomatisch verlaufen. Häufig kommt es jedoch nach einer Inkubationszeit von 2 bis 12 Tagen zu nichtblutigem Durchfall mit Bauchschmerzen und Krämpfen. Typischerweise wird dieser nach etwa 1 bis 3 Tagen als Manifestation einer

hämorrhagischen Kolitis blutig, was die Patienten veranlasst, medizinische Hilfe in Anspruch zu nehmen (30, 73). Als mögliche systemische Komplikation kann ein hämolytisch-urämisches Syndrom (HUS) auftreten, gekennzeichnet durch die Trias von einem Hämatokrit unter 30 % mit Hinweis auf Hämolyse im Blutausstrich, einer Thrombozytenzahl unter $150 \times 10^9 / L$ und einem erhöhtem Serum-Creatinin (73). HUS gilt momentan als eine der häufigsten Ursachen für akutes Nierenversagen und Dialysepflichtigkeit im Kindesalter (30, 63). Auch wenn HUS in den Industrieländern in nur 5 % tödlich verläuft (30, 73), so tragen doch bis zu 30 % der Betroffenen Langzeitschäden im Sinne von chronischen renalen und zerebralen Funktionseinschränkungen davon (17, 30, 73). Von therapeutischer Seite aus verfolgt man einen symptomatischen Ansatz. Neben intravenöser (Re-)Hydratation sind auch intensivmedizinische Maßnahmen wie Hämofiltration und Dialyse sowie, falls nötig, Transfusionen von Erythrozyten- bzw. Thrombozytenkonzentraten indiziert. Nach derzeitigem Stand des Wissens wird eine antibiotische Therapie wegen der möglichen massiven Freisetzung von bakteriellen Toxinen nicht empfohlen (63, 73, 82). Auch wenn die Infektionen in der Regel sporadisch auftreten (73), so kommt es doch auch immer wieder zu größeren Ausbrüchen wie 1996 im Japanischen Sakai (16) oder 2006 in mehreren Staaten der USA (41). Der häufigste und am besten untersuchte, aber bei weitem nicht der einzige Serovar ist O157:H7 (14, 42-44, 53, 58).

1.3 Die HUSEC-Kollektion

Um einen Überblick über die Diversität der mit HUS assoziierten EHEC zu erhalten, wurde am Institut für Hygiene in Münster eine Stammsammlung aller mit HUS-assoziierten EHEC-Varianten, die HUSEC-Kollektion, zusammengestellt. Hierzu wurden im Zeitraum von 1996 bis 2006 von 524 - nicht zueinander in epidemiologischer Beziehung stehenden - Patienten mit HUS ebenso viele EHEC als einziger pathogener Keim aus Stuhlproben isoliert. Die anschließende Serotypisierung klassifizierte 355 als dem Serotyp O157:H7/H⁻ zugehörig. Die übrigen 169 Isolate gehörten 34 unterschiedlichen non-O157 Serotypen an. Da die Phylogenie der O157 Stämme bereits gut untersucht war, wurden alle 169 non-O157 sowie nur 10 O157 Stämme als Kontrolle per Multi Locus Sequence Typing (MLST) umfassend phylogenetisch untersucht. Die von ursprünglich 41 auf 42 Stämme erweiterte Kollektion repräsentiert also die gesamte Breite an Serotypen und deren Sequenztypen, die im genannten Zeitraum von 11 Jahren von deutschen HUS-Patienten isoliert worden sind (44).

1.4 Phylogenie von EHEC

Ziel einer phylogenetischen Analyse ist es, eine Systematik zu finden, die die tatsächlichen evolutionären Beziehungen der untersuchten Organismen darstellt. Im Zusammenhang mit den in dieser Arbeit untersuchten EHEC stellt sich die Frage, wie aus dem apathogenen Kommensalen *E. coli* die hochpathogene HUSEC-Gruppe entstand und welche Verwandtschaftsbeziehungen innerhalb der HUSEC-Kollektion bestehen.

Frühere phylogenetische Untersuchungen z. B. durch Feng et al. zeigten, dass *E. coli* O157:H7 im Sinne einer schrittweisen Evolution von einem O55:H7-ähnlichen Vorfahren durch Akquisition verschiedener Virulenzfaktoren wie Stx1 und Stx2 entstanden sind (14). In diesem Szenario, welches auf Ergebnissen einer Multilocus Enzyme Electrophoresis-Analyse basiert, wird O55:H7 einer mit A bezeichneten Urgruppe zugeordnet, aus der sich die pathogeneren Gruppen B (O157:H⁻) sowie C (O157:H7) entwickelten (36, 80). Neuere Studien zur Phylogenie von EHEC O157:H7/H⁻ basierend auf DNA-sequenzbasierten Methoden bestätigten diese schrittweise Entstehung von EHEC O157:H7/H⁻ aus O55:H7 und konnten eine weitere Verfeinerung der Phylogenie herbeiführen. So lassen sich die momentan existierenden *E. coli* O157:H7 Stämme der C-Gruppe in drei Untergruppen ordnen (36, 64, 65), welche alle zur etablierten EHEC1-Klade gehören. Leopold et al. analysierten 2009 per „backbone-concatenome-analysis“ 1113 synonyme Single Nucleotide Polymorphism (SNP)-Loci in sieben evolutionsgeschichtlich interessanten O55:H7, O157:H⁻ und O157:H7 Stämmen (36). Ein SNP ist eine Variation eines einzelnen Basenpaares, also eine durch Mutation bedingte Variation des genetischen Codes. In einer weiteren Arbeit von 2010 konnten diese Autoren zeigen, dass das Konzept der limitierten Cladogenese auch für Cluster 2 zutrifft, somit also die Evolution von Cluster 1 über Cluster 2 zu Cluster 3 angenommen werden kann (37).

Die systematische Analyse von SNPs dagegen kann sowohl erhellende Beiträge zur evolutionären Geschichte nah verwandter Mikroorganismen liefern (2, 55), als auch Hinweise auf die Verknüpfung von bakteriellem Genotyp und Krankheitsausprägung bzw. Verlauf bei Ausbrüchen geben (42, 60). In stabilen Regionen des Genoms wie dem Kerngenom stellen Detektion und Vergleich von SNPs eine sehr belastbare Möglichkeit dar, diskriminatorische Marker für eine phylogenetische Analyse auch von genetisch monomorphen Erregern zu erhalten. Die erste Beschreibung des Einsatzes von SNPs zur Typisierung eines monomorphen Taxons erfolgte 1997 am Beispiel des *Mycobacterium tuberculosis* Komplexes durch Sreevatsan et al (68).

Im Gegensatz zur relativ gut untersuchten Gruppe der *E. coli* O157, gibt es kaum Arbeiten zu non-O157 EHEC. Multilocus Sequenztypisierungs-Analysen (MLST) konnten jedoch bereits zeigen (41), dass non-O157 EHEC ein deutlich breiteres Spektrum an Stämmen umfasst. Bei der MLST, die eine Adaption des Konzepts der multi-locus-enzyme-electrophoresis (MLEE) ist, können auf Basis der Sequenzdiversität einiger ausgewählter "housekeeping" Gene mikrobiologische Taxa mit hoher inter- und intralaboratorischer Reproduzierbarkeit eingeordnet werden. Leider ist die Auflösung für eine sichere phylogenetische Analyse von genetisch monomorphen Bakterien nicht ausreichend hoch (1), sodass hier zusätzliche Verfahren, wie z. B. SNP-basierte Verfahren, notwendig sind, um ein ausreichendes Diskriminationsvermögen zu erreichen. Der große Bedarf derartiger Verfahren wird durch die zunehmende Bedeutung von non-O157 EHEC betont, die zeigen, dass in Deutschland und Österreich EHEC O26:H11 in 15,4% der Fälle als ursächlicher Erreger von HUS - somit als zweithäufigster nach O157 - festgestellt wurde (84). Daten des Robert-Koch-Instituts (RKI) aus dem Jahre 2008 bestätigen dies und geben sogar O26 in 19 % der untersuchten Fälle als Verursacher der dem RKI übermittelten HUS-Erkrankungen an, also häufiger als O157 und O103 mit jeweils 15 %.

1.5 Zielsetzung

Aus diesen Gründen soll in der vorliegenden Arbeit ein belastbares Verfahren zur Diskrimination und Diagnostik des gesamten Spektrums der ätiologisch für HUS verantwortlichen Organismen etabliert werden, um die phylogenetische Entstehung insbesondere der non-O157 EHEC anhand von SNP-Analysen in 90 Genen des Kerngenoms genauer zu charakterisieren. Hierzu sollen die folgenden Fragen beantwortet werden:

- Sind in allen der 42 Stämmen der HUSEC-Kollektion alle 90 Zielgene vorhanden?
- Lassen sich Serotyp/Serogruppen-spezifische SNP-Muster erkennen?
- Gibt es Regionen im Genom, in denen besonders viele SNPs detektiert werden können?
- Welches phylogenetische Clustering ergibt sich nach Concatenom-Analyse der detektierten SNPs im Vergleich zur Serotypisierung und zur MLST-Analyse?

2 Material und Methoden

Es wurden Standardgeräte, -chemikalien und -materialien (Glas- und Plastikwaren) der Firmen 4titude (Surrey, Vereinigtes Königreich) ABgene (Genf, Schweiz), Applied Biosystems (jetzt Life Technologies) (Foster City, Ca, USA), Biometra (Göttingen), BioRad-Laboratories (München), Biozym (Wien, Österreich), Braun (Melsungen), Eppendorf (Hamburg), Merck (Darmstadt), Roth (Karlsruhe) Sigma (Osterode am Harz) Thermofischer (Waltham, MA, USA) und Sarstedt (Nümbrecht) verwendet. Spezielle Materialien, Geräte, Reagenzien und Chemikalien sind unter 2.1.1 bis 2.2.4 aufgeführt.

2.1 Untersuchungsgut und Materialien

2.1.1 Chemikalien und Enzyme

Tabelle 1: Chemikalie und Enzyme

Artikelbezeichnung	Artikelnummer	Hersteller
3130 POP 7 Polymer	4352759	ABI
10 x genetic Analyzer Buffer EDTA	402824	ABI
Hi-Dye-Formamide	4311320	ABI
Ethanol Absolute	8006	Baker
Ethidiumbromid Solution 100 mg/ml	161-0433	BioRad
LE Agarose	840004	Biozym
NaCl	65827	Honeywell
Lysozym 50.000 U/mg	1.052.810.001	Merck
Li Chrosolv Wasser	1.153.332.500	Merck
Isoamylalkohol	1009791000	Merck
Glycerin etwa 87 %, reinst	1.040.912.500	Merck
Exonuclease I (E.coli) 3000 Units	MO 293 S	New England Biolabs
Chloroform 99 % DAB	6340.1	Roth
Ethanol 96 % DAB reinst	P 075.1	Roth
Natriumchlorid	3957.3	Roth
R2523 RED Taq Ready Mix	R2523-100RXN	Sigma
Shrimp Alcaline Phosphatase 5000 Units	E 70092X	USB Amersham
5xSeq.Buffer aus Tris HCL pH 9 (400 mM)		Eigenherstellung
MgCl ₂ 10 mM		Eigenherstellung

2.1.2 Bakterienstämme

Die verwendete DNA entstammt der HUSEC-Kollektion des Instituts für Hygiene in Münster. Die bei -70 °C gelagerten Stämme wurden per Drei-Ösen-Ausstrich auf eine Blutagarplatte überimpft und danach 24 h bei 37 °C bebrütet. Die DNA-Extraktion aus den so gewonnenen Kulturen erfolgte nach dem Protokoll von Embden et al. (76).

Tabelle 2: Stämme der HUSEC-Kollektion unter Angabe des Serotyps sowie des MLST Sequenztyps (ST) und klonalen Komplexes (CC)

Alias Name	Serotyp	MLST ST (CC)	Alias Name	Serotyp	MLST ST (CC)
HUSEC001	O111:H10	43 (10)	HUSEC022	O145:H ⁻	137 (32)
HUSEC002	Ont ^a :H ^{-b}	330 (10)	HUSEC023	O112:H ⁻	40 (40)
HUSEC003	O157:H7	11 (11)	HUSEC024	O73:H18	69 (69)
HUSEC004	O157:H ⁻	11 (11)	HUSEC025	O55:Hnt	101 (101)
HUSEC005	O55:H7	335 (11)	HUSEC026	O113:H21	56 (56)
HUSEC006	O157:H ⁻	587 (11)	HUSEC027	O163:H19	679 (469)
HUSEC007	O103:H2	17 (20)	HUSEC028	O128:H2	25 (n.a.) ^e
HUSEC008	O103:H ⁻	17 (20)	HUSEC029	O70:H8	39 (n.a.)
HUSEC009	OR ^c :H2	17 (20)	HUSEC030	O98:H ⁻	306 (n.a.)
HUSEC010	O119:H2	20 (20)	HUSEC031	OR:H ⁻	306 (n.a.)
HUSEC011	O111:H8	16 (29)	HUSEC032	O136:Hnt	329 (n.a.)
HUSEC012	O111:H ⁻	16 (29)	HUSEC033	O145:H25	342 (n.a.)
HUSEC013	O26:H11	21 (29)	HUSEC034	O91:H21	442 (n.a.)
HUSEC014	O26:H ⁻	21 (29)	HUSEC035	O121:H19	655 (n.a.)
HUSEC015	OR:H11	21 (29)	HUSEC036	O145:H ⁻	659 (n.a.)
HUSEC016	Ont:Hnt ^d	21 (29)	HUSEC037	O104:H21	672 (n.a.)
HUSEC017	O26:H11	27 (29)	HUSEC038	Ont:H21	672 (n.a.)
HUSEC018	O26:H11	29 (29)	HUSEC039	O76:H19	675 (n.a.)
HUSEC019	OR:H11	29 (29)	HUSEC040	O174:H21	677 (n.a.)
HUSEC020	O26:H11	396 (29)	HUSEC041	O104:H4	678 (n.a.)
HUSEC021	O145:H28	32 (32)	HUSEC042	O165:H25	119 (n.a.)

^a Ont: O-Antigen nicht typisierbar

^b H⁻: H-Antigen nicht vorhanden

^c OR: O Antigen liegt als „rough“ vor, d.h. die Verlust-Mutanten sind nicht in der Lage in den Ablauf des Komplementsystems einzugreifen.

^d Hnt: H-Antigen nicht typisierbar

^e n.a.: nicht verfügbar

2.2 Methoden

Für die phylogenetische Analyse wurden insgesamt Sequenzen aus 90 verschiedenen ORFs bestimmt. Die ORFs wurden anhand ihres möglichst ubiquitären Vorkommens in *E. coli* sowie der Verteilung im Referenzgenom des EHEC O157:H7 Stammes Sakai ausgewählt. Die Liste der untersuchten ORFs sowie die Primersequenzen für die Amplifikations- und Sequenzierungsreaktionen sind in Tabelle 3 im Anhang aufgeführt.

2.2.1 Polymerase-Chain-Reaction – Amplifikation

Die PCR ist eine enzymatische Methode, DNA *in vitro* für die nachfolgende Sequenzierung zu vervielfältigen. Die PCR-Reaktion wurde in folgendem Ansatz durchgeführt (hier für eine Reaktion dargestellt):

- 12,5 µl Red-Taq-Ready (Sigma, Red-Taq Ready Mix)
- 7,5 µl HPLC-H₂O (Merck, Li Chrosolv Wasser)
- 2 µl Primer forward, Konzentration 5 pMol/µl (= 5 µM) (Sigma)
- 2 µl Primer reverse, Konzentration 5 pMol/µl (= 5 µM) (Sigma)
- 1 µl DNA des Stammes (entspricht ca. 10 ng DNA)

Thermocyclerprogramm:

- 94 °C für 45 sec.
 - 60 °C für 60 sec.
 - 72 °C für 90 sec.
 - 72 °C für 10 min.
- } 35 Zyklen

Zum Nachweis eines PCR-Produkts wurde eine Agarose-Gelelektrophorese durchgeführt.

2.2.2 PCR-Produktaufreinigung mit EXOSAP

Um die nachfolgende Sanger-Sequenzierung des PCR-Produkts nicht zu stören, müssen zunächst nicht-inkorporierte Primer und dNTPs entfernt werden. Dies erfolgt mittels einer enzymatischen Reaktion, nach deren Ablauf die Enzyme durch Erhitzen inaktiviert werden. Zur Aufreinigung des Produktes der Amplifikations-PCR werden zu 5 µl PCR-Produkt 2 µl EXOSAP, bestehend aus 1 µl Exonuklease (New England Biolabs, Ipswich, MA, USA) (1 U/ µl) und 1 µl Shrimp Alkaline Phosphatase (USB Amersham, Santa Clara, CA, USA) (1 U/µl) hinzugegeben. Die Reaktion wird dann für 45 min. bei 37 °C inkubiert und anschließend das Enzym bei 80 °C für 10 min. inaktiviert.

2.2.3 Sequenzierungs-PCR

Die Sequenzierungs-PCR läuft ebenfalls in einer zyklischen enzymatischen Reaktion ab. Bei der hier durchgeführten Methode von Sanger et al. (62) werden zusätzlich mit Fluoreszenz-Farbstoffen markierte Didesoxynukleosidtriphosphate (ddNTPs) eingesetzt. Wird ein solches ddNTP in den DNA-Strang eingebaut, kommt es zum Syntheseabbruch. Bei der Kapillarelektrophorese im Sequenzer werden die Fragmente nach Länge aufgetrennt und mit Hilfe eines Lasers zur Fluoreszenz angeregt, so dass anhand der verschiedenen Farbstoffe detektiert werden kann, welches Nukleotid sich an welcher Position des untersuchten Abschnitts befindet (57, 62, 67).

Sequenzierungs-PCR-Ansatz für eine Reaktion:

- 0,5 µl Premix (Applied Biosystems)
- 1,8 µl 5x ABI-Seq.-Puffer (40 % Tris HCl, 1 % MgCl₂, 59 % HPLC-H₂O)
- 2,0 µl Primer (forward oder reverse)
- 3,7 µl HPLC-H₂O
- 2,0 µl aufgereinigtes PCR-Produkt

Thermocyclerprogramm:

- 96 °C für 10 sec.
 - 60 °C für 2 min.
- } 25 Zyklen

2.2.4 Aufreinigung des Sequenzierungsproduktes

Damit es bei der Detektion im Sequenzer nicht zu unspezifischen Signalen kommt, müssen nicht eingebaute ddNTPs und Salze entfernt werden. Dies geschieht mittels einer Ethanol-fällung. Hierzu wird zunächst pro Probe ein 40 µl Ansatz Ethanol/Ammoniumacetat-Lösung vorbereitet:

- 1,5 µl 3 M Ammoniumacetat, pH 4,6
- 31,25 µl 96 %iger C₂H₅OH
- 7,25 µl HPLC-H₂O

Zu dieser Lösung werden 10 µl Sequenzierungs-PCR-Produkt hinzugegeben, ca. 15 sec. gründlich vermischt und anschließend 30 Minuten bei Raumtemperatur zur Präzipitation stehen gelassen. Nachfolgend wird die 96-Well-Platte orientiert für 30 Minuten bei 2754 g zentrifugiert. Im Anschluss wird die 96-Well-Platte zunächst getrocknet, dann 150 µl 70 %iger

Ethanol hinzupipettiert und wiederum für 10 min bei 2754 g zentrifugiert. Danach wird die Flüssigkeit entfernt und die Proben in einer Vakuumzentrifuge 5 min lang bei 45 °C getrocknet. Die so getrockneten Proben werden dann in 20 µl Hi-dye-Formamide (ABI) gelöst und im ABI Prism 3130xl Genetic Analyzer sequenziert.

2.2.5 Sequenzauswertung

Die Auswertung der Chromatogramme erfolgte mit dem Programm Ridom Seqsphere, Version 0.9 beta (Ridom GmbH, Münster).

2.2.6 Statistische Analyse

Das Auftreten von genetischen Varianten in einem ORF ist diskret verteilt, da an jeder Position die Ergebnisse a) Mutation oder b) keine Mutation möglich sind. Deskriptiv können somit arithmetisches Mittel, Standardabweichung, Varianz und Konfidenzintervall wie folgt ermittelt werden:

Arithmetisches Mittel:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Varianz:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$

Standardabweichung :

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Für das Konfidenzintervall um das arithmetische Mittel wurde ein Bereich von 3s gewählt, die Wahrscheinlichkeit für einen Fehler 1. Art beträgt also $\alpha=0,01$.

Da das Vorliegen von Mutationen als Bernoulli-Experiment mit der Anzahl von n= Länge des ORFs Wiederholungen desselben aufgefasst werden kann, liegt eine Binomialverteilung und somit ein Sonderfall einer Normalverteilung vor.

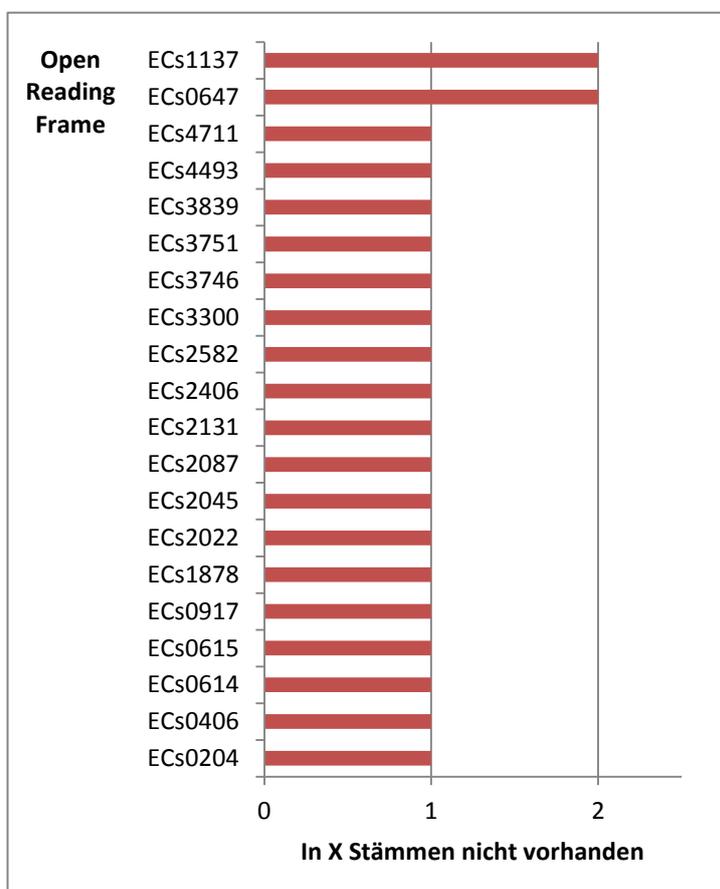
2.2.7 Software

- Seqsphere, Version 0.9β (Ridom GmbH)
- *MEGA4*: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 (70)
- *MEGA5*: Molecular Evolutionary Genetics Analysis (MEGA) software version 5.0 (72)
- SplitsTree4 Version 4.11.3 (27)

3 Ergebnisse

Insgesamt wurden für diese Arbeit alle 42 Stämme der HUSEC-Kollektion hinsichtlich der Anwesenheit von 90 Kerngenom-ORFs untersucht und Sequenzen von internen Fragmenten der nachgewiesenen ORFs für weiterführende phylogenetische Analysen bestimmt. Es zeigte sich, dass nicht alle Stämme der HUSEC-Kollektion über sämtliche Zielregionen verfügen. Von den 90 untersuchten Genen waren 70 in allen Stämmen vorhanden, die übrigen 20 Gene fehlten in ein bis zwei Stämmen.

Abbildung 1: Auf der Abszisse die ORFs, die nicht in allen Stämmen nachzuweisen waren, auf der Ordinate die Anzahl der Stämme, in denen dies der Fall war



In die weiterführenden phylogenetischen Analysen basierend auf den DNA-Sequenzen der einzelnen ORF-Abschnitte wurden nur die Fragmente aus den ORFs aufgenommen, die in allen untersuchten Stämmen vollständig nachgewiesen werden konnten:

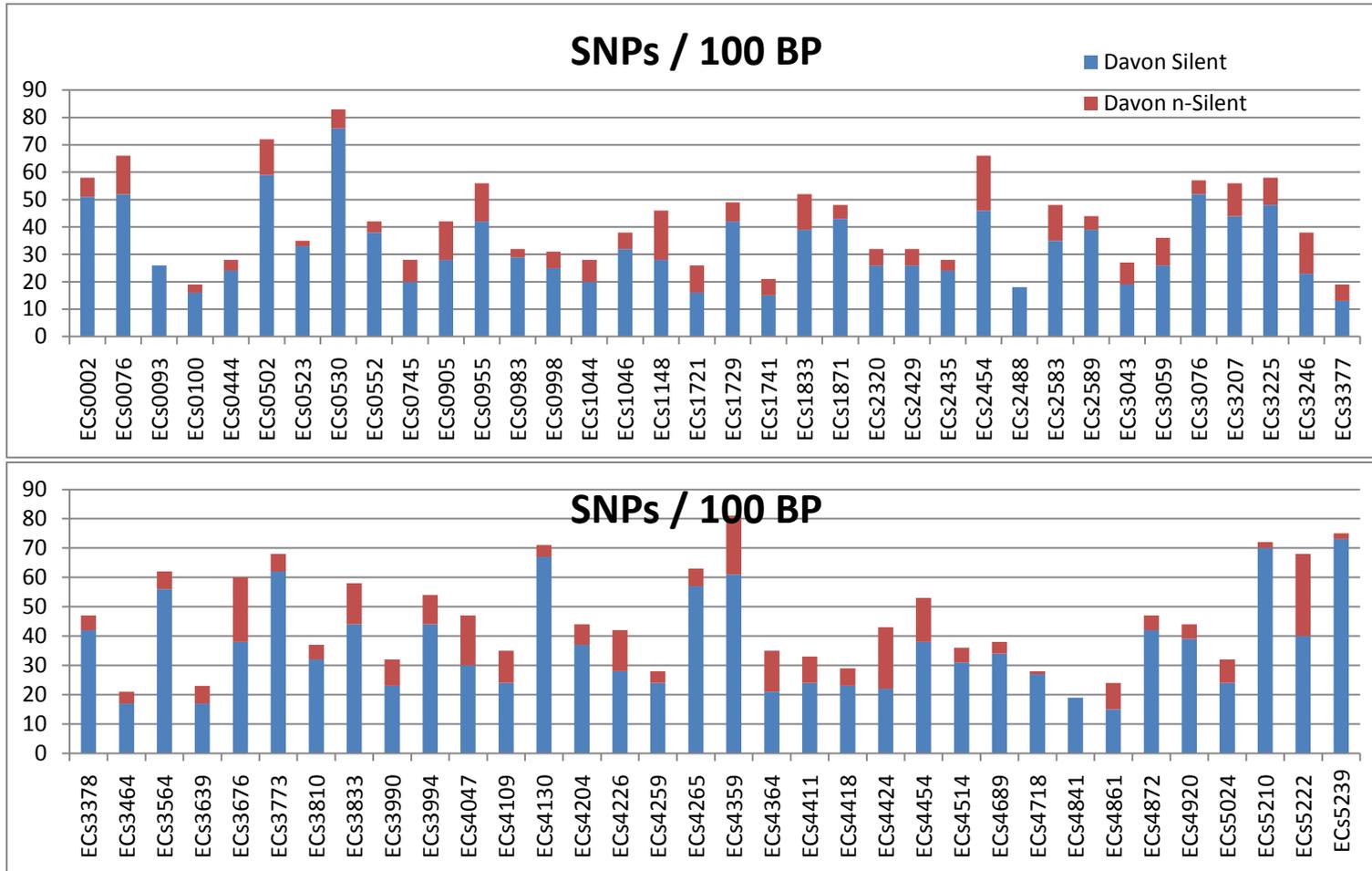
ECs0002, ECs0076, ECs0093, ECs0100, ECs0444, ECs0502, ECs0523, ECs0530, ECs0552, ECs0745, ECs0905, ECs0955, ECs0983, ECs0998, ECs1044, ECs1046, ECs1148, ECs1721, ECs1729, ECs1741, ECs1833, ECs1871, ECs2320, ECs2429, ECs2435, ECs2454, ECs2488,

ECs2583, ECs2589, ECs3043, ECs3059, ECs3076, ECs3207, ECs3225, ECs3246, ECs3377, ECs3378, ECs3464, ECs3564, ECs3639, ECs3676, ECs3773, ECs3810, ECs3833, ECs3990, ECs3994, ECs4047, ECs4109, ECs4130, ECs4204, ECs4226, ECs4259, ECs4265, ECs4359, ECs4364, ECs4411, ECs4418, ECs4424, ECs4454, ECs4514, ECs4689, ECs4718, ECs4841, ECs4861, ECs4872, ECs4920, ECs5024, ECs5210, ECs5222, ECs5239.

Insgesamt erfüllten 70 ORFs dieses Kriterium und wurden für die weiterführenden Analysen verwendet. In den 70 untersuchten ORFs (konkateniert 39788 bp) ließen sich an 3034 Positionen SNPs detektieren, welche zur weiteren Analyse als synonyme (sSNP) oder non-synonyme SNPs (nsSNP) klassifiziert wurden. Von den 3034 SNPs, die im Vergleich zur Referenzsequenz festgestellt wurden, wurden 2438, entsprechend 80,36 %, als sSNPs und 596, entsprechend 19,64 %, als nsSNPs klassifiziert. Zunächst wurde die Verteilung der SNPs über alle eingeschlossenen ORFs bestimmt. Hierzu wurden die Genfragmente konkateniert und die Zahl der SNPs in 100 bp großen Abschnitten bestimmt. Im Durchschnitt liegen pro 100 Basenpaare insgesamt 7,8 Mutationen vor, bzw. 6,1 sSNPs und 1,5 nsSNPs.

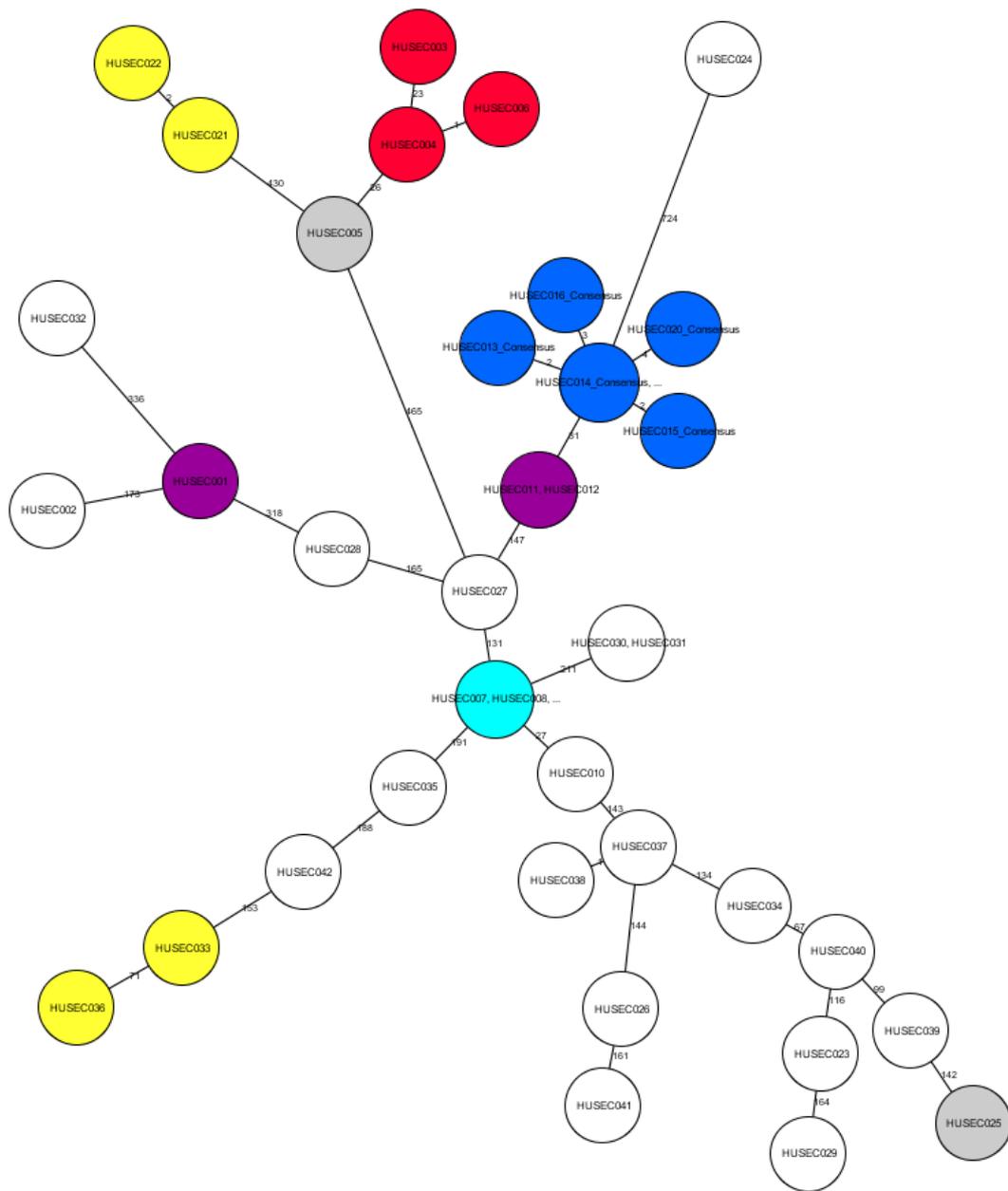
Für eine detaillierte Analyse wurde nun jeder einzelne ORF analysiert. Der unterschiedlichen Länge der Zielgene geschuldet wurde die Häufigkeit der Polymorphismen pro 100 Basenpaare dargestellt. Es wurden in allen Zielgenen Varianten detektiert, eine signifikant höhere Mutationsrate als im arithmetischen Mittel wurde allerdings in keinem der untersuchten ORFs festgestellt. Abbildung 2 zeigt die absolute Anzahl sowie die Verteilung von sSNPs bzw. nsSNPs pro untersuchtem ORF.

Abbildung 2: SNPs / 100 Basenpaare sowie Aufteilung sSNP vs. nsSNP



In der weiteren Analyse konnten für verschiedene Serogruppen spezifische SNPs festgestellt werden. Für Serogruppe O157 waren dies 18 Positionen, für O26 29 Positionen, für O103 (bzw. HUSEC 007-009 = molekular O 103) 27 Positionen. Im Einzelnen ist dies in Tabellen 4 bis 6 im Anhang dargestellt.

Als nächstes wurden die Daten verwendet, um das Verwandtschaftsverhältnis zwischen den einzelnen Stämmen zu untersuchen. Hierbei wurde zunächst ein mittels modifiziertem Algorithmus von Kruskal (15, 34) generierter minimum-spanning tree erstellt (Abbildung 3). Es können 35 Gruppen differenziert werden. Das Ergebnis der Serotypisierung der jeweiligen Stämme wurde zur besseren Übersicht wie in der Legende angegeben farblich codiert mit dargestellt.



- O157
- O103
- O26
- O55
- O111
- O145

Abbildung 3: Minimum-spanning tree der HUSEC-Kollektion. Jeder Kreis repräsentiert einen Stamm der HUSEC-Kollektion, die unterschiedlichen Farben stellen die Serogruppen dar. Die Nummer auf den Linien steht für die Anzahl an Varianten, durch die sich die Stämme unterscheiden.

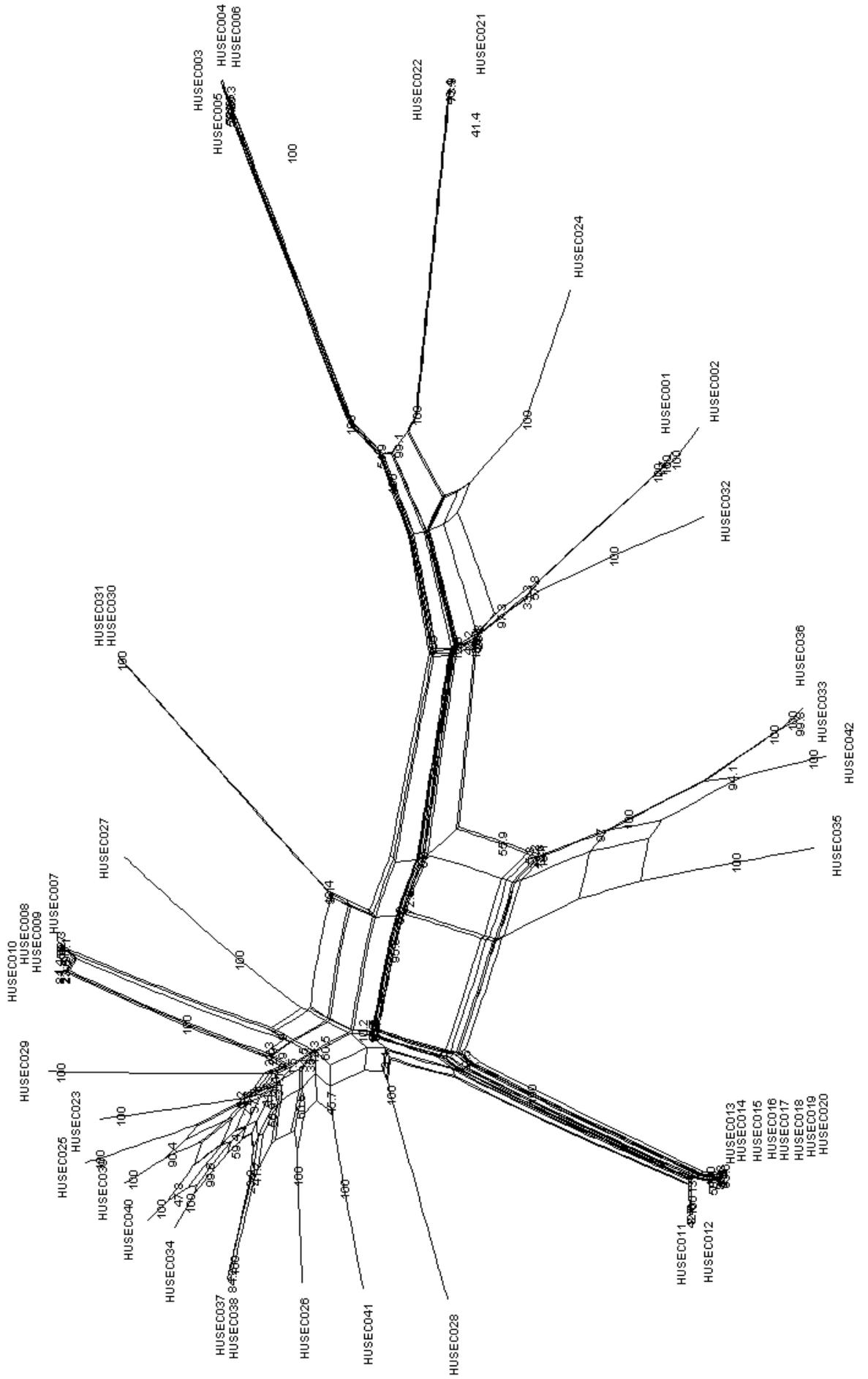
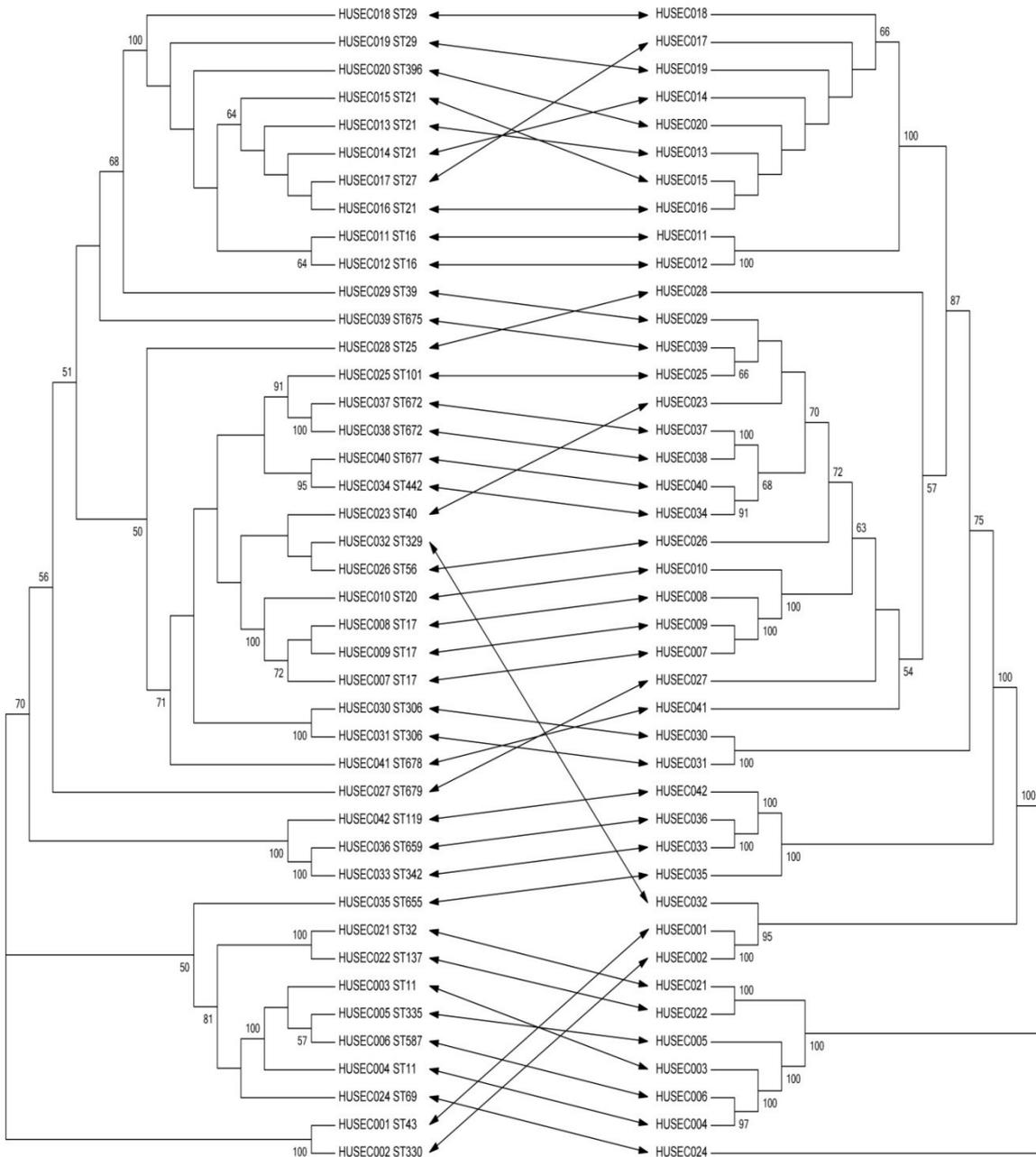


Abbildung 4: Konfidenznetzwerk mit bootstrap values

Um die komplexen phylogenetisch informativen Signale genauer aufzulösen und mögliche Rekombinationsereignisse darzustellen wurde mittels des Programms „SplitsTree“ die in Abbildung 4 gezeigte Ansicht eines „Konfidenz-Netzwerks“ (27) erstellt.

Schließlich wurden die hier generierten SNP-Daten mit bereits vorliegenden MLST-Daten verglichen. Hierzu wurden jeweils ein qua neighbour-joining Algorithmus (61, 71) erstelltes und mittels der „bootstrap-consensus Methode“ (13) verifiziertes Phylogramm auf Basis des SNP-Concatenoms (39788 bp) sowie ein mittels MLST (3423 bp) erstelltes gegenübergestellt (Abbildung 5). Hierbei zeigt sich, dass bei der Mehrheit der Stämme die Gruppierung durch beide Methoden weitgehend konkordant ist. Lediglich HUSEC 032 wird in den unterschiedlichen Phylogrammen wesentlich anders angeordnet.

Abbildung 5: Neighbour-joining bootstrap-consensus tree auf Basis von MLST-Daten links vs. SNP-Daten rechts



4 Diskussion

In der vorliegenden Arbeit wurde die phylogenetische Entwicklung und genetische Struktur der 42 Stämme umfassenden HUSEC-Kollektion anhand von 90 Zielgenen untersucht, die zufällig über das Genom verteilt sind und aus dem Kerngenom von EHEC O157:H7/H⁻ stammen. Hierbei wurden jeweils interne Genabschnitte mit einer Länge von 492 bis 618 bp untersucht. Interessanterweise konnten nur 70 der 90 Gene in allen Stämmen nachgewiesen werden, was schon einen ersten deutlichen Hinweis auf die Diversität der HUSEC-Kollektion gab. In den 70 ubiquitär nachweisbaren ORFs konnten an 3034 Positionen SNPs detektiert werden, was ebenfalls die hohe genomische Diversität unterstreicht.

Nicht in allen untersuchten Stämmen der HUSEC-Kollektion konnten sämtliche Zielgene detektiert werden, einige lagen nach Definition von Touchon et al. als sog. Pseudogene vor (75). Selbige besagt, dass das Leseraster der Zielsequenz durch Mutation einer einzelnen Base verschoben ist. Somit ist der genetische Code zwar konserviert, das kodierte Protein aber funktionslos. In diesem Zusammenhang wäre die Funktion des kodierten Proteins interessant, da, falls dieses für essentielle Funktionen des Zellstoffwechsels verantwortlich zeichnet, von einem Sequenzierungsfehler ausgegangen werden muss (75). In anderen Sequenzen fanden sich multiple Nonsensmutationen bzw. das Zielgen war gar nicht vorhanden. Für die meisten untersuchten Gene liegt die Funktion nicht gesichert vor, somit ist die Auswertung an dieser Stelle sicherlich limitiert. Der Anteil von 77 % im Untersuchungsgut ubiquitär vorkommender Gene korreliert nicht mit der Beobachtung von Touchon et al., dass das Kerngenom weniger als die Hälfte des durchschnittlichen *E. coli*-Genoms repräsentiert (75). Eine mögliche Erklärung ist, dass hier eine - zumindest in Bezug auf das ätiologisch verantwortete Krankheitsbild homogene - Gruppe von Stämmen untersucht wurde und somit die beobachtete Diversität auf Genomebene eingeschränkt ist.

Eine traditionelle Methode zur Differenzierung von EHEC, die Serotypisierung, beruht auf der An- oder Abwesenheit von Antigenen auf der Zelloberfläche, von Geißeln und von Kapseln und deren Reaktion mit spezifischen Antiseren. Der Vorteil dieser Methode besteht darin, dass sie mit relativ einfachen Mitteln zu bewerkstelligen und international standardisiert ist und somit den Vergleich von Ergebnissen zwischen Laboratorien ermöglicht. Ein Nachteil ist, dass Bakterien die Fähigkeit besitzen, den Ausdruck ihrer Eigenschaften unvorhersehbar zu ändern. Somit können unabhängige Isolate desselben Stammes phänotypisch variieren, die

Diskriminationsfähigkeit ist relativ gering (74). Aus diesem Grund ist interessant, ob basierend auf den hier gewonnenen SNP-Daten eine molekulare Differenzierung unterschiedlicher Serotypen möglich ist. In der Tat konnten für die wichtigsten Serotypen (u. a. O157:H7/H₇, O26:H11/H₁₁) serotypenspezifische SNPs identifiziert werden (Tabellen 4 bis 6 im Anhang). Hierbei ist allerdings zu beachten, dass diese SNPs natürlich mit größeren Stammkollektionen der jeweiligen Serotypen validiert werden müssen, um eine Stammspezifität auszuschließen.

Neben der Möglichkeit einer molekularen Serotypisierung können SNP-Daten grundsätzlich auch für die weitere Subtypisierung von Erregern eingesetzt werden. Die ist v. a. interessant für monomorphe Erreger, die in vielen der herkömmlichen Typisierungsmethoden wie der pulsed-field-gel-electrophoresis (PFGE) (69) nicht ausreichend differenziert werden können. Konkurrieren muss die SNP-Typisierung bei monomorphen Erregern beispielsweise mit der multi-locus-VNTR-analysis (MLVA) (31), clustered-regular-interspaced-short-palindromic-repeats (CRISPR) (10, 19), sowie multi-locus-sequence-typing (MLST) (39, 40). Da diese Verfahren auf der Analyse von häufig nicht-kodierenden Repeatregionen im Genom der Erreger beruhen, ist jedoch kontrovers diskutiert, in wieweit man diese Verfahren zur phylogenetischen Analyse heranziehen kann. So sind MLVA und CRISPR laut Achtman nicht geeignet, einen abschließenden Überblick der Evolution von genetisch monomorphen Erregern zu geben: MLVA auf Grund der hohen Mutationsrate und ergo hohen Homoplasierate sowie einiger Probleme bei der interlaboratorischen Reproduzierbarkeit (1, 38), CRISPR da - trotz einiger vielversprechender Resultate - die Auflösung ob der limitierten Anzahl von variablen kurzen palindromischen Wiederholungen (interspaced short palindromic sequences) möglicherweise nicht ausreichend ist (1).

Studien an O157 Stämmen mittels PFGE (50), octamer-based genome scanning (32) und MLVA (49) aus klinischen und tierischen Proben zeigten eine genetische Diversität auf. Weitere Untersuchungen legten nahe, dass Selbige weitestgehend auf Phagen-assoziierte Duplikationen, Insertionen und Deletionen zurückzuführen ist (35, 54, 65). Gene des sogenannten Kerngenoms (Gene, die sowohl in *E. coli* K-12 als auch *E. coli* O157 Sakai nachgewiesen wurden) dagegen zeigen eine große Uniformität von über 99 % der Nukleotidsequenzen (23, 50, 56, 83). Dieser hohe Grad der Genkonservierung führt auch zur Zuordnung der EHEC O157 zur Gruppe der monomorphen Organismen, die nur durch gezielte Sequenzanalysen einiger weniger Polymorphismen überhaupt differenziert werden können (1).

Die als O145 klassifizierten Stämme HUSEC 021 und 022 mit H-Phase H28 bzw. H⁻ sind im Phylogramm direkt benachbart, die anderen beiden zu dieser Serogruppe gehörigen analysierten Stämme HUSEC 033 und 036 mit H-Phase H25 bzw. H⁻ sind als weitere Subgruppe jedoch ebenso eng verwandt, liegen aber auf einem von ersterem Paar deutlich entfernten phylogenetischen Ast. Quasi identisch ist das Bild der Serogruppe O111, wo sich die Stämme HUSEC 001 mit H10 und HUSEC 011 sowie HUSEC 012 mit H8 bzw. H⁻ als Untergruppen gegenüberstehen. Die Vertreter der Serogruppe O55 dagegen scheinen im SNP-basierten Phylogramm willkürlich verteilt zu sein. Berücksichtigt man jedoch die Tatsache, dass die o. g. Stämme zwar der gleichen Serogruppe, aber durch ihre unterschiedlichen H-Phasen verschiedenen Serotypen zugeordnet sind, wird offensichtlich, dass man mittels SNPs nicht nur auf Serogruppenebene, also der O-Phase, sondern auch auf Serotypenebene, also der H-Phase differenzieren kann.

Die Verteilung der SNP-Positionen über das analysierte Genom ist in Abbildung 2 dargestellt. Es wurden in allen Zielgenen Varianten detektiert, eine überzufällig (>3s) hohe Mutationsrate bezogen auf das arithmetische Mittel wurde allerdings in keinem ORF festgestellt. Dies deckt sich gut mit der Annahme, dass das über die mannigfaltigen Facetten der *E. coli* konservierte Kerngenom nur einem sehr geringen Selektionsdruck ausgesetzt ist, bzw. die Proteine, für die es kodiert im mikrobiellen Stoffwechsels essentielle Funktionen wahrnehmen, deren mutationsbedingter Ausfall mit dem Fortbestand des mutierten Klons nicht vereinbar sind (75). Sog. „Hotspots“, wie z. B. Leopold et al. sie für das *rfb-gnd*-Segment beschrieben, konnten im Rahmen dieser Arbeit nicht als Stelle signifikant erhöhten SNP-vorkommens verifiziert werden (36, 48, 78). Es ergibt sich das in Abbildung 3 als minimum-spanning tree und in Abbildung 4 als Konfidenznetzwerk gezeigte phylogenetische Clustering. Diese untermauern unter anderem die in früheren Arbeiten aufgestellte These der sequentiellen Evolution von O157:H7 aus einem O55:H7 Vorfahren (36).

Mittels MLST können 30 unterschiedliche Sequenztypen bestimmt werden, der SNP-basierte Ansatz schafft dank höherer Auflösung 5 weitere, also insgesamt 35 Stämme, zu differenzieren. Auch können einige Aufzweigungen der phylogenetischen Bäume mit höheren „bootstrap-consensus“-Werten untermauert bzw. korrigiert werden. Es zeigt sich, dass in beiden Phylogrammen die Anordnung der Stämme weitestgehend ähnlich ist, also auch mit dem deutlich kleineren Ansatz der MLST (7 Gene) bereits eine gute Abbildung der Verwandtschaftsverhältnisse erreicht wurde. Auffällig ist, dass der zuoberst angeordneten

Aufzweigung im neighbour-joining tree auf Basis des SNP-Concatenoms HUSEC 024 und im auf MLST-Sequenzdaten fußenden HUSEC 001 bzw. HUSEC 002 zugeordnet wird. Innerhalb des Phylogramms gruppiert lediglich HUSEC 032 wesentlich anders. Das SNP-basierten Phylogramm erbringt hinsichtlich des Hervorgehens von O111:H8 aus O26 das gleiche Ergebnis wie im Shigatoxnet für die Erreger der dort als EHEC 2 Gruppe bezeichneten Klade. O111:H10 dagegen wurde sowohl von O111:H8 als auch den Vertretern von O26 weiter entfernt angesiedelt. Insgesamt erscheinen die Bilder, soweit bei den unterschiedlichen Schwerpunkten der Stammauswahl beurteilbar, weitestgehend kongruent.

Da einige Studien vermuten lassen, dass O157 EHEC und non-O157 EHEC eine nicht unbedeutende Zahl von Virulenzfaktoren teilen (6, 11, 44, 52), sequenzierten Ogura et al. 2009 drei EHEC Stämme, namentlich O26, O111 und O113, und verglichen die Sequenzen mit O157 sowie 15 weiteren bereits bekannten *E. coli* Sequenzen, um die Theorie der parallelen Evolution der EHEC Stämme zu prüfen. Die neu untersuchten EHEC Stämme wiesen ein größeres Genom auf als andere Stämme. Weiterhin zeigte sich eine signifikante Übereinstimmung in quasi EHEC-spezifischen Virulenzgenen wie Stx und anderen, so dass allen begutachteten Stämmen ein ähnliches pathogenes Potenzial attribuiert werden kann. Auf Grund der verschiedenen genetischen Strukturen der ähnlichen Virulenzfaktor-Sets deuten Ogura et al. diese Elemente als Hinweise auf eine unabhängige und komplexe Evolution der EHEC (53).

Die momentan existierenden *E. coli* O157:H7 Stämme der C-Gruppe lassen sich drei Untergruppen zuordnen (36, 64, 65,), welche alle zur etablierten EHEC1-Klade gehören. Leopold et al. analysierten 2009 per „backbone-concatenome-analysis“ 1113 synonyme SNP-loci in sieben evolutionsgeschichtlich interessanten O55:H7, O157:H⁻ und O157:H7 Stämmen. Es konnte dargelegt werden, dass die Stämme, welche Cluster 1 und 3 bilden, eine eingeschränkte - und nicht-zufällige - Auffächerung in weitere Unterarten bieten. Es konnte mit hinreichender Sicherheit gezeigt werden, dass im Cluster 1 der Wildtyp von O157:H7 einem von zwei vom „Ahnen“ ausgehenden Ästen zugeordnet werden darf (36). In einer weiteren Arbeit von 2010 konnten Leopold et al. zeigen, dass das Konzept der Cladogenese auch für Cluster 2 zutrifft, somit also die Evolution von Cluster 1 über Cluster 2 zu Cluster 3 angenommen werden kann (37).

Dürfen nun diese im Rahmen dieser Arbeit erstellten Phylogramme als bedeutsam bzw. belastbar angenommen werden, oder wird das phylogenetische Signal durch Störgrößen wie

Rekombination, Konjugation bzw. Transduktion ausgelöscht? In der Literatur zu diesem Problem gibt es durchaus kontroverse Standpunkte. Einerseits wird die durch sowohl MLEE (24, 51) als auch DNA-Marker-basierte (8, 12, 18, 81) populationsgenetische Studien erstellte klare Unterteilung der verschiedenen *E. coli* Stämme in vier phylogenetische Gruppen (A, B1, B2 und D), die unterschiedliche ökologische Nischen besetzen, klar als Anhalt für hohe Klonalität gewertet. Andererseits zeigen Sequenzanalysen ein nicht zu vernachlässigendes Aufkommen an Rekombination, horizontalem Gentransfer und Deletionen (20, 81), so dass die Übereinstimmung zweier *E. coli* - Genome weniger als 70 % betragen kann (79). Im Kerngenom jedoch wurden weniger als 3 % Sequenzdiversivität gefunden (79). Touchon et. al. konnten durch diverse Simulationen und Distanzermittlung nach Robinson Foulds (59) beim Vergleich verschiedener phylogenetischer Bäume darlegen, dass trotz der genannten Faktoren eine robuste, verwertbare Phylogenie aus *E. coli* - Sequenzdaten erstellt werden kann (75).

Die Wichtigkeit, das Wissen um die Phylogenie auch der non-O157 HUSEC zu erweitern, wird durch neuere Arbeiten betont. So wurde kürzlich EHEC O26:H11 in 15,4 % der Fälle in Deutschland und Österreich als ursächlicher Erreger von HUS - somit als zweithäufigster nach O157 - festgestellt (84). Konkordant zu diesem Ergebnis ist der Bericht des infektionsepidemiologischen Jahrbuchs des Robert-Koch-Instituts (RKI) aus dem Jahre 2008. Dort wird O26 in 19 % der untersuchten Fälle als Verursacher der dem RKI übermittelten HUS-Erkrankungen genannt, also häufiger als O157 und O103 mit jeweils 15 %. Ebenfalls in diesem Kontext bedeutsam ist der jüngste Ausbruch von HUS durch einen extrem virulenten O104:H4 in Deutschland, der über 830 Erkrankungen und 46 Todesfälle verursachte. Dieser Serotyp war zuvor nicht bei Tieren festgestellt worden und nur sporadisch mit HUS assoziiert (45).

Aus diesen Gründen sind weiterführende Kenntnisse über die Phylogenie aller HUS-assoziierten EHEC essentiell, um basierend auf Typisierungsdaten mögliche Reservoirs und Übertragungswege zu identifizieren sowie spezifische diagnostische Tests zu etablieren. Die vorliegende Arbeit gibt dazu einen Ausgangspunkt für zukünftige Studien mit größeren Stammkollektiven der verschiedenen Serotypen, die mit HUS assoziiert sind.

5 Literaturverzeichnis

- (1) Achtman M. (2008) Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol*; 62(1):53-70.
- (2) Alland D, Whittam TS, Murray MB, Cave MD, Hazbon MH, Dix K, et al. (2003) Modeling bacterial evolution with comparative-genome-based marker systems: application to *Mycobacterium tuberculosis* evolution and pathogenesis. *J Bacteriol*; 185(11):3392-3399.
- (3) Bielaszewska M, Dobrindt U, Gärtner J, Gallitz I, Hacker J, Karch H, et al. (2007) Aspects of genome plasticity in pathogenic *Escherichia coli*. *Int J Med Microbiol*; 297(7-8):625-639.
- (4) Bielaszewska M, Köck R, Friedrich AW, Von Eiff Christof, Zimmerhackl LB, Karch H, et al. (2007) Shiga Toxin-Mediated Hemolytic Uremic Syndrome: Time to Change the Diagnostic Paradigm? *PLoS ONE*; 2(10):8.
- (5) Bielaszewska M, Zhang W, Tarr PI, Sonntag A, Karch H. (2005) Molecular Profiling and Phenotype Analysis of *Escherichia coli* O26:H11 and O26:NM: Secular and Geographic Consistency of Enterohemorrhagic and Enteropathogenic Isolates. *J Clin Microbiol*; 43(8):4225-4228.
- (6) Brooks JT, Sowers EG, Wells JG, Greene KD, Griffin PM, Hoekstra RM, et al. (2005) Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983-2002. *J Infect Dis*; 192(8):1422-1429.
- (7) Clarke SC. (2001) Diarrhoeagenic *Escherichia coli*-an emerging problem? *Diagn Microbiol Infect Dis*; 41(3):93-98.
- (8) Clermont O, Bonacorsi S, Bingen E. (2000) Rapid and Simple Determination of the *Escherichia coli* Phylogenetic Group. *Appl Environ Microbiol*;66(10):4555-4558.
- (9) Cohen MB. (2002) Shiga toxin-producing *E. coli*: two tests are better than one. *J Pediatr*; 141(2):155-156.
- (10) Cui Y, Li Y, Gorge Olivier, Platonov ME, Yan Y, Guo Z, et al. (2008) Insight into Microevolution of *Yersinia pestis* by Clustered Regularly Interspaced Short Palindromic Repeats. *PLoS ONE*; 3(7):10.
- (11) Eklund M, Scheutz F, Siitonen A. (2001) Clinical isolates of non-O157 Shiga toxin-producing *Escherichia coli*: serotypes, virulence characteristics, and molecular profiles of strains of the same serotype. *J Clin Microbiol*; 39(8):2829-2834.
- (12) Escobar-Páramo P, Sabbagh A, Darlu P, Pradillon O, Vaury C, Denamur E, et al. (2004) Decreasing the effects of horizontal gene transfer on bacterial phylogeny: the *Escherichia coli* case study. *Mol Phylogenet Evol*; 30(1):243-250.
- (13) Felsenstein J. (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*; 39:783-791.
- (14) Feng PCH, Monday SR, Lacher DW, Allison L, Siitonen A, Keys C, et al. (2007) Genetic diversity among clonal lineages within *Escherichia coli* O157:H7 stepwise evolutionary model. *Emerg Infect Dis.*; 13(11):1701-1706.

- (15) Francisco AP, Bugalho M, Ramirez M, Carricco JA. (2009) Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics*; 10(1):152.
- (16) Fukushima H, Hashizume T, Morita Y, Tanaka J, Azuma K, Mizumoto Y, et al. (1999) Clinical experiences in Sakai City Hospital during the massive outbreak of enterohemorrhagic *Escherichia coli* O157 infections in Sakai City, 1996. *Pediatrics international : official journal of the Japan Pediatric Society*; 41(2):213-217.
- (17) Garg AX, Suri RS, Barrowman N, Rehman F, Matsell D, Rosas-Arellano MP, et al. (2003) Long-term renal prognosis of diarrhea-associated hemolytic uremic syndrome: a systematic review, meta-analysis, and meta-regression. *JAMA*; 290(10):1360-1370.
- (18) Gordon DM, Clermont O, Tolley H, Denamur E. (2008) Assigning *Escherichia coli* strains to phylogenetic groups: multi-locus sequence typing versus the PCR triplex method. *Environ Microbiol*; 10(10):2484-2496.
- (19) Grissa I, Vergnaud G, Pourcel C. (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics*; 8:172.
- (20) Guttman DS, Dykhuizen DE. (1994) Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science*; 266(5189):1380-1383.
- (21) Gyles CL. (2007) Shiga toxin-producing *Escherichia coli*: an overview. *J Anim Sci*; 85(13 Suppl):E45-62.
- (22) Harmsen D, Claus H, Witte W, Rothgänger J, Claus H, Turnwald D, et al. (2003) Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for spa repeat determination and database management. *J Clin Microbiol*; 41(12):5442-5448.
- (23) Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, et al. (2001) Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA res*; 8(1):11-22.
- (24) Herzer PJ, Inouye S, Inouye M, Whittam TS. (1990) Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J Bacteriol*; 172(11):6175-6181.
- (25) Heuvelink AE, van Heerwaarden C, Zwartkruis-Nahuis JTM, van Oosterom R, Edink K, van Duynhoven Y T H P., et al. (2002) *Escherichia coli* O157 infection associated with a petting zoo. *Epidemiol Infect*; 129(2):295-302.
- (26) Hildebrand JM, Maguire HC, Holliman RE, Kangesu E. (1996) An outbreak of *Escherichia coli* O157 infection linked to paddling pools. *Communicable disease report.CDR review*; 6(2):R33-6.
- (27) Huson DH, Bryant D. (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*; 23(2):254-267.
- (28) Kaper JB, Nataro JP, Mobley HL. (2004) Pathogenic *Escherichia coli*. *Nature reviews.Microbiology*; 2(2):123-140.
- (29) Karch H, Mellmann A, Bielaszewska M. (2009) Epidemiology and pathogenesis of enterohaemorrhagic *Escherichia coli*. *Berl Munch Tierarztl Wochenschr*; 122(11-12):417-424.

- (30) Karch H, Tarr PI, Bielaszewska M. (2005) Enterohaemorrhagic *Escherichia coli* in human medicine. *IJMM*; 295(6-7):405-418.
- (31) Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, et al. (2000) Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J Bacteriol*; 182(10):2928-2936.
- (32) Kim J, Nietfeldt J, Benson AK. (1999) Octamer-based genome scanning distinguishes a unique subpopulation of *Escherichia coli* O157:H7 strains in cattle. *Proc Natl Acad Sci U S A*; 96(23):13288-13293.
- (33) Klein EJ, Stapp JR, Clausen CR, Boster DR, Wells JG, Qin X, et al. (2002) Shiga toxin-producing *Escherichia coli* in children with diarrhea: a prospective point-of-care study. *J Pediatr*; 141(2):172-177.
- (34) Kruskal J. (1956) On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical society*; 5:1955-1957.
- (35) Kudva IT, Evans PS, Perna NT, Barrett TJ, Ausubel FM, Blattner FR, et al. (2002) Strains of *Escherichia coli* O157:H7 differ primarily by insertions or deletions, not single-nucleotide polymorphisms. *J Bacteriol*; 184(7):1873-1879.
- (36) Leopold SR, Magrini V, Holt NJ, Shaikh N, Mardis ER, Cagno J, et al. (2009) A precise reconstruction of the emergence and constrained radiations of *Escherichia coli* O157 portrayed by backbone concatenomic analysis. *Proc Natl Acad Sci U S A*; 106(21):8713-8718.
- (37) Leopold SR, Shaikh N, Tarr PI. (2010) Further evidence of constrained radiation in the evolution of pathogenic *Escherichia coli* O157:H7. *Infect Genet Evol.*; 10(8):1282-1285.
- (38) Lista F, Faggioni G, Valjevac S, Ciammaruconi A, Vaissaire J, le Doujet C, et al. (2006) Genotyping of *Bacillus anthracis* strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis. *BMC microbiology*; 6:33.
- (39) Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A*; 95(6):3140-3145.
- (40) Maiden MCJ. (2006) Multilocus sequence typing of bacteria. *Annu Rev Microbiol*; 60:561-588.
- (41) Maki DG. (2006) Don't eat the spinach--controlling foodborne infectious disease. *N Engl J Med*; 355(19):1952-1955.
- (42) Manning SD, Motiwala AS, Springman AC, Qi W, Lacher DW, Ouellette LM, et al. (2008) Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. *Proc Natl Acad Sci U S A*; 105(12):4868-4873.
- (43) Mead PS, Griffin PM. (1998) *Escherichia coli* O157:H7. *Lancet*; 352(9135):1207-1212.
- (44) Mellmann A, Bielaszewska M, Köck R, Friedrich AW, Fruth A, Middendorf B, et al. (2008) Analysis of collection of hemolytic uremic syndrome-associated enterohemorrhagic *Escherichia coli*. *Emerg Infect Dis.*; 14(8):1287-1290.
- (45) Mellmann A, Harmsen D, Cummings Ca, Zentz EB, Leopold SR, Rico A, et al. (2011) Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS one*; 6(7):e22751.

- (46) Nübel U, Roumagnac P, Feldkamp M, Song J, Ko KS, Huang Y, et al. (2008) Frequent emergence and limited geographic dispersal of methicillin-resistant *Staphylococcus aureus*. *Proc Natl Acad Sci U S A*; 105(37):14130-14135.
- (47) Nataro JP. (2005) Enteroaggregative *Escherichia coli* pathogenesis. *Curr Opin Gastroenterol*; 21(1):4-8.
- (48) Nelson K, Selander RK. (1994) Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (*gnd*) in enteric bacteria. *Proc Natl Acad Sci U S A*; 91(21):10227-10231.
- (49) Noller AC, McEllistrem MC, Pacheco AGF, Boxrud DJ, Harrison LH. (2003) Multilocus variable-number tandem repeat analysis distinguishes outbreak and sporadic *Escherichia coli* O157:H7 isolates. *J Clin Microbiol*; 41(12):5389-5397.
- (50) Noller AC, McEllistrem MC, Stine OC, Morris JG, Boxrud DJ, Dixon B, et al. (2003) Multilocus sequence typing reveals a lack of diversity among *Escherichia coli* O157:H7 isolates that are distinct by pulsed-field gel electrophoresis. *J Clin Microbiol*; 41(2):675-679.
- (51) Ochman H, Selander RK. (1984) Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol*; 157(2):690-693.
- (52) Ogura, Yoshitoshi and Ooka, Tadasuke and Asadulghani, Terajima J, Nougayrede J, Kurokawa K, Tashiro K, Tobe T, et al. (2007) Extensive genomic diversity and selective conservation of virulence-determinants in enterohemorrhagic *Escherichia coli* strains of O157 and non-O157 serotypes. *Genome Biol*; 8(7):R138.
- (53) Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, et al. (2009) Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proc Natl Acad Sci U S A*; 106(42):17939-17944.
- (54) Ohnishi M, Terajima J, Kurokawa K, Nakayama K, Murata T, Tamura K, et al. (2002) Genomic diversity of enterohemorrhagic *Escherichia coli* O157 revealed by whole genome PCR scanning. *Proc Natl Acad Sci U S A*; 99(26):17043-17048.
- (55) Pearson T, Busch JD, Ravel J, Read TD, Rhoton SD, U'Ren JM, et al. (2004) Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proc Natl Acad Sci U S A*; 101(37):13536-13541.
- (56) Perna NT, Plunkett G, Burland V, Mau B, Glasner JD, Rose DJ, et al. (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature*; 409(6819):529-533.
- (57) Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, et al. (1987) A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science*; 238(4825):336-341.
- (58) Reid SD, Herbelin CJ, Bumbaugh AC, Selander RK, Whittam TS. (2000) Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature*; 406(6791):64-67.
- (59) Robinson DF, Foulds LR. (1981) Comparison of phylogenetic trees. *Math Biosci*; 53(1-2):131-147.
- (60) Roumagnac P, Weill F, Dolecek C, Baker S, Brisse S, Chinh NT, et al. (2006) Evolutionary history of *Salmonella typhi*. *Science*; 314(5803):1301-1304.
- (61) Saitou N, Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*; 4(4):406-425.

- (62) Sanger F, Nicklen S, Coulson AR. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci*; 74(12):5463-5467.
- (63) Scheiring J, Andreoli SP, Zimmerhackl LB. (2008) Treatment and outcome of Shiga-toxin-associated hemolytic uremic syndrome (HUS). *Pediatr Nephrol*;23(10):1749-1760.
- (64) Shaikh N, Holt NJ, Johnson JR, Tarr PI. (2007) *Fim* operon variation in the emergence of Enterohemorrhagic *Escherichia coli*: an evolutionary and functional analysis. *FEMS Microbiol Lett*; 273(1):58-63.
- (65) Shaikh N, Tarr PI. (2003) *Escherichia coli* O157:H7 Shiga toxin-encoding bacteriophages: integrations, excisions, truncations, and evolutionary implications. *J Bacteriol*; 185(12):3596-3605.
- (66) Slutsker L, Ries AA, Greene KD, Wells JG, Hutwagner L, Griffin PM. (1997) *Escherichia coli* O157:H7 diarrhea in the United States: clinical and epidemiologic features. *Ann Intern Med*; 126(7):505-513.
- (67) Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, et al. (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* ; 321(6071):674-679.
- (68) Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, et al. (1997) Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci*; 94(18):9869-9874.
- (69) Swaminathan B, Barrett TJ, Hunter SB, Tauxe RV. (2001) PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance. *Emerg Infect Dis.* ; 7(3):382-389.
- (70) Tamura K, Dudley J, Nei M, Kumar S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*; 24(8):1596-1599.
- (71) Tamura K, Nei M, Kumar S. (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci*; 101(30):11030-11035.
- (72) Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*; 28(10):2731-2739.
- (73) Tarr PI, Gordon CA, Chandler WL. (2005) Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. *Lancet* ; 365(9464):1073-1086.
- (74) Tenover FC, Arbeit RD, Goering RV. (1997) How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. *Infect Control Hosp Epidemiol.*; 18(6):426-439.
- (75) Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, et al. (2009) Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS genetics*; 5(1):e1000344.
- (76) van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. (1993) Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol*; 31(2):406-409.
- (77) Varma JK, Greene KD, Reller ME, DeLong SM, Trottier J, Nowicki SF, et al. (2003) An outbreak of *Escherichia coli* O157 infection following exposure to a contaminated building. *JAMA*; 290(20):2709-2712.

(78) Wang L, Huskic S, Cisterne A, Rothmund D, Reeves PR. (2002) The O-antigen gene cluster of *Escherichia coli* O55:H7 and identification of a new *UDP-GlcNAc C4* epimerase gene. *J Bacteriol*; 184(10):2620-2625.

(79) Welch RA, Burland V, Plunkett G, Redford P, Roesch P, Rasko D, et al. (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci*; 99(26):17020-17024.

(80) Whittam TS, Wilson RA. (1988) Genetic relationships among pathogenic *Escherichia coli* of serogroup O157. *Infect Immun*; 56(9):2467-2473.

(81) Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, et al. (2006) Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol*; 60(5):1136-1151.

(82) Wong CS, Jelacic S, Habeeb RL, Watkins SL, Tarr PI. (2000) The risk of the hemolytic-uremic syndrome after antibiotic treatment of *Escherichia coli* O157:H7 infections. *N Engl J Med*; 342(26):1930-1936.

(83) Zhang W, Qi W, Albert TJ, Motiwala AS, Alland D, Hyytia-Trees EK, et al. (2006) Probing genomic diversity and evolution of *Escherichia coli* O157 by single nucleotide polymorphisms. *Genome Res*; 16(6):757-767.

(84) Zimmerhackl L, Rosales A, Hofer J, Riedl M, Jungraithmayr T, Mellmann A, et al. (2010) Enterohemorrhagic *Escherichia coli* O26:H11-Associated Hemolytic Uremic Syndrome: Bacteriology and Clinical Presentation. *Semin Thromb Hemost*; 36(6):586-593.

6 Lebenslauf

7 Danksagung

8 Anhang

Tabelle 3: Liste der untersuchten ORFs (Nomenklatur anhand der Referenzsequenz von EHEC O157:H7 Sakai (GenBank Accession-Nr. NC_002695) sowie Sequenzen der verwendeten Primer für jeden ORF

ORF	Primer 5'-3'	Sequenz des Primers (5'-3')	Primer 3'-5'	Sequenz des Primers (3'-5')	Amplicongröße, bp	Länge analysiertes Fragment, bp
ECs0002	0002_1558_F	GCATGGCCTAAATCTGAAAA	0002_2214_R	AGCGCAAACCTTTCTCTCA	695	618
ECs0076	0076_42_F	GAAATGGCCGGTCACAGC	0076_682_R	CACCTGACGAAACCACCTTT	677	584
ECs0093	0093_293_F	GCGTTTATTCTGGCGATCAT	0093_915_R	CACCGACATACCCAGTTCT	661	531
ECs0100	0100_276_F	GCTTGGGCATCGATAACATT	0100_851_R	GTCGTCCTGGAAGGTCACAT	614	504
ECs0204	0204_40_F	CTCCGGGAGCTGTTTTACA	0204_693_R	TCGCGTACAGTCAATTTTGC	692	585
ECs0406	0406_61_F	TGGCACCGATCTGATGATTA	0406_676_R	GCCTCAGCCATTTTCATTGAT	654	545
ECs0444	0444_203_F	CTGGCTCAATTCGCCTTAC	0444_781_R	AACTGCGGAACCGTTTGATA	607	511
ECs0502	0502_115_F	ACTGGTTCACCAAAAGTGG	0502_715_R	TCGAAACGAGCATCAATACG	639	569
ECs0523	0523_843_F	GTATCGAGTGGGAAGCGTTG	0523_1447_R	TCTTTTTGCTGCATCACTGG	643	548
ECs0530	0530_238	CGAAGCGCTTTATCAGGAAC	0530_899_R	ATGCTGGGTTTTACGTTTTCG	700	584
ECs0552	0552_265_F	GCGCTTCAGCCTGTAAAAAC	0552_885_R	CCCATTCTCATTGTTGTCG	659	548
ECs0614	0614_42_F	CGGAAGATACTGCGTGAAT	0614_621_R	ACAAAATCAGCCATTCCAG	618	548
ECs0615	0615_140_F	TTCATATTCAGCCACACCA	0615_769_R	TTTTGATGATCTCGTGCTG	668	558
ECs0647	0647_31_F	GACTTATCACGCCACATC	0647_654_R	CAAGCGGTAGGGATGATGA	661	568
ECs0745	0745_309_F	GTGCCAGCTCTTTCAGAAC	0745_942_R	GACCCGTCATACCATGATCC	672	560
ECs0905	0905_574_F	CGCGATAATCCCTAAGTTG	0905_1187_R	GACACCGCGCTTAATGAGAT	652	562
ECs0917	0917_107_F	CTGCAAGACAACTCGACCA	0917_763_R	TAGTTGATTCAGGTTGTC	695	604
ECs0955	0955_366_F	TGCCACAAAATATTGCCAAA	0955_973_R	TAATCGTTGGTGACGTTCA	646	541
ECs0983	0983_281_F	CTTGGCCTGATGATTGGTTT	0983_898_R	GCCACCGGATATAGCGTAAA	656	539
ECs0998	0998_81_F	GGTTGTATGGCCTGGTGAGT	0998_673_R	TGTTCAAGCTGAGCAACGTC	631	522
ECs1044	1044_19_F	TTACGCCGAAACAGGAGTT	1044_594_R	CGATACGCTGATTGGTTGTG	613	544
ECs1046	1046_155_F	ATTCAGTGAAGCGCGGTCT	1046_730_R	CCAAACGCCAGCAATAAAAT	614	544
ECs1137	1137_1518_F	CCGCTTTCTTACTGCGTTC	1137_2111_R	ATCGATCTGCTTCGCTGGT	632	530
ECs1148	1148_1491_F	TCCTGCAATTCAGCTGACG	1148_2089_R	TGCAGCAAATGGTGATGAAT	637	557
ECs1721	1721_338_F	GAAGCCAGTAAAGGCGACAG	1721_993_R	GGGAAGCAGCCAGACTACT	694	600
ECs1729	1729_484_F	CGGTGGTTTTGTTCTTCTT	1729_1102_R	TTCCATTCCGGATTGTTTTT	657	556
ECs1741	1741_406_F	TGAATGCAGTCTGCTGGTC	1741_982_R	ACGAACATCCGAAACTCAC	615	516
ECs1833	1833_326_F	TTCTTCAATCTGCCGCTCTT	1833_898_R	GTCAGGCTTTACTGGCGAAG	611	518
ECs1871	1871_401_F	CTGAAAGCGACCTTCTACCG	1871_986_R	GGCACTGGACCGTATCAGTT	624	554
ECs1878	1878_393_F	TGTTGTGGCGATGAATGAT	1878_976_R	ATTTTCAGCCCTCAACTG	622	537
ECs2022	2022_36_F	TTTTGAGGGTGCGAATAAG	2022_645_R	CGCGAAGAAGGTGTTGATT	648	557
ECs2045	2045_191_F	GGCGGGCTATCTCTATCTT	2045_793_R	AGTTCACCAGGGGTGTTGAG	641	561
ECs2087	2087_281_F	CACAATCTGCCGAGATACA	2087_893_R	TAACGTTACGCGACGTCCA	650	567
ECs2131	2131_76_F	CAAGCGAGTTACCTGCATCA	2131_736_R	CGCAAGAACGTTTTTCCATT	699	583

ECs2320	2320_280_F	TCAGAGCGTCATCTTCAACG	2320_891_R	CATCAATTTGGCCGACTTTT	650	557
ECs2406	2406_587_F	GGCCTCACACCAGAACAGAT	2406_1202_R	CATGATCATTTCGCGTACCG	654	550
ECs2429	2429_315_F	AAGCAAGCGGTGAGCAGTAT	2429_882_R	TATCGTCATGGGAGCACAGA	606	510
ECs2435	2435_40_F	GAACATCGTGGTTCGTTG	2435_666_R	GTTTCATCACCCAGCTTTGC	665	554
ECs2454	2454_38_F	AACAAAGTTCGCAAGCAA	2454_657_R	CTCTACCTGTGCGGTGATTG	658	536
ECs2488	2488_310_F	TGGTCTGTTCTGACTGACG	2488_936_R	AGTAACCGGTTTCGTTGTCG	665	587
ECs2582	2582_278_F	GTTCCAGACACGGACCAGAT	2582_887_R	TTTGAGCGCAATGACTTGAC	648	534
ECs2583	2583_495_F	TCGCCTATGCGAGAAGTTTT	2583_1066_R	CGCATTGGGTTATTGTTTTTG	611	527
ECs2589	2589_1054_F	CGGCATTATTCTCGCCATT	2589_1706_R	GTTGGTGGCAATTCGCTAT	691	574
ECs3043	3043_240_F	TCATCGAAACCGATGATTGA	3043_822_R	GGACGTTTCTGATGCGTTTT	621	509
ECs3059	3059_51_F	GGACGTTTCAGGAAGGCATA	3059_669_R	GCAACTGAAACTGCCACAGA	657	556
ECs3076	3076a_179_F	GTGCTGGCACACGTTAAAGA	3076a_834_R	GGGTGTCGCCAGTAATCAGT	694	564
ECs3076	3076b_1147_F	CAACGTTCCGGTACAGGTTT	3076b_1722_R	CGACGAAAACGACCTTCG	612	524
ECs3207	3207_386_F	CATCGCCATCTGCATACAAC	3207_1043_R	CTGGATACCACTGGCCTCAT	696	581
ECs3225	3225_63_F	GCACAGGCGGTAACATAACC	3225_647_R	GCATGGTCAGACGGAAAAC	623	528
ECs3246	3246_302_F	CGACGTACCAACTGATCCTG	3246_923_R	GACTTCACCGGTCAAATGAA	661	557
ECs3300	3300_365_F	AAATTCGCCACCATCTTAC	3300_983_R	GCTGTTGAATCCTTGCTGT	657	541
ECs3377	3377_470_F	ACCAATGGCGGATTTACTG	3377_1098_R	GCATAACCAGGCTCCAATTC	667	571
ECs3378	3378_242_F	CCAGCAATCGGCAGTAAAGT	3378_836_R	GCCGATCTTGCTTCAACATT	633	538
ECs3464	3464_407_F	CGGATTCCTCGATTTGTA	3464_1003_R	TAATGACTCCGGCACAAC	635	539
ECs3564	3564_287_F	CGCGATGTGATGCTGTTAT	3564_898_R	AGTTTGCCGTTTCATCCAC	650	543
ECs3639	3639_69_F	CCCAGAGCTTCTCGATACG	3639_681_R	AGGCATGAACACTGCTGTTG	651	560
ECs3676	3676_1166_F	ATCGGTCAGCGGTAACGTG	3676_1784_R	GAAGCTGTGGAGCACAAACA	657	551
ECs3746	3746_719_F	CCGCTCTATATCGTGACCT	3746_1302_R	AGGTGCCGTCACAGAAAATC	622	522
ECs3751	3751_2250_F	ATGCTGATGGCACCTTAACC	3751_2873_R	CAGGCTGAAAACGGTGATT	662	554
ECs3773	3773_726_F	TGAAAGTCGGCTGTATGCTG	3773_1351_R	ATATCGCCAGTACCGTCGTC	664	564
ECs3810	3810_1293_F	AGGGACGGTTTGTGAGTAC	3810_1932_R	GAGCATGGACGCAGTACAGA	678	585
ECs3833	3833_299_F	CGCATCTACCAGGGCTTTAG	3833_913_R	GTGACTCCGCAACCAATCT	653	545
ECs3839	3839_474_F	CAAGCAACTCGCGGATTAT	3839_1047_R	ATTTTTGCGCGGTATTCACT	612	542
ECs3990	3990_378_F	CGCGGTAACGGTTATGGAT	3990_1018_R	GGCTGGAAGTGCTGAAAGAC	679	564
ECs3994	3994_524_F	TGGCAGCGCTTAATTTTCT	3994_1136_R	TCGGAAGAATTACCGATTGC	651	548
ECs4047	4047_149_F	TGTACGAACCGGTTACCAT	4047_743_R	GCGACGAAGTTTTCCAGTA	633	521
ECs4109	4109_39_F	GCCCAGGGCGATATCTTT	4109_666_R	GGATCGTTCGACCTGTTTA	664	571
ECs4130	4130_367_F	CCGCTACCAGAGTCGTCTTC	4130_1002_R	GCAATTGGGCGTTAATTGTC	674	593
ECs4204	4204_235_F	ACCTCGTCTGGTGGTGTTC	4204_806_R	AATAATCAGCGTCGGTTGG	610	514
ECs4226	4226_37_F	AACCAATCGCTTCGTACACC	4226_671_R	CGCATGACGCAGTTTAAAGA	673	603
ECs4259	4259_288_F	TTCAATACTGCGTCCAGCAC	4259_884_R	CAGCAATATCGCGAGATCAA	635	536
ECs4265	4265_233_F	GTGGTGAGCGTTGAGATTGA	4265_823_R	GCGCTGCGTAATAATCCTTC	629	507
ECs4359	4359_780_F	ATCCAGACACCGAGTCCAC	4359_1384_R	TGTTCTGGCAGTTGATGGTC	643	554
ECs4364	4364_20_F	AGCACGACTTCGCTGCAA	4364_662_R	AAGATTGCCGAACAATTTGG	679	585
ECs4411	4411_428_F	TTTTAGCTGCCAGCCTTTGT	4411_1047_R	GGCTGCCTTTAGTGTTCAGG	658	536
ECs4418	4418_150_F	CGCTCCTCAATTTGGTGTTC	4418_781_R	TCAATATCGAAATCGCACCA	670	561
ECs4424	4424_906_F	GTATCGATCTGCGGTTGGT	4424_1564_R	GGATGCTGAAGCTTGGTCTC	697	579
ECs4454	4454_42_F	TGACACTCCTTCTGGCTTC	4454_639_R	TTTCCGCATACCGTCTTTA	636	541
ECs4493	4493_361_F	GCAACCGACTCACTGACGTA	4493_956_R	TATAATGCGGGCGATGATT	634	527

ECs4514	4514_239_F	GGCCATATTGAGCTGGGTAA	4514_864_R	TGGCCTGCTTTTGGATTTTC	664	551
ECs4689	4689_322_F	GCTGGTGATTATGGGGCTAA	4689_979_R	TCCATTTGAGAGGTGGAAT	696	568
ECs4711	4711_22_F	GTCTAAACCCCGCCAAC	4711_654_R	CATACTGGCTGGTGTGGTG	669	584
ECs4718	4718_366_F	GAGAGTTCTGGCTGCAAACC	4718_965_R	AATCATCAGGAAGGCACGAC	638	523
ECs4841	4841_181_F	CCGTTACAGCGTTTCTGACA	4841_786_R	CCATACGGAAGCCAGAATA	644	564
ECs4861	4861_177_F	TCACAAAATTGCGTCAGGTC	4861_829_R	GCAACGTCAAGCGTAATGAA	691	588
ECs4872	4872_234_F	CCTTAGCCATTGCCCTGTTA	4872_828_R	GCACAATCAGCACAATACCG	633	547
ECs4920	4920_376_F	TGAACTGGGGTACGTGATGA	4920_965_R	GCCGTGACCAAGGTTAAAGA	628	531
ECs5024	5024_1136_F	GGCGTAAGTACCCACTTCCA	5024_1730_R	ATTGCCAAAGCGGTAGAAGA	633	540
ECs5210	5210_149_F	AAGCAAGGCATTGAGCTGAT	5210_772_R	TCCAGCAAACCTCCCATTTC	662	576
ECs5222	5222_284_F	GTCAACTCCGCCATCACTT	5222_913_R	TGGCTAATCCGCTATATCAGAAA	671	522
ECs5239	5239_468_F	AAGGCTTATGGGCGAAAGAT	5239_1046_R	CAGCCACAGGCTGATTAAGA	617	522

Tabelle 4: Serogruppenspezifische SNPs O26

ORF	SNP-Position	Base
ECs0002	1702	C
ECs0002	1803	G
ECs0002	1851	A
ECs0002	1971	C
ECs0002	1974	G
ECs0002	1980	A
ECs0002	2037	G
ECs0002	2061	C
ECs0002	2133	C
ECs0002	2183	G
ECs0076	873	G
ECs0076	801	G
ECs0100	340	C
ECs0523	1224	C
ECs0983	567	T
ECs2429	698	G
ECs2583	177	C
ECs3676	350	C
ECs4204	681	G
ECs4359	1714	C
ECs4411	636	G
ECs4454	101	A
ECs4718	593	C
ECs4841	747	C
ECs4920	418	C
ECs5222	543	A
ECs5239	801	T
ECs5239	804	C
ECs5239	985	C

Tabelle 5: Serogruppenspezifische SNPs O157

ORF	SNP-Position	Base
ECs0076	1212	C
ECs0100	406	G
ECs0444	407	C
ECs0502	342	T
ECs0905	453	C
ECs1046	358	C
ECs2454	693	A
ECs2589	771	A
ECs3207	291	C
ECs3246	1113	C
ECs3377	226	T
ECs3676	147	C
ECs3773	1100	A
ECs4359	1726	G
ECs4364	1079	G
ECs4424	192	T
ECs4920	568	G
ECs5210	366	T

|

Tabelle 6: Serogruppenspezifische SNPs O103

ORF	SNP-Position	Base
1044	1746	G
1833	414	G
3059	1242	C
3810	663	T
3810	372	G
3810	351	A
3810	315	A
3810	309	T
3833	568	G
3833	543	A
3833	540	G
3833	537	A
3833	534	G
3833	531	C
3833	528	T
3833	519	G
3833	510	C
3833	495	A
3833	492	A
3833	426	G
3833	351	G
3833	303	T
3833	195	T
3833	192	T
4204	323	T
4418	329	T
5222	230	C