

Diskussion

Den Beginn der Arbeit stellte die kritische Auswahl geeigneter Zielgene zur Generierung von DNA-Sequenzen aus dem *Escherichia (E.) coli*-Genom dar. Da Rekombinationen bei Housekeeping-Genen, insbesondere bei solchen, die für metabolische Enzyme kodieren, seltene Ereignisse sind (Boyd and Hartl, 1997), wurden ausschließlich diese Gene in Betracht gezogen. Bei metabolischen Enzymen ist u.a. die Aminosäure-Substitution in der NAD⁺-bindenden Domäne geringer als in der katalytischen (Boyd et al., 1994; Boyd, 1999). Weiterhin sollten keine Gene ausgewählt werden, die in direkter Nachbarschaft zu Genen stehen, die einem erhöhten Selektionsdruck ausgeliefert sind, z. B. durch den Wirt. Die Gene sollten zudem in Genomregionen lokalisiert sein, die keine sogenannten „Hotspots“ aufweisen bzw. in der Nähe derartiger Bereiche liegen. Ein weiteres Kriterium war auch, ob das Gen bereits von anderen Arbeitsgruppen für phylogenetische Untersuchungen genutzt worden war, denn damit könnte eine erste Abschätzung der eigenen Ergebnisse in einem deutlicheren Bild erscheinen. Ebenso war die Expressionsrate des untersuchten Gens von Bedeutung. So zeigen hochexprimierte Gene eine verminderte Mutationsrate, denn deren Codon-Nutzung ist sehr stringent. Weiterhin sollten die ausgewählten Gene möglichst gleichmäßig über das gesamte Chromosom hin verteilt liegen. Die Wahl fiel aus den genannten Gründen schließlich auf die Gene *trpC*, *mdh*, *putP* und *cadB*. Die Auswahl der letztlich für die PCR-Amplifizierung benötigten Oligonukleotidpaare erfolgte anhand der DNA-Sequenz des K-12-Referenzstammes MG1655 (Acc. No. AE000203).

Bezüglich des *trpC*-Gens von Enterobakterien lagen z.B. bereits Daten vor. Unter anderem konnte mit Hilfe der analysierten *trpC*-Gene eine enge phylogenetische Verwandtschaft zwischen *E. coli*, *Klebsiella* sowie *Citrobacter* belegt werden. Analysen anderer Housekeeping-Gene widerlegten diese Daten jedoch (Reeves and Stevenson, 1989; Bisercic et al., 1991; Bisercic and Ochman, 1993b; Bisercic and Ochman, 1993a; Nelson and Selander, 1994).

Das Gen für die Malatdehydrogenase (*mdh*) wurde ausgewählt, da zu Beginn der Dissertation bereits die entsprechenden Gene von fünf *E. coli*-Stämmen beschrieben waren (K-12 Acc.No. AE000404) und auch die Häufigkeit des Locus im Chromosom war bekannt (Ochman and Selander, 1984a; Boyd et al., 1994; Blattner et al., 1997; Smith, 1998), so dass die bereits vorhandenen DNA-Sequenzen in die Auswahl für eine Primergestaltung einbezogen werden konnten. Darüberhinaus hatte es den Vorteil, dass die interessanten

Untersuchungen von Pupo et al. (Pupo et al., 1997; Pupo et al., 2000b) über evolutionäre Verwandtschaften zwischen pathogenen und nicht-pathogenen *Escherichia coli*-Stämmen, die sowohl mittels Multilocus-Enzymelektrophorese (MLEE) als auch mittels *mdh*-Sequenzanalysen durchgeführt worden waren, mit herangezogen werden konnten (verwendete Gensequenz Allel EC No. 1.1.1.37; 939bp; Position 72,86min; (Boyd et al., 1994).

Ähnliche Gründe führten auch zur Auswahl des Prolinpermease-Gens (*putP*), welches von Nelson und Selander (Nelson and Selander, 1992) bereits für evolutionäre Untersuchungen bei *E. coli* und *Salmonella* verwendet wurde (Lecointre et al., 1998). Das *put*-Operon ist bei Minute 23,25 des *E. coli*-Chromosoms lokalisiert und umfasst neben einer Kontrollregion, einem Segment von annähernd 420 bp, noch den Locus für das *putA*, einem zweiten Strukturgen dieses Operons. Die Prolinpermease ist ein integrales Membranprotein, welches den Transport von Prolin in die Zelle vermittelt. In der Membran wird es durch das bifunktionelle Enzym Oxidasedehydrogenase (*putA*) zu Glutamat umgewandelt, welches dann als N- und C-Quelle dient. Die Expression von *putP* (1.509 bp) ist durch die endogene Prolinkonzentration reguliert. Die Aktivität dieses Enzyms ist jedoch für das Bakterium nicht lebenswichtig.

Das *cad*-Operon liegt, entgegen der üblichen Leserichtung, bei Minute 93,5, also nahe dem *origin of replication* (*ori*). Es besteht aus den Genen *cadA* (2.183bp), *cadB* (1.335 bp) sowie *cadC* (1.539 bp), wobei sich zwischen *cadC* und *cadB* der Promotor für *cadBA* befindet – als *Pcad* bezeichnet. Dieses Operon kodiert eine Lysindecaboxylase (*cadA*) und ein hydrophobes Protein, welches als Lysin-Cadaverin-Austauscher (*cadB*) dient. Direkt stromaufwärts liegt *cadC*, das jedoch getrennt transkribiert wird und ein Protein kodiert, welches für die *cadBA*-Transkription erforderlich ist. Die Funktion der Lysindecaboxylase besteht darin bei niedrigem externen pH-Wert durch die Decarboxylierung, CO₂ und das biogene Amin der Aminosäure Lysin, Cadaverin, zu bilden. Auf diese Weise wird der externe pH-Wert erhöht. CadC ist ein zytoplasmatisches Membranprotein, dessen N-Terminus im Cytoplasma lokalisiert ist und als ein positiver Regulator an die *cadBA* Promotorregion (Neely and Olson, 1996) bindet. Die extrazelluläre Domäne hingegen ist wahrscheinlich an der Signalübertragung bei niedrigem externen pH beteiligt (Neely et al., 1994; Casalino et al., 2003). Cadaverin, also das Produkt der Lysindecarboxylierung, ist hingegen ein negativer Regulator der *cadBA*-Expression.

Der *trpC*-Locus (Acc. No. AE000224), ebenfalls Bestandteil eines Operons, stellt das vierte Housekeeping-Gen dar, welches im Rahmen dieser Arbeit sequenzanalysiert wurde. Das Gen umfasst 1.359 bp und befindet sich bei Position 28,37 min im *E. coli*-Chromosom

(Sharp, 1991; Milkman and Bridges, 1993). Bezüglich dieses Gens lagen bereits umfangreiche phylogenetische Untersuchungen vor, die u.a. die Analyse von ECOR-Stämmen beinhalteten (Lecointre et al., 1998). Das *trpC*-Genprodukt ist Bestandteil eines bifunktionalen Enzyms, welches zusammen mit dem Transkript von *trpB* (Tryptophan-Synthetase B-Untereinheit) einen Multi-Enzymkomplex im Rahmen der Tryptophanbiosynthese von *E. coli* bildet.

In den vorliegenden Untersuchungen konnte mittels Multilocus-Sequenztypisierung (MLST) bei allen vier Loci eine deutliche Rahmenbildung nachgewiesen werden. Das Phylogramm, in dem die Daten aller Housekeeping-Gensequenzdaten zur Sicherung der Ergebnisse kombiniert wurden, ergab hierbei sechs Cluster (Äste). Die untersuchten Stämme sind phylogenetisch eng verwandt. Die geringste Anzahl von Veränderungen war beim *trpC* nachweisbar. Das mag darin begründet sein, dass das *E. coli*-Chromosom nur eine Kopie dieses Gens aufweist, weshalb das Gen hochkonserviert ist. Es wurden keine Ereignis-Akkumulationen nachgewiesen (Milkman and Bridges, 1993). Um so höher sind die Abweichungen bei den Stämmen IHIT0072 (O157:H7) und dem zweiten γ -Intimin-positiven Stamm IHIT0304 (O145:H28) hier zu bewerten. Weiterhin ist auch der Stamm 72/90-56 (O5:K⁻) hierbei interessant, da er mit seinen Mutationen gleichfalls deutlich von den anderen Stämmen abweicht.

Bei den anderen Stämmen ist eine Unterscheidung kaum möglich. Lediglich die Stämme RW1372 und RW1374 (O103:H2) sowie IHIT2087 (O26:H11), die im Vergleich zum Locus des Stammes MG1655 eine Mutation mehr aufweisen, weichen davon ab.

Deutlicher ausgeprägt ist die Clusterbildung hingegen bei *cadB*. Ein wesentlicher Grund dafür ist die völlig unerwartete Tatsache, dass sieben Stämmen dieses Gen fehlt. Hierbei handelt es sich zum einen um alle Stämme mit der Intiminv variante η ³² (IHIT0554, IHIT0597, IHIT2115 und IHIT2430), der Intiminv variante γ (IHIT1703, Serovar O111:H2) sowie um die beiden Stämme mit den ζ -Intiminen (IHIT1190 und IHIT1968). Interessanterweise ist das Fehlen dieses Gens bei *Shigella spp.* mit einer gesteigerten Virulenz der jeweiligen Stämme assoziiert (Maurelli et al., 1998). Dies wurde durch Rekonstituierung des *cad*-Operons bewiesen, denn diese führte zu einer Virulenzreduktion der jeweiligen *Shigella*-Stämme im Kaninchen-Infektionsversuch. Es scheint sich hier um einen bislang in der Literatur wenig beachteten Mechanismus der Virulenzsteigerung zu handeln, nämlich, dass in einem „dynamic“ Genom nicht nur der Erwerb im Sinne einer Pathogenitätsinsel, sondern auch der Verlust von DNA-Regionen zur Virulenzsteigerung führen kann. Durch die

³² ε_2 -Int

fehlende Bildung von Cadaverin in *cadA*-negativen *Shigella*-Stämmen fällt die Enterotoxin-inhibierende Wirkung des Cadaverins weg (Blum et al., 1994; Hacker et al., 1997; Maurelli et al., 1998). Offensichtlich unterliegt dieses Enzym im Laufe der Evolution einem stärkeren Selektionsdruck, als dies bei unserer Auswahl des entsprechenden Zielgens bekannt war.

Interessanterweise zeigen bezüglich dieses Gens der O157:H7-Stamm IHIT0072 und der K12-Referenzstamm MG1655 die höchste Verwandtschaft. Zusammen mit den Serovaren O145:H28 (IHIT0304), O5:K⁻ (72/90-56) und noch etwas entfernter On.t.:H⁺ (IHIT0608) bilden diese Stämme ein eigenes Cluster. Die anderen untersuchten *E. coli*-Stämme zeigten hingegen keine Unterschiede. Keine der ermittelten Distanzen (Programm distances HUSAR, DKFZ, Heidelberg) ergab eine Identität unterhalb von 95 %. Auch dieses Ergebnis weist darauf hin, dass *cadB* über Rekombination zwischen einzelnen *E. coli*-Linien ausgetauscht werden kann.

Obwohl die MLST diese deutlichen Unterscheidungen offenbart, treten die Häufungen von Nukleotidänderungen bei den Housekeeping-Genen derart langsam auf, dass die allelen Profile von Bakterienisolaten über die Zeit genügend stabil bleiben. Sie sind demnach ideale Zielsequenzen für weltweite epidemiologische Studien. Ein weiterer Vorzug ist die relativ leichte Probengewinnung, die sogar aus klinischem Material heraus möglich ist. Auch die vorliegenden Ergebnisse belegen, dass genotypische Methoden zur Erregertypisierung präziser sind als phänotypische. Nicht zuletzt ist zu bedenken, dass unter bestimmten Bedingungen die Expression der zu untersuchenden phänotypischen Eigenschaften stark variieren kann.

Neben den o.g. Housekeeping-Genen wurden in der vorliegenden Arbeit auch Virulenzgene mittels MLST analysiert. Dieser ungewöhnliche Ansatz wurde gewählt, um die phylogenetische Entwicklung der mobilen Pathogenitätsinsel *Locus of enterocyte effacement* (LEE) nachzuvollziehen. Ziel war es, Hinweise auf Zeitpunkte zu erhalten, wann die mutmaßlich verschiedenen LEE-Pathogenitätsinseln in die jeweiligen unterschiedlichen *E. coli*-Phylotypen inseriert wurden. Als Zielgene dienten zwei Gene des LEE, das *eae*- und das *espB*-Gen. Diese beiden Gene waren die ersten identifizierten Gene, die als maßgeblich für die Fähigkeit von *E. coli*-Stämmen zur Auslösung der Attaching-and-Effacing-Läsion (AE-Läsion) (EPEC) identifiziert wurden (Clarke et al., 2003). Während die Funktion des *eae*-kodierten Intimins sehr gut charakterisiert ist, ist die Rolle des *espB* kodierten EspB weiterhin nicht restlos geklärt. Zwar wird das Protein in die Zielzelle sezerniert, aber seine Rolle ist

nicht detailliert bekannt. Zur Zeit wird davon ausgegangen, dass dieses Protein zusammen mit EspD einen Komplex bildet, der in die Wirtszellmembran integriert eine Pore formt. Durch diese Pore werden dann schließlich die bakteriellen Effektorproteine in die Wirtszelle eingeschleust. Diese Überlegungen beruhen aber eher auf Homologieschlüssen zu anderen bakteriellen Effektorproteinen, nämlich den *Yersinia*-Proteinen YoB/D, als auf experimentelle Daten (Clarke et al., 2003). Das Intimin ist ein äußeres Membranprotein, welches für die Anheftung der Bakterien an die Darmschleimhaut von entscheidender Bedeutung ist.

Die Auswahl dieser Gene wurde aufgrund zweier Aspekte getroffen. Zum einen deuten neuere Untersuchungen darauf hin, dass der Insertionsort der Pathogenitätsinsel LEE im jeweiligen *E. coli*-Genom mit der phylogenetischen Herkunft der jeweiligen Stämme assoziiert ist (Wieler et al., 1997; Sonnenberg and Whittam, 2001). Bislang sind drei Insertionsorte bekannt, die tRNA-Gene *selC*, *pheU* und *pheV* (Benkel et al., 1997; Jores et al., 2001; Rumer et al., 2003).

Zum anderen erhofften wir uns von der phylogenetischen Analyse des Genoms mittels MLST von Housekeeping-Genen im Zusammenhang mit der Analyse von zwei Genen des LEE, weitere Auskünfte über die Entstehung der LEE-positiven STEC-Stämme. Aus den in dieser Arbeit ermittelten Daten konnten weitere Hinweise auf einen Zusammenhang zwischen Pathovar, d.h., Intimintyp und Serogruppe gemacht werden.

Das *eae*-Gen diente bereits weiterhin mehreren Autoren als Zielgen, was zur Beschreibung von inzwischen insgesamt fünf Intiminsubtypen (α -Int, β -Int, γ -Int, δ -Int und ε -Int) führte. Stämme, die dem von McGraw (McGraw et al., 1999) etablierten Cluster EPEC1 angehören, bilden typischerweise α -Intimin, die Stämme der EPEC2-Gruppe β -Intimin. Diese EPEC sind verschieden von jenen, welche das γ -Intimin bilden. Das γ -Intimin wird im übrigen von den EHEC-Stämmen der Serovar O157:H7 sowie deren Vorläuferstämmen der EPEC-Serovar O55:H7 gebildet (Wolf, 1997; Adu-Bobie et al., 1998a; Jores et al., 2003). Unterschiede in den Sequenzen innerhalb des Intimin-Gens finden sich vor allem am C-terminalen Ende (10 % Sequenzunterschied zwischen α -Intimin des EPEC Stammes E2348/69 und des β -Intimins innerhalb der C-terminalen Region). Dagegen ist die N-terminale Region hochgradig konserviert. Identische *eae*-Gene wurden von Whittam et al. auch bei den EPEC2-Stämmen DEC11a (O128:H2) und DEC12a (O111:H2) durch MLEE identifiziert (Whittam, 1998). Die Mosaikstruktur des Gens erleichtert den Austausch von Abschnitten, die in Kontakt zur Wirtsabwehr stehen, ohne erheblichen funktionellen Verlust.

Dieses Phänomen ist eine Erklärung für die starken Variationen. Das Intimin ist als ein äußeres Membranprotein einem hohen Selektionsdruck unterworfen, entweder durch Anpassungsveränderung an äußere Reize, wie zum Beispiel auch unterschiedliche Wirte (Tzipori et al., 1995) oder durch antigene Veränderungen, um der Immunabwehr des Wirtes auszuweichen bzw. diese zu unterlaufen. Da die C-terminale Domäne den Rezeptorbindungsanteil kodiert (Frankel et al., 1994), ist verständlich, dass auch die größeren Unterschiede in dieser Region liegen, obgleich die Rezeptoren für das Intimin keine eukaryotischen (Yu and Kaper, 1992) sind, sondern ein Protein, das gleichfalls von den Bakterien gebildet und in die Oberfläche der eukaryotischen Zellen eingeschleust wird (Kenny et al., 1997). Es handelt sich um den ebenfalls vom LEE kodierten *Translocated Intimin Rezeptor* (Tir). Auch dessen Gen, das stromaufwärts des *eae* liegt, zeichnet sich durch eine ausgeprägte Diversität aus. Aufgrund des funktionellen Zusammenhangs zwischen Tir und Intimin muss hier von einer Co-Evolution ausgegangen werden. Die Mosaikgen-Struktur erleichtert auf jeden Fall Mutationen oder Rekombinationen, die zu Variationen in der Aminosäuresequenz führen, ohne dass funktionelle Änderungen entstehen müssen. Hierfür gibt es auch andere Beispiele. Zu nennen sind unter anderem die Antibiotikaresistenz von *Neisseria meningitidis* gegenüber Penicillin, die Phasenvariation des Flagellin (Li et al., 1994), oder die große Zahl der somatischen O-Antigene von *Salmonella enterica* (Reeves, 1992).

Auch die Unterschiede bei den übrigen in dieser Arbeit analysierten Intiminen konzentrierten sich auf das 3'-Ende der Gene (Yu and Kaper, 1992; An et al., 1997; Oswald et al., 2000), so dass sich zu den fünf bekannten (s.o.: α -Int, β -Int, γ -Int, δ -Int und ϵ -Int) Intimine, drei weitere neue Intimine aus dieser Arbeit anschließen (ζ -Int, η -Int und θ -Int sowie Variationen des γ -Int). Bei dieser Nennung von acht Typen sind die durch Antikörper (Agin and Wolf, 1997) differenzierten Subtypen α_1 , α_2 , β_1 , β_2 , γ_1 und γ_2 nicht berücksichtigt. Diese nachgewiesenen neuen Intimine, ζ -Intimin wurden bei den EHEC Serogruppen O84, O92, O119 sowie O150 gefunden, die η -Int Typen bei den Serogruppen O4, O80, O157 und On.t., welche alle unbeweglich sind und die Variation möglicherweise des α -Int bei der Serogruppe On.t. (mutmaßlich aber auch ein weiteres gänzlich neues Intimin), hier als θ -Int (theta) bezeichnet, sind phylogenetisch eher zu den Stämmen des γ -Intimintyps zu zählen.

Die Sequenzlängen der verschiedenen Intimine sind dabei deutlich unterschiedlich mit Längen von 889 bp bei den Vertretern der ϵ -Int Gruppe, über die Gruppe der β -Int mit 862 bp hin zum ζ -Intimin mit 858 resp. 859 bp und schließlich der γ -Int Gruppierung mit 850 bp, 847 bp sowie 846 bp. Die Ergebnisse der vorliegenden Arbeit stellen den überzeugenden Beleg dafür dar, dass tatsächlich klare Assoziationen zwischen den einzelnen Intimintypen und dem Insertionsort des LEE bestehen. Dies wurde auch durch die Beschreibung der drei bis dato unbekannt Intimintypen unterstützt (Jores et al., 2003; Rumer et al., 2003).

Das Intimin ist ein Protein mit einer hochvariablen Region am C-terminalen Ende, auf der die o.g. Einteilung in Typen nach griechischen Buchstaben beruht. Weiterhin weist das Intimin eine hochkonservierte N-terminale Region auf (Louie et al., 1993; Louie et al., 1994; Agin and Wolf, 1997; Adu-Bobie et al., 1998a; Louie et al., 1998; Morabito et al., 1998; Boyd, 1999; McGraw et al., 1999). Zudem ist es ein Mosaikgen, d.h., es besteht aus Anteilen unterschiedlicher phylogenetischer Herkunft, weshalb auch die spannende Frage der Herkunft des LEE wie auch die des Intimin-Gens *eae*, weiterhin ungeklärt bleibt. Mit großer Sicherheit kann allerdings von einer Übertragung von Teilen des Mosaikgens durch Phagen aufgrund der Insertionsorte und des G+C-Gehaltes ausgegangen werden.

Bisherige Untersuchungen anderer Autoren zeigen, dass sich einzelne *E. coli*-Pathovaren nach der Aufnahme von Virulenzgenen unabhängig voneinander entwickelten. So konnten mittels MLEE zwei große EPEC und zwei große EHEC-Cluster nachgewiesen werden (McGraw et al., 1999; Reid et al., 2000). Innerhalb dieser Gruppen bestehen wiederum Untergruppen, die anhand des LEE-Insertionsortes identifiziert werden können.

Eine andere Zielstruktur, die sich für phylogenetische Untersuchungen eignet, ist das Flagellin. Hier besteht eine starke Konservierung in der Flagellindomäne, was darauf hindeutet, dass die Geißel als Bewegungsorganelle bei Bakterien tiefe evolutionäre Wurzeln besitzt. Die *fli*-Gene gehören neben den *fla* und *flg* von *Salmonella typhimurium* und *Escherichia coli* zu den am besten untersuchten Genen. So führten Reid (Reid et al., 1999) Studien zur molekularen Evolution des Flagellins durch. Hierbei zeigte sich, dass die N- und C-terminale Region der Allele weitestgehend konserviert ist, wohingegen die zentrale Region polymorph und hochverschieden ist. Die Vergleiche der *fliC*-Sequenzen enthüllten auch bei diesem Gen eine Mosaikgenstruktur, die das Vorkommen eines horizontalen DNA-Transfers zwischen Stämmen anzeigt. Phylogenetische Analysen zeigten, dass die *fliC*-Sequenzen der

O157:H7- und O55:H7-Serovaren nahezu identisch sind und deutlich unterschiedlich zu denen aus *E. coli*-Stämmen, die H6- und H2-Flagellenantigene exprimieren. Als ein Ergebnis der fehlenden Flagellinexpression hat ein unbeweglicher sorbitol-fermentierender O157 Klon vermutlich schnell multiple Mutationen im *fliC* akkumuliert. Wang (Wang et al., 2000) stellten weitergehend Sequenzunterschiede bei *Escherichia coli* H7 *fliC*-Genen fest, die für ein molekulares Typisierungsschema für *E. coli* O157:H7 Stämme nutzbar sind. Aus diesem Grunde wurden in den vorliegenden Untersuchungen auch die *fliC*-Gene der *E. coli*-Stämme einbezogen. Es zeigte sich zum einen, dass jeder H-Typ eines *E. coli*-Stammes ein eigenes Restriktionsmuster bei Verdau des *fliC* mit der Restriktionsendonuklease ergab. Damit wurden die Daten von Fields (Fields et al., 1997) bestätigt. Interessant war aber insbesondere, dass die jeweiligen *fliC*- bzw. H-Typen mit den phylogenetischen Gruppen korrelierten. Der Grund für diese Korrelation ist spekulativ (eine Sequenzanalyse und erweiterte MLST kann Teil einer weiterführenden Arbeit sein). Aber es bietet sich die Hypothese an, dass die Flagellin-Gene, die ja ebenfalls einen TypIII-Sekretionsmechanismus darstellen, mit dem LEE co-evolutioniert sind. Eventuell könnte das TypIII-Sekretionssystem des LEE auch aus dem der Flagellen hervorgegangen sein. Da in der vorliegenden Arbeit jedoch in die Analysen keine Gene des TypIII-Sekretionssystems des LEE einbezogen wurden, gibt es für diese Hypothese keinen Beweis. Manche Autoren gehen davon aus, dass der LEE selber eine Mosaikstruktur aufweist (Cid et al., 2001). Dies wäre ein Hinweis darauf, dass einzelne Einheiten des LEE gegeneinander austauschbar wären, was wiederum die eigene Hypothese unterstützt.

Die Proteine, die vom LEE-kodierten TypIII-Sekretionssystem ausgeschleust werden, werden als „*E.coli* secreted proteins“³³ bezeichnet. Derzeit sind fünf solcher Proteine bekannt (EspA, B, D, E, und F³⁴); (Clarke et al., 2003). Die Gene sind stromabwärts des *eae*-Locus lokalisiert. Diese sezernierten Proteine beeinflussen nach Injektion in die Wirtszelle deren Signaltransduktion. Das vom *espB* kodierte EspB war das erste Protein, von dem eine intrazelluläre Lokalisation beschrieben wurde (Donnenberg et al., 1993; Foubister et al., 1994; Kenny et al., 1995; Donnenberg et al., 1997a; Donnenberg et al., 1997b; Deibel et al., 1998; Ebel et al., 1998; Kresse et al., 1999). Mutanten des *espB* sind nicht in der Lage, eine vollständige Attaching and Effacing (AE)-Läsionen auszubilden. Weiterhin ist das EspB

³³ sezernierte Proteine, die in die Signalübertagung des Wirtes verwickelt sind

³⁴ durchschnittliche Ähnlichkeit zwischen EPEC und EHEC liegt bei 93,9 %

hoch immunogen und die bislang charakterisierten Allele besitzen eine ausgeprägte Diversität.

Hochkonserviert sind die ersten fünf (N-terminal: MNTID) und die letzten vier (C-terminal: RLAG) Aminosäuren des EspB. Auch beim *esp*-Gen handelt es sich um eine Aneinanderreihung von drei Genen (siehe *cad*-Operon), die eine Mosaikstruktur erkennen lassen (homologe Proteine des Typ III Systems bei anderen Spezies). Für die Sequenzanalyse wurde ein Amplikon von ca. 500 bp gewählt, wobei die Auswahl der Oligonukleotidprimer anhand konservierter Regionen getroffen wurde. Das *esp*-Operon (ORF *espA*, *espD* und *espB*) liegt ca. 6.000bp downstream des *eae* in der gleichen Leserichtung auf dem LEE. Dieses Gen wurde zum einen wegen seines gemeinsamen Auftretens mit dem *eae* und der Virulenzsteigerung und zum anderen wegen der großen Datenmenge ausgewählt, die für eine Abschätzung der Ähnlichkeiten notwendig ist. Aber auch für eine ausreichende Primergenerierung sind möglichst viele Sequenzdaten erforderlich. Trotzdem konnte bei fünf Stämmen (IHIT0067 537/89-1 (O84:NM), IHIT3669 (O84:H2), IHIT1190 (O92:NM), IHIT1968 (O119:H25) sowie IHIT3000 (O150:NM)) kein Amplikon synthetisiert werden, so dass bei diesen Stämmen die *espB*-Gene vollständig sequenzanalysiert wurden. Die Analyse dieser fünf *espB*-Gene war auch insofern interessant, als dass die Gene einen 150 bp längeren Leserahmen besaßen (950 statt 800 bp) als dies bis zu diesem Zeitpunkt in der Literatur bzw. den Datenbanken bekannt war. Weiterhin besaßen diese Stämme alle ein Typ ζ (zeta) Intimin.

Die phylogenetische Analyse der *espB*-Gene ergab jedoch eine geringere Auftrennung der einzelnen Cluster als jene der *eae*-Gene. Das bereits bei der *eae*-Sequenzanalyse gefundene β -Cluster war zwar ebenfalls deutlich abgesetzt, beinhaltete aber zusätzlich das ε -Cluster mit der Variante des η -Typs. Dies spricht dafür, dass die verschiedenen Intimin-Typen eine schnellere Entwicklung genommen haben als die des EspB. Es besteht aber auch die Möglichkeit, dass der analysierte DNA-Bereich des *eae*-Gens einer höheren Rekombinationsrate ausgesetzt ist. Da *eae* ein Mosaikgen darstellt, ist diese Möglichkeit durchaus vorstellbar. Allerdings sind die hier durchgeführten Analysen nicht umfangreich genug, um diese Annahmen zu be- oder widerlegen.

Ein überraschendes Ergebnis lieferte Stamm IHIT0608 (O_{n.t.}:H+), für den die Analyse einen isolierten Cluster ergab. Eine weitere überraschende Singularität bezüglich seiner Ähnlichkeit/Verwandtschaft fand sich auch bei IHIT1703 (O11:H2), dessen Intimin als eine Variation des γ -Intimintyps identifiziert wurde. Die beiden anderen γ -Intimin-Stämme IHIT0072, ein klassischer O157:H7-Stamm sowie der O145:H28-Stamm IHIT0304 sind deutlich näher am β -Cluster lokalisiert als dies nach der *eae*-Typisierung zu erwarten war. Die noch verbleibenden Stämme des auch beim *eae* durch eine auffällige Homogenität gekennzeichneten Stämme des ζ -Clusters sind sogar zu 100 % identisch.

Tatsächlich stimmen die *eae*-Daten nicht vollständig mit den Clusteranalysen der *espB*-Gensequenzen überein. Dies ist wahrscheinlich auf die Mosaikgenstruktur des *eae*-Gens zurückzuführen, da sich dessen Anteile unabhängig voneinander entwickelt haben. Beide Genprodukte sind jedoch aufgrund ihrer innigen Interaktion mit der Wirtszelle gegenüber der Wirtsabwehr sehr stark exponiert, so dass aus Sicht der Immunselektion durchaus eine parallele Entwicklung zu erwarten gewesen wäre. Es besteht aber auch die Möglichkeit, dass die direkte Insertion von EspB in die Wirtszelle eine geringere Immunantwort auslöst als dies beim Intimin der Fall ist. Intimin ist nämlich ein äußeres Membranprotein, welches ständig mit den Wirtszellen interagiert und bereits in der frühen Phase der Besiedlung für die Anheftung der STEC-Bakterien verantwortlich ist. Deshalb finden sich in Seren von rekonvaleszenten Rindern und Menschen auch hohe Antikörpertiter gegen Intimin (Jenkins et al., 2000).

Im Prinzip ist jedoch eine ähnliche Aufteilung wie bei dem *eae*-Locus zu erkennen. Auch hier finden sich die dominantesten Intimintypen wieder zusammen in einer Gruppe. Die Gruppe der β -Intimine umfasst wiederum die gleichen Stämme, diesmal liegen jedoch auch die Stämme innerhalb der "Groß"-Gruppe, die das ε -Intimin bzw. Variationen davon tragen (Abb. 13). Davon grenzt sich deutlich eine Gruppierung solcher Stämme ab, die γ -Intimine besitzen. Eine Ausnahme stellt der Stamm IHIT1703 (O111:H2) dar, der nicht mehr so eng bei dem Cluster der γ -Intimine liegt. Eine weitere Sonderstellung nimmt nun auch der Stamm IHIT3641 (O17,77:H18) ein, der nicht mehr bei dem Cluster der β/ε -Int Stämme liegt, sondern innerhalb der Stämme des ζ -Intiminclusters. Diese interessante Konstellation ist Anlass für die zukünftige Untersuchung weiterer Stämme dieses Typs. Der nicht-typisierte Stamm IHIT0608 besitzt die größten Übereinstimmungen in seiner *espB*-Sequenz mit dem δ -

Int Stamm 4221 (U66102) (An et al., 1997) und dem α -Int Stamm E2348/69 (Elliott et al., 1998). Die ζ -Intimgruppe liegt auch hier mit nahezu vollständiger Identität zusammen.

Wegen der starken Sequenzunterschiede des *espB*-Gens dieser Stämme konnte eine eindeutige Zuordnung nur vorgenommen werden, nachdem das gesamte Gen sequenzanalysiert wurde. Dabei zeigte sich neben den deutlichen Unterschieden auch, dass das Gen einen deutlich kürzeren Leserahmen besitzt als dies bei den anderen *espB*-Genen der Fall war. Dies bestätigt die hohe Variationsbreite der *espB*-Gene und unterstützt die Annahme, dass es sich bei diesem Gen ebenfalls um ein Mosaikgen handelt (Cid et al., 2001).

Bei den eigenen Untersuchungen zeigten sich deutliche Gruppierungen sowohl innerhalb der Housekeeping-Gene, mit Ausnahme des hochkonservierten *trpC* Locus, als auch bei den Virulenzgenen. Alle hier untersuchten Serovaren wurden durch die unabhängige Aufnahme der Virulenzfaktoren, welche auf dem LEE kodiert sind, zu Pathovaren unterschiedlicher phylogenetischer Abstammung mit einem ursprünglich gemeinsamen Virulenzmuster. Das heißt, dass Stämme einer Serovar nicht notwendigerweise in das gleiche Cluster fallen mit anderen Stämmen derselben Serovar. Das widerspricht auch der Annahme, dass Klone einer Kategorie von Pathovaren monophyletisch sein müssen. Diese für das Verständnis der Phylogenie und der Entstehung von Klonen wichtigen Ergebnisse stimmen auch mit denen von Whittam (Whittam, 1998) überein.

Die von Pupo (Pupo et al., 2000b) gefundene EHEC-Gruppe der O157:H7-Stämme konnte in der vorliegenden Arbeit um einen nicht-O157:H7-Stamm erweitert werden, den O145:H28 Stamm IHIT0304. Der andere hier analysierte O157-Stamm liegt als ein nicht-H7-Stamm jedoch in einem anderen Cluster. Das bestätigt die Annahme, dass das Flagellin-Gen eher einen Marker für die jeweilige phylogenetische Gruppe darstellt als die O-Gruppe. In die Untersuchungen von Pupo (Pupo et al., 2000a) gingen keine O157:NM-Stämme ein. Durch die deutliche Abgrenzung einiger Stämme (IHIT0554, IHIT0597, IHIT2430, IHIT2115, IHIT1703, IHIT1190 sowie IHIT1968; Abb. 2) wird die Diversität besonders deutlich.

Die Abweichungen im G+C-Gehalt wie auch im Codon-Nutzungsgebrauch zeigen deutlich (Sharp and Li, 1987b; Sharp and Li, 1987a; Sharp, 1991; Bäumlner, 1997; Bäumlner et al., 1997; Smith and Smith, 1998), dass der LEE eine durch horizontalen Gentransfer aufgenommene DNA darstellt, also Fremd-DNA unbekanntem Ursprungs ist (Rumer et al., 2003). Die Herkunft der untersuchten Virulenzgene bleibt jedoch nach wie vor im Dunkeln,

und solange noch keine Herkunft bekannt ist, kann auch nur ab dem Zeitpunkt des Eintritts in einen *E. coli*-Klon eine Einschätzung der Evolution des jeweiligen Phylotyps vorgenommen werden. Der Hinweis auf einen solchen horizontalen Transfer bedeutet, dass empfängliche Bakterien einen neuen komplexen multigenen DNA-Block in das Chromosom in einem einzigen Evolutionsschritt erlangen können. Diese Möglichkeit, also die Aufnahme einer Vielzahl von zusätzlichen Virulenzfaktoren, führt zu einer sprunghaften Neuentwicklung eines Pathogens. Die Anordnung verschiedener Gene aus einer Vielfalt von natürlich vorkommenden *E. coli*-Stämmen hat das Vorkommen von Mosaiksegmenten (Bisercic et al., 1991; Milkman and Bridges, 1993; Guttman and Dykhuizen, 1994) bei der DNA belegt. Diese Arbeiten offenbaren, wie bakterielle Linien entstehen und sich in der Natur verbreiten, Mutationen anhäufen und Teile des Genoms durch lateralen Gentransfer und Rekombination ersetzen (Whittam, 1996). Das zeigt auch der Aufbau der LEE Merkmale in einer Mosaikstruktur (Milkman and Bridges, 1993; LeClerc et al., 1999; McGraw et al., 1999). So können die starken Variationen des Intimin und des EspB leicht erklärt werden.

Beim EHEC Stamm O157:H7 konnte die Arbeitsgruppe um Whittam darstellen, dass die Aufnahme von Shigatoxin-Genen mittels eines Bakteriophagen in einem γ -Intimin-produzierenden Vorfahren (O55:H7) einen hoch virulenten Erreger entstehen ließ (Selander et al., 1986; Feng et al., 1998; Whittam, 1998). Durch eine schrittweise Aufnahme eines *stx2* Phagen und Mutation an dem Strukturgen *uidA*+92 entstand aus einem humanpathogenen O55:H7 Stamm (ohne *stx*, aber mit LEE) ein O157:H7-EHEC-Stamm. Nach der Theorie von Whittam et al. (Kaper and O'Brien, 1998) entstand nach Verlust der Motilität ein O157:NM Stamm. Dieser wurde dann von einem *stx1*-positiven transduziert, und verlor die *E. coli*-typische Eigenschaft der Sorbitol-Fermentation. Weiterhin entstand ein O157:H7-Stamm mit den folgenden Merkmalen: Stx2-pos., Stx1-pos, GUD-neg. und SOR-neg. (Whittam, 1996) (Pupo et al., 2000a). Unsere Analysen des O157:NM-Stammes bestätigen diese Annahme nicht.

Der Virulenzfaktorbesatz der untersuchten Stämme deutet auf eine potentielle Gefährdung des Menschen als Zoonoseerreger hin. Oswald (Oswald et al., 2000) beschrieb seine gefundenen O103 Stämme des ϵ -Int Typ als Serogruppen, die eng mit verschiedenen Erkrankungen beim Menschen verbunden sind. Den Befund, dass das γ -Int mit mehreren für den Menschen hochpathogenen Serogruppen (O111, O145 und O157) vorkommt, konnte gleichfalls bestätigt werden, wobei der Stamm IHIT0304, die Serogruppe O145:H28, im Gegensatz zum IHIT0072 (O157:H7) keine übereinstimmende Einordnung durch MLEE und MLST erfahren konnte. Die phylogenetische Einordnung erfolgte mit Hilfe der ausgewählten Housekeeping-Gensequenzen sowohl bei MLEE (Whittam) als auch MLST (diese Arbeit). Der Stamm IHIT1703 mit der Serogruppe O111:H2 ist dabei in seiner Sequenz dem Stamm von Voss (Voss et al., 1998), einem O111:NM sehr ähnlich (96 %), der von Whittam (Reid et al., 2000) als ein STEC2 über die MLEE eingruppiert wurde. Weiterhin bestätigte Whittam in dieser Arbeit auch den O157:H7 Stamm IHIT0072 als STEC1 Klon. Die Stämme der Serogruppen O26:NM, On.t.:H⁺, O153:H⁺, O145:H⁺, O5:K⁻, O26:H11, On.t.:H11, O17,77:H18 und die beiden O118:H16 Stämme wurden dem β -Int Typ zugewiesen. Whittam konnte darüber hinaus noch mit Hilfe der MLEE die Stämme IHIT0084 (On.t.:H⁺) und IHIT2087 (O26:H11) als STEC2 identifizieren (persönliche Mitteilung). Eine übereinstimmende Gruppierung erfolgte bei dem Stamm IHIT1703 (gamma-Intiminträger) sowohl mit der MLST als auch der MLEE. Daneben konnte keine weitere Zuordnung in Gruppen über die MLEE durch Whittam erfolgen (persönliche Mitteilung).

Zur Analyse der Nukleotidsequenzen wurden in dieser Arbeit Parsimoniemethoden verwendet. Parsimoniemethoden versprechen den größten Erfolg, wenn die Merkmalsänderung von geringem Ausmaß ist. Selbst die Proteinebene stellt sich in dieser Arbeit als zu einheitlich dar, um mit Distanzmethoden zu arbeiten, mit denen keine verlässlichen Aussagen getroffen werden können. Dies gilt auch für die Nukleinsäureebene bezüglich der LEE-kodierten Gene, die horizontal erworben wurden. Bei einigen Stämmen wurde eine vollständige Identität festgestellt. Für einen phylogenetischen Baum ergeben sich daraus Probleme der Darstellung. Identische Sequenzen sind auf einem Ast gelegen und damit ohne weitere Aussagekraft. In den Tabellen (Tabelle 8 bis 13) sind die identischen sowie die abweichenden Sequenzen aufgelistet, auf denen die Generierung des Baumes beruht. Zur Erstellung wurde ein Alignment mit den Clustal zugrundeliegenden Algorithmen hergestellt. Im Anschluss daran wurde ein *bootstrapping* durchgeführt. D.h., es werden Replikanten erstellt, und dieser Vorgang wird bis zu 1000mal wiederholt, um eine optimale Übereinstimmung zu erzielen. Je häufiger der jeweilige identische Baum entsteht – dies ist in

den Graphiken anhand der Prozentzahlen 1-100 erkennbar – desto wahrscheinlicher stimmt der Baum mit der tatsächlichen Phylogenie überein. Diese Methoden setzen eine weitgehend konstante Zahl von Mutationen pro entwicklungsgeschichtlichem Intervall über den gesamten Baum voraus. Die eigentliche Analyse bei diesen sehr ähnlichen Sequenzen erfolgte abschließend im Vergleich mit den Sequenzen der Housekeeping-Gene, welche prospektiv im originären *E. coli*-Chromosom gelegen sind. Wie die Analysen belegen, sind die oberflächenexponierten Genprodukte Intimin und EspB wesentlich diverser, so dass deren Diversität diejenige der Housekeeping-Gene stark überlagert. Deshalb kann mit den in der vorliegenden Arbeit wenigen Sequenzdaten – es wurden lediglich vier Housekeeping-Gene analysiert - keine Abschätzung bezüglich der zeitlichen Insertion des LEE in dem jeweiligen *E. coli*-Phylotyp getroffen werden. Unabhängig davon liegt zur Zeit kein verlässlicher Algorithmus vor, mit dem solche horizontal übertragenen Sequenzen adäquat berechnet werden können.

Da Parsimoniemethoden die Zahl der notwendigen Evolutionsschritte zu minimieren versuchen (*parsimony*-Methoden), bedingen sie bei den zum Teil sehr ähnlichen Sequenzen die beste Auflösung, sind also für die hier durchgeführten Datenanalysen am besten geeignet. Der beste Baum nach dem *parsimony*-Prinzip ist gleichzeitig der wahrscheinlichste unter den Kriterien der Maximum-likelihood. Die Neighbor-Joining Methode (Saitou and Nei, 1987) ist zwar weniger rechenaufwendig, eignet sich aber nur für erste orientierende Untersuchungen.

Die fehlenden Sequenzen und Sequenzabschnitte (Längen, d.h., „Basenlücken“) wurden jeweils zu vollständigen Datensätzen bei der Eingabe in MEGA, wie auch bei Clustal, aufgefüllt und entsprechend der Vorgaben der Programme eingegeben.