

Molecular Basis of the Indole-Negative Reaction in *Shigella* Strains: Extensive Damages to the *tna* Operon by Insertion Sequences

Ferdousi Rezwan,¹ Ruiting Lan,² and Peter R. Reeves^{1*}

School of Molecular and Microbial Biosciences, University of Sydney,¹ and School of Biotechnology and Biomolecular Sciences, University of New South Wales,² Sydney, New South Wales, Australia

Received 24 March 2004/Accepted 5 August 2004

The molecular basis of the loss of tryptophan utilization (indole-negative phenotype) of *Shigella* strains, in effect clones of *Escherichia coli*, was investigated. Analysis of the *tna* operon of 23 *Shigella* strains representing each of the indole-negative serotypes revealed that insertion sequence-mediated insertion and/or deletions damaged the *tna* operon, leading to inability to convert tryptophan to indole. These events differ for cluster 1, cluster 3, and the outlier *Shigella* strains, confirming our previous observation of independent origins of these lineages from within *E. coli*. Parallel loss of the trait and prevalence of indole-negative strains suggest that the trait is deleterious in *Shigella* strains and advantages those without it.

Shigella and *Escherichia coli* strains are often extremely difficult to separate biochemically because there are aerogenic (gas-producing) *Shigella* and lactose-negative, anaerogenic, nonmotile *E. coli* strains (9). Generally *Shigella* strains are biochemically less active than *E. coli* and regarded as metabolically inactive biogroups of *E. coli*, both being in one species based on DNA homology (3). The evolutionary relationships of 46 *Shigella* strains representing each of the serotypes belonging to the four traditional *Shigella* species (subgroups), *S. dysenteriae*, *S. flexneri*, *S. boydii*, and *S. sonnei*, have been shown by sequencing of housekeeping genes to be in three clusters with five outliers, indicating that the *Shigella* phenotype has arisen independently in several lineages within *E. coli* (19) (note that we prefer for *Shigella* and the four species names to not be italicized, since we contend that they are in effect forms of *E. coli*; however, this is not consistent with this journal's policy).

With few or no exceptions, *Shigella* strains do not use lactose, lysine is not decarboxylated, and they are nonmotile. The molecular bases of these three properties have been studied in *Shigella*. Ito et al. (11) studied the lack of lactose fermentation and showed that *S. flexneri* 1 and 3 and *S. boydii* 2 and 4 do not have any of the three genes of the *lac* operon, while *S. dysenteriae* 1 contains *lacY* and *lacA* but not *lacZ*, and *S. sonnei* has all three genes but with the *lacY* permease gene defective. Lysine decarboxylase (LDC) activity is present in ~90% of *E. coli* strains but is uniformly absent in *Shigella* strains. The lack of LDC activity in representative *Shigella* strains *S. flexneri* 2a, *S. dysenteriae* 2, *S. boydii* 14, and *S. sonnei* was found to be due to deletion of the *cadA* gene for LDC with additional insertions and rearrangement in different strains (7, 15). The above three properties are generally negative in *Shigella* strains. However there are others that are negative in certain serotypes or species, for example indole production and mannitol utiliza-

tion (9), and the genetic basis of these changes has not been studied.

Indole production is often used to differentiate *E. coli* from other indole-negative enteric bacteria because 96% of *E. coli* are indole positive, whereas many enterobacterial species are negative in the indole reaction. In the case of *Shigella*, the indole reaction is consistently negative only in specific serotypes within each traditional species of *Shigella*, including 7 of the 10 *S. dysenteriae* serotypes, 9 of the 15 *S. boydii* serotypes, 1 of the 6 *S. flexneri* serotypes, and *S. sonnei* (9). *S. flexneri* 1 to 4 are variable with regard to this trait. When the indole-negative serotypes are rearranged according to their genetic relationships based on housekeeping gene sequences, it was found (19) that all except four indole-negative serotypes are in cluster 1, which includes only one indole-positive serotype (*S. dysenteriae* 7). Of the four indole-negative serotypes not in cluster 1, three are outliers (*S. dysenteriae* 1 and 10 and *S. sonnei*) while *S. boydii* 12 is a cluster 3 serotype. It seems that the indole-negative phenotype has evolved several times independently in *Shigella* (19).

Indole is a product of tryptophanase, which converts tryptophan to indole plus pyruvate and ammonia. The tryptophanase operon (*tna*) of *E. coli* K-12 is 3,144 bp in length including intergenic regions, containing three genes, *tnaL*, *tnaA*, and *tnaB*. *tnaA* of 1,431 bp and *tnaB* of 1,248 bp are the major structural genes coding for tryptophanase and tryptophan permease, respectively. Upstream of *tnaA* is *tnaL*, encoding a 25-residue leader peptide. The *tna* promoter upstream of *tnaL* includes a 28-bp σ^{70} binding site as well as a 20-bp binding site for cyclic AMP receptor protein (16). The operon is involved in use of tryptophan as carbon and nitrogen sources and is under catabolic repression (16). The loss of this property in *Shigella* strains is another example of loss of metabolic activities by *Shigella*. To investigate the basis of the indole-negative property and evolution of this *Shigella* phenotype, the *tna* operon was sequenced for representatives of all consistently indole-negative *Shigella* serotypes.

PCR sequencing of the *tna* operon. The entire *tna* operon, including the three structural genes (*tnaL*, *tnaA*, and *tnaB*) and the promoter region, was sequenced for each strain, with ex-

* Corresponding author. Mailing address: School of Molecular and Microbial Biosciences, University of Sydney, NSW 2006, Australia. Phone: 61 2 9351 6045. Fax: 61 2 9351 4571. E-mail: reeves@angis.usyd.edu.au.

ceptions where only the *tnaB* gene could be amplified. Primers for PCR and sequencing were based on the *E. coli* K-12 sequence (GenBank accession numbers AE000447 and AE000448) and designed to give overlapping amplicons of approximately 1 kb in length. PCR products were purified using a Wizard PCR purification system (Promega, Madison, Wis.) or a MO BIO PCR Clean-Up kit (MO BIO Laboratories, Solana Beach, Calif.). Samples were sequenced using dye terminator technology (Perkin-Elmer Cetus, Norwalk, Conn.) through the Sydney University and Prince Alfred Hospital Macromolecular Analysis Centre and an automated 377 DNA sequencer (Applied Biosystems, Burwood, Victoria, Australia).

DNA sequences were assembled and edited by using the programs PHRED, PHRAP, and CONSED (10). Further analysis was undertaken using programs available from the Australian National Genomic Information Service at the University of Sydney. Sequence comparisons were done using MULTICOMP (20).

The strains representing 23 indole-negative serotypes were described previously (19). They are referred to as D1, B1, and F1A for *S. dysenteriae* 1, *S. boydii* 1, and *S. flexneri* 1A, respectively, and so forth.

Indole-negative property in all *Shigella* strains is due to damage by IS-mediated insertion and or deletion. All cluster 1 strains except D7 are indole negative. Sequencing revealed that the indole-negative strains have at least one *IS1* sequence inserted in *tnaL* at base 55 (Fig. 1), which disrupted the *tnaL* gene. In D3 there is an additional insertion in the *tnaA* gene in the opposite orientation. In D3, D4, D11, D12, and D13, the original insertion sequence (IS) was flanked by a nine-base direct repeat (GACAATAAG), characteristic of *IS1* insertion (14). In B2, B4, B14, F6, and F6A, the insertion is followed by a 1,001-bp deletion to base 777 of the *tnaA* gene. The insertion and deletion damages both *tnaL* and *tnaA*.

A third group of strains comprising B1, B3, B6, B8, B10, B18, and D5 have one partial 192-bp *IS1* sequence followed by a full *IS1* sequence. All except B3 and D5 have a deletion of 49 bp including 21 bp of the *tnaL* gene. The *tnaA* gene and part of *tnaB* from B3 could not be amplified. B3 also has an IS insertion in the *tnaB* gene. In D5 there has been a further IS insertion close to that inserted in *tnaL* at base 55; an IS insertion in the *tnaL/tnaA* intergenic region was found by PCR using an IS primer and a *tnaL/tnaA* intergenic region primer. Sequencing from the *tnaA* end revealed a junction between an *IS1* insertion and the intergenic region at base 29. The *IS1* insertion was in the opposite orientation to that at base 55. However, the structural arrangement in the region between the IS insertions at *tnaL* and the intergenic region could not be determined, and as it is unique to D5, is not needed to follow the course of the changes. Also we were not able to amplify the *tnaA* gene or the 5' part of the *tnaB* gene. The boundaries of this deletion were not determined.

Although the IS elements present in the three groups of cluster 1 strains have a different configuration at position 55, it is most likely that the initial insertion event occurred in the common ancestor before the groups diverged. The complete *IS1* sequences were nearly identical, with some base substitutions, as indicated in Fig. 2.

Cluster 3 has only one consistently indole-negative strain,

B12. An *IS1* sequence was found inserted two bases before the σ^{70} binding site of the promoter region, separating it from the cyclic AMP receptor protein binding site, which seems to be sufficient to disrupt the expression of the *tna* operon. All three structural genes were intact.

Three of the five outlier strains, D1, D10, and SS, are indole negative. We sequenced only part of *tnaB* gene of D1, as the rest could not be amplified. The D1 strain has a six-base in-frame deletion in the *tnaB* gene but in a different position (base 649 to 654) from that of the cluster 1 strains (see below). Analysis of the genome sequence of the D1 strain sequenced by the Sanger Centre revealed that it too had the 6-bp deletion, and also an *IS2* inserted at base 91 of *tnaB* and the *tna* operon up to the *IS2* insertion site in *tnaB* is deleted.

No product could be amplified from *S. sonnei*. Analysis of the genome sequence of the *S. sonnei* strain sequenced by the Sanger Centre indicates that the entire operon is deleted starting from 374 bp upstream of *tnaL* to 486 bp downstream of *tnaB* and is replaced by an *IS1* element that probably mediated the deletion.

Two insertion sequences were found in D10, one at the end of *thdF*, the gene immediately upstream of the *tna* operon, and one at base 333 of the *tnaB* gene. The insertion at base 333 also led to the deletion of six bases. Interestingly, both insertions were due to *ISSfl7* of the *IS5* family (4, 14). We were able to obtain a PCR product from base 575 of *tnaA* but unable to PCR amplify a product from the start of the *tna* operon to that segment of *tnaA*, suggesting that there had been some restructuring which we did not resolve.

Deletions in the *tnaB* gene. Apart from the IS insertions and associated deletions described above, several other events lead to further damage to the *tna* operon (Fig. 1). A single base deletion at position 19 of the *tnaB* gene was observed in seven strains: D3, D4, D6, D9, D11, D12, and D13, which generated a stop codon at bases 52 to 54. An in-frame six-base deletion of bases 706 to 711 of the *tnaB* gene was observed in strains B1, B3, B6, B8, B10, B18, and D5. As described above, a different six-base deletion of bases 649 to 654 in the *tnaB* gene was observed in the outlier strain D1. B14 has a base substitution in *tnaB* at position 149, which generates a stop codon.

Sequence of events in cluster 1 strains. Variation observed in this study and multilocus sequence data from housekeeping genes allow us to look at the evolution of the cluster 1 strains. Cluster 1 strains were grouped into three subgroups in a previous study of eight housekeeping genes (19), but there were only a few base changes to support the branching pattern. When all the data are considered together it is clear that the subgroups are well supported, and we refer to them here as subclusters 1a, 1b, and 1c (Fig. 2). Subcluster 1a consists of D3, D4, D6, D9, D11, D12, and D13, supported by three mutations in housekeeping genes and two in the *tnaB* gene. Subcluster 1b consisting of B2, B4, B14, F6, and F6a is supported by three unique base changes in housekeeping genes and the IS-mediated 1,001-bp deletion of the 5' end of the *tna* operon. Subcluster 1c, consisting of B1, B3, B6, B8, B10, B18, and D5, is supported by two base changes in housekeeping genes, a six-base deletion in *tnaB*, and a double IS insertion at base 55 of the *tnaL* gene. Further branching within each subcluster is evident. No reverse or parallel changes were observed, giving confidence to the tree presented in Fig. 2.

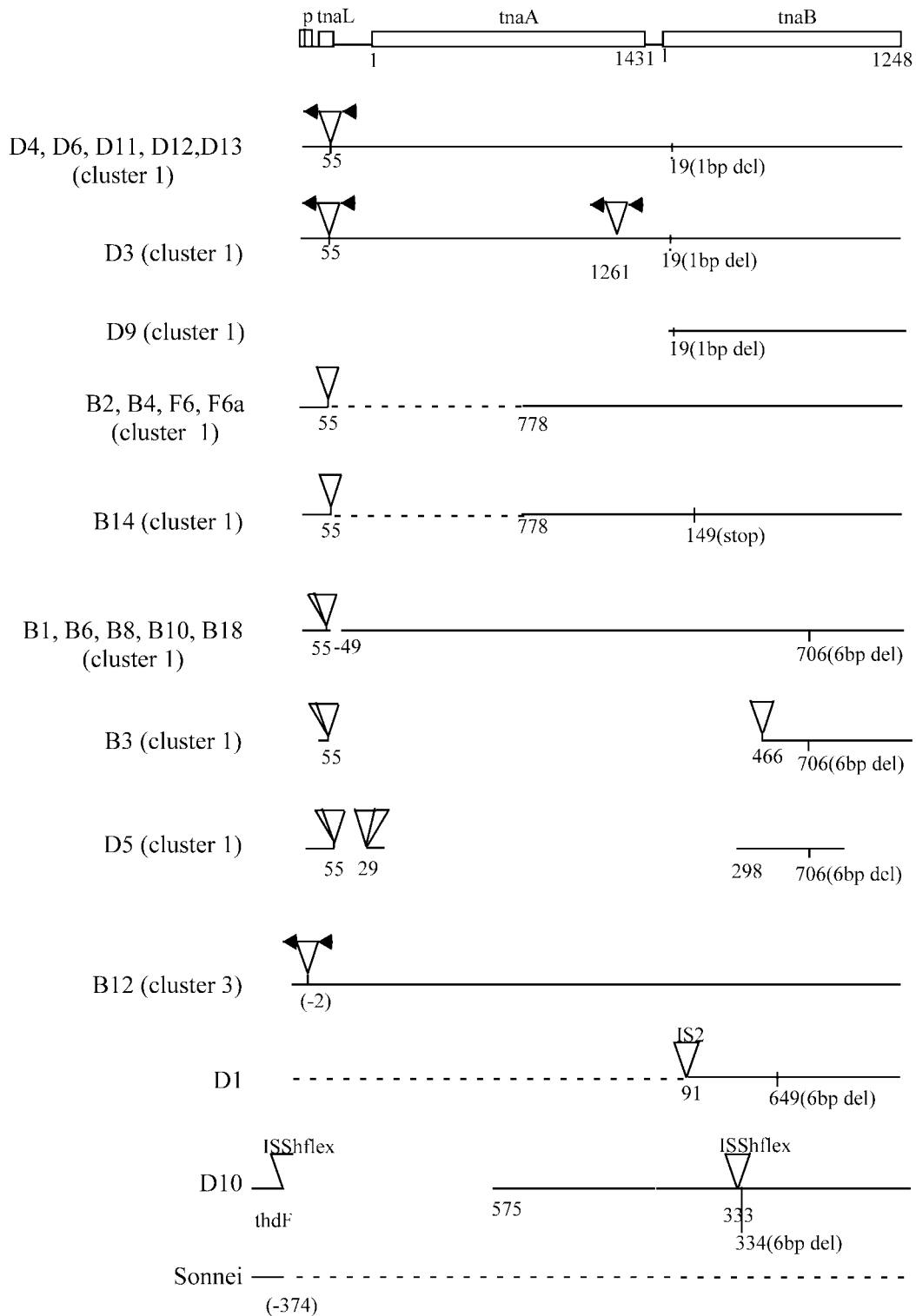


FIG. 1. Schematic diagram of the *tna* operon and genetic changes in *Shigella* strains. The map of the operon at the top is based on *E. coli* K-12. Below are shown 12 forms of the operon that can be distinguished in indole-negative *Shigella* serotypes. Data for D1 and *S. sonnei* are partly based on the incomplete genome sequences from the Sanger Centre. Blank regions indicate that PCR failed to amplify a fragment from that region while dotted lines indicate deletions with a defined end or ends. Triangles indicate IS insertions with base position marked below. Positions are given as the base pair in the gene, with negative values indicating the position relative to the start of the next gene or major feature (see the text). The name of the IS is also indicated if not *IS1*. Solid arrowheads at the ends of the IS indicate repeats of chromosomal DNA if present. For deletions the number of the first base is shown, with lengths in parentheses.

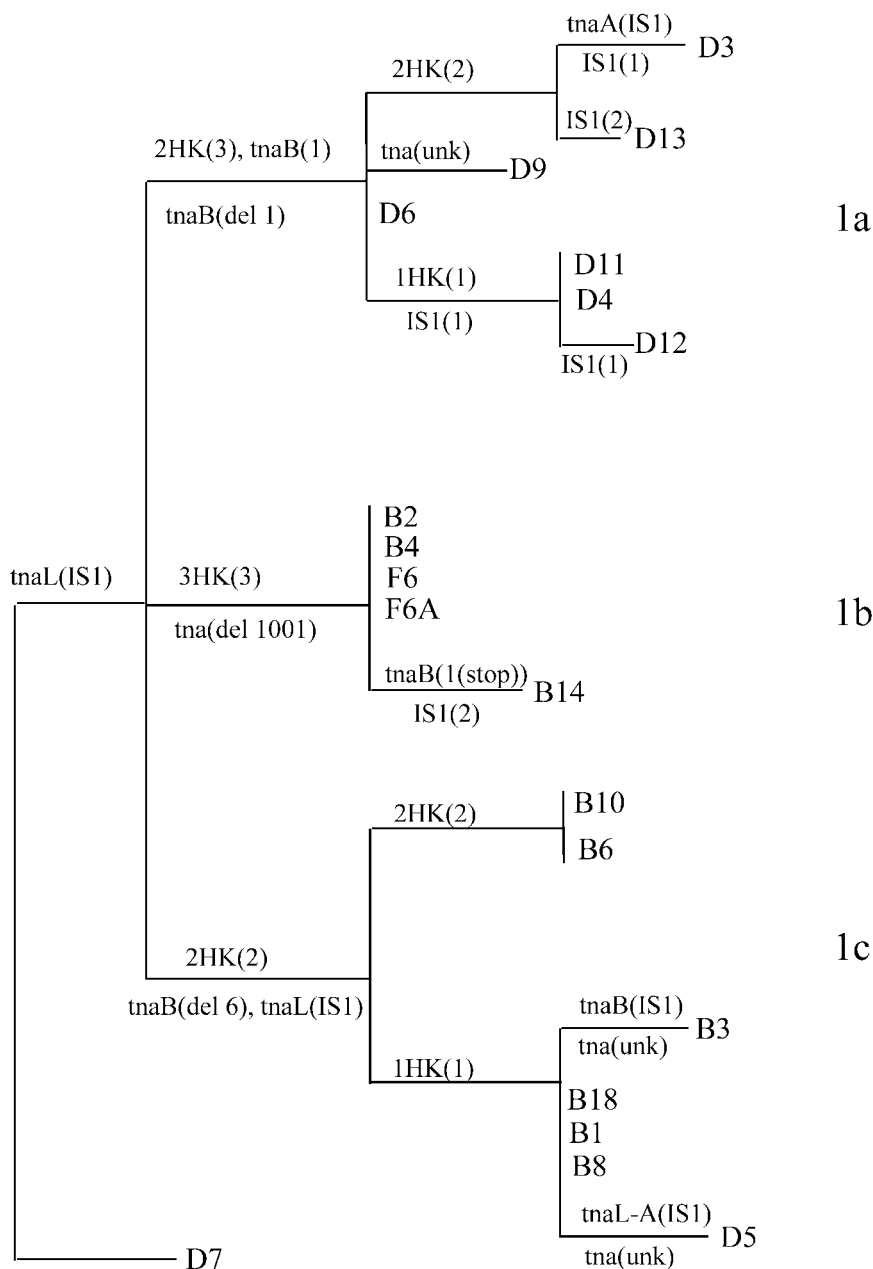


FIG. 2. Evolution of *Shigella* cluster 1 strains. A tree based on housekeeping gene sequences (19) is presented. Changes in housekeeping genes (19) and in the *tna* operon (this study) are indicated at the nodes. For example 2HK (3) means that three changes in two housekeeping genes supported that branch. Events in the *tna* operon are marked by the gene name with changes in parentheses. Regions that failed to amplify a product are treated as unknown events and shown as tna(unk). Changes that occurred in individual strains only are marked for the *tna* operon but not housekeeping genes, leading to the collapse of most external branches.

D7, the only indole-positive strain in cluster 1, is at the base of cluster 1 and in most genes studied appears to be almost unchanged from the common ancestor of cluster 1, with the *tna* operon still functional and only one housekeeping gene base change not present in all cluster 1 strains (19). The IS insertion at base 55 of *tnaL* is present in all cluster 1 strains except D7, indicating that it occurred early in the development of this group of serotypes. It is interesting that the D7 line did not

expand while the other 3 lineages (subclusters) expanded dramatically to give multiple serotypes.

Concluding comments. Analysis of the *tna* operon of strains representing each of the 25 *Shigella* serotypes that are consistently indole negative revealed at least nine IS-related events that damaged the *tna* operon, leading to a block in conversion of tryptophan to indole. These events are shown to be independent for cluster 1, cluster 3, and the three indole-negative

outlier strains, supporting our previous conclusion that these lineages have independent origins. Further, genetic changes in this operon in cluster 1 strains are consistent with a tree based on housekeeping genes, giving strong support for recognition of three subclusters in cluster 1.

The loss of catabolic functions in *Shigella* strains could be due to niche adaptation. *Shigella* strains occupy a very different niche from commensal *E. coli* and could be expected to shed functions not needed for life inside a eukaryotic cell (13), as a trade-off for optimal fitness (5). The parallel loss of a biochemical trait is indicative that the loss is likely to have an adaptive advantage (5, 6). This phenomenon of gene decay is also seen in other clones adapting to a new niche, for example the recently sequenced *Salmonella enterica* serovar Typhi genome has many pseudogenes (17) as does the *Yersinia pestis* genome (18). Both are clones adapted to live in a very specific part of the species niche (note that *Y. pestis* is a clone of *Yersinia pseudotuberculosis* [2]). However, *Shigella* strains of *E. coli* give us additional evidence, as there are several independent lineages and parallel loss is clearly evident for many traits.

With regard to reasons for frequent loss of the tryptophanase pathway in *Shigella* lineages, little is known of the effect of tryptophan utilization on host or host-pathogen interaction. Indole can act as an extracellular signal, activating a number of genes, and is toxic at high concentrations (23). In *E. coli*, the presence of tryptophanase has been shown to affect adherence to human epithelial cells and biofilm formation (8), showing that the enzyme can have effects on relevant functions. Four of the seven *Shigella* lineages within *E. coli* have lost the indole trait (D1, D10, *S. sonnei*, and the major part of cluster 1). In addition, one serotype of cluster 2 has lost it and several cluster 3 serotypes are often indole negative (9), suggesting that they are losing the trait. These include serotype 2, prevalent in both developed and developing countries. It appears that conversion of tryptophan to indole is deleterious for *Shigella* forms, such that those without it have an advantage. The most important lineages are affected. Cluster 1 can be viewed as very successful with expansion to have the most serotypes among the three clusters. *S. sonnei* is the major form in developed countries, while D1 is the major cause of epidemic-level disease in developing countries (1). However, it must be stated that any effect of tryptophan utilization and indole production on *Shigella* fitness or virulence is yet to be determined. Also, selection for loss is apparently not as strong as for loss of lactose fermentation, lysine decarboxylation, and motility, uniformly absent in *Shigella*, and for which the negative selection pressures are better understood (13). In the case of lysine decarboxylation, it has been shown that the property is very deleterious for *Shigella* (7, 15). The introduction of *cadA*, the gene encoding lysine decarboxylase, into *S. flexneri* 2a resulted in attenuation of virulence by inhibition of iron-regulated enterotoxins.

There are also several mutations and insertions in *tnaB*, and for cluster 1, where we have several subclusters and strong support for the tree as shown in Fig. 2, we can see that these occurred in strains that already lacked *tnaA* function. This could be due to a functional *tnaB* gene being deleterious in the absence of *tnaA*. The tryptophanase operon (*tna*) is induced by tryptophan and high levels of tryptophan are ordinarily required to obtain maximum *tna* operon induction (25). The

operon would be induced under conditions where there was significant level of tryptophan in the environment and “better” substrates for catabolism were not available. It is possible that under such circumstances the accumulation of tryptophan in the absence of tryptophanase is deleterious. Only a few cluster 1 strains, B2, F6A, F6, and B4, have no deletion, insertion, or frameshift mutation in *tnaB*, indicating that there is selection for loss of function.

IS elements are known to mediate various genetic rearrangements including inversions and deletions. Clearly IS elements were involved in most of the events affecting function of the *tna* operon in indole-negative *Shigella* strains. It is also evident that IS insertion lead to loss of other properties in *Shigella*. The loss of motility was found to be due to an IS insertion in the *flhD* operon in one case, and insertion of an IS element impaired the flagellin gene in another (21). This is perhaps not surprising as the *S. flexneri* 2A genome contains more than 300 IS elements and half of the invasion plasmid genome consists of IS sequences (4, 22). The IS content in the *S. flexneri* 2a genome is more than sevenfold greater than that of the non-pathogenic *E. coli* K-12 and enterohemorrhagic *E. coli* O157:H7 (12, 24). IS transposition seems to be a major force in the evolution of the *Shigella* phenotypes. The high level of IS transposition could be viewed as a mutator trait leading to acceleration of the adaptation under selection pressure for loss of function in a range of genes.

Nucleotide sequence accession numbers. The nucleotide sequences reported in this study have been deposited in the GenBank database (accession numbers AY746464 to AY746494).

This research is supported by a grant from the National Health and Medical Research Council of Australia. F.R. is supported by a University of Sydney postgraduate scholarship.

We thank the anonymous referees for comments and suggestions.

REFERENCES

- Acheson, D. W. K., and G. T. Keusch. 1995. *Shigella* and enteroinvasive *Escherichia coli*, p. 763–784. In M. J. Blaser, P. D. Smith, J. I. Ravdin, H. B. Greenberg, and R. L. Guerrant (ed.), *Infections of the gastrointestinal tract*. Raven Press, New York, N.Y.
- Achtman, M., K. Zurth, G. Morelli, G. Torrea, A. Guiyoule, and E. Carniel. 1999. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc. Natl. Acad. Sci. USA* **96**:14043–14048.
- Brenner, D. J. 1984. Enterobacteriaceae. In N. R. Krieg and J. G. Holt (ed.), *Bergey's manual of systematic bacteriology*, vol. 1. Williams & Wilkins, Baltimore, Md.
- Buchrieser, C., P. Glaser, C. Rusniok, H. Nedjari, H. D'Hauteville, F. Kunst, P. Sansonetti, and C. Parsot. 2000. The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. *Mol. Microbiol.* **38**:760–771.
- Cooper, V. S., and R. E. Lenski. 2000. The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* **407**:736–739.
- Cooper, V. S., D. Schneider, M. Blot, and R. E. Lenski. 2001. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *J. Bacteriol.* **183**:2834–2841.
- Day, W. A., R. E. Fernandez, and A. T. Maurelli. 2001. Pathoadaptive mutations that enhance virulence: genetic organization of the *cadA* regions of *Shigella* spp. *Infect. Immun.* **69**:7471–7480.
- Di Martino, P., A. Merieau, R. Phillips, N. Orange, and C. Hulen. 2002. Isolation of an *Escherichia coli* strain mutant unable to form biofilm on polystyrene and to adhere to human pneumocyte cells: involvement of tryptophanase. *Can. J. Microbiol.* **48**:132–137.
- Ewing, W. H. 1986. *Edwards and Ewing's identification of the Enterobacteriaceae*, 4th ed. Elsevier Science Publishers, Amsterdam, The Netherlands.
- Gordon, D., C. Abajian, and P. Green. 1998. CONSED—a graphical tool for sequence finishing. *Genome Res.* **8**:195–202.
- Ito, H., N. Kido, Y. Arakawa, M. Ohta, T. Sugiyama, and N. Kato. 1991. Possible mechanisms underlying the slow lactose fermentation phenotype in *Shigella* spp. *Appl. Environ. Microbiol.* **57**:2912–2917.

12. Jin, Q., Z. Yuan, J. Xu, Y. Wang, Y. Shen, W. Lu, J. Wang, H. Liu, J. Yang, F. Yang, X. Zhang, J. Zhang, G. Yang, H. Wu, D. Qu, J. Dong, et al. 2002. Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. *Nucleic Acids Res.* **30**:4432–4441.
13. Lan, R., and P. R. Reeves. 2002. *Escherichia coli* in disguise: molecular origins of *Shigella*. *Microbes Infect.* **4**:1125–1132.
14. Mahillon, J., and M. Chandler. 1998. Insertion sequences. *Microbiol. Mol. Biol. Rev.* **62**:725–774.
15. Maurelli, A. T., R. E. Fernandez, C. A. Bloch, C. K. Rode, and A. Fasano. 1998. “Black holes” and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **95**:3943–3948.
16. McFall, E., and E. B. Newman. 1996. Amino acids as carbon sources, p. 358–379. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella* cellular and molecular biology, 2nd ed., vol. 1. American Society for Microbiology, Washington, D.C.
17. Parkhill, J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. Holden, M. Sebahia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connor, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O’Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell. 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**:848–852.
18. Parkhill, J., B. W. Wren, N. R. Thomson, R. W. Titball, M. T. Holden, M. B. Prentice, M. Sebahia, K. D. James, C. Churcher, K. L. Mungall, S. Baker, D. Basham, S. D. Bentley, K. Brooks, A. M. Cerdeno-Tarraga, T. Chillingworth, A. Cronin, R. M. Davies, P. Davis, G. Dougan, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Leather, S. Moule, P. C. Oyston, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell. 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**:523–527.
19. Pupo, G. M., R. Lan, and P. R. Reeves. 2000. Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc. Natl. Acad. Sci. USA* **97**:10567–10572.
20. Reeves, P. R., L. Farnell, and R. Lan. 1994. MULTICOMP: a program for preparing sequence data for phylogenetic analysis. *Comput. Appl. Biosci.* **10**:281–284.
21. Tominaga, A., M. A.-H. Mahmoud, A. A. M. Al Mamun, and T. Mukaihara. 2001. Characterization of cryptic flagellin genes in *Shigella boydii* and *Shigella dysenteriae*. *Genes Genet. Syst.* **76**:111–120.
22. Venkatesan, M. M., M. B. Goldberg, D. J. Rose, E. J. Grotbeck, V. Burland, and F. R. Blattner. 2001. Complete DNA sequence and analysis of the large virulence plasmid of *Shigella flexneri*. *Infect. Immun.* **69**:3271–3285.
23. Wang, D., X. Ding, and P. N. Rather. 2001. Indole can act as an extracellular signal in *Escherichia coli*. *J. Bacteriol.* **183**:4210–4216.
24. Wei, J., M. B. Goldberg, V. Burland, M. M. Venkatesan, W. Deng, G. Fournier, G. F. Mayhew, G. Plunkett III, D. J. Rose, A. Darling, B. Mau, N. T. Perna, S. M. Payne, L. J. Runyen-Janecky, S. Zhou, D. C. Schwartz, and F. R. Blattner. 2003. Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect. Immun.* **71**:2775–2786.
25. Yanofsky, C., V. Horn, and P. Gollnick. 1991. Physiological studies of tryptophan transport and tryptophanase operon induction in *Escherichia coli*. *J. Bacteriol.* **173**:6009–6017.