Genotype and Phenotypes of an Intestine-Adapted Escherichia coli K-12 Mutant Selected by Animal Passage for Superior Colonization

Andrew J. Fabich
Mary P. Leatham
University of Rhode Island

See next page for additional authors

Follow this and additional works at: https://digitalcommons.uri.edu/cmb_facpubs

Terms of Use
All rights reserved under copyright.

Citation/Publisher Attribution
Available at: http://dx.doi.org/10.1128/IAI.01199-10
Genotype and Phenotypes of an Intestine-Adapted \textit{Escherichia coli} K-12 Mutant Selected by Animal Passage for Superior Colonization\textsuperscript{\dag}\textsuperscript{†}

Andrew J. Fabich,\textsuperscript{2} Mary P. Leatham,\textsuperscript{3} Joe E. Grissom,\textsuperscript{1} Graham Wiley,\textsuperscript{1} Hongshing Lai,\textsuperscript{1} Fares Najar,\textsuperscript{1} Bruce A. Roe,\textsuperscript{1} Paul S. Cohen,\textsuperscript{3} and Tyrrell Conway\textsuperscript{1,2,}\textsuperscript{*}

Advanced Center for Genome Technology, Department of Chemistry and Biochemistry, University of Oklahoma, Norman, Oklahoma 73019; Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma 73019; and Department of Cell and Molecular Biology, University of Rhode Island, Kingston, Rhode Island 02881

Received 10 November 2010/Returned for modification 11 January 2011/Accepted 6 March 2011

We previously isolated a spontaneous mutant of \textit{Escherichia coli} K-12, strain MG1655, following passage through the streptomycin-treated mouse intestine, that has colonization traits superior to the wild-type parent strain (M. P. Leatham et al., Infect. Immun. 73:8039–8049, 2005). This intestine-adapted strain (\textit{E. coli} MG1655\textsuperscript{*}) grew faster on several different carbon sources than the wild type and was nonmotile due to deletion of the \textit{flhD} gene. We now report the results of several high-throughput genomic analysis approaches to further characterize \textit{E. coli} MG1655\textsuperscript{*}. Whole-genome pyrosequencing did not reveal any changes on its genome, aside from the deletion at the \textit{flhDC} locus, that could explain the colonization advantage of \textit{E. coli} MG1655\textsuperscript{*}. Microarray analysis revealed modest yet significant induction of catabolic gene systems across the genome in both \textit{E. coli} MG1655\textsuperscript{*} and an isogenic \textit{flhD} mutant constructed in the laboratory. Catabolome analysis with Biolog GN2 microplates revealed an enhanced ability of both \textit{E. coli} MG1655\textsuperscript{*} and the isogenic \textit{flhD} mutant to oxidize a variety of carbon sources. The results show that intestine-adapted \textit{E. coli} MG1655\textsuperscript{*} is more fit than the wild type for intestinal colonization, because loss of \textit{FlhD} results in elevated expression of genes involved in carbon and energy metabolism, resulting in more efficient carbon source utilization and a higher intestinal population. Hence, mutations that enhance metabolic efficiency confer a colonization advantage.

\textit{Escherichia coli}, often one of the first bacteria to colonize infants (14, 16, 44, 45), is present in the gastrointestinal (GI) tracts of all humans (14, 53) and is found in the GI tracts of most mammals (8, 13, 35, 47). \textit{E. coli} resides and grows within the nutrient-rich mucus layer of the intestine, which provides many needed biosynthetic precursors, carbon sources, and electron acceptors (7, 15, 19, 28, 29, 39, 40). Colonization of the GI tract requires bacteria to effectively compete for these nutrients and maintain a growth rate at least equal to the 2-h turnover rate of the intestinal contents (9–11). We have shown that of the many traits that possibly make \textit{E. coli} such a remarkably successful intestinal colonizer, competition for carbon sources plays an important role (1, 7, 15, 29).

It has long been known that passage through the animal intestine can enhance virulence of certain pathogens (21, 26, 27), but the enhancement of colonization by animal passage only recently has been studied. A few years ago, we isolated a spontaneous mutant of wild-type \textit{E. coli} K-12, strain MG1655, following passage through the intestine of a streptomycin-treated mouse, which is a better colonizer than its wild-type parent strain (31). This intestine-adapted strain (\textit{E. coli} MG1655\textsuperscript{*}) grew faster on several different carbon sources than the wild type and was nonmotile due to deletion of the \textit{flhD} gene (31). FlhDC, the master regulator of flagellar synthesis, also has been implicated in control of several metabolic gene systems (34). Thus, there were two possible explanations for why \textit{E. coli} MG1655\textsuperscript{*} grows faster than the wild-type on several carbon sources. First, the \textit{flhD} mutant might save energy by not making flagella and not turning the flagellar motor. Second, loss of \textit{flhD} could cause upregulation of genes unrelated to motility, i.e., catabolic genes, that enhance colonization fitness (19). We subsequently obtained evidence that supported both of these possibilities (19). In the wild-type \textit{E. coli} MG1655 strain, an insertion element (ISI) immediately upstream of \textit{flhD} increases promoter activity, causing hypermotility and strong repression of some carbohydrate metabolism genes (2). This ISI element appears to be a hot spot for deletion of \textit{flhD} in the mouse intestine, as we isolated a number of closely related \textit{E. coli} MG1655 derivatives following animal passage, each with deletions of various lengths, beginning at the same base position immediately downstream of the ISI element (19). In addition, pathogenic \textit{Escherichia coli} O157:H\textsuperscript{−} strains with deletions knocking out FlhDC have been associated with up to 40% of patients with hemolytic uremic syndrome in Germany (41). Thus, it seemed important to further characterize the intestine-adapted \textit{E. coli} MG1655\textsuperscript{*} strain.

There are seven ISI, seven IS2, five IS3, one IS4, 11 IS5, one IS150, three IS186, four IS30, one IS600, and two IS911 elements in the \textit{E. coli} MG1655 genome (3, 51). Since IS elements are hot spots for bacterial speciation and genome evolution (32), we cannot rule out the possibility that other changes across the \textit{E. coli} MG1655 genome occurred during passage.
through the mouse GI tract. As deletion of certain E. coli regulatory genes (e.g., fnr, crp, creB) can result in unintended secondary deletions of flhD (22), it also seems possible that the inverse of this scenario could happen. For these reasons, we were interested in determining the extent of changes to the E. coli MG1655* genome and whether or not the flhD deletion alone is sufficient to explain its enhanced colonization fitness.

We have further characterized E. coli MG1655* using several high-throughput genomic analysis approaches and now show that the loss of flhD is sufficient to confer a colonization advantage in streptomycin-treated mice. Our results show that the intestine-adapted E. coli MG1655* strain is more fit than the wild type for intestinal colonization, because loss of FlhD causes elevated expression of genes involved in carbon and energy metabolism, resulting in enhanced metabolic efficiency, which is pivotal for success in host colonization.

MATERIALS AND METHODS

Bacterial strains and growth conditions. Bacterial strains were derived from the E. coli MG1655 Strr (streptomycin-resistant) strain, a K-12 strain, and are listed in Table 1. The intestine-adapted E. coli MG1655 Strr strain (MG1655*), obtained as a colony present in mouse feces at 15 days postfeeding (see the description of the mouse colonization experiment, below), was isolated and characterized as described previously (31). It has a 500-bp deletion immediately downstream of ISI in the regulatory region of flhD, extending into flhD (31). The flhD deletion strain was constructed by using the allelic replacement method of Datonko and Wanner (12) and contains a 546-bp deletion beginning immediately downstream of ISI in the regulatory region of flhD and extending into flhD. The deletion was replaced with a chloramphenicol resistance (Cat*) cassette, as described previously by Chang et al. (7). The mutation was verified by PCR analysis and by DNA sequencing of the PCR products (4). Cultures were routinely grown at 37°C on Luria-Bertani (LB) medium (33).

Mouse colonization experiment. Streptomycin-treated mice were used to study competition in the intestine between streptomycin-resistant wild-type E. coli and the isogenic flhD mutant strain. We previously described our use of this animal model in detail (30). Briefly, three male CD-1 mice (5 to 8 weeks old) were given streptomycin (5 g/liter) by drinking water containing streptomycin sulfate (5 g/ml) to distinguish the wild-type from the MG1655 Strr Nalr strain (MG1655*), extending into flhD (31). The deletion was constructed by using a blunt-end ligation of the E. coli MG1655 Strr strain (MG1655*), was used overnight on LB medium. After the mice ingested the bacterial sample was homogenized in 10 ml of 1% Bacto tryptone (Difco) and diluted in the same medium by using a DNeasy kit (Qiagen). The methods we used to generate shotgun and paired-end DNA libraries for the 454/Roche GS-FLX-Titanium sequencer are described in detail elsewhere (63). Briefly, the genomic DNA (50 μg) was sheared in a nebulizer at 30 lb/in² to generate a fragment length of 500 to 800 bp for the shotgun library. To generate the paired-end library, genomic DNA was sheared in a HydroShear (GeneMachines) apparatus to generate DNA fragments with an average length of 2.2 kb. DNA fragment length was assessed by analysis on a Caliper AMS-90 using an SE30 DNA LabChip. For generation of the shotgun library, the ends of the fragmented DNA were repaired by using DNA polymerase, the 454 adapters were ligated onto the fragment ends, and the ends were repaired a second time to obtain blunt-ended DNA, followed by quantification, dilution, and amplification via emulsion PCR, in preparation for pyrosequencing (63). The paired-end library utilized a common central linker to identify potential paired ends, which were purified and circularized, followed by nebulization, as described above, prior to emulsion PCR amplification in preparation for sequencing (63). The genome sequence was obtained from the two libraries by massively parallel pyrosequencing with a 454/Roche GS-FLX-Titanium system (36). Data were collected from four sequencing runs, which yielded 1.6 × 10⁸ reads that passed the 454/Roche filters, with an average length of 233 bases, for a total of ~3.7 × 10⁹ bases. The coverage depth, based on the size of the parent genome (3), was 75-fold.

Sequence assembly and validation. To search the E. coli MG1655* genome for mutations that may have been acquired during animal passage, the data were mapped to the published E. coli MG1655 RefSeq genome sequence (accession number NC_000913) by using GS Reference Mapper software (454/Roche), with one or more nt in the database, resulting in a total of 52 contigs ranging in length from 103 to 406,010 bases, with an average contig length of 88,609. Using this assembly strategy, repeat regions result in gaps (36), which explains why the genome was not closed completely.

Microarray experiments and statistics. The microarray protocols and data analysis used in this study were described previously (15). Briefly, total RNA was extracted from mid-logarithmic-phase cultures grown on MOPS 5-[N-(morpholino)propanesulfonic acid]-defined medium containing 0.2% (wt/vol) glucose as the sole carbon source (43) with gyratory shaking at 250 rpm. Culture samples were diluted into DNA-RNA Protect (Sierra Diagnostics), according to the Affymetrix prokaryotic labeling protocol. Biotin-labeled cDNA was fragmented and biotinylated (Enzo Kit, Roche II (Invitrogen) and random hexamers, according to the manufacturer’s specifications. The resulting cDNA was fragmented and biotinylated (Enzo Kit, Roche Diagnostics), according to the Affymetrix prokaryotic labeling protocol. Biotinylated cDNAs were hybridized to Affymetrix GeneChips containing microarrays for 16 h at 60°C. The custom GeneChips contained probe sets for all genes on the E. coli MG1655 genome. Hybridized arrays were stained using the Affymetrix protocol (ProkGE_WS2v2_450). Biological replicates of cultures were used for all microarrays. After hybridization, microarrays were scanned and .cel files were further analyzed as described previously (15). Briefly, the raw .cel files were processed in RNA (quartile normalization) (25) by using a Bioconductor R package (lumi, version 1.8.1, 3 September 2008) with an optional filter set to analyze only the subset of probe sets corresponding to MG1655 genes and intergenic regions. Microarrays of E. coli MG1655* and the E. coli MG1655 ΔflhDC:cat strain grown on glucose were used as the experimental condition, and the MG1655 Strr Nalr strain grown on glucose was used as the control.
condition. Fisher’s exact test was used to determine whether the modest gene expression changes examined were statistically relevant.

**BiOLOG GN2 MicroPlate assays.** BiOLOG assays involved an experiment design developed by Ihssen and Egli (24) and used with some modifications (61). Briefly, overnight cultures grown in MOPS minimal glucose medium were used to inoculate fresh medium (1% inoculum). When the cultures reached an optical density at 600 nm of 0.3, chloramphenicol (30 µg/ml) was added and the cells were washed three times in MOPS minimal medium lacking a carbon source and containing chloramphenicol. The washed cells were resuspended to an optical density at 600 nm of 0.3 and inoculated into the microplate. The BiOLOG GN2 microplates were placed in an OmniLog plate reader at 37°C and read every 15 min for 24 h. Each experiment was duplicated on separate days, and the data were analyzed.

**Accession numbers.** The sequence data are available at GenBank, accession number CM000960 AEFE01000000. The microarray data are available at the GEO repository, accession number GSE25106.

**RESULTS**

**Colonization fitness advantage of the E. coli MG1655 flhD mutant.** We previously showed that E. coli MG1655* grows 6 to 23% faster on several different sugars and about 15% faster on cecal mucus, as sole carbon and energy sources, than its wild-type parent. Moreover, E. coli MG1655* has a colonization advantage over the wild type (31). Subsequently, E. coli MG1655* was found to have an flhD deletion. When tested for its growth rate on mannose and ribose, we found that E. coli MG1655 flhD grows as rapidly as E. coli MG1655* under these conditions (31). We also found that the E. coli MG1655 flhD mutant grows faster than its wild-type parent on a mixture of 10 different sugars known to be present in mouse cecal mucus (19). Therefore, it was of interest to determine whether the E. coli MG1655 flhD mutant is a better colonizer than the wild type. To test this possibility, groups of 3 CD-1 mice were treated for 24 h with streptomycin (5 mg/ml) in their drinking water to open a niche for experimental colonization, and then the mice were fed 10⁵ CFU each of the E. coli MG1655 flhD::cat strain and the E. coli MG1655 Str⁰ Nal⁰ wild type; the Cm⁰ cassette was used to enumerate the E. coli MG1655 flhD::cat population, and the NaI cassette was used to count the wild-type population. As shown in Fig. 1, the E. coli MG1655 Str⁰ flhD mutant had about a 2-log (100-fold)-higher population than the E. coli MG1655 Str⁰ Nal⁰ wild type. From this result, we concluded that mutation of flhD is sufficient to confer a colonization advantage that is analogous to that of the intestine-adapted strain (31).

**Pyrosequencing of the intestine-adapted E. coli MG1655* genome.** Wild-type E. coli MG1655 has an IS1 element within its flhD regulatory region that increases transcription of the flhDC operon, resulting in increased production of the flagellar master regulator, which causes wild-type E. coli MG1655 to be hypermotile (2). Following passage through the streptomycin-treated mouse intestine, the majority of intestine-adapted derivative strains are nonmotile (19). Previously, we used PCR mapping to identify on the E. coli MG1655* genome an ~500-bp deletion that begins immediately downstream of the insB gene of the IS1 element and extends into the flhD gene (31). PCR mapping of 10 additional intestine-adapted derivatives of E. coli MG1655 found that each contains a deletion that begins at the same locus proximal to the IS1 element and extends at various lengths (~100 to 6,100 bp) into and beyond the flhDC operon (19). Since IS elements can promote genome rearrangements, they are hot spots for mutation (32, 42, 62), and there are approximately 50 transposon-related loci on the E. coli MG1655 strain (51), which represents significant potential for genetic change. Therefore, we sequenced the genome of intestine-adapted E. coli MG1655* to determine whether the flhD mutation is solely responsible for its colonization fitness advantage or if additional mutations occurred during animal passage.

To sequence the E. coli MG1655* genome, we used a massively parallel pyrosequencing strategy that is now routine in genome centers (63). The pyrosequencing summary statistics are shown in Table 2. We obtained 75-fold coverage of the genome and used GS Reference Mapper to align the reads against the E. coli MG1655 reference genome (RefSeq accession number NC_000913). The resulting assembly consists of

---

**TABLE 2. Summary statistics for pyrosequencing of E. coli MG1655***

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of instrument runs</td>
<td>4</td>
</tr>
<tr>
<td>No. of high-quality reads</td>
<td>1,586,667</td>
</tr>
<tr>
<td>Average read length (bases)</td>
<td>233</td>
</tr>
<tr>
<td>Total no. of contigs</td>
<td>52</td>
</tr>
<tr>
<td>Avg (range) contig size (bp)</td>
<td>88,609 (103 to 406,010)</td>
</tr>
<tr>
<td>Total no. of assembled reads in contigs</td>
<td>1,491,412</td>
</tr>
<tr>
<td>Total no. of sequenced bases in contigs</td>
<td>569,576,896</td>
</tr>
</tbody>
</table>
52 contigs averaging 88,609 bp and ranging between 103 and 406,010 bp long.

The \textit{flhD} gene was the only single-copy gene containing a deletion in the \textit{E. coli} MG1655* genomic sequence (Fig. 2), and it maps to a 500-bp gap between contigs 19 and 20 that begins 15 bp downstream of the \textit{insB} gene of the IS\textsubscript{1} element, extends across the \textit{flhD} regulatory region, and removes the first 195 bp of the \textit{flhD} coding sequence. Given that PCR mapping places the common endpoint of other \textit{flhD} deletions obtained by animal passage in the same region adjacent to the \textit{insB} locus (19), we propose that the corresponding gap in the \textit{E. coli} MG1655* genome sequence is due to a deletion promoted by the IS\textsubscript{1} element and selected because it confers a colonization fitness advantage. The remaining 51 gaps in the \textit{E. coli} MG1655* genome sequence cover regions containing genes that are highly repeated and include 26 insertion sequence genes (\textit{ins}), 18 rRNA genes (\textit{rrs} or \textit{rrl}), 5 transposon-related function genes (\textit{rhs}), and 2 prophage genes (\textit{ydfJ}); each gap represents sequences that are identical to between 2 and 11 other loci, making them impossible to close by pyrosequencing. The Newbler assembler software typically does not include high-copy-number repeat sequences and thus results in gaps.
that represent repeats, despite the inclusion of paired-end reads (36, 63). Nevertheless, the \textit{E. coli} MG1655* genome assembly is of very high quality, as roughly one-half of the regions containing IS elements were “closed.” Thus, while we are unable to rule out the possibility that additional deletions promoted by an insertion sequence did not occur, we can state with absolute certainty that \textit{flhD} is the only nonrepetitive gene that was deleted on the \textit{E. coli} MG1655* genome.

Our search for minor genotypic changes in the \textit{E. coli} MG1655* genome, by comparison to the \textit{E. coli} MG1655 reference genome, was facilitated by having high-quality reads and 75-fold coverage. A total of 8 possible mutations were identified: four were single-base substitutions, three were insertions of 1 or 2 bases, and one was a single-base deletion. These data are shown in Table 3. We closely examined the eight putative mutations and considered their potential involvement in colonization. Two changes were found in the \textit{ybE} coding sequence, and another was found in the intergenic region between \textit{ppiC} and \textit{yifN}. The former is a pseudogene, and the latter is unlikely to affect gene function; therefore, these three putative mutations are unlikely to impact colonization fitness. We found a single-base substitution in \textit{gvrA} that converts the aspartate at codon 87 to tyrosine. This mutation, which is known to confer fluoraurolidine resistance on a clinical isolate of \textit{E. coli} (10), is almost certainly the genotype responsible for the Nal r phenotype of the \textit{MG1655*} compared to in the wild-type strain. We also found a substitu-
tion in \textit{rpsL} gene that renders the \textit{E. coli} MG1655 parent strain resistant to streptomycin, as we previously reported (30). We then focused attention on the three genotypic changes that were identified within coding sequences of genes involved in metabolism. A 2-base deletion was identified in the \textit{gatC} gene, which encodes a component of the phosphoenolpyruvate (PEP)-dependent phospho-transferase system for galactitol uptake. We previously established that galactitol is not used by \textit{E. coli} MG1655 to colonize streptomycin-treated mice (7). Lastly, we identified a 2-base insertion in the \textit{gltP} gene that would inactivate the proton symporter for glutamate and aspartate (60). However, it is unlikely that transport function is lost, as \textit{E. coli} also has the \textit{gltIKL}-encoded ABC uptake system for amino acids in addition to the \textit{gltS}-encoded glutamate transporter.

To test directly the hypothesis that the genotypic changes to the \textit{E. coli} MG1655* genome, other than the \textit{flhD} deletion, did not contribute to its colonization advantage, we fed \textit{E. coli} MG1655* and the \textit{E. coli} MG1655 \textit{flhD} mutant to mice. If the two strains were found to cocolonize, this would exclude the possibility that mutations on the \textit{E. coli} MG1655* genome other than an \textit{flhD} mutation were responsible for its colonization advantage. As expected, we show in Fig. 1 that the two strains cocolonized, which indicates that \textit{E. coli} MG1655* and the \textit{E. coli} MG1655 \textit{flhD} mutant are identical in their ability to compete in the mouse intestine.

In summary, sequence analysis of the \textit{E. coli} MG1655* genome failed to reveal any mutations that could possibly cause the observed colonization advantage of this intestine-adapted strain, aside from the previously identified \textit{flhD} deletion. Since the data shown in Fig. 1 are consistent with the hypothesis that an isogenic \textit{flhD} mutant has a colonization advantage over \textit{E. coli} MG1655 analogous to that of the intestine-adapted strain (31), we conclude that the \textit{flhD} deletion is solely responsible for the colonization fitness advantage of \textit{E. coli} MG1655*.

### Table 3. Summary of minor genotypic changes on \textit{E. coli} MG1655* genome

<table>
<thead>
<tr>
<th>Gene affected</th>
<th>Mutation</th>
<th>Base position</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{ybE}</td>
<td>A to G substitution</td>
<td>547,964</td>
</tr>
<tr>
<td>\textit{ybE}</td>
<td>G insertion</td>
<td>547,835</td>
</tr>
<tr>
<td>\textit{gatC}</td>
<td>CC insertion</td>
<td>2,171,386</td>
</tr>
<tr>
<td>\textit{gvrA}</td>
<td>C to A substitution</td>
<td>2,357,184</td>
</tr>
<tr>
<td>\textit{rpsL}</td>
<td>T to G substitution</td>
<td>3,472,447</td>
</tr>
<tr>
<td>\textit{gltP}</td>
<td>G deletion</td>
<td>3,558,478</td>
</tr>
<tr>
<td>\textit{ppiC-yifN} intergenic</td>
<td>C to T substitution</td>
<td>3,957,957</td>
</tr>
<tr>
<td>\textit{gltP}</td>
<td>CG insertion</td>
<td>4,294,404</td>
</tr>
</tbody>
</table>

* Insertions or deletions of one or two bases and single-base substitutions.

\textit{b} Relative to \textit{E. coli} MG1655 reference sequence (accession no. NC_000913).
and on cecal mucus? Second, is the consistent with the faster growth of the strain on several sugars sugar catabolism gene systems in the intestine-adapted strain used to address two basic questions. First, is expression of the bioinformatics platform, as described previously (7, 15), and and hybridized to a custom Affymetrix DNA microarray. The state growth conditions. Total RNA was extracted, labeled, and hybridized to a custom Affymetrix DNA microarray. The resulting transcriptome data were analyzed on an in-house bioinformatics platform, as described previously (7, 15), and used to address two basic questions. First, is expression of the sugar catabolism gene systems in the intestine-adapted strain consistent with the faster growth of the strain on several sugars and on cecal mucus? Second, is the flhD deletion alone sufficient to explain the enhanced catabolic capacity of E. coli

**Induced catabolic regulon expression in the intestine-adapted strain.** To address the possibility that differences in its transcriptome cause E. coli MG1655* to grow faster on several sugars, we evaluated the expression levels of carbon catabolism gene systems by comparison to the wild-type parent strain, when grown on glucose minimal medium. We focused on 23 carbohydrates that were previously tested for their ability to support intestinal colonization (1, 7, 15, 29) or showed more rapid growth of E. coli MG1655* when growing on glucose, since the average log₂ expression ratios of the 23 regulons of cells grown on mannose, which is far less catabolite repressing (37), was –0.10 (data not shown). The modest yet consistently higher expression of catabolic regulons appeared to be significant, as the average log₂ expression ratio for all known regulons (RegulonDB) was 0.06, with 100 regulons expressed higher in E. coli MG1655* and 72 regulons expressed higher in the wild type (data not shown). The high incidence of elevated expression of catabolism genes extended beyond the 23 sugar regulons to include all 65 regulons that are involved in nutrient degradation (average log₂ expression ratio = 0.36), as opposed to the remaining 107 regulons that are not involved in degradation (average log₂ expression ratio = –0.11) (see Table S1 in the supplemental material). Thus, a simple transcriptome comparison of E. coli MG1655* and its wild-type parent revealed a general, modest increase in the transcript levels of genes involved in nutrient uptake and catabolism, including the 23 regulons of interest.

In situations where a transcriptome experiment is designed to identify relatively small gene expression differences in a data set where very few genes exceed statistically significant thresholds, generally accepted statistical metrics are ineffectual, including the 2-fold difference rule of thumb (64) or differences greater than 2 standard deviations from the mean of the expression ratios (11, 59). In the present case, for the entire data set the standard deviation from the mean of the log₂ expression ratios is 0.64, and only 22 genes exceeded this threshold by 2-fold (1.28), none of them from the 23 regulon set of interest (see Fig. S1 in the supplemental material). When the genes are considered individually, there are some regulons in which all members have a positive expression ratio and others in which

**FIG. 3.** Average regulon gene expression for carbon catabolism genes in E. coli MG1655* by comparison to the wild type. An online tool called E-ring (http://genexpdb.ou.edu/ering/) was used to average gene expression for genes involved in carbon metabolism when grown on glucose. E. coli regulons are represented on the x axis by the regulator for each carbohydrate that E. coli can metabolize. The y axis represents the average log₂ expression ratio obtained by averaging the expression ratios for each gene within the given regulon. *, P values of <0.005 in Fisher’s exact test.

---

**Gene Expression Ratios (log2)**

- E. coli
- MG1655*
- araC (arabinose)
- cdaR (glucarate, galactarate)
- exuR (hexuronates)
- galR (galactose)
- gntR (gluconate)
- lacI lactose)
- lldR (L-lactate)
- malT (maltose)
- mtlR (mannitol)
- nagC (N-acetylglucosamine)
- nanR (sialic acid)
- rbsR (ribose)
- rhaS (rhamnose)
- tdC (N-acetylgalactosamine)
- treR (trehalose)
- uhpA (glucos e-6-phosphate)

**Average regulon gene expression for carbon catabolism genes in E. coli MG1655* and the wild type, the expression**
some members have positive and others negative expression values (see Fig. S1). For all of these reasons, a traditional statistical approach cannot be used to establish significance in the data set. Therefore, we turned to an overrepresentation analysis, where lists of genes, predefined by RegulonDB, are analyzed to see which regulons are represented more than would be expected by chance (5). A similar approach based on overrepresentation of gene ontology annotations in microarray data has been successful (20). To evaluate the statistical probability that specific transcription factors are responsible for the observed transcriptional differences between the E. coli MG1655* and wild-type strains, we applied Fisher’s exact test, employing a two-tailed hypergeometric distribution to identify enriched transcription factors in RegulonDB that are overrepresented in the transcription data. The most overrepresented transcription factor of genes that exceeded the positive threshold was Crp (catabolite repression control). This was not surprising, since all but 13 genes in 2 of the 23 regulons under consideration also are under Crp control. Among the remaining 147 genes listed in RegulonDB as being Crp activated, 14 out of 23 of the corresponding transcription factors have P values of <0.005, as determined by Fisher’s exact test (Fig. 3). Thus, we conclude that the probability of involvement of most of the carbon catabolism transcription factors exceeds the 5% confidence level for control of the modestly elevated transcript levels, under catabolite-repressing growth conditions, for the corresponding catabolic gene systems.

The transcriptomes of E. coli MG1655* and the ΔflhDC strain are essentially identical. The most prominent feature of the E. coli MG1655* transcriptome experiments is substantially decreased expression of flagellar genes (see Fig. S2 in the supplemental material), particularly genes in the fli and flg operons. This result is consistent with the results of two related transcriptome analyses, published by others (50, 65), of flhD strains constructed for the purpose of characterizing the FlhDC and FliA regulons. To test the hypothesis that the flhD deletion in E. coli MG1655* is solely responsible for the transcription patterns detailed above, we constructed an isogenic strain in the same wild-type E. coli MG1655 genetic background by replacing the flhD allele with a Cmr cassette. Based on the above results, we would expect that the transcriptomes of the two strains are essentially identical. We tested this hypothesis by using Fisher’s exact test to analyze the regulatory networks that contributed the most significant differences between the E. coli MG1655* and E. coli MG1655 flhD mutant strains, in comparison to the wild-type control arrays. The P values of eight representative regulons, four above and four below the threshold value used in the analysis (1.7 linear fold change), are shown in Table 4. In each case, the P value is below the cutoff that indicates a 5% confidence level (P < 0.005). As shown in Fig. S3 in the supplemental material, the average gene expression ratios for these regulons are essentially identical. The conclusion that the transcriptomes of E. coli MG1655* and the E. coli MG1655 flhD mutant are essentially the same is further supported by the correlation values between the arrays. The correlation between the E. coli MG1655* and E. coli MG1655 flhD arrays was 0.962, which compares favorably to the correlation between replicate control arrays (wild type) of 0.964 (data not shown). The correlations between the control and E. coli MG1655* and between the control and the E. coli MG1655 flhD mutant were both 0.929 (data not shown). Thus, microarray comparison of E. coli MG1655* and the E. coli MG1655 flhD mutant grown under the same conditions indicates that their transcriptomes essentially are identical, with no features of the E. coli MG1655* transcriptome that would indicate obvious regulatory differences that could have been caused by unidentified mutations. We conclude that the flhD deletion acquired by E. coli MG1655* during animal passage is responsible solely for its transcriptional phenotype: expression of catabolic gene systems is modestly, yet broadly and consistently, elevated in the intestine-adapted strain.

Catabolome of an intestine-adapted strain compared to that of the wild type grown in glucose. To determine whether the changes in the transcriptome of E. coli MG1655* were reflected by its capacity to metabolize the indicated substrates, we used Biolog technology to measure the oxidation rates of 95 different carbon sources. To ensure that the results of the carbon source oxidation experiment were comparable to the microarray results, cultures were grown on minimal glucose medium as described above for the microarray analysis. The E. coli MG1655* and wild-type strains were essentially equal in the number of carbon sources utilized. However, we found that E. coli MG1655* oxidized 24 carbon sources to a greater extent than the wild-type parent strain (see Fig. S4 in the supplemental material). The data also show that the overall extent of carbon source oxidation was 21% higher for E. coli MG1655* (Fig. 4). Furthermore, the overall extent of carbon source oxidation was the same for E. coli MG1655* and the flhD deletion strain. Lastly, we examined an MG1655 strain in which the IS1 element was removed from the regulatory region of the flhDC promoter that is normally motile (19) and that has normal levels of FlhDC. We found the IS1 deletion strain had an intermediate level of carbon source oxidation compared to that of E. coli MG1655* and the wild type. Therefore, the overall extent of carbon source oxidation was highest in strains lacking FlhDC (the E. coli MG1655 flhDC mutant strain and E. coli MG1655*) and lowest in the wild-type strain (E. coli MG1655), which is hypermotile and overexpresses FlhDC. Thus, we conclude that loss of FlhDC in intestine-adapted E. coli MG1655* results in release of several carbon catabolism operons from FlhDC repression and thereby an enhanced carbon source oxidation phenotype.

### Table 4: Fisher’s exact test showing regulatory networks affected in E. coli MG1655* and the E. coli MG1655 flhD mutant compared to wild-type E. coli MG1655

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Regulon</th>
<th>Result for:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above</td>
<td>Crp</td>
<td>1.89E−03*</td>
</tr>
<tr>
<td></td>
<td>Fnr</td>
<td>4.01E−06</td>
</tr>
<tr>
<td></td>
<td>GadX</td>
<td>1.37E−03</td>
</tr>
<tr>
<td></td>
<td>IHF</td>
<td>4.67E−07</td>
</tr>
<tr>
<td>Below</td>
<td>FlhDC</td>
<td>1.28E−04</td>
</tr>
<tr>
<td></td>
<td>Fur</td>
<td>3.34E−06</td>
</tr>
<tr>
<td></td>
<td>PurR</td>
<td>4.81E−07</td>
</tr>
<tr>
<td></td>
<td>TyrR</td>
<td>1.66E−03</td>
</tr>
</tbody>
</table>

*P value cutoff of 0.005 corresponds to 5% confidence level.
FIG. 4. Catabolome analysis of E. coli MG1655* during growth on minimal glucose medium. Cells were grown to mid-log phase, harvested in chloramphenicol, loaded into BIOLOG GN2 microplates, and measured every 15 min for 24 h. The plot shows the overall extent of carbon source oxidation for wild-type E. coli MG1655 ("0.5 WT") and MG1655* ("1.0 WT") strains deleted for IS1 (IS1−) and flhD (ΔflhDC mutant), and the intestine-adapted strain (MG1655*). Error bars show the standard deviations.

DISCUSSION

Passage of E. coli MG1655 through the GI tract of the streptomycin-treated mouse selected for a strain that differed from the wild type in being more fit for colonization (31). We subsequently found that the intestine-adapted strain, E. coli MG1655*, acquired a deletion of the flhD gene that rendered it nonmotile and able to grow 6 to 25% faster than its wild-type parent in laboratory cultures on several sugars and on mouse cecal mucus (31). In a follow-up study, we found two reasons that E. coli MG1655* was a more successful colonizer (19).

First, mutational analysis showed that the flhD deletion apparently altered expression of genes not involved in motility that conferred enhanced colonization fitness. Second, the energy normally used by the wild type to synthesize flagella and power the flagellar motor was redirected to E. coli MG1655* growth (19). To confirm that the observed colonization fitness advantage resulted from deletion of flhD, we sequenced the E. coli MG1655* genome and used high-throughput approaches to characterize phenotypes relevant to its enhanced carbon catabolism. We found that the only E. coli MG1655* genomic deletion partially removed the flhD gene immediately downstream of an IS1 element that apparently promoted the deletion. We found two additional point mutations, one in a glycero catabolism regulatory gene and the other in the gatC gene, which could inhibit galactitol uptake. However, after careful consideration of the available evidence, we concluded based on evidence discussed in Results above that neither of them would positively impact colonization and hence would not provide a selective advantage. No doubt, mutations are likely to occur spontaneously during the 180 generations spent in the intestine (2 h generation time for 15 days), but in the highly selective environment of the intestine, beneficial changes in genotype would be selected, and detrimental changes would be eliminated from the population, leaving only neutral mutations and those that are advantageous, such as flhD.

Based on the above results, we then investigated the nature of the phenotypic changes promoted by deletion of flhD in an attempt to determine how they could provide a colonization advantage. Here, we observed that E. coli MG1655* had a generally and modestly elevated (1.3-fold) expression of carbon and energy metabolism genes, together with a corresponding 21% overall increase in oxidation of 24 carbon and energy sources. These genome-wide increases are of the same order of magnitude as the 6 to 23% higher growth rates on several carbon sources previously observed in laboratory cultures (19), which we propose to result in the 100-fold-higher intestinal population during competition against the wild type (Fig. 1).

Furthermore, we found no differences in the transcriptomes and catabolomes of E. coli MG1655* and an flhD mutant constructed in E. coli MG1655. To search for clues as to how deletion of flhD could result in the observed differences, we compared our transcriptome data to previously reported studies (49) in which deletion of flhD caused downregulation of genes involved in sugar acid metabolism and decreased growth on gluconate, thought to be mediated by the redox sensor-encoding aer gene but by an unknown genetic mechanism (23).

In distinct contrast, in our study we found that expression of sugar acid catabolism regulons was increased 1.3-fold, together with generally increased expression of a wide variety of sugar catabolism genes. The difference in growth conditions used in the two studies was revealing: the former results were obtained from cells grown on amino acids in Luria broth (49), and ours were from cells grown on glucose as the sole carbon source, as described above. As we noted, the vast majority of genes that showed altered expression in E. coli MG1655* compared to that of the wild type are under Crp control. Moreover, when grown on glucose, the average log2 expression ratio of the 23 regulons was 0.41, whereas for cells grown without significant catabolite repression on mannose, the average log2 expression ratio was −0.10. These data suggest that catabolite repression control might be involved in the faster growth of the intestine-adapted strain in the intestine, but for reasons that are not clear. Thus, this transcriptome analysis demonstrates that deletion of flhD results in modestly higher expression of carbon and energy metabolism gene systems, which also was reflected physiologically in the carbon source oxidation patterns of cells grown under catabolite-repressing conditions. We therefore conclude that this increase in gene expression is directly responsible for the colonization advantage of E. coli MG1655*.

In seeking to explain why E. coli MG1655* is a better colonizer, we note that deletion of flhD does not expand its catabolic capacity to include additional carbon sources not used by the wild type. Therefore, it is more likely that the enhanced expression and activity of catabolic pathways is responsible for the colonization fitness advantage of the intestine-adapted strain. Conversely, in many cases, we found that genetic lesions in metabolic pathways resulted in decreased colonization fitness (7, 15). Not all of the mutations we previously tested completely eliminated growth on a particular carbon source. Some mutations removed one redundant pathway and hence slowed the growth rate on a particular carbon source, while other mutations caused sugar phosphates to accumulate, which led to growth inhibition. Since decreased growth rate of a mutant can cause a colonization defect, it stands to reason that increased growth rate could be selected by virtue of being advantageous. These concepts are central to a corollary of
Freter’s nutrient-niche hypothesis (18), that the population of a species is proportional to the amount of its preferred carbon source. The results presented here indicate that more efficient carbon source utilization also results in a higher population size in the intestine. Mutations that enhance metabolic efficiency therefore confer a colonization advantage.

ACKNOWLEDGMENTS

This research was supported by Public Health Service grants AI48945 and GM095370 to T.C. and P.S.C. We thank Aaron Conrado for assistance with the animal experiments. Special thanks to personnel of the University of Oklahoma Advanced Center for Genome Technology for their hard work.

REFERENCES


