Enterotoxigenic Escherichia coli (ETEC) is a significant cause of morbidity and mortality in the developing world. ETEC-mediated diarrhea is orchestrated by heat-labile toxin (LT) and heat-stable toxins (STp and STh), acting in concert with a repertoire of more than 25 colonization factors (CFs). LT, the major virulence factor, induces fluid secretion after delivery of a monomeric ADP-ribosylase (LTA) and its pentameric carrier B subunit (LTB). A study of ETEC isolates from humans in Brazil reported the existence of natural LT variants. In the present study, analysis of predicted amino acid sequences showed that the LT amino acid polymorphisms are associated with a geographically and temporally diverse set of 192 clinical ETEC strains and identified 12 novel LT variants. Twenty distinct LT amino acid variants were observed in the globally distributed strains, and phylogenetic analysis showed these to be associated with different CF profiles. Notably, the most prevalent LT1 allele variants were correlated with major ETEC lineages expressing CS1 or CS2 or CS3, and the most prevalent LT2 allele variants were correlated with major ETEC lineages expressing CS5 or CS6 or CFA/I. LTB allele variants generally exhibited more-stringent amino acid sequence conservation (2 substitutions identified) than LTA allele variants (22 substitutions identified). The functional impact of LT1 and LT2 polymorphisms on virulence was investigated by measuring total-toxin production, secretion, and stability using in silico protein modeling. Our data show that LT2 strains produce 5-fold more toxin than LT1 strains (P < 0.001), which may suggest greater virulence potential for this genetic variant. Our data suggest that functionally distinct LT-CF variants with increased fitness have persisted during the evolution of ETEC and have spread globally.
riboylation of the stimulatory guanine-nucleotide-binding G protein α subunit (Gso), which leads to increased production of cAMP and deregulation of the cystic fibrosis transmembrane receptor (CFTR) ion channel, resulting in hypersecretion of electrolytes and water into the intestinal lumen, i.e., diarrhea (8).

Several studies of LT-producing ETEC strains—based on genetic, biochemical, and immunological characterization—have shown that LT is a heterogeneous family (6, 8, 15). Two families have been described: LT-I (including the human ETEC reference strain H10407) and the novel family LT-II. The LT-I expressed by ETEC strains isolated from human samples is highly similar to cholera toxin in terms of amino acid sequence, showing 80% sequence homology (6). LT-II (LT-Ila, LT-Iib, and LT-Iic) purified from buffalo stool samples is antigenically distinct from LT-I or cholera toxin (16). Subsequent sequencing analysis has validated such differences, showing high amino acid sequence divergence mainly in the LT-II mature B subunit, which shares only 15 to 16% identity with LT-I and cholera toxin (17). A previous study analyzed the DNA sequences of ETEC LT-I strains isolated from humans in Brazil; 16 LT-I types were identified and were termed LT1 to LT16 (15). These data revealed high levels of polymorphism, mainly in eltA. Since Lasaro et al. analyzed mainly Brazilian strains (15), we were interested in understanding the worldwide distribution of polymorphisms present in the eltAB operon among a geographically and temporally diverse set of clinical ETEC isolates, some of which belong to globally distributed persistent lineages (18). We analyzed the LT-I operons of 192 human ETEC strains isolated from several continents, including Asia, Africa, and Latin America, over 3 decades, both strains belonging to stable lineages and individual isolates with different colonization factor and toxin profiles, in order to evaluate the natural diversity of LT.

**MATERIALS AND METHODS**

**Bacterial strains.** A representative collection of 362 ETEC strains from the University of Gothenburg strain collection (comprising more than 3,500 ETEC strains) were subjected to whole-genome sequencing at the Wellcome Trust Sanger Institute (18); of these, 186 strains were positive for LT and were included in this study. The LT-ETEC strains were collected between 1980 and 2011 from 21 different countries. Strains were isolated from a diverse demographic, including patients younger than the age of 5 years, adults, and travelers and soldiers with acute diarrheal disease; some strains (n = 7) were also isolated from asymptomatic individuals. Six additional LT-expressing strains isolated in cases of diarrhea in Bolivia from 2002 to 2011 were also included in this study. All strains were from anonymous patients and were isolated from stool with informed consent. Permission to use the ETEC strain collection was granted by the Regional Ethical Board of Gothenburg, Sweden (Ethics Committee reference no. 088-10). Strains were characterized as ETEC by the expression of LT and/or ST as determined by GM1– enzyme-linked immunosorbent assay (GM1-ELISA) and inhibition ELISA, respectively, as well as by multiplex PCR. A dot blot assay was used for characterization of CFA/I, CS1 to CS8, CS12, CS14, CS17, CS19, and CS21 (19). BLASTn analysis was used to confirm the presence of CF operons and toxin genes within the genome of each ETEC isolate.

**Genomic sequencing and extraction of the eltAB gene.** ETEC strains were grown on horse blood agar plates overnight at 37°C. DNA was isolated from each strain in accordance with the instructions in the Wizard Genomic DNA kit (Promega). The genomic library preparation and DNA sequencing have been described by von Mentzer et al. (18), and genomic extraction of the eltAB gene was performed by nBLAST in this study. GenBank accession number S60731 was used for the eltAB genomic extraction.

**LT variant identification and phylogenetic analysis.** Multisequence alignment of 192 amino acid sequences translated from eltAB was performed using ClustalW. A concatenated sequence was constructed for phylogenetic analysis by subtracting the sequences corresponding to the signal peptides of the LTA and LTB subunits. The MEGA program (version 5.2) was used to extract the variables in the translated amino acid sequence of each strain. Sequences were compared to LT variants reported in previous studies: LT1 (15), LT2 (20), and LT3 to LT16 (GenBank accession numbers EU113242 [LT3], EU113243 [LT4], EU113244 [LT5], EU113245 [LT6], EU113246 [LT7], EU113247 [LT8], EU113248 [LT9], EU113249 [LT10], EU113250 [LT11], EU113251 [LT12], EU113252 [LT13], EU113253 [LT14], EU113254 [LT15], and EU113255 [LT16]) (15). Phylogenetic trees were generated in MEGA (version 5.2) using the neighbor-joining algorithm.

**GM1-ELISA.** A single-read GM1-ELISA for phenotypic demonstration and quantification of LT produced by a subset of 155 ETEC strains included in the study was adapted from the work of Svennerholm and Wiklund (21) with the following modifications. Briefly, 1 ml of culture was collected from a 5-h culture of an ETEC strain in Luria broth; cells were sonicated in phosphate-buffered saline (PBS) and were then transferred to GM1-coated ELISA plates in triplicate without dilutions. The plates were developed by following the protocol described in reference 21. All samples were tested in triplicate, and mean optical density (OD) values for the detection of LT in each strain were used to extrapolate the amount of LT produced. A subset of 31 strains (16 strains expressing LT1 and 15 strains expressing LT2) was selected for quantitative ELISA, which was performed using a 3-fold dilution series and recombinant cholera toxin subunit B (rCTB) as a reference as described previously (21). In addition, a test to determine the amounts of free B subunits relative to the amounts of the AB holotoxin in the culture was set up. Bacterial extracts from sonicated pellets and supernatant fractions were transferred to two GM1-coated microtiter plates. We used an in-house monoclonal antibody (mAb) for the LTB and CTB subunit (LT39:1:1) in the first plate (19) and another in-house monoclonal antibody specific for the LTA subunit (LT17A) in the second plate. Antibody LT39:1:1 detected both the free B subunit and fully assembled LT holotoxin, while antibody LT17A detected only assembled holotoxin. Quantitative ELISA using 3-fold dilutions was performed as described above. Assays were performed in duplicate, and the ratio of ABholotoxin to the total amount of Bholo was determined by measurement of the amount of product obtained in the ELISA targeting anti-CTA and anti-CTB. We assumed that anti-CTA detects the Bholo holotoxin bound through the Bsubunit-mediated binding to the GM1-coated plates, while anti-CTB detects both holotoxin and dissociated B subunits, i.e., the total amount of Bholo.

**LT1 and LT2 (AB5) modeling.** The predicted stability of the LT1 holotoxin was compared to that of the LT2 holotoxin using *in silico* modeling. The region of S190 on LTA was flexible; therefore, the coordinates for the LT2A part of this conformation was used to perform a potential protein–protein interface prediction with InterProSurf (24). The last conformation of the S190 on LTA was used for visualization. The LT2A part of this conformation was used to perform a 2-ns molecular dynamics (MD) simulations performed with the AMBER software package (23). The last conformation of the 2-ns MD simulation trajectories was used for further analysis and visualization. The LT2A part of this conformation was used to perform a potential protein–protein interface prediction with InterProSurf (24). The interactions of position 75 on LTB with the other region were missing in all available crystal structures. High-quality homology models of the LT1A-LT1B and LT2A-LT2B pentamers were generated using the MPACK software package (22). The underlying template for the models was a refined (1.95-Å) *E. coli* crystal structure of LTp (ILTS); the LT2 holotoxin was modeled onto the template and was then compared with LT1. The quality of the models was very high, since the sequence identity with the template was 97%. The LT1 and LT2 (AB5) models were individually soaked in a TIP3 water box. After initial energy minimizations, 2-ns molecular dynamics (MD) simulations were performed with the AMBER software package (23). The last conformation of the 2-ns MD simulation trajectories was used for further analysis and visualization. The LT2A part of this conformation was used to perform a potential protein–protein interface prediction with InterProSurf (24). The interactions of position 75 on LTB in both the LT1 and LT2 models were analyzed separately.

**Statistical analysis.** The different LT alleles determined by single-read ELISA were analyzed using one-way analysis of variance (ANOVA). Non-paired Mann-Whitney U tests were applied to the rest of the data. All tests were performed using GraphPad Prism, version 6.00 for Windows.
TABLE 1 Virulence gene profiles of ETEC strains included in this study

<table>
<thead>
<tr>
<th>Toxin and CF profile</th>
<th>No. (%) of ETEC strains (n = 192)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT CFA/I + CS21</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>LT/Sth CFA/I + CS21</td>
<td>5 (2.6)</td>
</tr>
<tr>
<td>LT CS1</td>
<td>3 (1.6)</td>
</tr>
<tr>
<td>LT/Sth CS1 + CS3</td>
<td>3 (1.6)</td>
</tr>
<tr>
<td>LT/Sth CS1 + CS3 + CS21</td>
<td>17 (8.9)</td>
</tr>
<tr>
<td>LT CS12</td>
<td>2 (1)</td>
</tr>
<tr>
<td>LT/Stp CS12</td>
<td>4 (2.1)</td>
</tr>
<tr>
<td>LT CS14</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>LT/ST CS14</td>
<td>2 (1)</td>
</tr>
<tr>
<td>LT CS17</td>
<td>7 (3.6)</td>
</tr>
<tr>
<td>LT/St CS17</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>LT/St CS19</td>
<td>11 (5.7)</td>
</tr>
<tr>
<td>LT/Sth CS2 + CS21</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>LT CS2 + CS3</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>LT/Sth CS2 + CS3</td>
<td>3 (1.6)</td>
</tr>
<tr>
<td>LT/Sth CS2 + CS3 + CS21</td>
<td>7 (3.6)</td>
</tr>
<tr>
<td>LT/ST CS21</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>LT/ST CS23</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>LT CS3 + CS21</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>LT/Sth CS3 + CS21</td>
<td>2 (1)</td>
</tr>
<tr>
<td>LT/Stp CS4 + CS6</td>
<td>2 (1)</td>
</tr>
<tr>
<td>LT/Sth CS5 + CS6</td>
<td>17 (8.9)</td>
</tr>
<tr>
<td>LT CS6</td>
<td>11 (5.7)</td>
</tr>
<tr>
<td>LT Cs6 + CS21</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>LT/Stp CS6 + CS21</td>
<td>5 (2.6)</td>
</tr>
<tr>
<td>LT Cs6 + CS8</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>LT CS7</td>
<td>6 (3.1)</td>
</tr>
<tr>
<td>LT CS8</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>LT CF-negative</td>
<td>54 (28.1)</td>
</tr>
<tr>
<td>LT/ST CF-negative</td>
<td>20 (10.4)</td>
</tr>
</tbody>
</table>

aCF, colonization factor.

(394x792)
only—which are rare combinations—were identified as LT8. The group IV variants found by Lasaro et al. included LT4 and LT6, which were not found in our study. LT4 is identical to porcine LT (LTp) and displays 3 additional amino acid changes in the B subunit from that of LT1 (15, 25). The LT4 variant is commonly found in porcine ETEC strains, and it is thus not surprising that we did not find it in our collection of strains from clinical isolates. Finally, the new group V included only the LT11 variant.

TABLE 2 Frequency and characterization of polymorphisms among natural variants of LT detected among ETEC strains analyzed in this study

<table>
<thead>
<tr>
<th>No.</th>
<th>LT variant</th>
<th>Alternative designation</th>
<th>No. (%) of ETEC strains (n = 192)</th>
<th>Amino acid substitution(s) in:</th>
<th>No. of amino acid replacements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A subunit</td>
<td>B subunit</td>
</tr>
<tr>
<td>1</td>
<td>LT1</td>
<td></td>
<td>78 (40.6)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>LT2</td>
<td>LT3 + R13H</td>
<td>48 (25)</td>
<td>S190L, G196D, K213E, S224T</td>
<td>T75A</td>
</tr>
<tr>
<td>3</td>
<td>LT3</td>
<td></td>
<td>6 (3.2)</td>
<td>K213E, R235G</td>
<td>R13H</td>
</tr>
<tr>
<td>4</td>
<td>LT7</td>
<td></td>
<td>2 (1)</td>
<td>P12S, S190L, G196D, K213E, S224T</td>
<td>T75A</td>
</tr>
<tr>
<td>5</td>
<td>LT8</td>
<td></td>
<td>7 (3.6)</td>
<td>T203A, K213E</td>
<td>R13H</td>
</tr>
<tr>
<td>6</td>
<td>LT11</td>
<td></td>
<td>7 (3.6)</td>
<td>M371I, T193A, K213E, I232 M</td>
<td>R18H, M37I</td>
</tr>
<tr>
<td>7</td>
<td>LT12</td>
<td>LT12 + R18H</td>
<td>2 (1)</td>
<td>R18H, M23I</td>
<td>R18H</td>
</tr>
<tr>
<td>8</td>
<td>LT13</td>
<td></td>
<td>4 (2.1)</td>
<td>H27N</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>LT17</td>
<td></td>
<td>12 (6.3)</td>
<td>G196D</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>LT18</td>
<td></td>
<td>1 (0.5)</td>
<td>S216T</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>LT19</td>
<td></td>
<td>3 (1.6)</td>
<td>D170N</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>LT20</td>
<td></td>
<td>1 (0.5)</td>
<td>H27Y</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>LT21</td>
<td></td>
<td>1 (0.5)</td>
<td>S190L, T193A, G196D, K213E, S224T</td>
<td>T75A</td>
</tr>
<tr>
<td>14</td>
<td>LT22</td>
<td></td>
<td>1 (0.5)</td>
<td>V103I</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>LT23</td>
<td></td>
<td>1 (0.5)</td>
<td>P12S</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>LT24</td>
<td></td>
<td>1 (0.5)</td>
<td>S228L</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>LT25</td>
<td></td>
<td>1 (0.5)</td>
<td>P12S, E229V</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>LT26</td>
<td></td>
<td>1 (0.5)</td>
<td>R237Q</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>LT27</td>
<td></td>
<td>1 (0.5)</td>
<td>R237Q</td>
<td>1</td>
</tr>
</tbody>
</table>

FIG 1 Phylogenetic analysis of the LT variants. An unrooted phylogenetic tree was used to determine the phylogenetic relatedness of LT variants, including the LT variants reported previously (LT1 to LT16) (15) and the new LT variants found in this study (LT17 to LT28). The tree was constructed by the neighbor-joining method using MEGA, version 5.2.
A majority of LT-ETEC strains that express known colonization factors belong to the two major LT variants LT1 and LT2, which have spread globally. Since the ETEC isolates in our study were collected over more than 3 decades from remote regions across the world, we were interested in determining if LT variants have evolved over time or show geographic clustering. Therefore, a phylogenetic tree was constructed based on the concatenated LTA and LTB peptides, and metadata were mapped back onto the tree. The overall result of the phylogenetic analysis revealed three distinct clusters, which were designated A, B, and C (Fig. 2). The topology of the tree shows that cluster A contained closely related LT variants belonging to group I. Cluster B included LT variants of groups III, IV, and V, which showed a distant branching, while cluster C included LT variants of group II. Interestingly, no clear relation was found with the country or year of isolation. However, the clusters shared distinct CF profiles. Cluster A is composed of two subclusters, designated A1 and A2. A1 harbored the majority of the isolates, whereas subcluster A2 contained 12 LT18 isolate with CS12 or CS6 + CS21. Cluster A1 harbored strains with diverse CF
profiles, including CS1 + CS3 (+ CS21), CS2 + CS3 (+ CS21), CS2 + CS1, CS3 + CS1, CS4 + CS6, CS6 + CS8, CS6 + CS21, CS7, CS17, CS19, and CS21 as well as CF-negative strains. Some of these strains belonged to major lineages of ETEC. Most of these cluster A strains in subclusters A1 and A2 had the LT1 allele, while a minority belonged to LT12, LT13, and LT17 to LT28. Single amino acid substitution variants of LT1, representing novel LT variants, were found mainly in single CF-negative ETEC isolates of cluster A (Fig. 2). Cluster A strains were isolated over 30 years from the Americas, Africa, and Asia. Hence, the LT1 variant of LT is a conserved variant that has persisted in several lineages, with different CF profiles that have spread globally over time. Cluster B harbored LT3, LT8, and LT11; the first two variants were found in CS1-, CS8-, and CS12-positive isolates, while LT11 was found only in CF-negative strains. The 19 ETEC strains of cluster B were isolated from the Americas and Asia during the period 1983 to 2009. Cluster C harbored lineages including CS5 + CS6-, CS14-, CFA/I + CS21-, CS21-, and CS23-positive isolates, as well as CF-negative strains with the majority expressing LT2 (except for 2 CF-negative isolates that expressed LT17 and LT22). Strains in cluster C were isolated from the Americas, Africa, and Asia over a period of 31 years, suggesting that LT2 has spread globally.

**Distribution of polymorphic sites along the LT protein.** The B subunit was much more conserved (only 2 amino acid substitutions) than the A subunit, which exhibited 22 amino acid changes. The A2 domain was slightly more diverse (13 amino acid substitutions) than the A1 domain (9 amino acid changes). Most of the amino acid substitutions in A1 were located between positions 12 and 37 (5 amino acid changes) and between positions 103 and 190 (4 amino acid changes), involving different structural folds within the protein, including an α-helix and β-sheets. Perhaps not surprisingly, no polymorphisms were found in the A1 subunit loop comprising residues 47 to 56, which covers the active site. These residues were also found to be under purifying selection, indicating that they are conserved (see Fig. S1 in the supplemental material). The 13 polymorphic sites of the A2 domain were distributed along the α-helix, which interacts with the B subunit; residues under positive selection were identified, but these changes were not significant (see Fig. S1 in the supplemental material). The R15H and T75A amino acid changes found in the B subunit were located in structures that form a turn and α-helix, respectively.

To analyze the potential impact of the amino acid substitutions, we modeled the LT1AB5 and LT2AB5 (Fig. 3a) complexes based on the crystal structure 1LTS. The model complexes were refined during a 2-ns MD simulation in an explicit water box. During the 2-ns simulation, the LTB domain pentamers were refined during a 2-ns MD simulation in an explicit water box.

**LT2-expressing strains produce significantly more LT than strains that express LT1.** The amino acid sequence differences in the various LT variants could have an impact on the stability and/or folding of the toxin itself and could therefore impair production and secretion (6). To examine this, we performed a single-read ELISA to assess total LT assembly by ETEC strains expressing different variants. A total of 155 ETEC strains were included in this analysis, representing 80.7% of the strains used in this study. As a preliminary test, bacterial cell lysates were analyzed by GM1-ELISA, and OD450 (optical density at 450 nm) values were normalized to bacterial numbers (an OD600 of 0.8 corresponds to 10^9 bacteria). Strains were categorized as high, medium, or low LT producers. The amounts of LT produced were high for LT2- and LT21-expressing strains (OD450 >0.5), medium for LT11 and LT13 (OD450 0.5 to 0.25), and low for LT1 and LT18 (OD450 <0.25) (Fig. 4).

More-detailed analyses of LT production and secretion by LT1 and LT2 strains were performed using quantitative GM1-ELISA. These analyses revealed that LT2 strains produced 5-fold more LT than LT1 strains (30.77 ng/ml versus 6.53 ng/ml) (P < 0.001). Similar results were obtained using the pellet and supernatant fractions (Fig. 5A and B). In the pellet fraction, LT2 ETEC produced 9-fold more LT than LT1 strains (P < 0.001), and in the supernatant fraction, LT2 ETEC produced 3-fold more LT than LT1 strains (P < 0.05). Next, the ability to secrete LT was analyzed as a percentage of the formed toxin found in the supernatant and was calculated from the toxin in the supernatant divided by total production in both the pellet and the supernatant multiplied by 100. When the secretion percentage was determined, almost equal values were found (50.29% for LT1 and 50.91% for LT2), and no statistical difference was found (Fig. 5C). Thus, secretion rates are similar for strains expressing LT2 and LT1.

**LT1 and LT2 toxin variants are equally stable.** Once the LT and LB subunits reach the periplasm, they assemble into the holotoxin. This formed holotoxin is remarkably stable; however, changes in the LT amino acid sequence could influence absolute stability (6). To determine whether LT1 and LT2 have differences in their stability, we measured the amount of LT1 and full folded LTB subunits in each isolate by GM1-ELISA. The ELISA was performed on 16 LT1 and 15 LT2 strains using two different monoclonal antibodies: one targeting the LT subunit specifically, which detects the intact LT holotoxin (when bound to GM1 via the B5 subunit), and a second targeting the total B subunit (which can detect both holotoxin and free B subunits bound to GM1 but without the A subunit). A ratio between the amounts of LTAB and LTB was calculated to infer LT stability. When the amounts of stable LT expressed by LT1 and LT2 strains were compared, the ratios were slightly

---

**January 2015 Volume 197 Number 2**

**Journal of Bacteriology**

**jb.asm.org**
different in the pellet and supernatant fractions (Fig. 6A). In the pellet, the ratios for the LT1 and LT2 holotoxins were 0.14 and 0.11, respectively. In the supernatant, LT2 strains reached a ratio of 0.19 and LT1 strains reached a ratio of 0.17. However, no statistical differences were found in the analysis of the amino acid differences between the two major LT variants; LT1 and LT2 did not affect the assembly of LT holotoxins in the periplasm or the stability of secreted, fully folded LT (Fig. 6B).

FIG 3 Structural analysis of the LT1 and LT2 variants. (a) The model of LT2 (AB5) is shown as a ribbon diagram, with select residues and regions represented by spheres and surface patches, respectively. The model was generated using the crystal structure LT8 as the template. The last conformation of a 2-ns MD simulation of the model is shown. The A and B subunits are represented by light blue and gray ribbons. Red spheres represent the A75 atoms on LT2B, and blue spheres represent the atoms of L190, D196, E213, and T224. Brown patches represent LT2A surface-exposed portions of residues that are predicted to be in protein-protein interface regions (Tyr24, Ser28, His45, Phe49, Asp50, Arg51, Gly52, Thr53, Gln54, Met55, Asn56, Gly69, Val71, Ser81, Leu82, Ser83, Leu84, Arg85, Ser86, His88, Leu89, Ala90, Gln92, Ser93, Ile94, Ser96, Gly97, Tyr98, Ser99, Thr100, Tyr102, Asn114, Val115, Asn116, Asp117, Val121, Tyr122, Ser123, Pro124, His125, Pro126, Tyr127, Glu128, Gln129, Glu130, Trp145, Tyr146, Arg147, Asn149, Phe150, Gly151, Val152, Ile153, Asp154, Gly155, Arg156, Leu157, His158, Ile173, Pro175, Ala176, Glu177, Asp178, Tyr180, Arg181, Arg193, Glu194, Gln195, Pro196, Trp197, Ile198, His199, His200, Ala201, Asn207, Leu208, Ser209, Asn223, Leu224, Ile227, Tyr228, Arg230, Glu231, Lys237, Arg238, Ile240, Phe241, Tyr244, Gln245, Tyr251, Asn252, Arg255, Glu257, and Leu258). (b) Structural fit of the LT1B (green ribbon) and LT2B (gray ribbon) pentamer models. The T or A residue at position 75 on the LTB subunit is indicated by red spheres. (c) LT1A contacts LT1B at position S224 at the inner top rim of the pentamer (blue spheres and bonds). Residue T75 (red spheres) on the LTB subunit makes only intramonomer contacts. Atoms in contact range (d < 3.5 Å) are shown in yellow spheres, and the side chain bonds of these residues are also shown in yellow. Residue K213 is solvent exposed. (d) Same as panel c but only for LT2.
DISCUSSION

ETEC strains are characterized by the production of one or both of LT and ST. However, ETEC is a highly diverse pathogen with regard to serotype, toxin type, and CF expression (1, 26). We were initially interested in determining whether LT evolves into more-virulent variants over time and whether we could follow this in a set of strains isolated over 3 decades. The recent discovery of several natural variants in the LT sequence among ETEC strains isolated in Brazil (15), as well as the emergence of novel variants of the related cholera toxin (27), prompted us to evaluate the diversity of LT (LT-I) in human ETEC strains on a global scale. We identified novel LT variants and could show that both known and novel LT variants are associated with specific ETEC CF profiles, but no association with time was found. In fact, we have recently reported that globally distributed ETEC lineages with stable toxin-CF profiles have been stable over substantial periods in areas of endemicity (18).

In our 192 human ETEC strains isolated from children and adults globally over 31 years, 20 natural variants of the translated LT were identified. This finding, together with previous studies (8, 28), demonstrated a remarkable genetic diversity of LT in ETEC strains with worldwide distribution. Lasaro et al. (15) reported 16 LT variants in a set of 51 Brazilian ETEC strains, and we found 8 of these 16 LT variants, as well as 12 additional novel variants, among ETEC strains isolated from areas of endemicity all over the world. These findings support previous results showing the heterogeneity and variability within the LT sequence and demonstrate even greater LT diversity than was previously known. Some LT variants identified (15) were not found in this study, suggesting that some variants may be nonclinical isolates, or they could have a local impact or circulate in specific geographic locations. This may also suggest that certain mutations occur temporarily but do not persist in the population of human ETEC strains. In support of this, most of the novel LT variants identified in our study were found only in a single strain, while the more common LT variants, such as LT1, LT2, LT11, LT13, and LT18, seem to be associated with ETEC lineages that are widespread and not only circulate among continents but also persist over time.

FIG 4 Differences in the amount of LT among strains expressing different LT variants as determined using GM1-ELISA. A total of 155 ETEC strains expressing 16 of 20 LT variants were tested for LT production. The number of strains expressing the respective LT variant is given above each box plot. The OD values came from the one-read ELISA. Results shown are averages of the OD values of triplicate samples. ODs of 0 to 0.25, 0.25 to 0.5, and >0.5 were considered to indicate small, medium, and large amounts of LT produced by ETEC strains, respectively. Statistical analysis was performed using ANOVA with Turkey’s multiple-comparison posttest by using Prism, version 6.00. For the statistical analysis, LT22, LT23, LT25, LT26, LT27, and LT28 were excluded due to the low numbers of samples. *, P < 0.05; **, P < 0.01; ***, P < 0.001.

FIG 5 Production and secretion of LT by ETEC strains expressing LT1 and LT2 variants as determined by quantitative GM1-ELISA. (A) Total production in LT1 and LT2 strains. (B) Comparison of LT production in LT1 and LT2 strains using the sonicated bacterial (Pellet) and supernatant (Sup) fractions. (C) Percentage of LT secretion. Results are means from duplicate independent tests of each strain. The statistical analysis was performed by the Mann-Whitney test using Prism, version 6.00. Horizontal lines indicate median values. *, P < 0.05; **, P < 0.01; ***, P < 0.001; n.s., not significant.
Lasaro et al. (15) also reported that the 16 LT variants were distributed in 4 different clusters. In agreement with their findings, our phylogenetic analysis, including the 16 previously reported LT variants and the 12 new LT variants identified in this study, showed a similar topology. However, we found that the LT11 variant formed a fifth group. This was even more obvious when the phylogenetic tree was based on the SNPs, where LT11 clearly branched off from the rest of the LT variants (see Fig. S2 in the supplemental material). We also observed that 11 of the 12 newly identified variants in this study belonged to group I—centering around LT1—and that most novel LT variants were closely related to LT1. The 12th novel variant fell into the divergent group II formed by LT2. While group I was rooted in LT1, group II was rooted in the LT2 allele, suggesting that LT1 and LT2 could be ancestors for their respective groups. Hence, groups I and II are the most important groups in human ETEC strains, since they encompass the majority of LT variants and strains with a wide-spread distribution, as well as being found in strains isolated over the entire study duration of 31 years.

We could not find a strong association between specific LT variants and the geographic distribution or year of isolation among the strains analyzed in this study, suggesting that similar polymorphisms in the LT gene could be present in different regions of the world and at different time points (Fig. 2). In contrast, we found a strong relation between the presence of specific LT variants and the CF profile. For instance, CS1, CS2, and CS3 were expressed only in LT1 strains, while CS5 + CS6 and CFA/I expression was associated with LT2-expressing strains. This finding suggests that there is a link between the acquisition of the LT gene and a particular colonization factor by means of lateral transfer of chromosome- and plasmid-borne genes. Our results are in agreement with previous observations showing that ETEC strains expressing the same virulence profile (toxin-CF) fall into the same clonal groups regardless of the place of isolation (18, 29–34). These data also suggest that a possible clonal expansion of ETEC strains expressing the LT variant ancestors LT1 and LT2 could have occurred by means of human migration and travel. In fact, we show that two clusters, A and C, make up the majority of the ETEC strains (Fig. 2). Cluster A is a highly diverse group that includes a large number of LT variants (group I) with a broad range of colonization factor profiles. Also, this cluster is the most polymorphic due to the high number of single amino acid substitutions among the LT sequences. However, the LT sequences of cluster A are all rooted in the LT1 variant, and strains expressing LT1 also express colonization factors such as CS1, CS2, CS4, CS17, and CS19, which were previously reported to belong to the CFA/I family with similar genetic and biochemical features (35–37). However, the strains that express variants related to LT1 were more often colonization factor negative and were present only in a single or few strains. It was reported that the presence of separated clusters is a consequence of recent genomic changes, suggesting that these related LT variants could have emerged and again disappeared recently, while strains with LT1 retain their colonization factors and are persistent virulent strains (33). The second largest cluster, cluster C, contains strains that express CFA/I, as well as CS5 + CS6. Fewer related LT variants are found within this group, but most derivatives from the ancestral variant LT2 were, again, CF negative.

Clusters A and C represent two divergent and prevalent populations of LT-ETEC strains. This suggests that since the majority of the colonization factors and toxin are normally encoded on plasmids, the different LT variants have been acquired together with certain colonization factors on the same plasmid or a compatible plasmid(s) (31, 38, 39). Although further analyses are required to demonstrate whether LT and colonization factors are physically located on the same plasmid, our data suggest that the alleles of both toxins and CFs are conserved within lineages and hence might have been acquired simultaneously by one ancestor strain at one point and then spread clonally. A previous report indicated that around 130 million years ago, before V. cholerae and E. coli diverged as species, LT genes were acquired by horizontal transfer (40). Also, it has been known that the LT sequence is flanked by insertion sequence (IS) elements, similar to those found next to genes encoding fimbriae, suggesting a general mechanism for the transmission of virulence-related genes (41, 42). Our data, together with the findings that ETEC strains with the same toxin-CF profile often are genetically related, suggest that LT acquisition is not due solely to horizontal gene transfer but rather is also due to lateral gene transfer.

When studying the natural diversity of LT, we observed more polymorphisms in the A subunit than in the B subunit, where only 2 amino acid substitutions were identified (in contrast to 22...
changes in the A subunit). A previous report (43) found that single mutations in the A subunit (K63, D53, K7, K104, K97, and K114) and the double mutation K7 and K97 caused a considerable decrease in the proportion of fully assembled molecules of LT. However, in our study, the four mutations identified in the LT2 A subunit apparently did not affect the assembly of the LT molecule, suggesting that these polymorphic sites are not involved in the formation of the AB complex. This is supported by the fact that these variants are present in clinical isolates from patients with diarrhea and hence are expected to express a virulent LT toxin. On the other hand, we identified a considerable number of polymorphic places in the A2 helix domain of the A subunit. This structure is located near the B pentamer and continues into the pore of the B subunit, creating points of hydrophobic interactions between A and B subunits. Here we found that S224T (LT2, LT7, and LT22) and S228L (LT26) in LTA are located in close proximity to A2-B interaction residues, i.e., close to T75A in LTB; such polymorphism could possibly affect the positioning of the A subunit during holotoxin assembly (44, 45). However, our in silico protein modeling does not suggest that the T75A substitution in LTB would affect the stability of the holotoxin. Based on our results, however, we cannot determine whether export to the periplasm or efficiency of assembly is affected by the amino acid substitutions. A previous study reported that deletion of the last 14 residues of the LTA subunit could dramatically affect holotoxin assembly but also that deletion of the last 4 amino acids could be critical for the stability of the toxin (46). We found that LT28 (n = 1) and LT23 (n = 1) have an amino acid change at residues in that critical region. These strains displayed very low levels of LT production, which may be related to a deficiency in holotoxin assembly due to a reduction in LT4-LTB interaction. However, the effect of polymorphism in this region needs to be studied in more detail.

The LTB subunit was more conserved than the LTA subunit, possibly reflecting host specificity, since the B subunit binds to host receptors. LTB binds to GM1, glycoproteins, and glycolipids, possibly reflecting host specificity, since the B subunit binds to the carbohydrate epitopes of the ABO blood group system (47), and specific amino acid substitutions can interfere with binding (6, 48, 49). For instance, amino acid changes at residues 46, 47, and 57 have been reported to diminish binding affinity, since they were located close to the binding pocket (25, 26). Additional mutations in the LTB sequence have been described before in LTp (isolated from pigs), and these polymorphisms resulted in reduced binding to human GM1 and blood sugars (8, 48). In this study, such mutations were not identified. We found amino acid changes at residues 13 (LT3 and LT8) and residue 75 (LT2) among high-LT-producing strains, which are not involved in direct binding to GM1, although residue 13 is close to a proposed binding site. A histidine at residue 13 was found in strains that clustered in group B, which are the closest relatives to porcine strains (28). These data corroborate our observation that strains expressing LT2 produce more toxin than strains expressing LT1 under laboratory conditions. However, whether this is the case in the human small intestine remains to be investigated.

In summary, ETEC strains that express either the LT1 or LT2 variant express the most prevalent colonization factors associated with the occurrence of diarrheal disease worldwide (2, 50), and major lineages expressing specific colonization factor profiles are linked to the two variants. Although LT2 strains express significantly larger amounts of LT than LT1 strains, both LT1 and LT2 ETEC strains are frequently and repeatedly identified in cases of severe diarrhea worldwide and over time, supporting their virulence and successful dissemination.

ACKNOWLEDGMENTS

This study was supported by Swedish Research Council grant K2012-56X-22209-01-3, VINNOVA grant 2011-03491, and a grant from Groschina’s Sky’s Foundation to Å.S. and by Swedish Foundation for Strategic Research (SSF) grant SB12-0072 to A.-M.S. and Å.S. The project was performed as part of the UMSA-IBMB Diarrheal Disease Project supported by the Swedish Agency for Research Economic Cooperation (SIDA) (to A.-M.S. and Å.S.), E.J. acknowledges financial support from the Swedish Institute and the International Science Programme (ISP). We also acknowledge RO1 NIAID AI0094001 funding to T.S.

We acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high-performance computing resources that have contributed to the research results reported in this paper (http://www.tacc.utexas.edu).

REFERENCES


