Novel type II heat-labile enterotoxin produced by human disease outbreak isolates of *Escherichia coli* is a member of the LT-IIa family and should not form a new group: clarification of nomenclature and precedent.

Michael G. Jobling.

Retired faculty, Department of Immunology and Microbiology, School of Medicine, University of Colorado Anschutz Medical Campus. RC1N, 12800 E 19th Ave, Aurora, CO 80045 USA.

*Michael.Jobling@UCDenver.edu* tel. (720) 427-2295

Comment on “*Escherichia coli* O8:H8 Carrying a Novel Variant of the Heat-Labile Enterotoxin LT2 Gene Caused Outbreaks of Diarrhea.” Kai Ishimaru, Mari Sasaki, Hiroshi Narimatsu, Yoko Arimizu, Yasuhiro Gotoh, Keiji Nakamura, Tetsuya Hayashi, Yoshitoshi Ogura. Open Forum Infectious Diseases, Volume 7, Issue 1, January 2020, ofaa021, [https://doi.org/10.1093/ofid/ofaa021](https://doi.org/10.1093/ofid/ofaa021)
Dear Editor,

I was very interested to read this report [1] describing outbreaks of human disease caused by type II (LT-II-producing) enterotoxigenic *Escherichia coli* (ETEC). While type II ETEC have been isolated from both symptomatic and healthy individuals [2], they have not previously been documented to cause outbreaks in humans, and the majority were isolated from animals [3]. This report is clear and convincing but does, however, have several issues that need addressing. First, the authors use non-standard nomenclature for the two types of heat-labile enterotoxins that are, by consensus, referred to as LT-I and LT-II [4], not LT1 and LT2. In fact, LT1 and LT2 have been used to denote specific lineages of LT-I toxins [5]. Second, I disagree with the assertion that the novel LT-II toxin “formed a branch distinct from known elt2 variants” (that was designated LT2d) since, while it is clearly novel, my analysis shows it is closely related to LT-IIa and should be described as an LT-IIa variant. Third, in a previous publication cited in the report [6], I noted the existence of a novel LT-II locus in an avian *E. coli* genome (B088; NZ_GG749133), distinct from all then known LT-IIa, b or c variants (Fig. 1C). This should take precedence for the LT-IIId designation.

One unique detail regarding LT2d is that its A subunit is seven residues shorter than all others (Fig. 1A), and lacks a recognizable R/KDEL sequence at its carboxyl terminus, important for full toxicity and intracellular trafficking [7]. Nevertheless, this novel LT-II toxin is clearly able to function both in vitro and in vivo. However, a global alignment of the complete LT2d locus with LT-IIa reveals that this is caused by a single bp deletion in the C-terminal coding region of LT2d A, changing the reading frame (Fig. 1A); a new stop codon terminates the A polypeptide, absent which the A gene would be translationally fused to the B gene. This is insufficient to designate this toxin as a new LT-II subgroup.

The report showed neighbor-joining trees for the A and B genes, separately, that produced trees with different branch points, used to claim LT2d forms a branch distinct from all other LT-II toxins. In analyzing the ancestry of these toxins, we must, however, consider the locus as a whole - both the A and B genes together - since this is the evolutionary unit upon which selection acts. A neighbor-
joining analysis of complete operons (Fig. 1B). clearly shows that LT2d, and a further novel variant I found in the NCTC10082 genome (NZ_UFZA01), are more closely related to each other and to LT-IIa than to any other LT-II. Comparing B polypeptides, which are the most diverse subunit, both LT2d and the NCTC10082 toxin each have two lengthy regions (of 31-50 residues, together encompassing 94% of the mature B subunits) with 90-98% identity to LTIIa B (Fig. 1C). This confirms they are descended from LT-IIa. This degree of identity and conservation is seen elsewhere only between toxin variants that are grouped into the LT-IIc family. I propose LT2d (and other homologous loci e.g. NCTC9109, NZ_UGDU01; Fig. 1B) be renamed LT-IIav1 (LT-IIa variant 1), and the toxin from NCTC10082 designated LT-IIav2, as novel members of the LT-IIa family. A 500-repetition bootstrap analysis supports this conclusion, as it shows >= 95% confidence in all nodes (100% for the LTIIc and new LT-IIa families) except for the LT-IIa family-b/d split (81%) and internal LT-IIc family nodes ; Fig. 1B). While the degree of difference (average changes per bp) for designating a new subgroup is somewhat arbitrary, these novel variants clearly group with LT-IIa at the operon level and have highly similar B subunits.

In conclusion, this is a very interesting report documenting a novel LT-II toxin, that is more accurately described as an LT-IIa variant than a novel LT-II group; the LT2d designation should be replaced with LT-IIav1, with LT-IId being reserved for the B088 toxin locus. I agree with their conclusion that “more attention should be paid to infections by E. coli strains producing LT(-II) with colonization factors”; this should be expanded to include all LT-II-encoding isolates, since it is clear from my analysis of other genomes that the LT-II family is even more diverse than currently accepted (awaiting demonstration of toxicity by genome-identified isolates). This study suggests that acquisition of a human colonization factor (plasmid) may be all that is required to confer on animal and environmental type II ETEC isolates the ability to colonize and produce disease in humans.

Conflict of Interest. Author declares no conflict of interest.
References

1. Ishimaru K, Sasaki M, Narimatsu H et al. *Escherichia coli* O8:H8 Carrying a Novel Variant of the Heat-Labile Enterotoxin LT2 Gene Caused Outbreaks of Diarrhea. *Open Forum Infect Dis* 7(1), ofaa021-ofaa021 (2020).

2. Seriwatana J, Echeverria P, Taylor DN et al. Type II heat-labile enterotoxin-producing *Escherichia coli* isolated from animals and humans. *Infect. Immun.* 56(5), 1158-1161 (1988).

3. Jobling MG, Holmes RK. Type II Heat-Labile Enterotoxins from 50 Diverse *Escherichia coli* Isolates Belong Almost Exclusively to the LT-IIc Family and May Be Prophage Encoded. *PLoS ONE* 7(1), e29898 (2012).

4. Holmes RK, Twiddy EM, Pickett C, Marcus H, Jobling MG, Pettijean F. The *Escherichia coli/Vibrio cholerae* family of enterotoxins. In: *Microbial Toxins in Foods and Feeds Molecular modes of action*, Pohland AE, Dowell VR, Jr., Richard JL (Eds). Plenum Press New York, NY 91-102 (1990).

5. Joffré E, Sjöling Å. The LT1 and LT2 variants of the enterotoxigenic *Escherichia coli* (ETEC) heat-labile toxin (LT) are associated with major ETEC lineages. *Gut microbes* 7(1), 75-81 (2016).

6. Jobling MG. The chromosomal nature of LT-II enterotoxins solved: a lambdoid prophage encodes both LT-II and one of two novel pertussis toxin-like toxin family members in type II enterotoxigenic *Escherichia coli* (ETEC). *Pathog. Dis.* 74(3), ftw001 (2016).

7. Lencer WI, Tsai B. The intracellular voyage of cholera toxin: going retro. *Trends Biochem. Sci.* 28(12), 639-645 (2003).
8. Felsenstein J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* 39(4), 783-791 (1985).

9. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. USA.* 101(30), 11030-11035 (2004).

10. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33(7), 1870-1874 (2016).
Figure 1.

A. Sequence comparison (ClustalW alignment) of the A-B gene junctions in LT2d and LT-IIa. DNA sequences shown in lowercase, translation in uppercase (LT2d, above, LT-IIa, below DNA sequence). * and x denote termination codons. Identities are shown by a period, differences highlighted with a teal background; single bp deletion (-) is highlighted with a red box. Orig RF denotes original reading frame for the LT2d A subunit sequence, now degenerated. Coding frame for B subunits has been maintained.

B. Neighbor-joining bootstrap analysis of evolutionary relationships of ClustalW aligned DNA sequences for all LT-II operons (start codon of A to stop codon of B genes), with LT-I as an outlier. The optimal tree with the sum of branch length = 1.93925762 is shown. The percentage of replicate trees in which the associated toxin operons clustered together in the bootstrap test (500 replicates) are shown next to the branches [8]. The tree is drawn to scale (bar shows 0.05 base substitutions per site), with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method [9]. The analysis involved 14 nucleotide sequences, including an additional novel LT-II toxin, LT-Ile, identified in the VREC0623 genome (NZ_UDBM01000062). All positions containing gaps and missing data were eliminated. There were a total of 1023 positions in the final dataset.

Evolutionary analyses were conducted in in MEGA7 [10] on ClustalW aligned sequences output from Clone Manager Professional v9.51, (SciEd software, Denver, CO). Red boxes group highly related toxins into families.
C. ClustalW amino-acid alignment of all LT-II B polypeptides to LT-IIa B. Amino-acids shown in single letter lowercase, with identities to LT-IIa B shown by a period, differences highlighted with a teal background. Black box encloses signal sequences, red boxes show regions of extreme conservation (90 – 98% identity) of LT2d (LT-Illav1), and LT-Illav2 B polypeptides, with LT-IIa B, encompassing all but 6 of the 100-residues of the mature B subunit. For clarity, only three of the six known LTIIc variants (representing the three LTIIc sub-branches, Fig. 1B) are shown.
