## ***Epidemiology and Infection***

Importance of Case Age in the Purported Association between Phylogenetics and Hemolytic Uremic Syndrome in *Escherichia coli* O157:H7 Infections

G. A. M. Tarr1, S. Shringi2, H. N. Oltean3, J. Mayer4,5, P. Rabinowitz6, J. Wakefield7, P. I. Tarr8, T. E. Besser2, A. I. Phipps4

1 Department of Pediatrics, University of Calgary Cumming School of Medicine

2 Veterinary Microbiology and Pathology, Washington State University

3 Washington State Department of Health

4 Department of Epidemiology, University of Washington

5 Department of Geography, University of Washington

6 Department of Environmental and Occupational Health Sciences and Center for One Health Research, University of Washington

7 Departments of Biostatistics and Statistics, University of Washington

8 Department of Pediatrics, Washington University in St. Louis School of Medicine

## **Supplementary Material**

**Assigning Phylogenetic Lineage to Non-SNP-typed Isolates**

In previous *E. coli* O157:H7 phylogenetic classifications used to assess the association with HUS, it has been common to use a single representative isolate from each PFGE subtype [1-3]. This practice masks variability among isolates with the same PFGE fingerprint (e.g. variability in demographics, location, etc.). Further, estimation of effects at the population level is compromised, because the isolates being analysed might not reflect the *E. coli* O157:H7 case population distribution. To accurately make inference at the population level, we sought to include all reported cases during the study period. Because we did not have sufficient resources to SNP-type all isolates, we leveraged the assumption inherent in the single-representative-isolate approach, although not generally made explicit: isolates with the same PFGE fingerprint belong to the same phylogenetic group.

Our sample contained 1160 isolates reflecting 355 unique PFGE patterns. We SNP-typed 793 of these isolates, covering 319 PFGE subtypes. The 36 PFGE subtypes not SNP-typed were either biochemically atypical or not present in the isolate bank. Atypical isolates were exclusively recovered in 2013 and 2014, the last 2 years of sampling. Missing isolates were predominantly (82%) from 2005 and 2006, the first 2 years of sampling. Of the 793 SNP-type isolates, 570 belonged to a PFGE subtype with multiple SNP-typed isolates. Among these 570, we examined which phylogenetic lineages the isolates had been assigned via SNP-typing. All but one PFGE subtype were assigned a consistent lineage. The one variable PFGE subtype was EXHX01.0047, associated with 82 isolates: 21 were not typed, 59 were typed to lineage IIa, and 2 were typed to lineage Ib (Supplementary Figure S2). In other words, only 2 of 570 isolates, or 0.4%, showed aberrant lineage assignment. With this, we felt that the assumption that isolates of the same PFGE subtype would be in the same lineage held adequately well to use the SNP-typing results to assign lineage to non-SNP-typed isolates. We were able to assign lineage to 328 additional isolates using this approach.

For the EXHX01.0047 PFGE type containing the two aberrant isolates, we examined the secondary PFGE patterns upon suggestion of a reviewer. Four secondary patterns within the EXHX01.0047 XbaI pattern were present in our sample: EXHA26.0015 (39 IIa, 14 untyped), EXHA26.0071 (7 IIa, 4 untyped), EXHA26.0332 (2 Ib, 11 IIa, 1 untyped), and EXHA26.3261 (1 IIa). Three isolates (1 IIa, 2 untyped) did not have secondary patterns typed. Both aberrant isolates typed to the EXHA26.0332 pattern. This pattern was shared by 12 other EXHX01.0047 isolates. However, 11 of those were SNP-typed to lineage IIa. Only one isolate with this combination of patterns was not SNP-typed and thus assigned lineage IIa based on PFGE pattern. Two additional isolates did not have a secondary pattern and were not typed, so they could also theoretically have been misclassified. None of these three cases was hospitalized or had HUS. The EXHX01.0047-EXHA26.0332 combination of patterns is one of the most common in the U.S., and EXHA26.0332 has been associated with several other XbaI patterns. It is not possible for us to be sure that lineage misclassification, if it occurred, was limited to at most the three isolates described above. However, all data available to us suggest low risk of lineage misclassification.

**Manning Clades**

 We used the 32-plex SNP assay used in the 2008 analysis of phylogenetic clades and HUS status by Manning et al. [2] to determine clade. Manning clades 1, 2, 3, 6, 7, and 8 were observed in our dataset, including a number of isolates that could not be distinguished between clades 2 and 3 (Table S1). Of the 480 isolates clade-typed, 6 resulted in an error. The remaining 474 isolates encompassed 190 of the 355 PFGE types present in our data. As with the Bono-Jung phylogenetic lineages, PFGE type was used to infer the Manning clade of untyped isolates. Multiple PFGE types had isolates in both clades 2 and 3. Given that some isolates were already indeterminate clade 2 or 3, we combined all 2 and 3 isolates into a clade 2/3 group. All isolates in the clade 2/3 group mapped to lineage Ib. Isolates in Manning clades 1, 6, and 7 were rare in our clinical dataset and were grouped for analysis. All mapped to the clinically rare lineages (Table S1). Clade 8 contained all lineage IIa and IIb isolates.

Consistent with its lineage-typing, the EXHX01.0047 PFGE type included 2 isolates of clade 2/3, and the remaining 38 clade-typed isolates from this PFGE type were clade 8. Therefore, when inferring Manning clade from PFGE type, EXHX01.0047 isolates were assigned clade 8. Discrepancies also existed for two PFGE EXHX01.0190 and two PFGE EXHX01.1921 typed isolates. The former included clades 6 and 7, so EXHX01.0190 isolates were assigned the “rare” clade group. The latter included clades 2/3 and 7. Because neither was clearly dominant, EXHX01.1921 isolates were not assigned a Manning clade (Supplementary Figure S2).

In GEE logistic regression analyses, Manning clade 2/3 was used as the reference group against which clade 8 and the clinically rare clades were compared. No HUS cases were infected with the clinically rare clades; results are not reported for this group. Substantial methodologic differences between our analysis of the Manning clade-HUS association and prior analyses include: 1) use of all isolates from a population-based sample, with correlated data methods used to account for the lack of independence among cases infected with isolates of the same PFGE-defined strain; 2) use of a consistent reference group to facilitate comparison across studies; and 3) validated HUS status with a standardized definition.

The association between the Manning clades and HUS resembled that of the association between the Bono-Jung lineages and HUS (Table S2). Relative to clade 2/3, the risk of HUS associated with clade 8 was consistent with no effect (OR 1.44; 95% CI 0.87, 2.39). The effect was attenuated toward the null after adjustment for age and sex. Stratification by age group suggested effect modification. Clade 8 had no discernible effect in 0-4 (OR 0.73; 95% CI 0.41, 1.31), 5-9 (OR 1.66; 95% 0.81, 3.39), or 10-19 (OR 3.25; 95% CI 0.68, 15.7) year-olds. As with lineages IIa and IIb, clade 8 strains appeared to increase the risk of HUS (OR 8.22; 95% CI 1.06, 63.6) in 20-59 year-olds but were isolated from none of the eldest cases.

## **Supplement References**

(1) **Iyoda S, et al.** Phylogenetic Clades 6 and 8 of Enterohemorrhagic Escherichia coli O157:H7 With Particular stx Subtypes are More Frequently Found in Isolates From Hemolytic Uremic Syndrome Patients Than From Asymptomatic Carriers. *Open forum infectious diseases* 2014; **1**(2): ofu061.

(2) **Manning SD, et al.** Variation in virulence among clades of Escherichia coli O157:H7 associated with disease outbreaks. *Proceedings of the National Academy of Sciences of the United States of America* 2008; **105**(12): 4868-4873.

(3) **Pianciola L, et al.** Genotypic characterization of Escherichia coli O157:H7 strains that cause diarrhea and hemolytic uremic syndrome in Neuquen, Argentina. *International journal of medical microbiology : IJMM* 2014; **304**(3-4): 499-504.

## **Supplementary Tables**

Supplementary Table S1. Correspondence of Bono-Jung Phylogenetic Lineages and Manning Clades

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Clade 1 | Clade 2 | Clade 2/3 | Clade 3 | Clade 6 | Clade 7 | Clade 8 |
| Lineage Ib | 1 | 140 | 22 | 62 | - | - | - |
| Lineage IIa | - | - | - | - | - | - | 132 |
| Lineage IIb | - | - | - | - | - | - | 85 |
| Rare lineages |  |  |  |  |  |  |  |
| Ia | - | - | - | - | 3 | - | - |
| IVa | - | - | - | - | - | 4 | - |
| IVa/Va | - | - | - | - | - | 1 | - |
| IVc | - | - | - | - | - | 11 | - |
| Va | - | - | - | - | - | 4 | - |
| Vb | - | - | - | - | - | 8 | - |
| VI var | - | - | - | - | - | 1 | - |

Number of isolates typed to determine Manning clade and their corresponding Bono-Jung lineages. Manning clade typing failed for 6 isolates: 3 for which Bono-Jung lineage typing also failed, 2 lineage IIa isolates, and 1 lineage IIb isolate.

Abbreviations: var, variant

Supplementary Table S2. Association of Manning Clades and HUS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | NHUS/Total | Odds Ratio | 95% Confidence Interval | *P* |
| Crude |  |  |  |  |
| Clade 2/3 | 25/404 | 1 | - | - |
| Clade 8 | 36/420 | 1.44 | 0.87, 2.39 | 0.15 |
| Adjusteda |  |  |  |  |
| Clade 2/3 | 25/402 | 1 | - | - |
| Clade 8 | 36/420 | 1.25 | 0.75, 2.07 | 0.40 |
| Age-stratified: 0-4 years oldb |  |  |
| Clade 2/3 | 14/84 | 1 | - | - |
| Clade 8 | 16/123 | 0.73 | 0.41, 1.31 | 0.29 |
| Age-stratified: 5-9 years oldb,c |  |  |
| Clade 2/3 | 7/61 | 1 | - | - |
| Clade 8 | 10/63 | 1.66 | 0.81, 3.39 | 0.17 |
| Age-stratified: 10-19 years oldb |  |  |
| Clade 2/3 | 1/62 | 1 | - | - |
| Clade 8 | 4/78 | 3.25 | 0.68, 15.7 | 0.14 |
| Age-stratified: 20-59 years oldb |  |  |
| Clade 2/3 | 1/143 | 1 | - | - |
| Clade 8 | 6/115 | 8.22 | 1.06, 63.6 | 0.043 |
| Age-stratified: ≥60 years oldb |  |  |
| Clade 2/3 | 2/52 | 1 | - | - |
| Clade 8 | 0/41 | 0 | 0, 0 | <0.001 |

Logistic regression using generalized estimating equations (GEE) of HUS status on Manning clades. No HUS occurred in the group of cases infected with rare clades, so results are not shown for this group.

a Model adjusted for age as a continuous variable and sex.

b Model adjusted for sex.

c Model with exchangeable correlation structure did not converge. Independent correlation structure used instead.

Abbreviations: HUS, hemolytic uremic syndrome

## **Supplementary Figures**

Supplementary Figure S1. Histograms of results from 10,000 simulations of randomly selecting one isolate per PFGE-defined strain. Effect size and p-value were obtained from GEE logistic regression of HUS on lineage, adjusted for age (continuous) and sex. Effect sizes can be exponentiated to obtain ORs. The lineage IIa OR exceeded 1 (effect size 0) in 53% of simulations, and p<0.05 in 0.11%. The lineage IIb OR exceeded 1 in 97% of simulations, and p<0.05 in 25%.

Abbreviations: GEE, generalized estimating equations; HUS, hemolytic uremic syndrome; OR, odds ratio; PFGE, pulsed field gel electrophoresis

|  |  |  |
| --- | --- | --- |
| **PFGE Type** | **Typed Lineages/Clades** | **Assigned Lineage/Clade of Untyped Isolates**  |

EXHX01.0047

2 lineage Ib

2 clade 2/3

Lineage IIa

Clade 8

59 lineage IIa

38 clade 8

Rare clades

1 clade 7

1 clade 6

EXHX01.0190

1 clade 2/3

No clade type

EXHX01.1921

1 clade 7

Supplementary Figure S2. Discordant PFGE types. Of 355 PFGE types, three showed discordant lineages and/or clades when SNP-typed. One PFGE type was typed to two separate Bono-Jung lineages. It was assigned to the more abundant lineage. Three PFGE types were SNP-typed to two Manning clades. One (EXHX01.0047) was assigned to the most abundant clade. Both typed isolates from PFGE type EXHX01.0190 typed to a rare clade. One each of the typed isolates from PFGE type EXHX01.0190 typed to clade 2/3 and clade 7, a rare clade. No clade type was assigned to untyped isolates from this PFGE type.