5/16/2003

Final Report

Project Title: Distribution of virulent and avirulent subclones of *E. coli* O157:H7 in the U.S.

Principal Investigator: Andrew K. Benson
Department of Food Science and Technology
University of Nebraska

Summary of Project:

The primary objective of the project was to develop and validate a test that can rapidly discriminate different subpopulations of *E. coli* O157:H7. High resolution genotyping methods have previously demonstrated the existence of subpopulations of O157:H7 which appear to have a bias as to their ability to cause human disease or to be transmitted to humans through contaminated meat. Preliminary studies indicated that Lineage I strains are more commonly isolated from humans than strains derived from Lineage II and that the bias was not regional, but national. This lead to the hypothesis that lineage II strains were less likely to be transmitted to humans from cattle or that they were less efficient at causing disease in humans. In order to test that hypothesis, a high-throughput genetic test needs to be developed and validate for use in epidemiological studies of the two lineages. This project used high-density OBGS analysis to identify lineage-specific markers which were then multiplexed into a single rapid PCR test. The test was validated against a large strain set previously tested by OBGS analysis.

Specific aims:

The specific aims of this project were:

1. To validate the SSP (now called Lineage Specific Polymorphism—LSP) assay using sets of O157:H7 strains that have previously been characterized
2. To use the LSP assay to conduct preliminary studies of prevalence of the subclones in from prevalence studies of E. coli O157:H7 in cattle
3. To use the LSP assay to examine distribution of the subclones among clinical isolates

The project was part of a larger O157:H7 comparative genomics effort in the Benson laboratory and was supported by a grant from USDA-NRICGP and a collaborative arrangement with Li-Cor, inc.

High-density OBGS analysis was first performed on a set of 40 different O157:H& strains that represent the diversity of lineage I and II populations. All lineage-specific polymorphisms (LSPs) were then cloned and sequenced along with a large number of class II and III LSPs (lineage-specific but non-conserved). Candidate LSPs that contain short, stable mutations were then tested as markers by designing specific primers upstream and downstream of the mutation. A set of 20 different candidates was identified and narrowed to a set of 6 that were highly stable and simple to multiplex into a single PCR test.

Summary of Results

The genome coverage phase of the project employed 170 different OBGS primer combinations on the set of 40 representative O157:H7 strains. Table 1 below indicates the number and distribution of polymorphisms that were identified in the combined analyses of all primer combinations. A total of 70 class I LSPs were identified, along with 897 class II and III LSPs. Thus far, 68 of the class I polymorphisms have been cloned and sequenced, of which 57 are in independent loci. Of the 57 independent loci, 9 had mutations that are suitable as discriminatory alleles (markers) by PCR assays (e.g. they are short insertions or deletions). Six of these were then multiplexed together into a Lineage-Specific Marker Assay (LMSA). We also included an additional 4 markers that are markers of the O157:H7 lineage (*eaeA, stx1, stx2*, and *hlyA*). The six LMSA markers were fluorescently labeled with a dye that fluoresces at 700nm while the four O157:H7 markers were labeled with a dye that fluoresces at the 800nm. This allows all ten primers to be multiplexed into a single PCR reaction and provides for a simplistic analysis of the PCR products on a two-color automated DNA sequencer (Fig. 1).

To validate the LMSA test, a panel of 216 different *E. coli* O157:H7 strains were compared by OBGS and LMSA. Data analysis was performed by scoring alleles from the LMSA test (1 for allele 1, 2 for allele 2) at each locus and then combining the score for all loci. Since lineage I strains always have allele 1 at each locus, the scoring scheme is lineage I if the sum of allele numbers is 6 and lineage II if the sum is greater than six. When compared to OBGS and cluster analysis, the LMSA results yielded identical lineage information as OBGS analysis on 98% of the samples. The distribution of genotypes is shown in table 3. It should be noted that we observed several instances of unique alleles (not allele 1 or allele 2) and these were always present in lineage II strains.

If we assume the simplest case of binary character states at each locus, which would allow for 64 different allele combinations, Z-test statistics indicate that the observed distribution of these four predominant genotypes is significantly deviant from that expected at randomness ($p<0.05$). Thus, the markers show relatively tight linkage and we conclude that despite the degree of genome diversity that can be observed in these strains, population structure, and hence phylogenetic signal, has not been completely abrogated by recombination.

We are currently in the process of using the assay for two different applications. The first application is to examine the distribution of the two lineages in a large number of clinical O157:H7 isolates. We are currently testing a set of >100 strains from the Nebraska State Public Health laboratory and are in the process of building collections from other public health laboratories. Secondly, we are using the LMSA test to examine distribution of the populations in animal production environments and testing whether certain populations of the organism are more efficient than others at causing "blooms" of infection among cattle.

Table 1. Distribution of Polymorphisms from high-density OBGS coverage of representative O157:H7 strains

Class	Distribution	Number of occurrences
I	lineage-specific, conserved	70
II	lineage-specific, non-conserved, monophyletic	106
III	lineage-specific, non-conserved, polyphyletic	791
IV	lineage-independent, polyphyletic	3409

Table 2. OBGS/O157 Marker Identification

Marker	**Allele**	**PCR product size**	**Polymorphic**	**Dye**
21_13C	1	127	Yes	700
	2	137		
22A_6C	1	157	Yes	700
	2	168		
22A_2C_2	1	386	Yes	700
	2	466		
22A_2C_4	1	311	Yes	700
	2	322		
	3	329		
12_13C	1	268	Yes	700
	2	277		
10_7C	1	207	Yes	700
	2	210		
	3	216		
Stx_1	1	176	No	800
Stx_2	1	250	No	800
Eae	1	380	No	800
HlyA	1	525	No	800

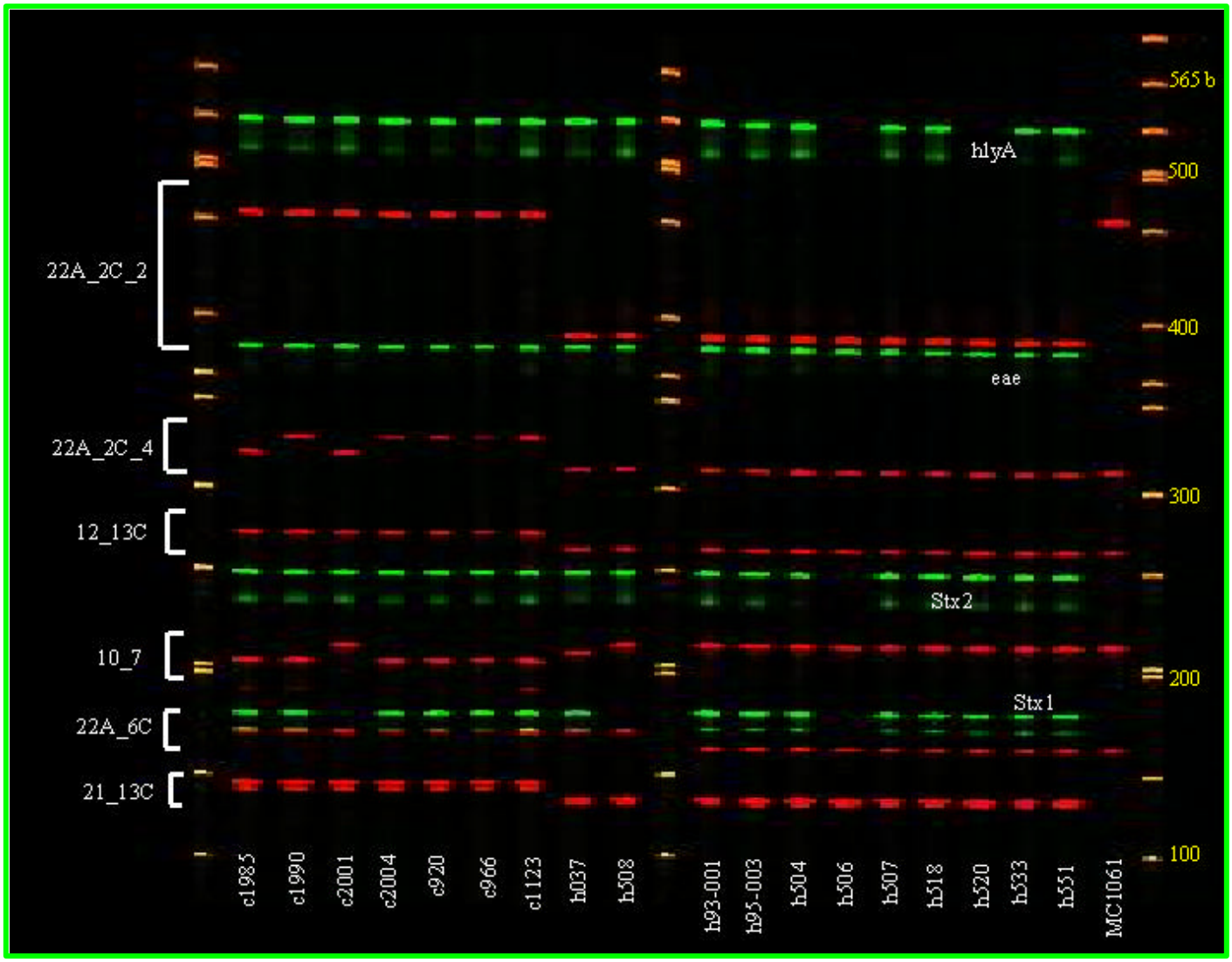

Fig. 1. Lineage-Specific Marker Assay. Shown is a colored image from electrophoresis of LMSA reactions products resulting from lineage I and II strains. Polymorphic LMSA markers are colored red and the range of allele sizes for each marker is indicated to the left of the image in white. The O157:H7-specific markers are colored green and each marker is labeled on the image (Stx1, Stx2, eae, and hlyA). Strain designations are indicated below each lane. Te three unlabeled lanes contain molecular size standards with the sizes indicated to the right of the image.

Table 3. Distribution of LMSA genotypes

Genotype	Number
111111	81
112111	5
211111	54
212111	28
221111	1
221213	2
221222	2
222111	5
222212	2
222213	2
222222	25
222223	2
222312	1
222322	1
231233	1
232221	1
232222	1
232233	1
232312	1
TOTAL	216