The Journal of Veterinary Medical Science

FULL PAPER

Public Health

Clarification of relationship between single-nucleotide polymorphism panels of Shiga toxin-producing *Escherichia coli* O157:H7/H- strains

Shinichiro HIRAI[1)]*, Eiji YOKOYAMA[1,2)], Yuh SHIWA[3,4)], Taichiro ISHIGE[3)], Naoshi ANDO[2)], Takeshi SHIMIZU[5)], Satoshi MURAKAMI[6)]

[1)]Center for Emergency Preparedness and Response, National Institute of Infectious Diseases, Tokyo, Japan
[2)]Division of Bacteriology, Chiba Prefectural Institute of Public Health, Chiba, Japan
[3)]NODAI Genome Research Center, Tokyo University of Agriculture, Tokyo, Japan
[4)]Department of Molecular Microbiology, Faculty of Life Sciences, Tokyo University of Agriculture, Tokyo, Japan
[5)]Department of Molecular Infectiology, Graduate School of Medicine, Chiba University, Chiba, Japan
[6)]Laboratory of Animal Health, Department of Animal Science, Tokyo University of Agriculture, Kanagawa, Japan

ABSTRACT. Eighty strains of enterohemorrhagic *Escherichia coli* O157:H7/H- were analyzed by three single-nucleotide polymorphism (SNP) panels using whole-genome sequencing data. The partial concordance of SNP types among the different SNP panels was observed on minimum spanning trees reconstructed with SNP data. As for lineage I/II strains, some of the clade 7 strains belonged to one unique SNP type as determined by three panels, suggesting that clade 7 should be divided into at least two genotypes, namely, the unique type and the rest. In addition, clade 8 contained two unique genotypes, which was consistent with the previous prediction. Similarly, for lineage II, clade 12 should be divided into three genotype strains. In contrast, many strains of several clades belonging to lineage I were clustered into the same node on each minimum spanning tree upon testing with the three SNP panels. Previous studies reported that lineage I diverged more recently than lineages I/II and II. Such low diversity in lineage I in this study may have arisen because this lineage has not accumulated SNPs because of its relatively recent divergence. Based on the concordance observed in this study, some of the previously published O157 genotype distribution data were successfully interpreted to clarify the clade distribution, which was well supported by previous literature.

KEYWORDS: Shiga toxin-producing *Escherichia coli* O157, single-nucleotide polymorphism (SNP) panel, whole-genome sequence (WGS)

J. Vet. Med. Sci.
84(10): 1399–1405, 2022
doi: 10.1292/jvms.22-0242

Received: 16 May 2022
Accepted: 7 August 2022
Advanced Epub:
20 August 2022

Shiga toxin-producing *Escherichia coli* O157:H7/H- (O157) is a significant foodborne pathogen in the public health field. O157 causes a wide variety of symptoms, from asymptomatic carriers to severe symptoms such as hemolytic uremic syndrome (HUS) and encephalopathy. It has been reported that the severity of the illness caused by O157 depends on the ability of the organism to produce Shiga toxin (Stx) 2 [2, 12].

Manning *et al*. [10] proposed a system for grouping O157 strains based on a panel of single-nucleotide polymorphisms (SNPs) that classified the organism into nine SNP genotypes, designated as clades. Among these O157 clades, clade 8 strains were thought to have strong pathogenicity because the incidence of HUS caused by them was significantly higher than that by strains in other clades. In addition, stronger pathogenicity of clade 6 strains was reported on the same basis [6]. In contrast, Etoh *et al*. [2] demonstrated weaker pathogenicity of clade 7 strains when the pathogenicity was compared based on the incidence of hemorrhagic colitis.

These findings suggest the possibility that O157 differently impacts public health in different areas if the dominant clade of the organism differs geographically. This hypothesis has been supported by the high incidence of HUS in Argentina, where clade 8 strains are prevalent, compared with the situation in Australia where other clade strains dominate [11]. Therefore, the investigation of O157

*Correspondence to: Hirai S: hirai@niid.go.jp, Center for Emergency Preparedness and Response, National Institute of Infectious Diseases, 4-7-1 Gakuen, Musashi-Murayama, Tokyo 208-0011, Japan
(Supplementary material: refer to PMC https://www.ncbi.nlm.nih.gov/pmc/journals/2350/)

clades should provide important information for public health. However, the SNP panel defined by Manning *et al*. [10] has been in need of revision. Our previous study indicated that clade 7 includes two lineages, which are phylogenetic groups, and those lineages were redefined as separate clades, namely, clades 7 and 12 (Supplementary Table 1) [3, 15]. However, our subsequent studies [4, 5] showed that clade 7 could be further divided into several clades. Clade 8 has also been reported to be divided into two groups [4, 5, 12].

The discriminatory power of molecular epidemiological methods can be used to further revise Manning's revised panel [3, 10]. Our previous studies [4, 5] showed that, using data from molecular epidemiological analysis, the classification of clades could be revised or clades could be subdivided. Some SNP panels other than Manning's revised panel [3, 10], which consist of a small number of SNPs detected in the whole genomes of O157, have been reported to be sufficient to differentiate the organism into separate genotypes [1, 8, 13]. Therefore, Manning's revised panel [3, 10] could be further revised by comparison with these SNP panels. Moreover, in recent years, O157 strain differentiation using whole-genome sequencing (WGS) data has been increasingly reported to be useful for molecular epidemiology [16]. However, even if O157 strains are differentiated by three SNP panels [1, 8, 13] using WGS data, the pathogenicity of these SNP types cannot be evaluated. When the relationship between the three SNP panels [1, 8, 13] and clades, whose pathogenicity has been evaluated, can be revealed, these panels should be able to provide data on the SNP types that are valuable for public health.

The purpose of this study is to reveal the relationship between SNP genotypes determined by three different SNP panels [1, 8, 13] and Manning's revised panel [3, 10]. SNP genotypes defined by these four SNP panels [1, 3, 8, 13] were compared using WGS data of 80 O157 genomes. Based on the results of comparison, we discussed the possibility of further revising the clade classification. Using previous reports, we also attempted to interpret the geographical distribution of SNP genotypes as determined by the SNP panels other than Manning's revised panel [3, 10].

MATERIALS AND METHODS

Selection of O157 strains

The strains tested in this study were selected as follows. A total of 1069 strains of O157 isolated in 1998–2013 in Chiba Prefecture, Japan, were investigated by SNP analysis as previously reported [4], and classified into lineages by lineage-specific polymorphism assay-6 [14], as also previously reported [15]. These strains were classified into clades based on data of SNP analysis and lineage-specific polymorphism assay-6, as described previously [3, 10]. The insertion sequence (IS) *629* distribution in these strains was investigated using IS-printing (Toyobo, Osaka, Japan), and the IS*629* distribution data were used to reconstruct a minimum spanning tree (MST) of the strains in each clade. A total of 78 non-sorbitol-fermenting strains that had no epidemiological links were selected without any intention for further analysis (Supplementary Figs. 1–8). Of these 78 strains, one clade 2 strain (CEC04072), one clade 6 strain (CEC09072), and one clade 12 strain (CEC04150) were isolated from cattle, while the other strains were isolated from humans. In addition, two strains (US2 and 86-24) isolated in the U.S. were added to the 78 strains, for a total of 80 O157 strains in this study (Table 1).

WGS analysis

WGS analysis of the selected O157 strains was performed using a next-generation sequencer as in our previous study [16]. The DNA of each strain was extracted using a NucleoBond Buffer Set III (TaKaRa, Kyoto, Japan) and NucleoBond AXG20 column (TaKaRa). Sequence libraries were prepared with a Nextera XT DNA Sample Prep Kit (Illumina, San Diego, CA, USA), and 100 cycles of dual index paired-end sequencing were carried out using the Illumina HiSeq2500 system (Illumina). The Illumina analysis pipeline (CASAVA 1.6.0) was used for image analysis, base calling, and quality score calibration. Reads were sorted by barcode and exported in FASTQ files.

Raw read data were deposited in the Sequence Read Archive (SRA) in the DNA Data Bank of Japan (DDBJ, Accession No. PRJDB4016). The FASTQ files were analyzed using CLC Genomics Workbench software, version 20 (CLC Bio, Aarhus, Denmark). Read data were mapped to a reference genome (*E. coli* O157:H7 strain Sakai, GenBank Accession No. NC_002695) with the "Non-specific match was ignored" option. SNPs were detected with a fixed ploidy variant detection method with the "Coverage and count filter with minimum coverage of 1" option to exclude ambiguous SNPs due to sequence reading errors.

Comparison of SNP panels

The SNP panels reported by Jung *et al*. (the Jung panel) [8], Strachan *et al*. (the Strachan panel) [13], and Clawson *et al*. (the Clawson panel) [1] were analyzed in this study (Supplementary Tables 2–4). Each SNP site in the three SNP panels was collected from the WGS data of this study using in-house Ruby scripts (ISO/IEC 30170). The collected SNPs were analyzed by MST using PopArt version 1.7 [9]. If a tested strain was clustered into the same node of a certain genotype strain by an SNP panel, the genotype of the tested strain was determined as the certain genotype. When a tested strain was clustered into a node that was separate from a node of a certain genotype by one SNP, the tested strain was designated as a single-locus variant (SLV) of the certain genotype.

RESULTS

Among the tested strains, lineage I strains tended to cluster into one large node on MSTs reconstructed by all of the tested SNP panels, although lineage I strains contained 4 different clades, namely, clades 1, 2, 3, and descendant 4/5 (Figs. 1–3). By the Strachan panel, 7 out of 8 clade descendant 4/5 strains were separately clustered into a different node from the large node where clade 1, 2, and 3 strains were clustered (Fig. 2). As for clade 2 strains, 9 of 11 strains showed the {Ib:Group Gvi:gen28} genotype [meaning {the

Table 1. Eighty strains of enterohemorrhagic *Escherichia coli* O157:H7/H- tested in this study

Lineage [a] (Consensus)	LSPA-6 code	Strain No.	Year of isolation	Genotypes determined by the following single-nucleotide polymorphism panel:			
				Manning's panel [b]	Jung's panel [c]	Strachan's panel [d]	Clawson's panel [e]
I	111111	CEC03058	2003	Clade 1	Ib	Group Gvi	gen31
(1N111N)	111111	CEC03063	2003	Clade 1	Ib	Group Gvi	gen31
	111111	CEC05072	2005	Clade 1	Ib	Group Gvi	gen28
	111111	CEC04009	2004	Clade 2	Ib	Group Gvi	gen28
	111111	CEC04072	2004	Clade 2	Ib	Group Gvi	gen28
	111111	CEC04177	2004	Clade 2	Ib	Group Gvi	gen28
	111111	CEC04184	2004	Clade 2	Ib	Group Gvi	gen28
	111111	CEC06064	2006	Clade 2	Ib	Group Gvi	gen28
	111111	CEC11100	2011	Clade 2	Ib	Group Gvi	gen28
	111111	CEC11143	2011	Clade 2	Ib	Group Gvi	gen31
	111111	CEC12014	2012	Clade 2	Ib	Group Gvi	gen28
	111111	CEC12025	2012	Clade 2	Ib	Group Gvi	gen28
	111111	CEC13068	2013	Clade 2	Ib	Group Gvi	gen28 or 32 SLV [f]
	111111	US2	1982	Clade 2	Ib	Group Gvi	gen28
	111111	CEC01398	2001	Clade 3	Ib	Group Gvi	gen29
	111111	CEC03095	2003	Clade 3	Ib	Group Gvi	gen28
	111111	CEC07016	2007	Clade 3	Ib	Group Gvi	gen29
	111111	CEC08110	2008	Clade 3	Ib	Group Gvi	gen28
	111111	CEC08112	2008	Clade 3	Ib	Group Gvi	gen28
	111111	CEC10001	2010	Clade 3	Ib	Group Gvi	gen28
	111111	CEC10007	2010	Clade 3	Ib	Group Gvi	gen28
	111111	CEC11077	2011	Clade 3	Ib	Group Gvi	gen29
	111111	CEC11153	2011	Clade 3	Ib	Group Gvi	gen29
	111111	CEC98075	1998	Clade 3	Ib	Group Gvi	gen28
	111111	86-24	1986	Clade descendant 4/5	Ib	Group Gv	gen28
	111111	CEC02107	2002	Clade descendant 4/5	Ib	Group Gv	gen28
	111111	CEC03106	2003	Clade descendant 4/5	Ib	Group Gv	gen28
	111111	CEC05085	2005	Clade descendant 4/5	Ib	Group Gv	gen28
	111111	CEC06041	2006	Clade descendant 4/5	Ib	Group Gvi	gen28
	111111	CEC07021	2007	Clade descendant 4/5	Ib	Group Gv	gen28
	111111	CEC08010	2008	Clade descendant 4/5	Ib	Group Gv	gen28
	111111	CEC08124	2008	Clade descendant 4/5	Ib	Group Gv	gen28
I/II	211111	CEC03027	2003	Clade ancestral 4/5	Ia	Group Gi or Giv [g]	gen15
(2N111N)	211111	CEC06082	2006	Clade ancestral 4/5	Ia	Group Giii	gen14
	211111	CEC11104	2011	Clade ancestral 4/5	IIb	Group Gi or Giv	gen19
	211111	CEC99085	1999	Clade ancestral 4/5	Ia	Group Gi or Giv	gen24
	211111	CEC09072	2009	Clade 6	Ia	Group Gi or Giv	gen24
	211111	CEC10106	2010	Clade 6	Ia	Group Gi or Giv	gen23
	211111	CEC99038	1999	Clade 6	Ia	Group Gi or Giv	gen24
	211111	CEC06016	2006	Clade 7	Ia	Group Gi or Giv	gen12, 22, 26 or 33 SLV
	211111	CEC06089	2006	Clade 7	Ia	Group Gi or Giv	gen14
	211111	CEC12094	2012	Clade 7	IVa	Group C	gen1 SLV
	211111	CEC13043	2013	Clade 7	Ia	Group Gi or Giv	gen12, 22, 26 or 33 SLV
	211111	CEC13046	2013	Clade 7	Ia	Group Gi or Giv	gen12, 22, 26 or 33 SLV
	211111	CEC13053	2013	Clade 7	Ia	Group Gi or Giv	gen12, 22, 26 or 33 SLV
	211111	CEC13069	2013	Clade 7	Ia	Group Gi or Giv	gen12, 22, 26 or 33 SLV
	221111	CEC01441	2001	Clade 7	IVb	Group D	gen33
	221111	CEC04007	2004	Clade 7	Ia	Group D	gen33
	221111	CEC04146	2004	Clade 7	Unable to type	Unable to type	gen14
	221111	CEC05092	2005	Clade 7	IVc	Group D	gen33
	221111	CEC99094	1999	Clade 7	Ia	Group Gi or Giv	gen12, 22, 26 or 33 SLV
	211111	CEC08139	2008	Clade 8	IIa	Group Fii	gen21
	231111	CEC08068	2008	Clade 8	IIb	Group Fi	gen19
	231111	CEC08090	2008	Clade 8	IIa	Group Fii	gen21
	231111	CEC08097	2008	Clade 8	IIb	Group Fi	gen19
	231111	CEC08114	2008	Clade 8	IIb	Group Fi	gen19
	231111	CEC08141	2008	Clade 8	IIb	Group Fi	gen19
	231111	CEC08142	2008	Clade 8	IIb	Group Fi	gen19
	231111	CEC08152	2008	Clade 8	IIa	Group Fii	gen21
	241111	CEC09077	2009	Clade 9	VII	Group A SLV	gen35

Table 1. Continued

Lineage [a] (Consensus)	LSPA-6 code	Strain No.	Year of isolation	Genotypes determined by the following single-nucleotide polymorphism panel:			
				Manning's panel [b]	Jung's panel [c]	Strachan's panel [d]	Clawson's panel [e]
II	222113	CEC00062	2000	Clade 12	Vb SLV	Group Ei	gen1
(Others)	212122	CEC03077	2003	Clade 12	Vb SLV	Group Ei	gen38
	242123	CEC03109	2003	Clade 12	Va	Group Eii	gen1
	222123	CEC04039	2004	Clade 12	Vb	Group Ei	gen9
	242222	CEC04150	2004	Clade 12	Vb	Group Ei	gen9
	222122	CEC04155	2004	Clade 12	VB	Group Ei	gen38
	211123	CEC04169	2004	Clade 12	Vb SLV	Group Ei	gen41
	222112	CEC05051	2005	Clade 12	Va	Group Eii	gen1
	222224	CEC05146	2005	Clade 12	Vb	Group Ei	gen8
	212114	CEC07107	2007	Clade 12	VI	Group B	gen36
	221223	CEC07142	2007	Clade 12	Vb	Group Ei	gen9
	252123	CEC08031	2008	Clade 12	Va	Group Eii	gen1
	221123	CEC08129	2008	Clade 12	Vb SLV	Group Ei or Eii SLV	gen1
	222222	CEC08162	2008	Clade 12	Vb	Group Ei	gen6
	22?222	CEC09011	2009	Clade 12	Vb	Group Ei	gen7
	212111	CEC09017	2009	Clade 12	VI	Group B	gen36
	221123	CEC09080	2009	Clade 12	Va	Group Eii	gen1
	221222	CEC98057	1998	Clade 12	Vb	Group Ei	gen7
	222123	CEC99030	1999	Clade 12	Vb	Group Ei	gen6
	242222	CEC99053	1999	Clade 12	Vb	Group Ei	gen6

a) Lineages were defined by the description of Yokoyama *et al*. (2012) [15]. b) Genotyping was carried out as described by Manning *et al.* (2008) [10] and by Hirai *et al*. (2013) [3]. c) Genotyping was carried out as described by Jung *et al*. (2013) [8]. d) Genotyping was carried out as described by Strachan *et al*. (2015) [13]. e) Genotyping was carried out as described by Clawson *et al*. (2008) [1]. f) "SLV" denotes single-locus variant. g) Strachan's single-nucleotide polymorphism panel cannot differentiate Group Gi from Group Giv.

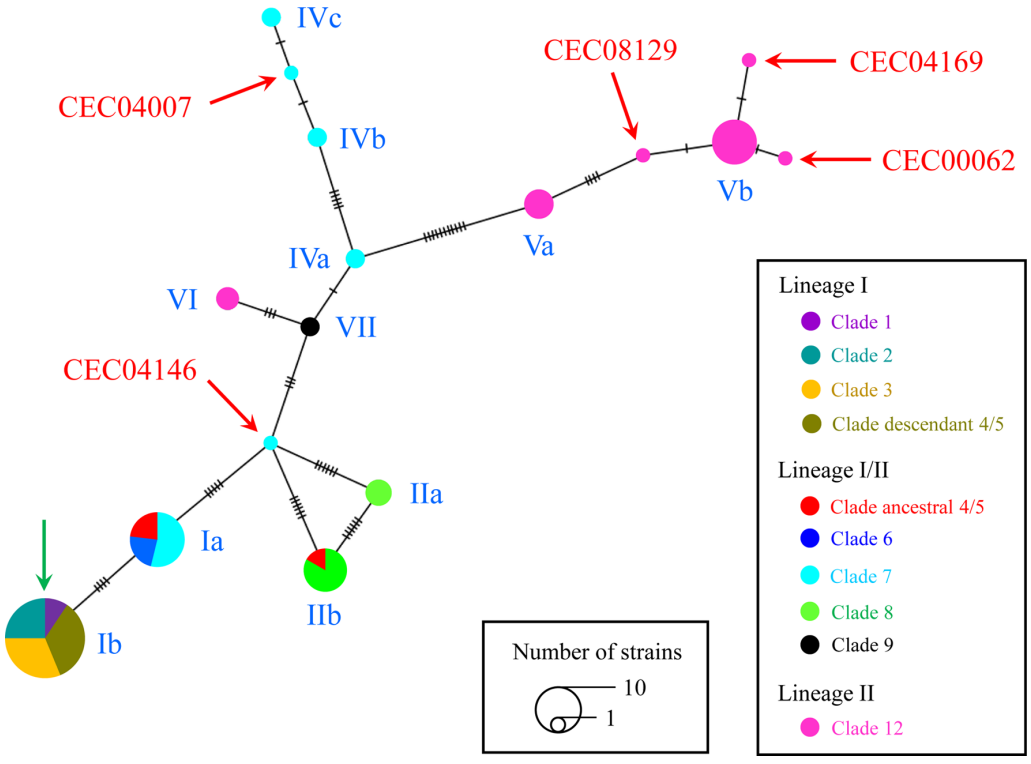

Fig. 1. A minimum spanning tree reconstructed using single-nucleotide polymorphisms (SNPs) following the work of Jung *et al*. [8]. A hatched bar on a line connecting two nodes indicates one SNP. A name in blue denotes a genotype determined by the SNPs. A strain in red denotes a strain that was not clustered with any genotype strains. The green arrow indicates a node where many strains of several clades belonging to lineage I were clustered.

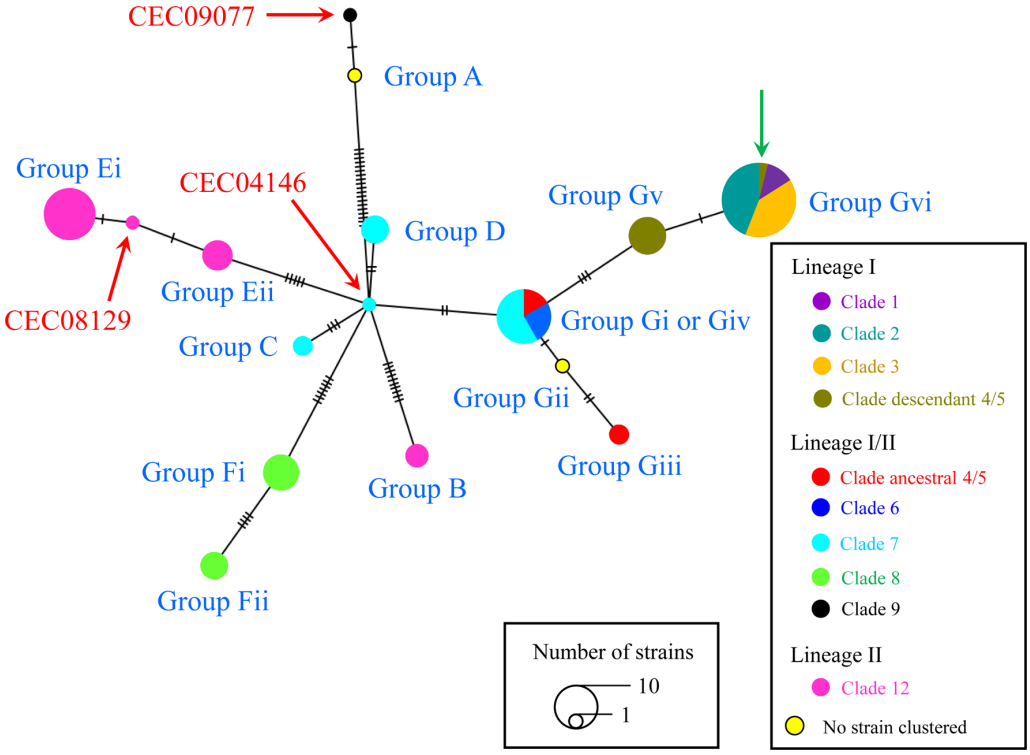

Fig. 2. A minimum spanning tree reconstructed using single-nucleotide polymorphisms (SNPs) following the work of Strachan *et al*. [13]. A hatched bar on a line connecting two nodes indicates one SNP. A name in blue denotes a genotype determined by the SNPs. A strain in red denotes a strain that was not clustered with any genotype strains. The green arrow indicates a node where many strains of several clades belonging to lineage I were clustered.

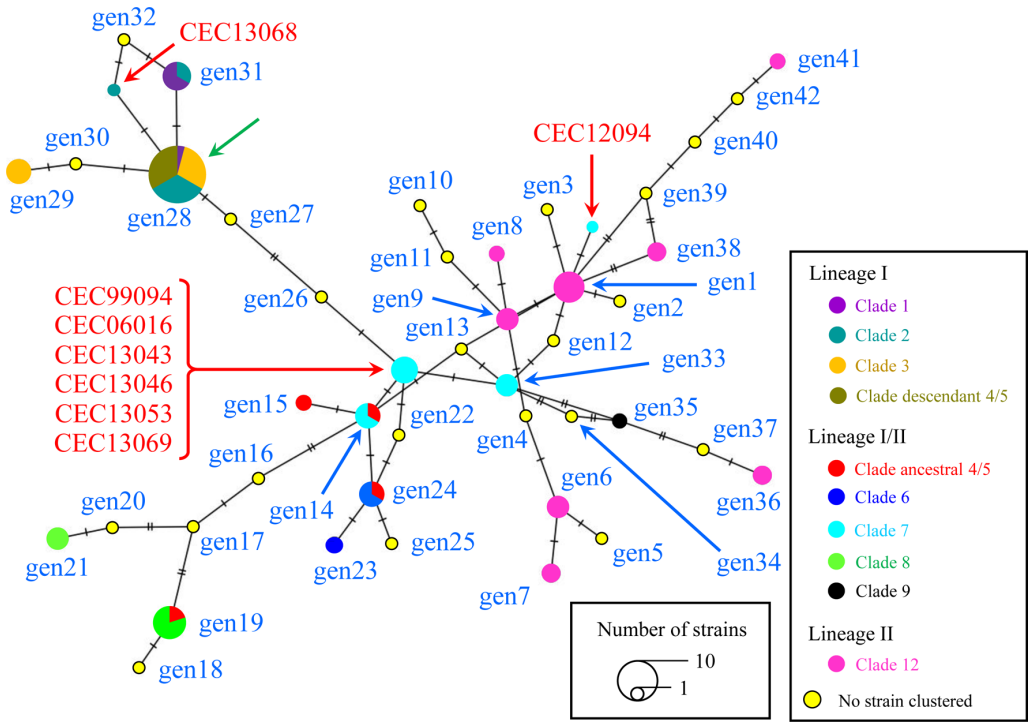

Fig. 3. A minimum spanning tree reconstructed using single-nucleotide polymorphisms (SNPs) following the work of Clawson *et al*. [1]. A hatched bar on a line connecting two nodes indicates one SNP. A name in blue denotes a genotype determined by the SNPs. A strain in red denotes a strain that was not clustered with any genotype strains. The green arrow indicates a node where many strains of several clades belonging to lineage I were clustered.

genotype by the Jung panel:that by the Strachan panel:that by the Clawson panel}]. However, 6 out of 10 strains of clade 3 were also divided into the same genotype (Table 1).

Among lineage I/II strains, clade 8 strains were separately clustered into two nodes on all MSTs (Figs. 1–3). In addition, the separation of clade 8 strains was exactly the same, and they had {IIa:Group Fii:gen21} and {IIb:Group Fi:gen19} genotypes (Table 1). As for clade 7 strains, three strains that were clustered into the group D node by the Strachan panel were clustered into the gen33 node by the Clawson panel, and no other strains were clustered into the gen33 node, although their genotypes by the Jung panel were different (Table 1). The remaining 6 out of 9 strains of clade 7 showed the same genotype, {Ia:Group Gi or Giv:gen12, 22, 26, or 33 SLV}, and this pattern was not shown in any other tested strains (Table 1).

Among lineage II strains, two strains of clade 12 had the {VI:Group B:gen36} genotype and no other tested strains showed this pattern. Among the remaining clade 12 strains, four had the {Va:Group Eii:gen1} genotype, while this pattern was not shown in any other tested strains.

DISCUSSION

The main purpose of this study was to reveal the relationship between SNP genotypes determined by three different SNP panels [1, 8, 13] and Manning's revised panel [3, 10]. We used three SNP panels that include a total of 140 SNPs. Among them, only one position (822,741) was shared, namely, between Jung's and Strachan's panels, which means that we actually used SNP panels with 139 SNP sites that were evaluated as effective for differentiating O157 strains [1, 8, 13]. Analysis of WGS data of the O157 genome using the large-scale SNP panels indicated the partial concordance of SNP types among the different SNP panels. Furthermore, in some clades, some of the O157 strains in the same clade have unique genotypes. These findings suggested that the 139 SNPs are sufficient to identify strains with a unique genotype in a clade.

Strains of clade 7, which was considered as the most dominant clade in Japan [15], should be divided into at least two genotype strains: {Ia:Group Gi or Giv:gen12, 22, 26, or 33 SLV} and others. This is consistent with our previous description [3] of there being some groups of clade 7 on an MST reconstructed using the IS*629* distribution. We also predicted the divergence of clade 7 because of the variety of *stx* genes possessed by strains of this clade [3]. Information on the dominant clade is important for local public health because of the relationship between the dominant clade of O157 and the severity of O157 infection in an area [11]. Therefore, further studies are necessary to demonstrate the division of clade 7 strains.

Similarly, clade 12 strains should be divided into at least three genotype strains: {VI:Group B:gen36}, {Va:Group Eii:gen1}, and others. This is consistent with our previous description [4], which showed some groups of clade 12 on an MST based on the IS*629* distribution. Previously, clade 12 was included in clade 7; however, clade 12 was newly differentiated from clade 7 based on the difference of lineage types [15]. Clade 12 was reported to be less pathogenic because of the possession of *stx2c* [2], leading to little public health interest and investigation. The fact that there are several populations in clade 12, as demonstrated by this study, may be due to the fact that adequate research on this clade has not been conducted to date.

Our results also corroborated the previous finding that clade 8 should be divided into two groups [4, 5, 12]. Clade 8 has been thought to have an important genotype because of its high likelihood of causing HUS [5–7, 12]. However, because it is known that the clade is composed of two different subgroups in terms of pathogenicity [5], the pathogenicity should be evaluated again after dividing clade 8 into two groups.

In contrast, differentiation of lineage I strains was obscure in this study. The three SNP panels investigated in this study did not show sufficient ability to differentiate clades of lineage I, such as clades 1, 2, 3, and descendant 4/5, and most of those strains were clustered into a large node on an MST reconstructed by the SNP panels. Lineage I is thought to be most recently diverged from lineage I/II strains [2, 3, 15], which may result in fewer SNPs that are effective for differentiating lineage I strains.

Interestingly, the results of our study indicate the possibility of interpreting the results of pathogenicity and the distribution of O157 in previous reports. Jung *et al*. [8] suggested that the Ia genotype was prevalent in Australia in their previous study, and this study showed that the Ia genotype contained lineage I/II clades, which is in agreement with the report by Mellor *et al*. [11]. Jung *et al*. [8] also reported that the Ib genotype was prevalent in Canada, suggesting that lineage I was prevalent there. In New Zealand, IIb and IVb genotypes were reported to be prevalent [8], and this study showed that the IIb genotype contained clade 8 strains and lineage IVb genotype contained clade 7 strains, suggesting that clades 8 and 7 were prevalent in New Zealand. This suggestion is concordant with other reports [7] showing that lineage I/II strains are dominant in New Zealand and that clades 7 and 8 are in lineage I/II.

Conversely, there was a discrepancy in the determination of the geographic distribution of O157 clades as follows. Jung *et al*. [8] reported that the Vb strains were prevalent in Japan, and found that the Vb genotype contained clade 12 strains. However, the report that clade 12 strains were most prevalent in Japan is contradicted by previous reports that lineage I strains were most prevalent in Japan [2, 15]. This difference may be due to sampling bias, as discussed by Jung *et al*. [8].

One limitation of this study may be derived from the use of short-read WGS analysis. In general, short-read WGS analysis cannot detect SNPs in paralogous genes because short-read WGS data cannot be precisely mapped. In this study, we did not use a read that can be mapped to multiple sites in a genome. If an SNP exists at such a site, it would be detected as a gap in the site. However, no gaps were observed in the results of this study (data not shown); therefore, the limitation mentioned above would not have had any impact on this study. Another limitation is the occurrence of ambiguous SNPs caused by incorrect polymerization. In this study, we used SNPs with low coverage because read data may not be sufficiently mapped for unknown reasons. If an ambiguous SNP occurs, it may cause an SLV of a certain genotype. The occurrence of an SLV may markedly affect the results obtained by the Clawson panel because many genotypes of the panel have only one specific SNP to differentiate from other genotypes. In contrast, the results

obtained by the Jung and Strachan panels may not be affected at all because these panels had no such genotype, with the exception of Groups Gv and Gvi.

In conclusion, this study revealed the relationship between SNP genotypes determined by three different SNP panels and Manning's revised panel. Partial concordance of several genotypes among these SNP panels was observed, suggesting that some clades can be further divided into subgroups. Clarification of the relationship made it possible to evaluate the pathogenicity of O157 strains via the SNP genotype determined by panels other than Manning's revised panel. It also enables interpretation of the clade distribution of O157 strains using data on the distribution of SNP genotypes determined using these panels in each region of the world.

CONFLICT OF INTEREST. The authors have no conflicts of interest to declare.

ACKNOWLEDGMENTS. Part of this study was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology (S0801025), Cooperative Research Grant of Genome Research for BioResource, NODAI Genome Research Center, Tokyo University of Agriculture in fiscal year 2014, and JSPS KAKENHI (grant number 26460540).

REFERENCES

1. Clawson ML, Keen JE, Smith TPL, Durso LM, McDaneld TG, Mandrell RE, Davis MA, Bono JL. 2009. Phylogenetic classification of *Escherichia coli* O157:H7 strains of human and bovine origin using a novel set of nucleotide polymorphisms. *Genome Biol* **10**: R56. [Medline] [CrossRef]
2. Etoh Y, Hirai S, Ichihara S, Maeda E, Yokoyama E, Sera N, Horikawa K, Yamamoto T. 2014. Evolutionary model of the divergence of enterohemorrhagic Escherichia coli O157 lineage I/II clades reconstructed from high resolution melting and Shiga-like toxin 2 analyses. *Infect Genet Evol* **24**: 140–145. [Medline] [CrossRef]
3. Hirai S, Yokoyama E, Yamamoto T. 2013. Linkage disequilibrium of the IS*629* insertion among different clades of enterohemorrhagic *Escherichia coli* O157:H7/H-strains. *Infect Genet Evol* **18**: 94–99. [Medline] [CrossRef]
4. Hirai S, Yokoyama E, Etoh Y, Seto J, Ichihara S, Suzuki Y, Maeda E, Sera N, Horikawa K, Sato S, Yamamoto T. 2015. Putative classification of clades of enterohemorrhagic *Escherichia coli* O157 using an IS-printing system. *Lett Appl Microbiol* **61**: 267–273. [Medline] [CrossRef]
5. Hirai S, Yokoyama E, Wakui T, Ishige T, Nakamura M. 2018. Enterohemorrhagic *Escherichia coli* O157 subclade 8b strains in Chiba Prefecture, Japan, produced larger amounts of Shiga toxin 2 than strains in subclade 8a and other clades. *PLoS One* **13**: e0191834. [Medline] [CrossRef]
6. Iyoda S, Manning SD, Seto K, Kimata K, Isobe J, Etoh Y, Ichihara S, Migita Y, Ogata K, Honda M, Kubota T, Kawano K, Matsumoto K, Kudaka J, Asai N, Yabata J, Tominaga K, Terajima J, Morita-Ishihara T, Izumiya H, Ogura Y, Saitoh T, Iguchi A, Kobayashi H, Hara-Kudo Y, Ohnishi M, Arai R, Kawase M, Asano Y, Asoshima N, Chiba K, Furukawa I, Kuroki T, Hamada M, Harada S, Hatakeyama T, Hirochi T, Sakamoto Y, Hiroi M, Takashi K, Horikawa K, Iwabuchi K, Kameyama M, Kasahara H, Kawanishi S, Kikuchi K, Ueno H, Kitahashi T, Kojima Y, Konishi N, Obata H, Kai A, Kono T, Kurazono T, Matsumoto M, Matsumoto Y, Nagai Y, Naitoh H, Nakajima H, Nakamura H, Nakane K, Nishi K, Saitoh E, Satoh H, Takamura M, Shiraki Y, Tanabe J, Tanaka K, Tokoi Y, Yatsuyanagi J. 2014. 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 Infect Dis* **1**: ofu061. [Medline] [CrossRef]
7. Jaros P, Cookson AL, Campbell DM, Besser TE, Shringi S, Mackereth GF, Lim E, Lopez L, Dufour M, Marshall JC, Baker MG, Hathaway S, Prattley DJ, French NP. 2013. A prospective case-control and molecular epidemiological study of human cases of Shiga toxin-producing *Escherichia coli* in New Zealand. *BMC Infect Dis* **13**: 450. [Medline] [CrossRef]
8. Jung WK, Bono JL, Clawson ML, Leopold SR, Shringi S, Besser TE. 2013. Lineage and genogroup-defining single nucleotide polymorphisms of *Escherichia coli* O157:H7. *Appl Environ Microbiol* **79**: 7036–7041. [Medline] [CrossRef]
9. Leigh JW, Bryant D. 2015. PopART: full-feature software for haplotype network construction. *Methods Ecol Evol* **6**: 1110–1116. [CrossRef]
10. Manning SD, Motiwala AS, Springman AC, Qi W, Lacher DW, Ouellette LM, Mladonicky JM, Somsel P, Rudrik JT, Dietrich SE, Zhang W, Swaminathan B, Alland D, Whittam TS. 2008. Variation in virulence among clades of Escherichia coli O157:H7 associated with disease outbreaks. *Proc Natl Acad Sci USA* **105**: 4868–4873. [Medline] [CrossRef]
11. Mellor GE, Sim EM, Barlow RS, D'Astek BA, Galli L, Chinen I, Rivas M, Gobius KS. 2012. Phylogenetically related Argentinean and Australian *Escherichia coli* O157 isolates are distinguished by virulence clades and alternative Shiga toxin 1 and 2 prophages. *Appl Environ Microbiol* **78**: 4724–4731. [Medline] [CrossRef]
12. Ogura Y, Mondal SI, Islam MR, Mako T, Arisawa K, Katsura K, Ooka T, Gotoh Y, Murase K, Ohnishi M, Hayashi T. 2015. The Shiga toxin 2 production level in enterohemorrhagic *Escherichia coli* O157:H7 is correlated with the subtypes of toxin-encoding phage. *Sci Rep* **5**: 16663. [Medline] [CrossRef]
13. Strachan NJC, Rotariu O, Lopes B, MacRae M, Fairley S, Laing C, Gannon V, Allison LJ, Hanson MF, Dallman T, Ashton P, Franz E, van Hoek AHAM, French NP, George T, Biggs PJ, Forbes KJ. 2015. Whole genome sequencing demonstrates that geographic variation of *Escherichia coli* O157 genotypes dominates host association. *Sci Rep* **5**: 14145. [Medline] [CrossRef]
14. Yang Z, Kovar J, Kim J, Nietfeldt J, Smith DR, Moxley RA, Olson ME, Fey PD, Benson AK. 2004. Identification of common subpopulations of non-sorbitol-fermenting, beta-glucuronidase-negative Escherichia coli O157:H7 from bovine production environments and human clinical samples. *Appl Environ Microbiol* **70**: 6846–6854. [Medline] [CrossRef]
15. Yokoyama E, Hirai S, Hashimoto R, Uchimura M. 2012. Clade analysis of enterohemorrhagic *Escherichia coli* serotype O157:H7/H- strains and hierarchy of their phylogenetic relationships. *Infect Genet Evol* **12**: 1724–1728. [Medline] [CrossRef]
16. Yokoyama E, Hirai S, Ishige T, Murakami S. 2018. Application of whole genome sequence data in analyzing the molecular epidemiology of Shiga toxin-producing *Escherichia coli* O157:H7/H. *Int J Food Microbiol* **264**: 39–45. [Medline] [CrossRef]