The Journal of Microbiology (2010) Vol. 48, No. 2, pp. 184-187
DOI 10.1007/s12275-010-9182-1

Sequence Analysis of the Gene Encoding H Antigen in *Escherichia coli* Isolated from Food in Morocco

Samira Badri[1]*, Aziz Fassouane[2], Ingrid Filliol[3], Mohammed Hassar[1], and Nozha Cohen[1]

[1]*Laboratoire de Microbiologie et d'Hygiène des Aliments et de l'Environnement, Institut Pasteur du Maroc, Casablanca 20100, Morocco*
[2]*Laboratoire de biochimie, Département de Biologie, Faculté des Sciences d'Eljadida, Université Chouaib, Doukkali 24000, Morocco*
[3]*Centre national de référence des Escherichia coli et Shigella, Unité de Biodiversité des Bactéries pathogènes émergentes, Institut Pasteur de Paris, Paris 75015, France*

(Received June 8, 2009 / Accepted October 14, 2009)

In order to develop other molecular method useful for typing of motile and non motile *Escherichia coli* strains, a total of 207 strains of *E. coli* (133 reference strains, 74 food strains) were characterized by analysis of sequences of their amplified flagellin-encoding (*fliC*) gene products. The collection of reference strains was used for database building of *fliC* gene sequences. Application of this identification system to 74 *E. coli* food isolates revealed a reproducible and clear cut classification with very good correlation to results obtained by *Hha*I restriction of the amplified flagellin gene. The proposed determination of *fliC* sequences variations should be helpful for epidemiological studies.

***Keywords*:** *E. coli*, H-antigen, *fliC* gene, sequence analysis, PCR-RFLP

Serotyping of *E. coli* O and H antigens has been shown to be suitable for identification of the major clonal types of pathogenic *E. coli* strains, such as O157 strains and others. Investigation of serotypes is important in investigations of outbreaks and for the spread and emergence of new *E. coli* types. The characterization of the O-antigen type only is not sufficient enough to identify potential pathogenic *E. coli* strains. Only some of the strains belonging to STEC O groups O26, O111, and O157 were shown to produce Shiga toxins (Leomil *et al.*, 2005; Guerra *et al.*, 2006; Kaufmann *et al.*, 2006). On the other hand, it was found that Shiga toxin production is closely associated with strains belonging to certain O:H serotypes, such as O26:H11, O111:H8, and O157:H7 (Whittam and Ake, 1993; Reid *et al.*, 2000; Zhang *et al.*, 2000).

However, several difficulties have been observed, in particular with H-serotyping of *E. coli*, when it was applied as a laboratory standard in routine praxis: 1) The expression of H-antigens can be sometimes delayed and dependent on various environmental signals (Kerridge, 1961; Adler and Templeton, 1967; Ratiner, 1999), 2) a great scale of cross-reactions happen to occur among *E. coli* which requires application of end-point dilution tests, and last but not least, 3) diagnostic H-sera are not commercially available and therefore not standardised. Therefore, several other approaches, in particular molecular approaches, were taken into consideration to characterise the O- and H-antigens of *E. coli* (Kilger and Grimont, 1993; Fields *et al.*, 1997; Dauga *et al.*, 1998; Reid *et al.*, 1999; Coimbra *et al.*, 2000; Machado *et al.*, 2000). One approach to subdivide *E. coli* was carried out on the basis of the flagella main protein gene *fliC* (Iino *et al.*, 1988; Joys, 1988; Kuwajima, 1988) in order to replace the time-consuming and non-standardised H-serotyping (Fields *et al.*, 1997; Reid *et al.*, 1999; Machado *et al.*, 2000).

In order to develop a specific and commonly applicable method for the typing of *E. coli* H-antigen, this study aimed first to analyze the complete nucleotide sequences of the *fliC* gene encoding flagellar antigen from *E. coli* reference strains for database building of *fliC* gene sequences, second to apply this identification system to 74 potentially pathogenic *E. coli* isolated from food in Morocco, and third to compare the results obtained by this method with those obtained by *fliC*-RFLP.

Materials and Methods

Bacterial strains

The *E. coli* reference strains used for database building of *fliC* gene sequences were from the collection described by (Machado *et al.*, 2000), which were characterized for their *fliC* types by PCR-RFLP typing. Seventy four *E. coli* isolates carrying virulence factors from the collection described by (Badri *et al.*, 2009) were tested to determine their H-types.

DNA preparation and amplification of *fliC* gene

Strains were streaked on tryptocasein soy agar (Sanofi Diagnostics Pasteur, Marnes-la-Coquette, France) and bacterial DNA was isolated using InstaGene Matrix kit (BioRad), following the instructions given by the manufacturer.

The *fliC* flagellin gene was amplified with the following two primers: 5′-CAA GTC ATT AAT AC (A/C) AAC AGC C-3′ (primer 1) and 5′-GAC AT(A/G) TT(A/G) GA(G/A/C) ACT TC(G/C) GT-3′ (primer 2). A/C, A/G, G/C, and G/A/C in parentheses indicate A or C, A or G, G or C, G or A or C (respectively). The synthesized product had to contain a mixture of oligonucleotides representing all these combinations (Dauga *et al.*, 1998).

* For correspondence. E-mail: samira_bdm@yahoo.fr; Tel: +212-68-30-67-18; Fax:+212-22-98-50-63

Restriction patterns and computer identification of isolates

PCR-generated DNA was purified using GFX PCR DNA and Gel Band Purification kit (Pharmacia). The purified PCR products *fliC* were cleaved with *Hha*I restriction endonuclease (Pharmacia) according to the manufacturer's instructions for 2 h at 37°C.

Using the Taxotron software, the restriction patterns of their amplified *fliC* gene were identified against databases containing the F types of reference strains (Coimbra *et al.*, 2000; Machado *et al.*, 2000) to conclude H types.

Nucleotide sequencing of *fliC* gene

The PCR products were purified with a QIAquick PCR purification kit (QIAGEN), used for sequencing by applying dye terminator chemistry PE Applied Biosystems, and separated on an automated DNA sequencer (Applied Biosystems). The sequences were analyzed with Bionumerics software.

Results

To investigate the ability of the *fliC* sequence analysis method to differentiate the *fliC* genes that encode the H antigens, the *fliC* gene of the reference strains has been sequenced and a database was built with the nucleotide sequences obtained. Application of this identification system to 74 *E. coli* food isolates carrying virulence factors were established. The results showed that prevalent H-types in *E. coli* strains tested were: H1 (5%), H2 (5%), H4 (23%), H5 (1%), H7 (5%), H8 (12%), H9 (5%), H10 (12%), H11 (3%), H18 (1%), H19 (2%), H21 (7%), H25 (1%), H26 (1%), H28 (1%), H29 (1%), H30 (1%), and H31 (8%). These strains were characterized by *Hha*I restriction of the amplified flagellin gene. Analysis of the restriction fragments were identified against databases containing the F types of reference strains (Coimbra *et al.*, 2000; Machado *et al*., 2000) to conclude H types (Fig. 1).

To correlate the F-types obtained by *fliC*-RFLP with H-types obtained by *fliC* sequence analysis, it was shown that the *fliC* genotypes of F-types derivatives of *E. coli* strains were similar to those obtained by *fliC* sequence analysis.

Discussion

In this work, we demonstrated that the *fliC* sequence analysis is a reliable, rapid, and easy-to-perform method for determination of the H types of *E. coli* isolates, including non motile strains that cannot be H typed by serotyping. Clearly identifiable and reproducible *fliC*-sequences were obtained for the *E. coli* strains isolated from food.

Moreover, serologically non-typeable H-antigens or H-antigen-negative (non-motile) strains could be classified into the H1 to H56 groups using the *fliC* sequences patterns. However, for international comparisons and application, *fliC* fingerprinting will require a standard operation protocol (e.g. as described in 'Materials and Methods'). Since the O- and H-antigens will remain an important biomarker for detection, population genetics, and epidemiological analysis of *E. coli*, their determination needs to be carried out by primary diagnostic laboratories and respective Reference Centres whether in its classical form or in its *fliC* fingerprinting version (Winstanley and Morgan, 1997). When new serogroups (O and H) will have to be established this will remain a task of the International Escherichia and Klebsiella Centre (WHO).

In addition to serotyping a range of other subtyping methods such as genotyping (molecular fingerprinting), electrotyping, and above all, virulence factor analysis has been applied (Swaminathan *et al.*, 2001). Earlier observations concerning particular pathogenic factors identified in particular serotypes (serovar) (Stenderup and Orskov, 1983) might have been misunderstood in the past as a serotype-related virulence with particular clinical implication. The definition of *E. coli* pathovars (such as EHEC, ETEC, EIEC, UPEC, etc.) according to their virulence make-up (Levine, 1987) does not imply principally a correlation between pathogenic factors and distinct serovars. This is not unexpected since most of the pathogenic factors are encoded by mobile genetic elements which allow the conversion of various serovars into the same pathovar (Brunder and Karch, 2000). However, the connection of distinct pathogenic factors with distinct serovars is indeed often observed among *E. coli* strains of clinical origin. Therefore, serotype grouping (O-and H-antigens) of pathogenic *E. coli* strains remains the first line of their investigation and is regarded as the essential first approach in subtyping pathogenic bacteria. Only when the serotype of clinical isolates is established the other molecular methods for subtyping and fingerprinting can be reasonably applied.

Acknowledgements

The authors are grateful to all collaborators from Centre National de Référence des *E. coli* et *Shigella*, Unite de Biodiversité des Bactéries Pathogènes Emergentes, Institut Pasteur Paris for their collaboration of the molecular analyses. We would like to thank the authorities of Casablanca for their help.

References

Adler, J. and B. Templeton. 1967. The effect of environmental conditions on the motility of *Escherichia coli*. *J. Gen. Microbiol.* 46, 175-184.

Badri, S., I. Filliol, I. Carle, M. Hassar, A. Fassouane, and N. Cohen. 2009. Prevalence of virulence genes in *Escherichia coli* isolated from food in Casablanca (Morocco). *Food Control.* 20, 560-564.

Brunder, W. and H. Karch. 2000. Genome plasticity in *Enterobacteriaceae*. *Int. J. Med. Microbiol.* 290, 153-165.

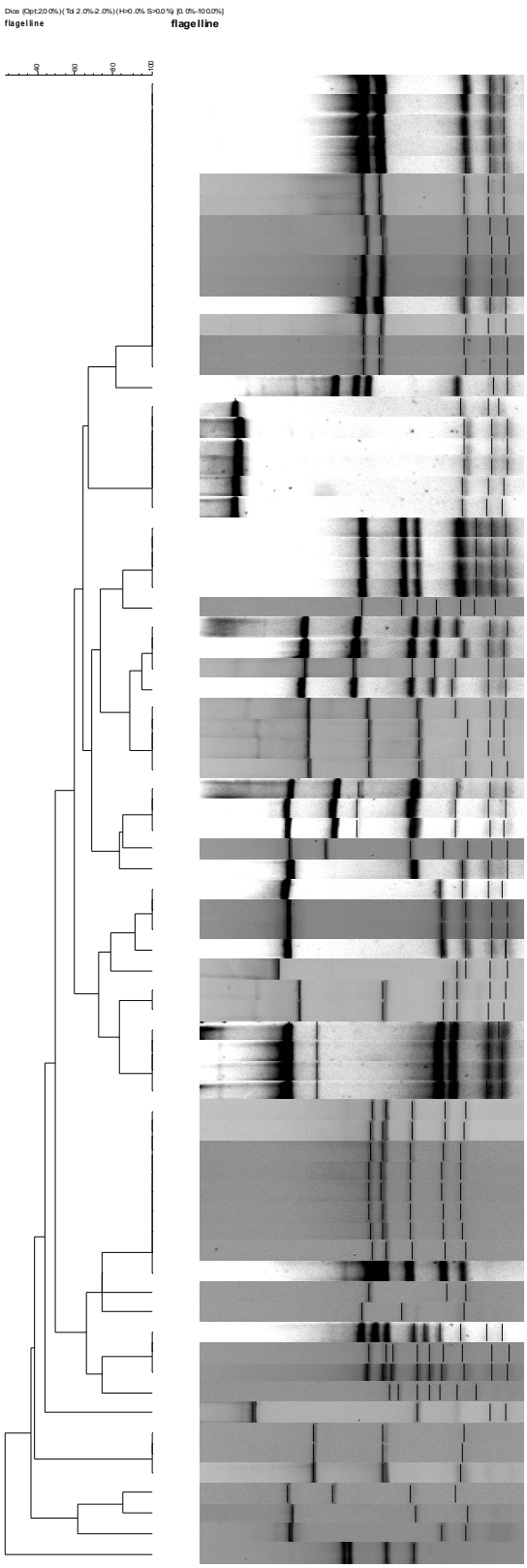
Coimbra, R.S., F. Grimont, P. Lenormand, P. Burguiere, L. Beutin, and P.A. Grimont. 2000. Identification of *Escherichia coli* O-serogroups by restriction of the amplified O-antigen gene cluster (*rfb*-RFLP). *Res. Microbiol.* 151, 639-654.

Dauga, C., A. Zabrovskaia, and P.A. Grimont. 1998. Restriction fragment length polymorphism analysis of some flagellin genes of *Salmonella enterica*. *J. Clin. Microbiol.* 36, 2835-2843.

Fields, P.I., K. Blom, H.J. Hughes, L.O. Helsel, P. Feng, and B. Swaminathan. 1997. Molecular characterization of the gene encoding H antigen in *Escherichia coli* and development of a PCR-restriction fragment length polymorphism test for identification of *E. coli* O157:H7 and O157:NM. *J. Clin. Microbiol.* 35, 1066-1070.

Guerra, B., E. Junker, A. Schroeter, R. Helmuth, B.E. Guth, and L. Beutin. 2006. Phenotypic and genotypic characterization of antimicrobial resistance in *Escherichia coli* O111 isolates. *J. Antimicrob. Chemother.* 57, 1210-1214.

Iino, T., Y. Komeda, K. Kutsukake, R.M. Macnab, P. Matsumura, J.S. Parkinson, M.I. Simon, and S. Yamaguchi. 1988. New unified nomenclature for the flagellar genes of *Escherichia coli* and *Salmonella typhimurium*. *Microbiol. Rev.* 52, 533-535.



Strains	H*	virulence genes	Origin
EcSa35	F4	iucD	Turkey
EcSa36	F4	iucD	Turkey
EcSa66	F4	iucD	Turkey
EcSa84	F4	iucD	Turkey
EcSa105	F4	iucD	Turkey
EcSa117	F4	iucD	Turkey
EcSa127	F4	iucD	Sausage
EcSa85	F4	astA	Sausage
EcSa92	F4	astA	Sausage
EcSa152	F4	st + hlyA	Well water
EcSa173	F4	st + hlyA	Turkey
EcSa103	F4	astA + pap	Turkey
EcSa19	F4	astA	Sausage
EcSa139	F4	astA	Turkey
EcSa148	F4	iucD	Well water
EcSa8	F28	astA	Turkey
EcSa112	F31	astA + iucD + ipaH	Turkey
EcSa40	F31	astA + iucD + ipaH	Turkey
EcSa47	F31	astA + iucD	Turkey
EcSa50	F31	astA + iucD	Turkey
EcSa72	F31	astA + iucD	Turkey
EcSa88	F31	astA + iucD	Turkey
EcSa142	F2	ipaH	Turkey
EcSa143	F2	ipaH	Turkey
EcSa145	F2	ipaH	Turkey
EcSa150	F2	ipaH	Turkey
EcSa51	F30	astA + iucD	Ground beef
EcSa37	F1	iucD	Turkey
EcSa75	F1	astA + iucD	Turkey
EcSa109	F1	iucD	Turkey
EcSa38	F1	iucD	Ground beef
EcSa185	F7	iucD + pap	Sausage
EcSa198	F7	stx1 + stx2 + eaeA + hlyA	Ground beef
EcSa201	F7	stx1 + stx2 + eaeA + hlyA	Ground beef
EcSa81	F7	astA	Turkey
EcSa136	F9	iucD	Turkey
EcSa42	F9	iucD	Turkey
EcSa91	F9	iucD	Ground beef
EcSa168	M1	astA	Sausage
EcSa135	F5	iucD	Turkey
EcSa158	F10	astA	Sausage
EcSa10	F10	astA	Sausage
EcSa25	F10	astA	Sausage
EcSa55	F10	astA	Sausage
EcSa183	F25	iucD	Sausage
EcSa188	F34	lt	Sausage
EcSa189	F34	lt	Sausage
EcSa87	F10	iucD	Turkey
EcSa64	F10	iucD	Ground beef
EcSa74	F10	iucD	Turkey
EcSa76	F10	iucD	Turkey
EcSa44	F8	eaeA	Turkey
EcSa48	F8	eaeA	Turkey
EcSa49	F8	eaeA	Ground beef
EcSa54	F8	eaeA	Ground beef
EcSa62	F8	eaeA	Ground beef
EcSa80	F8	eaeA	Ground beef
EcSa98	F8	eaeA	Sausage
EcSa78	F8	eaeA	Ground beef
EcSa169	F8	iucD	Sausage
EcSa16	F29	hlyA + astA	Turkey
EcSa178	F21	afa	Well water
EcSa21	F10	astA	Sausage
EcSa133	F19	pap	Sausage
EcSa7	F19	stx1 + hlyA + saa	Ground beef
EcSa14	F26	eaeA + astA + pap	Ground beef
EcSa116	F18	iucD	Sausage
EcSa140	F21	astA + iucD	Turkey
EcSa147	F21	astA + iucD	Turkey
EcSa125	F21	pap	Turkey
EcSa45	F9	astA + iucD	Sausage
EcSa110	F11	iucD	Turkey
EcSa33	F21	stx1	Sausage
EcSa2	F11	astA	Ground beef

Fig. 1. Dendrogram generated by Bionumerics software, showing H-antigen (F) generated with *Hha*I restriction of *fliC* PCR products and relevant characteristics of the 74 *E. coli* strains used in this study. The phenogram was constructed using the Dice coefficient and UPGMA analysis. The degree of similarity (%) is shown on the scale.

Joys, T.M. 1988. The flagellar filament protein. *Can. J. Microbiol.* 34, 452-458.

Kaufmann, T.C., A. Engel, and H.W. Remigy. 2006. A novel method for detergent concentration determination. *Biophys. J.* 90, 310-317.

Kerridge, D. 1961. The effect of environment on the formation of bacterial flagella, pp. 41-68. Microbial reactions environment. Cambridge University Press, Cambridge, UK.

Kilger, G. and P.A. Grimont. 1993. Differentiation of Salmonella phase 1 flagellar antigen types by restriction of the amplified *fliC* gene. *J. Clin. Microbiol.* 31, 1108-1110.

Kuwajima, G. 1988. Flagellin domain that affects H antigenicity of *Escherichia coli* k-12. *J. Bacteriol.* 170, 485-488.

Leomil, L., A.F. Pestana de Castro, G. Krause, H. Schmidt, and L. Beutin. 2005. Characterization of two major groups of diarrheagenic *Escherichia coli* O26 strains which are globally spread in human patients and domestic animals of different species. *FEMS Microbiol. Lett.* 249, 335-342.

Levine, M.M. 1987. *Escherichia coli* that cause diarrhea: Enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. *J. Infect. Dis.* 155, 377-389.

Machado, J., F. Grimont, and P.A. Grimont. 2000. Identification of *Escherichia coli* flagellar types by restriction of the amplified *fliC* gene. *Res. Microbiol.* 151, 535-546.

Ratiner, Y.A. 1999. Temperature-dependent flagellar antigen phase variation in *Escherichia coli*. *Res. Microbiol.* 150, 457-463.

Reid, S.D., C.J. Herbelin, A.C. Bumbaugh, R.K. Selander, and T.S. Whittam. 2000. Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* 406, 64-67.

Reid, S.D., R.K. Selander, and T.S. Whittam. 1999. Sequence diversity of flagellin (*fliC*) alleles in pathogenic *Escherichia coli*. *J. Bacteriol.* 181, 153-160.

Stenderup, J. and F. Orskov. 1983. The clonal nature of enteropathogenic *E. coli* strains. *J. Infect. Dis.* 148, 1019-1024.

Swaminathan, B., T.J. Barrett, S.B. Hunter, and R.V. Tauxe. 2001. Pulsenet: The molecular subtyping network for foodborne bacterial disease surveillance, United States. *Emerg. Infect. Dis.* 7, 382-389.

Whittam, T.S. and S.E. Ake. 1993. Genetic polymorphisms and recombination in natural populations of *Escherichia coli*, pp. 223-245. *In* N. Takahata and A.G. Clark (eds.), Mechanisms of molecular evolution, Sinauer Associates, Inc., Sunderland, Mass, USA.

Winstanley, C. and J.A. Morgan. 1997. The bacterial flagellin gene as a biomarker for detection, population genetics and epidemiological analysis. *Microbiology* 143, 3071-3084.

Zhang, W.L., M. Bielaszewska, J. Bockemuhl, H. Schmidt, F. Scheutz, and H. Karch. 2000. Molecular analysis of H antigens reveals that human diarrheagenic *Escherichia coli* O26 strains that carry the *eae* gene belong to the H11 clonal complex. *J. Clin. Microbiol.* 38, 2989-2993.