

Collaboration among EURL-VTEC and ANSES :

**Use of the CRISPR array for specific detection
and sub-typing of EHEC strains :**

Sabine DELANNOY, Aubin FLEISS, Patrick FACH

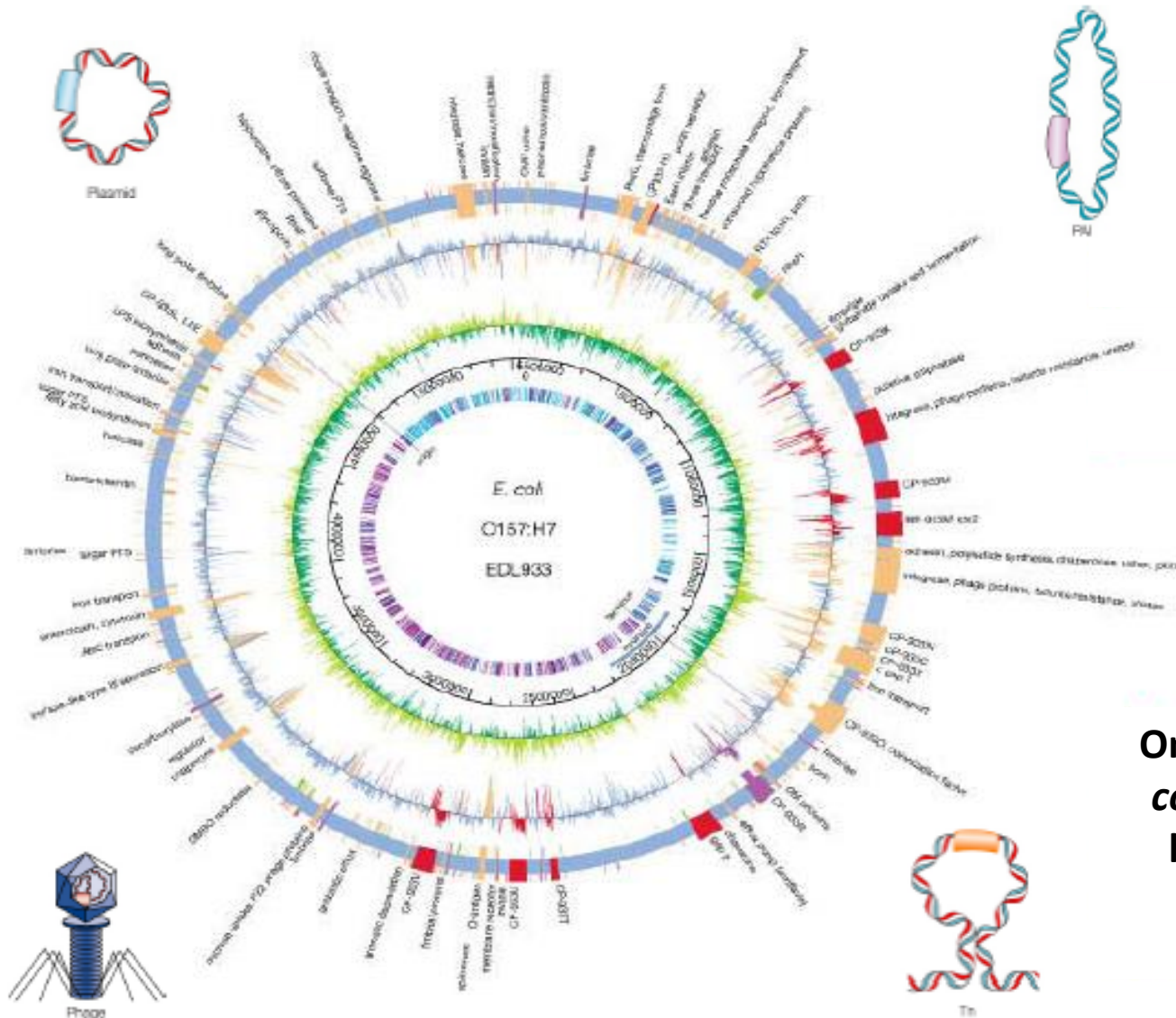
Platform IdentityPath

French Agency for Food, Environmental and Occupational Health & Safety (Anses)

Food Research Laboratory

14 Rue Pierre et Marie Curie,
94701 Maisons-Alfort, France

E. coli genome plasticity

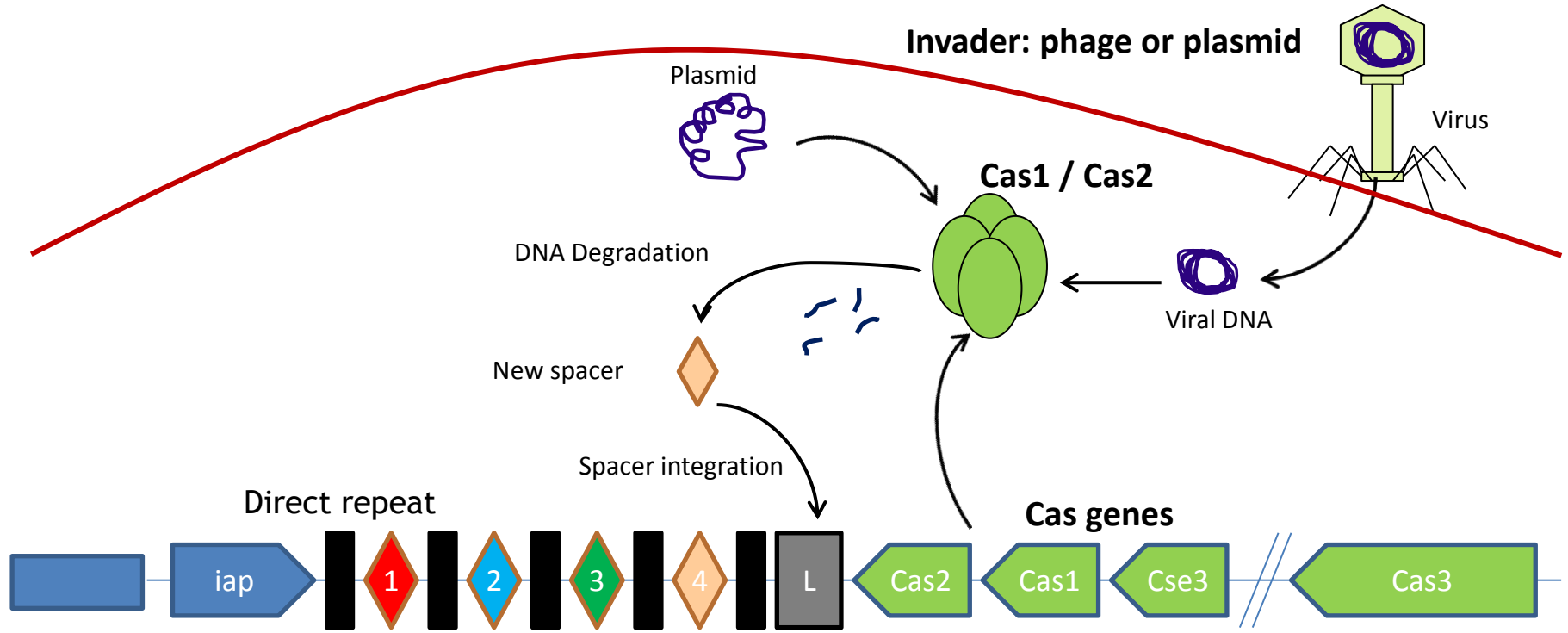


Genome comparison between pathogenic and non-pathogenic *E. coli*

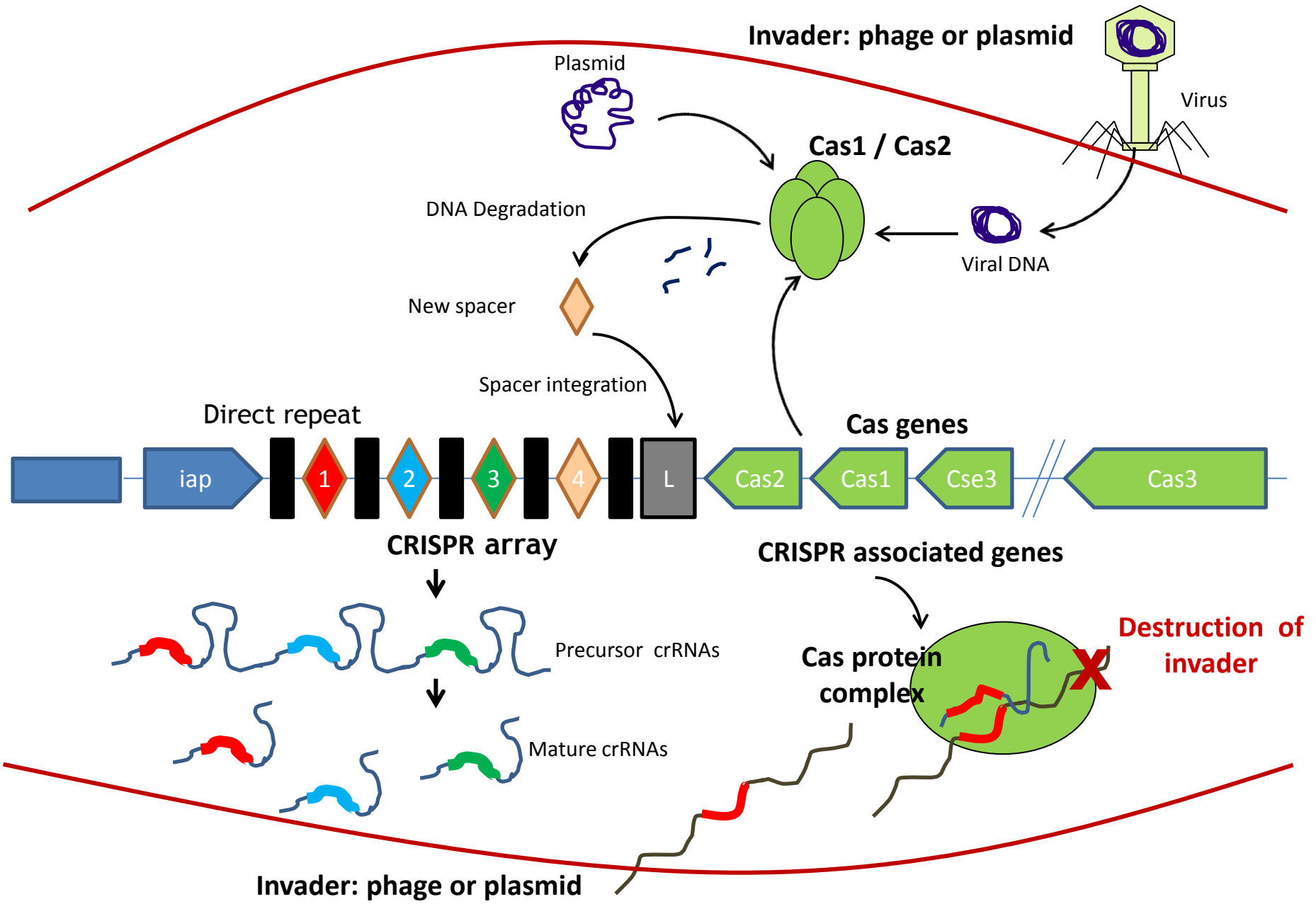
- E. coli* K12 sequences
- Genomic Islands
- Integrated prophages

Only 40% of the genome of *E. coli* (the « core » genome) is highly conserved between strains

CRISPR-mediated adaptive immunity

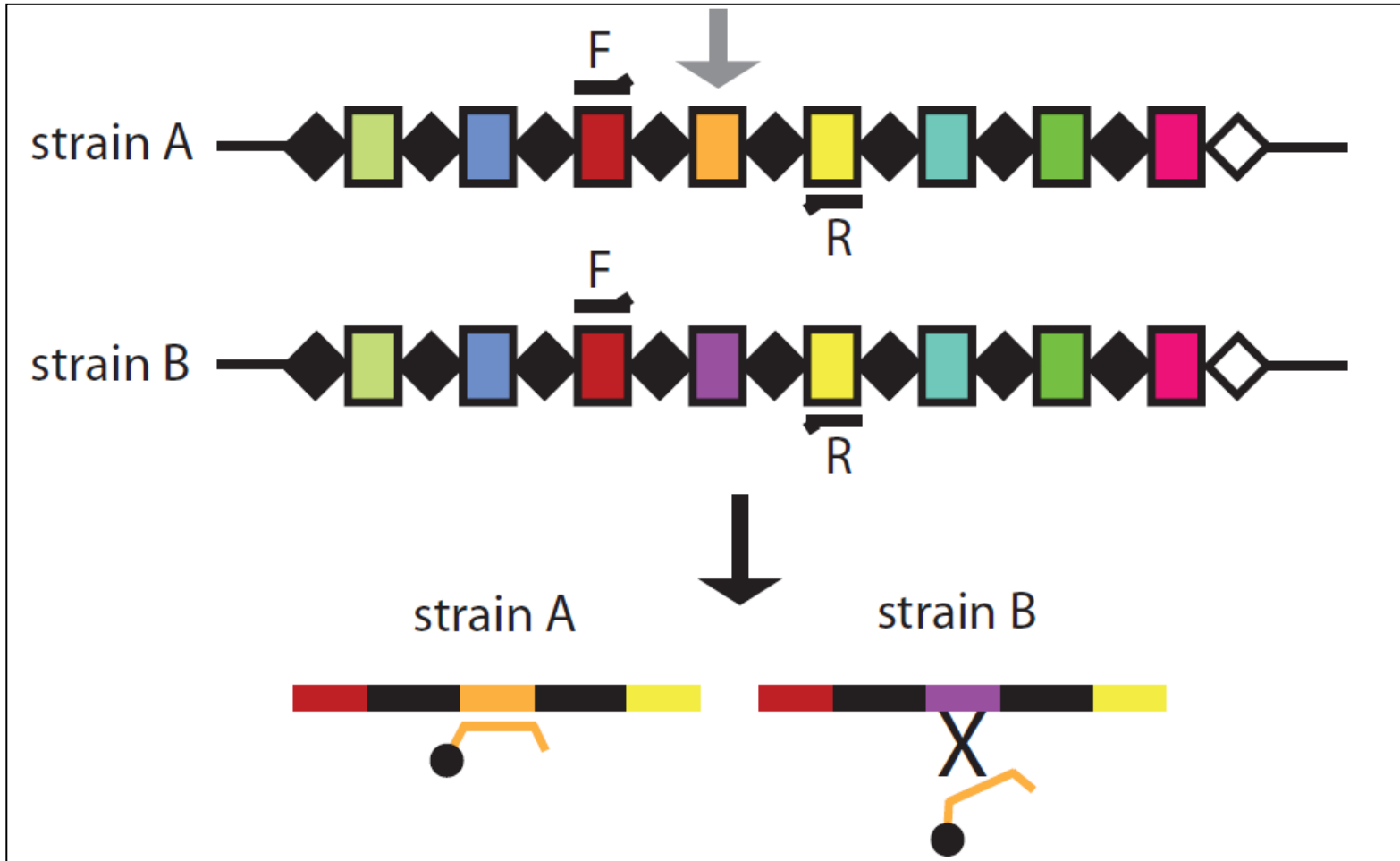


CRISPR-mediated adaptive immunity



Identification of spacers associated with pathogenic or predominant strains

Touchon et al. 2010: 85% of spacers are present in a single genome





Use of Clustered Regularly Interspaced Short Palindromic Repeat Sequence Polymorphisms for Specific Detection of Enterohemorrhagic *Escherichia coli* Strains of Serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19, O145:H28, and O157:H7 by Real-Time PCR

Sabine Dolanoy,^a Lothar Boutsin,^b and Patrick Fach^a

Arnes (French Agency for Food, Environmental and Occupational Health and Safety), Food Safety Laboratory, Maitrons-Allart, France,^a and National Reference Laboratory for *Escherichia coli*, Division of Microbial Toxins, Federal Institute for Risk Assessment (BfR), Berlin, Germany^b

We explored the genetic diversity of the clustered regularly interspaced short palindromic repeat (CRISPR) regions of enterohemorrhagic *Escherichia coli* (EHEC) to design simplex real-time PCR assays for each of the seven most important EHEC serotypes worldwide. A panel of 958 *E. coli* strains investigated for their CRISPR loci by high-throughput real-time PCR showed that CRISPR polymorphisms in *E. coli* strongly correlated with both O:H serotypes and the presence of EHEC virulence factors (*stx* and *eae* genes). The CRISPR sequences chosen for simplex real-time PCR amplification of EHEC strains belonging to the top 7 EHEC serogroups differentiated clearly between EHEC and non-EHEC strains. Specificity estimates for the CRISPR PCR assays varied from 97.5% to 100%. Sensitivity estimates for the assays ranged from 95.7% to 100%. The assays targeting EHEC O145:H28, O103:H2, and O45:H2 displayed 100% sensitivity. The combined usage of two simplex PCR assays targeting different sequences of the O26 CRISPR locus allowed detection of EHEC O26:H11 with 100% sensitivity. By combining two simplex PCR assays targeting different sequences of the EHEC O157 CRISPR locus, EHEC O157:H7 was detected with 99.56% sensitivity. EHEC O111:H8 and EHEC O121:H19 were detected with 95.9% and 95.7% sensitivity, respectively. This study demonstrates that the identification of EHEC serotype-specific CRISPR sequences is more specific than the mere identification of O-antigen gene sequences, as is used in current PCR protocols for detection of EHEC strains.

Shiga toxin-producing *Escherichia coli* (STEC) strains are a diverse group of *E. coli* strains belonging to over 400 O:H/C:H serotypes, some of which cause outbreaks and sporadic cases of food-borne illnesses ranging from diarrhea to hemorrhagic colitis (HC) and the hemolytic-uremic syndrome (HUS) (11, 15, 16). According to their pathogenicity for humans, the latter strains were also designated enterohemorrhagic *E. coli* (EHEC) (17, 18). Numerous cases of HC and HUS have been attributed to EHEC O157:H7 strains (25), but it has now been recognized that other STEC serotypes belong to the EHEC group (8, 24). Cumulative evidence from numerous countries indicates that up to 30 to 60% of human EHEC infections are caused by non-O157 EHEC strains (7). There are seven "priority" EHEC serotypes most frequently implicated in outbreaks and sporadic cases of HC and HUS (8, 24). These comprise serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19, O145:H28, and O157:H7 and their nonmotile derivatives.

Although regulations are disparate throughout the world, many food inspection programs aim at detecting STEC strains that pose a significant threat to human health in foods that are the most likely to disseminate EHEC and to be consumed raw or undercooked. Some beef products are of particular interest in that aspect. The U.S. regulations have precisely been revised to add 6 additional serogroups (O26, O103, O45, O111, O121, and O145) to the existing O157:H7 regulation (22, 23). This regulation imposes testing of specified meat products (i.e., ground beef, beef scraps, and machine-tenderized steaks) for these 7 EHEC serogroups. However, detection of non-O157 EHEC strains is particularly challenging because they have no phenotypical characteris-

tics that distinguish them from the large number of non-STEC strains that share the same habitats.

The current approach for detecting EHEC in food is to screen first for the presence of the *stx*₁/*stx*₂ genes and the *eae* gene (the latter is involved in the attaching and effacing phenotype) in DNA made from bacterial enrichment cultures. The CEN/ISO TS 13136 (8) and MLG 5B.01 (23) standard methods require the presence of both the *stx*₁/*stx*₂ and *eae* genes for further investigation of a suspected EHEC contamination. For this, specific sequences derived from the O-antigen genes associated with the seven priority serogroups are searched. However, the sequences derived from the O-antigen genes of the top 7 EHEC serogroups not only are present in EHEC strains but can be detected in a large number of *Stx*-negative *E. coli* strains as well (1). This sequential approach requires additional time and does not prove that EHEC strains are present in a sample, as multiple *E. coli* strains reacting with each of the PCRs will generate erroneous results. This is of particular interest in complex samples such as food, fecal, and environmental specimens (5, 9, 19). As a consequence, the CEN/ISO TS 13136 and MLG 5B.01 standard methods demand EHEC isolation from the sample to confirm that all the detected genes are associated

Received 7 August 2012; returned for modification 13 September 2012

Accepted 25 September 2012

Published ahead of print 3 October 2012

Address correspondence to Patrick Fach, patrick.fach@anses.fr.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JCM.02050-12

Development of CRISPR based PCR assays for specific detection of new virulent clones :



Specific Detection of Enterohemorrhagic *Escherichia coli* O104:H4 Strains by Use of the CRISPR Locus as a Target for a Diagnostic Real-Time PCR

Sabine Delannoy,^a Lothar Beutin,^b Ylanna Burgos,^b and Patrick Fach^a

Anses (French Agency for Food, Environmental and Occupational Health and Safety), Food Safety Laboratory, Maisons-Alfort, France,^a and National Reference Laboratory for *Escherichia coli*, Division of Microbial Toxins, Federal Institute for Risk Assessment (BfR), Berlin, Germany^b

In 2011, a large outbreak of an unusual bacterial strain occurred in Europe. This strain was characterized as a hybrid of an enterohemorrhagic *Escherichia coli* (EAEC) and a Shiga toxin-producing *E. coli* (STEC) strain of the serotype O104:H4. Here, we present a single PCR targeting the clustered regularly interspaced short palindromic repeats locus of *E. coli* O104:H4 (CRISPR_{O104:H4}) for specific detection of EAEC STEC O104:H4 strains from different geographical locations and time periods. The specificity of the CRISPR_{O104:H4} PCR was investigated using 1,321 *E. coli* strains, including reference strains for *E. coli* O serogroups O1 to O186 and flagellar (H) types H1 to H56. The assay was compared for specificity using PCR assays targeting different O104 antigen-encoding genes (*wbW*_{O104}, *wzX*_{O104}, and *wzY*_{O104}). The PCR assays reacted with all types of *E. coli* O104 strains (O104:H2, O104:H4, O104:H7, and O104:H21) and with *E. coli* O8 and O9 strains carrying the K9 capsular antigen and were therefore not specific for detection of the EAEC STEC O104:H4 type. A single PCR developed for the CRISPR_{O104:H4} target was sufficient for specific identification and detection of the 48 tested EAEC STEC O104:H4 strains. The 35 *E. coli* O104 strains expressing H types other than H4 as well as 8 *E. coli* strains carrying a K9 capsular antigen tested all negative for the CRISPR_{O104:H4} locus. Only 12 (0.94%) of the 1,273 non-O104:H4 *E. coli* strains (serotypes Ont:H2, O43:H2, O141:H2, and O174:H2) reacted positive in the CRISPR_{O104:H4} PCR (99.06% specificity).

More than 400 serotypes of Shiga toxin (Stx)-producing *Escherichia coli* (STEC) strains have been described as agents of disease in humans, and some of these have been shown to be associated with severe diseases, such as hemorrhagic colitis (HC) and hemolytic-uremic syndrome (HUS). These strains were called enterohemorrhagic *E. coli* (EHEC) and were found to carry additional virulence markers besides Stx, such as effectors encoded by the locus of enterocyte effacement (LEE) and various non-LEE-encoded effectors. A concept of molecular risk assessment (MRA) was developed by Karmali et al. (13) and Coombes et al. (9) that employs PCR for identification of human-pathogenic EHEC. Using the MRA approach for screening STEC collections (6, 8), an increasing number of emerging EHEC types was detected.

During spring 2011, Europe faced its largest STEC outbreak involving an emerging enterohemorrhagic *Escherichia coli* O104:H4 strain (1). This EHEC strain presents an unusual virulence pattern that combines the production of Stx2a with enterohemorrhagic adherence which is encoded by genes of the pAA plasmid and chromosomally carried genes of enterohemorrhagic *E. coli* (EAEC) strains (1, 10). This new type of EHEC was designated enterohemorrhagic *E. coli* since it shares virulence markers of both EHEC and EAEC strains. On the genome level (5, 17), the strain was found to be most closely related to an EAEC O104:H4 strain, strain 55989, that was isolated in Central Africa in 1995 (11). This hybrid EAEC STEC O104:H4 strain was found to be negative for the LEE-encoded effector and non-LEE-encoded effector (*nle*), both of which are presently being used by the current MRA approach to define human virulent EHEC types. Therefore, new diagnostic approaches needed to be developed for detection of EAEC STEC O104:H4 strains. The lack of unique biochemical traits of the hybrid EAEC STEC O104:H4 strains

makes their detection with cultural and phenotypical tests difficult and time-consuming. Therefore, rapid molecular testing methods allowing for timely detection of these strains are deemed highly desirable.

During the course of the O104:H4 outbreak investigation, multitarget PCR assays have been used for rapid screening of samples; however, all of these assays require cultural isolation of the bacteria to confirm that all gene targets are present in the same strain. The used PCR assays (4, 12, 21, 26) combine multiple pairs of primers targeting, for example, genes encoding Shiga toxin 2 (*stx*₂), O104 (*fliC*_{O104}) and H4 (*fliC*_{H4}) antigens, telurite resistance (*terD*), and AggR (*aggR*), which is the master regulator of EAEC plasmid, as well as chromosomally inherited virulence genes (18). However, none of these gene targets was unique to the O104:H4 outbreak strain. Therefore, samples containing a mixed flora of bacteria, such as those collected from environmental and food sources, did not allow prediction that all targets were present in the same bacterial strain. Hence, these assays were suitable only for bacterial isolates and have limited use with clinical, food, or environmental samples.

Based on nucleotide sequence analysis of the genome of EAEC STEC O104:H4, we identified in the clustered regularly interspaced short palindromic repeats (CRISPR) locus of the epidemic

Received 27 June 2012. Returned for modification 20 July 2012.

Accepted 6 August 2012.

Published ahead of print 15 August 2012.

Address correspondence to Patrick Fach, patrick.fach@anses.fr.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JCM.01656-12



Characteristics of Emerging Human-Pathogenic *Escherichia coli* O26:H11 Strains Isolated in France between 2010 and 2013 and Carrying the *stx*_{2d} Gene Only

Sabine Delannoy,^a Patricia Mariani-Kurkdjian,^{b,c,d} Stéphanie Bonacorsi,^{b,c,d} Sandrine Liguori,^{b,e} Patrick Fach^a

French Agency for Food, Environmental, and Occupational Health and Safety, Food Safety Laboratory, Platform IdentifiPath, Maisons-Alfort, France,^a AP-HP, Hôpital Robert Debré, Service de Microbiologie, CNR Associé *Escherichia coli*, Paris, France,^b IAME, UMR 1137, INSERM, Paris, France,^c IAME, UMR 1137, Université Paris Diderot, Sorbonne Paris Cité, Paris, France,^d Institut de Veille Sanitaire, Saint Maurice, France^e

Strains of *Escherichia coli* O26:H11 that were positive for *stx*₂ alone ($n = 23$), which were not epidemiologically related or part of an outbreak, were isolated from pediatric patients in France between 2010 and 2013. We were interested in comparing these strains with the new highly virulent *stx*_{2d}-positive *E. coli* O26 clone sequence type 29 (ST29) that has emerged recently in Europe, and we tested them by multilocus sequence typing (MLST), *stx*₂ subtyping, clustered regularly interspaced short palindromic repeat (CRISPR) sequencing, and plasmid (*elexA*, *katP*, *espP*, and *espD*) and chromosomal (*Z2098*, *espK*, and *espV*) virulence gene profiling. We showed that 16 of the 23 strains appeared to correspond to this new clone, but the characteristics of 12 strains differed significantly from the previously described characteristics, with negative results for both plasmid and chromosomal genetic markers. These 12 strains exhibited a ST29 genotype and related CRISPR arrays (CRISPR2a alleles 67 or 71), suggesting that they evolved in a common environment. This finding was corroborated by the presence of *stx*_{2d} in 7 of the 12 ST29 strains. This is the first time that *E. coli* O26:H11 carrying *stx*_{2d} has been isolated from humans. This is additional evidence of the continuing evolution of virulent Shiga toxin-producing *E. coli* (STEC) O26 strains. A new O26:H11 CRISPR PCR assay, SP_O26_E, has been developed for detection of these 12 particular ST29 strains of *E. coli* O26:H11. This test is useful to better characterize the *stx*_{2d}-positive O26:H11 clinical isolates, which are associated with severe clinical outcomes such as bloody diarrhea and hemolytic uremic syndrome.

Enterohemorrhagic *Escherichia coli* (EHEC) is responsible for gastrointestinal diseases such as diarrhea or bloody diarrhea and can lead to hemolytic uremic syndrome (HUS). The most common EHEC serotype associated with human disease is O157:H7. However, a growing number of human EHEC infections are caused by non-O157 EHEC strains (1–4). Among non-O157 EHEC strains, O26:H11 has emerged as the most common serotype associated with severe diarrhea and HUS worldwide (1–6).

EHEC O26:H11 strains are very dynamic; they can undergo frequent genetic rearrangements in their chromosome, virulence plasmids, and pathogenicity islands. They also have the ability to rapidly lose and acquire *stx*-carrying phages (7), which makes them highly adaptable and may account for their global spread. Until recently, EHEC O26:H11 strains isolated from humans mostly harbored Shiga toxin 1 (Stx1) (Stx1a subtype) only or, more rarely, Stx1a associated with the Stx2a subtype. In the middle 1990s, however, a new EHEC O26:H11 clone carrying the Shiga toxin Stx2a subtype alone emerged in Europe (8–18). This new clone has also been observed in South America (19) and in the United States (1). Shiga toxin-producing *E. coli* (STEC) strains carrying the *stx*₂ gene are usually associated with more severe outcomes (20). Indeed, this new O26:H11 clone appears highly virulent and is significantly associated with HUS (8, 9).

Multilocus sequence typing (MLST) analysis shows that the *stx*_{2d}-harboring *E. coli* O26:H11 strains are mostly divided into 2 related phylogenetic groups, i.e., sequence type 21 (ST21) (which also contains the EHEC O26:H11 strains harboring *stx*_{2d} alone or in combination with *stx*_{2a}) and ST29 (which contains the new highly pathogenic clone carrying *stx*_{2d} only) (9). The large EHEC plasmids encoding enterohemolysin (*elexA*), catalase peroxidase

(*katP*), serine protease (*espP*), and type II effector (*espD*) can be found in most EHEC O26:H11 strains (8, 9), and the presence of these specific plasmid virulence determinants can be used to distinguish the 2 clones. ST21 is characterized by the plasmid gene combination *elexA*+/*katP*+/*espP*+/*espD*-, while ST29 exhibits the distinctive combination *elexA*+/*katP*-/*espP*-/*espD*+ (9).

Chromosomally encoded virulence factors such as the locus of enterocyte effacement (LEE) effectors and some type III secretion system effectors were also found to be conserved in the phylogenetic group ST21 (21). In the context of a molecular risk assessment strategy, we previously described a combination of molecular markers for specific identification of EHEC and EHEC-like O26:H11 strains. Assays for these markers included *wzX*_{O26}, *fliC*_{O26}, *one-β*, *stx*, *espK*, and *arC*A single-nucleotide polymorphism (SNP) genotyping (21), as well as a set of PCR tests (SP_O26_C and SP_O26_D) targeting the *cus*

Received 8 August 2014. Returned for modification 16 September 2014.

Accepted 17 November 2014.

Accepted manuscript posted online 26 November 2014.

Citation: Delannoy S, Mariani-Kurkdjian P, Bonacorsi S, Liguori S, Fach P. 2015. Characteristics of emerging human-pathogenic *Escherichia coli* O26:H11 strains isolated in France between 2010 and 2013 and carrying the *stx*_{2d} gene only. J Clin Microbiol 53:486–492. doi:10.1128/JCM.02790-14.

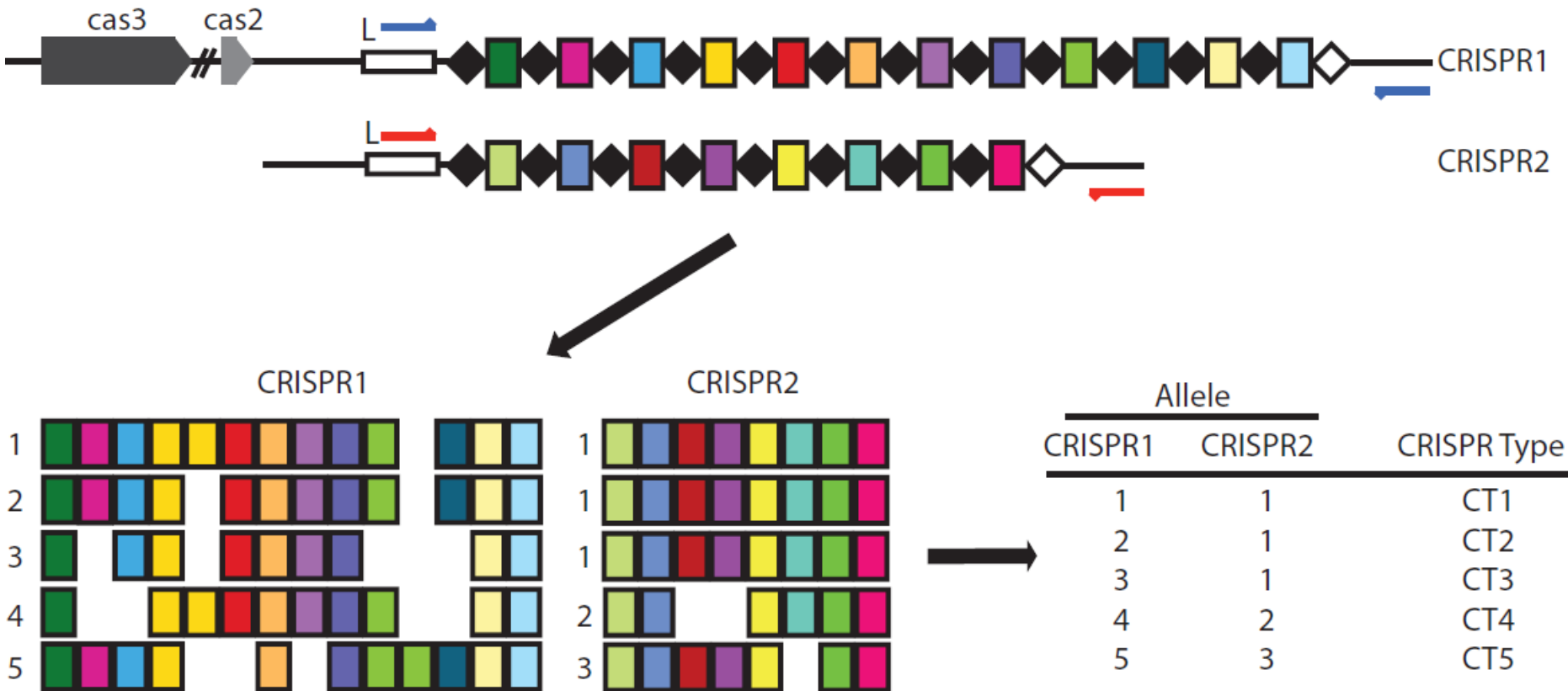
Editor: N. A. Leclercq

Address correspondence to Patrick Fach, patrick.fach@anses.fr.

Copyright © 2015, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JCM.02790-14

CRISPR typing of *Escherichia coli* :



Improved traceability of Shiga-toxin-producing *Escherichia coli* using CRISPRs for detection and typing

Sabine Delannoy¹ · Lothar Beutin² · Patrick Fach¹

Received: 30 June 2015 / Accepted: 16 September 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Among strains of Shiga-toxin-producing *Escherichia coli* (STEC), seven serogroups (O26, O45, O103, O111, O121, O145, and O157) are frequently associated with severe clinical illness in humans. The development of methods for their reliable detection from complex samples such as food has been challenging thus far, and is currently based on the PCR detection of the major virulence genes *stx1*, *stx2*, and *eae*, and O-serogroup-specific genes. However, this approach lacks resolution. Moreover, new STEC serotypes are continuously emerging worldwide. For example, in May 2011, strains belonging to the hitherto rarely detected STEC serotype O104:H4 were identified as causative agents of one of the world's largest outbreak of disease with a high incidence of hemorrhagic colitis and hemolytic uremic syndrome in the infected patients. Discriminant typing of pathogens is crucial for epidemiological surveillance and investigations of outbreaks, and especially for tracking and tracing in case of accidental and deliberate contamination of food and water samples. Clustered regularly interspaced short palindromic repeats (CRISPRs) are composed of short, highly conserved DNA repeats separated by unique sequences of similar length. This distinctive sequence signature of CRISPRs can be used for strain typing in several bacterial species including STEC. This review discusses how CRISPRs have recently been used for STEC identification and typing.

Keywords STEC · *E. coli* · CRISPR · Detection · Typing

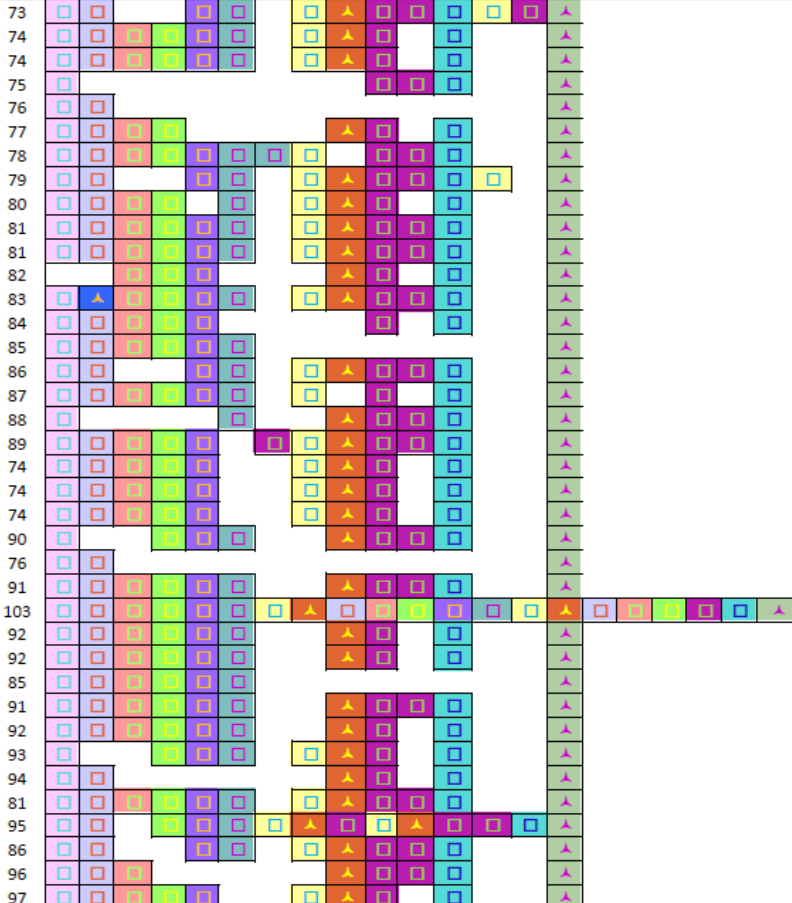
Introduction

Shiga-toxin-producing *Escherichia coli*

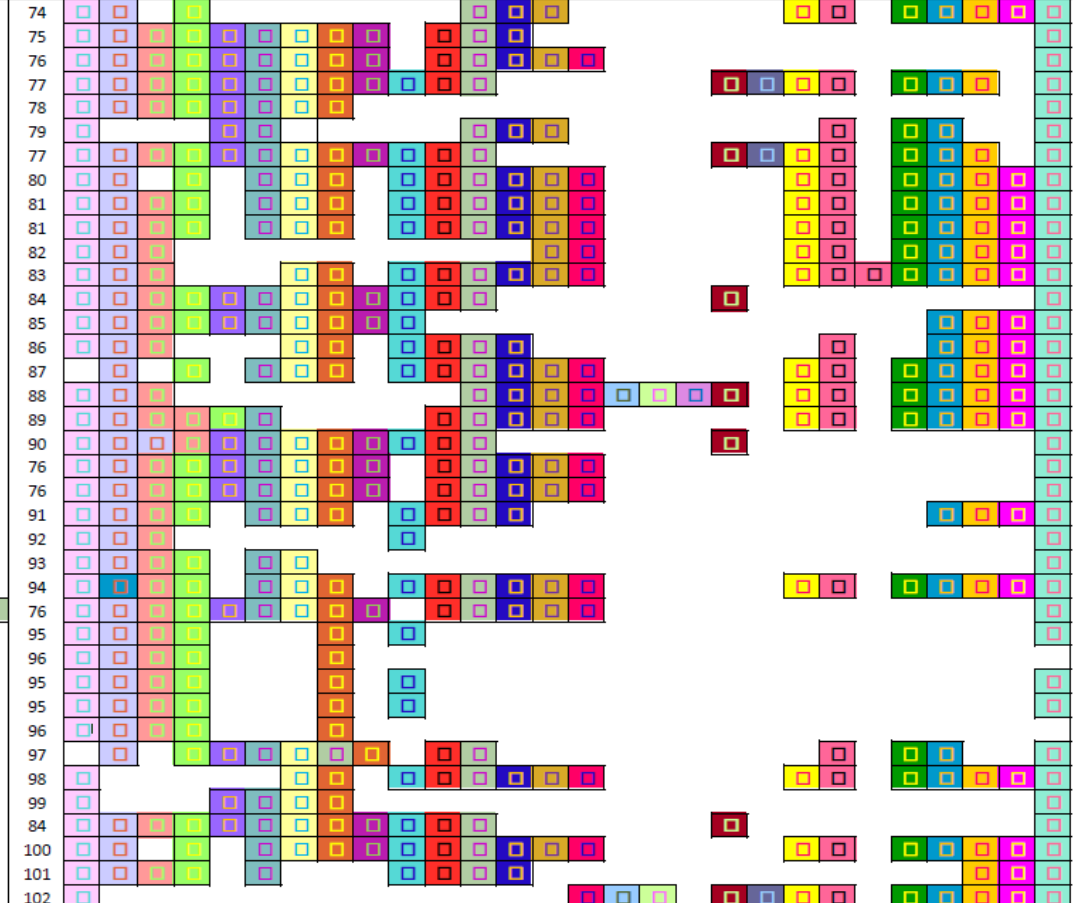
Shiga-toxin-producing *E. coli* (STEC) are food-borne pathogens which can cause various symptoms ranging from mild watery diarrhea to bloody diarrhea, hemorrhagic colitis, and the hemolytic uremic syndrome (HUS) (Nataro and Kaper 1998). Some strains, including the O104:H4 strain that caused the large 2011 epidemic, can also cause central nervous system complications with lasting neurological symptoms (Magnus et al. 2012; Ullrich et al. 2013). Among STEC strains, seven serogroups (O157, O26, O45, O103, O111, O145, and O121) are frequently associated with severe clinical illness in humans worldwide. Regulations for targeting STEC belonging to these serogroups in food have therefore been established in the European Union and the USA (EC 2013; FSIS 2013). Cattle are the main reservoir of STEC, and many human infections are caused by consumption of undercooked ground beef or unpasteurized milk and cheeses. However, contaminated water and sprouted seeds were implicated in some of the biggest outbreaks with STEC worldwide (e.g., O157:H7 in alfalfa sprouts in the USA (CDC 1997), O157:H7 in radish sprouts in Sakai, Japan (Michino et al. 1999; Watanabe et al. 1999), O157:H7 in spinach in the USA (CDC 2006), and O104:H4 in fenugreek

Development of a CRISPR sequences database for EHEC-top7

CRISPR1 Allele



CRISPR2a Allele

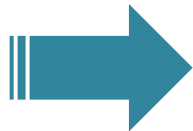


CRISPy Application

>FRIK2000

```
CGCGCTTACGTGGACGGCTCGCAATCTGGCTACTGGAAGTGCGTGCCGGTGTGTA
TGTTGGTGATACATCAAAACGTATTCGGGAGATGATCTGGCAGCAAATTACCCAAC
TGGCTGGTTGCGGAAATGTGGTGATGGCCTGGGGCGACCAATACCGAGTCGGGTTT
TGAATTCAGACCTGGGGAGAAAACAGACGTATTCCGGTGGATTTGGATGGGTTA
CGTTTGGTTTCTTTTCTTCCTGTTGATAATCAATAGGTTATGTGTTCTTTAAAAATAA
GGAAATGTTTGAATTTAGTTGGTAGATTGTTGATGTGGAATAAATTTGTTTAAAAC
AGATATGTATGCTTAGTGTGTTCCCCGCGCCAGCGGGGATAAACCGTCACCAAAC
AGTGACAAAAACTGTCACCAAAGTGTTCCCCGCGCCAGCGGGGATAAACCGCTCA
TATTCGGATTGATCGTGTGTTTCGGTTTGTGTTCCCCGCGCCAGCGGGGATAAAC
GGCCAGGGATTTGTTCAATCCAGCGTGCCGCTGTGTTCCCCGCATCAGCGGGGA
TAAACCGGGCGCACTGGATGCGATGATGGATATCACTTAGAATTCCCCGCCCTGC
GGTAGAACACCCAGCTCCCATTTTCCAACCATCAAGACGCCTTCGCCAACTCCCT
TCACCAA
```

Upstream Seq
A 1 A 2 A 3 B 4 C
Downstream Seq
Allelic Code



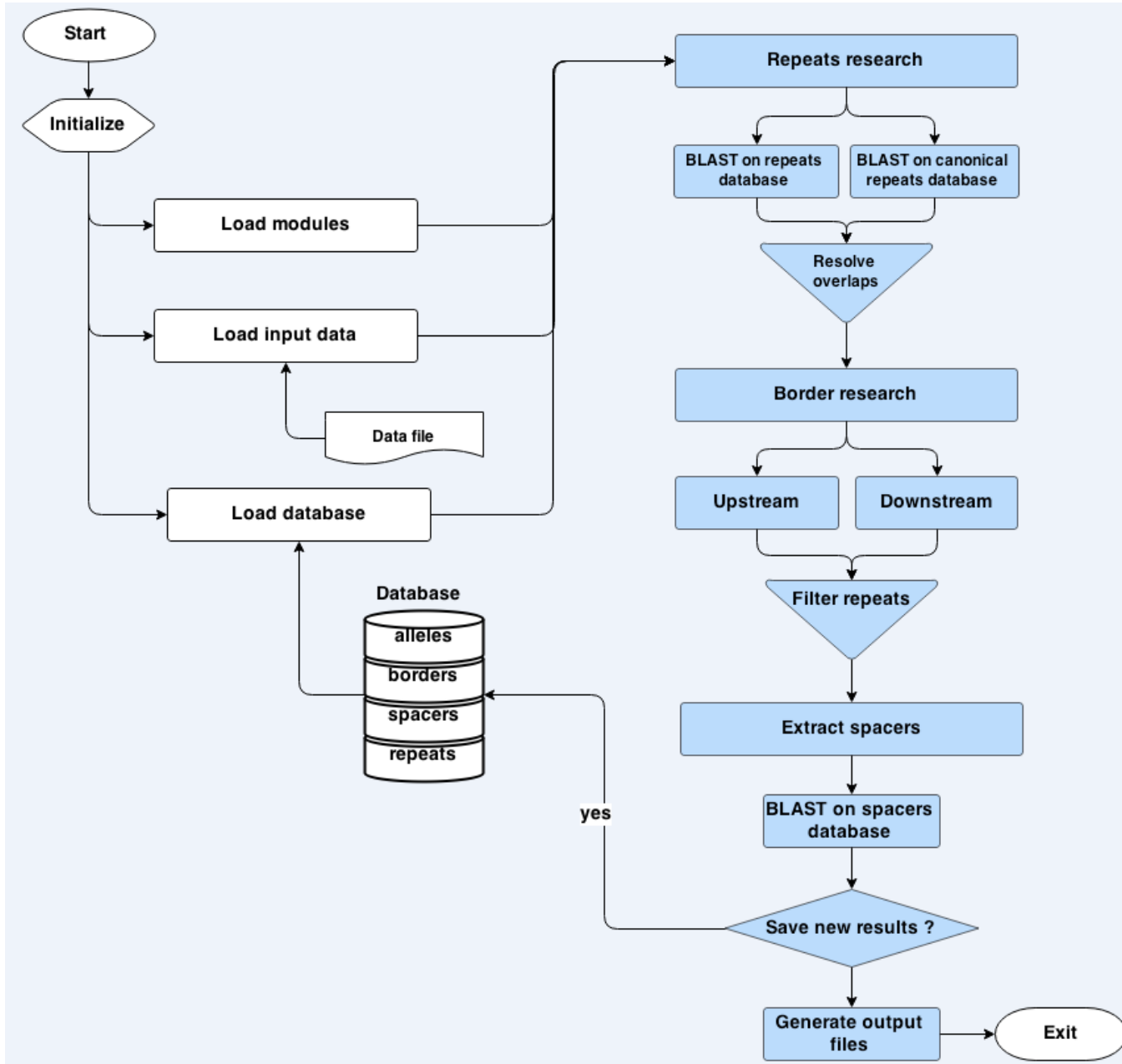
Data base



1

CRISPR Allele

CRISPy Application



CRISPy Application

Data outputs and Graphics

- Strains- Allelic code-Allele number -Timestamp

```
file:///C:/Users/...utput/output.html
file:///C:/Users/asleiss/Desktop/CRISPR_script_16/Python/output/output.html

03/06/2014 15:27:41 top

>FRIK2000
CGCGCTTACGTGGACGGCTCGCAATCTGGCTACTGGAAGTGCCTGCCGGTGTGTATGTTGGTGATACATC
AAAACGTATTCGGGAGATGATCTGGCAGCAAATACCCAACCTGGCTGGTTGCGGAAATGTGGTGATGGCC
TGGGCGACCAATACCGAGTCGGGTTTTGAATTTACAGACCTGGGGAGAAAACAGACGTATTCGGTGGATT
TGGATGGGTACGTTTGGTTCTTTTCTTCTCCTGTTGATAATCAATAGGTATGTGTTCTTAAAAATAAGGAA
ATGTTTGAATTTAGTTGGTAGATTGTGATGTGGAATAAATTTGTTAAAAACAGATATGTATGCTTAGTGTG
TTCCCGCGCCAGCGGGGATAAACCGTCACCAAAACAGTGACAAAACTGTCCACAAAGTGTTCGCCGC
GCCAGCGGGGATAAACCGCTCATATTCCGATTGATCGTGTGTTTCGGTTTGTGTTCCCGCGCCAGCGGG
GATAAACCGGCCAGGATTTGTTCAATCCAGCGTGCCGCTGTGTTCCCGCATCAGCGGGGATAAACCG
GGCGCACTGGATGCGATGATGGATATCACTTAGAATCCCGCCCTGCGGTAGAACACCCAGCTTCATT
TTCCAACCCATCAAGACGCCTTCGCCAACTCCCTTCAACAA

Upstream-A 1 A 2 A 3 B 4 C-Downstream
Allele : 1

Finished sequence analysis. Complete analysis duration was 0:00:01. 5 repeats were found among which 5 were within the
upstream/downstream borders. 0 new repeat(s) were identified.

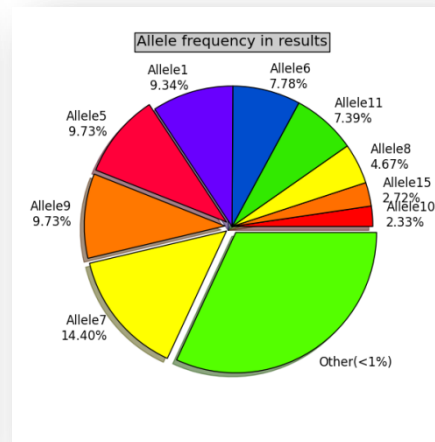
03/06/2014 15:27:41 top

>PA2
CCTCCGCGCTTACGTGGACGGCTCGCAATCTGGCTACTGGAAGTGCCTGCCGGTGTGTATGTTGGTGATA
CATCAAAAACGTATTCGGGAGATGATCTGGCAGCAAATACCCAACCTGGCTGGTTGCGGAAATGTGGTGATG
GCTGGGCGACCAATACCGAGTCGGGTTTTGAATTTACAGACCTGGGGAGAAAACAGACGTATTCGGTGG
ATTTGGATGGGTACGTTTGGTTCTTTTCTTCTCCTGTTGATAATCAATAGGTATGTGTTCTTAAAAATAAG
GAAATGTTTGAATTTAGTTGGTAGATTGTGATGTGGAATAAATTTGTTAAAAACAGATATGTATGCTTAGT
GTGTTCCCGCGCCAGCGGGGATAAACCGTCACCAAAACAGTGACAAAACTGTCCACAAAGTGTTCGCC
CGCGCCAGCGGGGATAAACCGCTCATATTCCGATTGATCGTGTGTTTCGGTTTGTGTTCCCGCGCCAGC
GGGGATAAACCGGCCAGGATTTGTTCAATCCAGCGTGCCGCTGTGTTCCCGCATCAGCGGGGATAAAA
CCGGCGCACTGGATGCGATGATGGATATCACTTAGAATCCCGCCCTGCGGTAGAACACCCAGCTTC
CATTTTCCAACCCATCAAGACGCCTTCGCCAACT

Upstream-A 1 A 2 A 3 B 4 C-Downstream
Allele : 1

Finished sequence analysis. Complete analysis duration was 0:00:01. 5 repeats were found among which 5 were within the
upstream/downstream borders. 0 new repeat(s) were identified.

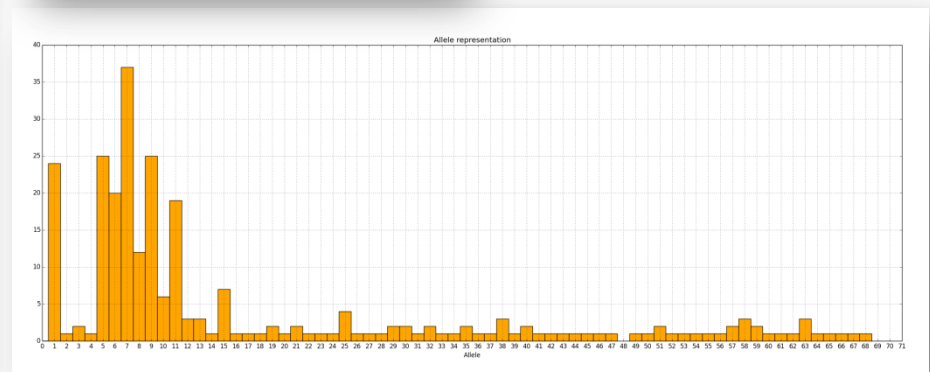
03/06/2014 15:27:41 top
```



Data outputs with HTML format

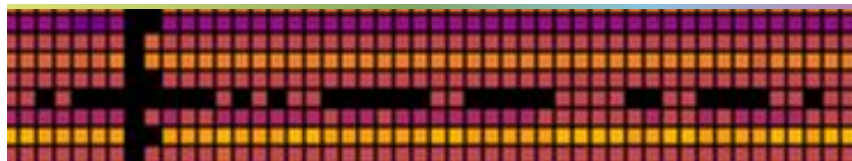
Diagram with sectors

Histogram



Conclusions & Perspectives

- The **data base of the CRISPR sequences** has been transferred to ISS
- CRISPy works with short sequences (PCR products = CRISPR array amplified by PCR)
- For Whole Genome Sequences the CRISPR array should be manually extracted (time consuming) and then analyzed by CRISPy
- ➔ Work is in progress with ISS to use CRISPy directly after Whole Genome Sequencing.
- Collaboration between **Anses** and **ISS** to improve CRISPY and **make it available from the ARIES platform.**



Thanks!



**Anses,
Laboratoire de Sécurité des Aliments
plateforme IdentityPath**

**14 Rue Pierre Curie,
94700 Maisons-Alfort, France**



Bio-informatical tools

- On-line (Université Paris-Sud)
 - Truncated loci
 - Incomplete detection
 - No nomenclature
- R Script (Yin et al. 2013)
 - No database
 - Interpretation biais
 - Not easy to use
 - Nomenclature will change with each dataset
- Need a better bioinformatic tool:
 - Fast
 - Objective
 - Stable nomenclature



CRISPR sequence diversity and applications

Strain	MLST	CRISPR type
CB6699	223	4
CB8531	223	7
CB8578	223	8
Ec253/02	223	10
Ec585/05	223	11
Ec254/01	223	12
CB6110	223	12
MDP09-27	223	14
MDP09-47	223	15
NV254	223	16
Ec226/04	223	19
188/06-28	223	20
226/99	223	21
07HMPA903	223	22
CB5250	223	23
571/05	223	24
1112/06	223	25
889/06	223	25
997/01	223	26
1108/01	223	27
CB9070	223	29
CB7612	223	30
CB7561	223	32
FP-054	223	33
Ec784	223	39
258/04	223	39
FP-120	223	44
MDP10-35	223	46
T842	223	48
TW01391	223	49

CRISPR typing is more resolutive than other typing methods:

- MLST
- Virulo-typing

Exemple : CRISPR-type diversity in a single ST.

Strain	MLST	CRISPR type
CB7267	820	1
EH41	820	9
98NK2	820	13
226/1	846	3

CB7960	846	6
CB2125	846	28
Ec596/05	846	40

Source : Feng et al., 2014