
METADETECT: DETECTION OF SHIGA TOXIN-PRODUCING *ESCHERICHIA COLI* WITH NOVEL METAGENOMICS APPROACHES AND ITS APPLICATION ON DAIRY FARMS IN FRANCE AND GERMANY

Thèse de Sandra **JAUDOU-DUREUIL**¹

Analysée par Anne **BRISABOIS**²

Directrice de thèse : Sabine **DELANNOY**, chargée de projet recherche, PHD HDR, Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (Anses)

Co-encadrement :

Joséphine **GRÜTZKE**, scientifique PHD, (Bundesinstitut für Risikowertung (BfR))

Patrick **FACH**, directeur de recherche, PHD, HDR, chef de l'unité *Escherichia coli* pathogènes (COLiPATH), Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (Anses)

La thèse présentée par Sandra Jaudou-Dureuil a pour objectif principal de développer des approches innovantes et performantes pour la caractérisation des *Escherichia coli* (*E. coli*) producteur de Shiga toxines (STEC) dans les produits au lait cru. En effet, au sein des bactéries de l'espèce *E. coli*, il faut distinguer les souches appartenant à la flore commensale de l'Homme, ayant potentiellement un rôle bénéfique, de celles qui présentent des facteurs de virulence leur conférant un caractère pathogène. Il existe donc une très grande diversité de pathovars selon leur pouvoir pathogène sur les différents organes et la présence de facteurs de virulence. Les STEC appartiennent aux *E. coli* diarrhéiques capables de produire des Shiga-toxines, parmi lesquels un sous-groupe de souches est responsable d'infections sévères chez l'Homme et, en particulier, chez les jeunes enfants qui peuvent développer un syndrome hémolytique et urémique (SHU) avec de graves conséquences cliniques. C'est le sous-groupe des EHEC (*E. coli* enterohémorragique), porteur du gène *eae* codant pour l'intimine responsable de l'adhésion, qui provoque des lésions d'attachement et d'effacement de l'épithélium intestinal. La contamination des aliments par les STEC s'effectue à partir d'un environnement souillé par les ruminants, principal réservoir. C'est ainsi que les produits dérivés tels que le lait cru et la viande de bœuf, peuvent être contaminés, mais aussi plus

¹ Thèse préparée pour obtenir le titre de docteur de l'Ecole nationale vétérinaire d'Alfort, spécialité : Microbiologie, Ecole doctorale n°581 Agriculture, alimentation, biologie, environnement et santé (ABIES), préparée dans l'Unité « COLiPATH - LSA Laboratoire de Sécurité des Aliments - Anses », présentée et soutenue à Maisons-Alfort, le 19 décembre 2023.

² Membre correspondant de l'Académie d'agriculture de France, section 6, « Sciences de la vie ».

largement d'autres produits comme les végétaux à partir des fèces ou déjections sur le sol ou l'eau d'irrigation des cultures.

Les symptômes engendrés par ce groupe de bactéries sont variables, allant de la diarrhée aqueuse bénigne à la diarrhée sanglante sévère pouvant évoluer vers le SHU. Le taux de mortalité lié aux STEC est estimé entre 3 et 5 % en Europe. Les jeunes enfants développent plus fréquemment des formes graves avec une atteinte rénale nécessitant une dialyse, ce qui fait de cette bactérie l'une des plus virulente parmi les bactéries transmises par les aliments. Ces épisodes sont fréquemment médiatisés avec des conséquences économiques non négligeables. C'est ainsi que récemment (janvier 2024), le morbier au lait cru a été responsable de SHU chez de jeunes enfants en France, mais aussi aux Etats Unis où du cheddar au lait cru a été responsables de SHU en février 2024. Par ailleurs, un épisode de cas groupés à STEC fortement médiatisé est survenu en France suite à la consommation de pizzas surgelées à l'origine de 56 cas dont 55 enfants avec des formes sévères de SHU. On peut ainsi considérer que les STEC sont des pathogènes à transmission principalement alimentaire et que les principales sources de contamination humaines par les STEC sont les viandes hachées et dérivés, suivies du lait cru et des produits à base de lait cru, les végétaux et les eaux de contact (baignades...).

Compte tenu de ces éléments, il apparaît nécessaire de pouvoir détecter et caractériser la présence de ces bactéries tout au long de la chaîne alimentaire et de façon indispensable sur les produits alimentaires mis sur le marché pour respecter la conformité de la réglementation européenne du Paquet Hygiène. Les travaux de la thèse s'orientent vers le développement de multiples approches permettant la caractérisation des souches STEC directement dans le lait cru, sans passer par les étapes laborieuses d'isolement des souches. En effet, la méthode actuelle de recherche des STEC dans les aliments préconise une première étape de détection par PCR des gènes de virulence *eae* et *stx* dans les bouillons d'enrichissement de l'échantillon. Lorsque cette étape s'avère positive, il est alors nécessaire de rechercher les souches en réalisant des étapes d'isolement sur milieu de culture, bien souvent infructueuse du fait de la flore microbienne associée et de l'absence de milieu d'isolement spécifique au STEC. De ce fait, la méthode conventionnelle actuelle ne permet pas de reconnaître une souche STEC à potentiel de virulence. La méthode développée lors des travaux de thèse, qui s'appuie sur la métagénomique, vise au contraire à identifier directement dans l'échantillon de lait cru, une souche présentant simultanément les gènes *eae* et *stx* dans son génome, signature d'une virulence avérée de la souche. Pour ce faire, le séquençage long reads, qui génère de longues séquences par la technologie ONT (Oxford Nanopore Technology), a été privilégié. Cet objectif ambitieux et complexe a nécessité de nombreux ajustements et développements technologiques avant d'aboutir à un schéma complet, depuis les étapes d'enrichissement jusqu'à l'analyse des données et la caractérisation complète *in-situ* dans l'aliment de la souche. Les principaux obstacles portaient notamment sur la nécessité d'obtenir une grande quantité d'ADN de haut poids moléculaire et exempt d'impuretés, sources d'artefacts pour le séquençage. Par ailleurs, un pipeline informatique dédié, appelé STECmetadetector, a été spécifiquement développé pour l'analyse des données ONT, qui extrait uniquement les séquences *Escherichia coli* pour les assembler et détecter la présence des deux gènes ciblés *eae* et *stx*. Enfin, l'un des obstacles limitant dans le processus développé était le ratio entre la souche STEC et les souches *E. coli* commensales, au-dessous duquel la caractérisation des

STEC n'est plus possible. Pour compléter le travail, une approche alternative basée sur l'intelligence artificielle de type machine learning a été développée permettant d'identifier la présence de STEC potentiellement porteurs des gènes *eae* et *stx* à l'aide de la reconnaissance d'autres marqueurs génétiques associés déjà identifiés en partie par l'équipe de recherche encadrante.

Ainsi, les travaux de thèse présentés ont permis de lever ou de contrer les différents obstacles ou limites basés sur des approches technologiques innovantes, permettant de mettre à disposition une méthode performante, efficace et sûre pour la détection des STEC et de leurs facteurs de virulence. Ils ont montré clairement tout le potentiel de la méta-génomique « long read » afin de caractériser des souches STEC positives pour le gène *eae* au sein de lait cru. Cette approche permet la caractérisation des souches sans avoir recours à l'étape d'isolement. Cependant, les conditions mentionnées (au moins 10^8 copies. mL⁻¹ de STEC après enrichissement et dix fois plus de STEC que d'autres souches *E. coli*) doivent être respectées. L'utilisation du machine learning permet néanmoins l'identification de souches STEC positives pour le gène *eae* lorsque le ratio obtenu n'est pas respecté mais que la quantité de STEC après enrichissement permet d'obtenir suffisamment de données en séquençage afin que les gènes soient présents dans l'assemblage. Cette méthode ne vise pas à remplacer l'étape de détection des gènes *stx* et *eae* par qPCR, mais à remplacer l'étape intensive d'isolement réalisée au sein de laboratoire de référence, afin de caractériser les souches potentiellement pathogènes. La méthode métagénomique « long read » développée présente un intérêt primordial en santé publique puisqu'elle permet d'identifier et de caractériser des souches STEC pathogènes dans le cas où la méthode de référence est en défaut et ne permet pas cette détection. Par ailleurs, la méthode a été développée en utilisant la matrice au lait cru, bien justifié par la fréquence de cas associés à la consommation de fromages au, lait cru et peut s'étendre sur d'autres matrices alimentaires ou d'autres échantillons (fèces par exemple) avec les ajustements nécessaires en lien avec le type de matrice. De plus, cette approche pourrait également être développée pour la détection d'autres bactéries pathogènes difficiles à isoler ou lorsque l'isolement fait défaut sur les milieux de culture sélectifs. Néanmoins, ces premiers travaux originaux devront être consolidés sur différents types d'échantillons naturellement contaminés en intégrant les aspects de mise en œuvre et de logistique dans les laboratoires, tout en maîtrisant le coût de l'analyse.

L'ensemble du processus global décrit dans la thèse correspond à un projet ambitieux et innovant avec un impact en sécurité sanitaire des aliments. Ce projet ouvre des perspectives intéressantes basées sur des nouvelles approches couplant la métagénomique et l'intelligence artificielle pour la détection et la caractérisation des STEC pathogènes dans les aliments. Les retombées attendues permettront à terme de prévenir les épisodes à répétition d'infections de cas humains à STEC et leurs conséquences sévères associées.

Le travail de thèse de Sandra Jaudou-Dureuil a été mené dans le cadre d'un partenariat franco-allemand entre l'Anses et le BfR (Bundesinstitut für Risikowertung, Institut fédéral allemand d'évaluation des risques sur les questions de sécurité alimentaire). Sandra Jaudou-Dureuil a effectué deux séjours de six mois à Berlin dans le cadre de sa thèse qui a été rédigée et présentée en anglais. Il a fait l'objet de cinq publications internationales dans des revues à haut facteur d'impact et de trois présentations internationales et deux communications au niveau national. Sandra Jaudou-Dureuil a remporté le deuxième prix de la meilleure

présentation orale au congrès international VTEC 2023 qui s'est tenu à Banff. En obtenant ce prix, les équipes de l'Anses et du BfR menant des travaux conjoints sur les STEC en métagénomique obtiennent une vraie reconnaissance de la communauté scientifique travaillant sur ce germe.

Pour conclure, l'Académie d'agriculture de France salue le travail de Sandra Jaudou-Dureuil qui présente une avancée notable et reconnue pour la sécurité des produits alimentaires de la fourche à la fourchette, en considérant spécifiquement la filière bovine, réservoir principal des STEC. Ce travail présente également un intérêt considérable pour les acteurs de la production et transformation en agro-alimentaire et pour les décideurs lors de l'analyse et de l'évaluation du risque associé à un produit alimentaire.

C'est pour ces raisons que les éléments apportés par ce travail de thèse justifient que cette analyse figure sur le site de l'Académie d'agriculture de France et dans le Mensuel, à titre de valorisation.