

Table des matières

Tab	le	e des matières	3
Abr	é	viations	4
Ren	n	erciements	5
l. Frai		Contexte de l'étude d'épidémiosurveillance génomique rétrospective de <i>Salmonella</i> Dublin en che-Comté	
II.		Objectifs et organisation de l'étude	7
III.		Une étude menée en 3 phases	8
1	.)	Détermination du panel de souches de l'étude (Phase I)	8
		Recensement et collecte des métadonnées des souches de S. Dublin	9
		Échantillonnage et sélection des souches	. 10
		Préparation et transfert des souches de S. Dublin	. 12
2	2)	Séquençage des 500 génomes sélectionnés (Phase II)	. 12
		Récupération des souches et remise en culture	. 12
		Extraction selon une méthode semi-automatique de l'ADN des souches et vérification de la qualité	. 13
3	3)	Apport des données exploitées et enseignements (Phase III)	. 13
		Traitement des données génomiques	. 13
		Arbre phylogénétique obtenu à partir des séquences d'ADN des 441 souches de S. Dublin, sur base de l'analyse SNP du core-genome : Arbre A	
		Interprétation de l'Arbre A à l'aide des métadonnées minimales	. 17
		Informations apportées par la construction de l'Arbre B (arbre phylogénétique constitué des souches de l'Arbre A et de 31 souches d'origine humaine)	
		Interprétation des cas ciblés de l'arbre A avec les métadonnées complètes	. 18
IV. mei	n	Conclusions de l'étude épidémiologique génomique rétrospective de <i>Salmonella</i> Dublin ée en Franche-Comté	. 19
Réf	é	rences citées	. 21
INA	N	EXE 1 : Méthodologie et processus de sélection aléatoire des souches de S. Dublin	. 23
INA	N	EXE 2 : Description des 72 clusters de l'Arbre A selon leur maillon d'origine	. 24

Abréviations

Abréviation	Explication
ANSES	Agence Nationale de sécurité de l'alimentation, de l'environnement et du travail
Cniel	Centre National Interprofessionnel de l'Economie Laitière
CTFC	Centre Technique des Fromages Comtois
GDS	Groupement de Défense Sanitaire
GT	Groupe de Travail
НА	Hygiène Alimentaire
NGS	Next Generation Sequencing (Techniques de séquençage de haut débit)
pb	paire de bases
PTF SCA	Plateforme de Surveillance de la Chaine Alimentaire (<u>www.plateforme-sca.fr</u>)
SA	Santé Animale
SNGTV	Société Nationale des Groupements Techniques Vétérinaires
SNP	Single Nucleotide Polymorphism (Polymorphisme pour un seul nucléotide)
URFAC	Union Régionale des Fromages d'Appellation d'Origine Comtois
WGS	Whole Genome Sequencing (Séquençage du génome entier)

Remerciements

Nous remercions l'ensemble des membres du groupe de travail de cet axe pour leur implication dans la réalisation de cette étude, à savoir Hélène AMAR (DGAL), Laurent GUILLIER (ANSES), Renaud LAILLER (ANSES), Ludovic MALLET (ANSES, (affiliation actuelle, Institut Claudius Régaud)), Nicolas RADOMSKI (Anses, (affiliation actuelle, IZSAM-GENPAT)), Madeleine DE SOUSA VIOLANTE (ACTALIA), Choreh FARROKH et Gaëtan PODEUR (Cniel), ainsi que les membres du groupe de suivi local : Florence ARNAUD (Syndicats AOP Mont d'Or et Morbier), Cédric CHAPUIS (GDS), Maryse COLOMB et Françoise POZET (LDA39), Jean-Marie DUCRET, Charlotte MARION et Lydie BEY (CTFC), Lionel GRISOT (SNGTV et GTV Bourgogne – Franche-Comté), Natacha WORONOFF-REHN (LVD25), Valérie MICHEL (ACTALIA) qui se sont fortement impliqués également dans l'interprétation des données obtenues au regard des métadonnées disponibles.

Nous remercions également François-Xavier WEILL (Institut Pasteur) pour ses précieux conseils méthodologiques durant la phase II de l'étude.

Ce projet a été financé par la Direction générale de l'alimentation et le Centre national interprofessionnel de l'économie laitière.







I. Contexte de l'étude d'épidémiosurveillance génomique rétrospective de Salmonella Dublin en Franche-Comté

Dans le cadre de la Plateforme de Surveillance Sanitaire de la Chaîne Alimentaire (<u>PTF SCA</u>), un premier groupe de travail (GT) portant sur la « Surveillance des salmonelles en filière laitière bovine » a été créé en juin 2017. Ce groupe de travail était lui-même organisé en deux sous-groupes thématiques :

- le premier (axe I) était consacré à la rédaction d'un <u>document d'aide méthodologique</u> pour surveiller *Salmonella spp.* en filière bovine de fabrication de fromages au lait cru; il est disponible sur le site de Plateforme SCA (https://www.plateforme-sca.fr/)
- le second (axe II) avait pour mission de construire un cahier des charges technique et financier pour la réalisation d'une étude d'épidémiosurveillance génomique vis-à-vis de Salmonella en filière fromagère.

Ce second sous-groupe a ainsi abouti, en janvier 2018, à définir les axes de réalisation d'une telle étude et à identifier un terrain et une filière fromagère volontaire pour y participer.

Le travail d'instruction mené dans le cadre de l'axe II s'est appuyé en premier lieu sur les données épidémiologiques disponibles concernant Salmonella spp. Parmi les près de 2 600 sérovars existant pour Salmonella enterica, un nombre important d'entre eux peut être retrouvé dans une grande diversité de matrices alimentaires dont les ovo-produits, les charcuteries, les volailles ou encore les produits laitiers fabriqués à partir de lait cru (Santé Publique France, 2021). Selon le bilan annuel des laboratoires du réseau Salmonella animé par l'ANSES, pour les fromages fabriqués à base de lait de vache, S. Dublin (34,3 %), S. Typhimurium (15,2 %), S. Kimuenza (13,1 %) et S. Enteritidis (12,1 %) étaient en 2016 les quatre principaux sérovars isolés sur les treize recensés dans ce type de produits. Les quatre sérovars S. Dublin (28,2 %), S. Montevideo (12,8 %), S. IV 40:z4 ,z23:- (11,4 %) et S. Typhimurium (8,7 %) quant à eux font partie des principaux sérovars isolés du lait de vache en France en 2016. Par ailleurs, les travaux du réseau Salmonella (ANSES) ont montré que, entre 2008 et 2015, la distribution des sérovars retrouvés dans le lait cru et les fromages au lait cru était très majoritairement dominée par S. Dublin, respectivement de 67% et 62.7%. Ce sérovar est réparti sur l'ensemble du territoire avec un tropisme en Franche-Comté et en Auvergne (Leclerc et al., 2019). Le sérovar Dublin est donc un danger préoccupant pour la filière laitière au lait cru. Il s'avère être aussi l'un des 20 sérovars les plus importants en santé humaine (CNR, Institut Pasteur 2019) et a été impliqué dans des bouffées épidémiques (entre 50 et 100 cas par an recensés au CNR). Ce sérovar, nettement bactériémique (isolé dans le sang), est majoritairement associé à des groupes à risque (personnes âgées et/ou immunodéprimées). Il peut s'avérer être la cause du décès de ces personnes fragiles (Crump et al., 2015, Ung et al, 2019).

Un bilan a également été mené sur les méthodes de typage utilisées afin de différencier les souches de *S.* Dublin dans un contexte épidémique. Les méthodes conventionnelles de génotypage, qui regroupent la macrorestriction d'ADN (comme la technique de PFGE (Pulse Field Gel Electrophoresis), l'analyse des motifs génétiques répétés (comme le MLVA (Multi Locus VNTR Analysis) ou le CRISPR) ou encore l'étude de la diversité allélique des gènes (comme le MLST-7 gènes ou autre core-génome MLST (cgMLST) (Multi Locus Sequence Type)) se sont avérées insuffisantes pour discriminer précisément les souches de *S.* Dublin dans un contexte d'investigation de cas humains en augmentation. Le WGS ou Whole Genome Sequencing, qui se base sur l'analyse de la séquence complète du génome bactérien, via la recherche des polymorphismes nucléotidiques (Single Nucleotide Polymorphism) associés à l'investigation épidémiologique de terrain, est la méthode la plus discriminante pour ce sérovar (Ågren et al., 2016). Ainsi, l'utilisation du WGS pour réaliser une étude épidémiologique rétrospective sur le sérovar Dublin, l'un des sérovars de *Salmonella enterica* subsp. *enterica* majoritairement associé aux

fromages au lait cru pour la période 2008 à 2015, présentait un intérêt certain dans le cadre du GT « Surveillance des salmonelles en filière laitière bovine ».

Dans ce contexte endémique et avec l'amélioration des méthodes de typage, la région Franche-Comté et les professionnels de la filière laitière au lait cru avaient comme volonté de sécuriser leur production afin de proposer au consommateur les produits les plus sûrs au niveau sanitaire. Plus précisément, la filière souhaitait améliorer les moyens techniques permettant d'explorer les cas de contamination de la chaîne de production fromagère. Cette filière s'est donc portée volontaire pour constituer le cadre d'une étude d'épidémiosurveillance génomique vis-à-vis de S. Dublin.

Construit avec les acteurs régionaux de cette filière (incluant les acteurs du domaine vétérinaire, de l'analyse en laboratoire, des filières fromagères (appui technique et représentants professionnels), l'interprofession laitière (Cniel), ACTALIA, l'Institut Pasteur et les membres opérationnels de la PTF SCA, le cahier des charges technique et financier de cette étude a pu ainsi être concrétisé, aboutissant à la réalisation de l'étude d'épidémiosurveillance génomique rétrospective de *Salmonella* Dublin en Franche-Comté qui s'est déroulée entre juin 2018 et février 2020.

II. Objectifs et organisation de l'étude

L'objectif général de cette étude rétrospective était de confronter l'outil analytique de typage, plus communément utilisé dans le domaine de la santé humaine, le WGS, pour appréhender la circulation clonale et les sources de contamination de S. Dublin dans une région productrice de fromages au lait cru. Au-delà des enseignements techniques de cette étude, enseignements qui pourraient être propres au secteur géographique étudié et à son fonctionnement, cette étude avait également pour objectif d'acquérir des enseignements méthodologiques quant à la réalisation d'étude épidémiologique utilisant le WGS. Ce dernier point est à souligner car, d'une part, il correspond aux attentes et productions de la PTF SCA pour optimiser les méthodes de surveillance des dangers microbiens et, d'autre part, le contexte de la présente étude se focalise principalement sur l'amont de la chaîne alimentaire allant du domaine de l'élevage laitier jusqu'à celui du produit fini, ici le fromage au lait cru.

Cette étude rétrospective a été organisée en plusieurs étapes techniques de travail, qui consistaient à :

- 1. Faire un inventaire des prélèvements et des souches de *S*. Dublin isolées entre 2015 et 2017 et conservées dans les laboratoires du territoire franc-comtois, et des métadonnées associées (= données décrivant le lieu, la date et le contexte de prélèvement, la matrice utilisée pour l'analyse, le laboratoire d'analyse, etc.) ;
- 2. Collecter toutes ou une partie de ces souches et leurs métadonnées (minimales et complètes);
- 3. Analyser toutes ou une partie de ces souches par NGS (Next Generation Sequencing);
- 4. Décrire les caractéristiques génomiques des souches analysées et rechercher des clusters (groupe de souches présentant une forte parenté génomique) par une analyse WGS;
- 5. Interpréter la présence de ces clusters à l'aide des métadonnées.

Ces étapes ont été regroupées en 3 phases principales, échelonnées ainsi :

- Une première Phase (Phase I), entre les mois d'Avril et Août 2018, devant aboutir à la sélection des souches de *S*. Dublin à séquencer ;
- Une seconde phase (Phase II), entre septembre et décembre 2018, d'obtention des données NGS des souches récoltées;

- Une troisième phase (phase III), entre janvier 2019 et février 2020, d'analyse des données NGS par méthode WGS afin de définir des clusters génomiques et les interpréter à l'aide des métadonnées. D'un point de vue organisationnel, ces différentes étapes ont été menées sous la responsabilité d'un chef de projet engagé pour coordonner cette étude, via un financement conjoint entre la DGAI (Direction Générale de l'Alimentation) et le Centre National Interprofessionnel de l'Economie Laitière (Cniel). Ce chef de projet a participé à la réalisation des différentes étapes décrites ci-dessus et également assuré la coordination globale du projet. En particulier, pour faciliter le déroulement, le suivi et l'appropriation des résultats de cette étude, un comité de suivi (ou groupe projet) spécifique a été créé. Ce comité de suivi était constitué :
- de représentants francs-comtois des différents maillons de la filière laitière locale, depuis le domaine animal et vétérinaire (représentants du GDS, de la SNGTV et du GTV Bourgogne – Franche-Comté) jusqu'au domaine fromager (URFAC, CTFC), sans oublier les laboratoires locaux (du secteur animal (LVD) et agro-alimentaire (LDA));
- de scientifiques de l'ANSES (Laboratoire de Sécurité des Aliments (LSAI) Unité GAMER (Genome Analysis, Modelling and Risk)- Maisons Alfort);
- du Cniel;
- et des membres opérationnels de la PTF SCA.

Constitué au plus d'une quinzaine de personnes, ce comité de suivi a été régulièrement informé de l'avancée des travaux par le chef de projet. Ce comité avait également pour rôle d'orienter ceux-ci lorsque nécessaire. Les membres francs-comtois de ce comité de suivi, de par leur connaissance du terrain d'étude et des pratiques locales, ont notamment fortement contribué et participé à l'interprétation des clusters à l'aide des métadonnées disponibles. L'ensemble du comité de suivi a eu aussi à cœur de préciser les enseignements méthodologiques d'une telle étude.

III. Une étude menée en 3 phases

1) Détermination du panel de souches de l'étude (Phase I)

Cette première phase du projet d'étude rétrospective a été scindée en plusieurs étapes.

Recensement des souches de S. Dublin

Collecte des métadonnées et anonymisation

Échantillonnage et sélection des souches En se basant sur les métadonnées recensées et sur les acteurs régionaux

Préparation et transfert des souches Après sélection, transfert des souches au laboratoire de l'ANSES

Recensement et collecte des métadonnées des souches de S. Dublin

Plus de 2 500 isolats de *Salmonella* Dublin, issus de prélèvements réalisés entre 2015 et 2017 ont été recensés via les 4 laboratoires d'analyse des échantillons de la filière fromagère au lait cru de Franche-Comté (Tableau 1).

1	LDA39	LIAL	LVD	ACTALIA 25
Années	2015-2016-2017	2016-2017	2015-2016- 2017	2017
Nombre de souches	1314	381	410	397
Maillon	SA + HA	НА	SA	SA + HA
Disponibilité des métadonnées	Minimales - OK	Minimales - OK	Minimales - OK	Minimales – OK
	Complètes - OK	Complètes - Partielles	Complètes - OK	Complètes - Partielles

SA = Santé Animale, HA = Hygiène Alimentaire

<u>Tableau 1</u>: recensement des souches de *S.* Dublin et des métadonnées associées, isolées entre 2015 et 2017 dans les 4 laboratoires de Franche-Comté de la filière laitière.

Les métadonnées de ces souches ont été également collectées. Elles sont composées de métadonnées minimales et complètes, de niveau d'information (Tableau 2) et de confidentialité différente. Les métadonnées minimales ont été fournies au chef de projet coordonnant l'étude. Les métadonnées complètes ne sont détenues le plus souvent que par le maillon professionnel.

Nature des métadonnées				
Minimales	Complètes			
 ✓ N° d'identité souche du détenteur initial (origine isolement) ✓ N° d'attribution de la souche par le laboratoire ✓ Identité du labo ayant fourni les souches ✓ Date de prélèvement ✓ Type de matrice (ex. : lait, lisier, fromage, etc.) ✓ Maillon prélevé (ex. : élevage, atelier, etc.) ✓ Département ✓ Niveau de surveillance (routine, renforcée, investigation) 	 = métadonnées minimales + ✓ Nature de l'échantillon : (version plus détaillée) • Environnement d'élevage • Animal • Type d'aliment (exemple : produit laitier) • Nature de l'aliment (type de fromage avec date de fabrication) ✓ Ville ou code postal ✓ Données de contextes associés aux prélèvements, ces données appartiennent à l'opérateur (ex. : le N° de l'élevage dans l'atelier laitier) 			

Tableau 2 : Nature des métadonnées minimales et complètes associées aux isolats de S. Dublin

A réception des métadonnées transmises par les 4 laboratoires, une harmonisation en a été nécessaire avant d'effectuer l'échantillonnage. Auparavant, 56 souches de *S.* Dublin déjà séquencées par le LSAL et issues de ces laboratoires, ont été retirées de la liste de souches pour l'échantillonnage. Leurs génomes ont été pris en compte lors de la recherche des clusters (Phase III). L'ensemble des 2 501

isolats intégrables au projet ont été anonymisés, le code étant détenu uniquement par le chef de projet chargé de cette étude.

Échantillonnage et sélection des souches

L'échantillonnage était un processus clé du projet. Deux approches complémentaires ont été proposées et retenues par le comité de suivi pour sélectionner le panel de souches de l'étude :

- Une approche aléatoire, basée sur les métadonnées minimales, de manière à maximiser la représentation des différents maillons de la chaîne de production et avec pour objectif de décrire la circulation clonale des souches;
- Une approche ciblée, de manière à incorporer des souches sélectionnées par les acteurs locaux et avec pour objectif de répondre à des questions spécifiques, telles que la recherche de flux sur des zones limitées.

Les principes, intérêt et résultats de ces deux types de sélection sont illustrés ci-dessous. L'annexe 1 détaille la méthode de sélection des 500 souches mise en œuvre.

Sélection aléatoire représentative (398 souches)

Intérêt : Permettre d'étudier la diversité génomique des souches circulantes de S. Dublin entre 2015 - 2017

Sélection ciblée (104 souches)



Principe - sélection aléatoire de 398 souches

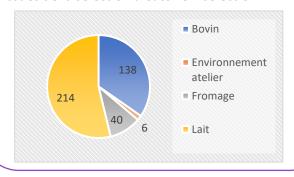
L'échantillonnage aléatoire a été réalisé par l'approche de Gower basé sur le calcul des distances (Gower, 1971). Cet algorithme, utilisé en écologie, calcule la dissimilarité entre les éléments d'une collection en se basant sur des données logiques, numériques, catégoriques et/ou de textes. En se basant sur cet algorithme, on obtient « k » groupes de souches à partir duquel on peut déterminer le nombre optimal de groupes pour étudier la diversité génomique de S. Dublin. Ces éléments sont visualisables sur le logiciel R avec les lignes de programme Elbow et Silouhette. On peut ainsi choisir n souches aléatoirement au sein des groupes de souches. L'approche permet au final d'obtenir une dissimilarité maximale entre les groupes et minimale au sein des groupes.

Principe - sélection ciblée de 104 souches

La sélection a été proposée par le CTFC, le LDA39 le GDS, le GTV et le LVD25 afin de fournir des cas communs en santé animale et en hygiène alimentaire pour répondre à des problématiques de terrain. Au total, 105 souches ont été proposées, parmi lesquelles 36 souches correspondent à l'étude de deux cas de contamination de coopératives, principalement entre 2016 et 2017, dans le Jura et le Doubs. Les 69 autres souches (sélection fournie par le LDA39) regroupaient des cas détectés en SA et HA de 2017. Un tri a été effectué par les opérateurs locaux pour étudier les cas les plus intéressants (par coopérative, aires géographiques-proximité des pâturages : lien environnemental. éventuel commerce entre élevages, estives...)



Résultat - Matrice d'origine des 398 souches issues de la sélection aléatoire - sélection B



Résultat - sélection ciblée de 4 cas

Le cas A concernait la récurrence d'isolement de S. Dublin au sein d'un même élevage, avec des manifestations cliniques particulièrement virulentes pour les bovins. Les cas C et E concernaient une hypothétique diffusion de salmonelles entre élevages et coopératives. Le cas D a été choisi pour étudier la diversité clonale de S. Dublin sur deux zones géographiques distinctes et s'intéresser à sa persistance et ses flux entre des élevages plus ou moins proches de bassins versants.

Schéma 1 : Représentation des deux voies de sélection des souches de l'étude d'épidémiosurveillance génomique rétrospective S. Dublin en Franche-Comté.

Préparation et transfert des souches de S. Dublin

Les identifiants des souches de S. Dublin ainsi sélectionnées ont été transmis aux laboratoires de Franche-Comté. Après mise en gélose longue conservation pour le transport, celles-ci ont été expédiées par les 4 laboratoires concernés vers le LSAL pour l'extraction de leur ADN.

2) Séquençage des 500 génomes sélectionnés (Phase II)

Cette phase a débuté par la réception effective des souches au LSAI, et s'est poursuivie par la réalisation des extractions des ADN et des contrôles qualité liés (selon les méthodes internes du LSAI). Ces extraits d'ADN ont été envoyés à un prestataire pour séquençage NGS (ICM - Unité Pitié Salpêtrière- Paris, technologie Illumina utilisée). Le synoptique de cette phase est présenté ci-dessous (schéma 2).

Récupération des souches et remise en culture

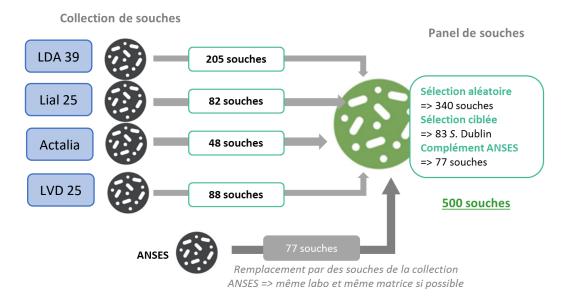


Schéma 2 : Synoptique des souches reçues pour l'étude

Extraction selon une méthode semi-automatique de l'ADN des souches et vérification de la qualité

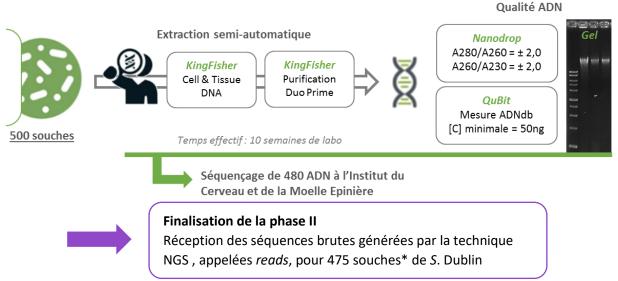


Schéma 3 : Étape d'extraction et de contrôle qualité des ADN des souches reçues pour l'étude

* les séquences de 5 souches n'ont pas pu être obtenues (suite à un criblage qualité des séquences brutes).

3) Apport des données exploitées et enseignements (Phase III)

La phase III du projet consiste en l'assemblage des séquences génomiques, leur comparaison et la réalisation d'arbres phylogénétiques. Deux arbres phylogénétiques ont été construits afin de mieux comprendre la similitude entre les souches de *Salmonella* Dublin en Franche-Comté et leur répartition selon différentes caractéristiques (métadonnées) :

- Un premier arbre, dit Arbre A, qui correspond aux souches retenues pour l'étude (sélection aléatoire (sélection B) et sélections ciblées (A, C, D et E) ;
- Un second arbre, dit Arbre B, qui correspond aux souches de l'arbre A, complété par des souches d'investigations de l'ANSES et des souches de cas humains, séquencées à l'Institut Pasteur en 2016 et disponibles sous la base de données Enterobase (code projet : PRJEB28817).

Traitement des données génomiques

À partir des *reads* bruts obtenus, un assemblage a été réalisé afin de vérifier la qualité des échantillons. Cet assemblage a été conçu avec ARTwork, un outil développé par l'équipe GAMER (ANSES) qui permet de construire un assemblage *de novo* à partir de *reads* bruts de séquençage.

Afin d'obtenir un résultat précis, les échantillons ont été sélectionnés selon certains critères de qualité. Pour cela, plusieurs éléments ont été pris en compte :

- La qualité de séquençage, notamment en regardant si l'assemblage est trop partitionné, si le nombre de paires de bases qui ne s'alignent pas sur la référence *Salmonella* Dublin est trop élevé et si le nombre d'indels (insertion et délétion) pour 100 paires de bases par rapport à la référence est trop grand. Si c'est le cas, alors l'assemblage est mauvais ou l'assemblage ne

corresponds pas à une Salmonella Dublin. Ces vérifications ont été réalisées grâce aux outils bio-informatiques QUAST (https://doi.org/10.1093/bioinformatics/btt086) et samtools (https://doi.org/10.1093/gigascience/giab008).

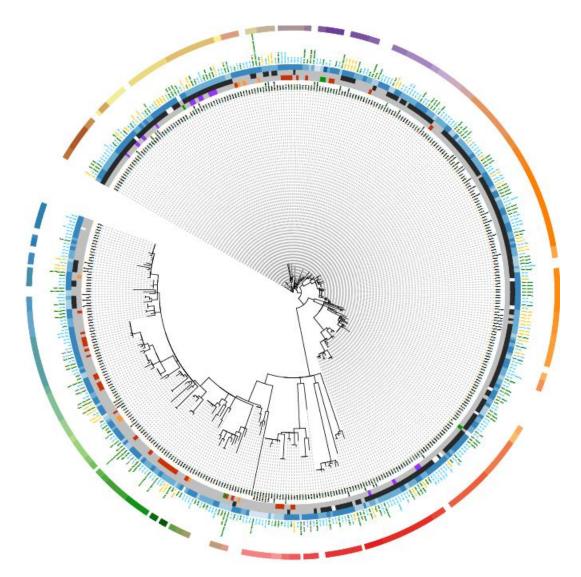
- La contamination des échantillons avec Confindr (https://doi.org/10.7717/peerj.6995), qui permet de vérifier s'il y a eu des contaminations (de la même espèce bactérienne ou non), et en regardant le nombre de bases non alignées sur la référence. Un échantillon d'Escherichia et un échantillon d'Enterobacter ont été trouvés parmi les échantillons.
- Le sérotypage, en vérifiant que l'échantillon était bien une Salmonella Dublin. Cela a été réalisé avec SeqSero (DOI: 10.1128/JCM.00323-15).

Les SNPs provenant d'événements de recombinaison homologue ont été supprimés de l'analyse. Ces critères de qualité permettent d'analyser les génomes sur un maximum de positions (environ 3,8 M positions pour l'arbre B) et d'avoir une analyse de SNPs robuste. Au final, après ces contrôles et compte-tenu des critères de qualité pris en compte, 441 séquences ont permis de construire l'Arbre A. Pour l'Arbre B, 480 séquences furent utilisées.

La phylogénie des isolats a été construite avec les outils iVARCall2 (Felten et al., 2017) et IQ-TREE (Nguyen et al., 2015). iVARCall2 est un outil développé par l'équipe GAMER qui permet d'identifier les SNPs du core-génome à partir des reads d'échantillons et d'une séquence de référence. La séquence de référence qui a été sélectionnée pour cette étude est Dublin_CT_02021853 (accession no. CP001144). iVARCall2 nous permet d'obtenir un alignement des SNPs retrouvés dans les différents échantillons, qui est ensuite utilisé pour construire l'arbre phylogénétique (Radomski et al., 2019, Vila Nova et al., 2019).

Arbre phylogénétique obtenu à partir des séquences d'ADN des 441 souches de S. Dublin, sur la base de l'analyse SNP du core-genome : Arbre A.

Afin de visualiser ces données, l'arbre phylogénétique A (Figure 1) a été découpé en clusters avec la librairie rPicone de R. Ce programme permet le découpage d'un arbre phylogénétique en clusters en se basant sur un nombre défini de SNPs. Pour l'étude, le seuil retenu était de 5 SNPs, c'est-à-dire que les génomes regroupés dans un même cluster ne pouvaient pas avoir plus de 5 SNPs de différence entre eux (par comparaison 2 à 2), c'est à dire 5 nucléotides de différence maximum sur la totalité de la séquence génétique comparée. Ce nombre de SNPs a été choisi au vu de la littérature scientifique. C'est un seuil consensuel, habituellement utilisé pour établir des clusters génétiques lors d'études de type « source - attribution » en santé humaine (Ung et al., 2019). Ce seuil de 5 SNPs est suffisamment bas pour éviter d'établir à tort des liens entre souches non reliées génétiquement.



<u>Figure 1</u>: Arbre phylogénétique A montrant les homologies génétiques entre 441 souches de S. Dublin, d'origine non humaines, issues de divers maillons de la chaîne de transformation fromagère (animal, environnement de l'élevage, transformation, fromage) isolées dans les départements du Jura et du Doubs entre 2015 et 2017. Arbre A présenté avec les longueurs de branches, inféré avec IQ-TREE. 1^{er} cercle interne : les sélections (SELEC B = sélection aléatoire) ; 2^{ème} cercle : les départements de prélèvements des échantillons ; 3^{ème} cercle : les années de prélèvements des échantillons ; 4^{ème} cercle : matrice des aliments (textes colorés) ; cercle extérieur : clusters des échantillons (72 clusters, les échantillons n'ayant pas de couleurs sont considérés comme des singletons)

Un grand nombre de clusters (72) ont été distingués par cette analyse. Ils montrent la diversité génétique des souches de *S*. Dublin sur la période considérée, en se basant sur l'ensemble du coregénome, c'est-à-dire la partie du génome commun à l'ensemble des souches étudiées (environ 4,2 M pb couverts à plus de 30X). Les clusters sont de tailles différentes (Annexe 2), le plus représenté d'entre eux regroupe 23 génomes et vingt clusters se composent chacun uniquement de deux génomes. 60 souches sont des singletons, c'est-à-dire qu'elles n'appartiennent à aucun des clusters. Au sein des clusters, des souches peuvent être identiques d'un point de vue génétique (ceci se visualise par l'absence de branches entre 2 souches).

La lecture des distances entre les souches d'un même cluster est expliquée sur l'exemple du cluster 32 (Figure 2). Les souches 2SD0144 et 2SD0170 ne présentent aucune branche entre-elles (ce que l'on appelle un « râteau »), et n'ont donc aucune différence génétique. 4SD0237 a une longueur de branche

unique, ce qui veut dire qu'elle présente une différence en plus par rapport aux souches 2SD0144 et 2SD0170. Si nous regardons plus globalement au niveau du cluster, 2SD0144, 2SD0170, 4SD0237 et 2SD0097 sont plus proches entres elles qu'avec toutes les autres souches du cluster. Cela s'illustre par une branche plus longue entre ces 4 souches et le reste des autres souches. Par ailleurs, les clusters sont définis avec un seuil de 5 SNPs, donc les différences entres les souches d'un même cluster ne peuvent pas excéder ce nombre.

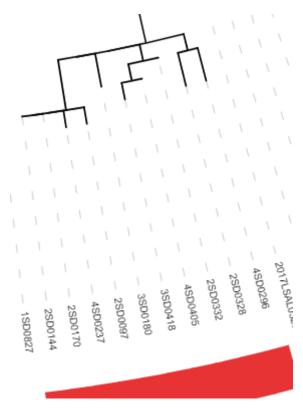


Figure 2 : Exemple de lecture de l'arbre avec le cluster 32

Interprétation de l'Arbre A à l'aide des métadonnées minimales

Un travail d'interprétation de l'Arbre A a été réalisé en utilisant les données minimales : les principaux enseignements sont présentés dans le Tableau 3.

Métadonnée minimale	Précaution sur la métadonnée	Interprétation de l'Arbre A
Année	Attention, la date associée à une souche correspond à sa date de prélèvement : cette date n'indique en rien la temporalité de la souche (en d'autres termes, elle a pu être isolée l'année N en un lieu mais pouvait s'y trouver déjà depuis plusieurs années sans avoir fait l'objet d'un prélèvement).	Une persistance des clusters est observée dans le temps : plusieurs facteurs explicatifs sont possibles : 1/ portage animal : portage passif, latent de ce sérovar que l'on sait bien adapté au bovin, 2/ persistance dans un écosystème global dont fait partie l'animal (par écosystème on entend : l'élevage avec le fumier, lisier, le secteur géographique avec un bassin versant, le voisinage) Un travail d'investigation sur une zone précise et sur une échelle de temps conséquente permettrait d'étayer les facteurs explicatifs de cette persistance. A consolider par un travail sur plusieurs zones précises. Il n'y a pas de séparation des clusters selon l'année.
Département	L'information département ne permet pas une localisation précise, une information plus précise (ex. : code postal) expliquerait certaines exceptions observées	Il existe une distinction entre les souches issues du département du Jura et du Doubs qui se trouvent dans des parties distinctes de l'arbre, avec toutefois des exceptions. Ces exceptions ne peuvent pas être expliquées sans l'information code postal.
Type de matrice	Une dizaine de données sont manquantes	Il n'existe pas de séparation des clusters selon les matrices Cela signifie que des souches de <i>S.</i> Dublin apparentées génétiquement peuvent être présentes chez l'animal et dans le lait, puis aller du lait au fromage. Cette observation est faite pour 16 des 72 clusters de la sélection A contenant plus d'une souche. (Annexe 2)

Tableau 3: Points de vigilance et conclusions tirées de l'interprétation de l'Arbre A avec les métadonnées minimales

Concernant les sélections ciblées des souches (A, C, D, E) représentant un total de 69 souches, elles ont été construites sur la base d'hypothèses et scénarios envisagés par les acteurs de terrain au regard des données écologiques (non analytiques) :

- Les souches issues de ces 4 sélections qui représentent des souches issues de cas d'investigation par les opérateurs locaux se répartissent à différents niveaux de l'arbre.
- Au sein de chaque sélection ciblée, les souches appartiennent à des clusters différents pouvant être assez éloignés (ex pour la sélection C (violet) ou E (orange)).
- Les souches de ces cas présentent des parentés génétiques avec d'autres souches de la sélection aléatoire. Le manque d'information exacte sur la nature de ces souches, dont leur contexte de prélèvement, ne permet pas d'émettre des hypothèses sur les liens écologiques (voies de circulation) entre les souches.

Informations apportées par la construction de l'Arbre B (arbre phylogénétique constitué des 441 souches de l'Arbre A et de 31 souches d'origine humaine)

Cet arbre B correspond à l'arbre A auquel 31 souches d'origine humaine, isolées en 2016 et 2017, ont été intégrées. L'ajout de ces 31 souches modifie légèrement la définition des clusters identifiés pour l'arbre A : 63 clusters, regroupant chacun des souches ne présentant pas plus de 5 nucléotides de différence sur leur core-génome, ont été identifiés (arbre non illustré).

Pour cet arbre B, 4 souches d'origine humaine sont également des singletons, c'est-à-dire qu'elles ont plus de 5 nucléotides de différence avec les séquences génétiques des autres souches.

Les souches humaines s'intègrent à divers endroits de l'arbre, indiquant des contaminations humaines multi-clusters, donc probablement multi-sources. Certaines souches d'origine humaine appartiennent au même cluster que des souches non humaines. Cependant, cette appartenance ne suffit pas pour établir que les souches non humaines sont à l'origine des cas humains. Il est nécessaire d'avoir estimé la diversité génétique sur la base d'un panel jugé suffisamment représentatif (ce qui est le cas de la présente étude au regard du nombre de souches séquencées et de la méthode de sélection des souches) et de prendre en compte les données d'enquêtes épidémiologiques (enquêtes alimentaires) pour juger de la probabilité d'existence de tels liens (ECDC, 2018, Ford et al., 2018, Ung et al, 2019, Vaughn et al., 2020).

Interprétation des cas ciblés de l'arbre A avec les métadonnées complètes

Un travail d'interprétation des cas ciblés avec les métadonnées complètes a été mis en œuvre par les acteurs locaux de Franche-Comté. Ceci a nécessité :

- Une organisation particulière de leur part, nécessitant le regroupement de tous les acteurs de la chaîne de production et de transformation (des vétérinaires d'élevage aux techniciens des fromageries, sans oublier les interlocuteurs des laboratoires d'analyses) au sein d'un groupe de travail local. Cette organisation dans un esprit « one health » avait pour objectif d'utiliser leurs connaissances de terrain, chaque spécialiste d'un maillon pouvant apporter et/ou complémenter les données et connaissances des autres spécialistes tout au long de la chaîne de production.
- La désignation d'un coordinateur de ces acteurs locaux pour mettre en œuvre ce travail d'analyse et son interprétation (recensement des métadonnées sur les souches de chacun des cas, organisation des réunions de travail et compte-rendu). Cette tâche a été menée par un agent du CTFC.

Le travail d'analyses des différents cas ciblés, mené a posteriori du travail initial d'investigations en utilisant la donnée nouvelle « appartenance des souches à un cluster cg-SNP » a eu pour conséquences de :

- Apprécier la diversité génomique des souches appartenant aux différents cas ciblés. Ces cas, considérés comme groupés sur la base d'une récurrence dans le temps et/ou sur l'isolement au sein d'une zone géographique restreinte, se subdivisent généralement dans plusieurs clusters génomiques, ces clusters pouvant être éloignés sur l'arbre phylogénétique.
- Attribuer l'origine exacte de contamination de produits finis par l'appartenance à un même cluster de souches isolées du maillon amont (l'élevage) et du lait mis en œuvre ou des

fromages fabriqués à une même date ou une date proche de celle de leur première mise en évidence.

- Ou, au contraire, infirmer l'attribution de l'origine de contamination entre un échantillon amont (issu de l'élevage) et aval (fromage), information basée sur une concomitance de date d'isolement de souches.
- Demander un travail complémentaire de recensement des métadonnées de souches n'ayant pas au départ été identifiées comme liées aux cas ciblés mais révélées ensuite comme appartenant à l'un des clusters dans lesquels se répartissaient les souches du cas ciblé étudié. L'obtention de ces métadonnées, a posteriori, s'avère chronophage et quelquefois non possible (conservation des informations détaillées, turn-over des opérateurs...).

A terme, l'analyse de l'ensemble de ces éléments peut permettre aux acteurs de comprendre plus précisément l'origine des contaminations en S. Dublin et la diffusion de souches sur le terrain, grâce à leurs connaissances conjointes des pratiques de l'amont à l'aval.

Par ailleurs, un travail de localisation géographique des souches des différents clusters a été mené sur la base de l'information du code postal (élément des données complètes détenu par le groupe d'acteurs locaux). Cette localisation géographique montre que la plupart des clusters se situent sur un périmètre géographique relativement restreint, ce qui ouvre des perspectives en termes de surveillance ou de gain en précision dans l'attribution d'une origine de contamination.

IV. Conclusions de l'étude épidémiologique génomique rétrospective de Salmonella Dublin menée en Franche-Comté

Le travail mené au sein de l'axe II « Étude épidémiologique génomique rétrospective de S. Dublin en Franche-Comté » du GT « Surveillance des salmonelles en filière laitière bovine » de la Plateforme de Surveillance de la chaîne alimentaire, peut être considéré comme un travail conséquent, notamment d'un point de vue :

- Humain, via l'implication conjointe et permanente des représentants des divers maillons de la production laitière et de la transformation fromagère de la zone d'étude au côté d'acteurs nationaux techniques, interprofessionnels ou scientifiques;
- Technique, par le nombre important de souches séquencées issues de la période et zone d'étude ayant nécessité au préalable un important travail de recensement des informations disponibles par l'ensemble des acteurs ;
- Financier, par les moyens mis en œuvre pour la coordination et la réalisation des différentes phases de ce travail, que ce soit par les financeurs directs de cette étude ou par l'investissement humain des différents partenaires de cette étude.

Ce travail était le premier de ce type mené dans cette configuration en France. Il a permis :

D'obtenir une vision globale de la diversité génomique des souches de Salmonella Dublin en filière laitière franc-comtoise isolées principalement entre 2015 et 2016. Ce travail a été obtenu par l'obtention et la valorisation des données de WGS de 441 souches, représentatives de la surveillance des différents maillons de la chaîne de production et de transformation laitière des deux départements du Jura et du Doubs sur la période considérée.

- De montrer la plus-value de ces données pour confirmer ou infirmer la proximité génomique entre les souches issues d'une situation de terrain (gain en précision) et pour une meilleure compréhension des voies de circulation des souches (gain en efficacité), notamment à travers le travail collaboratif mené sur les cas ciblés ;
- D'insister sur le besoin de collecter et de conserver les métadonnées (ou données écologiques) liées aux prélèvements et souches isolées, ces données devant être les plus complètes et les plus uniformes possibles pour faciliter l'identification de voies de dissémination des souches ;
- De souligner la force du collectif amont-aval dans ce travail d'interprétation où les connaissances des pratiques de chaque maillon de la chaîne de production et de transformation sont importantes à considérer, pour mieux comprendre les facteurs de dissémination des souches.

Des enseignements méthodologiques et des perspectives peuvent également être dégagés de cette étude : ils sont présentés dans le schéma ci-dessous.

Points forts pour mener une étude de ce type

Coordination: désignation d'un chef de projet

Groupe d'acteurs amont-aval, pluridisciplinaire : représentants des divers maillons de la chaîne de production et de transformation, et des laboratoires d'analyses, connaissant les pratiques locales; appui scientifique pour la constitution du jeu de données de souches, pour l'analyse et pour l'interprétation des données génomiques.

Disposer d'une collection de souches représentatives de la situation étudiée (zone, période de temps, origines...).

Disposer des métadonnées précises associées aux souches, renseignées selon une même structure (rubriques) dès le prélèvement de l'échantillon.

Points faibles – points de vigilance

Coût financier : du typage WGS

Délai de l'obtention de l'information WGS : non compatible avec une intervention d'urgence sur site (obtention des séquences brutes et traitement de l'information).

Interprétation des clusters :

Connaissance scientifique pour l'interprétation fine des données NGS.

Besoin d'associer l'information sur la proximité génomique des souches avec les métadonnées de ces souches pour étayer des voies de dissémination des souches.

Généralement a posteriori de la phase d'alerte pour des raisons de délai.

Perspectives

Connaissance de la diversité génomique globale des souches de S. Dublin isolées dans ce contexte : pourrait être valorisée pour gagner en précision et en rapidité sur la localisation potentielle de souches nouvelles isolées.

Mise à disposition de moyens pour enrichir l'arbre A par des souches nouvellement isolées : ceci nécessite que l'interprétation en routine des données WGS des nouvelles souches se fasse par cg-SNP et que ces souches puissent être intégrées dans l'arbre A. Ceci pose la question de l'organisation et des moyens nécessaires à une telle mise en place.

Valorisation de l'Arbre A par les acteurs locaux : le travail de recensement des métadonnées, d'analyse et d'interprétation des différents clusters pourra se poursuivre, sous réserve de disponibilité des acteurs locaux.

Références citées

Ågren E., Wahlström H., Vesterlund-Carlson C., Lahti E., Melin L., Söderlund R., 2016. Comparison of whole genome sequencing typing results and epidemiological contact information from outbreaks of Salmonella Dublin in Swedish cattle herds Infection Ecology and Epidemiology 2016, 6: 31782 http://dx.doi.org/10.3402/iee.v6.31782

Crump J., Sjölund-Karlsson M., Crump M., Gordon A., Parry C. 2015. Epidemiology, clinical presentation, laboratory diagnosis, antimicrobial resistance, and antimicrobial management of invasive Salmonella Infections. Clin Microbiol Rev doi:10.1128/CMR.00002-15.

European Centre for Disease Prevention and Control. Monitoring the use of whole-genome sequencing in infectious disease surveillance in Europe. Stockholm: ECDC; 2018. Doi: 10.2900/037665

Felten A., Vila Nova M., Durimel K., Mistou MY., Radomski N. 2017. First gene-ontology enrichment analysis based on bacterial coregenome variants: insights into adaptations of Salmonella serovars to mammalian- and avian-hosts. BMC Microbiol: 17(1):222. doi: 10.1186/s12866-017-1132-1

Ford L., Carter G., Wang Q., Seemann T., Sintchenko V., Glass K., Williamson D., Howard P., Valcanis M., Sotomayer Castillo C., Howden B., Kirk M. 2018. Incorporating Whole-Genome Sequencing into Public Health Surveillance: Lessons from Prospective Sequencing of Salmonella Typhimurium in Australia. Foodborne Pathogens and Disease: 15, DOI: 10.1089/fpd.2017.2352

Gower JC. 1971. A general coefficient of similarity and some of its properties. *Biometrics*, 27, 857-874.

Institut Pasteur et CHU Robert-Debré APHP. 2019. Rapport d'activité annuel 2019 Année d'exercice 2018. Centre National de Référence des Escherichia coli, Shigella et Salmonella. 165 pages

Leclerc V., Moury F., Morel V., Noel V., Lailler R. 2019. Le réseau Salmonella, un dispositif de surveillance des salmonelles de la fourche à la fourchette : bilan 2016. In : Bulletin épidémiologique, santé animale et alimentation n°89 – Numéro spécial SSA (12) – Décembre 2019, 8 pages.

Nguyen L., Schmidt H., von Haeseler A., Minh B. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol: 32:268-74. https://doi.org/10.1093/molbev/msu300

Radomski N., Cadel-Six S., Cherchame E., Felten A., Barbet P., Palma F., Mallet L., Le Hello S., Weill F-X., Guillier L., Mistou M-Y. 2019. A Simple and Robust Statistical Method to Define Genetic Relatedness of Samples Related to Outbreaks at the Genomic Scale – Application to Retrospective Salmonella Foodborne Outbreak Investigations. Front. Microbiol. 10:2413. https://doi.org/10.3389/fmicb.2019.02413

Ung A., Baidjoe A., Van Cauteren D., Fawal N., Fabre La., Guerrisi C., Danis K., Morand A., Donguy MP., Lucas E., Rossignol L., Lefèvre S., Vignaud ML., Cadel-Six S., Lailler Re., Jourdan-Da Silva N., Le Hello S. 2019 Disentangling a complex nationwide Salmonella Dublin outbreak associated with raw-milk cheese consumption, France, 2015 to 2016. Euro Surveill: 24(3):pii=1700703.

Santé Publique France. 2021. Surveillance des toxi-infections alimentaires collectives. Données de la déclaration obligatoire, 2019. Le point épidémio. Mars 2021, 12 pages.

https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-infectieuses-d-originealimentaire/toxi-infections-alimentaires-collectives/documents/bulletin-national/surveillance-destoxi-infections-alimentaires-collectives.-donnees-de-la-declaration-obligatoire-2019

Vaughn E., Vo Q., Vostok J., Stiles T., Lang A., Brown C., Klevens L., Madoff L. 2018. Linking Epidemiology and Whole-Genome Sequencing to Investigate Salmonella Outbreak, Massachusetts, USA, 2018. Emerg Infect Dis. 2020; 26(7):1538-1541. https://doi.org/10.3201/eid2607.200048

Vila Nova, M., Durimel, K., La, K., Felten A., Bessières P., Mistou MY., Mariadassou M., Radomski N. 2019. Genetic and metabolic signatures of Salmonella enterica subsp. enterica associated with animal sources at the pangenomic scale. BMC Genomics 20, 814 (2019). https://doi.org/10.1186/s12864-019-6188-x

ANNEXE 1 : Méthodologie et processus de sélection aléatoire des souches de S. Dublin

PANEL de Souches

Algorithme de Gower

Validation Elbow et Silouhette

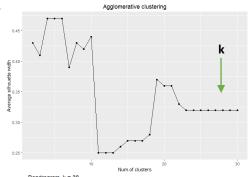
Observation de la répartition et pondération

Sélection du panel de 400 souches

SELECTION Aléatoire de 400 souches - Résultats obtenus

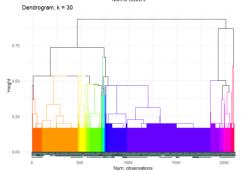
Panel de 2 101 souches se focalisant sur la surveillance de routine si possible Logiciel Rstudio Algorithme de Gower

La courbe ci-contre permet de déterminer le nombre optimal de groupes de souches (k). Ce nombre se situe à la fin du plateau symbolisée par une flèche. **Pour notre panel : k=30.**



Répartition des souches au sein des 30 groupes

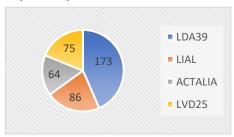
Le dendrogramme (fonction Elbow) ci-contre illustre la répartition non homogène des 2 101 souches au sein des 30 groupes. Afin de sélectionner 400 souches pour l'étude, il a été nécessaire de pondérer selon cette répartition.



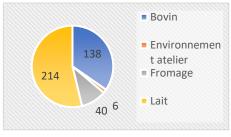
Sélection aléatoire des souches au sein des groupes

Les 2 101 souches sont réparties dans 30 groupes (tableau ci-contre). Les deux clusters les plus importants n°3 et n°12 correspondent à des prélèvements de placenta et de lait respectivement. La sélection est effectuée grâce à la fonction « alea() » en pondérant suivant la répartition (ex. : groupe n°12 = 41 % du panel = sélection de 165 souches). Les groupes représentant moins de 0,05% du panel sont exclus.

Répartition par labos



Répartition par matrice



	0 500	Num. observations	1500 2000
Groupes	Souches	Pourcentage	Sélection de
			400 souches
1	5	0,24%	1
2	64	3,05%	12
3	372	17,71%	71
4	31	1,48%	6
5	13	0,62%	2
6	3	0,14%	1
7	101	4,81%	19
8	197	9,38%	38
9	12	0,57%	2
10	1	0,05%	0
11	97	4,62%	18
12	868	41,31%	165
13	25	1,19%	5
14	6	0,29%	1
15	6	0,29%	1
16	1	0,05%	0
17	15	0,71%	3
18	1	0,05%	0
19	4	0,19%	1
20	69	3,28%	13
21	3	0,14%	1
22	123	5,85%	23
23	1	0,05%	0
24	5	0,24%	1
25	44	2,09%	8
26	5	0,24%	1
27	25	1,19%	5
28	1	0,05%	0
29	2	0,10%	0
30	1	0,05%	0

ANNEXE 2 : Description des 72 clusters de l'Arbre A selon leur maillon d'origine

Cluster n°	Bovin ou	Lait ou filtre à lait	Fromage ou	Nombre total de
	environnement d'élevage		environnement de transformation	souches du cluster
1	2	4	1	7
2	2			2
3	2			2
4	4	3		7
5	1	3		4
6		2	2	4
7		2		2
8		7		7
9	2	3		5
10	3	10	1	14
11	2			2
12	4	3	1	8
13	1	1	3	5
14	1		1	2
15	4	1		5
16		2		2
17	2			2
18	2			2
19	1	1		2
20		3		3
21	2	1	1	4
22	1	1		2
23	1	1		2
24		2		2
25	3			3
26	1	1		2
27	3			3
28	2	6		8
29		2		2
30	1	1	1	3
31		2	2	4
32	3	7		10
33	4	1		5
34	4	3		3
35	4	7		11
36	1	2		3 2
37	2	0	2	
38	4	9	2	15
39	3	10	1	4
40	2	10	1	13
41	2	2		2
42	2	1		3

Cluster n°	Bovin ou environnement d'élevage	Lait ou filtre à lait	Fromage ou environnement de transformation	Nombre total de souches du cluster
43	2	3	4	9
44		1	5	6
45	6	5	1	12
46	9	9	5	23
47		5	1	6
48	3		1	4
49	5	8	1	14
50	2			2
51		4		4
52	2	2		4
53	3	1		4
54		2		2
55	1	5	2	8
56		4		4
57	1	2		3
58	1		4	5
59	3	2		5
60		2		2
61	2	3		5
62	1	6		7
63		2	3	5
64	3			3
65	1	4	2	7
66	2			2
67	4	4	4	12
68	7	4	3	14
69	1	1	1	3
70	1	2		3
71	2	1		3
72	1	6		7
no cluster	32	22	6	60
Total	167	214	60	441

no cluster : souches n'appartenant à aucun cluster (singleton)

