

PDS-183020

Automatiser la classification des souches pour faciliter l'interprétation du
séquençage et aider à prioriser le contrôle du syndrome reproducteur et
respiratoire porcin

Rapport final soumis au:

Ministère de l'Agriculture, des Pêcheries et de l'Alimentation du Québec (MAPAQ)

Par:

Marie-Ève Lambert, D.M.V., Ph.D.

et

Sylvie D'Allaire, D.M.V., M.Sc., Ph.D.

Laboratoire d'épidémiologie et de médecine porcine (LEMP)

Département de sciences cliniques

Faculté de médecine vétérinaire (FMV)

Université de Montréal

Le 28 janvier 2021

© FMV-Université de Montréal, 2021

Table des matières

TABLE DES MATIÈRES _____	2
LISTE DES FIGURES _____	3
ABRÉVIATIONS _____	4
1. RAPPEL DU PROJET INITIAL _____	5
1.1 MISE EN CONTEXTE _____	5
1.2 OBJECTIFS _____	6
1.3 ÉQUIPE DE RÉALISATION _____	7
1.4 PARTENAIRES FINANCIERS _____	8
1.5 DURÉE DU PROJET _____	8
2. DÉROULEMENT DU PROJET _____	9
2.1. RÉSUMÉ DES ÉTAPES ET DES RÉALISATIONS _____	9
Préparation et extraction des données nécessaires à la classification _____	11
Réalisations 11	11
Automatisation des étapes nécessaires à l'estimation de la phylogénie _____	13
Réalisations 13	13
Automatisation des étapes nécessaires au groupement des souches _____	14
Réalisations 15	15
Développement et automatisation de rapports de validation _____	16
Réalisations 16	16
Développement et automatisation de rapports d'analyses sur les groupes de souche pour diffusion. _____	20
2.2. ACTIVITÉS DE DIFFUSION _____	24
3. ATTEINTES DES OBJECTIFS _____	25
4. RETOMBÉES POUR LE SECTEUR _____	26
5. SUIVI DES INDICATEURS DE RÉSULTATS _____	28
6. SUIVI _____	30

Liste des figures

Figure 1. Étapes du processus de classification des séquences en groupements génétiques.	10
Figure 2. Schématisation de la base de données relationnelle BD Clustering.	12
Figure 3. Répertoires générés pour structuration et archivage des fichiers.	16
Figure 4. Position relative des séquences du dernier trimestre au sein de l'arbre complet (vue rectangulaire, avec agrandissement).	18
Figure 5. Groupements génétiques identifiés par le système de classification.	19
Figure 6. Nombre de sites porcins par cluster par année au Québec.	22
Figure 7. Nombre de sites porcins par cluster par année en Montérégie.	22

Abréviations

ARC&E	Area-based regional control and elimination
ARN	Acide ribonucléique
BSR	Bootstrap rapide
CDPQ	Centre de développement du porc du Québec
CLUSTAL O	Clustal Omega
FMV	Faculté de médecine vétérinaire
LEMP	Laboratoire d'épidémiologie et de médecine porcine
PCR	Réaction de polymérisation en chaîne
ORF	Cadre de lecture ouvert
SRRP	Syndrome reproducteur et respiratoire porcin

1. Rappel du projet initial

1.1 Mise en contexte

Le Syndrome reproducteur et respiratoire porcin (SRRP) est une menace sanitaire causant des pertes annuelles supérieures à 50 millions de dollars au secteur porcin québécois. Le virus ARN responsable de cette maladie mute rapidement et augmente grandement le risque de contamination des troupeaux. En effet, le virus se transmet de très nombreuses façons et les vaccins commerciaux n'offrent qu'une protection limitée.

Le séquençage (code génétique du virus) est fortement encouragé sur le terrain afin de mieux comprendre la transmission du virus entre les sites porcins et implanter des mesures de contrôle adaptées. Le Laboratoire d'épidémiologie et de médecine porcine (LEMP) a récemment mis au point et testé une méthode innovatrice pour classifier les virus SRRP en différents groupes, en vue de simplifier l'analyse et l'interprétation des résultats de séquençage. Toutefois, une automatisation des différentes étapes nécessaires au processus de classification est essentielle avant de pouvoir l'incorporer à un système de surveillance continue de la maladie, le processus actuel étant trop long et coûteux. Des rapports d'analyse incorporant les résultats de la classification doivent également être conçus et automatisés. Ainsi, de l'information sur la circulation des populations virales pourrait alors être transmise aux différents intervenants sur le terrain.

1.2 Objectifs

Objectif général

Adapter une méthodologie issue des travaux de recherche sur la classification des souches du virus SRRP en vue de l'analyse continue de la dynamique des populations virales au Québec et d'une communication efficace de l'information générée à l'industrie.

Objectifs spécifiques

- A) Automatiser les différentes étapes du processus de classification préalablement développé pour permettre la prise en charge par un seul opérateur spécialisé et à un coût moindre.
- B) Développer des prototypes de rapports incorporant les résultats d'analyse des groupes de souches qui serviront d'outils d'aide à la décision.
- C) Être en mesure de rapporter et diffuser l'information produite aux intervenants et décideurs de l'industrie.

1.3 Équipe de réalisation

Responsable du projet

Marie-Ève Lambert, DVM, PhD

Professeure adjointe

Département de sciences cliniques, Faculté de médecine vétérinaire, Université de Montréal

3200 Sicotte, Saint-Hyacinthe, Québec, Canada, J2S 2M2

Téléphone: 450 773-8521 # 8674

Courriel: marie-eve.lambert@umontreal.ca

Collaboratrice

Sylvie D'Allaire, DVM, MSc, PhD

Professeure associée

Département de sciences cliniques, Faculté de médecine vétérinaire, Université de Montréal

3200 Sicotte, Saint-Hyacinthe, Québec, Canada, J2S 2M2

Téléphone: 450 773-8521 ext. 8473

Courriel: sylvie.dallaire@umontreal.ca

Analyste-programmeur

Pascal Audet

Département de sciences cliniques, Faculté de médecine vétérinaire, Université de Montréal

3200 Sicotte, Saint-Hyacinthe, Québec, Canada, J2S 2M2

Courriel: pascal.audet@umontreal.ca

Biologiste moléculaire

Benjamin Delisle

Département de sciences cliniques, Faculté de médecine vétérinaire, Université de Montréal

3200 Sicotte, Saint-Hyacinthe, Québec, Canada, J2S 2M2

Courriel: benjamin.delisle@umontreal.ca

1.4 Partenaires financiers

Ce projet a été réalisé grâce au soutien de plusieurs partenaires financiers.

**Ministère de l'Agriculture, des Pêcheries
et de l'Alimentation du Québec (MAPAQ)**



Les Éleveurs de porcs du Québec



L'Association CLÉ-Montérégie Inc.



Laboratoire d'épidémiologie et de médecine porcine



1.5 Durée du projet

2 ans : 28 janvier 2019 au 28 janvier 2021

2.Déroulement du projet

2.1. Résumé des étapes et des réalisations

Le projet visait à automatiser les différentes étapes nécessaires au processus de classification des souches du virus SRRP en vue d'en créer un processus coordonné et fluide. Les sections suivantes démontrent les différentes étapes qui ont été franchies ainsi que les réalisations associées à chaque étape.

La figure 1 illustre les différentes étapes constituant le processus global de classification. Le projet visait non seulement à réaliser les étapes du processus de classification afin d'obtenir des résultats sur les groupes de souches en circulation au Québec, mais surtout de les enclencher automatiquement en vue de diminuer le nombre d'opérateurs et le temps en ressources humaines nécessaire ainsi que le risque d'erreur inhérent aux nombreuses manipulations. L'automatisation des étapes était nécessaire en vue de pouvoir classer les souches régulièrement à un coût raisonnable dans un objectif de surveillance.

Le processus de classification des souches en différents groupements génétiques est essentiellement divisé en trois portions, comportant chacune plusieurs étapes à automatiser (Figure 1).

- Préparation des données (Bleu)
- Phylogénie (Gris)
- Clustering (Vert)

Au sein de ces trois portions, plusieurs flèches visibles assurent les connexions entre les différentes étapes. Ainsi, chaque flèche représente, en quelque sorte, une action ou procédure qui devait être automatisée par le présent projet. Le rapport documente donc, pour chacune des portions du système de classification, les différentes étapes ainsi que les réalisations accomplies dans le cadre de ce projet.

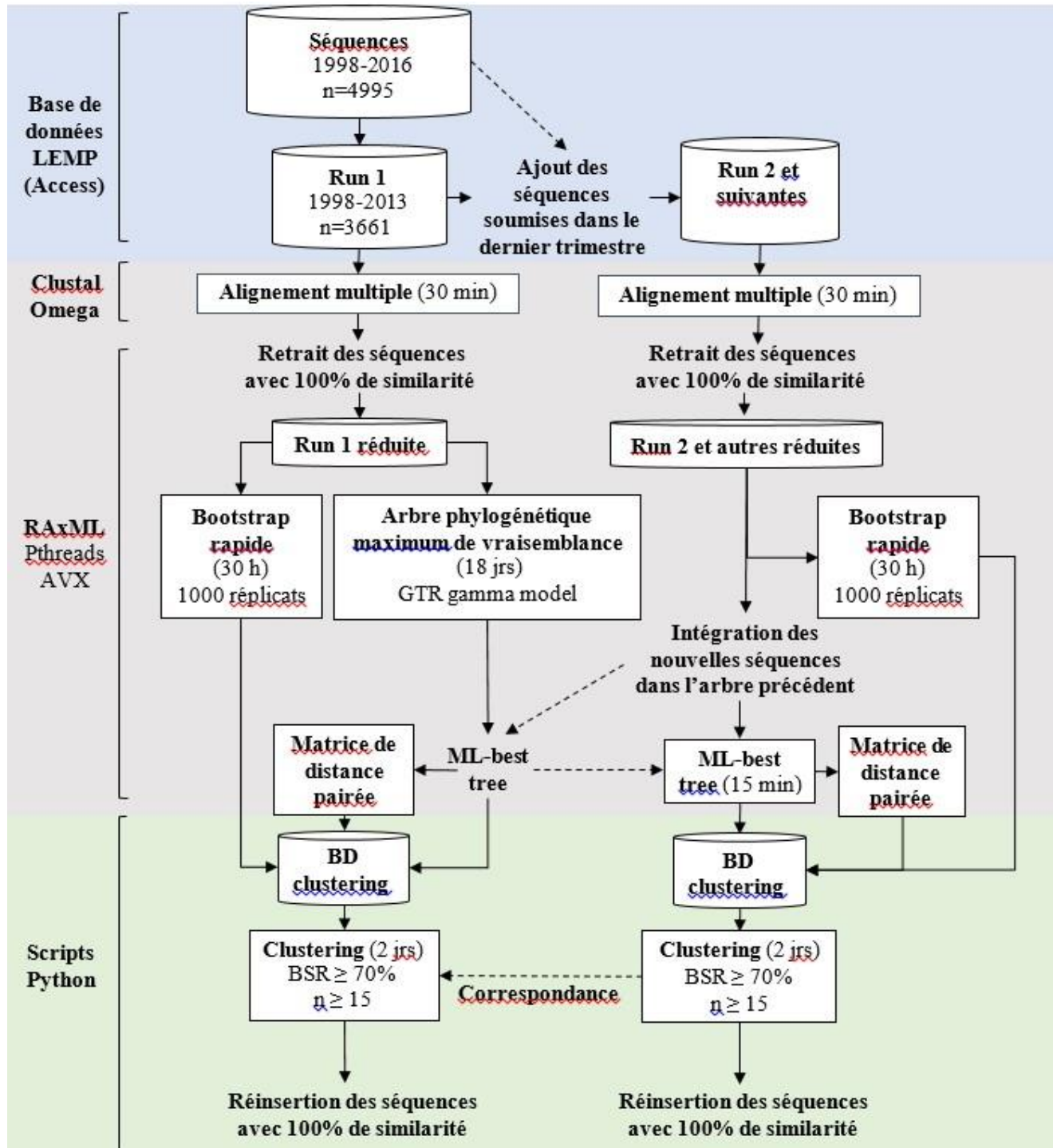


Figure 1. Étapes du processus de classification des séquences en groupements génétiques.

Préparation et extraction des données nécessaires à la classification

Cette étape visait à automatiser l'extraction des séquences du virus SRRP soumises au cours du dernier trimestre (série de 603 paires de bases, ATCG) ainsi que la date d'échantillonnage, deux informations nécessaires au processus de classification des séquences. Ces données devaient être extraites à partir de la base de données Access du LEMP. Cette dernière comprend présentement 7500 séquences et informations associées; donc 858 ont été colligées au cours du présent projet.

Réalisations

- Un script Python a été programmé pour sélectionner automatiquement les séquences à ajouter au processus de classification, soit les nouvelles séquences soumises durant le dernier trimestre. Le script permet également d'exporter les fichiers contenant les séquences, formatés au standard Fasta, multi-séquences.
- Une base de données MySQL (BD clustering) a été développée pour y inclure les données nécessaires à chacun des processus de classification. Cette base de données est schématisée à la Figure 2.
- Les données extraites de la base de données Access du LEMP sont d'abord introduites dans la table (tbl_key). Outre la séquence, des informations connexes sont également extraites et transférées dans la BD clustering :
 - Données d'identification du site : NIL, coordonnées géographiques, région ;
 - Autres informations telles que la participation à une zone de contrôle et le type de production.

Ces dernières données ne sont pas nécessaires au processus de classification, mais pourront être utilisées dans la production de rapports.

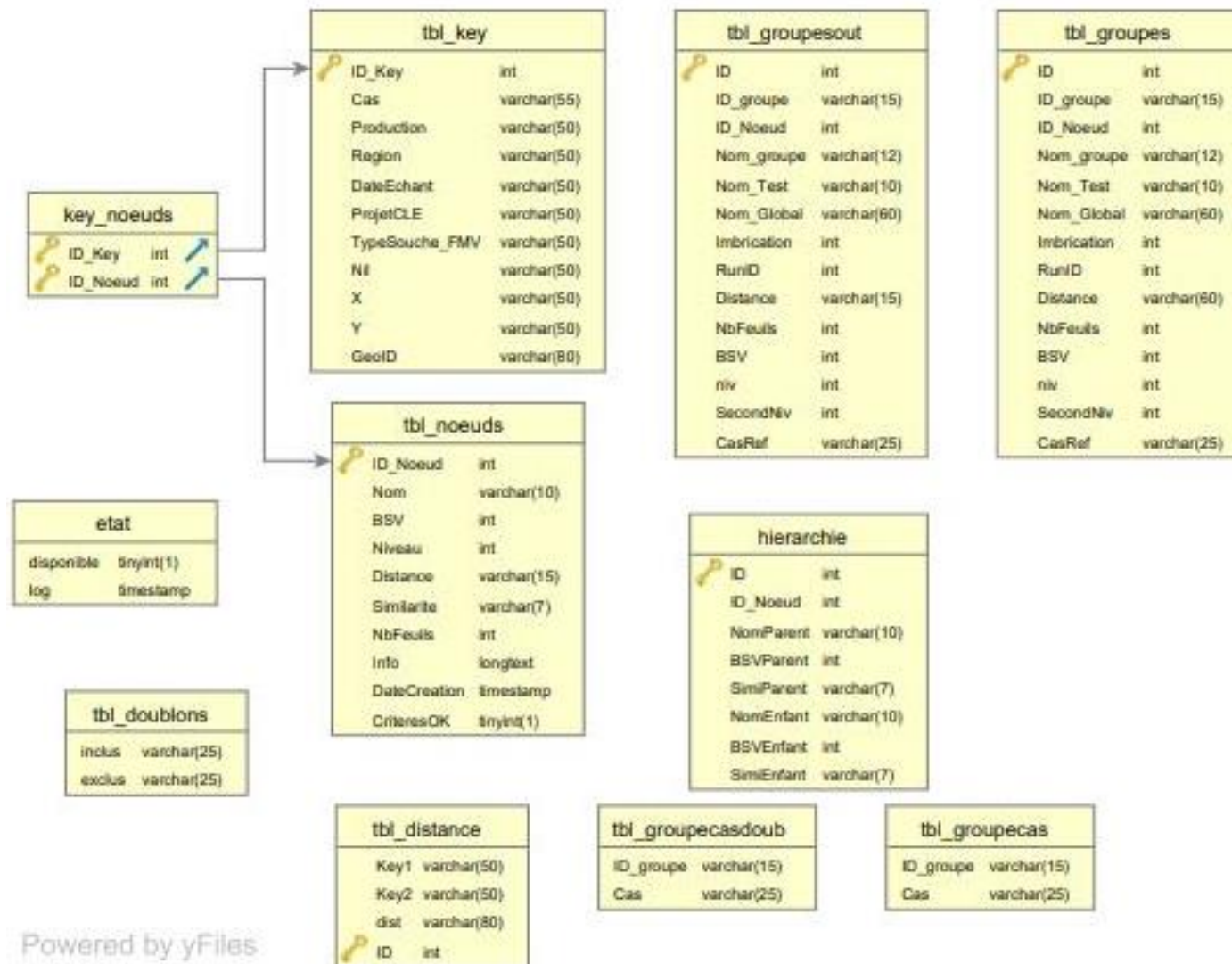


Figure 2. Schématisation de la base de données relationnelle BD Clustering.

Automatisation des étapes nécessaires à l'estimation de la phylogénie

Voici un rappel des différentes étapes relatives au calcul de la phylogénie et qui constituent la 2^e portion du processus de classification. Les étapes reliées à la phylogénie sont illustrées en gris dans la figure 1.

Le projet visait donc à enclencher automatiquement et séquentiellement les étapes suivantes :

- Alignement des séquences par Clustal Omega
- Détection de doublons (séquences dont la similarité est de 100%)
 - Conservation de la première séquence soumise (pré-requis à RaxML).
 - Archive de la liste de doublon pour usage ultérieur.
- Phylogénie
 - Estimation des paramètres du modèle d'évolution (1000 répliquats)
 - Estimation de la topologie par la méthode du maximum de vraisemblance.
- Stabilité des branches
 - Estimation de la stabilité des branches via Bootstrap rapide (1000 répliquats).
- Matrice de distance
 - Calcul de la distance évolutionnaire entre chaque paire de séquences faisant partie de l'arbre.
- Incorporation des résultats dans la base de données MySQL (Fig.1)
 - Calcul de la distance évolutionnaire entre chaque paire de séquences faisant partie de l'arbre.

Réalisations

Une application GUI a été développée pour démarrer automatiquement les étapes énoncées précédemment. Ainsi, plusieurs tâches faisant appel à des logiciels ou codes différents sont maintenant successivement enclenchées. L'application a été testée au cours de 4 trimestres consécutifs où la classification des séquences a été mise à jour afin de s'assurer de la bonne marche de l'application. Une fois le processus enclenché, un temps de calcul d'environ 1.5 jr est par la suite nécessaire avant d'obtenir les résultats. Ce temps de calcul ne peut être diminué, puisqu'il s'agit du temps nécessaire à l'ordinateur pour calculer. Une fois les calculs terminés, les résultats sont

automatiquement archivés dans la base de données MySQL (Fig 2) dans les tables Doublons, Nœuds, Hiérarchie et Distance prévues à cet effet.

Cette application a permis :

- Une diminution du nombre total d'opérateurs. À l'origine, à la fois la responsable du projet, un programmeur et un biologiste moléculaire, soit trois personnes ayant des aptitudes différentes, devaient intervenir à un moment ou à un autre sur les différentes étapes énoncées. Grâce à ce projet, un seul opérateur, soit le programmeur, peut enclencher tout le processus.
- Une diminution du temps nécessaire à la réalisation. La synergie qui devait précédemment exister entre les différents opérateurs responsables de certaines des étapes n'est maintenant plus nécessaire. Également, le risque d'erreur pouvant survenir en raison des nombreuses manipulations de fichiers a considérablement diminué. De plus, une diminution du temps en ressources humaines a été observée. Seules quelques minutes sont maintenant nécessaires pour démarrer le script.
- Une diminution des interactions entre les intervenants. En raison de la COVID, l'accès aux locaux par plusieurs employés était interdit. Le développement de l'application a permis la poursuite du projet en limitant certains délais qui auraient pu être occasionnés.

Automatisation des étapes nécessaires au groupement des souches

Une fois la phylogénie estimée, les groupements génétiques se retrouvant dans l'arbre doivent être identifiés. Par la suite, chaque séquence soumise doit être associée au groupement génétique correspondant. Cette portion est schématisée en vert dans la figure 1. Les étapes suivantes devaient être automatisées séquentiellement dans le présent projet afin d'attribuer un groupement génétique aux séquences.

- Groupement des séquences (clusters)

Algorithme groupant les séquences en fonction de la stabilité du groupement dans l'arbre (bootstrap rapide ≥ 70) ainsi que le nombre de séquences (≥ 15). Le système identifie non seulement les clusters majeurs (près de la racine de l'arbre), mais également les clusters mineurs, soit des clusters imbriqués au sein des clusters principaux.

- Correspondance des groupements

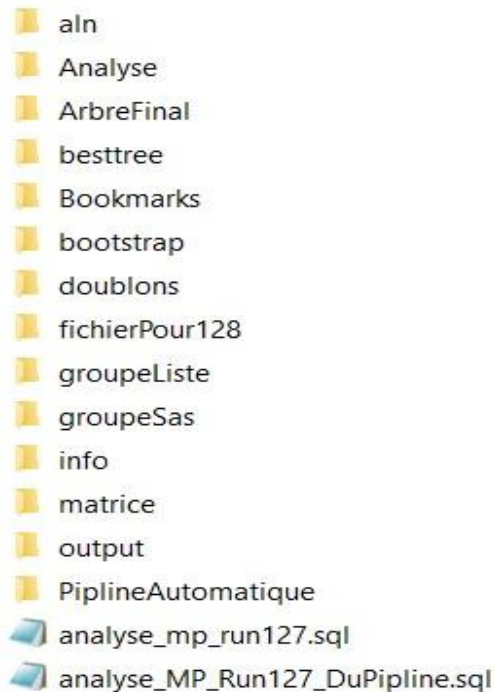
Lors d'un processus de classification, le système identifie une liste de groupements génétiques respectant l'algorithme de classification. Comme les étapes doivent être réenclenchées à chaque trimestre en vue de classer les nouvelles séquences soumises au cours du trimestre, une concordance entre les groupes trouvés d'un trimestre à l'autre doit être effectuée. Globalement, il s'agit d'établir la concordance entre un minimum de 52 clusters majeurs et 99 clusters mineurs ou imbriqués et ce, à chaque processus de classification. Des algorithmes doivent alors être programmés afin de faire cette concordance.

- Structuration et archivage des résultats

Les données et fichiers produits par le système de groupements doivent être structurés et archivés pour être utilisés efficacement lors de la production de rapports.

Réalisations

Un premier module interagit avec la base de données MySQL (Fig 2) et active un script Python préalablement développé par le LEMP en vue de grouper les séquences virales en fonction de différents critères précédemment énoncés. Un deuxième module a été programmé pour faire rapidement la concordance des groupements identifiés par le système. Cela permet également de colliger les résultats des groupements dans la base de données MySQL (BD Clustering) dans les tables GroupesOut, Groupes et Groupes Cas (Fig 2). La table GroupesCasDoub est peuplée automatiquement grâce à un algorithme qui attribue le groupement aux séquences de la table doublon. Ces dernières séquences devaient être retirées de l'arbre afin de calculer la phylogénie, mais il est essentiel d'attribuer un groupement à ces séquences puisqu'elles doivent être intégrées aux rapports de résultats produits par le système (décrit ultérieurement).



Alors que les données produites par le système sont colligées dans une banque de données, les différents types de fichiers produits par le système de classification sont automatiquement enregistrés dans différents répertoires pour faciliter le travail ultérieur. Il est à considérer que plus de 100 fichiers de formats différents sont produits à chaque processus de classification et ces fichiers sont par la suite archivés dans plusieurs répertoires différents sur un réseau partagé par plusieurs utilisateurs.

Figure 3. Répertoires générés pour structuration et archivage des fichiers.

Développement et automatisation de rapports de validation

Plusieurs rapports de validation ont ensuite été conçus afin que le LEMP puisse examiner les résultats. Les rapports permettent de suivre l'évolution du nombre de séquences et de la stabilité de chacun des groupements identifiés. Ces rapports n'ont pas été conçus pour diffusion à l'industrie, mais plutôt dans un but de surveillance et validation des résultats obtenus par l'équipe du LEMP. Voici les différents types de rapports développés.

Réalisations

A) Rapports : chiffrer excel

Des scripts Python ont été programmés en vue d'exporter ces rapports.

A1. Liste des clusters identifiés

- Liste permettant de répertorier tous les clusters identifiés depuis le début de la classification. Chacun est également associé à une séquence de référence pour faciliter le repérage dans le système informatique.

A2. Suivi de la grosseur et de la stabilité au fil du temps

- Chiffrier permettant de suivre pour chacun des clusters l'évolution des deux critères utilisés pour la formation des groupes, soit le nombre de séquences et la stabilité (bootstrap). Grâce à ce rapport, il est possible d'observer si certains groupements augmentent rapidement comparativement à d'autres, ou alors si la stabilité d'un groupement commence à diminuer, ce qui peut indiquer une subdivision éventuellement en sous-clusters.

B) Rapports : arbre phylogénique

B1. Arbre dynamique

Des scripts ont été développés pour incorporer les groupements et autres informations associées à la séquence directement à l'arbre phylogénétique qui est ensuite lu par des logiciels de phylogénie. Il est donc maintenant possible de naviguer directement dans l'arbre pour examiner rapidement les résultats de classification qui prennent la forme d'étiquette jointe à chaque séquence de l'arbre phylogénétique. L'étiquette comprend d'abord le groupement génétique, la date d'échantillonnage, le numéro d'identification de lieu (NIL), le nom du site porcin, la région ainsi que le nom du projet CLÉ. Les résultats ne peuvent être présentés afin de respecter la confidentialité.

En zoomant sur certaines portions de l'arbre, il est alors possible d'identifier certains groupements particuliers en agrandissant spécifiquement une portion de l'arbre (Fig. 4) : par exemple, suite à la détection d'un nouveau groupement, ou alors si un groupement prend de l'expansion de manière significative.

Des listes des nouvelles séquences soumises dans le cadre du dernier trimestre sont également générées automatiquement pour produire des requêtes rapides dans un logiciel de phylogénie et identifier leurs emplacements dans l'arbre en utilisant plusieurs configurations.

B2. Arbre statique

Des représentations visuelles de l'arbre phylogénétique et des groupements génétiques ont été programmées avec le logiciel R. Ces dernières figures présentent les résultats du dernier processus de classification qui a pu être effectué dans le cadre du présent projet et regroupant un total de 6964 séquences (110 doublons) qui ont été soumises entre 1998 et le 30 septembre 2020 (Fig. 5). L'arbre nous permet de visualiser l'emplacement et l'importance relative des groupements génétiques.

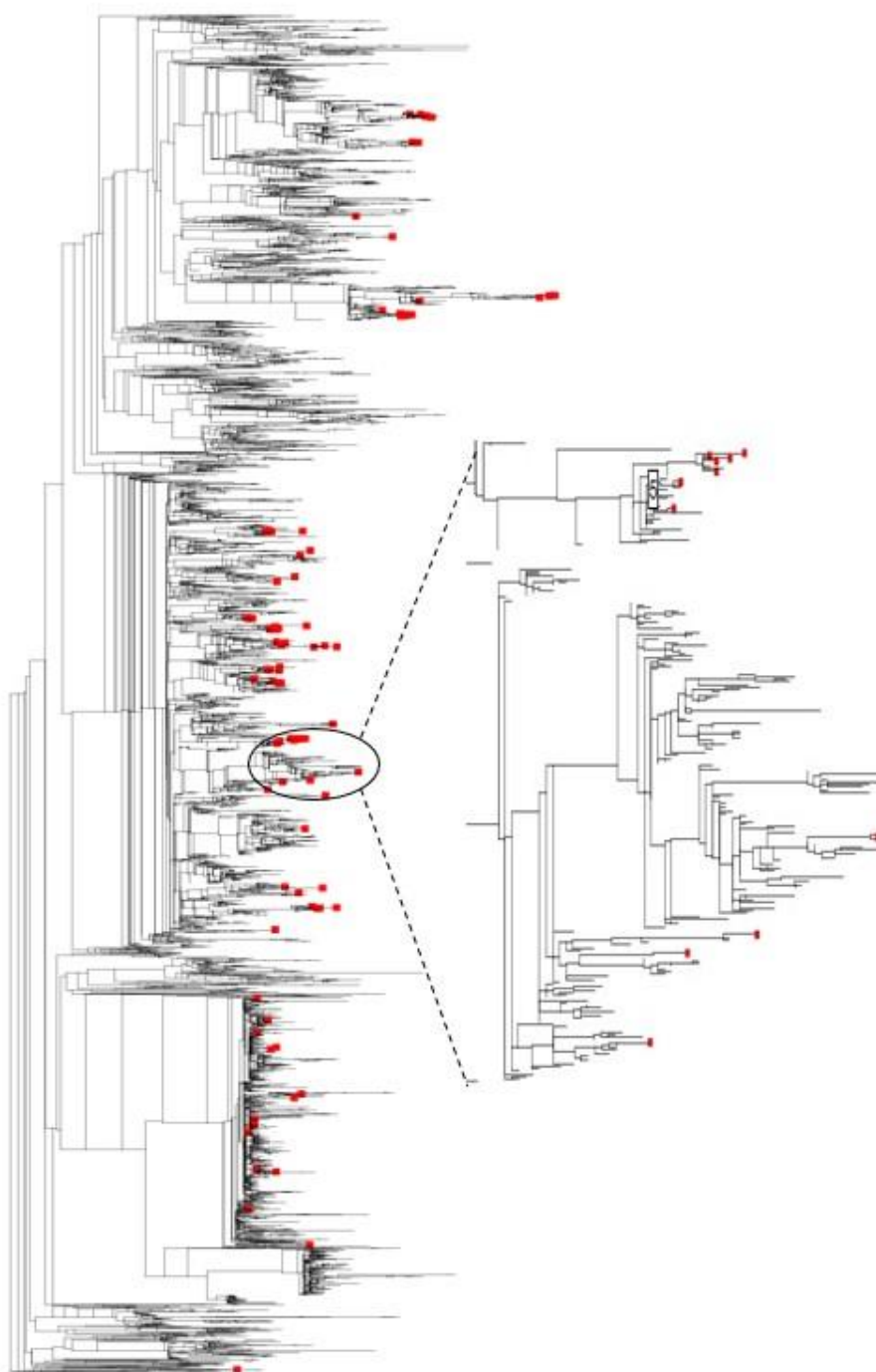
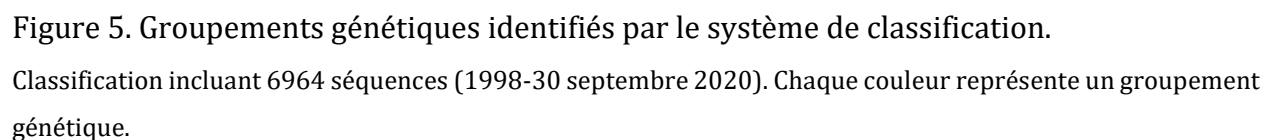


Figure 4. Position relative des séquences du dernier trimestre au sein de l'arbre complet (vue rectangulaire, avec agrandissement).

Noir : 6964 séquences soumises entre 1998 et le 30 septembre 2020. Rouge : 131 séquences soumises au cours du dernier trimestre 1 octobre au 31 décembre 2020.



Développement et automatisation de rapports d'analyses sur les groupes de souche pour diffusion.

Les étapes suivantes visaient à développer un rapport non-nominatif qui permettrait de visualiser l'évolution temporelle des groupements génétiques et ce, à différentes échelles géographiques. Un prototype devait être créé en consultant un groupe d'utilisateurs ciblés, soit les vétérinaires et producteurs de la CLÉ-Montérégie. Le but était d'identifier l'information pertinente à représenter, en visant le format le plus simple possible, tout en protégeant la confidentialité.

La production du rapport devait par la suite être automatisée en vue de pouvoir produire simultanément le même type rapport à différentes échelles géographiques : provinciale, régionale et à l'échelle des zones de contrôle. Le LEMP dessert actuellement 7 projets de contrôle pour les rapports nominatifs de séquençage, il était donc pertinent d'automatiser la production de ces rapports puisqu'un total de 13 rapports sont nécessaires pour la surveillance des souches à différentes échelles géographiques.

Réalisations

Un script a été créé dans le logiciel R en vue de produire un prototype. Le script interagit avec la base de données MySQL, effectue des sélections en fonctions de l'échelle géographique visée et, pour chacun des groupements identifiés par le système de classification, effectue une compilation du nombre de sites porcins ayant soumis par année, ainsi que par trimestre.

Le script produit séquentiellement un rapport pour le Québec et le même type de rapport pour chacune des 5 régions administratives ayant un nombre significatif de séquences soumises annuellement, afin que les sites ayant soumis des séquences ne puissent être identifiés individuellement. Il s'agit des régions suivantes : Montérégie, Centre du Québec, Chaudière-Appalaches, Estrie et Lanaudière. Le script peut également produire des rapports pour chaque zone de contrôle telle que la CLÉ-Montérégie qui était le groupe ciblé dans le cadre de ce projet. Les figures 6 et 7 illustrent des exemples de diagrammes à bulles qui peuvent maintenant être produits à différentes échelles géographiques. Concernant le prototype produit spécifiquement pour la CLÉ-Montérégie, groupe d'utilisateurs cibles du présent projet, il a été envoyé directement par courriel. Il est à noter que des exemples sont présentés au sein de ce rapport afin d'obtenir une image tangible

des rapports développés, et non pour en faire une diffusion à large échelle ou sur d'autres plateformes.

Interprétation des graphiques produits

Dans les diagrammes à bulles, chaque ligne représente un groupement génétique détecté par le système de classification. Les groupements ou clusters génétiques sont identifiés par des numéros
Ex.: CL_29 = Cluster 29 ou CL_29_02 = Sous-cluster 2, imbriqué à l'intérieur du cluster 29.

Chaque bulle comporte un chiffre, soit le nombre de sites ayant obtenu au moins une séquence d'un groupe donné durant la période de temps. La taille de la bulle varie en fonction du nombre de sites.
Ex 3 : Dans la Figure 6, nous pouvons noter qu'en 2019, 18 sites porcins ont soumis une séquence appartenant au cluster 29 sous-cluster 21.

Pour les 6 dernières années en date de production du rapport, les résultats représentent un cumulatif annuel. Pour l'année en cours, les résultats sont exprimés par trimestre. T1 = Janvier à Mars, T2 = Avril à Juin, T3 = Juillet à Septembre, T4= Octobre à Décembre. Alors que tous les groupes de souches sont visibles dans le rapport québécois, seuls les groupes observés dans une région donnée ou dans une zone de contrôle précis sont comptabilisés.

À titre d'exemple, voici informations pouvant être obtenues suite à la consultation rapport provincial (Fig 6) :

- Les souches sauvages qui ont été séquencées au cours des 6 dernières années, soit de 2015 à 2020, se distribuent dans 34 groupements génétiques (clusters). Il est à noter que l'année 2020 n'est pas encore complétée. Le processus de classification devra être réenclenché afin d'obtenir les résultats du dernier trimestre (T4).
- Les clusters ne figurant pas sur le graphique (ex. CL-4 ou 5) ont été détectés antérieurement par le système, soit avant 2015.
 - N.B. Le CL-25 ne figure pas sur le graphique puisqu'il rassemble les souches apparentées à des souches vaccinales.

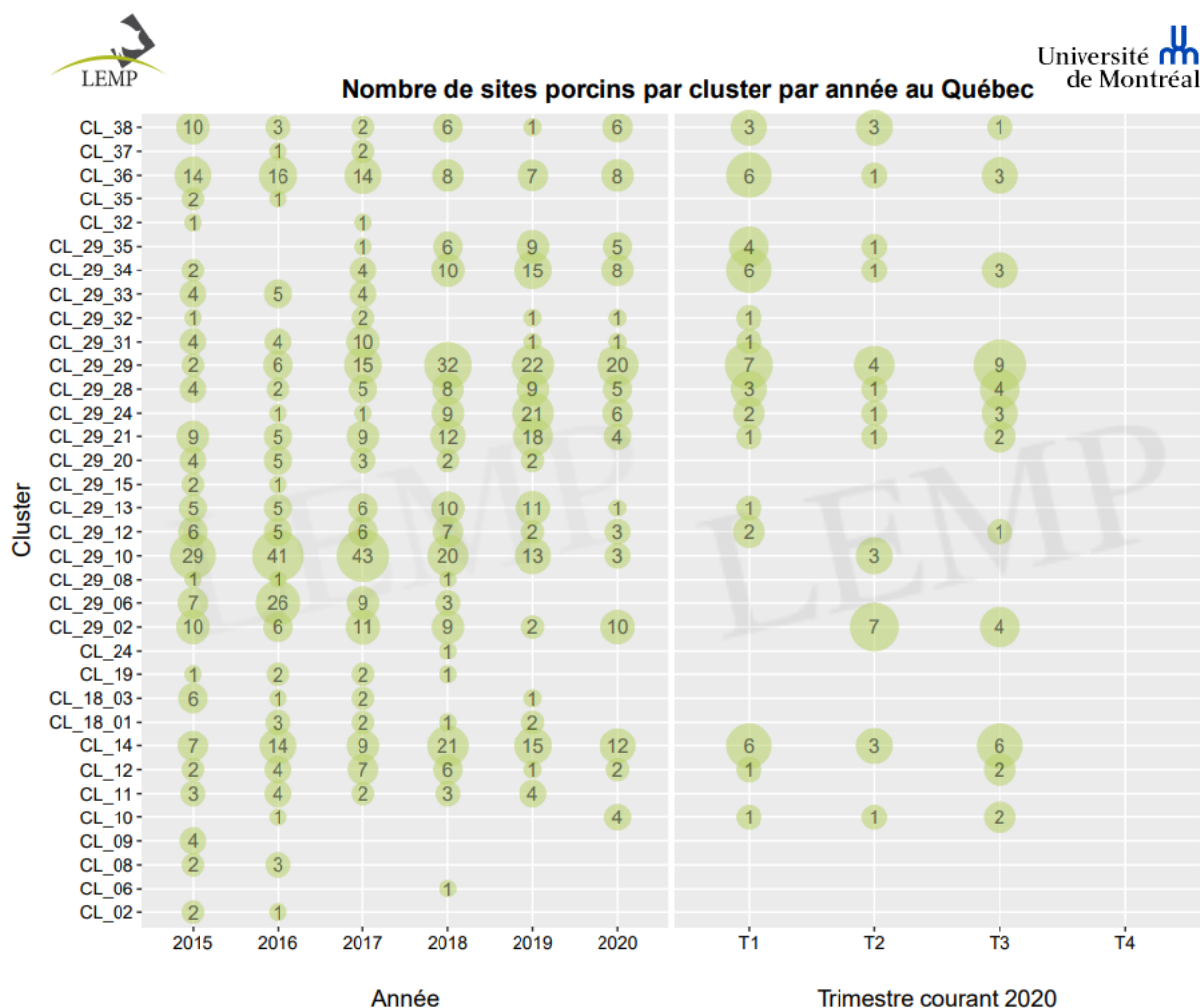


Figure 6. Nombre de sites porcins par cluster par année au Québec.



Figure 7. Nombre de sites porcins par cluster par année en Montérégie.

- Le cluster 29 s'est divisé en plusieurs sous-clusters au fil du temps ; un total de 17 sous-clusters ont été détectés entre 2015 et 2020.
- Le cluster 18 s'est également divisé en sous-clusters au fil du temps.
- Une très grande proportion de la diversité des souches au Québec s'exprime par les sous-clusters du cluster 29.
 - Entre 2015 et 2020, les séquences appartenant aux clusters CL-29_29 et CL-29_10 ont été détectées sur un très grand nombre de sites porcins.
- Plusieurs clusters n'ont été détectés que sur 1 ou 2 sites porcins par année.
 - Ex. CL-02, CL-06, CL-10, CL-32, CL-35...
- Une différence dans la prévalence de sites ayant soumis des séquences appartenant à un cluster donné est visible au cours des années.
 - Les 5 clusters les plus prévalents entre 2015 à 2017.
 - CL-29_10, CL-36, CL-29_06, CL-14 et CL-29_29.
 - Les 5 clusters les plus prévalents entre 2018 et 2020
 - CL-29_29, CL-14, CL-29_34, CL_29_10 et CL29_24.
 - Alors que le CL_29_10 semble en diminution entre 2015 et 2020, le CL-29_29 semble en augmentation.
- La figure 7 nous informe plutôt des types de souches retrouvées spécifiquement en Montérégie. Ainsi ce ne sont pas tous les clusters détectés au Québec qui circulent en Montérégie.

2.2. Activités de diffusion

Environ 70% du temps consacré à ce projet était relié à la programmation des différentes routines présentées alors que 30% du temps a été employé au développement de rapports. Par conséquent, la diffusion possible était limitée pendant la première année et demie du projet. De plus, l'équipe de réalisation doit rapporter un retard important principalement relié à la COVID en seconde année du projet, forçant une restructuration des activités de recherche à distance. Ainsi, le développement des rapports de validation et de diffusion a été finalisé plus tardivement qu'anticipé. Néanmoins, des mises à jour sur l'évolution du présent projet ont été faites dans le cadre des rencontres avec la CLÉ-Montérégie. Un contact régulier avec la coordonnatrice de la CLÉ-Montérégie permettait néanmoins d'assurer un suivi pour le présent projet.

Le prototype de rapport non nominatif a été envoyé aux membres de la CLÉ-Montérégie pour évaluation finale par courriel, incluant des lignes directrices pour l'interprétation du rapport.

Le prototype de rapport provincial non nominatif illustrant les différents types de souche en circulation au Québec a été intégré au site internet du LEMP. Une page web arborant une mise en contexte permet l'accès aux résultats obtenus.

<http://www.lemp.ca/index.php/site/page/view/surveillance>

Des logiciels spécialisés dans la création de liste d'envoi de courriels ont été évalués, cependant le manque de temps pour mettre en application un tel système a limité la mise en application. Également, une réflexion a été faite sur la meilleure plateforme pour diffuser les résultats québécois. Ces derniers pourraient, le cas échéant, être inclus à des listes de diffusions existantes telles que le bulletin du RAIZO porcin du MAPAQ. Les différents intervenants seront consultés éventuellement pour choisir la meilleure option advenant une suite au présent projet (voir section suivi).

Une présentation sera livrée le 9 février 2021 en vue de présenter certains rapports non nominatifs développés dans le cadre de ce projet. La présentation aura lieu dans le cadre du rendez-vous annuel du Réseau Santé Sud-Ouest.

Le rapport final du projet est disponible via le site internet du LEMP à l'adresse suivante :

<http://www.lemp.ca/index.php/site/page/view/publications>

3. Atteintes des objectifs

Voici un rappel des objectifs spécifiques énoncés en début de rapport.

- A) Automatiser les différentes étapes du processus de classification préalablement développé pour permettre la prise en charge par un seul opérateur spécialisé et à un coût moindre.

Le 1^{er} objectif est atteint. Une interface permettant d'opérer le système de classification par un seul opérateur a été développée. Cet accomplissement a nécessité 70% de temps consacré à ce projet. Plusieurs rapports de validation utilisés à l'interne ont également été développés.

- B) Développer des prototypes de rapports incorporant les résultats d'analyse des groupes de souches qui serviront d'outils d'aide à la décision.

Le 2^e objectif est atteint. Des prototypes de rapports non nominatifs permettant de visualiser l'évolution temporelle des groupes ont été développés. Le format utilisé a été choisi en fonction des principaux besoins énoncés par la CLÉ-Montérégie. Cet accomplissement a nécessité 30% du temps consacré à ce projet.

- C) Être en mesure de rapporter et diffuser l'information produite aux intervenants et décideurs de l'industrie.

Le 3^e objectif est atteint. Les calculs nécessaires à la production des rapports non nominatifs ont été automatisés. Il est donc possible de rapporter et diffuser l'information produite à l'industrie régulièrement, soit à chaque trimestre. Spécifiquement dans le cadre de ce projet, le rapport non nominatif produit a été acheminé par courriel à la CLÉ-Montérégie pour consultation finale. Le rapport québécois permettant de visualiser l'évolution des souches au fil du temps a été intégré à une plateforme web gérée par le LEMP.

4. Retombées pour le secteur

Retombées à court terme

Le projet a permis de décrire les souches sauvages en circulation au Québec, en Montérégie et dans la CLÉ-Montérégie. Globalement, une meilleure compréhension des déplacements de population permet d'émettre des hypothèses sur l'épidémiologie de l'infection, ce qui maximisera les bénéfices tirés des efforts de diagnostic soutenus sur le terrain. En effet, des sommes considérables sont investies dans le séquençage par les éleveurs et le projet permet de bonifier les investissements à l'échelle collective. Ainsi, en séquençant, les éleveurs obtiennent des réponses à l'échelle de leur site de production.

Par exemple :

- Est-ce une souche vaccinale ou une souche sauvage ?
- Est-ce une souche en recirculation ou une nouvelle souche?

Grâce à ce projet les éleveurs et différents intervenants de l'industrie peuvent également obtenir de l'information sur les groupes de souches en circulation au Québec.

Par exemple :

- Combien y a-t-il de groupes de souches sauvages ?
- Quel est le groupe de souches le plus prévalent au Québec?
- Est-ce que ce sont les mêmes groupes de souches qui circulent dans deux régions différentes ?
- Y'a-t-il des nouveaux groupes en émergence ?

Le présent projet a permis à l'industrie de se munir d'un outil de surveillance qui peut être utilisé par différents intervenants pour fins de communication. En effet, alors que beaucoup d'outils existants partagent des informations nominatives sur les sites de production, limitant donc la diffusion au-delà des ententes de partage d'informations signées, le rapport québécois produit dans le cadre de ce projet est accessible à tous. Il s'agit donc d'un outil de communication standardisé qui peut être accessible à l'ensemble des éleveurs, vétérinaires et autres intervenants qui ont tous un rôle à jouer en matière de prévention et contrôle de la maladie. En effet, les intervenants de l'industrie pourront facilement visualiser les différentes familles du virus au Québec ainsi que l'évolution au fil du temps et organiser efficacement les interventions de contrôle à grande ou petite échelle. Par exemple, les régions ayant des populations virales différentes devraient être gérées séparément. Les résultats

alimenteront les futures démarches de concertation en vue de la surveillance et du contrôle de la maladie à l'échelle provinciale.

Retombées à moyen terme

Avant ce projet, la description des groupes de souches en circulation ne pouvait se faire que dans un cadre de recherche en raison des investissements de temps et d'argent requis. L'automatisation des différentes étapes nécessaires à la classification des souches a permis l'utilisation du système de classification préalablement développé dans un contexte de surveillance, où des rapports peuvent être mis à jour à chaque trimestre. Les retombées seront alors maximisées dans la mesure où le système est maintenant en roulement, afin de générer de nouveaux résultats à chaque trimestre. Ainsi, une suite à ce projet permettrait de profiter pleinement des réalisations accomplies dans le cadre de ce projet. Dans un même ordre d'idée, le système pourrait également générer des rapports régionaux standardisés pour les différentes régions et zones de contrôle au Québec.

5. Suivi des indicateurs de résultats

Nombre de mises à jour du système, soit la classification des séquences nouvellement soumises (cible 3, 2e année).

- Indicateur surpassé : le processus de classification a été réalisé pour les 4 trimestres de 2019 et 3 trimestres en 2020 comme en témoignent les rapports présentés dans les sections précédentes. Un total de 7 processus de classification ont été réalisés dans le cadre de ce projet.

Nombre de participations aux réunions de la CLÉ-Montérégie (cible 4, 1e année) et (cible 4, 2e année).

- Indicateur partiellement atteint. Deux membres de l'équipe ont participé chacun à deux rencontres pour un total de 4 participations. Ces participations ont eu lieu dans la première et deuxième année de projet. Un contact régulier avec la coordonnatrice de la CLÉ-Montérégie permettait néanmoins d'assurer un suivi pour le présence projet.
- La COVID ayant amené des délais considérables, dans l'apprentissage du travail à distance, du temps a plutôt été investi dans la configuration des ordinateurs pour rendre possible le travail à distance, le travail d'équipe dans le logiciel teams et l'établissement d'une nouvelle routine de travail avec les employés. L'argent a plutôt été investi dans la structuration des activités à réaliser maintenant à distance.

Nombre de présentations à des groupes de producteurs ou autres intervenants (cible 1/an)

- Indicateurs partiellement atteint principalement en raison de la COVID. Tel que mentionné précédemment, une conférence sera réalisée le 9 février 2021 en vue de présenter certains rapports non nominatifs développés dans le cadre de ce projet. La présentation aura lieu dans le cadre du rendez-vous annuel du Réseau Santé Sud-Ouest.

Nombre de rapports ou autres produits de diffusion utilisant l'information issue de ce projet, soit les groupes de souches (mise à jour sur le site web du LEMP, feuillets d'information d'autres organismes, courriels d'information, etc.)

- Indicateur atteint. Le rapport non-nominatif québécois a été mis à jour (3 trimestres) sur le site internet du LEMP. Les éleveurs, vétérinaires et intervenants ont donc accès aux résultats provinciaux. Des informations concernant la manière d'interpréter le graphique sont également fournies sur la page web.
<http://www.lemp.ca/index.php/site/page/view/surveillance>
- Un courriel a été envoyé aux membres de la CLÉ-Montérégie, partenaire du présent projet.
- Le rapport final du projet a été déposé sur le site internet du LEMP dans on onglet prévu à cet effet. <http://www.lemp.ca/index.php/site/page/view/publications>

6. Suivi

Bien qu'aucun suivi ne soit nécessaire dans le cadre du projet, il importe de mentionner que le bénéfice à long terme de l'investissement du présent projet dépendra essentiellement des montants qui seront attribués au roulement de ce nouvel outil de surveillance. En effet, le projet a permis d'assurer à ce que le système de classification préalablement développé dans un contexte de recherche puisse être davantage automatisé pour assurer la faisabilité des mises à jour des groupements génétiques aux 3 mois. Également, le système peut maintenant prendre en charge la production de rapports non nominatifs à différentes échelles géographiques. Cependant, des ressources financières supplémentaires seront nécessaires pour assurer le roulement des outils développées, assurer une mise à jour des groupements génétiques et rapporter les résultats à l'industrie dans les prochains trimestres, voire années.