UTILISATION DU SÉQUENÇAGE À HAUT DÉBIT POUR L'IDENTIFICATION D'ORGANISMES PATHOGÈNES DES PLANTES

19-009-2.2-C-IRDA

Durée du projet : 02/2019 au 12/2023

RAPPORT FINAL

Rédigé par :

Richard Hogue, Ph.D., IRDA
Thomas Jeanne, M.Sc., IRDA

21 décembre 2023

Décembre 2023

ÉQUIPE DE RÉALISATION :

Institut de recherche et de développement en agroenvironnement (IRDA) :
Richard Hogue, Ph.D., Chercheur, Responsable scientifique
Thomas Jeanne, M.Sc., Professionnel de recherche
Joël D'Astous-Pagé, M.Sc., Professionnel de recherche
Vanessa Villeneuve, Technicienne de laboratoire

Université Laval, Centre de recherche Centre Hospitalier Universitaire de Québec :
Arnaud Droit, Ph.D., Chercheur
Clément Plessis, M.Sc., Professionnel de recherche
Emeric Texeraud, M.Sc., Professionnel de recherche

Laboratoire d'expertise et de diagnostic en phytoprotection (LEDP):
Antoine Dionne, M.Sc., Phytopathologiste
Marion Berrouard, Technicienne de laboratoire

Centre de recherche sur les Grains (CÉROM) :
Tanya Copley, Ph.D., Chercheuse

Agriculture et Agroalimentaire Canada (AAC) :
Wen Chen, Ph.D., Chercheuse

Décembre 2023

UTILISATION DU SÉQUENÇAGE À HAUT DÉBIT POUR L'IDENTIFICATION D'ORGANISMES PATHOGÈNES DES PLANTES
19-009-2.2-C-IRDA

1. RÉSUMÉ DU PROJET

Ce projet a consisté à répondre à l'objectif principal qui vise à évaluer la capacité d'identifier par séquençage à haut débit (SHD) des organismes phytopathogènes bactériens et fongiques qui infectent des pommes de terre et des plantes en grandes cultures, et en cultures maraîchères. En partenariat avec le Laboratoire d'expertise et de diagnostic en phytoprotection du MAPAQ (LEDP) et le Laboratoire de phytopathologie du CEROM, nous avons obtenu et traité des échantillons provenant de nos partenaires entre 2019 et 2022. Nous avons validé les protocoles d'extraction d'acides nucléiques nécessaires selon le type d'échantillon, développé les protocoles de séquençage haut débit sur plateforme MiSeq et MinION (Livrable L3). Au total cinq systèmes de détection ont été retenus pour l'évaluation sur MiSeq et deux systèmes sur MinION. Des tests additionnels ont également permis d'évaluer l'approche NanoMiSeq permettant de soumettre moins d'échantillons en même temps à la phase de séquençage. Cette approche sera plus adaptée au débit analytique du LEDP en saison estivale. Nous avons de plus évalué une approche de préparation de librairie en une seule étape qui permet de réduire les coûts d'opération reliés à l'analyse par SHD.

Nous avons également développé un nouvel outil informatique permettant de créer des bases de références taxonomiques en regroupant des séquences de plusieurs bases de données de référence publiques et selon le système d'amplification utilisé pour le SHD. Cet outil (ASVMaker) a été publié dans la revue « Plants » (Annexe 2-4). La base de données taxonomiques de référence spécifiques produite avec cet outil s'intègre dans notre stratégie de double identification (application pour les données obtenues par MiSeq et NanoMiSeq)

Nos résultats montrent un fort potentiel des approches SHD pour identifier les genres pathogéniques ciblés (Livrable L2). Pour certains d'entre eux (essentiellement des genres fongiques), il est possible d'identifier des espèces pathogéniques. Dans des cas intermédiaires, l'identification en deux étapes (bases de données de référence publiques, puis base de données de référence spécifiques + ASVMaker) permet d'améliorer la précision des identifications par rapport à l'utilisation seule des bases de données de référence publiques.

La plateforme web PhytoSHD a été créée pour permettre au LEDP de facilement traiter des données issues du SHD (Livrables L1 et L4). Cette interface visuelle permet de traiter, stocker et visualiser les résultats obtenus. Les visuels sont dynamiques et adaptés au besoin identifié par le LEDP. L'architecture de cette application web permet également d'intégrer d'autres pipelines d'analyses dans le futur.

UTILISATION DU SÉQUENÇAGE À HAUT DÉBIT POUR L'IDENTIFICATION D'ORGANISMES PATHOGÈNES DES PLANTES
19-009-2.2-C-IRDA

Table des matières

Liste des figures

Liste des tableaux

2. OBJECTIFS ET APERÇU DE LA MÉTHODOLOGIE

2.1. Objectifs

L'objectif général est de démontrer que les méthodes SHD permettent d'identifier simultanément, rapidement et précisément les organismes phytopathogènes responsables des principales maladies des pommes de terre ainsi que des plantes de grandes cultures et de cultures maraîchères.

Les objectifs spécifiques sont:

1) faire une évaluation précise par organisme phytopathogène, de la capacité d'identification par le SHD comparativement à une approche diagnostique conventionnelle;

2) établir des processus adaptés à des approches de diagnostic ou de détection impliquant l'utilisation du SHD qui permet un traitement rapide, précis et économique;

3) permettre une interprétation rapide et accessible des données de SHD par le développement d'une interface web intuitive;

4) faire le transfert et la validation de la stratégie de diagnostic par SHD et l'intégration d'une base de données de référence spécialisée et entraînée qui permettra d'accroître la vitesse de traitement des données de SHD.

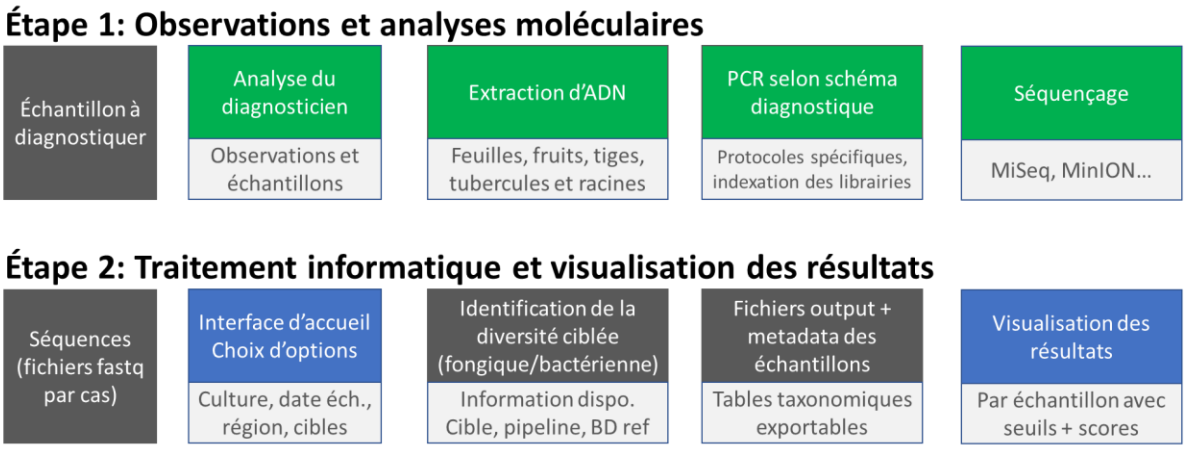

Figure 1. Schéma du processus de traitement d'un cas diagnostic par SHD

2.2. MÉTHODOLOGIE

Note : Plusieurs livrables de ce projet étant associés à des protocoles, certaines sections de la méthodologie peuvent être associées à des résultats. L'annexe 1 liste tous les protocoles et fiches techniques livrés dans ce projet.

Préparation des échantillons

La stratégie mise en place pour l'analyse par SHD a consisté à utiliser un échantillon prélevé par les diagnosticiens le plus similaire à celui qui servira à rendre un diagnostic par la méthode conventionnelle au LEDP et au CEROM. Chaque échantillon a été codifié par un numéro de cas unique et un fichier annexe a été complété pour recueillir toutes les informations pertinentes sur le cas échantillonné. Par la suite, deux sous-échantillons ont été préparés afin de pouvoir évaluer la variabilité de la détection par SHD. Chaque sous-échantillon préparé pour l'extraction d'ADN a été pesé. Tous les sous-échantillons ont été conservés à -20°C avant l'extraction d'ADN.

Extraction des ADN

Pour les tiges, les feuilles, les collets et les fruits, nous avons utilisé la trousse DNeasy Plant Mini (QIAGEN, Toronto, Canada) déjà utilisée au LEDP. Pour les tissus racinaires, nous avons utilisé la trousse DNeasy PowerSoil Pro (QIAGEN, Toronto, Canada).

Les validations méthodologiques suivantes ont été réalisées :

<u>Pour les tiges, feuilles, collets, fruits :</u>

Pour les tissus prélevés dur tige, feuille, collet ou fruit, nous avons évalué la trousse DNeasy Plant Mini (QIAGEN, Toronto, Canada) déjà utilisée au LEDP. L'approche technique fait intervenir l'utilisation d'une matrice de silice en complément d'une agitation initiale avec un appareil de type Fastprep-24™ (MPBiomedical, Solon, OH, E.U.). La limitation de la quantité de matériel initial peut rendre difficile l'évaluation directe de la performance de l'extraction, car les concentrations d'ADN génomiques sont souvent inférieures à 10 ng μL^{-1}. Pour pallier cette difficulté technique, il est recommandé d'ajouter un contrôle d'extraction pouvant permettre de facilement évaluer l'efficacité de l'extraction.

<u>Pour les racines :</u>

Pour des tissus racinaires, nous avons évalué une nouvelle trousse DNeasy PowerSoil Pro qui est une version mieux adaptée aux échantillons racinaires par rapport à la trousse de base pour les sols DNeasy PowerSoil (QIAGEN, Toronto, Canada) déjà utilisée au LEDP. Nous l'avons comparée à la trousse FastDNA Spin kit for Soil (MPBiomedical, Solon, OH, E.U.). Nous utilisons cette dernière trousse au laboratoire d'écologie microbienne de l'IRDA. La nouvelle trousse DNeasy PowerSoil Pro de Qiagen a l'avantage de pouvoir être utilisée par les robots d'extraction QIAcube

(QIAGEN) du LEDP et présente des correctifs pour améliorer l'efficacité d'extraction des groupes fongiques. L'évaluation de cette trousse a fait l'objet d'une fiche technique (Annexe1-1 MAP_QIASOLPRO_Fiche_v1).

Pour les tubercules de pomme de terre :
Les tissus de type tubercules de pomme de terre présentent une spécificité particulière compte-tenu de la grande quantité d'amidon que le prélèvement de pelures trop épaisses peut entraîner. Pour l'extraction d'ADN, des prélèvements de pelures superficielles sont visées si la trousse DNeasy Plant Mini est utilisée. S'il est nécessaire d'extraire une plus grande quantité de tissus, il est important de ne pas prélever trop d'amidon.

Préparation des librairies

Préparation des librairies pour séquençage MiSeq et NanoMiseq

(Plus de détails sont disponibles sur les protocoles détaillés mentionnés en annexe et fournis au LEDP lors de formations - Protocole-LEM-LIB-MiSeq-2022.04)

1. Matériel nécessaire

- Polymérase TAQ haute-fidélité : Q5® High-Fidelity DNA Polymerase (Whitby, ON)
- Billes pour les purifications: Axygen™ AxyPrep Mag™ PCR Clean-up Kits (New York, NY, E.U)
- Amorces pour l'amplification spécifique d'une cible (avec adaptateurs pour les index)
- Amorces pour l'indexation : Illumina (Nextera)/IBIS-ULAVAL

2. Étapes à suivre

a) Amplification des ADN : Selon le protocole détaillé fourni en annexe (A1-1) et en utilisant les amorces du tableau 1.
b) Après purification du premier PCR et une normalisation des amplicons obtenus, une seconde amplification de 10-15 cycles en double indexation permettant d'indexer facilement jusqu'à 388 échantillons séparément.
c) Une seconde de purification et de normalisation et la génération d'un pool d'amplicons indexés pour le séquençage.

Tableau 1. Liste des amorces utilisées pour les librairies MiSeq/NanoMiSeq

Groupe	Nom	Sens	Amorce	Reference	Sequence
Eucaryotes	EUC	F	Euk-Illumina-F	Comeau 2011	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCY GCGGTAATTCCAGCTC
	EUC	R	Euk-Illumina-R	Comeau 2011	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT AYGGTATCTRATCRTCTTYG
Champignons	BITS	F	BITS-ITS1	Bokulich 2013	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAC CTGCGGARGGATCA
	BITS	R	B58S3-ITS1	Bokulich 2013	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GAGATCCRTTGYTRAAAGTT
Bactéries	BACTV4V5	F	515FB-EMP	Parada 2016	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGT GYCAGCMGCCGCGGTAA
	BACTV4V5	R	926R-EMP	Apprill 2015	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GGACTACNVGGGTWTCTAAT
Fusarium	EF1A	F	Fa-150-ill	Cobo-D´ıaz 2019	ACACTCTTTCCCTACACGACGCTCTTC CGATCTCCGGTCACTTGATCTACCAG
	EF1A	R	Ra-2-ill	Cobo-D´ıaz 2019	GTGACTGGAGTTCAGACGTGTGCTCTT CCGATCTATGACGGTGACATAGTAGCG
Oomycètes	OOM	F	MC1-ill	Mugnier et Grosjean, 1995	ACACTCTTTCCCTACACGACGCTCTTC CGATCTAAGTTAAAGTCGTAACAAGG
	OOM	R	MC4-ill	Mugnier et Grosjean, 1995	GTGACTGGAGTTCAGACGTGTGCTCTT CCGATCTCATCCACTGCTGAAAGTTG

- D'autres amorces des régions alternatives ont été évalués mais ne sont pas mentionnés dans ce tableau en raison de leur plus faible performance (Par exemple : Région V6-V8 pour le 16s ARN des bactéries; région ITS2 des champignons, gène rps10 pour les oomycètes.

<u>Préparation des librairies pour séquençage sur MiSeq/ NanoMiseq en une seule étape</u>

(Plus de détails sont disponibles sur les protocoles détaillés mentionnés en annexe et fournis au LEDP lors de formations - Protocole-LEM-LIB-MiSeq-OneStep-2023.02)

Le protocole de préparation de librairie MiSeq/NanoMiSeq peut être adapté pour amplifier et indexer des ADN en une seule étape. Pour cela il est nécessaire d'utiliser des amorces sens (*Foward*) et anti-sens (*Reverse*) spécifiques qui incluent les index. Si cette approche offre des avantages au niveau du temps et des coûts de matériel, au laboratoire la gestion des amorces et des index dans la phase de séquençage peut-être plus complexe. De plus, il sera important d'évaluer avec le temps si le niveau de détection est équivalent.

Préparation des librairies pour séquençage sur MinION

(Plus de détails sont disponibles sur les protocoles détaillés mentionnés en annexe et fournis au LEDP lors de formations : LIB-ONT-2023.02 et LIB-ONT-2023.10)

1. Matériel nécessaire

Note importante : La technologie MinION évoluant très rapidement, les kits et cellules utilisés dans le projet ont évolués entre les résultats présentés dans ce rapport et les analyses effectuées avec le LEDP dans la phase de transfert.

- Native Barcoding Kit 24 V14 (SQK-NBD114.24)
- NEBNext Ultra II End repair/dA tailing Module (NEB, catalogue # E7546)
- NEBNext Quick Ligation Module (NEB, E6056)
- NEB Blunt/TA Ligase Master Mix (NEB, M0367)
- R10.4.1 Flow Cell (FLO-MIN114)
- Tube de 1,5 ml de type DNA LoBind

2. Étapes à suivre (voir le protocole détaillé mentionné en annexe)

- Préparation des ADN
- Amplication de la région ciblée (16S rARN; ITS, voir tableau 2)
- Purification des produits PCR
- Dosage des produits PCR
- Réparation des fragments amplifiés (end-prep)
- Ajout des barcodes (Native barcode ligation)
- Ligation de l'adaptateur de séquençage
- Chargement de la flow cell R10

Tableau 2. Liste des amorces utilisées pour les librairies MinION

Groupe	Nom	Sens	Amorce	Reference	Séquence
Bactéries	16S-ONT	F	SDBact0008c S20-ONT	Matsuo 2021	TTTCTGTTGGTGCTGATATTGC AGRGTTYGATYMTGGCTCAG
	16S-ONT	R	1492R-ONT	Matsuo 2021	ACTTGCCTGTCGCTCTATCTTC CGGYTACCTTGTTACGACTT
Champignons	ITS1-ONT	F	ITS-1F-ONT	Gardes et Bruns 1993	TTTCTGTTGGTGCTGATATTGC CTTGGTCATTTAGAGGAAGTAA
	ITS1-ONT	R	ITS-4-ONT	Gardes et Bruns 1993	ACTTGCCTGTCGCTCTATCTTC TCCTCCGCTTATTGATATGC

Analyse de séquences SHD

Cette section renseigne sur le traitement informatique de base pour traiter les données produites à la section précédente. Ce traitement est également automatisé via l'application PhytoSHD.

<u>Traitement des données issues du séquençage MiSeq ou NanoMiSeq</u>

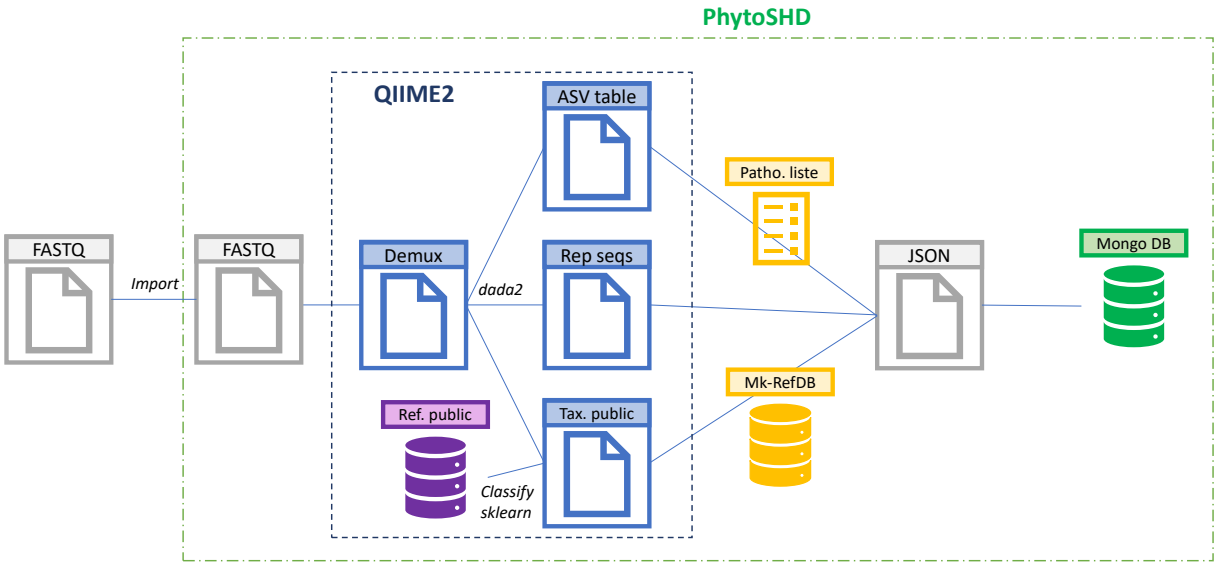

Figure 2. Schéma résumant les principales étapes du traitement bio-informatique (MiSeq/NanoMiSeq)

Description détaillée :

1. Environnement informatique et prérequis : L'ensemble des systèmes de détection SHD (tableau 1) ont été soumis à une analyse bio-informatiques permettant d'identifier les séquences variantes pour chaque cible recherchée. Ce traitement bio-informatique a été réalisé sous environnement Linux et les outils bio-informatiques QIIME2 [4]. Pour l'automatisation des processus et l'intégration des données avec l'interface web, les différentes phases de traitements bio-informatique ont été rassemblées dans des scripts BASH ou Snakemake en utilisant des dockers spécifiques (https://www.docker.com/). Les détails de l'application web PhytoSHD sont décrit à la section 2.3
2. Filtration et génération d'une table d'ASV : Une étape importante du traitement initial des séquences consiste à filtrer les séquences pour obtenir des variants de séquence d'amplicon (*Amplicon Sequence Variants*, ASV) également appelés variants de séquence exacte (ESVs). Historiquement les méthodes de filtration ne permettaient que d'identifier des unités taxonomiques opérationnelles (*Operational Taxonomic Unit*, OTU) qui présentaient des similitudes de séquence à un niveau de 97% d'homologie. Un OTU regroupait des séquences d'origine biologiques, mais également les séquences qui contenaient des erreurs liées aux processus de SHD qui étaient plus élevées avec les anciennes technologies de SHD (ex. Pyroséquençage). Un OTU regroupait ainsi plusieurs séquences homologues à 97%, tandis qu'un ASV identifie un variant de séquence biologique spécifique. L'analyse des ASVs permet ainsi une meilleure définition des microbiomes et facilite la comparaison de résultats SHD obtenus d'études différentes. Aujourd'hui, les avancées technologiques des équipements et des logiciels permettent d'optimiser la filtration des données de séquençage en intégrant des approches d'apprentissage machine dans la filtration des données. Nous avons utilisé l'approche DADA2 [5] pour réaliser cette étape. En remplacement de la notion d'OTU, nous avons ainsi adopté la notion d'ASV.
3. Bases de de données de référence taxonomique publiques : Chaque variant de séquence d'amplicon (ASV) détecté doit ensuite être associé à une identification taxonomique. Il est nécessaire de se baser sur des bases de données de référence publiques qui regroupent des séquences références. Il existe des bases de données de référence publiques généralistes qui permettent l'identification d'organismes bactériens (GreenGenes 13.8 [6]; SILVA138 [7]), des organismes fongiques (UNITE8.2 [8]) ou d'autres

organismes eucaryotiques (SILVA138; UNTE8.2-EUK). D'autres bases de données de référence sont plus spécifiques et permettent d'étudier plus en détail certains groupes microbiens [Fusarium (EF1A-DB); Oomycètes (OOM-AAC)]. Les algorithmes permettant d'identifier la taxonomie la plus proche pour une séquence donnée sont également importants. Selon l'approche utilisée, les performances peuvent être très variables en termes de rapidité et de spécificité. Pour l'attribution taxonomique des ASV nous avons utilisé l'approche «classify-sklearn » implémentée dans QIIME2 pour la première identification sur les bases de données de référence publiques.

4. Bases de données de référence spécifiques obtenue avec ASVMaker : Une base de données de référence spécifiques pour 9 genres bactériens et 38 genres fongiques a été créée (voir section 2.3.2). Cette base de données de référence spécifiques est utilisée en double identification avec les bases de données de référence publiques pour les régions de l'ADN ribosomale pour les bactéries (16S rADN), et les champignons (ITS) ou en base de données unique pour le gène du Facteur d'élongation alpha (EF1alpha) pour les champignons du genre *Fusarium*. Cette seconde identification est directement effectuée en language « python » avec une recherche à 100% d'homologie et 100% de couverture.
5. Table taxonomique : Les tables taxonomiques sont générées pour chaque cible SHD évaluée. L'information détaillée des ASV (table d'ASV) a été regroupée par niveau d'identification taxonomique (du règne jusqu'à l'espèce lorsque possible).

<u>Traitement des données issues du séquençage sur MinION</u>

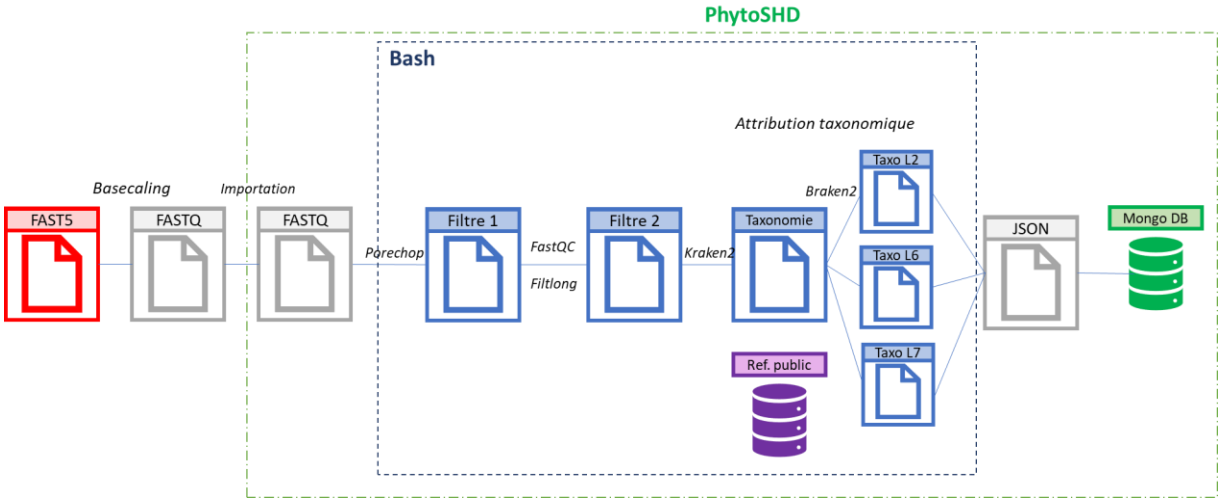

Figure 3 Schéma résumant les principales étapes du traitement bio-informatique (MinION)

Description détaillée :

1. Environnement informatique et prérequis:

Le traitement des données générées par MinION est plus complexe que pour les données de MiSeq. Le traitement initial des fichiers FAST5 pour générer des fichier FASTQ nécessite un ordinateur avec une très bonne carte graphique pour utiliser les performance GPU dans le traitement des données. Par la suite les fichier FASTQ peuvent être traités dans un environnement standard via Conda ou un Docker spécifique dans l'application PhytoSHD

2. Étape de « basecalling » pour traiter les fichiers FAST5 :

Cette étape est réalisée avec « Guppy » avec le modèle « SUPer high accuracy model » (SUP). Elle permet de traduire le signal brut généré par MinION en séquence et qualité. Les outils utilisent des réseaux de neurones nécessitant la puissance de calcul GPU. Nous avons utilisé une carte graphique NVIDIA (min Gbi RAM, CUDA version 6.1 et plus). D'autres outils de « basecalling » sont compatibles (ex. Dorado»)

3. Filtration et génération d'une table d'OTU :

Les séquences sont premièrement traitées avec « Porechop » selon les paramètres par défaut, pour retirer les adaptateurs. Une seconde étape de filtration avec « Fitlong » permet de filtrer les séquences selon des paramètres par défaut en spécifiant une longueur minimale pour les séquences (bactéries : 1400, champignons 300 pb).

4. Table taxonomique :

L'étape d'attribution taxonomique est réalisée avec l'outil « Kraken2 » à partir de séquences publiques de référence. L'outil « Braken2 » permet de finaliser les identifications possibles et de déterminer les proportions relatives des taxons à un niveau taxonomique donné (ex. genre).

3. RÉSULTATS SIGNIFICATIFS OBTENUS

3.1. Base de données de référence spécifique adaptée aux données de SHD

Outil de création de base de données de référence spécifiques – ASVMaker

L'outil ASVMaker (Figure 4) a été créé dans la première partie du projet. Cet outil est particulièrement important dans le projet, car il permet de nous assurer de répondre aux impressions qui peuvent demeurer avec la simple utilisation des bases de données de référence publics tels que SILVA et UNITE. Avec les bases de données de référence publiques, la taxonomie est associée selon un consensus de similarité de 99 ou 97%. Donc plusieurs ASV peuvent avoir la même taxonomie de consensus, par exemple au genre sans avoir plus d'information. Avec ASVMaker et les séquences de la base de données de référence spécifiques, les correspondances de 100% sont accessibles et permettent de discriminer des variants qui ont la même taxonomie initiale obtenue à l'aide d'une base de données de référence publique. On peut alors identifier plus précisément la présence d'organismes pouvant être problématiques pour les cultures ciblées et qui sans l'usage d'ASVMaker auraient été confondues avec d'autres ASV ne permettant pas ainsi l'identification spécifique de l'ASV représentant un agent pathogène.

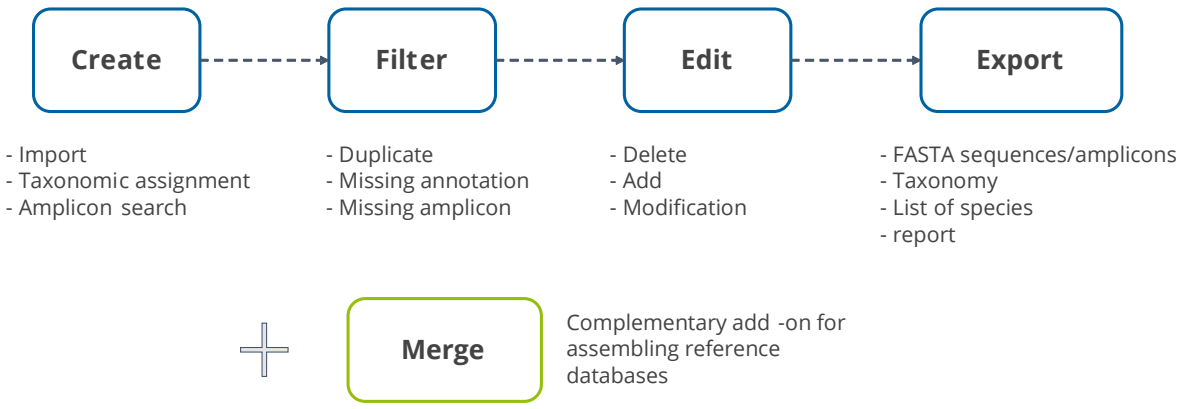

Figure 4. Fonctionnalités de l'outil ASVMaker

Base de données Q2_MkrefDB V1 (Bactérie 16S – Champignons ITS)

À l'aide du nouvel outil ASVMaker, une base de données de références spécifiquse a été créée pour améliorer l'identification des genres associés à des organismes pathogènes des plantes pour les espèces ciblées dans le projet (Tableau 5). Nous

avons initialement ciblé 38 genres fongiques et 9 genres bactériens (Figure 5). Les séquences publiques de trois bases de données de référence publiques ont été téléchargées, puis les étapes décrites dans la figure 4 ont été appliquées afin d'obtenir des séquences d'amplicon unique (donc spécifiques aux amorces utilisées). Les attributions taxonomiques ont été normalisées et les identifications différentes présentant des séquences identiques ont été catégorisées sous des appellations « SA » pour « shared amplicon ». Cela permet de documenter les identifications possibles pour une même séquence d'amplicon.

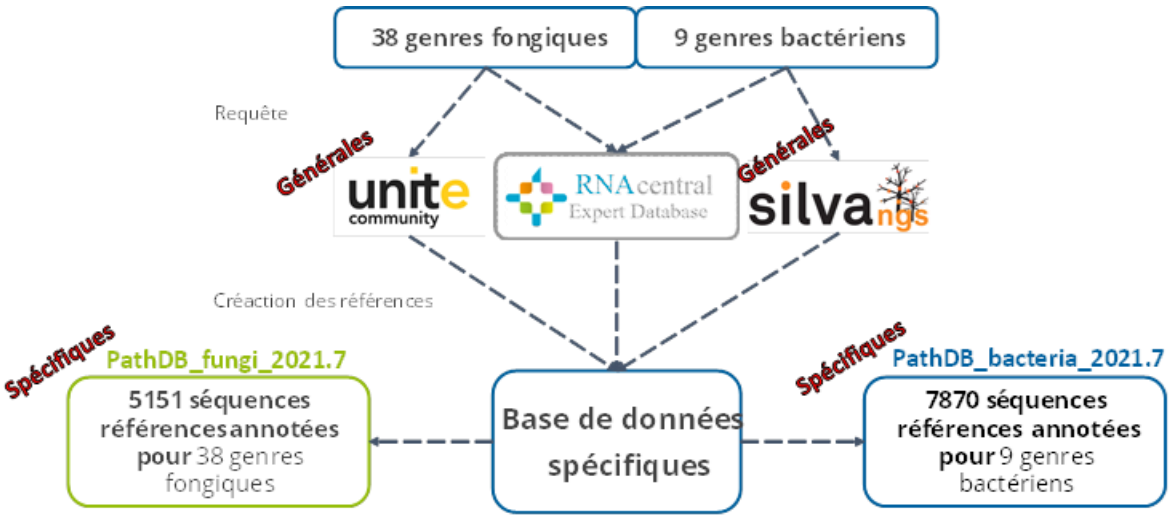

Figure 5. Schéma de la constitution de la base de données de références spécifiques

Les figures 6 et 7 illustrent la taille moyenne des ASV références qui constituent la base de données de références spécifiques pour les attributions taxonomiques par SHD. On observe que pour les champignons (système BITS) la taille des fragments est très variable. Pour les bactéries la taille des fragments est très constante.

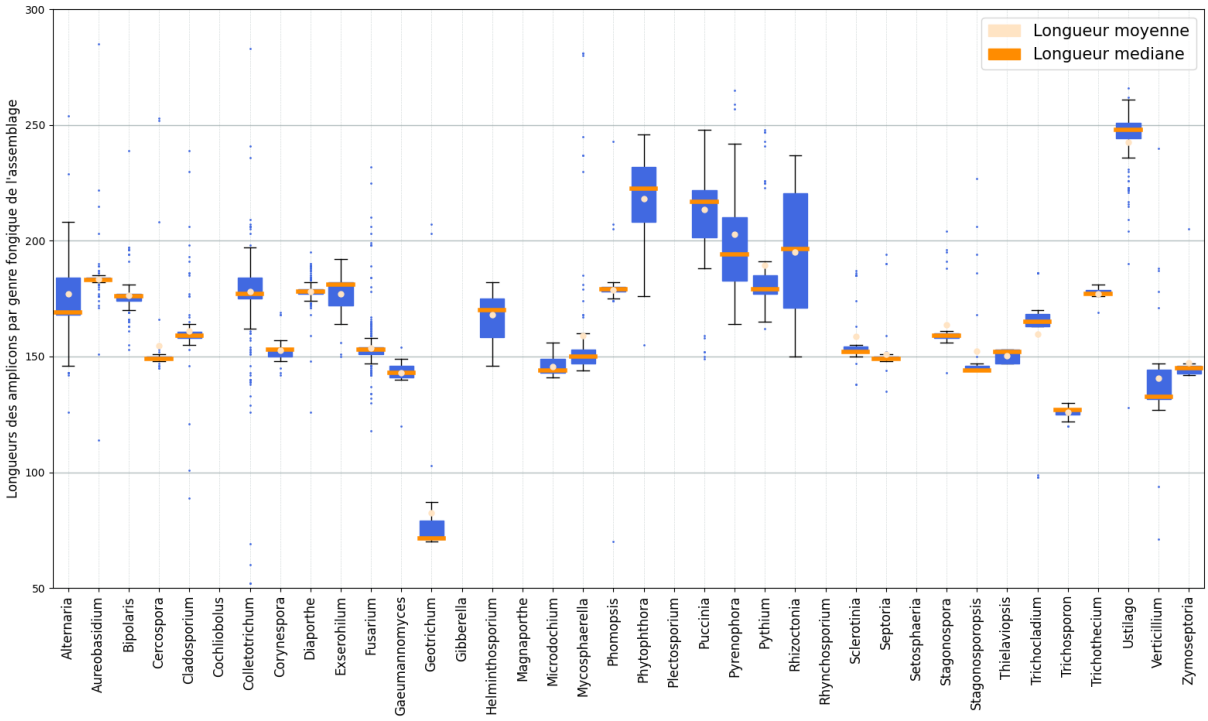

Figure 6. Taille moyenne des fragments selon les genres fongiques étudiés pour le système d'amplification utilisé (BITS)

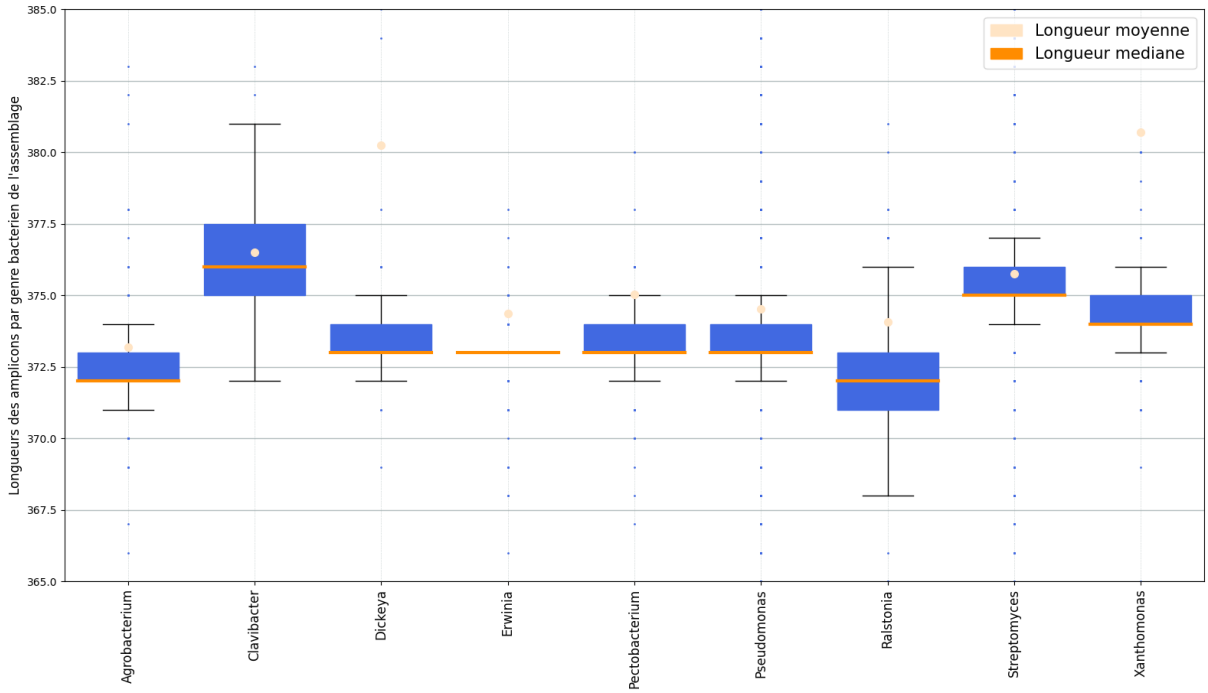

Figure 7. Taille moyenne des fragments selon les genres bactériens étudiés pour le système d'amplification utilisé (BACTV4V5)

Les figures 8 et 9 illustrent le nombre de séquences finales (variant unique) qui constituent la base de données de références spécifiques. On observe que l'assemblage des données initialement obtenu avec les bases de données de référence publiques et traitées par ASVMaker permet d'obtenir une diversité plus importante qu'avec les seules banques de données de référence publiques (Silva et RNAcentral) traitées individuellement. On observe également que tous les genres microbiens ciblés ne présentent pas la même diversité génétique pour les régions d'amplification utilisées.

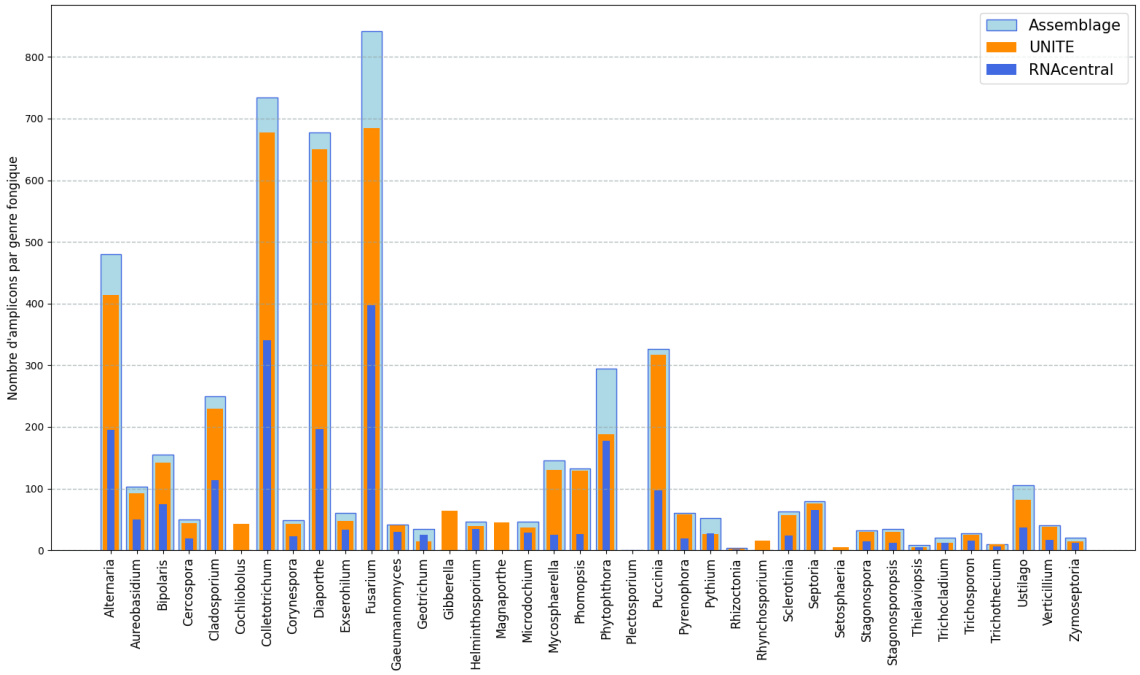

Figure 8. Nombre de séquences fongiques et contributions selon les sources de données

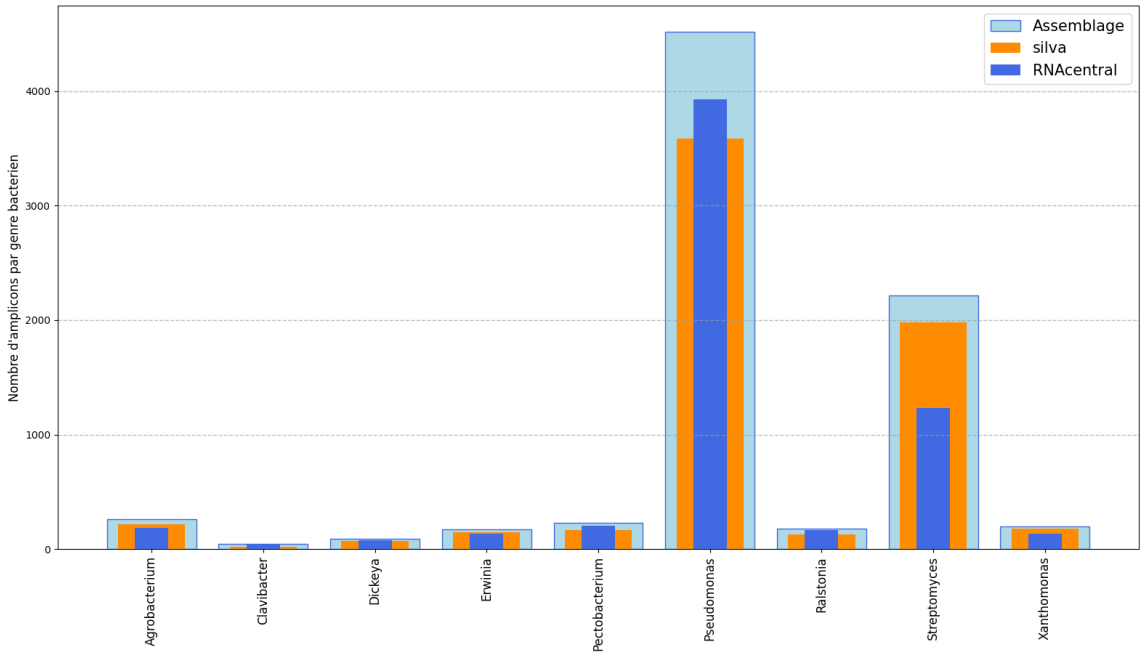

Figure 9. Nombre de séquences bactériennes et contributions selon les sources de données

Base de données pour les *Fusarium* sp. (Gène EF1-alpha)

L'outil ASVMaker permet également de générer des bases de données de référence spécifiques complètes pour des gènes pour lesquels aucune base de données de référence publique n'est disponible. Nous avons créé une base de données de référence spécifique adaptée pour le traitement de données SHD pour le gène EF1-alpha pour les fragments amplifiés avec les amorces décrites dans le tableau 1. Au total 7340 séquences ont été récupérées sur les bases de données de référence publiques. 855 variants de la région amplifiée ont été retenus. 39 de ces variants présentaient des homologies (2 espèces maximum). Au total 81 descriptions différentes dont 42 décrites à l'espèce ou au complexe et 39 avec des homologies.

3.2. Bilan des détections par approche conventionnelle et SHD

Bilan des détections par groupe de culture

Entre 2019 et 2021, nous avons pu traiter 180 échantillons en cultures maraîchères, 276 en grandes cultures (Soya, Maïs, Céréales) et 135 en pomme de terre. Dans la majorité des cas, nous avons analysé un échantillon par cas de diagnostic, mais certains cas diagnostiques peuvent présenter plusieurs échantillons ou des réplicas techniques.

Au total 772 échantillons ont été analysés (excluants échantillons pour les mises au point et le transfert de connaissances)

Les tableaux 3 et 4 résument les comparaisons entre les verdicts obtenus par la méthode conventionnelle et par l'approche SHD par groupe de cultures.

Trois classes de verdict ont été déterminées en collaboration avec le LEDP:

Divergence : Pas de concordance entre les approches conventionnelles et SHD

Concordance : Concordance entre les approches conventionnelles et SHD

Gain : Précision plus importante avec l'approche SHD (ex. Conventionnelle : genre microbien; SHD : espèce microbienne)

Les cas de divergence peuvent correspondre à plusieurs situations. La première peut-être une limitation liée à la prise d'échantillon. En effet, lors du prélèvement, nous n'avions pas accès exactement au même prélèvement. Pour certaines maladies, le symptôme peut être petit et difficile à discerner, cela limite les possibilités de prélèvement. Dans d'autres cas, il peut être difficile d'identifier précisément l'organisme pathogène. C'est le cas pour des organismes qui présentent des similarités avec des organismes d'autres genres. Dans ces cas il est possible d'être limité à une identification à la famille ou à des niveaux taxonomiques supérieurs.(voir tableau 5).

Tableau 3. Comparaison entre verdict conventionnel et verdict SHD (cas analysés par le LEDP)

Culture	Nombre d'échantillons	Proportions (%)		
		Divergence	Concordance	Gain
Autres	8	25	63	13
Céréales	47	17	49	34
Crucifères	43	21	47	33
Cucurbitacées	54	13	35	52

Solanacées	75	13	48	39

Tableau 4. Comparaison entre détection conventionnelle et détection SHD (Cas analysés par l'IRDA et le CEROM

Culture	Nombre d'échantillons	Proportions (%)		
		Divergence	Concordance	Gain
Pomme de terre IRDA	135	16	69	15
Soya CEROM	104	18	59	23
Maïs CEROM	52	27	63	10
Céréales CEROM	73	42	42	16

Bilan des détections par groupe d'organismes pathogènes

Le tableau 5 résume les niveaux de détection possibles selon les organismes pathogènes ciblés dans le projet.

Tableau 5. Niveau de détection des organismes pathogènes ciblés dans le projet et système de détection SHD recommandé.

Type Maladie**	Nom de l'agent pathogène	Code	Ident. Genre (g) Espèce (esps)	Commentaire
Bacterienne	*Clavibacter michiganensis*	B01	esps	
Bacteria	*Erwinia tracheiphila*	B02	g, esp	
Bacteria	*Pectobacterium sp.*	B03	g	
Bacteria	*Dickeya*	B04	g	Genres SA6
Bacteria	*Pseudomonas corrugata*	B05	g	
Bacteria	*Pseudomonas savastonoi*	B06	g	
Bacteria	*Pseudomonas syringae*	B07	g,esp	
Bacteria	*Xanthomonas cucurbitae*	B08	g	
Bacteria	*Xanthomonas raphani*	B09	g	
Bacteria	*Xanthomonas campestris*	B10	g	
Bacteria	*Agrobacterium tumefasciens*	B11		Pas de cas
Bacteria	*Streptomyces Scabies*	B12	g, esp	
Bacteria	*Ralstonia solanacearum*	B14	g	
Fongique	*Alternaria brassicicola*	F01	esp	
Fongique	*Alternaria cucumerina*	F02	g	
Fongique	*Alternaria solani*	F03a	esp	
Fongique	*Alternaria alternata*	F03b	g	
Fongique	*Alternaria tomatophila*	F04	g	

Type Maladie**	Nom de l'agent pathogène	Code	Ident. Genre (g) Espèce (esps)	Commentaire
Fongique	*Aureobasidium zeae*	F05		Pas de cas
Fongique	*Bipolaris sorokiniana*	F06	g	
Fongique	*Cercospora kikuchii*	F07		Pas de cas
Fongique	*Cercospora sojina*	F08		Pas de cas
Fongique	*Cercospora zeae-maydis*	F09		Pas de cas
Fongique	*Cladosporium*	F10	g	
Fongique	*Cladosporium cucumerinum*	F11	g	
Fongique	*Cochliobolus carbonum*	F12		Pas de cas
Fongique	*Cochliobolus heterostrophus*	F13		Pas de cas
Fongique	*Colletotrichum truncatum*	F14	g.esp	*dematium*
Fongique	*Colletotrichum coccodes*	F15	g.esp	
Fongique	*Colletotrichum graminicola*	F16		Pas de cas
Fongique	*Cosynespora cassiicola*	F17	g.esp	
Fongique	*Diaporthe phaseolorum*	F19	g,esp	
Fongique	*Fusarium oxysporum*	F20	g.esp	
Fongique	*Fusarium avenaceum*	F21	g.esp	
Fongique	*Fusarium solani*	F22	g.esp	
Fongique	*Fusarium equiseti*	F23	g.esp	
Fongique	*Fusarium graminearum*	F24	g.esp	
Fongique	*Fusarium poae*	F25	g.esp	
Fongique	*Gaeumannomyces graminis*	F26	g,esp	1 seul cas (tritici)
Fongique	*Geotrichum sp.*	F27		Pas de cas
Fongique	*Magnaporthe sp.*	F28		Pas de cas
Fongique	*Microdochium nivale*	F29	g,esp	*seminicola, bolleyi*
Fongique	*P. infestans*	F30	*	*
Fongique	*Phytophthora sp.*	F31	*	*
Fongique	*Plectosporium tabacinum*	F32		Pas de cas
Fongique	*Puccinia coronata f. sp. avenae*	F33	g.esp	
Fongique	*Puccinia graminis f. sp. tritici*	F34		Pas de cas
Fongique	*Puccinia triticina*	F35	g.esp	
Fongique	*Puccinia sorghi*	F35	g.esp	
Fongique	*Pyrenophora teres*	F36	g	
Fongique	*Pyrenophora tritici-repentis*	F37	g.esp	
Fongique	*Pythium*	F38	g.	*
Fongique	*Rhizoctonia*	F40		*
Fongique	*Rhizoctonia solani*	F41		*
Fongique	*Rhynchosporium secalis*	F42		Pas de cas

Type Maladie**	Nom de l'agent pathogène	Code	Ident. Genre (g) Espèce (esps)	Commentaire
Fongique	*Sclerotinia sclerotiorum*	F43	g,esp	
Fongique	*Septoria*	F44	g	
Fongique	*Septoria glycines*	F45	g	
Fongique	*Septoria tritici*	F46	g	
Fongique	*Setosphaeria turcica*	F47		Pas de cas
Fongique	*Stagonospora avenae*	F48	g	
Fongique	*Stagonospora nodorum*	F49	g	
Fongique	*Stagonosporopsis*	F50	g.esp	*trachelii, astragali*
Fongique	*Thielaviopsis basicola*	F51		Pas de cas
Fongique	*Ustilago*	F52	g.esp	*maydis*
Fongique	*Verticilium*	F53	g, esp	Pour certains variants, problème avec *Acremonium et Gibellulopsis*
Fongique	*Verticilium dahliae*	F53a	g, esp	
Fongique	*Verticilium albo-atrum*	F53b	g, esp	
Fongique	*helminthosporium solani*	F54	g.esp	
Fongique	*Colletotrichum orbiculare*	F56		Pas de cas
Eucaryote	*Spongospora subteranea*	E1	g.esp	

** Les bactéries ont été amplifiées par le système BACTV4V5 pour le MiSeq et 16S-ONT pour MinION. Les champignons ont été amplifiés par le système BITS pour MiSeq et ITS-ONT pour MinION. Les fusarium ont également été amplifiés avec le système EF1A et les oomycètes avec le système OOM.

L'intégration de gammes de référence des proportions relatives des détections pour chaque taxon détecté donne une information complémentaire au diagnosticien pour estimer l'importance de la présence d'un organisme pathogène détecté dans un échantillon (ANNEXE1-9 Référence_Taxon.pdf).

3.3. Comparatif Miseq/ NanoMiSeq/ MinION

Dans la seconde partie du projet, nous voulions comparer différentes approches de SHD pour proposer au LEDP des solutions adaptées à leur réalité. En effet, pour que les approches de SHD soient utilisables au LEDP, il est nécessaire de pouvoir réaliser des phases de séquençage régulièrement (au moins 1 fois par semaine) pour une gamme de 12 à 24 échantillons à la fois. L'approche MiSeq est plus adaptée pour des volumes importants (300-400 échantillons par phase d'analyse). Nous avons donc comparé l'approche MiSeq avec les approches NanoMiSeq et MinION. L'approche NanoMiseq est très proche de l'approche MiSeq, mais adaptée pour traiter entre 12 et

36 cibles par phase d'analyse. L'approche NanoMiseq présente l'avantage d'avoir le même niveau de qualité des séquences obtenues que l'approche MiSeq. De plus les scripts pour l'analyse informatique sont identiques et compatibles avec notre stratégie de double identification. L'approche MinION est très différente et permet d'être plus autonome, car elle peut se réaliser sur place au LEDP. Elle permet également de traiter entre 12 et 24 cibles par phase d'analyse. Cette approche à l'avantage de pouvoir séquencer des régions plus grandes (ex. l'ensemble du 16S rARN ou de la région ITS), mais la qualité des séquences obtenues (plus d'erreur) est moins bonne que celle du MiSeq. La phase de traitement bio-informatique est plus longue et nécessite plus de ressources informatiques (mémoire GPU). Nous avons effectué plusieurs mises au point pour pouvoir comparer ces approches de séquençage.
Nous avons comparé 24 échantillons (bactéries / champignons) en utilisant ces 3 approches et avons constaté que les résultats étaient identiques tant pour le MiSeq que pour le NanoMiseq. Les résultats étaient relativement similaires avec l'approche MinION pour les bactéries, mais non concluants pour les champignons. Cependant, ce premier essai sur le MinION a été effectué avec les cellules R9 et une première version du "pipeline" non optimisée pour les champignons. Dans les derniers mois du projet, en raison du changement de version des cellules et de la chimie employée avec celles-ci de la compagnie Oxford Nanopore pour le MinION, nous avons adapté nos protocoles (voir annexes) et revu complètement nos processus de traitement des données du MinION (Figure 3). Nos derniers tests ont été plus concluants pour les champignons, et les identifications obtenues par le MinION permettaient d'obtenir une identification plus similaire par rapport aux résultats obtenues par MiSeq. Nous avons toutefois noté des problèmes pour des échantillons à forte diversité (par exemple, racines) ou pour de faibles proportions de séquences cibles.

3.4. Outils de traitement et de visualisation des résultats (PhytoSHD)

Un objectif important du projet visait le développement d'une application web pour faciliter le traitement des données de SHD et en visualiser les résultats.
Les travaux ont été entamés dès la seconde année du projet. Les critères suivants ont été définis à la suite d'une consultation du LEDP :

1- L'application devra permettre le traitement des données SHD par des utilisateurs non bio-informaticiens.
2- L'application devra héberger les données SHD analysées (pas les données brutes)
3- L'application devra permettre de visualiser efficacement les données taxonomiques des cas diagnostiques analysés par SHD

4- La plateforme sous laquelle l'application sera développée devra permettre d'évoluer et d'intégrer d'autres processus de traitement informatique (script)
5- Au moment du développement de l'application, il n'avait pas été choisi si l'application serait hébergée localement ou par Cloud.

Fonctionnement de l'application PhytoSHD

L'application PhytoSHD présente quatre menus. Le premier Menu « Saisie des données » (Figure 10) est l'interface qui permet d'importer les fichiers bruts de SHD (fichiers FASTQ). On importe individuellement ou en groupe les fichiers et les métadata relatives à un échantillon. Les métadata représentent la liste de toutes les variables qui sont utiles pour constituer un verdict diagnostique, par exemple la liste peut regrouper la culture, la partie infectée de la plante, le type de symptôme exprimé, le lieu et la date de prélèvement, les conditions de culture, la variété de plantes et le type de régie de culture, etc. Avec les utilisateurs, nous avons établi qu'une liste de champs d'intérêt pour la base de données sera générée. Un fichier Excel permet d'inscrire les informations demandées et inclut également un onglet décrivant les étapes et les descriptions des champs.

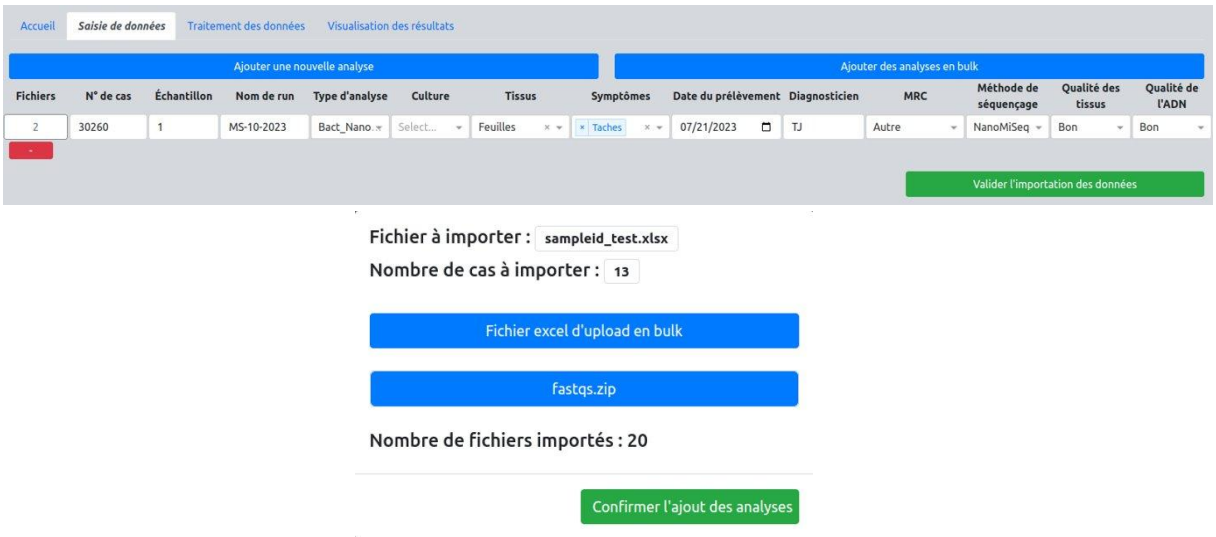

Figure 10. Menu « saisie des données » de l'application PhytoSHD

Le Menu « Accueil » donne un aperçu des derniers échantillons. Il est également possible d'effectuer une recherche spécifique pour un cas diagnostique. Depuis ce menu, il est possible de visualiser la qualité des séquences importées pour un échantillon, éditer ou supprimer un échantillon, lancer une analyse et visualiser les résultats d'une analyse (figure 11).

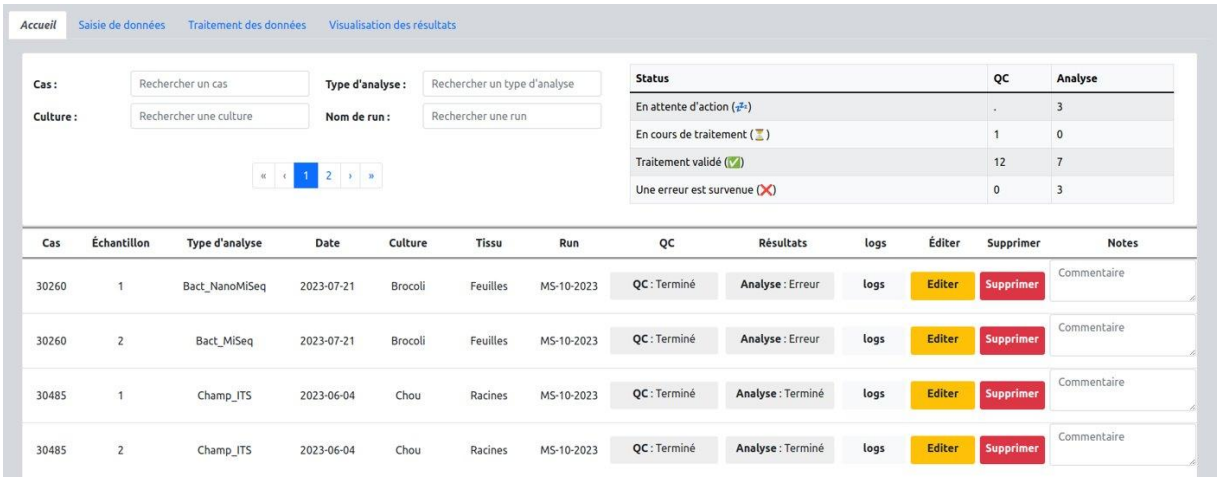

Figure 11. Menu « Accueil » de l'application PhytoSHD

La figure 12 montre un exemple de rapport de qualité. Cette fenêtre apparait lorsque l'on clique sur « QC ». Plusieurs visuels sont disponibles en faisant déroulé vers le bas.

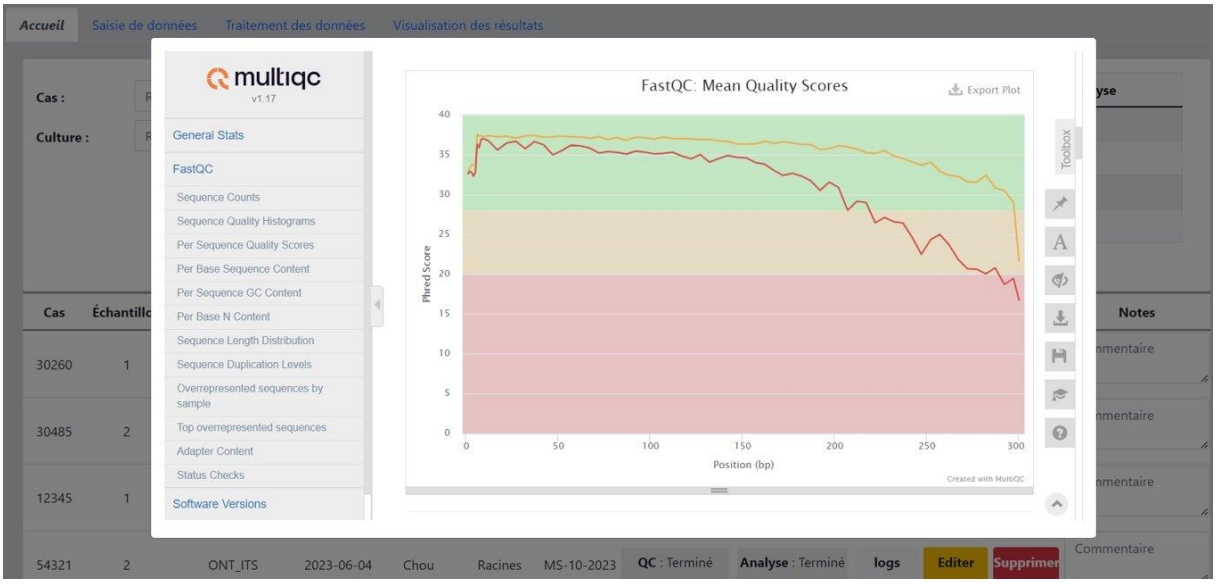

Figure 12. Outil intégré dans l'application PhytoSHD permettant de visualiser la qualité des séquences importées.

Le menu « Traitement des données » permet de lancer des analyses pour un échantillon ou pour un ensemble d'échantillons (figure 13). Les échantillons disponibles pour analyse sont affichés dans la première section et ils peuvent être sélectionnés individuellement ou tous à la fois. Il suffit ensuite de lancer l'analyse et un message de confirmation apparaîtra.

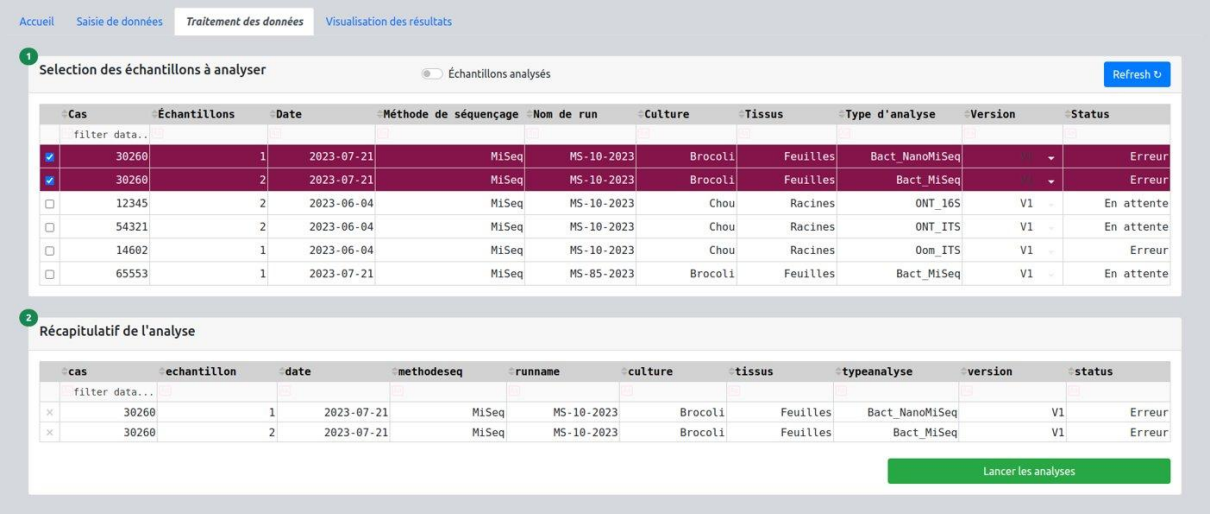

Figure 13. Menu « Traitement des données » de l'application PhytoSHD

Les figures 14 et 15 présentent le menu de visualisation des résultats. Il faut sélectionner le cas et l'analyse sur la droite et les résultats s'afficheront. Le graphique est dynamique. Cela signifie qu'en cliquant sur les sections d'intérêt, il est possible de zoomer et naviguer dans la taxonomie des espèces détectées. En choisissant l'option « pathogène », seuls les groupes pathogènes ciblés dans le projet apparaîtront. La barre faisant référence aux proportions relatives permet de filtrer les groupes trop importants ou trop faibles.

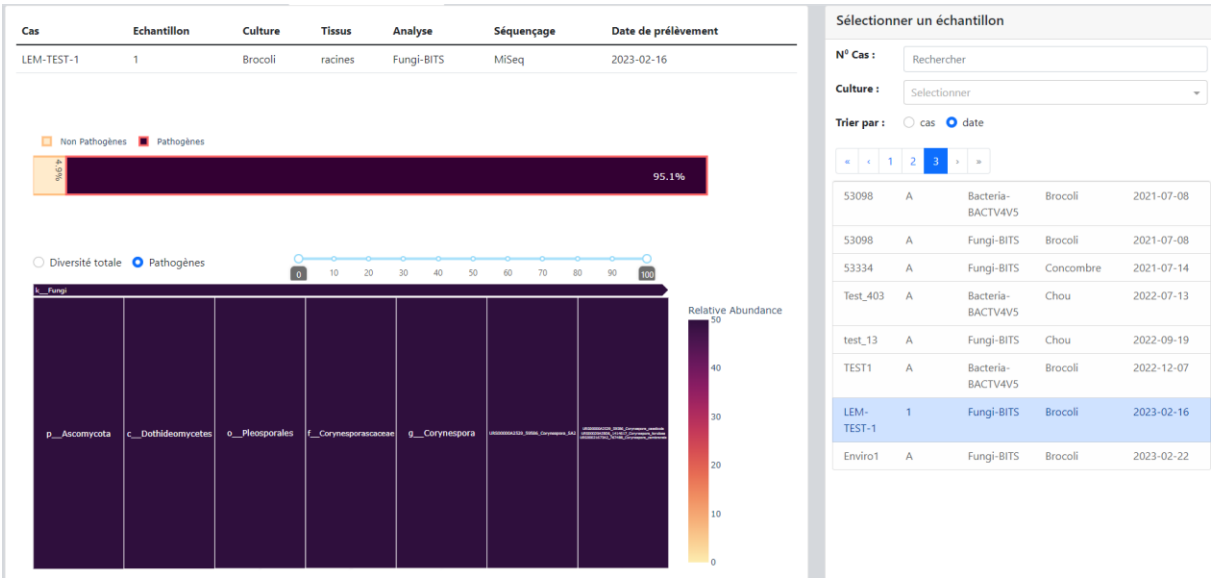

Figure 14. Menu « Visualisation des données » de l'application PhytoSHD

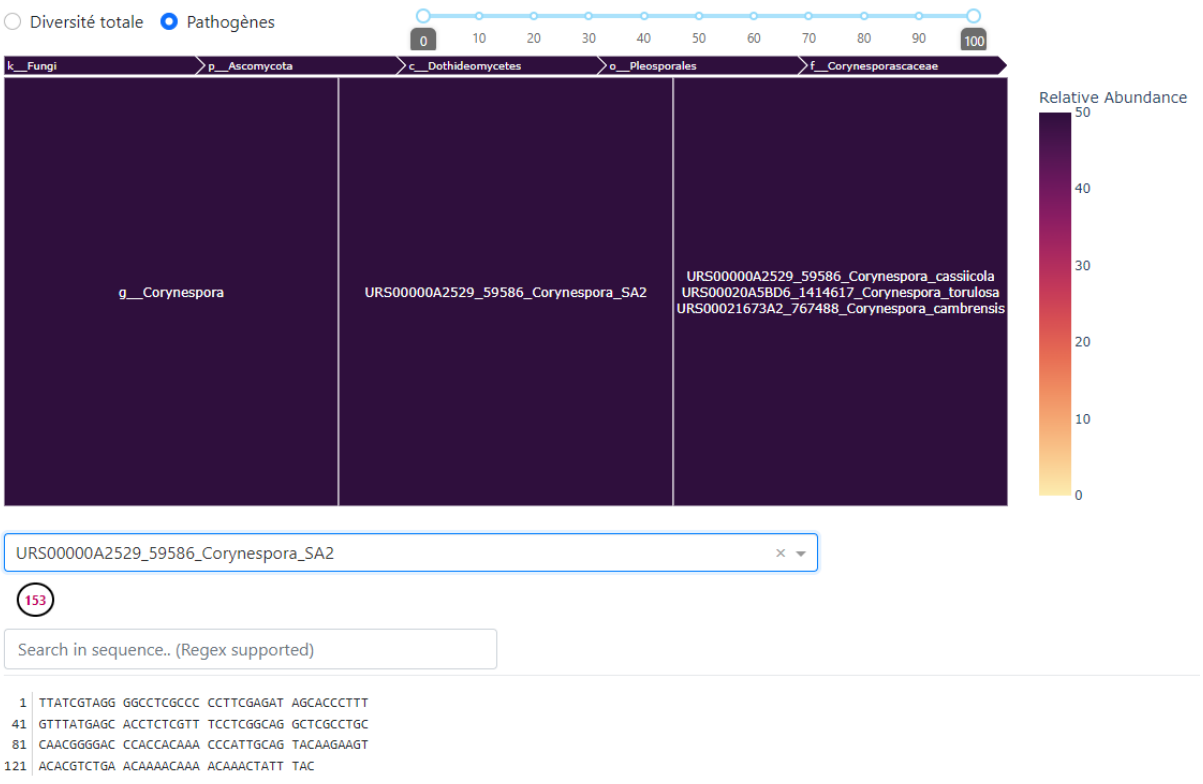

Figure 15. Menu « Visualisation des données » de l'application PhytoSHD – Zoom au niveau du genre

Architecture

Au niveau de la programmation, l'application PhytoSHD est construite avec un « back-end » (structure derrière l'application) et un « front-end » (ce qui est visuel). Un ordonnanceur tourne en permanence pour détecter les tâches lancées par l'utilisateur au niveau de l'interface (en rouge). Des Dockers (en mauve) sont autant de boîte autonome qui traite spécifiquement les données selon les types d'analyses. Par exemple, il y a un docker pour traiter les données provenant de la plateforme MiSeq et un Docker pour les données provenant de MinION. Cette structure permet de facilement ajouter d'autres types d'analyses. Les fichiers « config » et les bases de références sont à l'extérieur des Dockers pour facilement pouvoir les modifier en cas de besoin. Enfin, une base de données MongoDB permet de rassembler toute l'information relative aux métadonnées et aux résultats (vert).

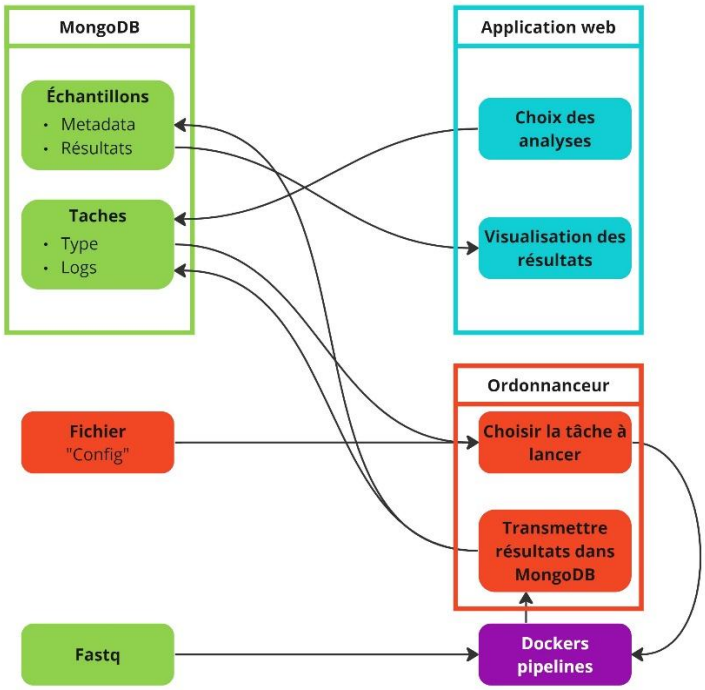

Figure 16. Schéma de l'architecture de l'application PhytoSHD

3.5. Analyse économique

Un bilan économique a été effectué en deux phases en collaboration avec le LEDP. Nous avons dans un premier temps déterminé une grille tarifaire permettant d'identifier les coûts en matériel et le temps humain pour la réalisation des analyses. Nous avons réalisé cette évaluation conjointement avec l'équipe du LEDP. Nous avons déterminé deux catégories de coûts. La première est associée à un contexte de basse saison au LEDP et la seconde est associée à un contexte de haute saison où le nombre d'échantillons traités hebdomadairement est supérieur, engendrant un coût unitaire

inférieur pour une même analyse (la quantité de consommables utilisés par analyse est moindre lorsque le volume d'échantillons à traiter simultanément est élevé). Pour chaque type d'analyse, un nombre d'échantillons à traiter simultanément a été identifié pour les deux catégories. Cette table de coûts est fournie en Annexe 1-8 (Tarification_SHD_CONV.pdf).

Elle est établie selon des coûts déterminés en janvier 2022. Dans un second temps, nous avons recueilli les détails des étapes techniques des approches techniques permettant d'établir un diagnostic. Au total, les 132 cas traités en 2019 ont été catégorisés.

En reliant les deux types de données, nous avons pu comparer le coût total de traitement et les temps techniques nécessaires pour les approches conventionnelles et SHD. La figure 17 présente une synthèse de ce bilan. Le coût de base utilisé pour le calcul des ressources humaines est un salaire journalier de 354$/jour (personnel technique – incluant les avantages sociaux)

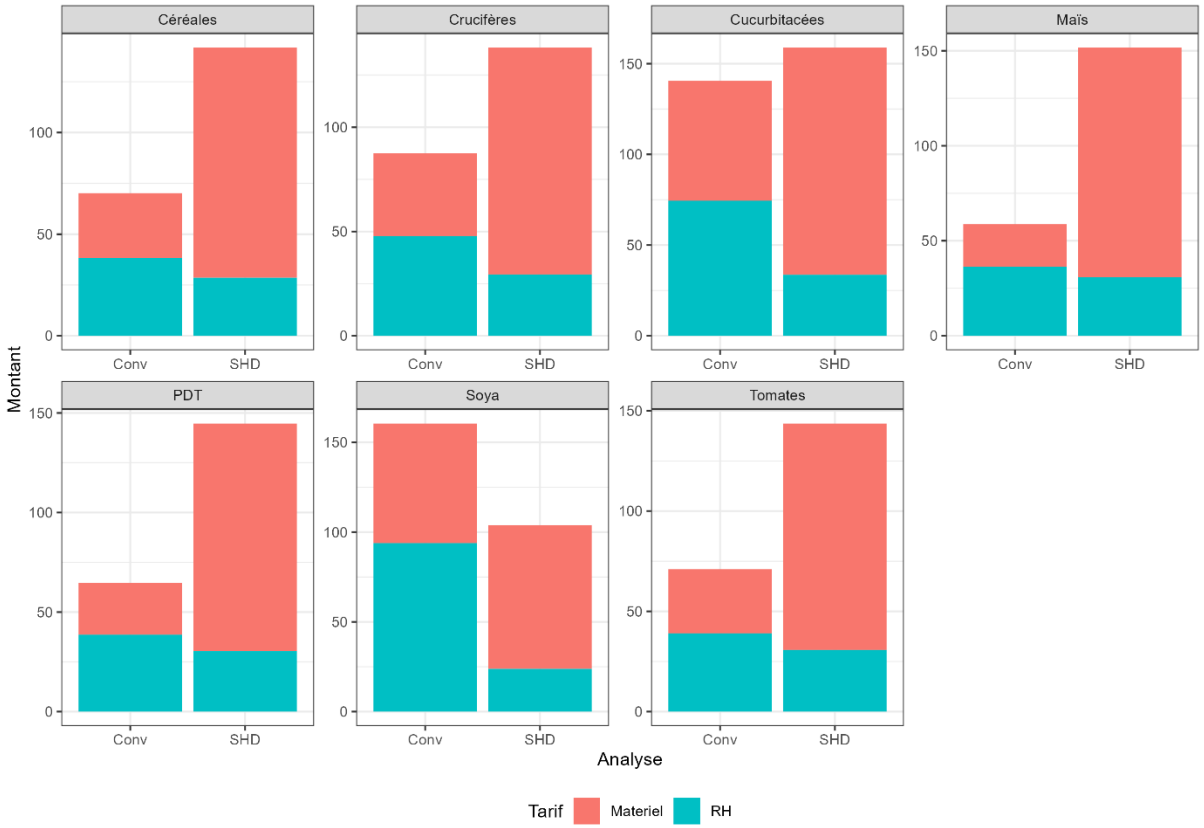

Figure 17. Coût d'analyse ($CA) par la méthode conventionnelle (Conv) ou SHD (SHD) en période estivale. Les calculs sont basés sur les cas reçus en 2019 (n=132). Pour l'approche SHD, les cas suspectés bactériens comptabilisent 1 seule cible, les cas fongiques 2 cibles et les cas bactériens ou fongiques 3 cibles (ex. BACTV4V5, ITS, EF1alpha)

NOTES : Les coûts mentionnés sont adaptés aux contraintes du LEDP et peuvent différer des coûts de mise au point ou d'opération au laboratoire d'écologie microbienne de l'IRDA. Les coûts d'approche conventionnelle ne prennent pas en considération les coûts associés à la gestion des témoins qui peuvent être importants pour des approches de microbiologie ou les détections quantitatives par qPCR. Le temps alloué à la formation du personnel n'est également pas pris en compte. À titre d'exemple, plusieurs mois sont nécessaires pour la formation du personnel pour l'identification visuelle des champignons (analyses conventionnelles) alors que la formation à la préparation d'un échantillon à séquencer ne prend que quelques jours. Cet aspect n'a pas été tenu en compte en raison de la variation de ce facteur selon le roulement de personnel. On peut tout de même supposer que les coûts associés au ressources humaines sont grandement sous-estimés au niveau des analyses conventionnelles.

4. TUTORIEL

Dans la première version de l'application Phyto_SHD un tutoriel décrivant les étapes pour l'importation, le traitement et la visualisation des données avait été implanté (Figure 18). Dans la version finale de l'application, l'étape d'importation ayant été simplifiée, et les étapes étant plus intuitives, il a été décidé d'inclure les instructions pour l'utilisation dans le modèle de fichier Excel qui sera utilisé pour l'importation des données. Ce fichier inclus les listes préétablies, les cultures, tissus, types d'analyses et symptômes ainsi qu'un onglet avec les instructions. Pour l'identification des noms des fichiers FASTQ, un programme (FastqFinder.exe) a été créé et transféré au LEDP (Figure 19).

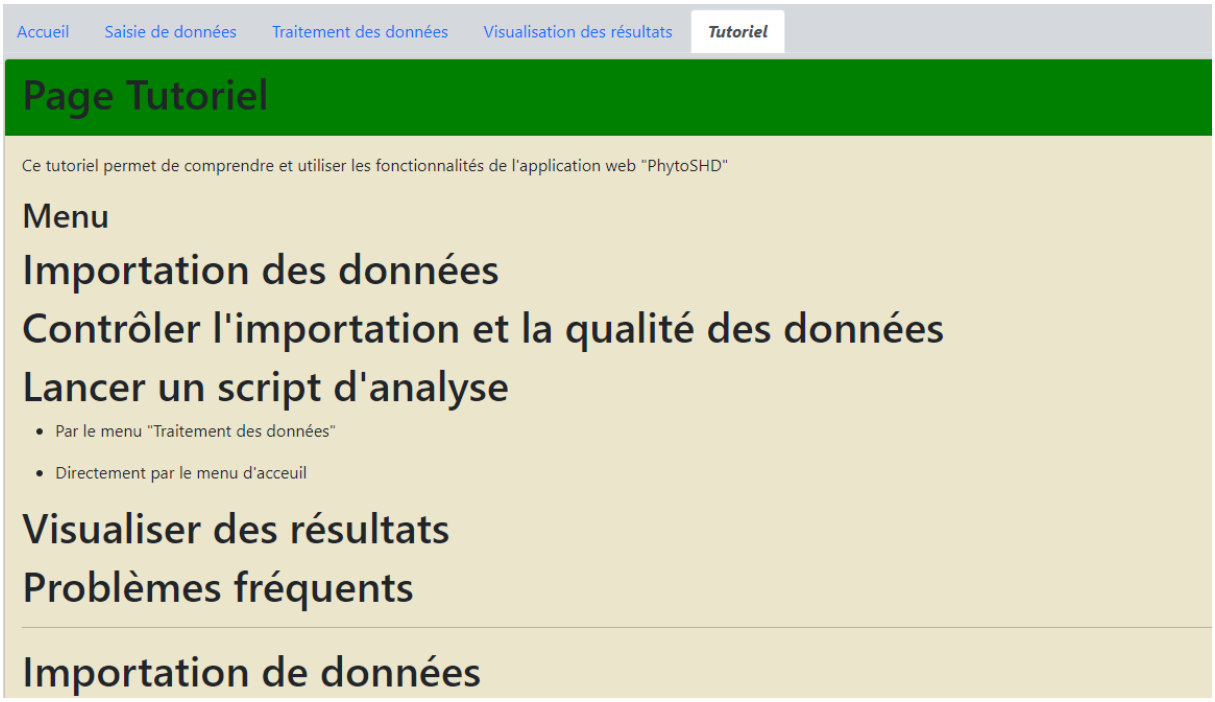

Figure 18. Tutoriel de l'application PhytoSHD

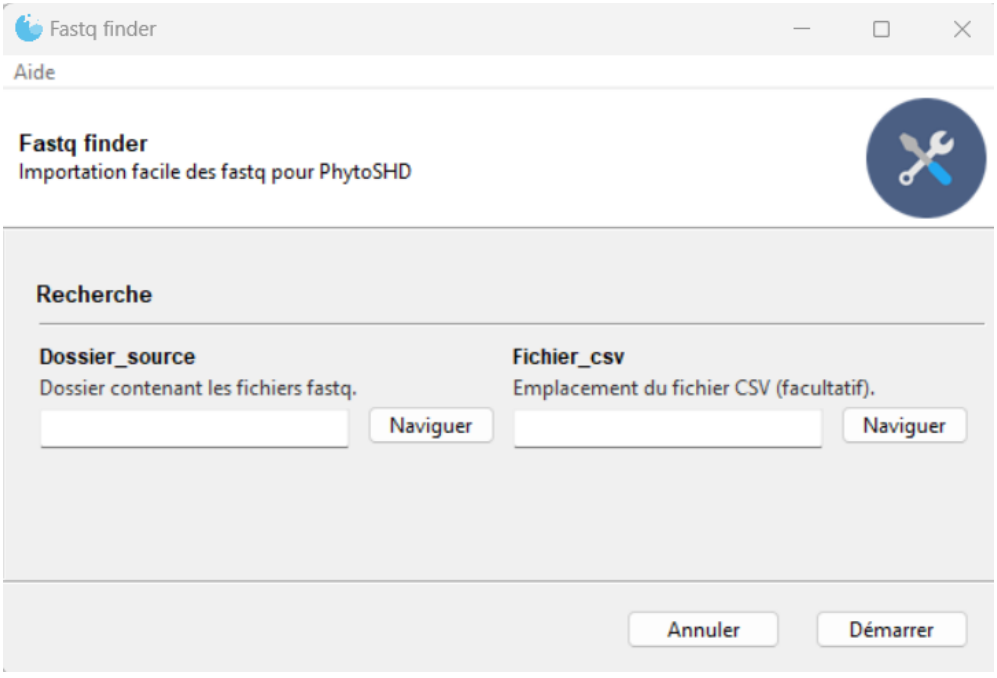

Figure 19. Application « FastqFinder » pour faciliter l'identification des fichiers FASTQ

5. ACTIVITÉS DE DIFFUSION ET TRANSFERT AUX UTILISATEURS

5.1. Diffusions

La liste des diffusions effectuées dans le projet est présentée dans le tableau 6

Tableau 6. Activités de diffusion

Nom de l'activité	Type de diffusion	Événement	Date	Public	Réf.
CPS	Présentation orale	CPS2021	05/2021	Chercheurs, professionnels	A2-1
SPPQ	Présentation orale	SPPQ2021	09/2021	Chercheurs, professionnels	A2-2
RAP-Cucurbitaceae	Présentation orale	RAP 2022	03/2022	Professionnels, agronomes	A2-3
Article	Article scientifique	Plants (journal)	09/2023	Chercheurs	A2-4

5.2. Transfert au LEDP

Durant le projet, plusieurs activités de formations ont été réalisées pour :

1. Former aux techniques d'extraction d'ADN pour des tissus végétaux et des racines. Nous avons produit deux fiches techniques permettant de documenter les mises au point réalisées pour établir les protocoles de référence. (Annexe1-1 et Annexe1-2). Nous avons organisé deux rencontres en 2020 à cet effet. Les fiches sont mentionnées dans la section annexe du rapport et disponibles en version PDF via un lien cloud.
2. Nous avons également formé le personnel du LEDP pour les techniques de préparation de librairies pour séquençage sur MiSeq et NanoMiSeq. Nous avons organisé des formations en 2022 et 2023 en incluant un protocole détaillé. (Voir mention en annexe 1-3; 1-4; 1-5).
3. Pour la dernière phase du projet, nous avons formé le personnel du LEDP pour l'utilisation de la plateforme web PhytoSHD pour importer, traiter et consulter les résultats obtenus à partir des données de séquençage à haut débit.
4. Enfin, nous avons organisé une activité de formation pour la préparation de librairie et le séquençage sur la plateforme MinION incluant un protocole détaillé (voir mention en annexe1-6).

5. Lors de la dernière phase du projet, 354 échantillons (cibles) destinés au séquençage à haut débit ont été réalisés pour nous assurer du transfert au personnel du LEDP. Cela inclut 210 échantillons (cibles) pour du MiSeq/NanoMiSeq et 144 pour du MinION

6. PERSPECTIVES

Base de données de références spécifiques : Lors de la compilation des cas diagnostiqués par groupe de culture, certains variants génétiques pouvaient présenter des différences au niveau du genre entre la base de données de références publiques et la base de données de références spécifiques. Ces différences s'expliquent la plupart du temps par un manque de séquences pour le groupe ciblé dans la base de données de références publiques, par des homologies proches ou des nomenclatures qui ont changé. Pour améliorer la base de données de références spécifiques, il serait pertinent d'ajouter les séquences références des genres Plectosphaerella, Acremonium, Ulocladium, Stenotrophomas, Pseudoperonospora, Pantoea, Serratia, Mycosphaerella, Aureobasidium, Stagonospora, Phaeosphaeria. En les intégrant dans notre base de données de références spécifiques, nous pourrons documenter les problématiques spécifiques pouvant survenir pour l'identification des variants génétiques (ASV) de ces genres.

MinION : Le traitement informatique effectué dans le cadre de ce projet, le développement de l'application PhytoSHD est spécifiquement conçu pour les données obtenues à partir des cellules R10. Initialement, nous avions élaboré des protocoles pour les cellules R9, mais le fournisseur des kits utilisés a décidé de ne plus les produire. Au cours de la dernière année du projet, nous avons donc ajusté nos protocoles afin de pouvoir utiliser les cellules R10 et les kits en version V14. Les données obtenues à partir des cellules R10 présentent une meilleure qualité par rapport à celles obtenues avec les cellules R9.

D'importants progrès ont récemment été réalisés dans le traitement en duplex des données brutes MinION (fast5/pod5), ce qui pourrait potentiellement améliorer davantage la qualité des séquences. Par conséquent, cela pourrait également accroître la précision des attributions taxonomiques pour certaines espèces bactériennes et fongiques.

7. PERSONNE-RESSOURCE

Richard Hogue, richard.hogue@irda.qc.ca
Thomas Jeanne, thomas.jeanne@irda.qc.ca

8. REMERCIEMENTS

Ce projet a été réalisé en vertu du sous-volet 2.2 du programme Prime-Vert 2018-2023 et il a bénéficié d'une aide financière du ministère de l'Agriculture, des Pêcheries et de l'Alimentation (MAPAQ).

9. RÉFÉRENCES

[1] Apprill A, Mcnally S, Parsons R, and Weber L. 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb Ecol 75: 129–37.
[2] Parada AE, Needham DM, and Fuhrman JA. 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol 18: 1403–14.
[3] Bokulich N a and Mills D a. 2013. Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. Appl Environ Microbiol 79: 2519–26.
[4] Bolyen E, Rideout JR, Dillon MR, et al. 2018. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. PeerJ Preprints.
[5] Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nature methods, 13: 581-583.
[6] DeSantis TZ, Hugenholtz P, Larsen N, et al. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72: 5069–72.
[7] Quast C, Pruesse E, Yilmaz P, et al. 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res 41: 590–6.
[8] Koljalg U, Nilsson RH, Taylor AFS, et al. 2013. Towards a unified paradigm for sequence-based identification of fungi. Mol Ecol 22: 5271–7.

10. ANNEXES

ANNEXE1
Liste des protocoles et fiches fournies au LEDP

1- Fiche comparaison d'extraction ADN de racines : QIASOILPRO.Racines.2021.6.pdf

2- Fiche comparaison kit d'extraction vs diversité bactérienne et fongique : QIASOILPRO.Sols.2021.6.pdf

3- Protocole de préparation de librairie pour MiSeq : Protocole-LEM-LIB-MiSeq-2022.04.pdf

4- Protocole de préparation de librairie pour NanoMiSeq : Protocole-LEM-LIB-MiSeq-2023.02.pdf

5- Protocole de préparation de librairie pour NanoMiSeq – OneStep : Protocole-LEM-LIB-MiSeq-OneStep-2023.02.pdf

6- Protocole de préparation de librairie pour MinION : Protocole-LEM-LIB-ONT-2023.02.pdf (Version R9/R10 chimie version 10) Protocole-LEM-LIB-ONT-2023.10.pdf (Version R10 chimie version 14)

7- Liste des index utilisés pour l'exaction en deux étapes pour MiSeq/NanoMiSeq: Index-LEM-MiSeq.pdf
8- Tarification_Conv_SHD.pdf
9- Références_Taxon.pdf

L'ensemble de ces fichiers électroniques ont été partagés avec le LEDP.

ANNEXE2
Preuves de diffusions

1- CPS :

2- SPPQ :

3- RAQ Cucurbitaceae :

4- Article ASVmaker : publié dans la revue « Plants » numéro spécial (Phytomicrobiome Research for Disease and Pathogen Management)

plants

MDPI

Article

ASVmaker: A New Tool to Improve Taxonomic Identifications for Amplicon Sequencing Data

Clément Plessis [1,2], Thomas Jeanne [1,2,*], Antoine Dionne [3], Julien Vivancos [3], Arnaud Droit [2] and Richard Hogue [1]

[1] Institut de Recherche et de Développement en Agroenvironnement, Québec, QC G1P 3W8, Canada
[2] Computational Biology Laboratory, CHU de Québec—Université Laval Research Center, Québec City, QC G1V 4G2, Canada
[3] Laboratoire d'Expertise et de Diagnostic en Phytoprotection, Ministère de l'Agriculture, des Pêcheries et de l'Alimentation du Québec (MAPAQ), Québec City, QC G1P 3W6, Canada
* Correspondence: thomas.jeanne@irda.qc.ca

Abstract: The taxonomic assignment of sequences obtained by high throughput amplicon sequencing poses a limitation for various applications in the biomedical, environmental, and agricultural fields. Identifications are constrained by the length of the obtained sequences and the computational processes employed to efficiently assign taxonomy. Arriving at a consensus is often preferable to uncertain identification for ecological purposes. To address this issue, a new tool called "ASVmaker" has been developed to facilitate the creation of custom databases, thereby enhancing the precision of specific identifications. ASVmaker is specifically designed to generate reference databases for allocating amplicon sequencing data. It uses publicly available reference data and generates specific sequences derived from the primers used to create amplicon sequencing libraries. This versatile tool can complete taxonomic assignments performed with pre-trained classifiers from the SILVA and UNITE databases. Moreover, it enables the generation of comprehensive reference databases for specific genes in cases where no directly applicable database exists for taxonomic classification tools.