

**Projet de fin d'étude**  
**Master Sciences & Techniques**  
**Gestion et Conservation de la Biodiversité**

**Caractérisation des gènes de virulence du**  
***botrytis cinerea* sur le fraisier**

**Présenté par : Asmae Masbah**

**Encadré par :**

- Pr. BOUCHAMMA El-Ouazna
- Mr. LABHILILI Mustapha,
- Mme BENTATA Fatiha

**Soutenu devant le jury composé de :**

**Le :19/06/2022**

- Pr. BOUCHAMMA El-Ouazna (FST-Fès)
- Dr. LABHILILI Mustapha (INRA-Rabat)
- Pr. MIKOU Karima (FST-Fès)
- Pr. DERRAZ Khalid (FST-Fès)
- Dr. BENTATA Fatiha (INRA-Rabat)

**Année universitaire : 2021/2022**

## *Avant-propos*

Le présent travail a été réalisé dans le cadre du Master Gestion et Conservation de la Biodiversité coordonné par le professeur Abderrahim Lazraq.

Il est le fruit d'un stage effectué au sein l'Institut National des Recherches Agronomiques de Rabat. Ce dernier comprend deux unités de recherche : Unité de recherche sur l'environnement et la conservation des ressources naturelles et l'Unité de recherche sur la biotechnologie. L'Unité de Biotechnologie élabore également des projets en partenariat avec les universités et les institutions de recherche. Ses relations de coopération nationale et internationale visent principalement :

- le renforcement de la recherche pour le développement agricole ;
- le renforcement de stratégies de recherche afin d'être compétitive à l'échelle internationale ;
- le développement de plateformes de recherche multi-institutionnelles
- le renforcement des capacités humaines par la formation et le perfectionnement des chercheurs et techniciens (formation diplômante et stages) ;
- l'encadrement et le transfert de technologie aux établissements nationaux ;
- l'apport de l'expertise et la formation continue ;
- le renforcement des structures et des équipements ;
- le renforcement de la documentation en périodiques, livres et logiciels informatiques et bio-informatiques ;
- les publications et participations aux manifestations scientifiques.

Mon stage de fin d'études a été effectué en 6 mois et dirigé par Madame Bouchamma El Ouazna, professeur de la Faculté des Sciences et Techniques de Fès, Madame Bentata Fatiha et Monsieur Mustapha Labhilili de l'Institut National des Recherches Agronomiques de Rabat.

## *Remerciements*

Ce manuscrit marque la fin de mes cinq ans d'études passés au sein de la Faculté des Sciences et Techniques de Fès. Bien plus qu'une formation de recherche, ce mémoire aura également été une fabuleuse aventure humaine. Ainsi, je tiens à exprimer toute ma gratitude aux nombreuses personnes qui ont rendu ce projet possible.

Je tiens tout d'abord à remercier mon encadrante, Madame Bouchamma El Ouazna, de m'avoir orienté tout au long de mon travail. Merci pour tous les conseils et tous nos échanges, merci pour cet enthousiasme et ce dynamisme communicatifs que vous mettez dans votre travail. Enfin, ce fut un plaisir de travailler avec vous tant pour vos qualités professionnelles qu'humaines.

Je remercie vivement Monsieur Labhilili Mustapha et Madame Bentata Fatiha pour m'avoir accueilli dans l'Unité de Bio-informatique et Biotechnologie et pour m'avoir encadré durant ce travail. Je les remercie également pour leur soutien et leur accompagnement.

Je remercie chaleureusement Madame Gaboun Fatima pour son soutien sans faille dans les nombreuses difficultés que j'ai pu rencontrer, soutien sans lequel bien peu de choses auraient été rendues possibles. Votre disponibilité et vos généreux secours au cours de certains de mes moments difficiles ont été d'une très grande qualité et d'un immense réconfort.

Merci aussi à Fatima Zahrae Lamkhath pour les travaux que nous avons menés ensemble et l'ambiance dans laquelle ils se sont déroulés.

Un grand merci également à Hamza Hraoui pour m'avoir supportée, soutenue et encouragée tout au long de cette période et de m'avoir fait bénéficier de ses précieux conseils.

Mes remerciements vont également à mes chers parents, mon frère et ma sœur grâce à qui, entre autres, je suis arrivée aujourd'hui à cette étape capitale et décisive de ma vie.

## *Résumé*

Le *botrytis cinerea* est un champignon nécrotrophe et polyphage capable de provoquer la pourriture grise sur plusieurs centaines d'espèces végétales. Les pertes engendrées par cette maladie sont importantes notamment sur des espèces économiquement importantes comme la fraise, la vigne ou encore la tomate.

Parmi les facteurs de virulence identifiés chez le *botrytis cinerea* se trouvent les gènes de virulence. Dans le présent travail il s'agit de 13 gènes de virulence collectés depuis NCBI, qui ont été ensuite classés selon des familles de gènes.

Des études comparatives supplémentaires ont été réalisées par le logiciel « MEGAX ». Nous avons étudié le fonctionnement des séquences nucléotidiques ainsi que les séquences protéiques via l'alignement, la détermination des fonctions des différents gènes, via « InterPro », leurs domaines conservés et la construction de l'arbre phylogénétique permettant ainsi de déterminer la distance entre les gènes. Le développement d'amorces, par la fonction « pick primer » sur NCBI, a été réalisé ensuite sur les séquences protéiques permettant leur amplification.

**Mots clés :** *Botrytis cinerea*, gènes de virulence, NCBI, MEGAX, séquences nucléotidiques, séquences protéiques, familles de gènes, alignement, domaines conservés, arbre phylogénétique, distance entre les gènes, développement d'amorces, amplification.

## *Liste des abréviations*

**ADN** : Acide DésoxyriboNucléique

**DNTP** : DésoxyriboNucléotide TriPhosphate

**DdNTP** : Di-DésoxyriboNucléotide TriPhosphate

**NGS**: Next Generation Sequencing

**ARN**: Acide Ribonucléique

**NCBI** : National Center for Biotechnology Information

**SNP**: Single Nucleotide Polymorphism

**MEGAX**: Molecular Evolutionary Genetics Analysis

**PCR**: Polymerase Chain Reaction

**SSR**: Simple Sequence Repeats

## *Liste des tableaux*

<b>Tableau 1</b> : Composition nutritionnelle de la fraise (Sources : USDA,2016 ; *Ciquial ANSES 2013). Error! Bookmark not defined.	
<b>Tableau 2</b> : Teneurs des différentes familles polyphénoliques chez la fraise (Source : Aprifel.fr) ..... Error! Bookmark not defined.	
<b>Tableau 3</b> : Paramètres choisis pour la recherche des SSRs.....	<b>33</b>
<b>Tableau 4</b> : Classification des gènes selon leur hôte et fonctions .....	<b>35</b>
<b>Tableau 5</b> : Classification des gènes de virulence en familles .....	<b>36</b>
<b>Tableau 6</b> : Classification des gènes de virulence selon les codes d'accèsion des séquences nucléotidiques.....	<b>37</b>
<b>Tableau 7</b> : Classification des gènes de virulence selon les codes d'accèsion des séquences protéiques .....	<b>39</b>
<b>Tableau 8</b> : Paramètres du choix du meilleur modèle de l'arbre phylogénétique sur MEGAX .....	<b>44</b>
<b>Tableau 9</b> : Description de quelques paires d'amorces des gènes étudiés générés par pickprimer. ..	<b>46</b>
<b>Tableau 10</b> : Les types SSR des différentes séquences des gènes de virulence du <i>Botrytis cinerea</i> .	<b>47</b>
<b>Tableau 11</b> : Classification des marqueurs SSR selon leur type .....	<b>48</b>

## Liste des figures

<b>Figure 1</b> : Structures d'une fleur de fraisier et d'une fraise .....	5
<b>Figure 2</b> : Ravageurs affectant le fraisier .....	10
<b>Figure 3</b> : Cycle de développement du <i>Botrytis cinerea</i> aux différents stages du fraisier (Duval, 1994) .....	12
<b>Figure 4</b> : Productions végétales infectées par le <i>Botrytis cinerea</i> .....	13
<b>Figure 5</b> : Le principe de la technologie Illumina (Raous, 2013).....	17
<b>Figure 6</b> : Séquençage <i>Paired-end</i> (Botero-Castro, 2014).....	18
<b>Figure 7</b> : Séquençage <i>Single-end</i> (Botero-Castro, 2014).....	18
<b>Figure 8</b> : Evolution des techniques de séquençage (Botero-Castro, 2014).....	19
<b>Figure 9</b> : Assemblage par référence (Botero-Castro, 2014).....	20
<b>Figure 10</b> : Site web NCBI ( <a href="https://www.ncbi.nlm.nih.gov/">https://www.ncbi.nlm.nih.gov/</a> ) .....	23
<b>Figure 11</b> : Paramètres de collecte des séquences nucléotidiques ( <a href="https://www.ncbi.nlm.nih.gov/">https://www.ncbi.nlm.nih.gov/</a> ) .....	25
<b>Figure 12</b> : Le logiciel MEGAX .....	26
<b>Figure 13</b> : Le logiciel CLC Genomics Workbench .....	26
<b>Figure 14</b> : Assistant de licence permettant d'accéder aux options de licence (depuis CLC Genomics Workbench Manual) .....	27
<b>Figure 15</b> : Coûts du CLC Genomics Workbench.....	27
<b>Figure 16</b> : Base de données InterPro ( <a href="https://www.ebi.ac.uk/interpro/">https://www.ebi.ac.uk/interpro/</a> ).....	28
<b>Figure 17</b> : Alignement multiple sur MEGAX .....	28
<b>Figure 18</b> : Illustration des étapes d'alignement multiple des séquences de nucléotides .....	29
<b>Figure 19</b> : Alignement par Muscle sur MEGAX .....	29
<b>Figure 22</b> : Développement d'amorces via NCBI ( <a href="https://www.ncbi.nlm.nih.gov/">https://www.ncbi.nlm.nih.gov/</a> ).....	31
<b>Figure 23</b> : Présentation de l'interface du développement d'amorces sur NCBI ( <a href="https://www.ncbi.nlm.nih.gov/">https://www.ncbi.nlm.nih.gov/</a> ).....	31
<b>Figure 24</b> : Illustration des étapes d'identification des amorces ( <a href="https://www.ncbi.nlm.nih.gov/">https://www.ncbi.nlm.nih.gov/</a> )	32
<b>Figure 25</b> : Présentation de l'interface de l'alignement multiple des différentes séquences nucléotidiques sur MEGAX .....	38
<b>Figure 26</b> : Présentation de l'interface de l'alignement multiple des différentes séquences protéiques sur MEGAX .....	40

<b>Figure 27</b> : Présentation de l'interface de l'alignement par Muscle des différentes séquences nucléotidiques sur MEGAX .....	40
<b>Figure 28</b> : Présentation de l'interface de l'alignement par Muscle des différentes séquences protéiques sur MEGAX.....	41
<b>Figure 29</b> : BlastP des séquences protéiques des 13 gènes de virulence du <i>Botrytis cinerea</i> .....	41
<b>Figure 30</b> : Alignement des séquences protéiques de la famille Bcg sur MEGAX.....	42
<b>Figure 31</b> : Domaines conservés des séquences protéiques de Bcg générés par Muscle sur MEGAX .....	42
<b>Figure 32</b> : Alignement des séquences protéiques de la famille Bcpg sur MEGAX.....	42
<b>Figure 33</b> : Domaines conservés des séquences protéiques de la famille Bcpg par Muscle sur MEGAX .....	42
<b>Figure 34</b> : Alignement des séquences protéiques de la famille BcATG sur MEGAX.....	42
<b>Figure 35</b> : Alignement des séquences protéiques de la famille BcATG par Muscle sur MEGAX .....	43
<b>Figure 36</b> : Alignement des séquences nucléotidiques de la famille BcVEL sur MEGAX.....	43
<b>Figure 37</b> : Alignement des séquences protéiques de la famille BcVEL par MUSCLE sur MEGAX .....	43
<b>Figure 38</b> : Illustration du paramètre du choix du meilleur modèle de l'arbre phylogénétique .....	44
<b>Figure 39</b> : Paramètres à modifier pour le traçage de l'arbre phylogénétique sur MEGAX.....	45
<b>Figure 40</b> : Arbre phylogénétique des différentes séquences de nucléotides générés par MEGAX .....	45
<b>Figure 41</b> : Ajustement des différents paramètres des amorces .....	46
<b>Figure 42</b> : Fichier.statistics décrivant les marqueurs SSR.....	49

# Table des matières

<b>Introduction générale</b> .....	<b>2</b>
A. Le fraisier .....	4
I. Histoire : .....	4
II. Biologie de la plante : .....	4
<b>1. Botanique</b> : .....	<b>4</b>
<b>2. Génétique et qualité nutritionnelle</b> : .....	<b>5</b>
a. Etude génomique du fraisier : .....	5
b. Marqueurs génétiques et moléculaires : .....	6
i. Les marqueurs microsatellites ou SSR : .....	6
ii. Les marqueurs SNP : .....	6
c. Qualité nutritionnelle : .....	7
III. Cycle de développement du fraisier : .....	8
IV. Production du fraisier : .....	8
<b>1. Production à l'échelle mondiale</b> : .....	<b>8</b>
<b>2. Production nationale</b> : .....	<b>8</b>
V. Ravageurs et maladies : .....	9
B. Botrytis cinerea : .....	11
I. Nomenclature et systématique : .....	11
II. Cycle de développement de Botrytis cinerea : .....	11
III. Gamme d'hôte : .....	13
IV. Pouvoir pathogène du Botrytis cinerea : .....	13
V. Importance économique de la maladie : .....	14
VI. Facteurs influençant le développement du champignon : .....	14
C. Analyse génomique .....	15
I. Le génome et les variations génétiques : .....	15
II. Données de séquençage des acides nucléiques : .....	15
<b>1. Séquençage Sanger (séquençage de première génération)</b> : .....	<b>15</b>
<b>2. Séquençage de seconde génération</b> : .....	<b>15</b>
a. Séquençage par Illumina : .....	16
• Préparation de la banque d'ADN : .....	16

	• Amplification en pont : .....	16
	• Séquençage par terminateurs réversibles : .....	16
b.	Séquençages Paired-end et Single-end : .....	18
	• Définition du séquençage <i>Paired-end</i> : .....	18
	• Définition du séquençage <i>Single-end</i> : .....	18
<b>3.</b>	<b>Séquençage de troisième</b>	
	<b>génération : .....</b>	<b>18</b>
III.	Analyse génomique après séquençage NGS : .....	19
<b>1.</b>	<b>Assemblage des séquences génomiques : .....</b>	<b>19</b>
a.	Assemblage par référence ou <i>mapping</i> : .....	20
b.	Assemblage de novo : .....	20
<b>2.</b>	<b>Annotation génomique : .....</b>	<b>21</b>
a.	Annotation relationnelle : .....	21
b.	Annotation fonctionnelle : .....	21
c.	Annotation structurale : .....	21
I.	Récupération des génomes via le programme NCBI : .....	23
<b>1.</b>	<b>Présentation du programme : .....</b>	<b>23</b>
	• Format GenBank : .....	24
	• Format FASTA : .....	24
<b>2.</b>	<b>Collecte des gènes de virulence du <i>Botrytis</i> : .....</b>	<b>24</b>
II.	Les outils bio-informatiques : .....	25
<b>1.</b>	<b>Logiciel MEGAX : .....</b>	<b>25</b>
<b>2.</b>	<b>Logiciel CLC Genomics Workbench : .....</b>	<b>26</b>
III.	Analyse des séquences nucléotidiques et protéiques des différents gènes : .....	27
<b>1.</b>	<b>Détermination des fonctions des différents gènes : .....</b>	<b>27</b>
<b>2.</b>	<b>Alignement multiple : .....</b>	<b>28</b>
<b>3.</b>	<b>Alignement par MUSCLE sur MEGAX : .....</b>	<b>29</b>
<b>4.</b>	<b>Les domaines conservés : .....</b>	<b>30</b>
<b>5.</b>	<b>Arbre phylogénétique : .....</b>	<b>30</b>
<b>6.</b>	<b>Développement des amorces : .....</b>	<b>30</b>
<b>7.</b>	<b>Recherche des marqueurs SSR : .....</b>	<b>32</b>
I.	Classification des gènes de virulence collectés : .....	35

II.	Les domaines conservés, les amorces associées et les arbres phylogénétiques des différents gènes de virulence du botrytis cinerea : .....	37
1.	<b>Alignement multiple :</b> .....	<b>37</b>
2.	<b>Alignement par Muscle :</b> .....	<b>40</b>
a.	Alignement des séquences nucléotidiques : .....	40
b.	Alignement des séquences protéiques : .....	41
3.	<b>Alignement des séquences protéiques des différentes familles de gènes :.</b>	<b>42</b>
a.	Famille de gènes Bcg : .....	42
b.	Famille de gènes Bcpg : .....	42
c.	Famille de gènes BcATG : .....	42
d.	Famille des gènes BcVEL : .....	43
4.	<b>Arbre phylogénétique :</b> .....	<b>43</b>
5.	<b>Le développement d'amorces :</b> .....	<b>45</b>
6.	<b>Les marqueurs SSR:</b> .....	<b>47</b>
<b>Conclusion générale</b> .....		<b>49</b>
<b>Références</b> .....		<b>50</b>
<b>Annexes</b> .....		<b>55</b>

# **Introduction générale**

## Introduction générale

Le fraisier est une plante herbacée vivace de grande qualité nutritive, riche en vitamine C, B1 et sels minéraux. Le croisement de deux espèces de fraisiers sauvages, *Fragaria chiloensis* (originaire du Chili) et *Fragaria virginiana* (originaire de l'est de l'Amérique du Nord) a produit l'hybride fertile utilisé pour la production commerciale de fraises (Darrow, 1966). Créé en Europe en 1766, *Fragaria x ananassa Duchesne* s'est diversifié au fil du temps en plusieurs variétés et cultivars. Il existe plus de 100 cultivars couramment utilisés pour la culture commerciale à travers le monde. Les cultivars se distinguent par leur résistance aux maladies, leur constitution, la couleur de leurs fleurs, le moment de fructification ainsi que la taille, la forme, la fermeté, la couleur et le goût des fruits (Bostanian et coll., 2005).

Le fraisier est en permanence exposé à des infections par des agents pathogènes. Néanmoins, il a élaboré au cours de son évolution des moyens de défense qui lui permet de s'adapter afin de contrer efficacement les agressions. Cependant lorsque les mécanismes de résistance des plantes sont inefficaces contre un agent pathogène, celui-ci réussit à infecter son hôte et il en résulte le développement d'une maladie (Selon l'Organisation Mondiale de la Santé).

La pourriture grise est l'une des deux maladies principales du fraisier avec l'oïdium. Les dégâts sur les fruits et sur la partie supérieure du rhizome sont les signes que l'infection est grave.

Le *Botrytis cinerea*, agent responsable de la pourriture grise du fraisier, peut entraîner la destruction partielle ou totale du fraisier. Il peut s'attaquer à différents stages de développement du fraisier et l'infection par les conidies peut se produire durant toute la saison de croissance. Ce champignon présente différents gènes de virulence présentant différentes fonctions et peuvent être spécifiques pour certains hôtes.

L'objectif de cette étude est de caractériser les différents gènes de virulence du *Botrytis cinerea*, responsable de la pourriture grise du fraisier, afin de développer les amorces spécifiques pour chaque gène de virulence. Cet objectif se décline en trois questions scientifiques :

- (1) Quelles sont les différents gènes de virulence et quels sont leur hôte ?
- (2) Quelles sont les différentes fonctions de chaque gène ? Quelles sont les amorces spécifiques pour chaque gène ?
- (3) Comment générer un arbre phylogénétique ? Et comment visualiser son arbre afin d'estimer le pourcentage d'homologie entre les différents gènes ?

# **Partie I**

## **Etude bibliographique**

## **A. Le fraisier**

### **I. Histoire :**

Depuis la Préhistoire, l'homme consomme des fraises issues de fraisiers sauvages, les fraises des bois (*Fragaria Vesca*). Les premières fraises mises en culture datent du XIV<sup>e</sup> siècle avec la plantation de 12000 fraisiers dans le jardin du Louvre. Il s'agissait là de fruits très parfumés mais avec un très petit calibre.

Au XVI<sup>e</sup> siècle, les explorateurs qui découvrirent l'Amérique s'émerveillèrent devant la grosseur des fruits. C'est probablement Jacques Cartier qui introduisit en France le fraisier écarlate ou fraisier de Virginie (*Fragaria Virginiana*). Il se répandit en Angleterre et en France notamment en Bretagne.

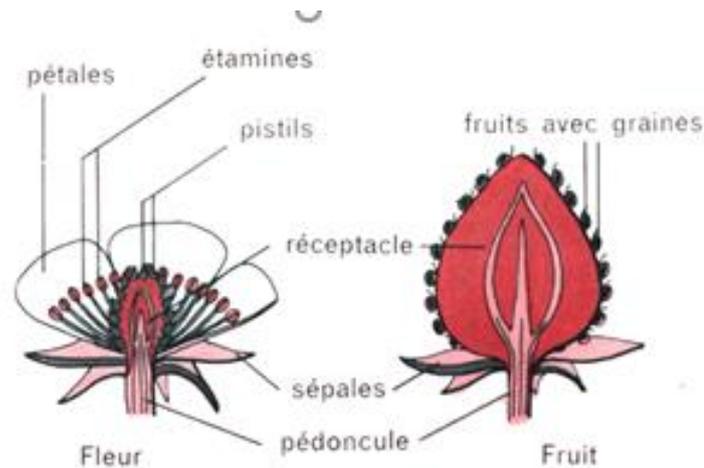
Au XVIII<sup>e</sup> siècle, la présence simultanée de *Fragaria Virginiana* et *F. chiloensis*, une autre espèce d'origine américaine, a permis d'obtenir les plants cumulant les avantages des deux espèces, la capacité de production de stolons de *F. virginiana* et l'arôme et le bon calibre de *F. chiloensis*. En 1766, Duchesne est le premier à émettre l'hypothèse que ces nouvelles plantes étaient des hybrides interspécifiques issus de l'intersection entre *F. virginiana* et *F. chiloensis* en les appelant *F. ananassa* car il trouvait que le fruit sentait l'ananas (Gaston, 2010).

### **II. Biologie de la plante :**

#### **1. Botanique :**

Le fraisier (*Fragaria × ananassa*) est une plante de petite taille herbacée vivace appartenant à la famille botanique des Rosacées. Les rameaux du fraisier sont des stolons qui s'enracinent en produisant des rosettes de feuilles afin de produire un nouveau pied. Les fleurs sont composées de cinq sépales, cinq pétales, une vingtaine d'étamines et plusieurs carpelles. Le fruit est un faux fruit de couleur rouge ou jaune blanchâtre, ce qui dépend des variétés, d'une forme ovoïde charnu. Les vrais fruits sont les akènes qui sont des minuscules fruits secs. Chacun contient une graine et pourrait donc donner vie à une nouvelle plante.

A l'extrémité de chaque akène, on distingue le style qui se termine par le stigmate sur lequel se posent les grains de pollen (Gaston, 2010).



**Figure 1 :** Structures d'une fleur de fraisier et d'une fraise

L'inflorescence est une cyme bipare à croissance déterminée. La fleur primaire, la plus grande, produira le plus gros fruit de la hampe. Chez les variétés modernes, des fleurs secondaires se développent sur chacune des deux ou trois branches formées de façon opposée et en dessous de la fleur primaire sur l'axe floral principal. Elles présentent un stade de différenciation moins avancé, qui correspond à cinq-huit jours de retard par rapport au développement de la fleur primaire chez le diploïde (Hollender et al. 2012).

## 2. Génétique et qualité nutritionnelle :

### a. Etude génomique du fraisier :

Le genre *Fragaria* regroupe 20 espèces sauvages de différents niveaux de ploïdie, du diploïde (le cas du fraisier des bois *F. vesca* ( $2n=2x=14$ )) à octoploïde ( $2n=8x=56$ ). Une espèce décaploïde, *F. iturupensis*, a très récemment été décrite (Hummer et al., 2009). Ainsi, parmi les 20 espèces de fraisier, 11 sont diploïdes, cinq tétraploïdes, une hexaploïde, deux octoploïdes et une décaploïde.

Un consortium scientifique américano-anglo-ibérico-français a mis 18 mois pour décoder le génome de la fraise des bois (*Fragaria vesca*). Cette dernière possède un "petit" génome composé de deux jeux de 7 chromosomes ; tandis que toutes les autres possèdent un génome hypertrophié (jusqu'à huit jeux de 7 chromosomes). Le génome de la fraise sauvage comprend environ 35 000 gènes (Gaston, 2010).

Le génome octoploïde du fraisier cultivé ( $2n=8x=56$ ), *F. x ananassa*, compte parmi les plus complexes des espèces cultivées. En contraste, la taille de base du génome du fraisier diploïde ( $2n=2x=14$ ), 200 Mb, compte parmi les plus petites des espèces cultivées. La taille de la plante est petite, le temps de régénération est rapide puisqu'un cycle de vie peut être complet en quatre mois, et l'espèce est

facilement transformable. De plus, sa faculté de produire des stolons permet la multiplication végétative d'un génotype, les séquences de génome du fraisier sont disponibles sur les bases de données National Center for Biotechnology Information (NCBI).

Parmi les espèces diploïdes, *F. vesca* ( $2n=2x=14$ ) présente des caractéristiques qui en font un modèle très intéressant pour le genre *Fragaria*, mais aussi pour d'autres membres de la famille des Rosacées (Slovin et al., 2009). Ainsi, cette espèce est développée en tant que système efficace pour étudier la génétique et la génomique des fraisiers mais également des autres espèces de la famille des Rosacées (Rabinowicz and Slovin, 2007 ; Shulaev et al, 2008).

Des études de cartographie comparée ont déjà montré qu'il était possible de transférer des marqueurs développés sur le fraisier diploïde aux autres genres de la famille des Rosacées, comprenant des arbres (pomme, poirier et cerisier) et le rosier ainsi que le fraisier cultivé octoploïde (Sargent et al, 2007 ; Rousseau-Gueutin et al, 2008).

#### **b. Marqueurs génétiques et moléculaires :**

Un marqueur génétique est une séquence d'ADN repérable spécifiquement. En contrôle du transfert de gène, le marqueur est un gène associé au gène d'intérêt, codant une caractéristique détectable facilitant ainsi le repérage des cellules auxquelles la transgénèse a réussi. La détection d'un marqueur génétique peut s'effectuer par hybridation d'une sonde complémentaire, ou par son expression phénotypique. Les marqueurs peuvent être de différentes formes : STS, RFLP, microsatellites, SNP, EST, gènes ...

##### **i. Les marqueurs microsatellites ou SSR :**

Les SSR (Simple Séquence Repeats) ou marqueurs microsatellites sont des séquences répétées en tandem d'un motif de 2 à 10 paires de bases (Pb) et de taille généralement inférieure à 100 pb. Ces microsatellites offrent un polymorphisme très élevé dû à des variations dans le nombre de motifs répétés. Des amorces encadrant le microsatellite permettent d'amplifier des fragments de taille différente. Ce polymorphisme de taille peut être révélé par une migration électrophorétique sur gel d'acrylamide en condition dénaturante.

##### **ii. Les marqueurs SNP :**

Les marqueurs SNP (Single Nucleotide Polymorphism) : polymorphisme d'un seul nucléotide. Variation stable de la séquence d'ADN génomique, portant sur une seule base toutes les 100 à 300 bases environ du génome affectant au moins 10% de la population. Beaucoup de SNP n'ont pas d'implications fonctionnelles mais ils définissent un locus unique dans le génome et sont polymorphes.

### c. Qualité nutritionnelle :

Les feuilles du fraisier contiennent des acides organiques tels que l'acide ascorbique (vitamine C), des composés phénoliques, notamment des flavonoïdes de type flavones, des tanins ellagiques, une huile essentielle constituée de terpénoïdes qui sont des monoterpènes (linalol, alpha-terpinéol) et le salicylate de méthyle (Philippe, 2016).

La fraise contient 90% de l'eau dans la matière fraîche alors que les principaux nutriments sont présents en faible quantité (tableau 1). Ce fruit dispose également de fibres qui régulent le taux de sucre dans le sang, riche en manganèse, vitamine B9 et vitamine C d'une teneur de 65mg/100g ce qui est analogue à la valeur trouvée chez l'orange et le kiwi. Les polyphénols représentés par les flavonoïdes et les acides phénoliques qui ont une activité antioxydante réalisée via des réactions non enzymatiques afin de réguler la présence de radicaux libres et notamment des ROS (Philippe, 2016).

**Tableau 1 :** Composition nutritionnelle de la fraise (Sources : USDA, 2016 ; \*Ciquel ANSES 2013).

Composants	Teneur moyenne (g/100g)
Eau	90,95
Glucides	4,89
Fibres	2
Acides organiques	1,05
Protéines	0,67
Lipides	0,30
AG insaturés	0,198
AG saturés	0,015
Vitamine C	58,8
Glucose	1,99
Fructose	2,44
Saccharose	0,47
Acide Citrique	0,82*
Acide Malique	0,24*

**Tableau 2 :** Teneurs des différentes familles polyphénoliques chez la fraise (Source : Aprifel.fr)

Polyphénols	Quantité (pour 100g MF)
Flavonoïdes	84.47 mg
Acides phénoliques	12.74 mg
Stilbènes	0.35 mg
Polyphénols totaux	97.56 mg

### III. Cycle de développement du fraisier :

Le fraisier se multiplie végétativement en produisant des stolons qui sont des ramifications dont les deux premiers entre-nœuds sont anormalement allongés. Ces stolons sont produits par les bourgeons axillaires qui selon les conditions environnementales produisent des ramifications secondaires de la tige principale ou des stolons. De nombreux clones d'un même génotype peuvent être obtenus à partir des stolons. Les conditions photo-thermiques affectent la production des stolons qui est en général stimulé par des jours longs (>14-16h) et des températures élevées (>17-20°C) (OM Heide, 2013).

La reproduction sexuée, pérennisée par la graine, est possible mais peu fréquente. Elle est utilisée pour créer de nouvelles variétés (Philippe, 2016).

### IV. Production du fraisier :

#### 1. Production à l'échelle mondiale :

Selon les chiffres publiés par l'Organisation des Nations Unies pour l'alimentation et l'agriculture, la production mondiale de la fraise a augmenté, entre 2008 et 2018, de 39.4%. En 2008, la Chine a produit 1,9 million de tonnes de ce fruit. En 2018, ce chiffre a été dépassé à 3 millions de tonnes.

En Turquie, les producteurs ont cultivé 261.000 tonnes de fraise en 2008. En 2018, ce chiffre a été dépassé à 441.000 tonnes. Les États-Unis produisent le 2<sup>ème</sup> plus grand nombre de fraises au monde.

En Europe, c'est l'Espagne qui produit le plus de fraises. Avec plus de 281.000 tonnes en 2008, ce pays est toujours en concurrence avec d'autres pays. Ceux-ci comprennent le Maroc, l'Égypte, la Turquie et le Mexique. Ce volume était passé à près de 345.000 tonnes en 2018 (AgriMaroc, 2020).

#### 2. Production nationale :

D'après des études réalisées en 2014 par l'Institut Agronomiques et Vétérinaires, le fraisier a été introduit au Maroc en 1930 dans le but était de faire des essais afin de diversifier la production nationale. Les sols adaptés, l'eau abondante et de bonne qualité, la main d'œuvre qualifiée et l'incitation aux investissements ont permis l'extension du fraisier aux périmètres irrigués du Loukkos et du Gharb. Les superficies occupées par le fraisier sont actuellement de presque 3300 ha avec 80% au Loukkos.

Le Maroc a détenu une part de marché de 2% dans les exportations mondiales en 2009 et il a été classé 11<sup>ème</sup> exportateur de la fraise ; il exporte principalement vers la France, l'Espagne, Allemagne et le Royaume Uni. En 2013 la production nationale a été de 145 mille tonnes/an. Au Maroc, de nombreuses variétés ont été plantés (Camarosa, Fortuna, Splendor, Ventana, Sabrina, Lusa, Festival, etc), avec dominance de Festival et Camarosa, ensuite viennent Splendor et Fortuna.

#### V. **Ravageurs et maladies :**

La production intensive de la culture du fraisier a mené à l'accroissement d'agents pathogènes dans le sol dont le botrytis (la pourriture grise), l'oïdium et l'anthracnose font partie. Ces champignons sont très nuisibles et néfastes pour le fraisier ce qui fait l'objet de nombreuses recherches depuis bon nombre d'années. De plus, le bromure de méthyle utilisé pour traiter ces agents pathogènes est graduellement interdit à l'échelle mondiale (Haddaoui, 2020).

En 2012-2013, les principales maladies du fraisier ont été l'oïdium causé par *Sphaerotheca macularis*, la pourriture grise causée par *Botrytis cinerea*, la maladie des taches pourpres causée par *Ramularia sp.* et l'anthracnose causée par *Colletotrichum fragariae* (IAV, 2014).

Le fraisier peut même être attaqué même par des virus, des bactéries, des acariens, des nématodes, des ravageurs ou les mauvaises herbes (Martin et al., 2004 ; Li and Yang, 2011).

Ainsi, plusieurs facteurs favorisent le développement des maladies fongiques et l'apparition de plusieurs pathogènes responsables de la dépréciation des fraises, des symptômes courants de la maladie, de la résistance des plantes ainsi que des méthodes de dépistage.

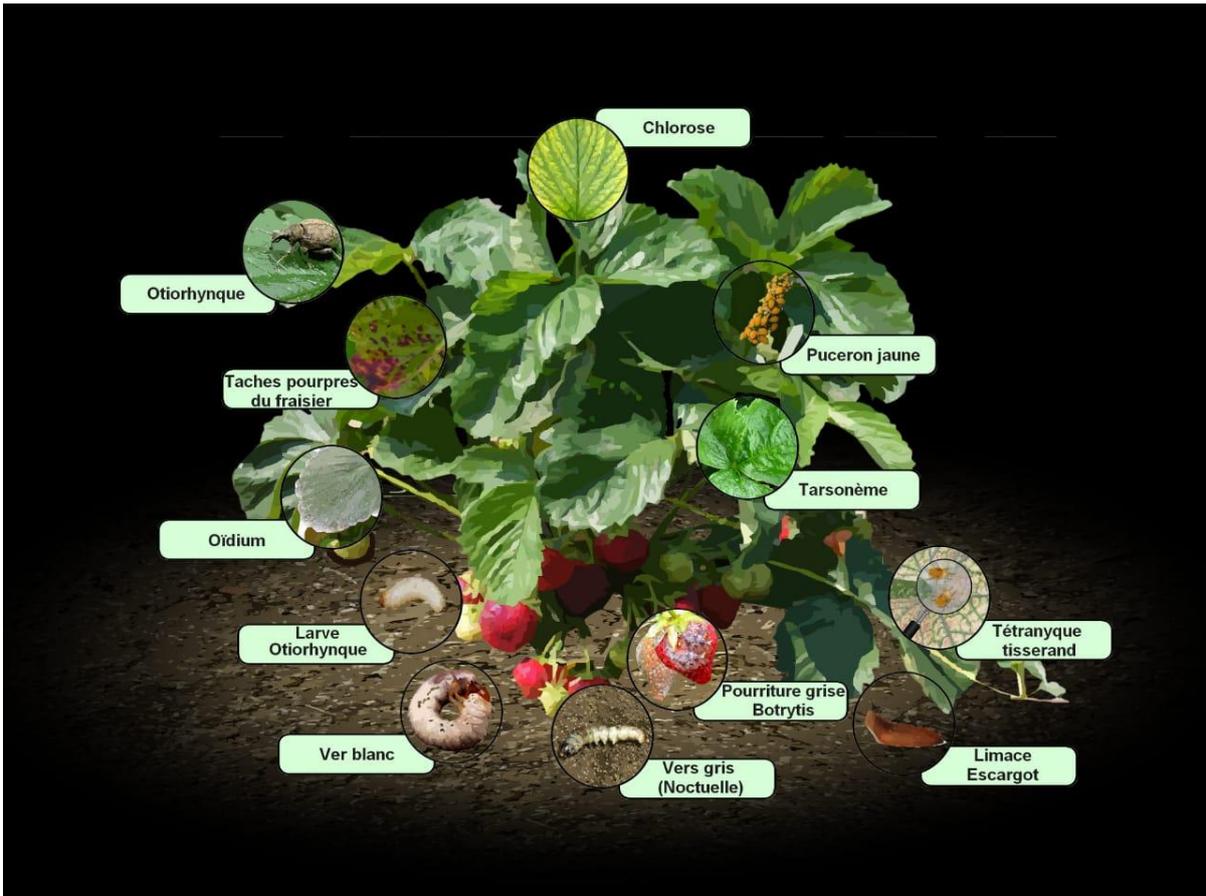


Figure 2 : Ravageurs affectant le fraisier

## **B. Botrytis cinerea :**

### **I. Nomenclature et systématique :**

Le nom *Botrytis cinerea* a été donné par Persoon en 1801 à un agent pathogène de la vigne. Le terme *botrytis* a été forgé à partir du grec ce qui signifie « grappe de raisin » et du suffixe néolatin -*itis* indiquant les maladies. L'épithète spécifique *cinerea* vient du latin signifiant « cendre » en raison de la couleur cendrée des spores (Barnett et Hunter, 1998).

*Botrytis cinerea* est un champignon pathogène, polyphage et ubiquiste, il est l'agent responsable de la pourriture grise sur de nombreuses plantes cultivées et sauvages.

Le nom *Botrytis cinerea* a été proposé par Elias Magnus Fries, le botaniste suédois qui fut le fondateur de la systématique des champignons. *Botrytis cinerea* connaît une double classification selon la forme sous laquelle il se reproduit :

- Forme anamorphe (imparfaite) : *Botrytis cinerea* (asexuée), appartenant à la classe des Hyphomycètes, ordre des Moniliales et famille des Moniliaceae.
- Forme téléomorphe (parfaite) : *Botrytinia fuckeliana* (sexuée) appartenant à la classe des Discomycètes, ordre des Léotiales et famille des Sclerotiniaceae.

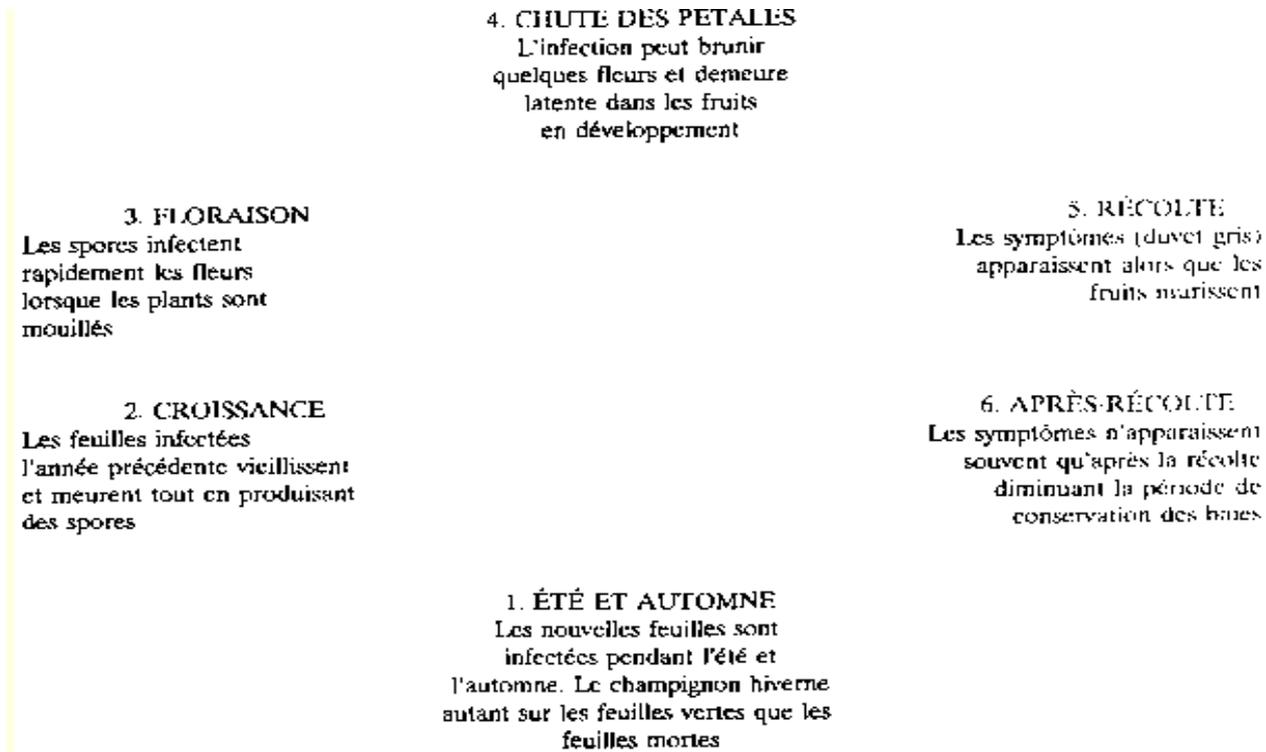
### **II. Cycle de développement de Botrytis cinerea :**

Le *Botrytis cinerea* fructifie en donnant des conidies, qui prennent une part importante dans la dissémination du champignon, quand les conditions deviennent favorables pour son développement. Le développement des conidies se fait par la production de conidiophores dressés en touffes souvent étendues, constituant un feutrage intense gris (Holz et al., 2004).

Lorsque les conditions deviennent défavorables au développement de mycélium et de conidies, des sclérotés se forment. Ils sont constitués par du mycélium agrégé blanchâtre. En vieillissant, ils durcissent et deviennent noirâtres (figure 4). Ils sont composés d'un cortex de cellules épaisses formant une mince barrière de cellules pseudo-parenchymateuses et d'une large medulla centrale composée d'hyphes filamenteux (Coley-Smith, 1980).

Dans le cas des cultures en plein champs, les conidies sont produites dès le printemps et dans le cas des cultures sous serre elles peuvent être produites en continu. Ainsi leur libération est favorisée par un climat humide ensuite elles sont transportées par la pluie, le vent ou les insectes.

Le cycle de développement du *Botrytis cinerea* passe par plusieurs étapes : la pénétration de la plante hôte, la formation de lésions primaires, l'expansion des lésions et la macération des tissus puis la sporulation.



**Figure 3 :** Cycle de développement du *Botrytis cinerea* aux différents stades du fraisier (Duval, 1994)

Il passe l'hiver sous plusieurs formes : conidies, mycélium et sclérote qui est considéré chez *Botrytis* comme étant l'inoculum produit et dispersé. Cet inoculum est issu des débris végétaux atteints, il peut se conserver dans le sol, sous l'écorce et même dans les bourgeons. Le champignon ne semble pas infecter les graines, mais il peut se propager avec semences contaminées par des morceaux de débris végétaux infectés par le champignon ou par des sclérotés (Holz et al., 2004).

Macroscopiquement, la présence du *B. cinerea* est repérable par la production massive de macroconidies asexuées (de 3 à 6 noyaux) portés par des conidiophores. Le champignon forme un épais « duvet » gris à la surface des tissus infectés. Les macroconidies sont dispersées par le vent, les insectes ou par les projections et gouttelettes d'eau et permettent la contamination d'autres organes ou d'hôtes sains (Fillinger et Elad, 2005).

Pour une meilleure croissance, sporulation, libération et germination des spores, le *Botrytis cinerea* a besoin de milieux frais et humides. Après que les spores germent, elles pénètrent dans les tissus par des blessures dans la plante hôte et produisent ainsi un mycélium sur les vieux pétales de fleurs, sur les feuillages mourants et sur la tige. Le mode de progression de *Botrytis cinerea* dans les tissus de l'hôte est assez particulier : il détruit les cellules végétales au fur et à mesure et même au-devant de sa progression.

### III. Gamme d'hôte :

Le *Botrytis cinerea* peut infecter plus de 596 genres végétaux pour environ 1400 espèces de plantes, il n'a pas de spécificité apparente de l'hôte (Elad et al., 2007).

L'agent pathogène affecte de nombreuses productions végétales d'importance économique en culture en plein champ ou sous serre, comme : la poire, la cerise, la fraise, le kiwi, le raisin et la pomme en production fruitière, la carotte, la courgette, l'aubergine et la tomate en production légumière ou des plantes ornementales comme la rose ou le cyclamen. Ce champignon engendre des pertes économiques énorme (Dean et al., 2012).



**Figure 4 :** Productions végétales infectées par le *Botrytis cinerea*

### IV. Pouvoir pathogène du *Botrytis cinerea* :

Les dommages sont variables selon la plante hôte. Toutes les parties de la plante peuvent être affectées (Feuilles, tiges, fleurs, semences, bulbes, etc.), sauf les racines.

Les symptômes de la maladie causée par le botrytis cinerea peuvent apparaître aussi bien avant la récolte que sur des fraises déjà récoltées. Ils apparaissent de façon aléatoire sur la plante puis se généralise sur tout le fraisier si les conditions sont favorables pour son développement du champignon (El Oirdi, 2009).

Toutes les parties de la fleur peuvent être affectés et une sporulation brunâtre en forme de duvet apparait sur ces tissus. Les fleurs peuvent présenter un brunissement ou une pourriture humide recouverte d'un duvet gris. La moisissure atteint également la hampe florale et les parties infectées se dessèchent ou pourrissent. Le botrytis cinerea peut même causer la fonte de semis (Bayer-Agri.fr, 2008).

L'infestation commence à se manifester sur les fleurs en fin de floraison, elles se nécrosent puis de taches brunes apparaissent et un duvet grisâtre prolifère. Le même symptôme caractéristique de feutrage gris se développe sur les fruits qui finissent par se dessécher et se momifier.

Les feuilles se couvrent de taches brunes et les tiges se dessèchent avant de provoquer la mort des rameaux, ainsi le champignon peut atteindre le rhizome ce qui entraîne la mort du fraisier (Ajouz, 2009).

#### **V. Importance économique de la maladie :**

*Botrytis cinerea* entraîne une destruction partielle ou totale de la plante hôte et dans certains cas de toute la récolte. Les pertes économiques mondiales causées par cet agent pathogène sont énormes, de plus le développement rapide et insidieux de cet agent pathogène engendre chaque année la destruction de récoltes sur des centaines d'hectares.

*Botrytis cinerea* a été classé comme le deuxième champignon phytopathogène le plus important en raison de son importance scientifique et économique (Dean et al., 2012).

#### **VI. Facteurs influençant le développement du champignon :**

Différents facteurs interviennent dans les premiers stades d'infection, du développement du *Botrytis cinerea* et de la sporulation du champignon (Elad and Yunis, 1993 ; Yunis et al., 1990).

Parmi ces facteurs, les exigences nutritives qui constituent un facteur clé dans le développement du *Botrytis cinerea* puisque les conidies présentent peu de réserves énergétiques d'où le besoin d'une autre source de nutriments pour son développement. Les conditions environnementales, en particulier l'humidité et la température, jouent un rôle très important dans l'infection de la plante par le champignon et son développement. D'après O'Neill et al., (1997), la phase de sporulation est favorisée par une forte humidité relative et l'interruption de ces conditions entraîne un retard de sporulation. Cependant, d'après Hawker (1950), *B. cinerea* semble mieux sporuler en conditions sèches plutôt que dans une atmosphère saturée. Le vent va aussi avoir un effet mécanique sur la dissémination des conidies.

## **C. Analyse génomique**

### **I. Le génome et les variations génétiques :**

La molécule d'ADN est le support de l'information génétique pour la très grande majorité des organismes vivants. Du point de vue moléculaire, c'est une chaîne orientée de nucléotides liés par des liaisons covalentes. L'ADN contient 4 nucléotides différents A, C, G et T. La succession des différents nucléotides le long d'un brin d'ADN forme une séquence orientée. Du point de vue informatique, une molécule d'ADN est vue comme un texte construit sur un alphabet à 4 lettres, on parle d'une séquence d'ADN. Le génome d'un organisme est l'ensemble de ces textes.

Les génomes subissent cependant constamment des mutations par divers mécanismes dont certaines échappent aux mécanismes de réparation de l'ADN. Ces mutations, si elles se produisent dans les cellules de la lignée germinale et si elles sont transmises aux descendants, pourront éventuellement se répandre dans la population au fil des générations. Ces différences qu'on peut observer entre les génomes d'individus appartenant à une même espèce sont alors appelées des variations génétiques.

### **II. Données de séquençage des acides nucléiques :**

Le séquençage consiste à identifier l'enchaînement des nucléotides d'une séquence d'ADN, ou d'une manière plus générale, d'un génome complet (Maxam et Gilbert, 1977).

#### **1. Séquençage Sanger (séquençage de première génération) :**

Le séquençage de Sanger (synthèse enzymatique) repose sur le procédé de la réplication de l'ADN. Les scientifiques font des copies de brins d'ADN. Ensuite, ils notent les nucléotides ajoutés. Ainsi, ils peuvent observer la séquence des nucléotides. Il s'appuie sur la faculté que possèdent les ADN polymérases qui est de constituer le brin complémentaire d'un brin d'ADN matrice en présence des dNTPs. Il ajoute aussi au milieu, des ddNTPs marqués par fluorescence qui sont des terminateurs d'élongation de façon aléatoire dans la réaction. La lecture de la séquence se fait après migration des fragments de différente taille obtenus, sur gel de polyacrylamide (Sanger et al., 1977).

#### **2. Séquençage de seconde génération :**

La seconde génération est apparue en 2005 en réponse du prix élevé et au faible débit du séquençage de première génération. Ces technologies récentes permettent de séquencer l'ADN et l'ARN beaucoup plus rapidement que les méthodes précédentes comme le séquençage de Sanger, et comme tels ont révolutionné l'étude de la génomique et de la biologie moléculaire.

Toutes les technologies NGS présentent 3 étapes communes (Radford et al., 2012) :

La préparation de banques : les banques sont créées en utilisant une fragmentation aléatoire de l'ADN suivie de la liaison avec des petites séquences spécifiques qui sont appelées adaptateurs.

L'amplification : la banque est amplifiée à partir de fragments d'ADN simple brin grâce à des méthodes d'amplification clonale et de PCR.

Le séquençage : l'ADN est séquencé afin de déterminer l'enchaînement de nucléotides.

#### **a. Séquençage par Illumina :**

La technologie de séquençage illumina est une technologie basée sur l'incorporation réversible de nucléotides fluorescents dont l'extrémité 3'OH est désactivée, cette caractéristique permet donc, l'ajout d'une seule base par cycle et par lecture optique de la fluorescence (Raous, 2013).

- **Préparation de la banque d'ADN :**

Une banque d'ADN simple brin liée à des adaptateurs sur les deux extrémités est préparée, grâce à ces adaptateurs les fragments d'ADN de la banque peuvent se fixer par hybridation et aléatoirement sur les amorces qui recouvrent la flow cell (Sengenès, 2012).

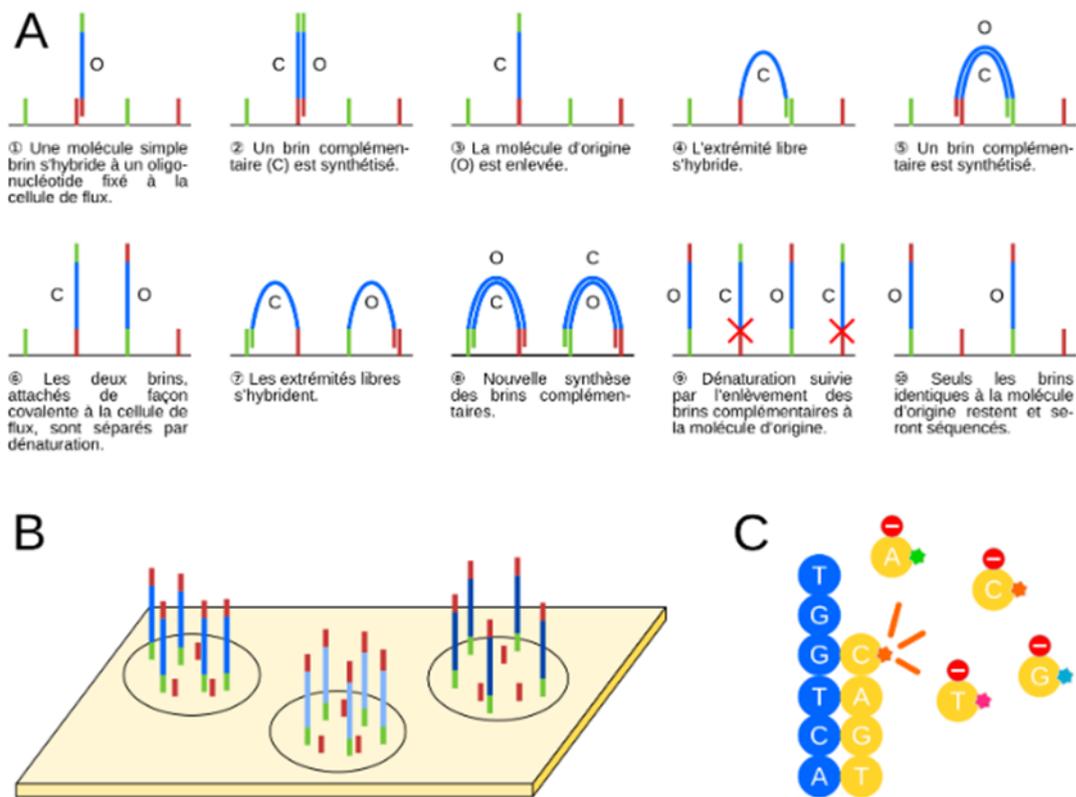
- **Amplification en pont :**

Après fixation de l'un des fragments de la banque préparée sur l'une des amorces de la flow cell, un nouveau brin complémentaire est synthétisé par une polymérase, et celui d'origine est éliminé par dénaturation. Ce nouveau brin se recourbe ensuite, pour s'hybrider à une amorce adjacente formant un pont. Le complément du brin est synthétisé par la polymérase formant un pont double brin, puis par dénaturation les deux copies sont ainsi séparées (Sengenès, 2012).

Le cycle d'amplification en pont se répète jusqu'à formation d'un cluster, puis en dernier lieu les brins anti-sens sont éliminés (Sengenès, 2012).

- **Séquençage par terminateurs réversibles :**

Cette étape commence par le blocage de l'extrémité 3' libre des fragments d'ADN, puis l'amorce de séquençage s'y hybride. Ensuite des nucléotides bloqués et marqués par fluorescence sont ajoutés, une fois que l'un d'entre eux est incorporé, la fluorescence émise est captée puis le fluorophore et le bloqueur sont clivés laissant la place pour l'ajout d'un nouveau. Après chaque cycle une base peut être identifiée (Sengenès, 2012).



**Figure 5** : Le principe de la technologie Illumina (Raous, 2013)

(A) Représentation schématique du processus d'amplification en pont (bridge amplification) des clusters sur la cellule de flux (flow cell, trait noir). Les adaptateurs et les oligonucléotides complémentaires présents sur la cellule de flux sont indiqués en rouge et vert, les fragments d'ADN à séquencer en bleu. La molécule s'étant initialement fixée à la cellule de flux, et les brins dont la séquence est identique, sont notés O ; les brins complémentaires sont notés C. À l'étape 9, les brins complémentaires à la molécule d'origine sont éliminés par clivage chimique d'un nucléotide modifié dans l'oligonucléotide « rouge » fixé à la cellule de flux (croix rouges).

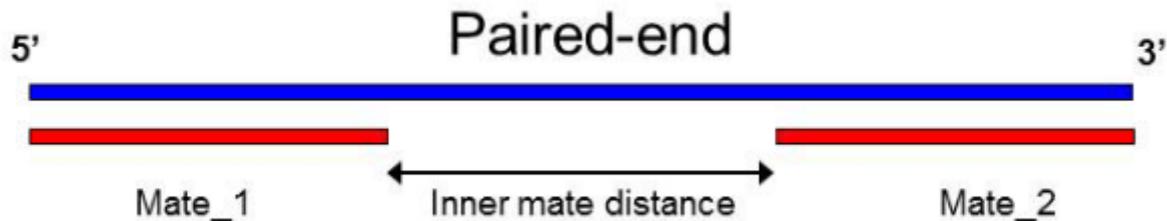
(B) Différentes molécules d'ADN sont amplifiées puis séquencées en même temps sur une même cellule de flux.

(C) Réaction de séquençage. Une polymérase synthétise un brin complémentaire (jaune) au brin situé sur la cellule de flux (bleu) en incorporant des nucléotides qui portent des groupes fluorescents (bleu, rose, orange, ou vert) et un groupe « stop » qui bloque la polymérisation. S'ensuit une étape d'imagerie afin d'identifier le nucléotide incorporé. Ensuite, les groupes fluorescents et « stop » sont enlevés et le processus se répète.

**b. Séquençages Paired-end et Single-end :**

- **Définition du séquençage *Paired-end* :**

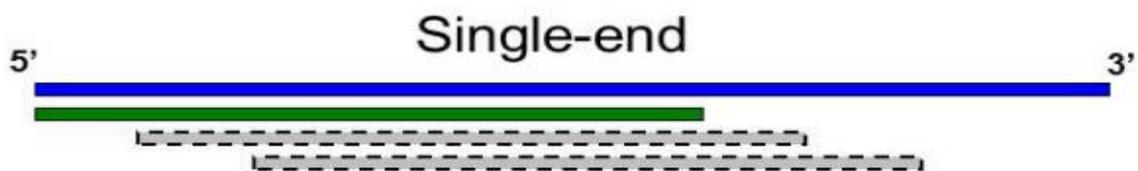
Ce type de séquençage consiste à séquencer un fragment d'ADN à partir de l'extrémité 5' puis l'extrémité 3' pour former un read reverse et un autre forward. Compte tenu de la taille de l'insert et de la longueur du *read*, les fragments séquencés peuvent être séparés ou chevauchés (Escalona et al., 2016).



**Figure 6:** Séquençage *Paired-end* (Botero-Castro, 2014)

- **Définition du séquençage *Single-end* :**

Le séquençage *Single-end* repose sur le séquençage d'un fragment d'ADN à partir d'une seule extrémité (Escalona et al., 2016).

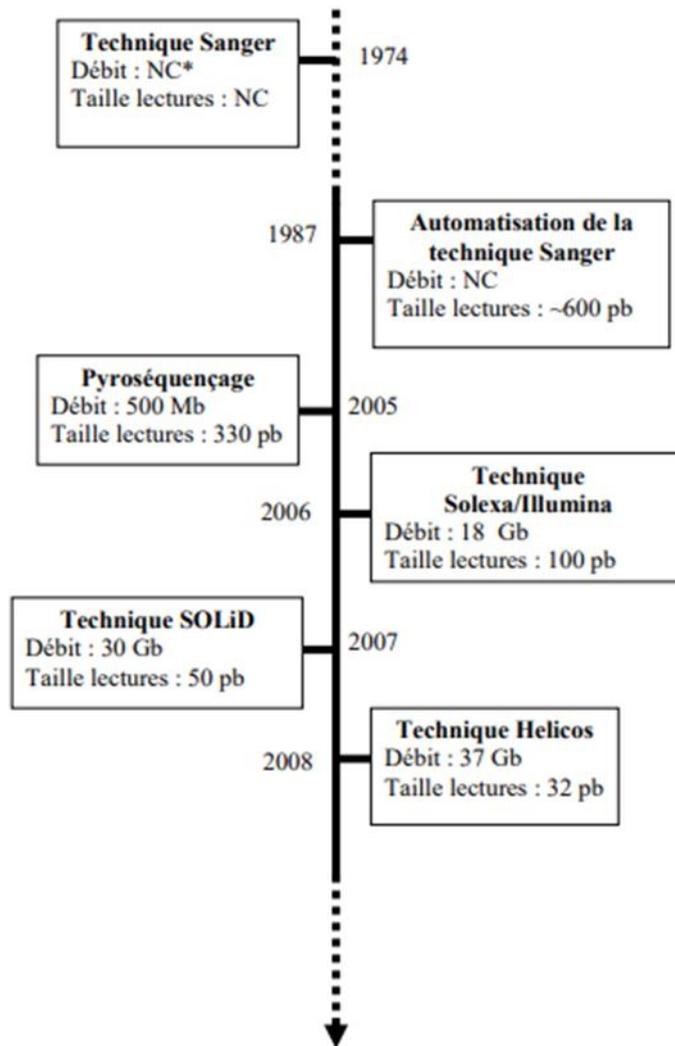


**Figure 7:** Séquençage *Single-end* (Botero-Castro, 2014)

3. Séquençage de troisième génération :

Ce type de séquençage connu sous le nom de séquençage *next-next generation* est le résultat de plusieurs recherches développées par des entreprises afin de trouver des systèmes de séquençage plus rapide et à débit plus élevé (Moorthie et al., 2011).

Le principe de la 3<sup>ème</sup> génération est le séquençage d'une molécule d'ADN sans passer par l'étape de la pré-amplification.



**Figure 8** : Evolution des techniques de séquençage (Botero-Castro, 2014)

### III. Analyse génomique après séquençage NGS :

La technologie NGS a nécessité le développement d'outils de bio-informatique spécifiques, capables d'analyser les données générées par les différentes méthodes. La bio-informatique joue un rôle fondamental dans l'analyse des génomes après leurs séquençages, en raison de la quantité et de la diversité importante de l'information à traiter. Les séquenceurs haut débit peuvent générer des millions, voir des milliards, de séquences courtes appelées reads (Peterlongo, 2016). Ces reads représentent une importante source d'information, accessible qu'après passage par plusieurs étapes primordiales.

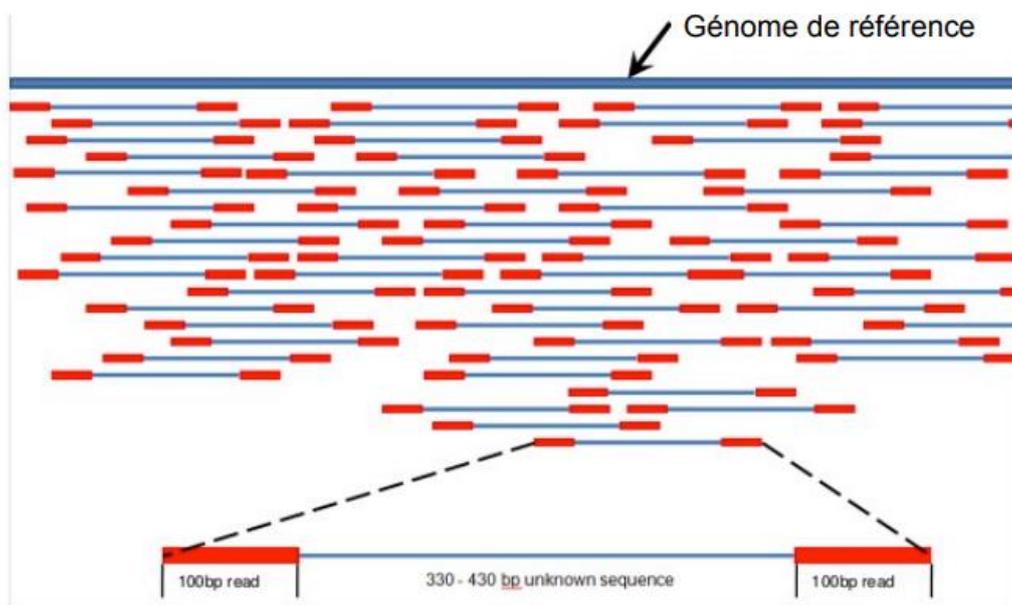
#### 1. Assemblage des séquences génomiques :

Pour assembler un génome deux méthodes différentes sont possibles, assemblage par références ou assemblage de novo (Pop, 2009).

Dans le cas de l'assemblage par références, l'utilisation d'une séquence d'un génome proche de celui étudié est nécessaire contrairement à l'assemblage de novo qui n'exige aucune référence pour la reconstruction des génomes (Pop, 2009).

**a. Assemblage par référence ou *mapping* :**

C'est un processus qui nécessite une séquence de référence afin d'assembler un génome, ainsi les reads obtenus sont alignés en fonction de leur localisation sur un génome de référence dont on déduit la structure du génome séquencé (Pop, 2009).



**Figure 9 :** Assemblage par référence (Botero-Castro, 2014)

**b. Assemblage de novo :**

La reconstitution du génome ne peut se faire qu'à partir des reads obtenus après le séquençage d'ADN. Pour ce faire, des logiciels appelés assembleur sont utilisés. Ces programmes à l'aide de divers algorithmes, arrivent à assembler plusieurs reads en contigs. Chaque contig représente une portion de la séquence finale que nous cherchons à reconstituer (Théroux, 2015).

La plupart des algorithmes d'assemblage de novo de génomes les plus utilisés, appartiennent à deux grandes familles, les algorithmes de type (OLC) et ceux de type (DBG) (Théroux, 2015).

Les assembleurs de type OLC sont plus adaptés pour les reads de grande taille (Théroux, 2015). Ces algorithmes commencent par la comparaison des reads les uns aux autres à la recherche de régions qui se chevauchent. Contrairement à l'assembleur de type OLC, celui de type graphe DeBruijn (DBG) est mieux adapté pour les reads de grande taille (Théroux, 2015).

## 2. Annotation génomique :

L'annotation structurale vise à associer le séquençage d'un génome à une information biologique exploitable ce qui permet l'identification des éléments d'intérêts dont les séquences codant des protéines et des molécules d'ARN ainsi que les séquences répétées et séquences régulatrices de l'expression des gènes font parties (Beyne, 2008).

### a. Annotation relationnelle :

L'annotation relationnelle a pour objectif l'exploitation des informations plus délicates que les informations rattachées aux séquences dans le but d'en sortir les relations qui peuvent exister entre les éléments prédits et caractérisés auparavant (Beyne, 2008).

### b. Annotation fonctionnelle :

Ce type d'annotation intervient afin d'attribuer les fonctions potentielles aux gènes identifiés lors de l'annotation structurale (Beyne, 2008).

### c. Annotation structurale :

L'annotation structurale consiste à identifier les éléments d'intérêts dans un génome. Ces éléments peuvent inclure les séquences codant des protéines et des molécules d'ARN (ARNt, ARNr, ARNsn...), les pseudogènes ainsi que certaines séquences répétées et séquences régulatrices de l'expression des gènes (Beyne, 2008).

# **Partie II**

## **Matériel et méthodes**

Dans ce chapitre, nous représentons les outils et les approches bio-informatiques ainsi que les méthodologies suivies au cours de notre travail de recherche.

Afin d'établir un bon workflow d'analyse génomique adéquat à l'analyse du génome du champignon, il est nécessaire de choisir le bon assortiment de programmes informatique en se basant sur le schéma général de traitement et d'analyse des données.

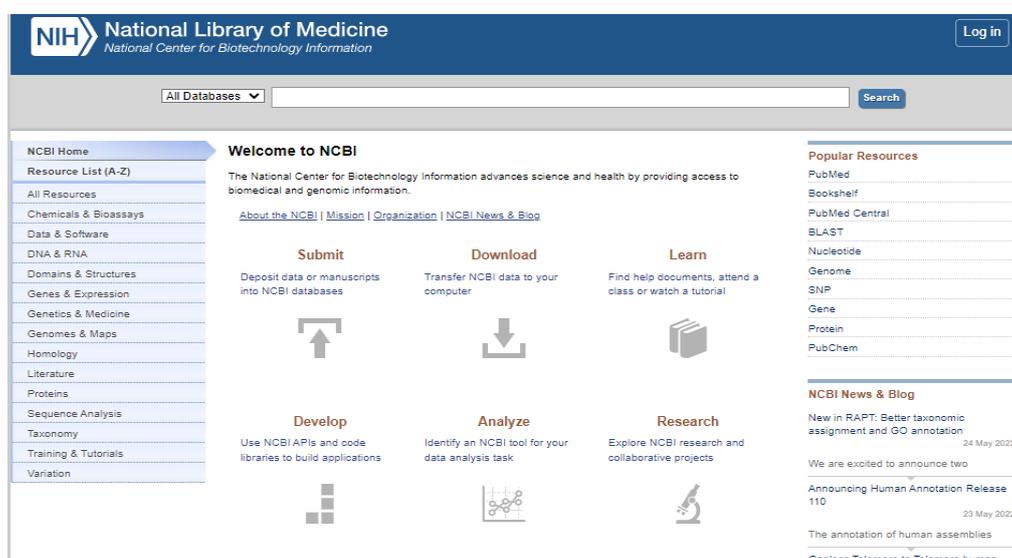
## I. Récupération des génomes via le programme NCBI :

Les bases de données utilisées lors de ce travail sont issues de banques de données publiques telles que GenBank. Ces banques sont générées par NCBI (National Center for Biotechnology Information).

### 1. Présentation du programme :

Le National Center for Biotechnology Information est un institut national américain pour l'information biologie moléculaire qui a pour mission le développement de logiciels d'analyse des données de génome ce qui facilite les recherches en biologie informatique.

Le NCBI n'est réellement pas une base de données mais il développe des bases de données publiques telles que : GenBank, PubMed et RefSeq (Altschul et al., 1990).



**Figure 10** : Site web NCBI ( <https://www.ncbi.nlm.nih.gov/>)

Le GenBank est la banque de données de séquences par excellence. Créée, maintenue et distribuée par le NCBI, elle contient une collection annotée de toutes les séquences nucléotidiques et protéiques publiques.

Deux formats différents de séquences peuvent être récupérés depuis NCBI :

- **Format GenBank :**

Ce format se compose d'une section d'annotation et une section de séquence. La section d'annotation est subdivisée en deux parties, Header « en tête » où on trouve les informations générales sur la séquence (L'identifiant, type de molécule, longueur en paires de bases, les références des articles, l'origine biologique, le numéro d'accèsion et la date de création) et les caractéristiques « Features » où on trouve les propriétés associées à toute ou une partie de la séquence. La section de séquence commence par le mot « ORIGIN » suivi de la séquence et elle est marquée à la fin par une ligne avec "/" (Annexe 1).

- **Format FASTA :**

Le format FASTA, signifie « fast all », est l'un des formats les plus répandus et les plus pratiques utilisés par de nombreux programmes d'analyse de séquences, tels que la similarité entre les séquences (Blast). Il se compose d'une ligne de définition ou de description suivie de la séquence. La ligne de description commence par un symbole supérieur à (>), qui est un élément crucial dans la forme FASTA et marque le début de chaque nouvelle séquence.

Le signe « > » et la première lettre de la ligne de définition ne doit contenir aucun espace. Les autres lignes sont constituées de lettres représentant des nucléotides ou des acides aminés de la séquence (Annexe 2).

## 2. Collecte des gènes de virulence du Botrytis :

Les gènes de virulence ont été collectés depuis des articles publiés sur NCBI et plus précisément sur PubMed. Les séquences nucléotidiques et protéiques ont été ensuite enregistrées dans un fichier Bloc-Notes sous format FASTA (Figure11).

NIH National Library of Medicine  
National Center for Biotechnology Information

Nucleotide

GenBank

### Botrytis cinerea vel1 gene for Velvet 1, strain B05.10

GenBank: HE977589.1  
FASTA Graphics

Go to:

LOCUS HE977589 1794 bp DNA linear PLN 15-JAN-2013  
DEFINITION Botrytis cinerea vel1 gene for Velvet 1, strain B05.10.  
ACCESSION HE977589  
VERSION HE977589.1  
KEYWORDS .  
SOURCE Botrytis cinerea (Botryotinia fuckeliana)  
ORGANISM [Botrytis cinerea](#)  
Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina;  
Leotiomycetes; Helotiales; Sclerotiniaceae; Botrytis.  
REFERENCE 1  
AUTHORS Schumacher,J., Pradier,J.M., Simon,A., Traeger,S., Moraga,J.,  
Collado,I.G., Viaud,M. and Tudzynski,B.  
TITLE Natural variation in the VELVET gene bcvel1 affects virulence and  
light-dependent differentiation in Botrytis cinerea  
JOURNAL PLoS ONE 7 (10), E47840 (2012)

Send to:

- Complete Record
- Coding Sequences
- Gene Features

Choose Destination

- File
- Clipboard
- Collections
- Analysis Tool

Download 1 item.

Format:

Show GI

Protein  
PubMed  
Taxonomy  
Full text in PMC

**Figure 11** : Paramètres de collecte des séquences nucléotidiques  
(<https://www.ncbi.nlm.nih.gov/>)

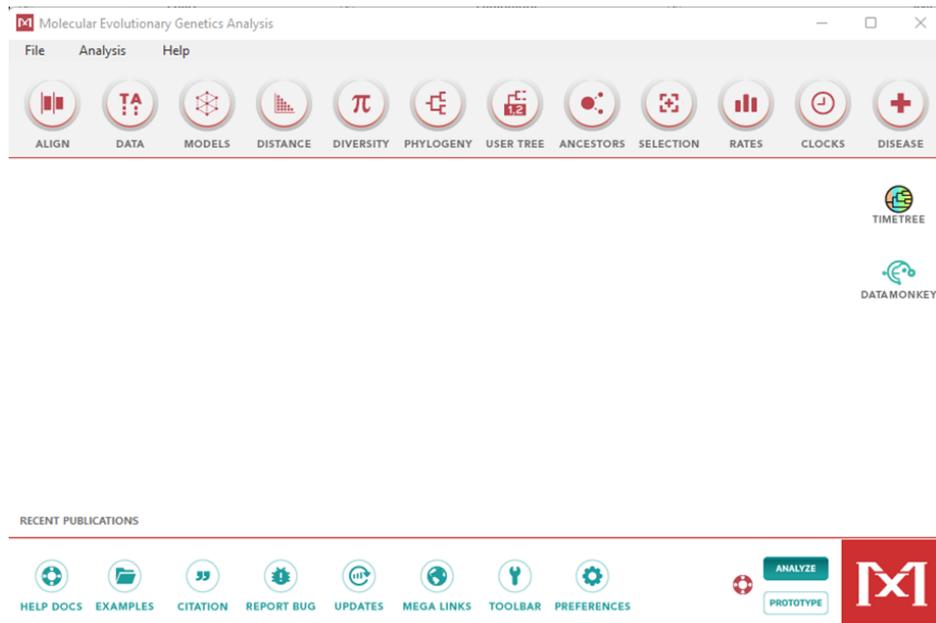
## II. Les outils bio-informatiques :

### 1. Logiciel MEGAX :

Dans cette étude le logiciel MEGAX qui a été utilisé pour réaliser l'alignement multiple des différents gènes collectés.

Le logiciel Molecular Evolutionary Genetics Analysis (MEGA) fournit des outils pour effectuer plusieurs types d'analyses. Ce logiciel comprend un large répertoire de programmes afin d'aligner des séquences et les assembler, déduire des arbres évolutifs, estimer des distances et des diversités génétiques, déduire des séquences ancestrales, calculer des arbres temporels et tester la sélection (Kumar et al. 2016). Au cours des 25 dernières années, l'utilisation de MEGA dans l'analyse évolutive a été citée dans plus de cent mille études dans divers domaines biologiques.

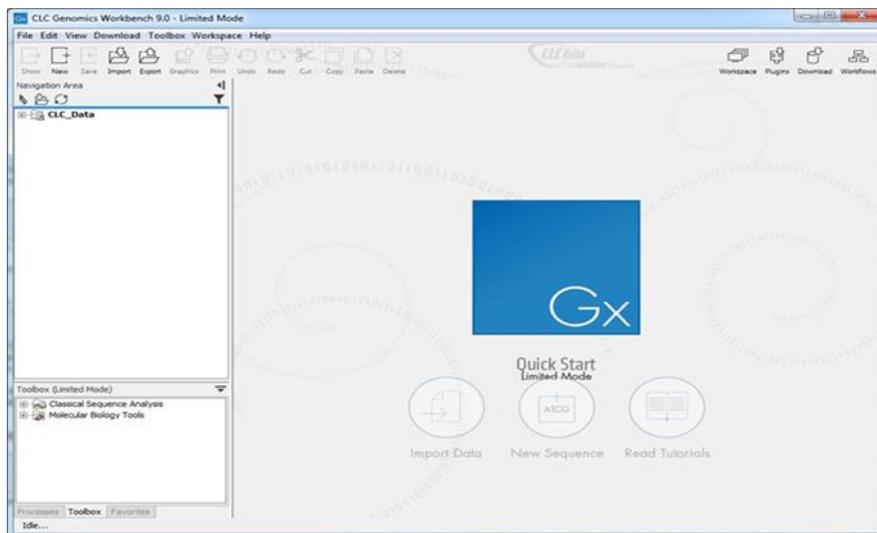
Ainsi, il est téléchargeable depuis le site suivant : [https://www.megasoftware.net/show\\_eua](https://www.megasoftware.net/show_eua) .



**Figure 12** : Le logiciel MEGAX

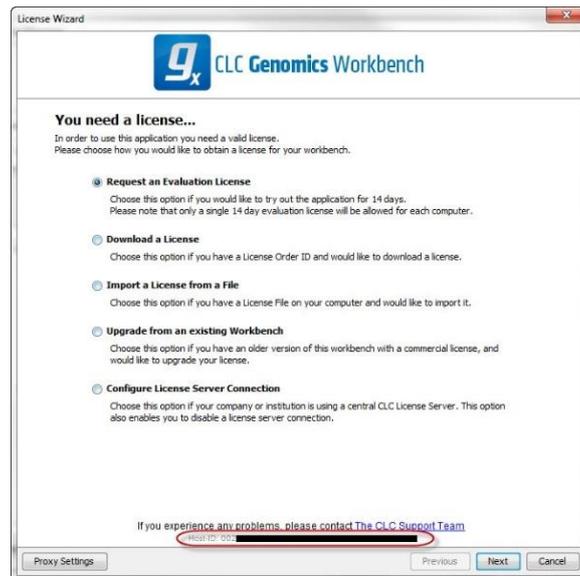
## 2. Logiciel CLC Genomics Workbench :

Pour l'alignement multiple et l'élaboration d'un arbre phylogénétique plusieurs autres logiciels peuvent être utilisés comme le CLC Genomics Workbench.



**Figure 13** : Le logiciel CLC Genomics Workbench

Le CLC Genomics Workbench est développé pour Windows, macOS et Linux. Le logiciel pour soit plate-forme téléchargeable sur <https://digitalinsights.qiagen.com/downloads/product-downloads/>. Lorsque vous démarrez CLC Genomics Workbench pour la première fois sur votre système, ou après l'installation d'une nouvelle version majeure, l'assistant de licence vous sera présenté (**Figure 14**).



**Figure 14 :** Assistant de licence permettant d'accéder aux options de licence (depuis CLC Genomics Workbench Manual)

Le CLC Genomics Workbench sert, à son tour, d'aligner des séquences nucléotidiques ou protéiques, déterminer les domaines conservés, tracer l'arbre phylogénétique et au développement des amorces.

Le seul inconvénient de ce logiciel, c'est qu'il est payant (Figure 15).

Software .	Company .	Cost (USD) <sup>a</sup> .
Avadis NGS	Strand Scientific Intelligence	\$4500
CLC Genomics Workbench	<b>CIC bio, Qiagen</b>	<b>\$5500</b>
CodonCode Aligner	CodonCode	\$720
Genamics Expression	Genamics	\$295

**Figure 15 :** Coûts du CLC Genomics Workbench

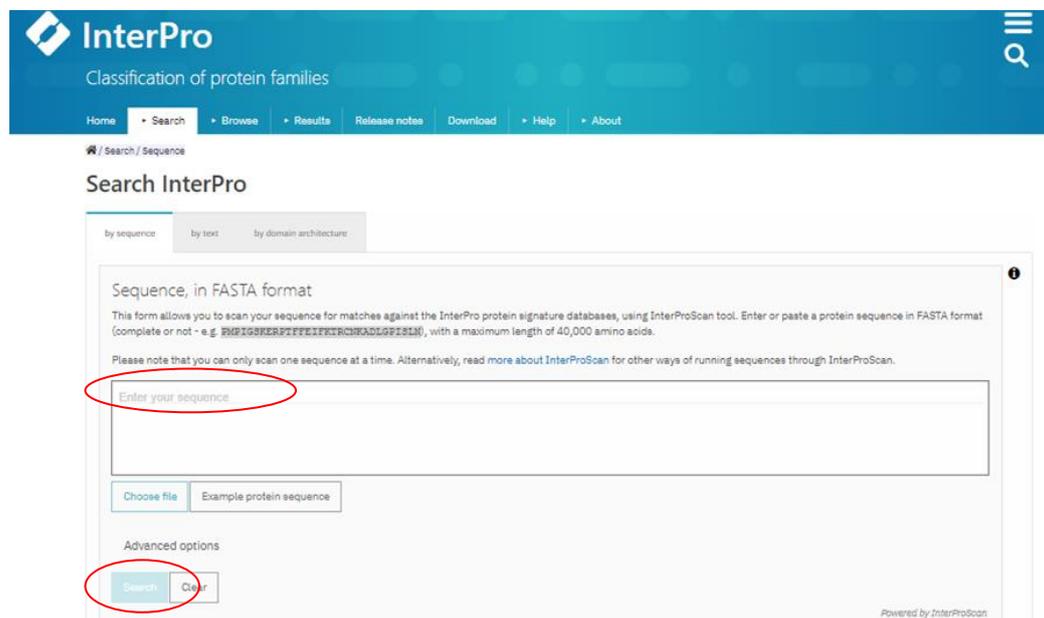
### III. Analyse des séquences nucléotidiques et protéiques des différents gènes :

#### 1. Détermination des fonctions des différents gènes :

Les gènes indiquent à chaque cellule son rôle dans l'organisme. Sur leur ordre, elles synthétisent des protéines : c'est la traduction du code génétique. Chaque protéine a un rôle différent d'une autre.

Une base de données, InterPro, a été utilisée dans notre étude des fonctions de gènes. InterPro permet la classification des protéines en fonction de la présence de domaines fonctionnels, répétitions et

signaux grâce à une recherche automatisée dans plusieurs bases de données (Pfam, SUPERFAMILY, PANTHER...).

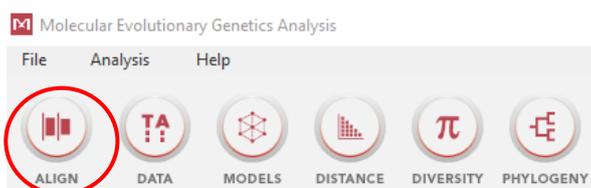


**Figure 16 :** Base de données InterPro (<https://www.ebi.ac.uk/interpro/>)

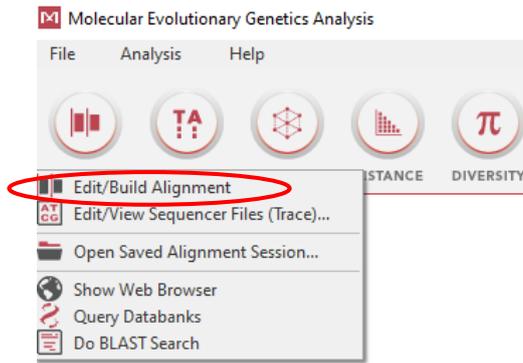
## 2. Alignement multiple :

L'alignement multiple est un outil de bio-informatique nécessaire pour structurer des protéines et prédire leurs fonctions, la phylogénie et d'autres tâches afin d'analyser une séquence d'ADN ou de protéine. Cette méthode consiste à aligner plusieurs séquences nucléotidiques ou protéiques et révèle les caractéristiques conservées ainsi qu'il permet de mieux comprendre les processus évolutifs ayant généré ces séquences.

L'alignement multiple permet de caractériser les familles de nucléotides par l'identification de régions d'homologie partagées, déterminer la séquence consensus de plusieurs séquences alignées afin d'établir des relations et de la phylogénie.



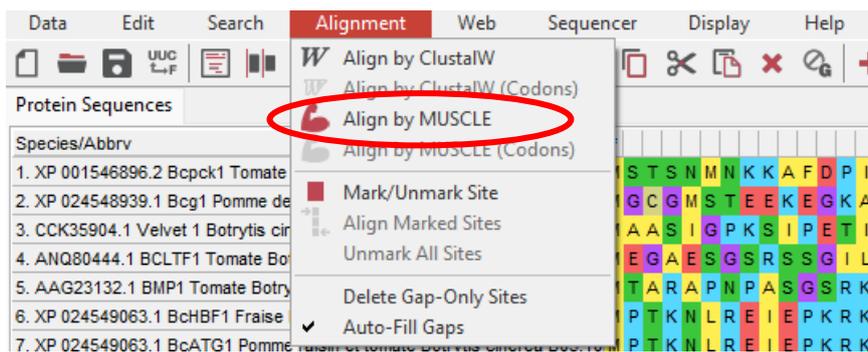
**Figure 17 :** Alignement multiple sur MEGAX



**Figure 18 :** Illustration des étapes d’alignement multiple des séquences de nucléotides

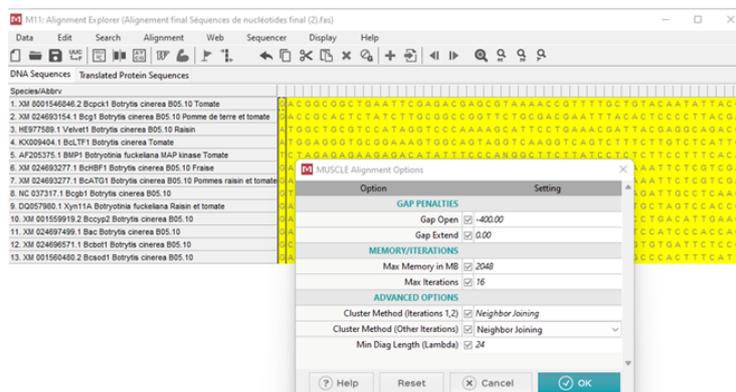
### 3. Alignement par MUSCLE sur MEGAX :

MUSCLE est un nouveau programme informatique pour créer de multiples alignements de séquences d’acides aminés ou de nucléotides. La vitesse et la précision de MUSCLE ont été comparées à celles de ClustalW et T-Coffee et ont obtenu le classement le plus élevé en termes de précision dans tous les tests (Edgar, Robert C., 2004).



**Figure 19 :** Alignement par Muscle sur MEGAX

Pour obtenir un bon alignement par MUSCLE, il faut ajuster certains paramètres illustrés par la figure ci-dessous.



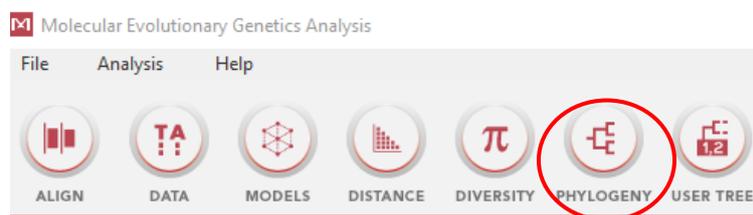
**Figure 20 :** Illustration des paramètres de l’alignement par MUSCLE

#### 4. Les domaines conservés :

Une séquence conservée ou un domaine conservé est une séquence, d'acides nucléiques ou d'acides aminés, qui est restée relativement similaire ou identique le long de l'arbre phylogénétique dans une ou plusieurs familles de protéines (Poitras et al., 2002).

#### 5. Arbre phylogénétique :

Les relations phylogénétiques entre les séquences permettent d'établir une matrice de distances génétiques entre plusieurs espèces. Pour tracer un arbre phylogénétique une nouvelle approche consiste à utiliser des logiciels et des applications qui reposent sur des modèles et des mécanismes évolutifs différents (Mariadassou, 2009).



**Figure 21** : Elaboration d'un arbre phylogénétique sur MEGAX

#### 6. Développement des amorces :

La conception d'amorces Oligo nucléotidiques appropriées est importante dans différentes applications, séquençage d'ADN et l'hybridation.

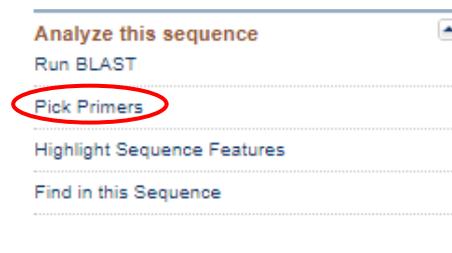
De nombreux programmes sont disponibles pour la conception des amorces prenant l'exemple des programmes bio-informatiques les plus utilisés et qui sont gratuits pour l'analyse des séquences d'acides nucléiques : Pick primer, Bioedit, Genfisher2...(Depuis <https://toptipbio.com/real-time-pcr-primer-blast/>. Mai 2022).

Dans notre étude, on a eu recours à Pick primer afin de développer les meilleures amorces permettant une meilleure amplification.

Les paramètres pris en compte pour la conception d'amorces sont les suivants :

- **Longueur d'amorce** : La longueur d'amorces doit être de 16 à 24 nucléotides.
- **Taux de GC** : Le pourcentage de GC donne des informations concernant la température de fusion. Le pourcentage optimal de l'amorce en GC est de 40 à 60%. Les nucléotides GC doivent être répartis uniformément le long de l'amorce.

- **Tm des oligonucléotides :** Pour de meilleurs résultats la Tm doit être de 52°C à 60°C alors que les températures plus basses ne donnent pas de bons résultats. Ainsi les amorces de températures de fusion supérieures à 65°C doivent être évitées en raison du risque d'hybridation secondaire. Il est donc conseillé de choisir une température de 60°C pour la réaction de séquençage d'hybridation et d'élongation.
- **La taille du produit PCR :** La taille du produit PCR idéale est de 70-200 paires de bases.
- **Séquence :** Une complémentarité entre les amorces ne doit pas exister afin de ne pas former les épingles à cheveux ou à l'autre amorce utilisée dans les réactions de PCR.



**Figure 22 :** Développement d'amorces via NCBI (<https://www.ncbi.nlm.nih.gov/>)

Primer pair 1									
	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	TTGATTGGTGACGACGAGCA	Plus	20	1514	1533	59.97	50.00	3.00	2.00
Reverse primer	GCGCGAAGCAAGATGAGAAG	Minus	20	1955	1936	59.97	55.00	4.00	0.00
Product length	442								
Products on intended targets									
>XM_001546846.2 Botrytis cinerea B05.10 Bcpc1 (Bcpc1), mRNA									
product length = 442									
Forward primer	1	TTGATTGGTGACGACGAGCA	20						
Template	1514	.....	1533						
Reverse primer	1	GCGCGAAGCAAGATGAGAAG	20						
Template	1955	.....	1936						
Primer pair 2									
	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	TGTTTTGCTTGCTGCATCCC	Plus	20	157	176	59.97	50.00	5.00	1.00
Reverse primer	CGAAGGACCAACTGCTCT	Minus	20	622	603	59.97	55.00	3.00	0.00
Product length	466								
Products on intended targets									
>XM_001546846.2 Botrytis cinerea B05.10 Bcpc1 (Bcpc1), mRNA									
product length = 466									
Forward primer	1	TGTTTTGCTTGCTGCATCCC	20						
Template	157	.....	176						
Reverse primer	1	CGAAGGACCAACTGCTCT	20						
Template	622	.....	603						

**Figure 23 :** Présentation de l'interface du développement d'amorces sur NCBI (<https://www.ncbi.nlm.nih.gov/>)

The image shows the Primer3 web interface. The 'PCR Template' section has 'XM\_037339890.1' entered in the 'Enter accession number or FASTA sequence' field. The 'Primer Parameters' section shows 'PCR product size' set to 70-1000, '# of primers to return' set to 10, and 'Primer melting temperatures (Tm)' set to 57.0-63.0. The 'Exon/intron selection' section shows 'Exon junction span' set to 'No preference' and 'Exon junction match' set to 7-8. The 'Primer Pair Specificity Checking Parameters' section shows 'Search mode' set to 'Automatic', 'Database' set to 'Refseq mRNA', and 'Organism' set to '1964551'. The 'Get Primers' button is circled in red.

**Figure 24 :** Illustration des étapes d'identification des amorces  
(<https://www.ncbi.nlm.nih.gov/>)

## 7. Recherche des marqueurs SSR :

Afin de rechercher les marqueurs SSR, on a recours au traitement des séquences nucléotidiques des différents gènes par MISA. C'est un logiciel sous Windows pour la recherche automatique et la prédiction des microsatellites, automatisé à la conception des amorces, l'identification et localisation des marqueurs microsatellites parfaits ainsi que les microsatellites composés qui sont interrompus par un certain nombre de bases (Duran et al., 2009). MISA nécessite un ensemble de séquences au format FASTA et un fichier de paramètres qui définit la taille de l'unité et le nombre de répétition minimum de chaque SSR (tableau 3), en possédant plusieurs paramètres d'entrés.

Le nombre de répétitions des motifs varie d'une espèce à l'autre, on parle de taille par exemple (CA)15, (CA)11 ou (CA)19. Les microsatellites montrent également un polymorphisme de structure qui correspond à un changement d'une base en une autre à l'intérieur du motif, par exemple le microsatellite (CA)16 a subi une mutation de A en T dans la 14ème répétition (CA)13 CT (CA)2.

**Tableau 3 : Paramètres choisis pour la recherche des SSRs**

Taille de l'unité	Nombre minimum de répétitions
<b>1 (mono)</b>	10
<b>2 (di)</b>	6
<b>3 (tri)</b>	5
<b>4 (tetra)</b>	5
<b>5 (penta)</b>	5
<b>6 (hexa)</b>	5

Les valeurs des paramètres pris en compte sur MISA sont comme suite :

- $T_m^\circ$  : 50°C à 61°C
- Taille des amorces : 20 – 21 pb
- Pourcentage en GC% : 30 à 60%

L'exécution du programme engendre 2 types de fichiers séparés :

- Fichier. Xls : Contient les microsatellites identifiés avec leurs types, leurs localisations, les amorces associées à chaque SSR avec plusieurs paramètres, tels que, la taille d'amorces, la teneur en GC, la  $T_m^\circ$  et la taille du produit PCR.
- Fichier. Statistics : Résume les différentes statistiques d'un type de microsatellites.

# **Partie III**

## **Résultats et discussions**

## I. Classification des gènes de virulence collectés :

Les gènes de virulence collectés depuis PubMed ont été classés dans le tableau ci-dessous avec un total de 13 gènes en mentionnant les codes d'accès des séquences nucléotidiques et des séquences protéiques, l'hôte et les différentes fonctions de chaque gène.

**Tableau 4 :** Classification des gènes selon leur hôte et fonctions

Nom du gène	Forme du Botrytis	Code d'accèsion	Hôte	Fonctions
BcPCK1	Botrytis cinerea B05.10	Nucléotide: XM_001546846 .2 Protéine: XP_001546896.2	Tomate	-Catalyse la décarboxylation réversible de l'oxaloacétate pour produire du phosphoenolpyruvate et du CO <sub>2</sub> , -Synthèse du glycogène à partir du glucose.
Bcg1	Botrytis cinerea B05.10	Nucléotide: XM_024693154 .1 Protéine: XP_024548939.1	Pomme de terre Tomate	-Liaison au complexe de sous-unité bêta/gamma de la protéine G. -Rôle important dans la morphogénèse
Bcvel1	Botrytis cinerea B05.10 (Botryotinia fuckeliana)	Nucléotide: HE977589.1 Protéine: CCK35904.1	Vitis	-Formation de sclérotés et la production de toxines.
BcLTF1	Botrytis cinerea	Nucléotide: KX009404.1 Protéine: ANQ80444.1	Tomate	-Régulation de la transcription sur matrice d'ADN.
BMP1	Botryotinia fuckeliana	Nucléotide: AF205375 Protéine: AAG23132.1	Tomate (leaves)	-Catalyse de la phosphorylation d'un résidu d'acide aminé dans une protéine, généralement selon la réaction : une protéine + ATP = une phosphoprotéine + ADP.
BcHBF1	Botrytis cinerea B05.10	Nucléotide: XM_024693277 .1 Protéine: XP_024549063.1	Tomate et fraise	-Production de sclérotés, -Indispensable pour l'adaptation au stress osmotique et oxydatif de B. cinerea ainsi que pour l'intégrité de la paroi cellulaire
BcATG1	Botrytis cinerea B05.10	Nucléotide: XM_024693794 .1 Protéine: XP_024549581.1	Pomme, raisins et tomate	_____
Bcgb1	Botrytis cinerea B05.10	Nucléotide: NC_037317.1 Protéine: XP_024550184.1		-Endopolygalacturonase -Formation de sclérotés

<b>Xyn11A</b> endo-beta-1,4-xylanase precursor	Botrytis cinerea (Botryotinia fuckeliana)	Nucléotide: DQ057980.1 Protéine: AAZ03776.1	Raisins et feuilles de tomate	-Dégradation de la paroi cellulaire, -Activité hydrolase, hydrolysant les composés O-glycosyle.
<b>Bccyp2</b>	Botrytis cinerea B05.10	Nucleotide: XM_001559919.2 Protéine: XP_001559969.1		-Activité peptidyl-prolyl cis-trans isomérase, -Isomérisation des protéines peptidyl-prolyl.
<b>Bac</b>	Botrytis cinerea B05.10	Nucléotide: XM_024697499.1 Protéine: XP_024553314.1		-Transduction du signal, Adenylate cyclase, - Rôle important dans les processus de croissance et de pathogénicité du champignon
<b>Bcbot</b>	Botrytis cinerea T4 SAS56	Nucléotide: XM_024696571.1 Protéine: XP_024552382.1		-Biosynthèse de toxines, P450 monooxygénase
<b>Bcsod1</b>	Botrytis cinerea B05.10	Nucléotide: XM_001560480.2 Protéine: XP_001560530.1		-Génération de H2O2, superoxide dis-mutase

Ces gènes de virulence peuvent être classés en familles différentes ce qui est représenté dans le tableau 4.

**Tableau 5** : Classification des gènes de virulence en familles

Familles de gènes		Gènes			
<b>Bcg</b>	Bcg1	Bcg2	Bcg3 :		
<b>Fonction : Gα sous-unités G-protéines</b>	Nucléotide : XM_024693154.1 Protéine : XP_024548939.1	Nucléotide : XM_024697040.1 Protéine : XP_024552855.1	Nucléotide : XM_024697564.1 Protéine : XP_024553380.1		
<b>Bcpls</b>	Bcpls1				
<b>Fonction : Tétraspandin</b> <b>Pénétration de l'appressorium</b>	Nucléotide : XM_001551683.2 Protéine : XP_001551733.2				

<b>Bcpg</b>	Bcpg1 Nucléotide : NC_037323.1 Protéine : XP_001550077.1	Bcpg2 Nucléotide : XM_024696956.1 Protéine : XP_024552772.1	Bcpg3 Nucléotide : XM_024692630.1 Protéine : XP_024548409.1		
<b>Fonction : endo-polygalacturonase</b> <b>Dégradation de la paroi cellulaire</b>					
<b>Bcatg</b>	Bcatg2 Nucléotide : XM_001552778.2 Protéine : XP_001552828.2	Bcatg13 Nucléotide : XM_024696860.1 Protéine : XP_024552673.1	Bcatg8 Nucléotide : XM_001558783.2 Protéine : XP_001558833.1	Bcatg5 Nucléotide : XM_024695217.1 Protéine : XP_024551010.1	
<b>BcVEL</b>	BcVEL1 Nucléotide: HE977589.1 Protéine: CCK35904.1	BcVEL2 Nucléotide: HE977591.1 Protéine : CCK35906.1	BcVEL3 Nucléotide: HE977592.1 Protéine : CCK35907.1	BcVEL4 Nucléotide: HE977594.1 Protéine : CCK35909.1	BcLAE1 Nucléotide : XM_024692832.1 Protéine : XP_024548614.1

## II. Les domaines conservés, les amorces associées et les arbres phylogénétiques des différents gènes de virulence du botrytis cinerea :

### 1. Alignement multiple :

L'alignement multiple est réalisé par le logiciel MEGAX qui nous a permis d'étudier le degré de similitude des séquences des gènes de virulence dont les résultats sont représentés par les figures (24 et 25).

Les gènes de virulence alignés sur MEGAX sont au nombre de 13 gènes différents avec leurs codes d'accès, ils sont ensuite classés dans les tableaux ci-dessous (Tableau 5 et 6).

**Tableau 6 :** Classification des gènes de virulence selon les codes d'accès des séquences nucléotidiques

Gènes de virulence	Code d'accès des séquences nucléotidiques
Bcpk1	XM_001546846.2
Bcg1	XM_024693154.1
Velvet1	HE977589.1
BcLTF1	KX009404.1

BMP1	AF205375
BcATG1	XM_024693794.1
BcHBF1	XM_024693277.1
Bcgb1	NC_037317.1
Xyn11A	DQ057980.1
Bccyp2	XM_001559919.2
Bac	XM_024697499.1
Bcbot1	XM_024696571.1
Bcsod1	XM_001560480.2

Puis on a aligné les différentes séquences nucléotidiques sur MEGAX après les avoir récupérés sur Bloc-notes depuis NCBI sous format FASTA.



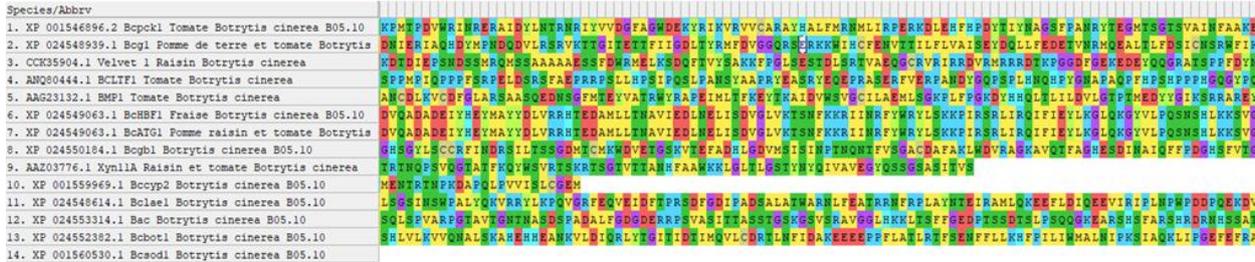
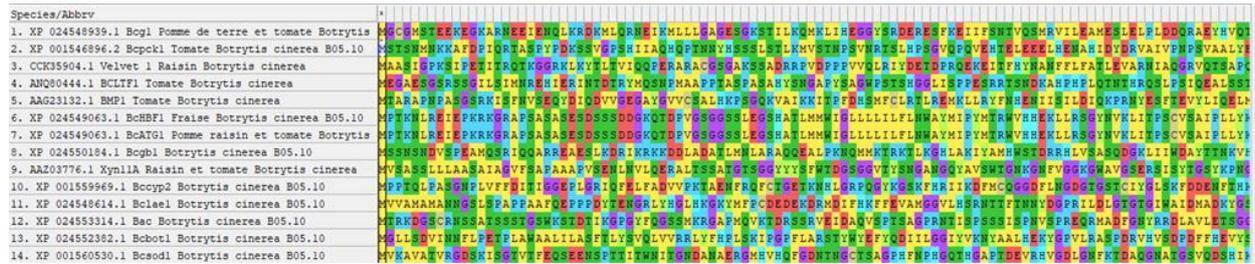
**Figure 25 :** Présentation de l'interface de l'alignement multiple des différentes séquences nucléotidiques sur MEGAX

Ensuite on a aligné les différentes séquences protéiques classées dans le tableau ci-dessous.

**Tableau 7** : Classification des gènes de virulence selon les codes d'accèsion des séquences protéiques

Gènes de virulence	Code d'accèsion des séquences Protéiques
Bcpk1	XP_001546896.2
Bcg1	XP_024548939.1
Velvet1	CCK35904.1
BcLTF1	ANQ80444.1
BMP1	AAG23132.1
BcATG1	XP_024549581.1
BcHBF1	XP_024549063.1
Bcgb1	XP_024550184.1
Xyn11A	AAZ03776.1
Bccyp2	XP_001559969.1
Bac	XP_024553314.1
Bcbot1	XP_024552382.1
Bcsod1	XP_001560530.1

Puis on a aligné les différentes séquences protéiques sur MEGAX après les avoir récupérés sur Bloc-notes depuis NCBI sous format FASTA.

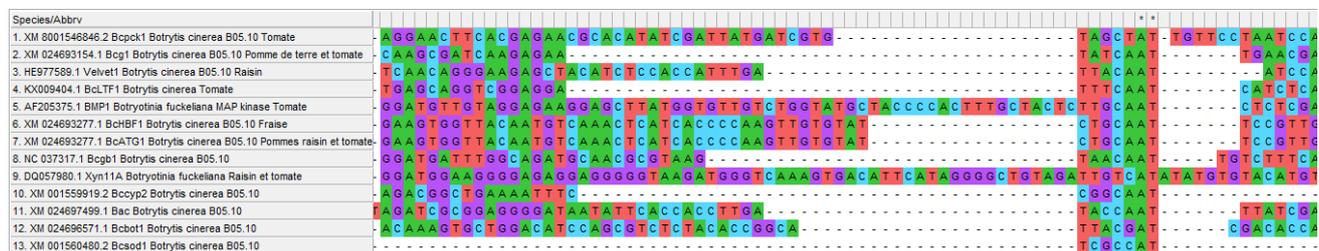
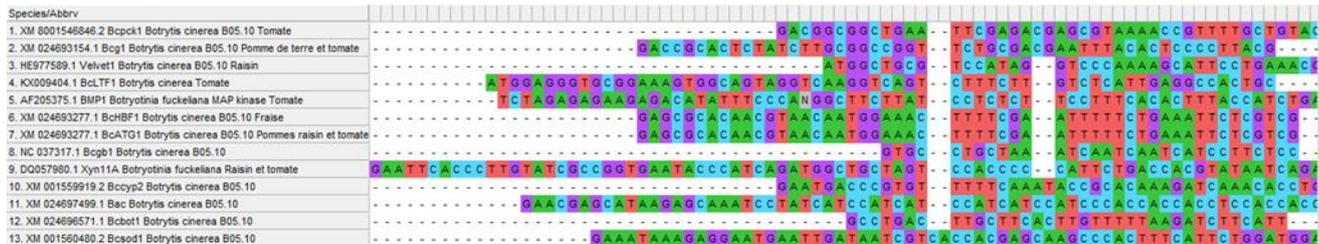


**Figure 26 :** Présentation de l'interface de l'alignement multiple des différentes séquences protéiques sur MEGAX

2. Alignement par Muscle :

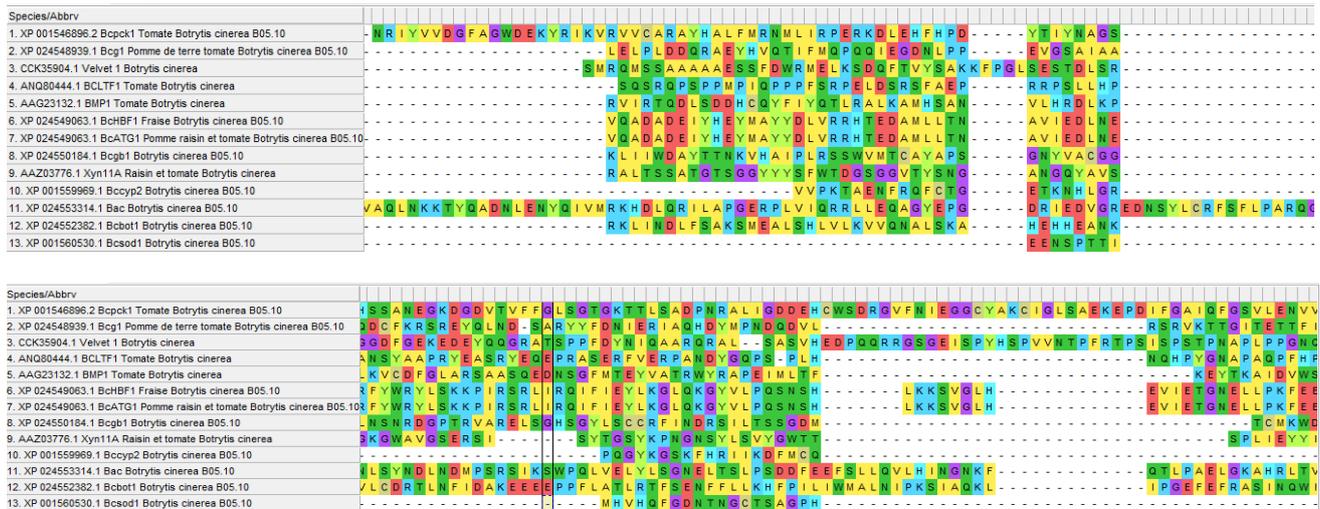
MUSCLE est un nouveau programme informatique pour créer de multiples alignements de séquences de protéines. Il est utilisé dans notre travail afin d'identifier les zones de concordance entre les différentes séquences de gènes de virulence du *Botrytis cinrea*.

**a. Alignement des séquences nucléotidiques :**



**Figure 27 :** Présentation de l'interface de l'alignement par Muscle des différentes séquences nucléotidiques sur MEGAX

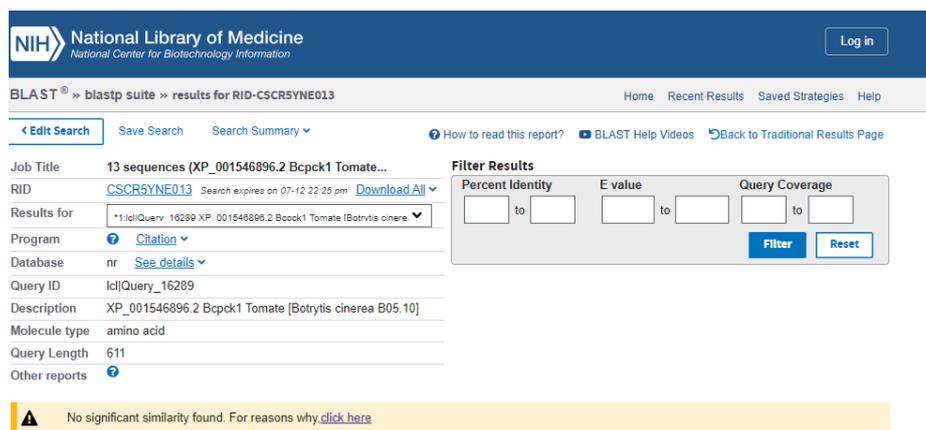
## b. Alignement des séquences protéiques :



**Figure 28 :** Présentation de l'interface de l'alignement par Muscle des différentes séquences protéiques sur MEGAX

Après avoir aligné les 13 séquences des gènes de virulence du *Botrytis cinerea*, on constate qu'il y'a pas de similitude remarquable entre eux.

On prend par la suite le code d'accèsion des 13 séquences protéiques et on implique un Blast protéique (BlastP) afin de savoir les noms des domaines conservés «specific hits» et leur emplacement sur les séquences. On a obtenu ensuite le résultat représenté par la figure 28.



**Figure 29 :** BlastP des séquences protéiques des 13 gènes de virulence du *Botrytis cinerea*

D'après les résultats qui précèdent on conclut que les gènes de virulence sont différents et ne sont pas identiques. Mais ils peuvent être similaires au point de vue fonction ou bien selon la nature chimique.

Les divergences entre les séquences de gènes de virulence du *Botrytis cinerea* sont interprétées comme résultant de mutation.

3. Alignement des séquences protéiques des différentes familles de gènes :

a. Famille de gènes Bcg :

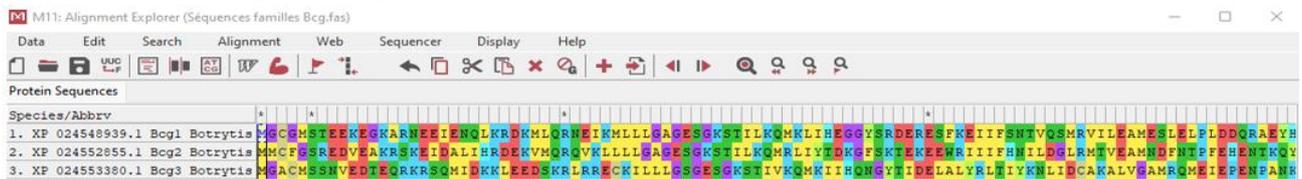


Figure 30 : Alignement des séquences protéiques de la famille Bcg sur MEGAX



Figure 31 : Domaines conservés des séquences protéiques de Bcg générés par Muscle sur MEGAX  
Ces domaines conservés représentent les similarités entre les différents gènes de la famille Bcg.

b. Famille de gènes Bcpg :



Figure 32 : Alignement des séquences protéiques de la famille Bcpg sur MEGAX

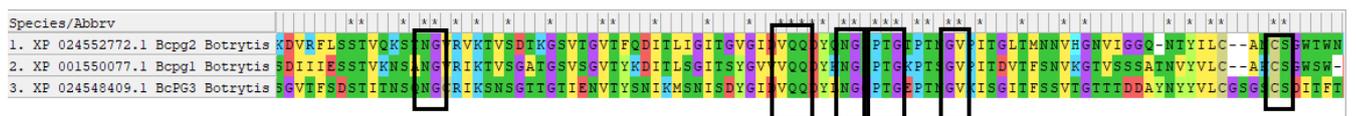


Figure 33 : Domaines conservés des séquences protéiques de la famille Bcpg par Muscle sur MEGAX

c. Famille de gènes BcATG :

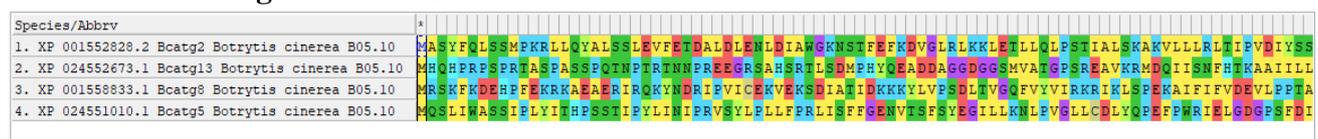


Figure 34 : Alignement des séquences protéiques de la famille BcATG sur MEGAX



**Figure 35 :** Alignement des séquences protéiques de la famille BcATG par Muscle sur MEGAX

**d. Famille des gènes BcVEL :**



**Figure 36 :** Alignement des séquences nucléotidiques de la famille BcVEL sur MEGAX

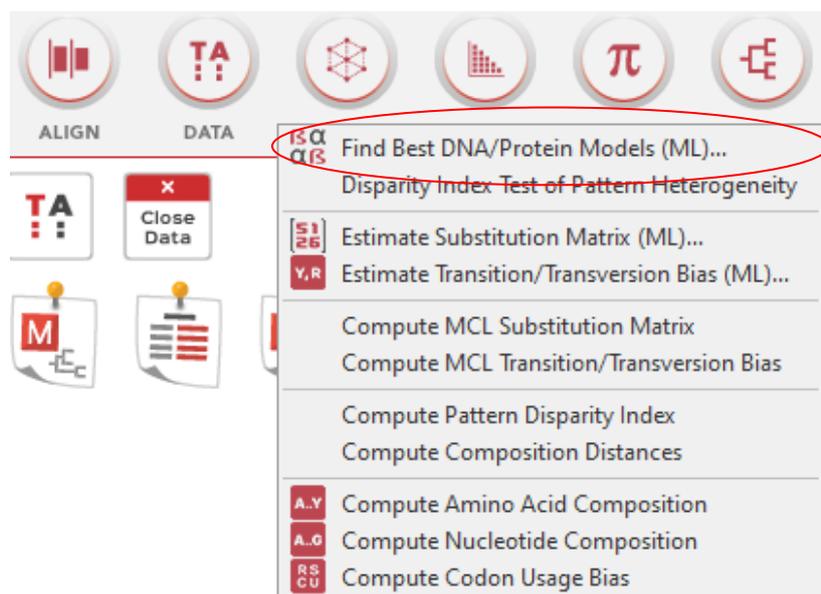


**Figure 37 :** Alignement des séquences protéiques de la famille BcVEL par MUSCLE sur MEGAX  
 Pour la famille BcVEL, les gènes ne représentent pas de similitude vu qu'il n'existe pas de domaines conservés entre les séquences protéiques.

4. Arbre phylogénétique :

A partir des figures précédentes (26 et 27) de l'alignement certains gènes présentent une similitude dans leurs parties conservées. En outre, pour d'avantage d'éclaircissement les arbres phylogénétiques ont été établis sous la base de données de l'alignement multiple toujours par le même logiciel MEGAX.

Avant de tracer l'arbre phylogénétique, on doit chercher les paramètres à modifier pour avoir un meilleur modèle et le plus raisonnable de notre arbre présentant les différentes séquences et ceci est possible sur le logiciel MEGAX (figure 36).



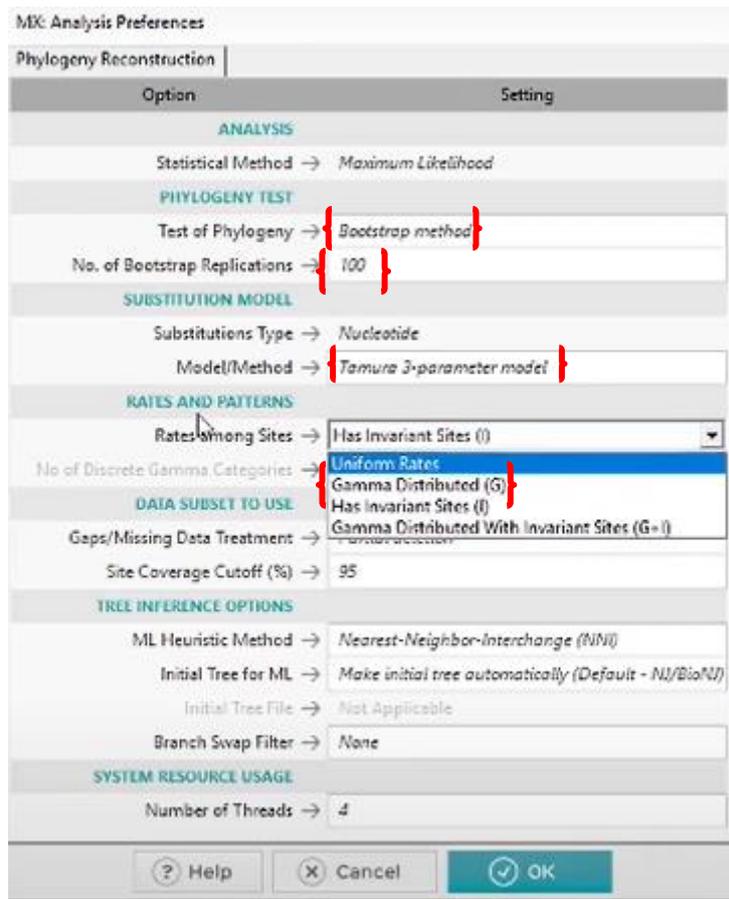
**Figure 38 :** Illustration du paramètre du choix du meilleur modèle de l'arbre phylogénétique

Par la suite, on a comme résultat un tableau qu'on ouvre sur Excel, il représente différents paramètres du choix du meilleur modèle.

**Tableau 8 :** Paramètres du choix du meilleur modèle de l'arbre phylogénétique sur MEGAX

1	A	B	C	D	E	F	G	H	I	J	K	L	A
2	Model	#Param	BIC	AICc	InL	Invariant	Gamma	R	Freq A	Freq T	Freq C	Freq G	
2	T92+G	28	99257,2791	99016,82082	-49480,38994	n/a	4,04781183	1,36517064	0,27735103	0,27735103	0,22264897	0,22264897	
3	HKY+G	30	99281,87628	99024,24543	-49482,09927	n/a	4,00341995	1,37533524	0,281771017	0,272931043	0,229411172	0,215886768	
4	TN93+G	31	99292,46501	99026,24803	-49482,09901	n/a	4,00296757	1,37539806	0,281771017	0,272931043	0,229411172	0,215886768	
5	GTR+G	34	99308,25151	99016,27674	-49474,10837	n/a	3,99861725	1,37683708	0,281771017	0,272931043	0,229411172	0,215886768	
6	GTR+G+	35	99310,0122	99009,45171	-49469,69409	0,003332323	4,57318943	1,36144703	0,281771017	0,272931043	0,229411172	0,215886768	
7	T92	27	99395,81526	99163,94341	-49554,95265	n/a	n/a	1,1246511	0,27735103	0,27735103	0,22264897	0,22264897	
8	T92+G+I	29	99402,71812	99153,6735	-49547,81482	0,005391681	6,91452847	1,89168615	0,27735103	0,27735103	0,22264897	0,22264897	
9	GTR+I	34	99403,73894	99111,76418	-49521,85209	0,005690278	n/a	1,14280065	0,281771017	0,272931043	0,229411172	0,215886768	
10	HKY	29	99419,59849	99170,55387	-49556,25501	n/a	n/a	1,12830183	0,281771017	0,272931043	0,229411172	0,215886768	
11	HKY+G+	31	99424,4247	99158,20772	-49548,07886	0,005418659	6,95270779	1,89081203	0,281771017	0,272931043	0,229411172	0,215886768	
12	TN93+G+	32	99432,10944	99157,30643	-49546,6266	0,005392055	6,93212055	1,87842906	0,281771017	0,272931043	0,229411172	0,215886768	
13	TN93	30	99444,05645	99186,4256	-49563,18936	n/a	n/a	1,23191882	0,281771017	0,272931043	0,229411172	0,215886768	
14	GTR	33	99446,15363	99162,76469	-49548,35406	n/a	n/a	1,12687217	0,281771017	0,272931043	0,229411172	0,215886768	
15	T92+I	28	99528,09782	99287,63953	-49615,7993	0,006361212	n/a	1,70842326	0,27735103	0,27735103	0,22264897	0,22264897	
16	K2+G	27	99543,60705	99311,7352	-49628,84855	n/a	4,22387296	1,30620459	0,25	0,25	0,25	0,25	
17	HKY+I	30	99548,24755	99290,6167	-49615,28491	0,006397906	n/a	1,70763383	0,281771017	0,272931043	0,229411172	0,215886768	
18	TN93+I	31	99555,64667	99289,42969	-49613,68984	0,006378603	n/a	1,69727398	0,281771017	0,272931043	0,229411172	0,215886768	
19	K2	26	99677,10256	99453,81725	-49700,89093	n/a	n/a	1,09327214	0,25	0,25	0,25	0,25	
20	K2+G+I	28	99718,59642	99478,13814	-49711,0486	0,00572523	6,68365374	1,91460252	0,25	0,25	0,25	0,25	
21	K2+I	27	99848,41437	99616,54252	-49781,25221	0,006706794	n/a	1,72911954	0,25	0,25	0,25	0,25	
22	JC+G	26	100226,9642	100003,6789	-49975,82176	n/a	7,55765831	0,5	0,25	0,25	0,25	0,25	
23	JC+G+I	27	100230,9807	100006,8907	-49970,44542	0,005460465	23,6648554	0,5	0,25	0,25	0,25	0,25	

D'après les résultats obtenus, le meilleur modèle à choisir c'est le T92+G avec T92=Tamura 3-parameter et G=Gamma distribution et le « Test of phylogeny » est le « Bootstrap method » qui rééchantillonne notre alignement de séquence et reconstruit un arbre de vraisemblance maximum et il le fait cent fois.



**Figure 39 :** Paramètres à modifier pour le traçage de l'arbre phylogénétique sur MEGAX



**Figure 40 :** Arbre phylogénétique des différentes séquences de nucléotides générés par MEGAX

##### 5. Le développement d'amorces :

Le développement d'amorces est conçu à partir des différents gènes étudiés selon l'outil « pickprimers », les résultats sont représentés par le tableau 7 qui présente les primers développés, les numéros d'accèsion des gènes, les « Forward » primers et les « reverse » primers, la teneur en GC, la

température d'hybridation  $T_m$  ainsi que la taille en paire de bases. Il faut mentionner que le bon primer retenu c'est le premier pour chaque gène.

Dans la section des paramètres d'amorce, c'est possible d'entrer nos propres séquences d'amorce. Ceci est idéal pour visualiser où sur le gène les amorces se lient. Mais nous laisserons ce blanc car nous ne connaissons pas encore les séquences d'amorces. Pour la PCR en temps réel, une taille de produit PCR optimale se situe entre 70 et 300 paires de bases pour une bonne révélation sur gel d'acrylamide et une taille au-delà de 300 paires de bases pour une bonne révélation sur gel d'agarose.

Au bas des paramètres de l'amorce, on peut spécifier la plage de température de fusion de l'amorce ( $T_m$ ). En règle générale, la température d'hybridation de la réaction PCR est d'environ 3 à 5°C inférieure de celle de la  $T_m$  de notre paire d'amorces. Par conséquent, une paire d'amorces avec une  $T_m$  de 62°C devrait, en théorie, avoir une température d'anneau de 57°C à 60°C.

**Figure 41 :** Ajustement des différents paramètres des amorces

(<https://www.ncbi.nlm.nih.gov/pickprimer>)

**Tableau 9 :** Description de quelques paires d'amorces des gènes étudiés générés par pickprimer.

Gènes	Amorces sens / antisens	$T_m$	%GC	Taille en pb	Taille du produit PCR
<b>BcPCK1</b>	F : CTATCGACAAACCACCCAAGC R : GCCATCTTTGAGGTGTAACCG	59.26 59.26	52.38 52.38	21 21	131
	F : TGGCTGAGAAGATTGAGCAC R : CTTGGAGGTTGAAGGTCTCG	58.06 57.91	50.00 50.00	20 20	111
<b>Bcg1</b>	F : AGAAGCTTCCCGTTAGTCCT R : GCTGGTTGAGACTGACGAAT	57.87 57.96	50.00 50.00	20 20	148
	F : ACGATCAAGATGTCCTTCGC R : CGGAGATAGCGACCAAGAAG	58.07 57.87	50.00 55.00	20 20	177
<b>Bcvel1</b>	F: GCAGACACTGATACCGCTTT R: TGTCGTGGATCGGTTTCATC	58.27 58.00	50.00 50.00	20 20	99
	F: CCTGAAACGATTACGAGGCA R: AAAGCGGTATCAGTGTCTGC	58.00 58.27	50.00 50.00	20 20	169
<b>BcLTF1</b>	F: GCATGTATGGCGATGACAGA R: CAGCGCGCTATTCAATTGTT	58.13 57.82	50.00 45.00	20 20	123
	F: TGGCTCAAGTTTGTCTCCA R: CTTATTGCGTGGAATGGCG	57.27 58.17	45.00 50.00	20 20	119
<b>BMP1</b>	F: ACGTGCATTTGTTGTAACCT R: GCAAAGAGAAGCGTAGCAAT	56.18 56.81	40.00 45.00	20 20	71

	F: GGGATGATGGAGGATGGAATG R: CAGGTTACAACAAATGCACGTA	57.92 57.52	52.38 40.91	21 22	157
<b>BcHBF1</b>	F: GCATTTGTACCTCGCTCAGA R: ATGTCGACTAGTGCTTGCTG	57.99 58.00	50.00 50.00	20 20	200
	F: GCCCAAGTTTGAAGAGGTCA R: TCTGAGCGAGGTACAAATGC	58.02 57.99	50.00 50.00	20 20	120
<b>BcATG1</b>	F: CGACAAAGAGATCGCAAGGA R: GCTCGGCTTCAGTTATGGAA	58.00 57.98	50.00 50.00	20 20	145
	F: AACCGGGTACCAATAATCGC R: ATTGGCAAGCCTTGAAGGAA	58.04 57.99	50.00 45.00	20 20	138
<b>Bcgb1</b>	_____	_____	_____	_____	_____
<b>Xyn11A</b>	F: TATTAGTGAGTGGTGAGTGGG R: ACCACCACTACACCCATTTG	47.62 50.00	47.62 50.00	21 20	122
	F: TTAGTGAGTGGTGAGTGGG R: CCACCACTACACCCATTTGT	55.97 57.71	52.63 50.00	19 20	119
<b>Bccyp2</b>	F: CCAATTTGAGCTCTTCGCCG R: TCTTGGTTTCGCCGGTACAA	59.90 59.89	55.00 50.00	20 20	77
	F: ACCGTACGCTTTGGACCTTT R: TTCTTGGTTTCGCCGGTACA	59.98 59.98	50.00 50.00	20 20	269
<b>Bac</b>	F: GCGGCTCTACTTCCATCCTC R: AGCTGCTTGAAAAACGCAG	59.97 59.97	60.00 50.00	20 20	253
	F: TTGTCCGGTGTGATTCTCCG R: GAGGATGGAAGTAGAGCCCG	60.04 59.97	55.00 60.00	20 20	196
<b>Bcbot</b>	F: GCGGCTCTACTTCCATCCTC R: AGCTGCTTGAAAAACGCAG	59.97 59.97	60.00 50.00	20 20	253
	F: CATCATGCAAGTCCTGTGCG R: TTGGTTCGTTCTTGGGGATGG	59.90 59.96	55.00 55.00	20 20	484
<b>Bcsod1</b>	F: TACCGTTGTTGTCCACTCCG R: TGGTCTGGTTCCAGCGTTAC	59.97 59.97	55.00 55.00	20 20	88
	F: GTAACGCTGGAACCAGACCA R: AGACACAGCCGGCAAATAA	59.97 59.89	55.00 50.00	20 20	145

## 6. Les marqueurs SSR:

Après avoir lancer la recherche des marqueurs SSR sur MISA, on a obtenu deux fichiers, le premier qui est le fichier.xls il est représenté par le tableau ci- dessous.

**Tableau 10 :** Les types SSR des différentes séquences des gènes de virulence du *Botrytis cinerea*

ID	SSR nr.	SSR type	SSR size	start	end
AF205375.1_BMP1_Botryotinia_fuckeliana_MAP_kinase_Tomate	1 p1	(T)12	12	2671	2682
XM_024693277.1_BcATG1_Botrytis_cinerea_B05.10_Pommes_raisin_et_tomate	1 p1	(A)10	10	113	122
XM_024693277.1_BcHBF1_Botrytis_cinerea_B05.10_Fraise	1 p1	(A)10	10	113	122
NC_037317.1_Bcgb1_Botrytis_cinerea_B05.10	1 p5	(ACACT)5	25	267	291
DQ057980.1_Xyn11A_Botryotinia_fuckeliana_Raisin_et_tomate	1 p6	(ATTGGA)6	36	883	918
DQ057980.1_Xyn11A_Botryotinia_fuckeliana_Raisin_et_tomate	2 p2	(AT)9	18	1207	1224
DQ057980.1_Xyn11A_Botryotinia_fuckeliana_Raisin_et_tomate	3 p1	(T)10	10	4787	4796
XM_024697499.1_Bac_Botrytis_cinerea_B05.10	1 p3	(CCA)6	18	60	77

Il existe plusieurs types de marqueurs SSR tout dépend du motif répété. Un motif de 1 paire de base dont le plus souvent on a répétition du A ou de T est appelé mono-nucléotidique, un motif avec 2 paires de base est appelé di-nucléotidique, celui de 3 paires de base est appelé tri-nucléotidique ainsi le motif avec 4 paires de base répétés est appelé tétra-nucléotidique. La longueur de ces séquences, c'est-à-dire le nombre de répétitions, varie selon l'espèce mais aussi d'un individu à un l'autre et d'un allèle à

l'autre chez un même individu voire d'une cellule à l'autre lors d'une erreur lors de la réplication de l'ADN.

D'après le tableau 10, on classe les marqueurs SSR selon le tableau ci-dessous.

**Tableau 11** : Classification des marqueurs SSR selon leur type

ID	Marqueur SSR	Type
AF205375.1_BMP1_Botryotinia_fuckeliana_MAP_kinase_Tomate	(T)12	Mono-nucléotidique
XM_024693277.1_BcATG1_Botrytis_cinerea_B05.10_Pommes_raisin_et_tomate	(A)10	Mono-nucléotidique
XM_024693277.1_BcHBF1_Botrytis_cinerea_B05.10_Fraise	(A)10	Mono-nucléotidique
NC_037317.1_Bcgb1_Botrytis_cinerea_B05.10	(ACACT)5	Penta-nucléotidique
DQ057980.1_Xyn11A_Botryotinia_fuckeliana_Raisin_et_tomate	(ATTGGA)6	Hexa-nucléotidique
	(AT)9	Di-nucléotidique
	(T)10	Mono-nucléotidique
XM_024697499.1_Bac_Botrytis_cinerea_B05.10	(CCA)6	Tétra-nucléotidique

L'analyse a montré que toujours les motifs mono-nucléotidiques avec 50% suivis par les motifs penta-nucléotidiques, hexa-nucléotidiques, tétra-nucléotidiques et di-nucléotidiques avec un même pourcentage de 12,5%. Des centaines d'amorces ont été générées pour toutes les séquences analysées. Le motif de types mono-nucléotidiques a été pris comme critère de choix, vue l'abondance de ces motifs chez *Botrytis cinerea*.

Pour le fichier statistics qui résume les différentes statistiques des microsatellites il est représenté par la figure ci-dessous.

```

Specifications
=====
Sequence source file: "df.fasta"

Definement of microsatellites (unit size / minimum number of repeats):
(1/10) (2/6) (3/5) (4/5) (5/5) (6/5)

Maximal number of bases interrupting 2 SSRs in a compound microsatellite: 100

RESULTS OF MICROSATELLITE SEARCH
=====
Total number of sequences examined:          13
Total size of examined sequences (bp):      35244
Total number of identified SSRs:            8
Number of SSR containing sequences:         6
Number of sequences containing more than 1 SSR: 1
Number of SSRs present in compound formation: 0

Distribution to different repeat type classes
-----

Unit size Number of SSRs
1         4
2         1
3         1
5         1
6         1

Frequency of identified SSR motifs
-----
Repeats  5    6    7    8    9    10   11   12   total
A         -    -    -    -    2
T         -    -    -    -    1    1
AT        -    -    -    1    1
CCA       1
ACACT    1
ATTGGA   1

Frequency of classified repeat types (considering sequence complementary)
-----
Repeats  5    6    7    8    9    10   11   12   total
A/T      -    -    -    -    3    1    4
AT/AT    -    -    -    -    1
ACC/GGT  1
ACACT/AGTGT 1
AATTCC/AATTGG 1

```

**Figure 42 :** Fichier.statistics décrivant les marqueurs SSR

Généralement, les marqueurs SSR sont considérés comme des locus spécifiques et par conséquent, toute amplification utilisant une paire d'amorce SSR devrait générer une ou deux bandes respectivement pour un allèle homozygote ou hétérozygote.

## Conclusion générale

La bio-informatique, la génétique, la génomique fonctionnelle et structurale sont des approches expérimentales complémentaires pour identifier des gènes d'intérêt *in silico* ou *in vivo* et *in vitro*.

L'étude des gènes *in silico* est une étape importante puisqu'elle permet de guider l'interprétation des résultats expérimentaux et de suggérer de nouvelles expériences.

Dans notre cas, les procédés bio-informatiques utilisés pour la conception d'amorces à partir des gènes candidats à la virulence que nous avons identifiés sont :

- La recherche et identification des gènes candidats,
- L'alignement,
- La construction d'un arbre phylogénétique,
- Le développement d'amorces,
- La conception de marqueurs SSR à l'aide de logiciel MISA.

Les résultats de notre étude, ont montré l'implication de plusieurs gènes de virulence différents à la pourriture grise du fraisier. L'arbre phylogénétique a regroupé les gènes des espèces ayant une similitude remarquable entre eux qui serait probablement due à l'ancêtre commun de tous les gènes étudiés.

Enfin, il nous semble indispensable dans l'étude du génome des champignons d'intégrer les différentes sources d'informations au sein de l'environnement informatique afin d'essayer de comprendre un peu le mystère de l'unité du vivant.

## Références

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J., 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215, 403-410,

Barnett H.L., Hunter B.B. (1998) *Illustrated genera of imperfect fungi*, fourth edition, APS press, Minnesota, USA, 218 p.

Sakhr Ajouz (2009). Estimation du potentiel de résistance de *Botrytis cinerea* à des biofongicides.

Braun, P.G. et J.C. Sutton. 1987. Inoculum sources of *Botrytis cinerea* in fruit rot of strawberries in Ontario. *Revue canadienne de phytopathologie*, 9 :1-5.

Braun, PG et Sutton, J. C 1988. Infections cycles and population dynamics of *Botrytis cinerea* in strawberry leaves. *Can. J. plant pathol.* 10: 133\_14.

Christoph CARLEN, Anne-Marie POTEL, Pamela CRESPO, Zo-Noroso ANDRIANJAKA-CAMPSet André ANÇAY, Agroscope, 1964 Conthey, Suisse. Facteurs de variation de la qualité gustative et nutritionnelle des fraises.

Coley-Smith, J.R. 1980. Sclerotia and other structures in survival, p. 85-114, in: *The biology of Botrytis*. J. R. Coley-Smith, K. Verhoeff and W. R. Jarvis, eds. Academic Press, London, UK.

Dean, R., Van Kan, J. A., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D., Rudd, J. J., et al. (2012). The top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol*, 13(4), 414-430. doi:10.1111/j.1364-3703.2011.00783.x. doi:http://dx.doi.org/10.1016/S0022-2836(05)80360-2).

Elad and Yunis, 1993; Yunis et al., 1990 Effect of nutrition on susceptibility of cucumber, eggplant, and pepper crops to *Botrytis cinerea*.

Elad Y, Williamson B, Tudzynski P and Delen N (2004) *Botrytis* spp. and diseases they cause in agricultural systems – an introduction. *Botrytis: Biology, Pathology and Control* (Elad Y, Williamson B, Tudzynski P & Delen N, eds), pp. 1–6. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Elyse Poitras and Alain Houde (2002). *Reviews in Biology and Biotechnology By The Moroccan Society of Biology in Canada. La PCR en temps reel: principes et applications.*

Escalona, M., Rocha, S., Posada, D., 2016. A comparison of tools for the simulation of genomic next-generation sequencing data. *Nature reviews. Genetics* 17, 459-469, doi :10.1038/nrg.2016.57.

Fidel Botero-Castro (2014). *Systématique, phylogénie et évolution moléculaires de la famille Phyllostomidae (Mammalia : Chiroptera) : une approche mitogénomique comparative.*

Fillinger, S., & Elad, Y. (2015). *Botrytis – the fungus, the pathogen and its management in agricultural systems*: Springer International Publishing.

Florian Philippe (2016). *Etude des sources de variation des composés d'intérêt nutritionnel chez la fraise.*

Gaston, Amélia 2010. Etude et compréhension du déterminisme génétique et moléculaire de la remontée florale chez le fraisier.

Haddaoui et Olfa, 2020. Étude d'un concept novateur de culture du fraisier en buttes semi-permanentes : évaluation d'un substrat adapté.

Holz, G., Coertze, S., and Williamson, B. 2004. The ecology of Botrytis on plant surfaces, p. 9-27, in: Botrytis: biology, pathology and control. Y. Elad, B. Williamson, P. Tudzynski and N. Delen, eds. Kluwer Academic Press, Dordrecht, The Netherlands.

How To Create Real-Time PCR Primers Using Primer-BLAST <https://toptipbio.com/real-time-pcr-primer-blast/>. Mai 2022.

<https://Bayer-Agri.fr>, 2018. Botrytis de la fraise, pourriture grise au printemps. Mai 2022.

Janet P Slovin, Kyle Schmitt and Kevin M Folta<sup>2</sup> 2009. An inbred line of the diploid strawberry *Fragaria vesca* f. *semperflorens* for genomic and molecular genetic studies in the Rosaceae.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 33(7):1870–1874.

Li L. and Yang H., 2011. First report of strawberry necrotic virus in China. *Plant Disease* 95 (9) : 1198.

Mahendra Mariadassou (2009). Robustesse des arbres phylogénétiques. Thèse Doct. Sci., Université Paris Sud XI, France.

Martin R. R., Tzanetakis I. E., Barnes J. E. and Elmhirst J. F., 2004. First report of strawberry latent ring spot virus in strawberry in United States and Canada. *Plant Disease* 88 (5): 575.

Mathias Choquer, Elisabeth Fournier, Caroline Kunz, Caroline Levis, Jean-Marc Pradier, Adeline Simon<sup>1</sup> et Muriel Viaud (2007) Botrytis cinerea virulence factors: new insights into a necrotrophic and polyphageous pathogen..

Mathieu Rousseau-Gueutin, Estelle Lerceteau-Köhler, Laure Barrot, Daniel James Sargent, Amparo Monfort, David Simpson, Pere Arús, Guy Guérin, Béatrice Denoyes-Rothan, 2008. Comparative Genetic Mapping Between Octoploid and Diploid *Fragaria* Species Reveals a High Level of Colinearity Between Their Genomes and the Essentially Disomic Behavior of the Cultivated Octoploid Strawberry.

Maxam, A. M., Gilbert, W., 1977. A new method for sequencing DNA. Proc Natl Acad Sci U S A 74, 560-4.

Mohamed El Oirdi (2009). Facteurs qui contrôlent le pouvoir pathogène chez *Botrytis cinerea*, Sherbooke, Québec, Canada.

O.M. Heide, 2015. Physiology and genetics of flowering in cultivated and wild strawberries.

Peterlongo, P., 2016. Lire les lectures : analyse de données de séquençage. Thèse en bio-informatique [q-bio.QM]. Université rennes1. Disponible depuis : <https://hal.inria.fr/tel-01278275/file/HDR.pdf>.

Pop, M., 2009. Genome assembly reborn: recent computational challenges. Briefings in Bioinformatics 10, 354-366, doi:10.1093/bib/bbp026.

Rabinowicz and Slovin, 2007; Shulaev et al, 2008. Flower and early fruit development in a diploid strawberry, *Fragaria vesca*.

Radford, A. D., Chapman, D., Dixon, L., Chantrey, J., Darby, A. C., Hall, N., 2012. Application of next-generation sequencing technologies in virology. The Journal of General Virology 93, 1853-1868, doi:10.1099/vir.0.043182-0.

Raous, A. S., 2013. Le séquençage haut débit : principe et applications en virologie. Pour le doctorat vétérinaire. Université Paris-Est Créteil Val de Marne, Île-de-France. Disponible depuis : <http://theses.vet-alfort.fr/telecharger.php?id=1845>.

Sanger, F., Nicklen, S., Coulson, A. R., 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74, 5463-7.

Sengenès, J., 2012. Développement de méthodes de séquençage de seconde génération pour l'analyse des profils de méthylation de l'ADN. Thèse en biologie moléculaire. Université Pierre et Marie Curie - Paris VI, 2012. Français. Disponible depuis : [https://tel.archivesouvertes.fr/file/index/docid/743905/filename/These\\_Jennifer\\_SENGENES.pdf](https://tel.archivesouvertes.fr/file/index/docid/743905/filename/These_Jennifer_SENGENES.pdf).

Théroux, J.-F., 2015. Développement de méthodes d'assemblage de génomes de novo adaptées aux bactéries endosymbiotes. Mémoire de maîtrise en bio-informatique. Université de Montréal. Disponible depuis :

[https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/13126/Theroux\\_Jean-Francois\\_2015\\_memoire.pdf?sequence=2&isAllowed=y](https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/13126/Theroux_Jean-Francois_2015_memoire.pdf?sequence=2&isAllowed=y). Juin 2022.



## Annexes

### Annexe 1 : Format GenBank

Exemple du gène Bcg1 :

```
LOCUS       XM_024693154                1260 bp    mRNA    linear    PLN 17-APR-2018
DEFINITION  Botrytis cinerea B05.10 Bcg1 (Bcg1), mRNA.
ACCESSION   XM_024693154
VERSION     XM_024693154.1
DBLINK      BioProject: PRJNA20061
            BioSample: SAMN02953628
KEYWORDS    RefSeq.
SOURCE      Botrytis cinerea B05.10
  ORGANISM  Botrytis cinerea B05.10
            Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina;
            Leotiomycetes; Helotiales; Sclerotiniaceae; Botrytis.
REFERENCE   1 (bases 1 to 1260)
  AUTHORS   Van Kan,J.A., Stassen,J.H., Mosbach,A., Van Der Lee,T.A., Faino,L.,
            Farmer,A.D., Papisotiriou,D.G., Zhou,S., Seidl,M.F., Cottam,E.,
            Edel,D., Hahn,M., Schwartz,D.C., Dietrich,R.A., Widdison,S. and
            Scalliet,G.
  TITLE     A gapless genome sequence of the fungus Botrytis cinerea
  JOURNAL   Mol. Plant Pathol. 18 (1), 75-89 (2017)
  PUBMED   26913498
REFERENCE   2 (bases 1 to 1260)
  AUTHORS   Staats,M. and van Kan,J.A.
  TITLE     Genome update of Botrytis cinerea strains B05.10 and T4
  JOURNAL   Eukaryotic Cell 11 (11), 1413-1414 (2012)
  PUBMED   23104368
REFERENCE   3 (bases 1 to 1260)
  AUTHORS   Anselem,J., Cuomo,C.A., van Kan,J.A., Viaud,M., Benito,E.P.,
            Couloux,A., Coutinho,P.M., de Vries,R.P., Dyer,P.S., Fillinger,S.,
            Fournier,E., Gout,L., Hahn,M., Kohn,L., Lapalu,N., Plummer,K.M.,
            Pradier,J.M., Quevillon,E., Sharon,A., Simon,A., ten Have,A.,
            Tudzynski,B., Tudzynski,P., Wincker,P., Andrew,M., Anthouard,V.,
            Beever,R.E., Beffa,R., Benoit,I., Bouzid,O., Braut,B., Chen,Z.,
            Choquer,M., Collemare,J., Cotton,P., Danchin,E.G., Da Silva,C.,
            Gautier,A., Giraud,C., Giraud,T., Gonzalez,C., Grossetete,S.,
            Guldener,U., Henrissat,B., Howlett,B.J., Kodira,C., Kretschmer,M.,
            Lappartient,A., Leroch,M., Levis,C., Mauceli,E., Neuveglise,C.,
            Oeser,B., Pearson,M., Poulain,J., Poussereau,N., Quesneville,H.,
            Rascle,C., Schumacher,J., Segurens,B., Sexton,A., Silva,E.,
            Sirven,C., Soanes,D.M., Talbot,N.J., Templeton,M., Yandava,C.,
            Yarden,O., Zeng,Q., Rollins,J.A., Lebrun,M.H. and Dickman,M.
  TITLE     Genomic analysis of the necrotrophic fungal pathogens Sclerotinia
            sclerotiorum and Botrytis cinerea
  JOURNAL   PLoS Genet. 7 (8), e1002230 (2011)
  PUBMED   21876677
REFERENCE   4 (bases 1 to 1260)
  CONSRTH   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2018) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   5 (bases 1 to 1260)
  AUTHORS   van Kan,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2017) Phytopathology, Wageningen University,
            Droevendaalsesteeg 1, Wageningen 6708PB, Netherlands
REMARK      Annotation update by submitter
```

---