# Study of viral diversity in Poaceae-based communities with contrasted plant richness

**Auteur :** Debue, Virginie
**Promoteur(s) :** Massart, Sébastien
**Faculté :** Gembloux Agro-Bio Tech (GxABT)
**Diplôme :** Master en bioingénieur : chimie et bioindustries, à finalité spécialisée
**Année académique :** 2018-2019
**URI/URL :** http://hdl.handle.net/2268.2/7676

LIÈGE université
Gembloux
Agro-Bio Tech

# Study of viral diversity in *Poaceae*-based communities with contrasted plant richness

**VIRGINE DEBUE**

**TRAVAIL DE FIN D'ETUDES PRESENTE EN VUE DE L'OBTENTION DU DIPLOME DE MASTER BIOINGENIEUR EN CHIMIE ET BIOINDUSTRIES**

**ANNEE ACADEMIQUE 2018-2019**

**PROMOTEUR : SEBASTIEN MASSART**

[1] Dans ce cas, l'autorité académique est représentée par le(s) promoteur(s) membre du personnel(s) enseignant de GxABT.

LIÈGE université
Gembloux
Agro-Bio Tech

# Study of viral diversity in *Poaceae*-based communities with contrasted plant richness

**VIRGINE DEBUE**

**TRAVAIL DE FIN D'ETUDES PRESENTE EN VUE DE L'OBTENTION DU DIPLOME DE MASTER BIOINGENIEUR EN CHIMIE ET BIOINDUSTRIES**

**ANNEE ACADEMIQUE 2018-2019**

**PROMOTEUR : SEBASTIEN MASSART**

# Acknowledgements

First of all, I would like to thank my promoter, Sebastien Massart, for allowing me to do my master thesis in his unit and for allowing me to deepen my knowledge of bioinformatics and phytopathology, for answering my questions and for supervising my work. Thank you for your advice when writing the work.

I would also like to warmly thank François Maclot, my supervisor, who has been able to support me throughout this work. Thank you for being so available for answering all my questions. Thank you also for brightening up my laboratory days with your many anecdotes, for giving me many tips and comments to improve the quality of my written work.

Let me thank Integrated and Urban Phytopathology Laboratory of the ULiège-Gembloux Agro-Bio Tech University and its staff for welcoming me during my master thesis. And a special thank you to Angelo, for his advice during my manipulations. Thanks also to Arnaud and Lucie for taking the time to answer my questions about bioinformatics.

A special thanks to you, my family, who supported me throughout my studies. Thanks to Julien who supported and encouraged me during this period, as well as Laurie, Lucas, Laetitia, Gauthier, Arnaud, Cyril, Kevin, Nicolas, Charles for your support since we joined the faculty. I also wanted to thank the residents of the "Chiée des glands", life has never been so beautiful with you. Thanks also to the AG committee 2015 and the soft team for many moments shared in recent years.

Master's thesis performed in the Integrated and Urban Phytopathology Unit of the ULiège-Gembloux Agro-Bio Tech University

# Abstract

Study of plant viruses is essential to address yield loss in crops due to diseases. However, viruses are most often studied in crops while viruses present in wild areas can also have a significant impact if the virus moves from one plot to another according to the different modes of transmission specific to each virus. It is therefore essential to better understand virus behaviour in these non-cultivated areas, and virus adaptation in their environment, through the identification of new virus species, new host plants and study of virus vectors. This will allow to better control disease risks to crops in case of virus transmission to them from wild reservoirs. In this study, prevalence of three plant viruses was examined, i.e. *Barley yellow dwarf virus* (*Luteovirus, Luteoviridae*), a new nepovirus (*Secoviridae*) candidate and a new waikavirus (*Secoviridae*) candidate three different plant communities (monoculture, pasture and grassland with high ecological value) within the Natural Park of Burdinale Mehaigne (Antheit, Province of Liège, Belgium). Virus prevalence was studied in plant communities as a whole as well as within some specific *Poaceae* species (*Poa trivialis* L. and *Lolium perenne* L.). Virus prevalence was studied using RT-PCR techniques to determine the presence of the virus in plant samples randomly collected from the plots. Co-infections between these three virus species were also analyzed. Bioinformatics analyses were also carried out on nepovirus and waikavirus candidates in order to determine if they represent some new virus species. Results showed a high prevalence (80-90%) of nepoviruses in wild plant communities, while any symptoms were observed. Bioinformatics analyses allowed study phylogeny of different consensus viral sequences established for each plant community. These sequences were then compared with each other but also with reference sequences from NCBI for *Nepovirus*, *Waikavirus* and *Sequivirus* genera (all belonging to *Secoviridae* family). Bioinformatics study showed that nepovirus is potentially a new virus species similar to *Tomato black ring virus* and *Beet ringspot virus*. The waikavirus shows a significant genetic difference with other waikaviruses. A last part of the project was to investigate host range of *White clover mosaic virus*. This virus, known to only infect *Fabaceae* plants such as clovers (*Trifolium repens* L.), was also detected by high throughput sequencing in *Lolium perenne* L. in a pasture in Héron (Province of Liège). Virus detection by RT-PCR in clovers and ryegrass confirmed sequencing data and allowed to extend host range of this virus species to *Poaceae*.

**Keywords:** plant virus ecology – metagenomics – wild *Poaceae* – crops – virus prevalence

# Résumé

L'étude des virus des plantes est essentielle pour remédier aux pertes de rendement des cultures dues aux maladies. Cependant, les virus sont le plus souvent étudiés dans les cultures alors que les virus présents dans les zones sauvages peuvent également avoir un impact significatif si le virus se déplace d'une parcelle à une autre selon les différents modes de transmission propres à chaque virus. Il est donc essentiel de mieux comprendre le comportement viral dans ces zones non cultivées et l'adaptation des virus dans leur environnement, par l'identification de nouvelles espèces virales, de nouvelles plantes hôtes et l'étude des vecteurs de virus. Cela permettra de mieux contrôler les risques de maladies pour les cultures en cas de transmission du virus à partir de réservoirs sauvages. Dans cette étude, la prévalence de trois virus végétaux a été examinée, à savoir le *Barley yellow dwarf virus* (*Luteovirus, Luteoviridae*), un nouveau nepovirus (*Secoviridae*) candidat et un nouveau waikavirus (*Secoviridae*) candidat dans trois communautés végétales différentes (monoculture, pâturage et prairie à haute valeur écologique) dans le Parc naturel Burdinale Mehaigne (Antheit, Province de Liège, Belgique). La prévalence du virus a été étudiée dans l'ensemble des communautés végétales ainsi que dans certaines espèces spécifiques de *Poaceae* (*Poa trivialis* L. et *Lolium perenne* L.). La prévalence du virus a été étudiée à l'aide de techniques de RT-PCR pour déterminer la présence du virus dans des échantillons de plantes prélevés au hasard sur les placettes. Les co-infections entre ces trois espèces de virus ont également été analysées. Des analyses bio-informatiques ont également été effectuées sur des candidats nepovirus et waikavirus afin de déterminer s'ils représentent de nouvelles espèces virales. Les résultats ont montré une prévalence élevée (80-90%) de nepovirus dans les communautés végétales sauvages, alors que des symptômes ont été observés. Les analyses bio-informatiques ont permis d'étudier la phylogénie de différentes séquences virales consensuelles établies pour chaque communauté végétale. Ces séquences ont ensuite été comparées entre elles, mais aussi avec des séquences de référence provenant de NCBI pour les genres *Nepovirus, Waikavirus* et *Sequivirus* (tous appartenant à la famille des Secoviridae). Une étude bio-informatique a montré que le nepovirus est potentiellement une nouvelle espèce de virus similaire au *Tomato black ring virus* et au *Beet ringspot virus*. Le waikavirus présente une différence génétique significative avec les autres waikavirus. Une dernière partie du projet consistait à étudier la gamme d'hôtes du *White clover mosaic virus*. Ce virus, connu pour n'infecter que des plantes de *Fabaceae* telles que le trèfle (*Trifolium repens* L.), a également été détecté par séquençage à haut débit sur *Lolium perenne* L. dans un pâturage à Héron (Province de Liège). La détection du virus par RT-PCR sur trèfle et ray-grass a confirmé les données de séquençage et a permis d'étendre la gamme d'hôtes de cette espèce de virus aux *Poaceae*.

**Mots-clés** : écologie des phytovirus - métagénomique - *Poaceae* sauvages - cultures - prévalence du virus

# Table of contents

# Table of figures

# Table of tables

# 1. Introduction

Study of plant viruses and their impact on cultivated species represent a significant issue in our society today. Indeed, as chemicals (usually used to control the virus vectors) are increasingly challenged, and resistances to insecticides appears, it becomes essential to better understand the mechanisms of virus multiplication, transmission and damage on crops. What is less studied, however, is the presence of these viruses in wilder environments as plants are more tolerant of virus infection in nature and usually show few or no symptoms. However, the presence of these viruses in meadows or pastures near crop fields could have an impact on these crops because some viruses could move from one plot to another through vectors actions, and therefore infect these crops [1].

*Poaceae* species are very common in meadows and pastures but also in crops because they are useful in the manufacture of basic products such as flour. It also serves to feed livestock. Studying the viruses in these types of plants is thus essential in order to preserve these crops and pastures [2] [3].

Moreover, virus diversity within wild plant communities remains poorly studied, many new viruses are discovered each year, as well as numerous variants or known virus species. Study of these viruses can, therefore, lead to the discovery of new virus genera and species. Then, it would be interesting to study their real impact on these plants [4].

Creation and evolution of new virus analysis techniques have also had a significant impact on their discovery and classification. Genomics and metagenomics allow a more in-depth and accurate study of the genome of these different viruses. It is now possible to quickly establish phylogenetic trees of these viruses and to study more easily their links with other viruses, but also to study the differences between the genomes. The development of bioinformatics makes these analyses accessible to many researchers and facilitates work in this field. Much progress has been made in this area in recent years [5].

This work is part of the PhD program of François Maclot: "Impact of ecosystem diversity on the *Poaceae* virome". The high throughput sequencing data used for the bioinformatics analyses are from the analyses carried out in previous years as part of this study. The plots studied are also the same. In this work, several aspects of the study of viruses in wild plants were addressed. On the one hand, the characterization of two new potential viral species close to *Waikavirus* and *Nepovirus* (*Secoviridae* family) by bioinformatics (in particular the Geneious program). The presence of these viruses in several different plant communities (fields, pastures and meadows of high ecological value) and certain species of *Poaceae* (*Lolium perenne* L., *Poa trivialis* L.) in three different sites (Antheit, Heron and Latinne) of the Burdinale-Mehaigne Natural Park (province of Liège) was studied and consensus genomes were established for each of these plant communities. For the study of nepoviruses, ryegrass (*Lolium perenne* L.) was chosen because, during the first sampling, it was the species with the highest prevalence of nepovirus infection. Then, the consensual sequences were compared with each other and with the reference genome of waikaviruses and nepoviruses to determine whether the viruses were similar between the different plant communities and whether or not they were close to the references. In addition, therefore, establish whether a new species or genus of the virus has been found.

The second significant aspect of the work was to study the presence and prevalence of these viruses in these different plant communities. Primers were developed from the consensus sequences and, using RNA extraction, and RT-PCR techniques, the presence of these viruses was determined in individual plants. Then, the data produced made it possible to study their prevalence in the plots, to study their presence or absence in certain major or minor species and even to determine specific co-infection profiles. This has highlighted some of the characteristics of these viruses.

Another part of this work, based on the analysis of the same HTS data, was devoted to the study of the *White clover mosaic virus* (Genus *Potexvirus*, Family *Alphaflexiviridae*) which was exceptionally detected by sequencing in ryegrass (*Lolium perenne* L.) of Heron pastures, whereas this virus species is generally present in legumes such as clover (*Trifolium repens* L.). In fact, clover and ryegrass samples from Heron pastures, total RNA extraction and RT-PCR were collected to better understand the host range of this virus and its prevalence in pasture.

# 2. Bibliography

## 2.1. Diversity and taxonomy of plant viruses

### 2.1.1. Definition

A virus is defined as "*a set of one or more nucleic acid template molecules, either RNA or DNA, normally encased in a productive coat or coats of protein or lipoprotein, that is able to organize its own replication only within suitable hots cells*" [6]. Viruses are hence classified according to the composition of their nucleic acid type [7] (See Figure 1).

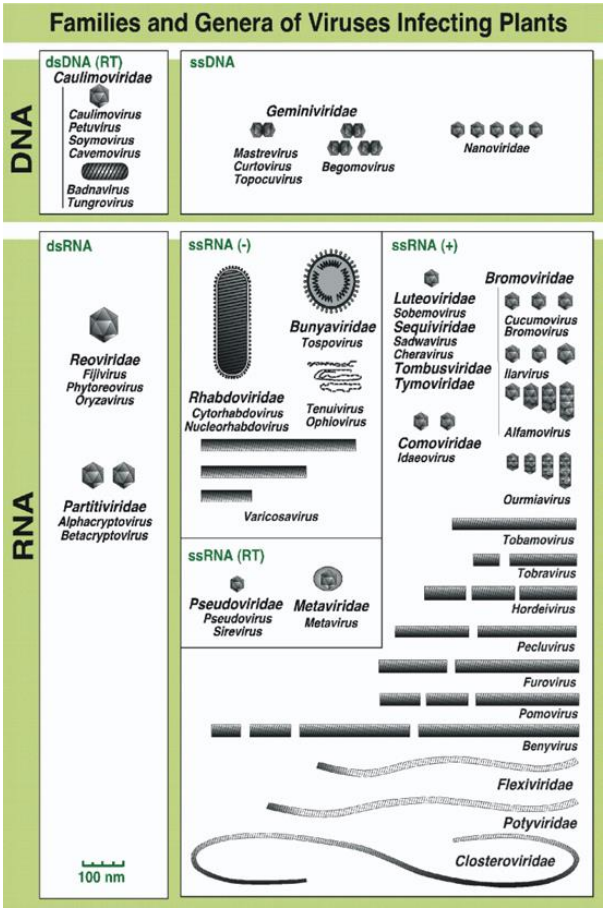



*Figure 1: Families and genera of viruses infecting plants [8]*

There is a vast diversity in plant viruses, and virus population are genetically very heterogeneous [8]. Several factors influence the genetic diversity of plant viruses. Several criteria are used to classify viruses. The classification may vary according to the criteria used. The main criteria are the vectors of the virus (means of transmission), hosts of the virus, the type of nucleic acid (RNA or DNA) and its properties, genome sequence of virus and proteins expressed by the virus [9].

Figure 2 shows a distribution of the different plant virus species recognized by the ICTV (International Committee on Taxonomy of Viruses). Some families have more different species than others, which shows the great diversity of viruses [10].

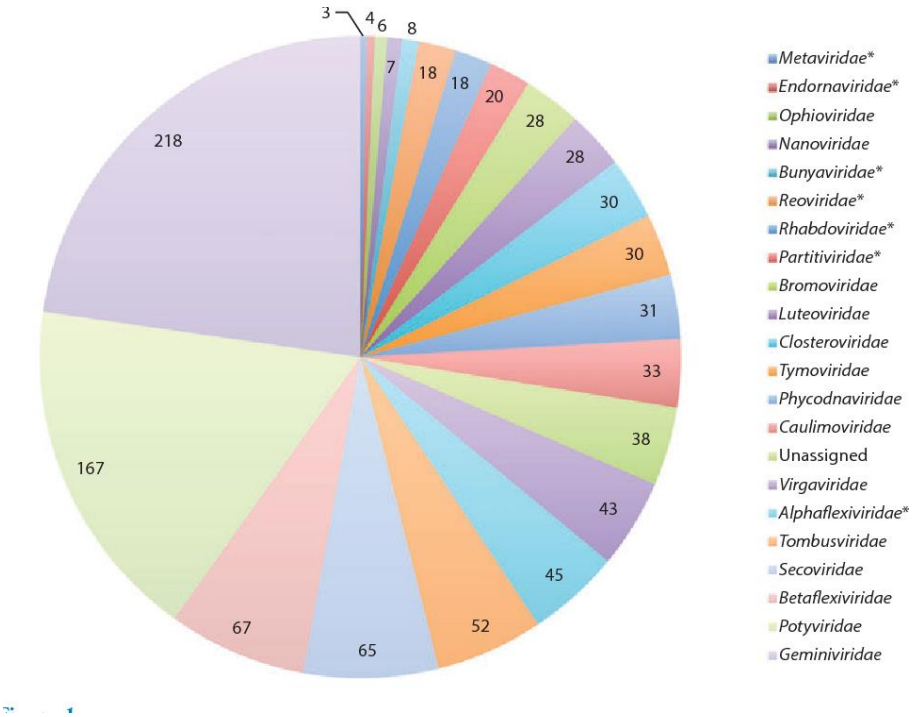



*Figure 2: Distribution of plant virus species by family [10]*

### 2.1.2. Structure of virus

Most plant virus species (62.2%) are constituted by ssRNA positive sense genome. Virus size can vary from 1kb per segment for multi-segment *Nanoviruses* to 20kb for *Closteroviruses*. A smaller proportion of plant viruses (23.4%) have a ssDNA form [11]. The genomes of viruses are very different from each other. However, some genes are found in nearly all genomes because they are essential to the life and reproduction of a virus. There are therefore three main parts: the replication part of the genome, a part for the transport from cell to cell of the virus and a last part, the capsid protein, which allows the assembly of the virus particle (see Figure 3) [12].

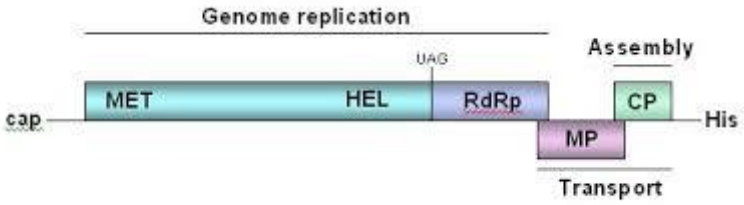



*Figure 3 : Different parts of tobamovirus genome (MET: methyltransferase, HEL: RNA helicase, RdRp: RNA-dependant RNA polymerase, MP: movement protein, CP: capsid protein) [12]*

### 2.1.3. Transmission and vectors

Viruses are transmitted by various types of transmission mode: insects, mechanical transmission, nematodes, seeds or fungal infection.

Virus vectors can have a significant impact on the dispersion and genetic variability of these viruses. Viruses which are transmitted at a short distance (vectors in the soil, contact between plants) have a patchy geographical distribution with a good structure in space. For viruses that are transmitted over a long distance (leafhoppers, etc.), the geographical structure of the virus distribution is less organized [13].

Virus transmission by insects occurs in two different patterns: non-persistent and persistent [14]. Non-persistent viruses are detected after insects have fed on an infected plant. Insect's infection is usually of short duration, but the vector can immediately infect another plant. This mode of transmission is present in most aphids. Persistent viruses are transmitted by insects after a period of latency or the virus cannot be transmitted. The virus stays in the body for a long time and can, therefore, be transmitted over a long period [15].

There are many different transmission mode or vectors for each virus. This also makes it easier for the virus to resist to selection pressure. Viruses can be transmitted by simple contact between an infected plant and a healthy plant (contact caused by human work, agricultural machinery, animals travelling through plots, etc.) [16], nematodes are also vectoring of some viruses [17]. Viruses that are transmitted in this way have a much smaller spread radius than what can be transmitted by insects (aphids, flies, leafhoppers, planthoppers, etc.). However, aphids remain one of the most critical vectors for plant viruses [18]. Some viruses are also transmitted by seeds, so the virus infects several generations of plants [19]. Viruses can also be transmitted through oomycetes [20] and fungal infection, for instance, Soil-borne wheat mosaic virus [21] (genus *Furovirus;* family *Virgaviridae* [22]).

There are therefore so-called "horizontal" and other “vertical” propagations. The type of spread also depends on the type of virus transmission vectors. A horizontal propagation corresponds to is a transmission between plants of the same generation. Vertical transmission refers to transmission by sexual or asexual (just by cuttings) reproduction of the plant, for example, transmission by seeds. A virus is then transmitted from generation to generation [23]. Transmission can also be done alternatively between horizontal and vertical transmission (See Figure 4) [24].

### 2.1.4. Plant-virus interactions in nature

Interactions between viruses and their host plants are very involved in nature. Indeed, wild plants can be infected by viruses without having any visible symptoms or these changes do not really impact the life of the plant. Many wild plants grow with viruses without really reporting any external symptoms. This makes it difficult to identify infected plants just by observing the plants. Also, wild plants may not have the same symptoms as cultivated plants [25] [26].

In addition, virus diversity is different between cultivated and wild plants. In wild plant environments, virus diversity is often higher than in cultivated plant communities [27] [26]. However, grapevines show that they are multi-infected with viruses while they are cultivated species [28].

It may also be that the virus is not too virulent because to survive, the plant must remain alive. So, he has no interest in his host disappearing. [27]

Plants can adapt over time to the presence of certain viruses, and a selection pressure may also be present in these environments. If the plant is not able to resist the virus, it cannot have the time to reproduce, and therefore, the weakest elements are eliminated [29].

### 2.1.5. Plant-virus-vector interactions

It is important not only to study viruses, environment or host but to examine it as a general interaction. Indeed, the virus cannot develop without a host, but the transmission of the virus and its evolution also depends on its environment and sometimes on the vectors present to transmit this virus [30]. Viruses must be studied with all the parameters that impact their life cycle. Indeed, as mentioned above, there is already a great impact if it is a wild plant or a cultivated plant. In studying viruses and virus ecology (virus ecology is defined as the study of the interactions between viruses, the environment in which they evolve and other organisms [31]), it is also important to take into account external factors: plant diversity, differences between the plant communities where plants are found, the different vectors that can transmit the virus and all other transmission means that can have an impact on the life of the virus [4]. It is also necessary to study the direct impact of the presence of the virus on the life cycle of the plant [4] [30].

The type of transmission can have an impact on the virulence of the virus, defined here "*as the harm it can inflict on its host*" [32]. Horizontal transmission of the virus will tend to increase the virulence of the virus because transmission depends on external vectors that are not always infected by the presence of the virus in their bodies. Vertical transmission will reduce the virulence of the virus because the plant may be reduced in its replication capacity by the presence of the virus in its body. If infected with a virus, a plant may have more difficulty reproducing, and the virus will, therefore, be less easily and effectively transmitted [32].

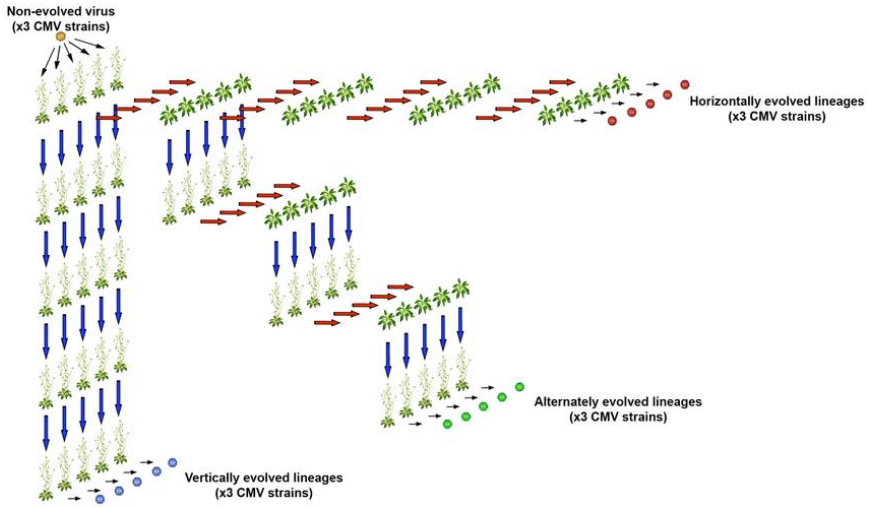



*Figure 4 : Vertical and horizontal transmission of plant virus [24]*

Viruses are able to adapt very quickly, resulting in rapid evolution of their genome and behaviour. They must adapt quickly to withstand the selection pressure. RNA viruses seem to be the most rapidly adapting and highly diverse due to a replication mechanism that causes a lot of error and creates many mutations and very similar species. DNA viruses also show some diversity, but the mechanisms that generate them are less well known at present [33].

### 2.1.6. Taxonomy

The taxonomy of viruses has evolved over time, in particular, thanks to new technologies allowing progress for virus classification. Virus classification began when the composition of virions began to be studied and understood in 1930. The study of the various components such as vectors and hosts allowed to set up some virus classifications, but which differed according to working committees or institutions. This is why the International Committee on Taxonomy of Viruses (ICTV) was created in 1966, in order to obtain a clear and universal classification and thus facilitate the study of viruses [34].

The taxonomy of viruses then followed the first criteria based essentially on virus biology, as in Figure 5. Thanks to genome studies, new criteria for virus classification have been developed, such as phylogeny and divergences between different genomes. Biological criteria continue to be also used. Nowadays, scientists and ICTV are reconsidering virus classification accepting new species only based on a sequenced genome using metagenomics. Indeed, technologies are still evolving, and therefore, the taxonomy of viruses follows this evolution [35].

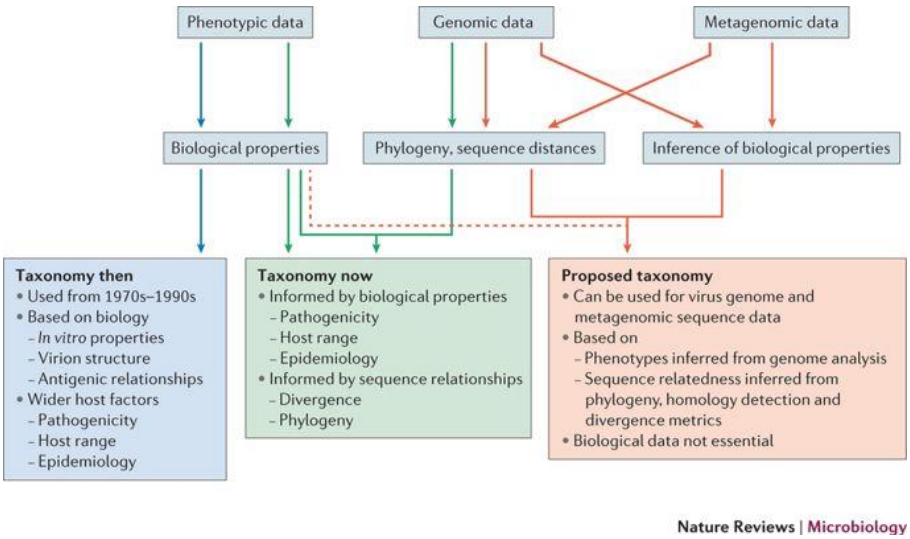



*Figure 5 : Past, present and future of taxonomy virus: Technologies evolution [35]*

## 2.2. Viruses infected *Poaceae* plants

### 2.2.1. *Poaceae* family

*Poaceae* is one of the most diversified plant families, containing 700 genera and 10,000 species. *Poaceae* plants are found in most areas of the Earth. They can grow at very various altitudes, soil types and temperatures. *Poaceae* include rice, maize, cereals, which are essential crops for human food [2]. *Poaceae* are very important in the human food sector because 60% of the calories consumed come from only 3 plants: rice, corn and wheat [36]. Cereal cultivation began more than 10,000 years ago, and more than thirty species have been tamed by humans. In addition to feeding many different populations on Earth, *Poaceae* are also the staple food of many wild animals or livestock. *Poaceae* are therefore essential on Earth and studying their viruses is essential. *Poaceae* are also used in the pharmaceutical and pesticide sectors, in sugar production but also in many other sectors [37].

### 2.2.2. Viruses structure of *Poaceae*

Almost 200 viruses infect *Poaceae,* but this figure may still change as new viruses are discovered regularly [2]. As shown in Figure 6, two main capsid structures (used to protect viral nucleic acids) are the icosahedral and helical structures (see Figure 7). The size of the capsid

depends essentially on the length of the virus genome. In Figure 6, the second graph shows that the most common type of nucleic acid represented in the *Poaceae* virus family is single-strand RNA in positive sense (RNAss+), which are directly translated by the cell++ [2] [38].

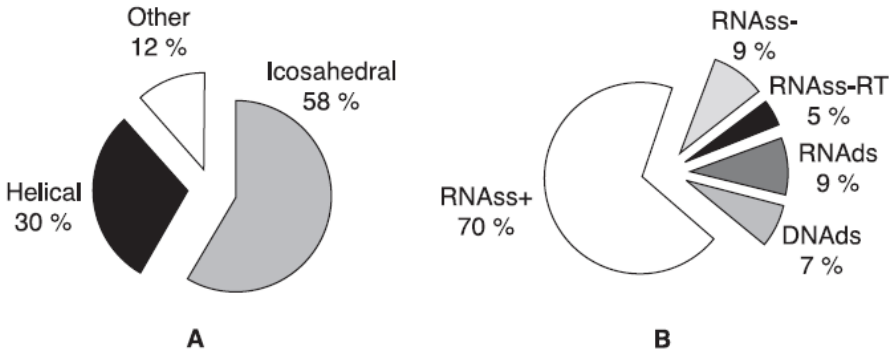



*Figure 6 : Percentage of each structure of capsid(A) and nucleic acid type (B) of viruses infecting Poaceae family [2]*

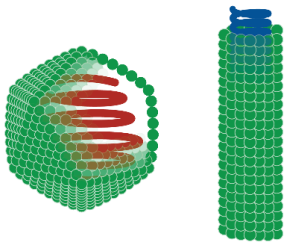


*Figure 7: On the left is the icosahedral structure and on the right is the helical structure, the two structures most represented in Poaceae viruses [39]*

### 2.2.3. Transmission and vectors

Like most viruses, *Poaceae* viruses are transmitted by different types of agents. Their mode of transmission depends on the virus species, and a virus can have several modes of transmission. But the most common agents for virus transmission in *Poaceae* are nematodes and arthropods [2].

#### *2.2.3.1. Mechanical transmission*

Mechanical transmission through contact between two plants is infrequent in *Poaceae*, but grazing animals can be a cause of mechanical transmission because they break several plants when eating and therefore transmission of the virus is easier. Two examples of a virus with mechanical transmission: *Barley stripe mosaic hordeivirus* (Family: *Flexiviridea*, genus: *Potexvirus*) with plant to plant contact and *Cocksfoot mottle sobemovirus* (genus : *Sobemovirus*) with grazing animals[2].

#### *2.2.3.2. Arthropod transmission*

Most *Poaceae* viruses can be transmitted by arthropods, in particular, insects such as homopters (e.g. aphids, leafhoppers). Viruses transmitted by homopterous have three main characteristics: a virus type is transmitted by a single family of homopterous (although there are some exceptions), no viruses are transmitted by vectors from more than one family of homopterous and the mode of transmission remains the same for a virus type (See Figure 8) [40].

| Arthropod group | Common name | Virus genera | Transmission mode |
|---|---|---|---|
| *Acari Eriophyidae* | mites | *Rymovirus*<br>*Tritimovirus* | C |
| *Insecta*: | | | |
| *Coleoptera Chrysomelidae* | beetles | *Machlomovirus*<br>*Sobemovirus* | NC / C |
| *Homoptera Aphididae* | aphids | *Carlavirus*<br>*Closterovirus*<br>*Cucumovirus*<br>*Luteovirus*<br>*Polerovirus*<br>*Potyvirus*<br>*Sobemovirus* | NC<br>NC<br>NC, capsid strategy<br>C<br>C<br>NC, helper strategy<br>NC |
| *Homoptera Pseudococcidae* | mealybugs | *Badnavirus*<br>*Ampelovirus* | NC<br>NC |
| *Homoptera Cicadellidae* | leafhoppers | *Cytorhabdovirus?*<br>*Nucleorhabdovirus*<br>*Marafivirus*<br>*Mastrevirus*<br>*Phytoreovirus*<br>*Tungrovirus*<br>(cica) *Tenuivirus*<br>*Waikavirus* | P<br>P<br>P<br>C<br>P<br>NC, helper strategy<br>P<br>NC, helper strategy |
| *Homoptera Fulgoroidea* | *planthoppers* | *Cytorhabdovirus*<br>*Fijivirus*<br>*Nucleorhabdovirus?*<br>*Oryzavirus*<br>*Tenuivirus* | *P*<br>*P*<br>*P*<br>*P*<br>*P* |
| *Thysanoptera* | thrips | *Tospovirus* | P |

*Figure 8 : Transmission mode and arthropod group for some virus genera (NC: noncirculative mode, C: circulative mode, P: propagative mode) [2]*

There are several ways in which viruses are transmitted by arthropods (see Figure 9), based on Harris (1977) classification [41]:

- Non-circulative and semi-persistent: The virus is placed in the insect oesophagus and binds to receptors. There is no virus circulation within the insect [42].
- Non-circulative and non-persistent: The virus binds to the insect's stylus; the receptors can directly recognize the capsids (capsid strategy), or there may be helper factors that make the virus recognized even if it does not match the receptor [2].
- Circulative, non-propagative: the virus is located in the insect's salivary gland and comes from the digestive system. There is no virus reproduction in the insect [42].
- Circulative, propagative: The virus infects many of the insect's organs, and there may be a reproduction of the virus [43].

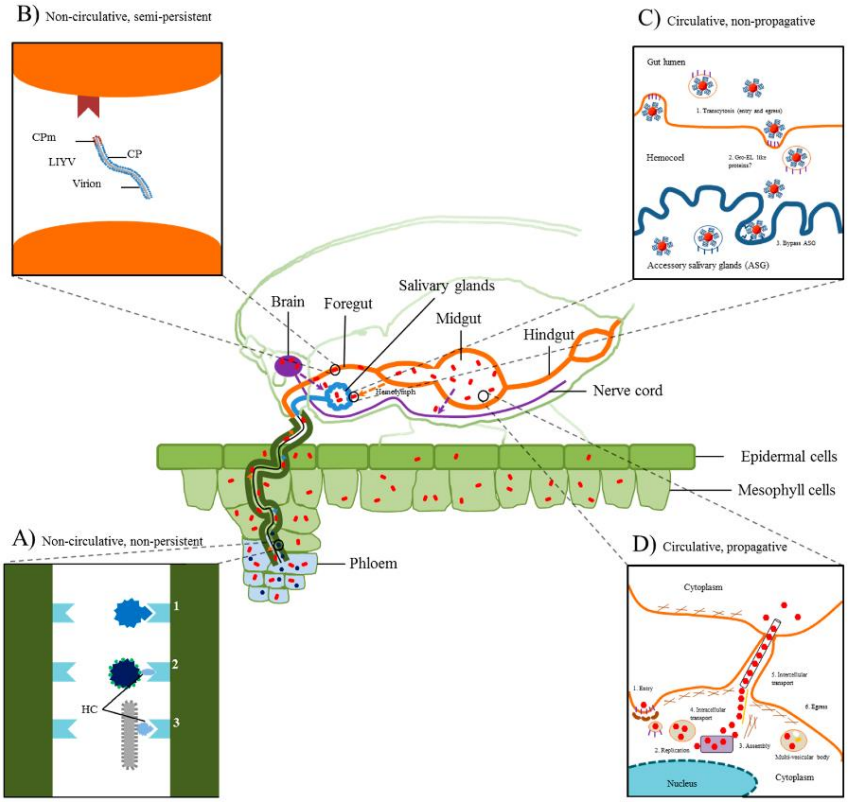



*Figure 9 : Different transmission of plant virus by arthropods [42]*

#### *2.2.3.3. Nematode transmission*

Only two *Poaceae* virus genera and only a few species from these genera are transmitted by nematodes while four thousand plant virus species are known to be transmitted by this route [44]. The two genera concerned are Nepoviruses and Tobraviruses. Viruses transmitted by this route are transmitted very locally because nematodes do not travel long distances. It is, therefore, often possible to observe patches of diseased plants in the plots studied [2].

#### *2.2.3.4. Seed transmission*

The transmission of viruses by seeds allows vertical transmission from generation to generation. It is a simple and effective way for the virus to be transmitted and to stay from generation to generation and not disappear with the death of the plant [45].

Viruses generally infect all plant organs, but only fifteen viruses infecting *Poaceae* can be seed transmitted because some viruses cannot move beyond the phloem and therefore the virus never reaches the seeds. The infection of the embryo must also be done in the first stages of growth of the plant because it allows better transmission of the virus. However, not all seeds are always infected [2]. The transmission of the virus to the seed also depends on the number of cytoplasmic connections that exist between the mother plant, flower and seed. The more cytoplastic connections, the more effective the transmission of the virus to the seed [46].

### 2.2.4. Agronomic aspects and control of diseases

To control the transmission of viruses by insects, it is possible to use pesticides that are very effective against insect vectors and therefore reduce virus spread [47]. However, some arthropods show resistance to certain pesticides, and therefore, there are problems in controlling virus transmission [48]. Other solutions are being considered, such as GMOs, but they are still

being studied and not accepted everywhere. Other techniques are used, such as integrated pest management using predators or parasites of the vectors to reduce virus transmission [2].

This work is focused on *Nepovirus,* and *Waikavirus* genera, as well as *White clover mosaic virus,* found to infect *Poaceae*:

### *2.2.5. Nepovirus*

Nepoviruses are (+) ssRNA viruses belonging to *Picornavirales* order, *Secoviridae* family and *Comoviridae* subfamily [49].

Nepoviruses were not known to infect *Poaceae* until the discovery of Arabic mosaic virus on winter barley in 1991. Today, the virus is present in *Poaceae* family [2] but also in *Vitaceae* family [50].

Virion structure should be studied at the family level. *Secoviridae* virion is formed by an icosahedral symmetry, in a non-enveloped (without a protective protein capsid [51]) form and with a length of 25 to 30 nm., *Comovirinae* sub-family is constituted by 2 RNAs, as shown in Figure 10. RNA2 length varies significantly in the different subgroups of nepoviruses. The coat protein is found on RNA 2, while 3C-like proteinase (Pro) and RNA-dependent RNA polymerase (Pol) are on RNA1 [52]. RNA 1 has the necessary information for replication of the virus, while RNA 2 has the information for cell to cell movement and coat protein (See Table 1) [53].

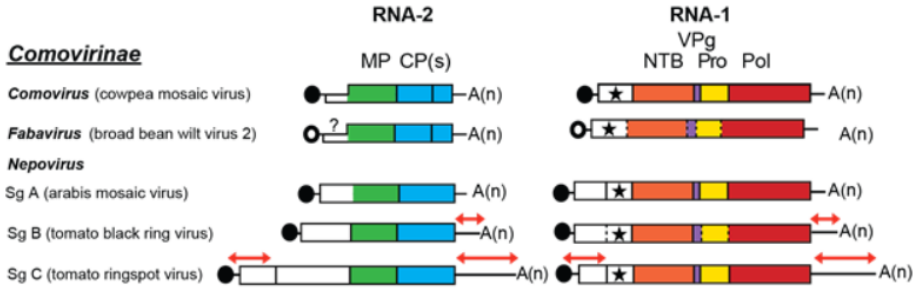



*Figure 10 : Genome organisation of Comovirinae [52]*

*Table 1: Polyproteins of Nepovirus genome [53]*

| **Polyprotein of RNA1** | | **Polyprotein of RNA 2** | |
|---|---|---|---|
| **Peptides** | **Size (kDa)** | **Name** | **Size (kDA)** |
| Protease co-facteur | 63 | Movement protein | 93 |
| Nucleotide-binding protein | 72 | Coat protein | 57 |
| Genome-linked protein | 2.3 | | |
| Protease | 23 | | |
| Polymerase | 92 | | |

Three subgroups are present within *Nepovirus* genus: Subgroup A with RNA-2 of 3,700-4,000 nt in length, subgroup B with a length of 4,400-4-700 nt in length and finally subgroup C with a length of 6,400-7,300 nt in length. On the other hand, RNA1 size is 7.5 kb. The three subgroups are divided according to viral genomes, organization and the cleavage zones [54]. These subgroups are illustrated in Figure 11.

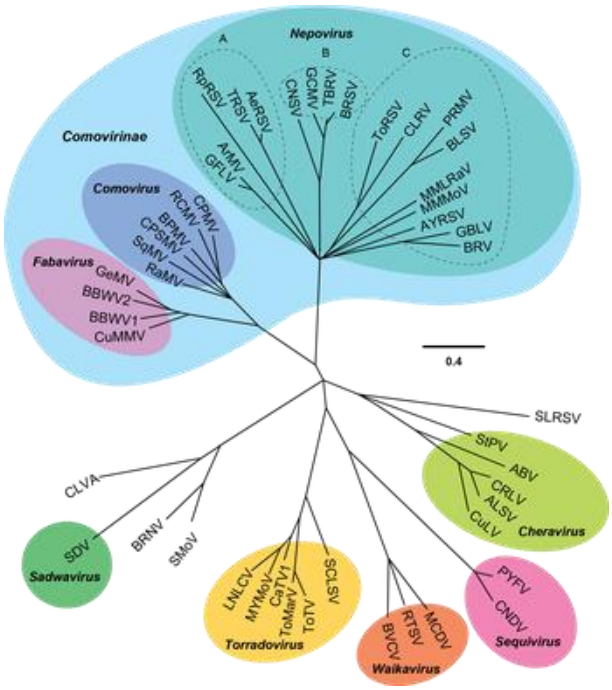



*Figure 11 : Phylogentic tree of Secoviridea family [52]*

Various symptoms of *nepovirus* infection can be observed, most commonly ringspot (see Figure 12) , as well as mottling and spotting [54].


*Figure 12: Ringspot of Tobacco ringspot virus [55]*

Vectors of *Nepovirus* are nematodes living in soil, in particular, the *Longiviridae* family (genus: *Longidorus, Xiphinema*) [56]. Pesticides can be used to eliminate nematodes in fields soil and suppress the transmission of nepoviruses. But it is more and more complicated to suppress nematodes totally as pesticides have been overused in last years and some are now banished. Nematodes could hence become a significant problem for agriculture in the future, and new alternatives should be found [2]. Due to their ability to adapt, nepoviruses are also transmissible through pollen [54] (e.g. *Tomato black ring virus* in *Rubus* spp. and *Raspberry ringspot virus* in *Fragaria* spp. [57] ) and mites for one species: Blackcurrant reversion virus [58].

This high capacity for adaptation and evolution has for origin their genome-based of RNA and deriving factors such as [59] intra- and interspecific recombination between genomes of different species that can then create a new species or genotype [60] [61], a very high negative selection and finally [62], or a rearrangement between different viruses that increases genetic variability [63].

### *2.2.6. Waikavirus*

Waikaviruses are (+) ssRNA viruses belonging to *Picornavirales* order and *Secoviridae* family. This monopartite virus has ~12 kb genome length. Its virion has a diameter length of 30 nm and is close to Sequiviruses for which they share genome structure [64] [65]. Conversely, *Waikavirus* genus includes fewer species than the *Nepovirus* genus, on the opposite situation in phylogenetic trees. (See Figure 11).

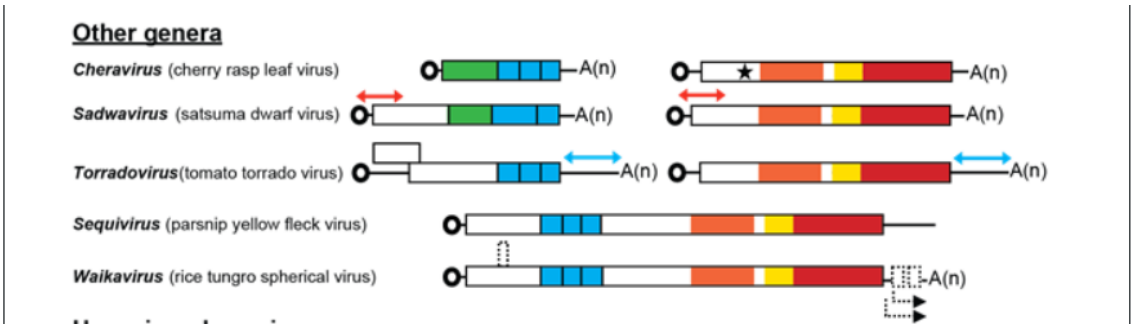



*Figure 13 : Organisation of genome of other genera of Secoviridae family [52]*

*Waikavirus* genome is monopartite but CP zones and the 3C-like proteinase (Pro) and RNA-dependent RNA polymerase (Pol) zones, respectively in blue, yellow and red in Figure 13, are in the same location as on nepovirus genome. The coat protein is divided into three parts in waikavirus; this specificity distinguishes them from nepoviruses [52].

These three coat proteins are assembled to form the icosahedral capsid (See. Figure 14). The virions are also non-enveloped [64].

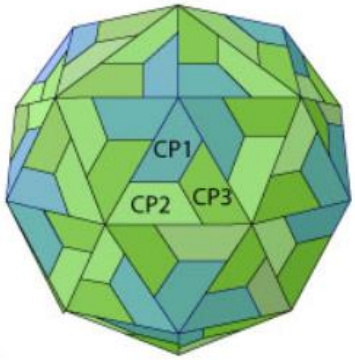



*Figure 14 : Structure of waikavirus virion [64]*

Waikaviruses hosts are *Poaceae* for two species of virus: *rice tungro spherical virus* and *maize chlorotic drawf virus*. A third virus (Anthriscus yellows virus) infects Eudicots [2].

Symptoms are vein yellowing, chlorotic stripes mosaic, and plant stunting and yellow flecks on natural hosts for *Maize chlorotic dwarf virus* (See Figure 15). For *Rice tungro spherical virus*, plants infected by the virus do not show symptoms in most types of rice used in crops. It is, therefore, more complicated to identify it [2].

*Figure 15 : Maize chlorotic dwarf virus symptom [66]*

Waikaviruses vectors in *Poaceae* are leafhoppers and aphids, in a semi-persistent manner. A virus-encoded helper protein is needed [67]. Some waikaviruses can help in the transmission of other viruses such as *Rice tungro spherical virus* for *Rice tungro bacilliform virus* (*Caulimoviridae* family) which can present essential issues as it can cause massive damage to rice crops [67]. There is no seed transmission reported for this virus [52].

### 2.2.7. New species or genera in *Secoviridae* family

Some criteria defined by the ICTV should be met in order to affirm a virus genome as a potential new species or genus within the *Secoviridae* family [52].

Criteria for a new genus consist an of different number of genomic RNAs, protein domains, coat proteins, or additional ORFs, which should be represented by a new branch in the phylogenetic tree presented in Figure 11. This tree compares the Pol-Pro regions of *Secoviridae* but only from a CG sequence to a GDD sequence that is about 500 bp. The entire Pol-Pro is not used for this comparison. Not all the criteria mentioned above have to be met simultaneously to qualify for a new genus [52].

On the other hand, criteria for new species consist on an identity percentage in amino acid sequences of less than 75% and 80% for the CP and the Pol-Pro respectively (when compared to other known virus species), new host species, different vectors, absence of cross-protein, differences in the antigenic reactions and for viruses divided into two parts, absence of this division. As with the genus, not all criteria must be met to have a new species [52].

### *2.2.8. White clover mosaic virus*

*White clover mosaic virus* (*Alphaflexiviridae* family, *Potexvirus* genus) is a (+) ssRNA virus with a genome length of 5.9 to 7 kb long. The virion is in the non-enveloped form. It has a helical shape of 470-1000 nm in length and 12-13 nm in diameter [68].


*Figure 16 : Structure of Potexvirus genome (RdRp : RNA-depedant RNA polymerase, TGB : Triple Gene Block, CP : capsid protein) [68]*

Coat protein is located at the end of the genome, on the side of the 3' part, this CP is between 0.8 and 1 kb [69]. There is also a large RNA-dependent RNA polymerase (RdRP) gene used for replication of RNA (150 to 181 kDa). Between ORF1 and ORF5, there are three additional ORFs: TGBs that are used for cell to cell movement of the virus. They are 25, 12 and 8 kDa, respectively (see Figure 16) [70].

This virus infects *Fabaceae* and has never been detected in *Poaceae*. It infects plants such as *Trifolium spp., Phaseolus vulgaris, Vicia faba, Vigna sinensis, Pisum sativum* and *Cucumis sativus.* It has different symptoms depending on the plant species. In clovers (*Trifolium repens* L.), symptoms consist of diffuse or sometimes clear mosaics on leaves. There may be necrotic lesions as well. Infected plants have different symptoms: mosaic and ringspot symptoms are the most common (see Figure 17) [71].

*Figure 17 : White clover mosaic virus symptoms [72]*

*White clover mosaic virus* is known to be transmitted mechanically, by grazing animals, tractor passages or any other factors that can damage the plant and allow this virus to pass from an infected plant to a healthy one [73]. There is a low capacity for virus transmission through seeds [74]. Insects do not appear to be a vector of this virus [71].

This virus is very well studied in New Zealand because clover is very important in the general economy of this country. Indeed, clover has several very important aspects of the preservation of the environment and agriculture. It fixes the nitrogen present in the atmosphere, has a high nutritional value for some cattle and even participates in the production of honey. This virus can cause significant damage to clovers (loss of 36.5% dry matter) and create significant economic losses [75] [76]. It is extensively studied in other countries, such as Korea [77].

Criteria for differentiating species in this genus (*Potexvirus*) are host differences for viruses, different serology, transmission between different species (species cannot protect themselves from the virus, sequence differences, i.e. identity percentage in amino acid sequences of less than 80% for coat protein or polymerase genes [70].

## 2.3. Bioinformatics and new technologies

### 2.3.1. History

The first step is genome sequencing. Sequencing is defined as *"the process of discovering the order in which nucleotides (i.e. chemical substances) are combined within DNA* " [78]. Several techniques have been used or are still used in the laboratories. The Sanger sequencing method is the first technique used for sequencing and is still used today [79]. The 454 pyrosequencing (Roche technology) was the first high throughput sequencing technology developed and was based on pyrosequencing (detects pyrophosphatase during DNA synthesis) [80]. This technique allowed to analyse long sequences but has disappeared due to its more expensive cost [81]. Illumina sequencing is a technology which uses the fluorescent emission from incorporated dye-labelled nucleotides method. The Illumina technique is a high-performance, fast and inexpensive technique. [82] Today, recent HTS techniques, as Oxford Nanopore Technology, have been developed. The nanopore technique consists in immersing the nanopores (transmembrane cellular proteins with small pores) in a conductive liquid, when a nucleotide derived from DNA passes through the nanopore, the current is modified. These current changes allow DNA sequences to be sequenced [83]. The technology used to sequence RNA in viruses is the Illumina technology.

Evolution of sequencing methods facilitated the sequencing of many genomes. Technology has become affordable in terms of cost. An Illumina sequencing costs $5.97 per Mb. Many different genomes have been studied, and the use of this technology is increasing. For example, it is now

possible to sequence a human genome in one day. The speed of analysis is also an important parameter [84].

It is important to distinguish next-generation sequencing (NGS) and high throughput sequencing (HTS): NGS is new generation sequencing, and HTS is high-throughput screening. It is important to differentiate between the two terms because it does not mean the same thing but are two important components of bioinformatics analysis. Today, HTS is preferred to NGS [85].

### 2.3.2. Vocabulary of bioinformatic

First, here are some definitions of terms used in the field of bioinformatics. These terms are useful for understanding the rest of this chapter (see Table 2).

*Table 2 : Vocabulary of bioinformatic [85]*

| Word | Definition |
|---|---|
| Scaffold | *Two or more contigs joined together using read-pair information* |
| *De novo* assembly | *Refers to the reconstruction of contiguous sequences without making use of any reference sequence* |
| Alignment | *Similarity-based arrangement of DNA, RNA or protein sequences. In this context, subject and query sequence should be orthologous and reflect evolutionary, not functional or structural relationships* |
| Contig | *A contiguous linear stretch of DNA or RNA consensus sequence. Constructed from a number of smaller, partially overlapping, sequence fragments (reads)* |
| Mapping | *A term routinely used to describe alignment of short sequence reads to a more extended reference sequence* |
| Read | *Short base-pair sequence inferred from the DNA/RNA template by sequencing* |
| RNA-Seq | *High-throughput shotgun transcriptome (cDNA) sequencing. Usually not used synonymously to RNA-sequencing which implies direct sequencing of RNA molecules skipping the cDNA generation step* |
| SNP | *Single-Nucleotide Polymorphism* |
| Annotation | *The computational process of attaching biologically relevant information to any sequence data* |
| Library | *Collection of DNA (or RNA) fragments modified in a way that is appropriate for downstream analyses, such as high-throughput sequencing in this case* |
| NGS | *High-throughput sequencing nano-technology used to determine the base-pair sequence of DNA/RNA molecules at much larger quantities than the previous end-termination (e.g. Sanger sequencing) based sequencing techniques* |
| Mate-pair | *Sequence information from two ends of a DNA fragment, usually several thousand base-pairs long* |
| Paired-end sequencing | *Sequence information from two ends of a short DNA fragment, usually a few hundred base pairs long* |

### 2.3.3. Bioinformatic and plant virus

New technologies change the study of a genetic sequence. High throughput sequencing has made significant progress in the field of plant virology. Indeed, thanks to new technologies, it is possible to identify and sequence small genomes of viruses. It is then possible to study the sequences of these, and this has made it possible to make a lot of progress in this field [86].

The advent of bioinformatics and these advances has made it possible to identify many viruses and to create significant databases that allow comparisons between viruses. This makes it easier to identify viruses found in plants. Virus analysis is complicated because nucleic acid samples are often contaminated by the molecules of the virus-host. Bioinformatics makes it possible to identify the viral sequences contained in samples after shotgun sequencing, for example. The evolution and automation of techniques will continue to evolve and will make it possible to analyze and discover new viruses and thus, further increase existing databases [87].

### 2.3.4. Bioinformatic steps

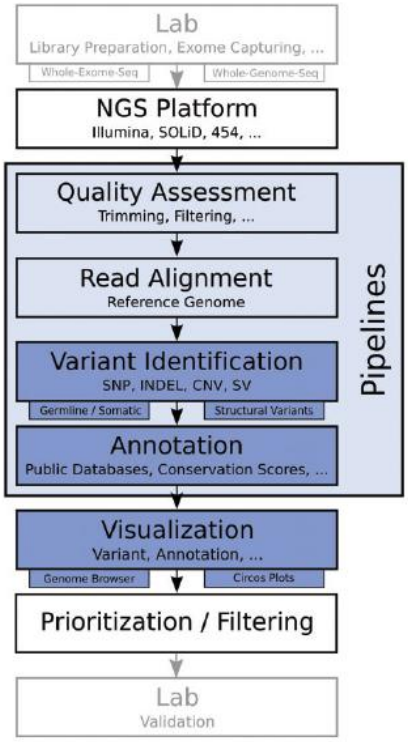



*Figure 18 : Workflow of bioinformatic [88]*

After sequencing, bioinformatic analyses can be performed, according to a work plan presented in Figure 18. It is first necessary to evaluate the quality of the sequencing with the trimming (trimming is used to process paired-ends) and remove the read duplicates present [88]. Duplicates are sometimes introduced during library preparation. They are due to different factors. It is important to remove them to avoid putting a bias in the rest of the analysis [89]. Then, the filtered reads are de novo assembled [90]. There are several types of assemblers: "SPAdes" which allows assembling the data of a single-cell but also to be a multicell assembler [91], "Velvet" that allows creating contigs from small reads [92], There are still other algorithms such as SOAP-denovo, MIRA, etc. [93]. The assembly is followed by the taxonomical assignment step, where reads are aligned with reference sequences using BLAST (Basic Local Alignment Search Tool). Then comes the annotation of the genome to recognize

the different functions and parts of useful sequence in the genome. BLAST are used to annotate the sequences obtained. The tool allows finding the similarities of the different zones of the sequence studied from a biological database. It also makes it possible to calculate the rate of similarities between regions. There are different types of BLAST: BLASTn to study nucleotides, BLAST protein to study the proteins in the sequence, BLASTx that compares with the nucleotide sequence translated with proteins and tBLASTn that does the opposite of BLASTx [94]. After all these steps, it is possible to carry out laboratory analyses to validate the sequences obtained [88]. It is also possible to perform SNP analyses to determine nucleotide variations [95].

In this study, data validation is performed after designing PCR primers. PCR (polymerase chain reaction) is used to amplify sequences between two primers. This amplification is achieved through different temperature cycles. Once the sequences have been amplified, they can be viewed on gels by fluorescence. A marker (fluorescent molecule like ethidium bromide) is used to highlight the presence of amplified sequences in an agarose gel. Ethidium bromide is an intercalating agent that will rise under UV lamps, and therefore, amplicons will be visible [96]. Real-time PCR is another type of PCR that allows DNA to be quantified in a sample. It allows quantifying the number of amplicons. This amount is calculated at each cycle of PCR in real-time. Amplicons are quantified using a fluorescent marker [97].

### 2.3.5. Consensus sequence formation

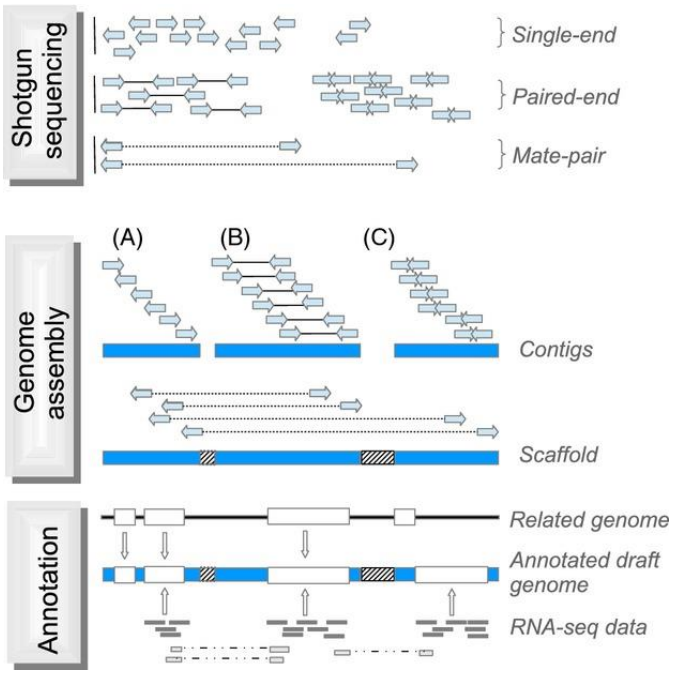



*Figure 19 : Assembly process and terminology [85]*

Figure 19 illustrates the formation of consensus sequences using bioinformatics. This consists of 3 main steps: shotgun sequencing, genome assembly and finally annotation. Shotgun sequencing consists of sequencing a small part of the target genome at random locations to ensure good coverage and high sequencing depth. After sequencing, a database is created. There are three different categories in shotgun sequencing: Single-end (500 to 1000 bp), paired-end (variable size with overlapping possibility) and mate-pair (~2 to 20 kb). These elements will make it possible to form “contigs” or “scaffolds”. These contigs or scaffolds are then assembled

by the bioinformatics program. Once the contigs are assembled, we obtain scaffolds. Thanks to the reference sequence database, it is possible to annotate the contigs and scaffolds obtained and thus locate the different important areas of a genome. It is also possible to compare areas with each other and carry out phylogenetic studies, for example [85].

It is possible to align nucleotide sequences and find the similarity between the sequences and make a comparative study between several sequences. These alignments can also be obtained with protein sequences. It is then possible to know if the proteins are identical between the different sequences. Alignment matrices are also generated to study the different areas preserved or not [98].

### 2.3.6. Phylogenetic tree

"*Phylogenics is in biology, the study of the ancestral relatedness of groups of organisms, whether alive or extinct*" [99]. A phylogenetic tree is a diagram composed of branches that show the evolution between the species studied (mammals, viruses, bacteria, etc.). It is a pairwise comparison between the differences and similarities of each species or different individuals in the same species studied [100]. The first phylogenetic trees were based on physical observations between species. The technique of phylogenetic trees has evolved considerably with genetic analyses, which makes it possible to create trees based on the genomes of the different organisms being compared [99].

Phylogenetic trees allow us to know and observe the differences or phylogenetic similarities between different viruses. It allows identifying different families or different genera within same family. Trees can identify possible new viruses if, for example, a new branch is very different from existing ones. It is important to note that it is not the only criteria for identifying new viruses. For each family, there are strict criteria for claiming to discover a new virus [101].

It seems interesting to compare the areas of preserved viruses when performing phylogenetic analysis. Indeed, it is through these areas that new virus species can be identified. The conserved areas must be sufficiently different from the references to claim the discovery of a new virus [101].

There are different parameters to interpret a phylogenetic tree to make a comparison between different tree:

- Number of bootstraps: bootstraps (0 to 100%) are associated with each branch of the phylogenetic tree. This indicates the number of times the branch was found during the different repetitions. This allows us to know the robustness of the phylogenetic tree model and the robustness of each branch. This indicates the robustness of each node to data changes [102].
- Type of phylogenetic tree:
    - The maximum likelihood method makes it possible to develop a model with the most relevant and probable data. This function promotes the highest likelihood of data. Trees obtained by this method work with the highest likelihood of data. PHYml is a program which works with this method. [103].
    - The distance-matrix method is a method that is based, as its name suggests, on the distance between genetic sequences. This is a quick method that is practised after the alignment of the sequences. The branches of the trees obtained by this method represent the importance of the genetic difference between the

sequences. Neighbour-joining is a distance-matrix method that works with an algorithm by a bottom-up clustering system [104].

- Phylogenetic trees using the Bayesian inference method are trees based on the posterior probabilities of the trees generated. They are easily interpretable trees and are very useful in the complex case study [104].
- The maximum parsimony method is not the most effective method. This method consists in reducing the length of the branches of the tree. The selected tree will always be the shortest, which is not always representative of the mutations or the real evolution of the sequences studied [104].

This work made it possible to study the presence of specific virus genera (*Nepovirus* and *Waikavirus*) in individual community plants; *Barley yellow drawf virus* was also studied. A bioinformatics analysis was also carried out on consensus sequences of nepovirus and waikavirus to determine the phylogeny of these viruses in relation to the reference sequences of these two genera. A study of *Lolium perenne* L. as a new host for *White clover mosaic virus* was also carried out.

# 3. Materials and methods

Presentation of different *Poaceae* -based plant communities studied in Belgium
Three different types of plots will be studied as part of this study:

- Field: cultivated area with only one species of *Poaceae*
- Pasture: non-cultivated area where livestock lives and eat (cows). Some notable species of *Poaceae.*
- Grassland: an area not cultivated, not occupied by animals, which is very poorly maintained by the farmer (cut once or twice a year). Has a great diversity of *Poaceae* (up to 16 different species).

## 3.1. Plant harvesting

### 3.1.1. Antheit grassland (*Lolium perenne* L.)

#### *3.1.1.1. Location*

The plant community studied is grassland with high biological value within the Natural of Burdinale-Mehaigne, in Antheit (Province of Liège, See Figure 20). This grassland is adjacent to two other *Poaceae*-based plant communities: a wheat field and an extensive pasture (See Figure 21).

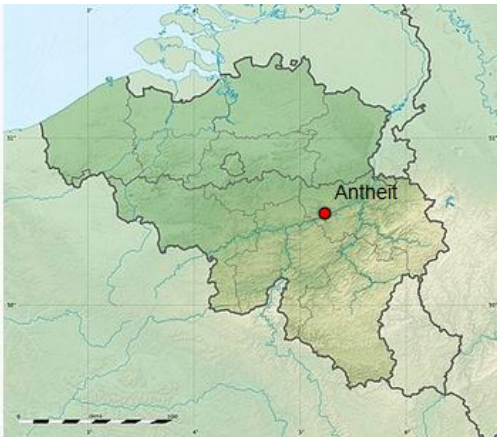



*Figure 20 : Antheit Location in Belgium [105]*

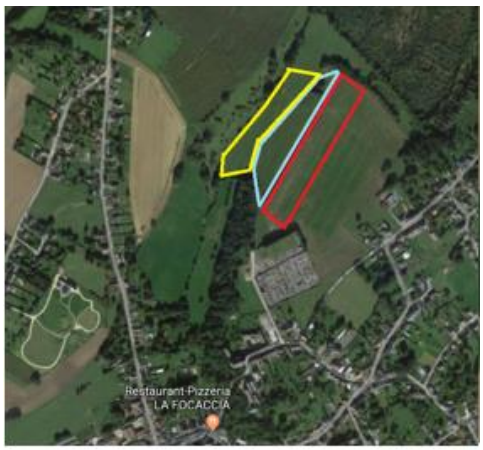



*Figure 21 : Representation of Antheit location, with the three adjacent plant communities examined: a wheat field (in red), an extensive pasture (in yellow) and a grassland with high biological value (in light blue).*

#### *3.1.1.2. Collecting*

In this grassland, 30 areas (15cmx15cm) were defined for harvesting ryegrass (*Lolium perenne* L.*)* plants (See Figure 22). Sampling was done following a zig-zag pattern; this allows a complete sampling on plot and to explore all plots to have a complete sampling. Soil and plants were collected in each area with a depth of 15 to 20 cm of soil to collect also nematodes in the soil, vector candidate of the potentially new nepovirus species. These samples all contained ryegrass plants and were placed in pots. These pots were then placed in a greenhouse and watered regularly to keep the plants alive. Ryegrass leaves and stems were collected for each area and placed in individual bags. Between each area, the hands of the person harvesting were disinfected with ethanol to avoid contamination. The plants were placed in individual bags immediately in a freezer at -80°C for quick freezing and virus storage. Samples will then be used for total RNA extraction and virus detection (see Section 3.2.1).

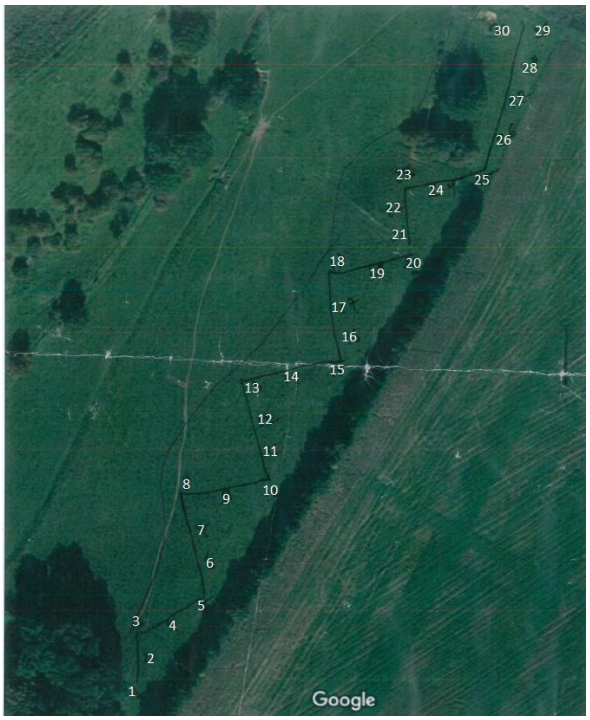



*Figure 22 : Collection plan in Antheit grassland*

The sampling scheme is different from that carried out in previous years. This is because this sampling was done to study nematodes in the soil and the zigzag sampling scheme is the best way to study nematodes [106].

### 3.1.2. Heron Pasture (*Lolium perenne* L. and *Trifolium repens* L.)

#### *3.1.2.1. Location*

The plant community studied is a pasture within the Natural of Burdinale-Mehaigne, in Héron (Province of Liège, see Figure 23). This pasture is adjacent to two other *Poaceae*-based plant communities: a wheat field and a grassland (see Figure 24).


*Figure 23 : Héron Location in Belgium [107]*


*Figure 24 : Representation of Héron location, with the three adjacent plant communities examined: a wheat field (in red), an extensive pasture (in yellow) and a grassland with high biological value (in light blue).*

#### *3.1.2.2. Collecting*

The collection plan shows the fifty areas where the plants were collected (see Figure 25). In each area, white clover (*Trifolium repens* L.) and ryegrass (*Lolium perenne* L.) plants were collected in sufficient quantity to allow laboratory analysis and placed in an individual numbered bag with the harvest area number. One person was in charge of harvesting the ryegrass and another of the white clover and between each area, the hands were disinfected with ethanol to avoid all contamination. Only leaves and stems were harvested. After harvesting all plants, these were placed in a refrigerated box to prevent plant degradation during transport. Once in the laboratory, the plants were placed directly in a freezer at -80°C for quick freezing and virus storage. Samples were then used for total RNA extraction and virus detection (see Section 3.2.1).

*Figure 25 : Collection plan in Héron pasture*

## 3.2. Laboratory work

### 3.2.1. Total RNA extraction

A total RNA extraction protocol was used on collected ryegrass and white clover plants. This protocol was adapted from Oñate-Sánchez L. et al. 2008 [108].

#### *3.2.1.1. Samples*

For the analyses, the samples used are either those taken from the Pasture Héron or Antheit grassland or samples previously taken and stored at -80°C for Antheit field, Antheit pasture (global plant community, *Lolium perenne* L. and *Poa trivialis* L.) and Antheit grassland (global plant community, *Lolium perenne* L., *Poa trivialis* L. and minor species). It is also important to note that areas 36-37-45-45-46-49 and 50 were not analysed in the global pasture plant community for all viruses.

There were no nepovirus and waikavirus analyses in the Antheit field because the previous sequencing showed that they were not present in this plant community.

#### *3.2.1.2. Homogenization*

Two hundred mg of leaves (preferably green) from each plant were collected and placed in an extraction bag. In each sachet, 2 ml of cell lysis solution - Cell lysis solution (2% SDS, 68 mM sodium citrate, 132 mM citric acid, 1mM EDTA) was added. The plant material was ground with a tissue homogenizer and then placed in the refrigerator (4°C) for 10 minutes. One ml of the juice obtained was taken and transferred to a 2 ml tube. Each tube is vortexed for 10 to 30 seconds.

#### *3.2.1.3. Phase separation*

The tubes were left at room temperature for 5 minutes for incubation. Then, 300µl of DNA/protein precipitation solution (4 M NaCl, 16 mM sodium citrate, 32 mM citric acid) was added. The tubes were vortexed for 10 to 30 seconds and then gently inverted several times.

The tubes were then left to incubate for 10 minutes at 4°C. The tubes were centrifuged at 13,000 rpm for 10 minutes at 4°C. Then, 800µl of supernatant was taken and placed in a 2 ml tube.

#### *3.2.1.4. RNA precipitation*

Nine hundred µl isopropanol (VWR) was added to each tube. The tubes were then delicately inverted several times. The tubes were placed in the centrifuge at 13,000 rpm for 4 minutes at 4°C. After this step, supernatant was discarded, and 1 ml of ethanol 70% (Merck) was placed in each tube. The tubes were placed at -20°C for storage.

At the end of this step, the tubes can be stored at -20°C for one year. This allows the extraction manipulation to be stopped and the other manipulations to be continued later.

#### *3.2.1.5. End of extraction*

The samples were taken out of freezer and centrifuged at 9,000 rpm for 5 minutes at 4°C. The supernatant was discarded, and tubes were dried in a hood for at least 10 minutes (until all ethanol was removed). Fifty µl of sterile water was added to each tube. The samples were placed on ice for at least 5 minutes to break up the pellet. The samples were ready for RT-PCR.

The first ten samples from each pool were tested using Nanodrop (Isogen) to determine the average RNA concentration of the samples and sample quality. There are two absorbance ratios to assess the quality of extraction and analysis. The ratio 260/280 evaluates the purity of DNA or RNA. For RNA, the ration of 2 is accepted for a pure sample. That measures intern pollution of a sample due to polysaccharides or polyphenols. The ratio 260/230 represents used to measure the purity of nucleic acid. That measure extern pollution in a sample (trizol, chloroform, EDTA) It is generally higher than the 260/280 ration and reaches a value of 2.2 [109]. Samples were then diluted in sterile water to reach an average RNA concentration of 100 µg RNA/ µl in order to obtain conclusive results for PCR. All dilutions were done on ice. Diluted samples were stored in ice for RT-PCR.

### 3.2.2. Reverse-transcription (RT)

Two different protocols were compared, using different reverse transcriptase enzyme: Superscript III (Invitrogen Thermo Fischer) and Tetro RT enzyme (Bioline).

#### *3.2.2.1. Superscript III*

There are two stages in this protocol.

*Table 3 : Mastermix 1 for Superscript III RT-PCR protocol*

| **Mastermix 1** | |
|---|---|
| **Reagent** | **µl/reaction** |
| Sterile water | 11 |
| Random hexamer (Thermo Fischer) | 1 |

All manipulations were done on ice. Ten µl of mastermix 1 (see Table 3) were placed in each well of a PCR plate. Two and a half µl of each sample were placed in the wells. The wells were hermetically sealed. The plates were placed in the thermal cycler (Doppio, VWR) for 10 minutes at 80°C. After this step, the plate was immediately put back into the ice.

*Table 4 : Mastermix 2 for Superscript III RT-PCR protocol*

| **Mastermix 2** | |
|---|---|
| **Reagent** | **µl/reaction** |
| 5 x first strand buffer (Thermofischer) | 4 |
| 0.1 M DTT (Thermofischer) | 2 |
| 10 mM dNTP mix (Eurogentec) | 1 |
| RNaseOUT (40U/µl) (Thermofischer) | 0.25 |
| Superscript III (200U/µl) (Thermofischer) | 0.5 |

After complete cooling of the PCR plate, 7.5µl of mastermix 2 (see Table 4) was added in each well. The plate has been replaced in the thermal cycler (Doppio, VWR) for the second cycle: 10 minutes at 25°C, 30 minutes at 48°C and finally 5 minutes at 95°C. The plates were stored directly on-site if the PCR was done immediately or in the refrigerator at 4°C.

#### *3.2.2.2. Tetro Reverse Transcriptase*

*Table 5 : Mastermix Tetro for Tetro Reverse Transcriptase RT-PCR protocol*

| **Mastermix Tetro** | |
|---|---|
| **Reagent** | **µl/reaction** |
| Random hexamer 40 µM (Thermo Fischer) | 1 |
| 10 mM dNTP mix (Eurogentec) | 1 |
| 5 x RT buffer | 4 |
| RNaseOUT (40U/µl) (Thermofischer) | 0.5 |
| Tetro Reverse Transcriptase (200U/µl) (Bioline) | 1 |
| Sterile water | 10 |

All manipulations were done on ice. Seventeen and a half µl of mastermix Tetro (see Table 5) were placed in each well of a PCR plate. Two and a half µl of each sample was placed in the wells. The wells were hermetically sealed. The plates were placed in the thermal cycler (Doppio, VWR) for this cycle: 10 minutes at 25°C, 30 minutes at 45°C and finally 5 minutes at 85°C. The plates were stored directly on-site if the PCR was done immediately or in the refrigerator at 4°C.

### 3.2.3. Polymerase chain reaction (PCR)

New PCR plates were used for this reaction. The mastermix PCR (see Table 6) was prepared and placed on the plates in a different location in the laboratory to avoid contamination. The hood was placed under UV light for at least 10 minutes before the start of manipulation. The primers used was designed on Geneious. (see section 3.3.7).

*Table 6 :Mastermix PCR for PCR protocol*

| Mastermix PCR | |
|---|---|
| **Reagent** | **µl/reaction** |
| Sterile water | 12 |
| 5 x PCR buffer (Bioline) | 5 |
| 50mM $MgCl_2$ (Bioline) | 2.5 |
| 10 mM dNTP mix (Eurogentec) | 1 |
| 20µM Forward primer (Sigma) | 1 |
| 20µM Reverse primer (Sigma) | 1 |
| Mango Taq (5U/µl) (Bioline) | 0.5 |

Here are the primers (see Table 7) for the four viruses analysed:

*Table 7 : Sequence, 5' Position, Tm and amplicon size for primers*

| Primer | Séquence (5'-3') | 5'Position | Tm (°C) | Amplicon size (bp) |
|---|---|---|---|---|
| BYDV F<br>BYDV R | F : CCCAGTCTATCGCAATGCCCAGC<br><br>R : GGTTCCGGTGTTGAGGAGTCTAC | 3104<br><br>3483 | 55°C | 379 |
| Waika F<br>Waika R | F : ACCCTCAAGTTCTTTCCACTT<br><br>R : ACTCCCTCTCCAGTATTGAA | 3251<br><br>3607 | 56°C | 356 |
| WCMV F<br>WCMV R | F : AAGTCTGAACTTACTGGTGACTCTG<br><br>R : GTCGGAAGGACCACGAATGAGG | 5467<br><br>5742 | 71°C | 275 |
| Nepo F<br>Nepo R<br>RNA2 | F : TGTGTCGGGAAATAAACTACAAGCA<br><br>R : GCAAAAGAGCCAAACTGGAATGGTA | 3775<br><br>4150 | 63°C | 375 |

Twenty-three µl of the PCR mastermix were placed in each well of the PCR plate, and then 2µl of the RT product from each sample were added. For blanco, 2µl of sterile water were added, and for the positive control, 2µl of the confirmed positive RTs were used.

The plates were placed in the thermal cycler (Doppio, VWR) to follow the following cycle (See Table 8):

*Table 8 : Cycle for PCR in a thermal cycler*

| Temperature | Time | Number of cycles |
|---|---|---|
| 94°C | 4 minutes | / |
| 94°C | 45 seconds | 35 cycles |
| Tm of the virus | 1 minute | |
| 72°C | 45 seconds | |
| 72°C | 10 minutes | / |
| 4°C | Infinity | / |

The plates were stored at 4°C in fridge after PCR.

#### *3.2.3.1. Visualization on gel*

A 1% agarose gel in TAE was prepared by adding 10 µl of GelRed (Biotium). The molecular weight marker used is GeneRuler 100bp (ThermoSientific)plus with 5µl per well. The gel photos are stored on the gel reader (Vilber).

## 3.3. Bioinformatic analysis

For analysis, Durandal and Geneious software were used.

### 3.3.1. Identification of contigs

For each plant community, the contigs were analysed by Durandal: Durandal is a cluster from ULiège which is usually used for bioinformatics analyses on numerous libraries in parallel, allowing faster implementation compared to software such as Geneious. In this case, Durandal is used to assemble the HTS data [110], and Geneious to annotate each contig. Once the annotation was done, the contigs belonging to the virus of interest genus (*Nepovirus* or *Waikavirus*) were placed in distinct folders. A distinction between RNA1 and RNA2 was also made for nepoviruses. The same process was used for the analysis of white clover mosaic virus.

### 3.3.2. Analysis contigs and sequence consensus

For each plant community, contigs were de novo assembled according to the following method: assembler was Geneious and sensitivity was medium sensitivity/fast. If de novo assembly was successful, sequences are preserved. If the assembly was not successful, analyses continued with all contigs.

Contigs are mapped to references of virus type analysed with the following parameters: mapper was Geneious and sensitivity was medium sensitivity/fast without iterations, maximum mismatch per read: 30%, best match: randomly. If some contigs did not match references, a BLASTn analysis was made on NCBI website[2] to know if contigs belonged to the genus studied. If the percentage of identity was too low, contigs are excluded from the analysis. Thanks to this mapping, a first consensus sequence was obtained for each plant community.

Raw reads are mapped to the identified virus contigs to determine if there was good coverage of each contig. Mapping parameters were as follows: mapping was Geneious and sensitivity was medium sensitivity/fast without iterations, maximum mismatch per read: 30%, best match: randomly. In case of doubt or poor coverage, contigs are excluded from the analysis.

Consensus sequences obtained and maintained for rest of the analyses are derived from the analysis of AP2 (Antheit pasture year 2) for *Nepovirus* RNA2 (4056 bp) and LG2 Poa (Poa in Latinne grassland year 2) for *Waikavirus* (11699 bp).

For *white clover mosaic virus*, the sequence consensus was in HP2 (Héron pasture year 2). It was the only plant community with this virus.

### 3.3.3. Sequence consensus with reads for each plant community

The reads of each plant community were mapped to the *Nepovirus* or *Waikavirus* consensus sequences. The mapping parameters were as follows: mapping used Geneious and sensitivity was medium sensitivity/fast without iterations, maximum mismatch per read: 30%, best match: randomly. A consensus sequence was obtained for each plant community containing reads of

[2] https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

the virus genera studied. The consensus sequences are placed in the same file for further analysis.

For each plant community, a second mapping with reads was performed with iteration 3 times to see if it was possible to extend the consensus sequence obtained. The mapping was done with the following parameters: sensitivity was medium sensitivity/fast without iterations, maximum mismatch per read: 30%, best match: randomly, iterate 3 times. After this mapping, the added sequences are checked by BLASTn on the NCBI website[3]. If the added sequence corresponds to the genus of the virus being studied, the consensus sequence is retained. Otherwise, the sequence is not used.

### 3.3.4. Phylogenetic analysis and tree construction

The analysis process is the same for *Waikavirus* or *Nepovirus*.

#### *3.3.4.1. Waikaviruses*

The nucleotide consensus sequences of each plant community are aligned using a MUSCLE alignment (default parameters): re-align sequence, eight iterations, sequence order: group sequences by similarities, distance measure: kmer 4_6, clustering method: UPGMB, tree rooting method: pseudo, sequence weighting scheme: CLUSTALW, Optimisation: Anchor.

A tree was then built based on the alignment. The sequences are cut after alignment, based on the shortest sequence for the construction of the tree. It is important that all sequences are the same length for the analysis. Tree parameters: Geneious Tree builder (default parameters): genetic distance model: Tamura-Nei, Tree build method: Neighbor-Joining, No outgroup, Resampling method: 100 bootstraps. The tree matrix was set as a percentage of identity.

This same process is then applied with the consensus nucleotide sequences of each plant community and references of waikaviruses and sequiviruses from NCBI database (see Table 9). The parameters of the alignment and construction of the tree were identical to the previous analysis.

[3] https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

*Table 9 : NCBI references of waikaviruses and sequiviruses with name of each virus*

| NCBI reference | Virus name |
|---|---|
| NC_040586 | Brassica napus RNA virus 1 isolate SP2S, complete genome |
| MH844554 | Brassica napus RNA virus 1 isolate SP2S, complete genome |
| NC_038320 | Carrot necrotic dieback virus strain Anthriscus, complete genome |
| NC_003628 | Parsnip yellow fleck virus, complete genome |
| NC_003626 | Maize chlorotic dwarf virus, complete genome |
| NC_001632 | Rice tungro spherical virus, complete genome |
| KT238881 | Bellflower vein chlorosis virus isolate CT1, complete genome |
| KC794785 | Rice tungro spherical virus isolate AP, complete genome |
| EU980442 | Carrot necrotic dieback virus strain Anthriscus, complete genome |
| D14066 (PYFPOLYP) | Parsnip yellow fleck virus gene for polyprotein, complete cds |
| AM234049 | Rice tungro spherical virus complete genome, West Bengal isolate |
| AM234048 | Rice tungro spherical virus complete genome, Orissa isolate |
| AY829112 | Maize chlorotic dwarf virus isolate M1, complete genome |
| AY362551 | Maize chlorotic dwarf virus strain Severe, complete genome |
| AB064963 | Rice tungro spherical virus genomic RNA, complete genome, strain:Vt6 |
| U67839 (MCU67839) | Maize chlorotic dwarf waikavirus strain Tennessee polyprotein mRNA, complete cds |
| M95497 (RTUPOLYP) | Rice tungro spherical virus polyprotein gene, complete cds |

#### 3.3.4.1.1. CP (coat protein) analysis

After alignment, it was possible to identify the coat proteins on the genome and to extract them. Coat proteins of the plant community sequences are compared with each other using a tree. The construction of this tree was done with the same parameters as above.

A second tree is built with coat protein references and consensus sequences. Always according to the same parameters.

#### 3.3.4.1.2. Pol-pro (polymerase- protease) analysis

After alignment, it was possible to identify the pol-pro zone on the genome and to extract them. The analyses were the same as for coat protein with the same analysis parameters.

#### 3.3.4.1.3. Protein analysis

Once the coat protein and the polymerase protease had been translated to obtain the respective protein sequences.

Once the protein sequences were obtained, an alignment for the CP and one for the Pol-pro was performed according to the following parameters: MUSCLE alignment (default parameters): re-align sequence, eight iterations, sequence order: group sequences by similarities, distance measure: kmer 4_6, clustering method: UPGMB, tree rooting method: pseudo, sequence weighting scheme: CLUSTALW, Optimisation: Anchor. This analysis is carried out with the consensus sequences of each plant community and references.

Once the alignment was complete, a tree was built for the CP and another for the Pol-pro with these parameters: Geneious Tree builder (default parameters): genetic distance model: Tamura-Nei, Tree build method: Neighbor-Joining, No outgroup, Resampling method: 100 bootstraps. The tree matrix is set as a percentage of identity.

#### *3.3.4.2. Nepoviruses*

The same analyses as that of waikaviruses were performed for RNA 1 and RNA 2. Only references change:

- RNA 1 references (downloaded from NCBI) (see Table 10):

*Table 10 : NCBI references of nepoviruses RNA1 with name of each virus*

| NCBI reference | Virus name |
|---|---|
| NC_040399 | Red clover nepovirus A isolate B46 segment RNA1, complete sequence |
| NC_033492 | Petunia chlorotic mottle virus segment RNA 1, complete sequence |
| NC_038767 | Mulberry mosaic roll leaf-associated virus isolate zj segment RNA1, complete sequence |
| NC_038765 | Melon mild mottle virus gene for polyprotein, complete cds, segment RNA 1 |
| NC_038762 | Aeonium ringspot virus segment RNA1, complete sequence |
| NC_034214 | Peach rosette mosaic virus isolate PRMV2 segment RNA1, complete sequence |
| NC_032270 | Soybean latent spherical virus isolate ND1 segment RNA1, complete sequence |
| NC_022798 | Potato black ringspot virus isolate PRI-Ec segment RNA 1, complete sequence |
| NC_018383 | Grapevine Anatolian ringspot virus RNA 1, complete genome |
| NC_017939 | Grapevine deformation virus RNA1, complete genome |
| NC_015492 | Grapevine Bulgarian latent virus segment 1, complete genome |
| NC_015414 | Cherry leaf roll virus RNA1, complete genome |
| NC_006057 | Arabis mosaic virus RNA 1, complete sequence |
| NC_005266 | Raspberry ringspot virus RNA1, complete genome |
| NC_005097 | Tobacco ringspot virus RNA 1, complete sequence |
| NC_003791 | Cycas necrotic stunt virus RNA 1, complete sequence |
| NC_003509 | Blackcurrant reversion virus RNA1, complete sequence |
| NC_003622 | Grapevine chrome mosaic virus RNA 1, complete sequence |
| NC_004439 | Tomato black ring virus RNA 1, complete sequence |
| NC_003693 | Beet ringspot virus RNA 1, complete sequence |

- RNA 2 references (downloaded from NCBI) (see Table 11):

*Table 11 : NCBI references of nepoviruses RNA2 with name of each virus*

| NCBI reference | Virus name |
|---|---|
| NC_040400 | Red clover nepovirus A isolate B46 segment RNA2, complete sequence |
| NC_033493 | Petunia chlorotic mottle virus segment RNA 2, complete sequence |
| NC_038768 | Mulberry mosaic roll leaf-associated virus isolate zj segment RNA2, complete sequence |
| NC_038766 | Melon mild mottle virus gene for polyprotein, complete cds, segment RNA 2 |
| NC_038761 | Aeonium ringspot virus segment RNA2, complete sequence |
| NC_034215 | Peach rosette mosaic virus isolate PRMV2 segment RNA2, complete sequence |
| NC_032271 | Soybean latent spherical virus isolate ND1 segment RNA2, complete sequence |
| NC_022799 | Potato black ringspot virus isolate PRI-Ec segment RNA 2, complete sequence |
| NC_018384 | Grapevine Anatolian ringspot virus RNA 2, complete genome |
| NC_017938 | Grapevine deformation virus RNA2, complete genome |
| NC_015493 | Grapevine Bulgarian latent virus segment 2, complete genome |
| NC_015415 | Cherry leaf roll virus RNA2, complete genome |
| NC_006056 | Arabis mosaic virus RNA 2, complete sequence |
| NC_005267 | Raspberry ringspot virus RNA 2, complete sequence |
| NC_005096 | Tobacco ringspot virus RNA 2, complete sequence |
| NC_003792 | Cycas necrotic stunt virus RNA 2, complete sequence |
| NC_003502 | Blackcurrant reversion virus RNA 2, complete sequence |
| NC_003621 | Grapevine chrome mosaic virus RNA 2, complete sequence |
| NC_004440 | Tomato black ring virus RNA 2, complete sequence |
| NC_003694 | Beet ringspot virus RNA 2, complete sequence |

#### *3.3.4.3. White clover mosaic virus*

For the phylogenetic analysis of white clover mosaic virus, reference sequences were as follows (see Table 12):

*Table 12 : : NCBI references of white clover mosaic virus with name of each virus*

| NCBI reference | Virus name |
|---|---|
| NC_003820 | White clover mosaic virus, complete genome |
| LC159490 | White clover mosaic virus genomic RNA, complete genome, strain: RC, isolate: Cheongdo-2 |
| LC159489 | White clover mosaic virus genomic RNA, complete genome, strain: RC, isolate: Cheongdo-1 |
| LC159488 | White clover mosaic virus genomic RNA, complete genome, strain: RC, isolate: Suwon |
| AB056720 | White clover mosaic virus genomic RNA, complete genome, strain: RC |
| AB669182 | White clover mosaic virus genomic RNA, complete genome, strain: WClMV-RC |

### 3.3.5. Other analysis for polymerase-protease zone waikavirus

A protein sequence that appeared to be found in most Waikavirus consensus protein sequences has been identified: "QA" -> +/- 150 bp -> "CG" -> +/- 500 bp -> "GDD" -> 200 bp -> STOP
The sequences were extracted for all plant communities and references except for LG2 Poa (Latinne grassland year 2) and AP2 (Antheit pasture year2) because this sequence scheme was not found in the protein sequence. A phylogenetic analysis was launched with the same parameter as before.

### *3.3.6.* NCBI tree *Secoviridae*

To allow a comparison with the phylogenetic tree present on the ICTV site for the *Secoviridea* family (Figure 11) [52], a phylogenetic analysis was done on pol-pro region of the consensus protein sequence of waikaviruses (LG2 Poa), obtained from the assembly of the contigs. It is the part between "CG" motif of the 3C-proteinase and "GDD" motif of the polymerase that is analyzed. Once this part was identified for the waikavirus consensus sequence, a phylogenetic analysis was launched with the same parameters as the ICTV: The tree was generated with PhyML (100 bootstrap replicates) in the TOPALi using a RtRev +I+G evolutionary model selected by Protest [52].

The list of references (from NCBI) was identical to that of the ICTV (see Table 13):

*Table 13 : NCBI references of Secoviridea with name of each virus*

| NCBI reference | Virus name | NCBI reference | Virus name |
|---|---|---|---|
| JX304792 | Aeonium ringspot virus | X15346 | Grapevine chrome mosaic virus |
| AB030940 | Apple latent spherical virus | D00915 | Grapevine fanleaf virus |
| AY30378 | Arabis mosaic virus | KC855266 | Lettuce necrotic leaf curl virus |
| JQ437415 | Arracacha virus B | U67839 | Maize chlorotic dwarf virus |
| AM087671 | Artichoke yellow ringspot virus | AB518485 | Melon mild mottle virus |
| U70866 | Bean pod mottle virus | KM229700 | Motherwort yellow mottle virus |
| D00322 | Beet ringspot virus | KC904083 | Mulberry mosaic leafroll associated virus |
| KT238881 | Bellflower vein chlorosis virus | D14066 | Parsnip yellow fleck virus |
| DQ344639 | Black raspberry necrosis virus | AF016626 | Peach rosette mosaic virus |
| AF368272 | Blackcurrant reversion virus | AB295643 | Radish mosaic virus |
| AB649296 | Blueberry latent spherical virus | AY303787 | Raspberry ringspot virus |
| AB084450 | Broad bean wilt virus 1 | X64886 | Red clover mottle virus |
| AF225953 | Broad bean wilt virus 2 | M95497 | Rice tungro spherical virus |
| EU980442 | Carrot necrotic dieback virus | AB009958 | Satsuma dwarf virus |
| KF533719 | Carrot torradovirus 1 | KU052530 | Squash chlorotic leaf spot virus |
| FR851461 | Cherry leaf roll virus | AB054688 | Squash mosaic virus |
| AJ621357 | Cherry rasp leaf virus | DQ143874 | Stocky prune virus |
| JN052073 | Chocolate lily virus A | AY860978 | Strawberry latent ringspot virus |
| X00206 | Cowpea mosaic virus | AJ311875 | Strawberry mottle virus |
| M83830 | Cowpea severe mosaic virus | U50869 | Tobacco ringspot virus |
| FJ194941 | Cucurbit mild mosaic virus | AY157993 | Tomato black ring virus |
| KT692952 | Currant latent virus | EF681764 | Tomato marchitez virus |
| AB073147 | Cycas necrosis stunt virus | L19655 | Tomato ringspot virus |
| AB084452 | Gentian mosaic virus | DQ388879 | Tomato torrado virus |
| FN691934 | Grapevine Bulgarian latent virus | | |

### 3.3.7. Primer design for PCR

To do PCR, it was necessary to design primers to determine the presence or absence of the targeted viruses in the samples analysed.

Primers of *Nepovirus* RNA2, *Waikavirus*, and *White clover mosaic virus* were chosen from the conserved sequences between the consensus sequences of the different locations. The primer was between 25 and 35 bp, and the sequence length is 300 to 400 bp (see Table 7).

The compatibility of the primer has been established through the Oligo Analyzer program. The difference in Tm of the two primers (Reverse and Forward) was a maximum of 2°C. The dG should not be less than -3 kcal/mol in absolute terms.

# 4. Results

## 4.1. PCR tests and geographical distribution

### 4.1.1. Global overview

*Table 14: Global overview of candidate nepovirus, candidate waikavirus and Barley yellow dwarf virus (BYDV) in Antheit ecosystems*

| | | | % prevalence in 50 individual plants | | |
|---|---|---|---|---|---|
| **Location** | **Ecosystem** | **Pool category** | **BYDV** | ***Nepovirus*** | ***Waikavirus*** |
| Antheit | Field | Global ecosystem | 6,00% | | |
| Antheit | Pasture | Global ecosystem | 29,55% | 59,09% | 4,55% |
| Antheit | Pasture | *Lolium perenne* L. | 6,00% | 86,00% | 20,00% |
| Antheit | Pasture | *Poa trivialis* L. | 56,00% | 76,00% | 36,00% |
| Antheit | Grassland | Global ecosystem | 72,00% | 66,00% | 38,00% |
| Antheit | Grassland | *Lolium perenne* L. | 22,00% | 90,00% | 10,00% |
| Antheit | Grassland | *Poa trivialis* L. | 24,00% | 80,00% | 50,00% |
| Antheit | Grassland | Minor species | 14,29% | 28,57% | 28,57% |

A total of 1082 samples were analyzed to obtain the prevalence of the different plant communities. As shown in Table 14 and Figure 26, it is possible to note that *Nepovirus* is the genus with the highest prevalence in all the plant communities studied in Antheit and that the ryegrass (*Lolium perenne* L.) is mainly infected. Antheit Grassland has a higher presence of viruses in plants except for BYDV in Antheit Grassland Poa and waikaviruses in Antheit Grassland Lolium.

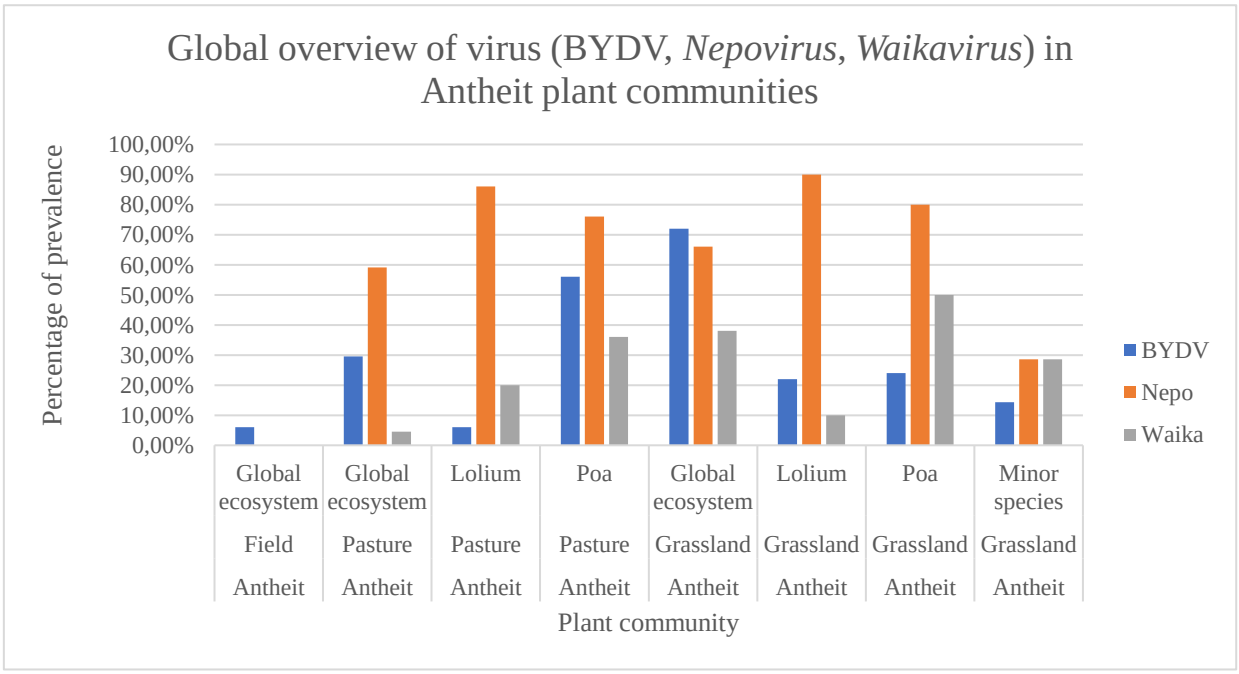



*Figure 26: Global overview of candidate nepovirus, candidate waikavirus and Barley yellow dwarf virus (BYDV) in Antheit ecosystems*

*Table 15: Prevalence of infection virus for all plant species in Antheit pasture and grassland (global plant communities)*

| **AP2 Global ecosystem** | | | |
|---|---|---|---|
| | **BYDV** | **Nepo** | **Waika** |
| ***Phleum pratense* subsp. *bertolonii*. (2)** | 50,00% | 50,00% | 0,00% |
| ***Cynosorus cristatus* L. (7)** | 42,86% | 100,00% | 0,00% |
| ***Holcus Lanatus* L. (6)** | 33,33% | 83,33% | 16,67% |
| ***Poa trivialis* L. (10)** | 30,00% | 20,00% | 10,00% |
| ***Lolium perenne* L. (11)** | 36,36% | 54,55% | 0,00% |
| ***Agropyron repens* (L.) P.Beauv. (1)** | 0,00% | 100,00% | 0,00% |
| ***Agrostis* sp. (6)** | 0,00% | 66,67% | 0,00% |
| **AG2 Global ecosystem** | | | |
| | **BYDV** | **Nepo** | **Waika** |
| ***Dactylis glomerata* L. (3)** | 0,00% | 0,00% | 0,00% |
| ***Cynosorus cristatus* L. (9)** | 100,00% | 44,44% | 11,11% |
| ***Holcus Lanatus* L. (12)** | 91,67% | 91,67% | 0,00% |
| ***Poa trivialis* L. (10)** | 100,00% | 70,00% | 20% |
| ***Lolium perenne* L. (12)** | 58,33% | 83,33% | 58,33% |
| ***Festuca rubra* L. (1)** | 100,00% | 0,00% | 0,00% |
| ***Agrostis spica-venti* (L.) P. Beauv. (3)** | 66,67% | 66,67% | 33,33% |

Table 15 shows which plant species were harvested during the ecosystem study. This table shows that the harvested *Dactylis* is not infected with any of the three viruses studied. Only three individuals of this species were studied. In Antheit pasture, only the *Poa trivialis* L. and *Holcus* which are infected by the waikavirus candidate. In the Antheit grassland, *Festuca* is infected only with BYDV. In this grassland, *Holcus* and *Festuca* are not infected with candidate waikavirus. Only one individual of the Festuca species has been studied. The Nepovirus candidate is present in all the individual of *Cynosorus* and *Agropyron*, only one individual in the Antheit pasture.

Figure 27 and Figure 28 show the result of a PCR gel for Antheit pasture year 2. It is thanks to its results that the prevalence has been calculated (presence or absence of white bands at the amplicon size level. Some bands are more apparent than others such as the 13 and 15 sample band for example.

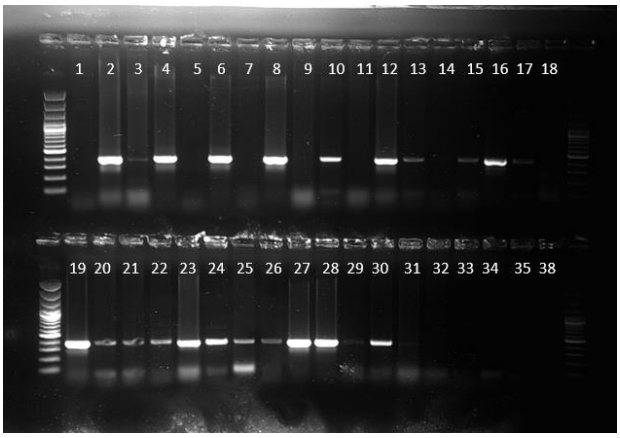



*Figure 27 : PCR gel picture for AP2 (Antheit pasture year 2). Sample from 1 to 35 and 38.*

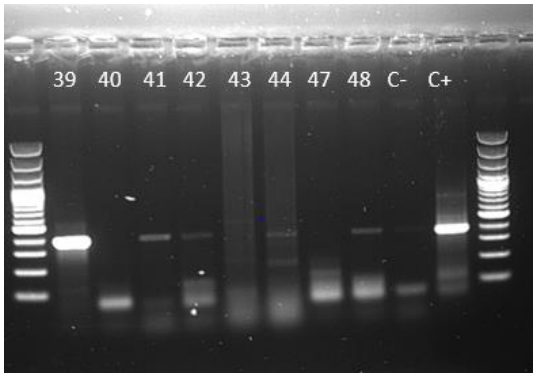



*Figure 28 : PCR gel picture for AP2 (Antheit pasture year 2). Sample from 39 to 44 and 47-48. Negative and positive control.*

### 4.1.2. Candidate nepovirus

#### *4.1.2.1. Prevalence*

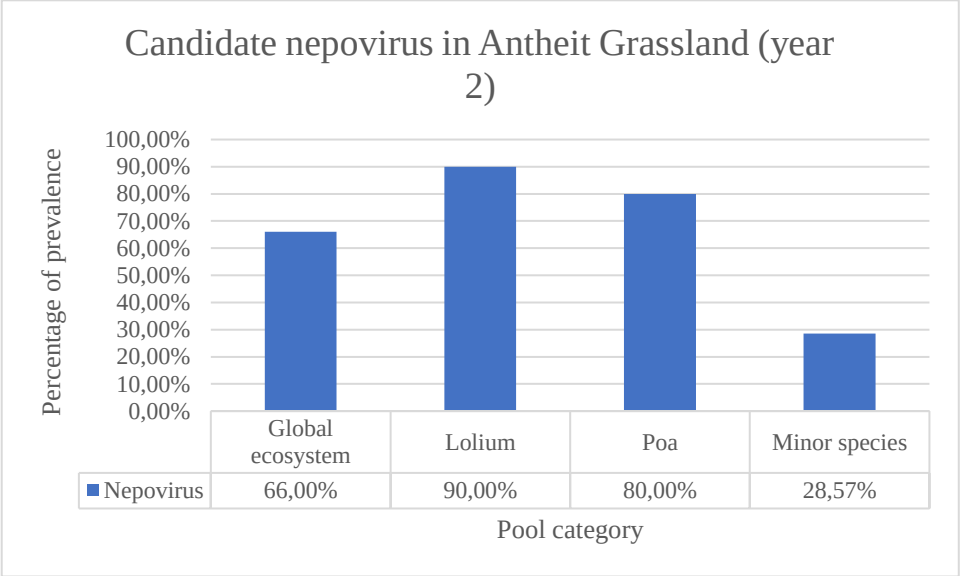



*Figure 29: Prevalence of candidate nepovirus in Antheit grassland (year2)*

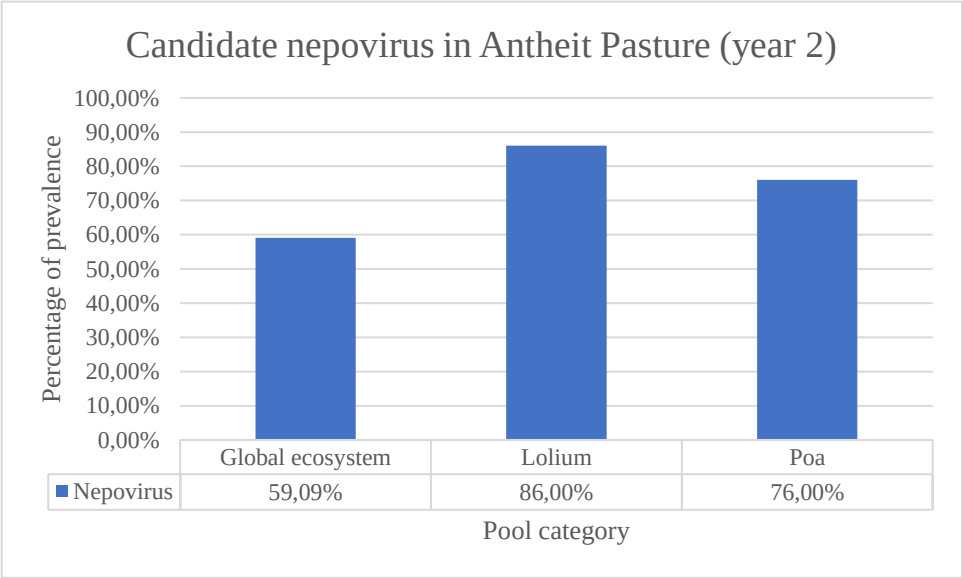



*Figure 30: Prevalence of candidate nepovirus in Antheit pasture (year2)*

As shown in Figure 29 and Figure 30, it is the ryegrass (*Lolium perenne* L.), which has the highest prevalence. Then it is *Poa trivialis* L., which is the most infected, followed by the global ecosystem. The prevalence is always lower in the pasture compared to the grassland. Minor species are the least infected species.

Analysis of ryegrasses collected in Antheit grassland this year (only thirty samples and not fifty) indicates a prevalence of 50% for the nepovirus candidate. This is 40% less than the previous year.

#### *4.1.2.2. Geographical distribution*

*Figure 31: Geographical distribution of nepoviruses in Antheit pasture (global ecosystem)*

*Figure 32: Geographical distribution of nepoviruses in Antheit pasture (Lolium perenne L.)*

*Figure 33: Geographical distribution of nepoviruses in Antheit pasture (Poa trivialis* L.*)*

Areas 1-32-35-38 are not contaminated by any nepoviruses (global ecosystem, *Lolium perenne* L. and *Poa trivialis* L.) in Antheit pasture. Areas 11-14-32-33-33-34-35-35-38-49 have not infected with the nepovirus candidate the case of the global ecosystem and the *Poa trivialis* L. Infected areas are present throughout the plot (see Figure 31, Figure 32, Figure 33) .

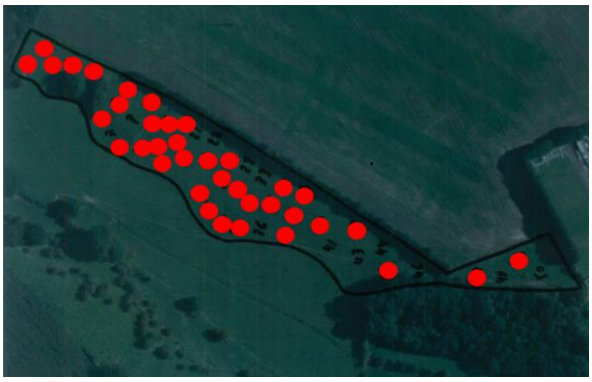


*Figure 34: Geographical distribution of nepoviruses in Antheit grassland (Poa trivialis* L.*)*

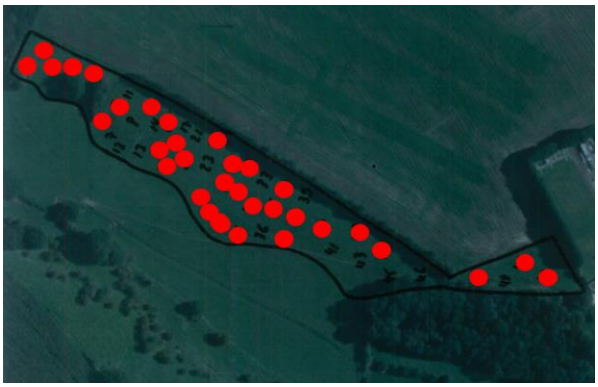


*Figure 35: Geographical distribution of nepoviruses in Antheit grassland (global ecosystem)*

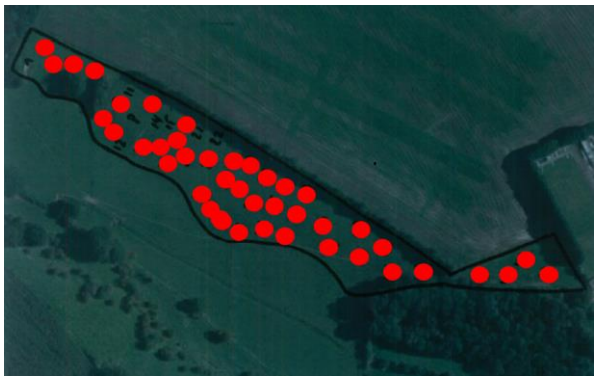


*Figure 36: Geographical distribution of nepoviruses in Antheit grassland (Lolium perenne L.)*

Figure 34, Figure 35 and Figure 36 show that zones 8 and 21 are not contaminated by the nepovirus candidate in Antheit grassland (global ecosystem, *Lolium perenne* L., *Poa trivialis* L.). Areas 8-11-12-14 in the global ecosystem and *Lolium perenne* L. are not infected. One area appears to be less infected, along the border with the field, in area 21-22. The borders of the 1-2-3 and 47-48-49-50 zones are highly infected.


*Figure 37: Geographical distribution of nepoviruses in Antheit grassland (Lolium perenne L.) (year 3)*

Two areas of contamination are highlighted by the ryegrass sampling in Antheit grassland for the third year: the area between point 23 and 29 and the area between 9 and 19. Fifteen areas are infected by the nepovirus candidate (see Figure 37).

### 4.1.3. Candidate waikavirus

#### *4.1.3.1. Prevalence*

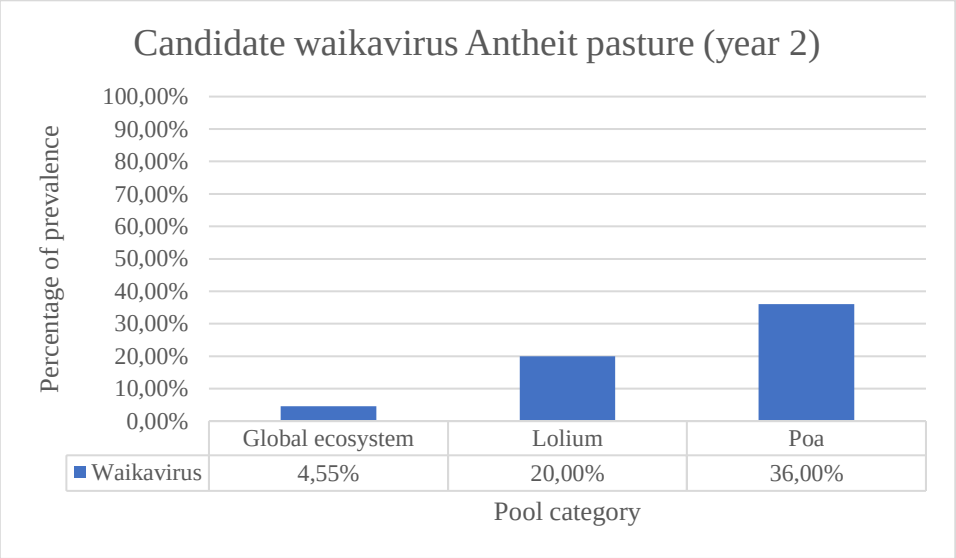



*Figure 38: Prevalence of candidate waikavirus in Antheit pasture (year2)*

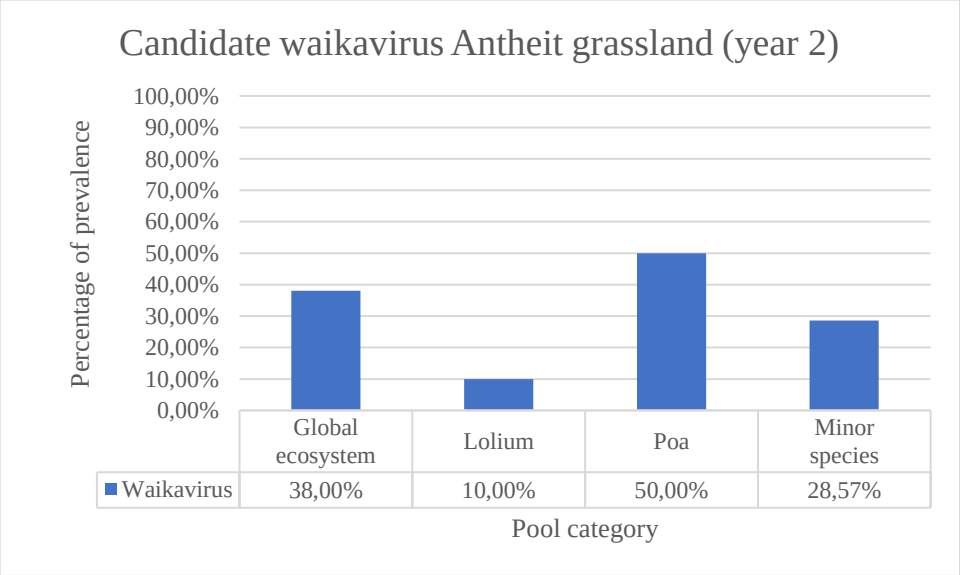



*Figure 39: Prevalence of candidate waikavirus in Antheit grassland (year2)*

In the case of the waikavirus candidate, rough bluegrass (*Poa trivialis* L.) is the most infected with a maximum prevalence of 50% in Antheit grassland. For this ecosystem, the ryegrass is the least infected. Minor species are 28.57% infected. In Antheit pasture, it is the analysis of the global ecosystem that gives the lowest prevalence, with the lowest prevalence for both ecosystems at 4.55% (see Figure 38, Figure 39)

#### *4.1.3.2. Geographical distribution*

*Figure 40: Geographical distribution of waikaviruses in Antheit pasture (global ecosystem)*

*Figure 41: Geographical distribution of waikaviruses in Antheit pasture (Lolium perenne L.)*

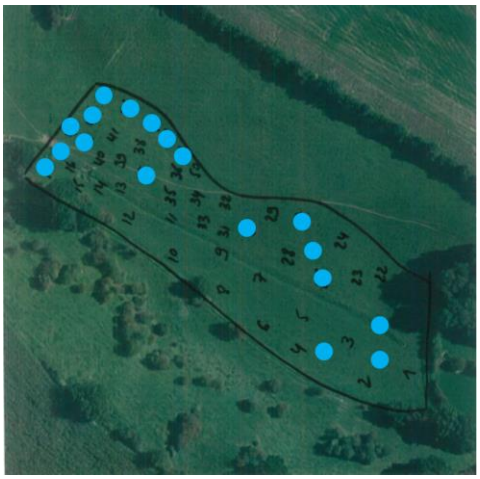


*Figure 42: Geographical distribution of waikaviruses in Antheit pasture (Poa trivialis* L.*)*

Figure 40, Figure 41, Figure 42 allow you to make different observations. The waikavirus candidate infects Antheit pasture less (see Figure 31, Figure 32, Figure 33 ). The distribution of infection is very random. An area of highest contamination is observed in the Poa in zone 18 and zones 42 to 49. Areas 21-48 and 49 are infected in ryegrass and rough bluegrass.

*Figure 43: Geographical distribution of waikaviruses in Antheit grassland (global ecosystem)*

*Figure 44: Geographical distribution of waikaviruses in Antheit grassland (Lolium perenne L.)*

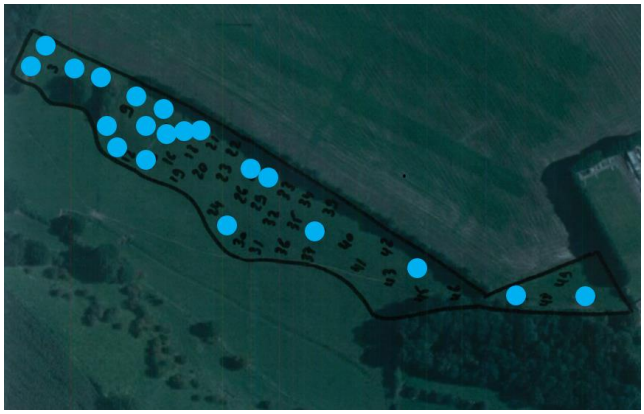


*Figure 45: Geographical distribution of waikaviruses in Antheit grassland (Poa trivialis* L.*)*

Figure 44 shows that Antheit grassland is very poorly infected and in two distinct areas for *Lolium perenne* L. Areas 47 to 50 are also infected for the global ecosystem and *Poa trivialis* L. (see Figure 43, Figure 45). The infected areas compared to the poll category are very different.

### 4.1.4. Co-infection

*Table 16: Percentage of different type of co-infection in Antheit pasture*

| | **AP2 Lolium** | **AP2 Poa** | **AP2 global ecosystem** |
|---|---|---|---|
| **BYDV-*Nepovirus*** | 6,00% | 20,00% | 22,73% |
| **BYDV-*Waikavirus*** | 0,00% | 2,00% | 0,00% |
| ***Nepovirus*-*Waikavirus*** | 12,00% | 8,00% | 0,00% |
| **Three viruses** | 0,00% | 26,00% | 2,27% |
| **No infection** | 10,00% | 14,00% | 31,82% |

*Table 17: Percentage of different type of co-infection in Antheit grassland*

| | **AG2 Lolium** | **AG2 Poa** | **AG2 global ecosystem** |
|---|---|---|---|
| **BYDV-*Nepovirus*** | 16,00% | 4,00% | 50,00% |
| **BYDV-*Waikavirus*** | 0,00% | 4,00% | 6,00% |
| ***Nepovirus*-*Waikavirus*** | 10,00% | 14,00% | 8,00% |
| **Three viruses** | 0,00% | 16,00% | 6,00% |
| **No infection** | 14,00% | 16,00% | 8,00% |

Co-infection tables (see Table 16, Table 17) show that no co-infection between BYDV and *Waikavirus* has been found in the ryegrass pool category. The same is true for cases of co-infection of the 3 viruses. No cases of BYDV-*Waikavirus* and *Nepovirus*-*Waikavirus* co-infection are observed in the Antheit pasture for the global ecosystem. The highest prevalence of co-infection is observed in the grassland for BYDV-*Nepovirus* co-infection at a value of 50%.

Figure 36 and Figure 44 show that all the ryegrass infected by the waikavirus candidate, are also infected by nepovirus, in Antheit pasture.

## 4.2. Bioinformatics

Alignment matrices (nucleotide and protein) are in appendix.

### 4.2.1. Candidate nepovirus

#### *4.2.1.1. Nucleotide alignment of the genome consensus sequences*

##### 4.2.1.1.1. RNA1

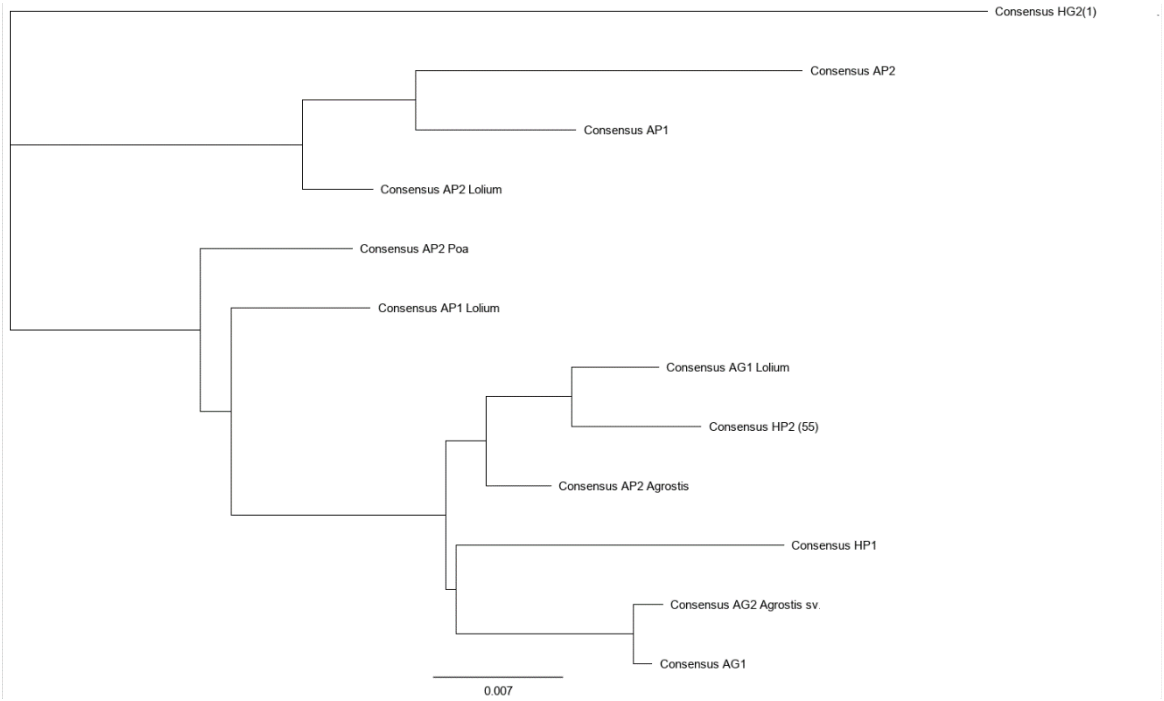



*Figure 46: Phylogenetic tree of consensus sequences of nepovirus RNA1 (complete genome)*

##### 4.2.1.1.2. RNA2

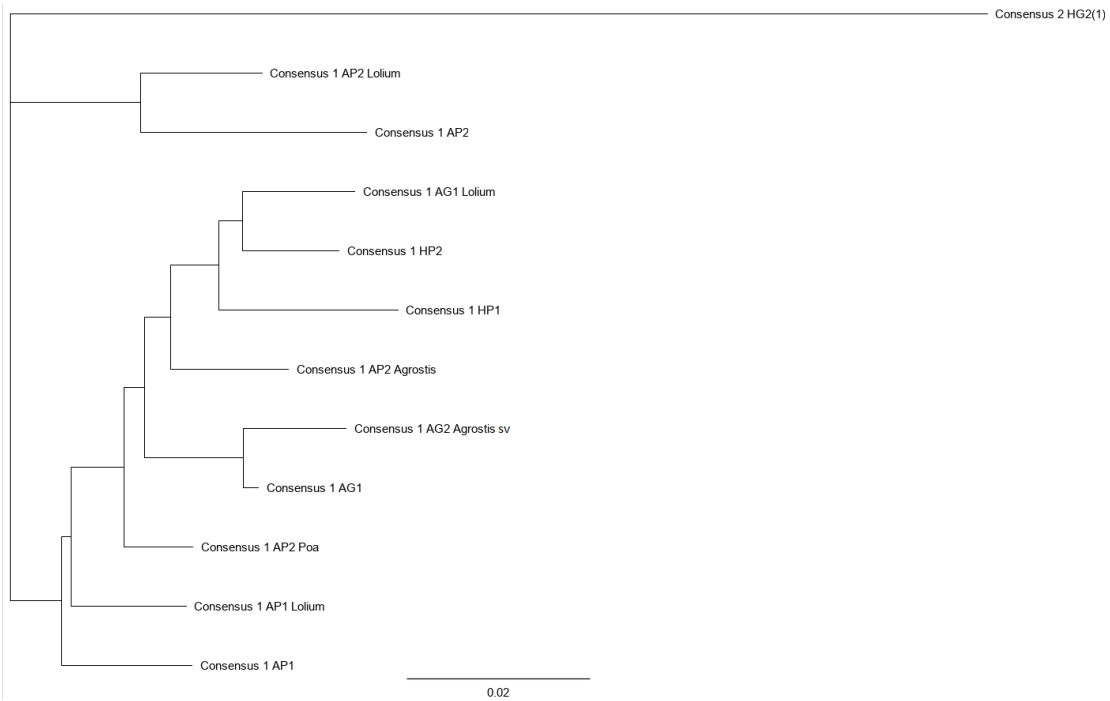



*Figure 47: Phylogenetic tree of consensus sequences of nepovirus RNA2 (complete genome)*

In Figure 46 and Figure 47, one branch is totally alone for RNA1 and RNA2; it is the consensus sequence of Heron grassland (year 2). A second demarcation is noticed with a branch that

contains Antheit pasture (year 2). The Antheit pasture year 1 sequence is in this branch for RNA1 but not for RNA2. In Appendix 1 and 2, the matrices give the percentages of identity between the consensus sequences that made it possible to build these trees. The lowest percentages of identity are observed for Heron grassland year2 (91.6% for RNA1 and 84.3% for RNA2). The highest identity percentages are observed for Antheit grassland Agrostis year 2 and Antheit grassland year 1 (99.7% for RNA1 and 97.6% for RNA2). The percentages of identity between consensus sequences are high and above threshold for species demarcation.

#### 4.2.1.1.3. Polymerase – protease (RNA1)

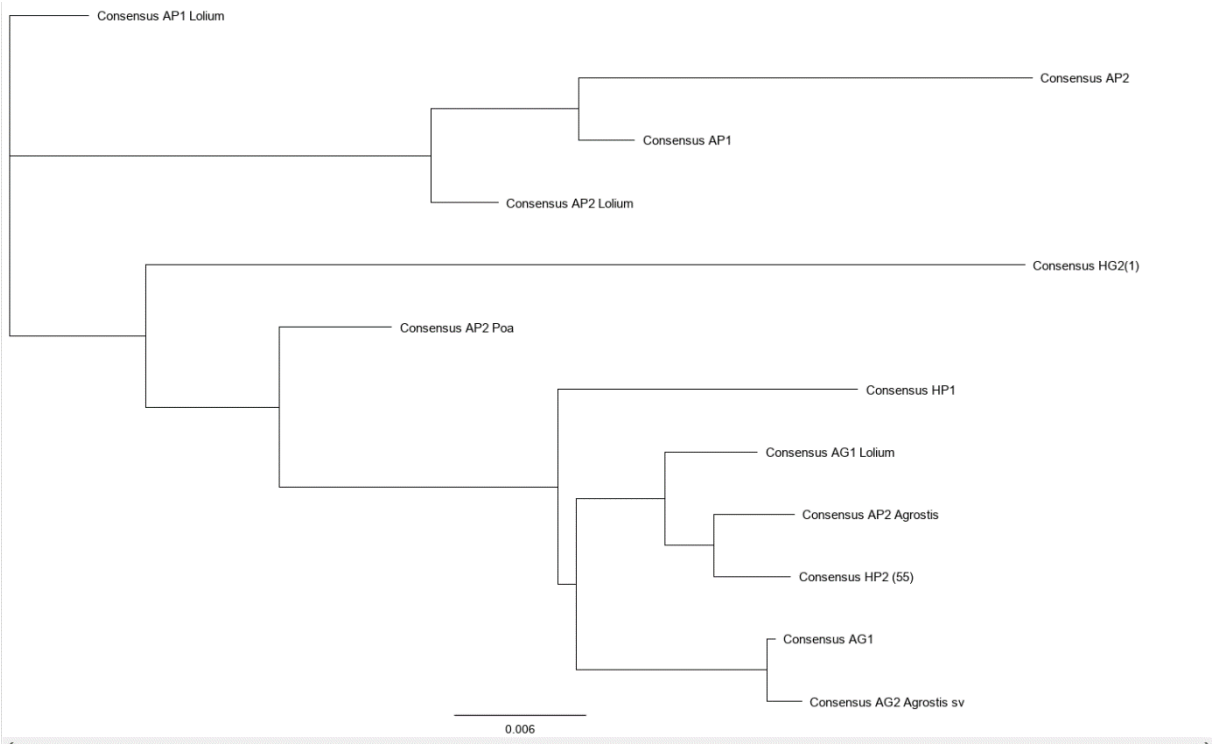



*Figure 48: Phylogenetic tree of consensus sequences of nepovirus RNA1 (polymerase-protease)*

Analysis of the nucleotides of the polymerase-protease zone located on RNA1 (See Figure 48) shows that the Antheit pasture branch is identical to that shown in Figure 46. There is another separate branch with the consensus sequence of Antheit pasture Lolium year1. In appendix 3, the lowest percentage is observed for Antheit pasture year2 (91.8%) and the maximum percentage is observed for Antheit grassland Agrostis year 2 and Antheit grassland year 1 (99.8%). The percentages of identity are high.

#### 4.2.1.1.4. Coat protein (RNA2)

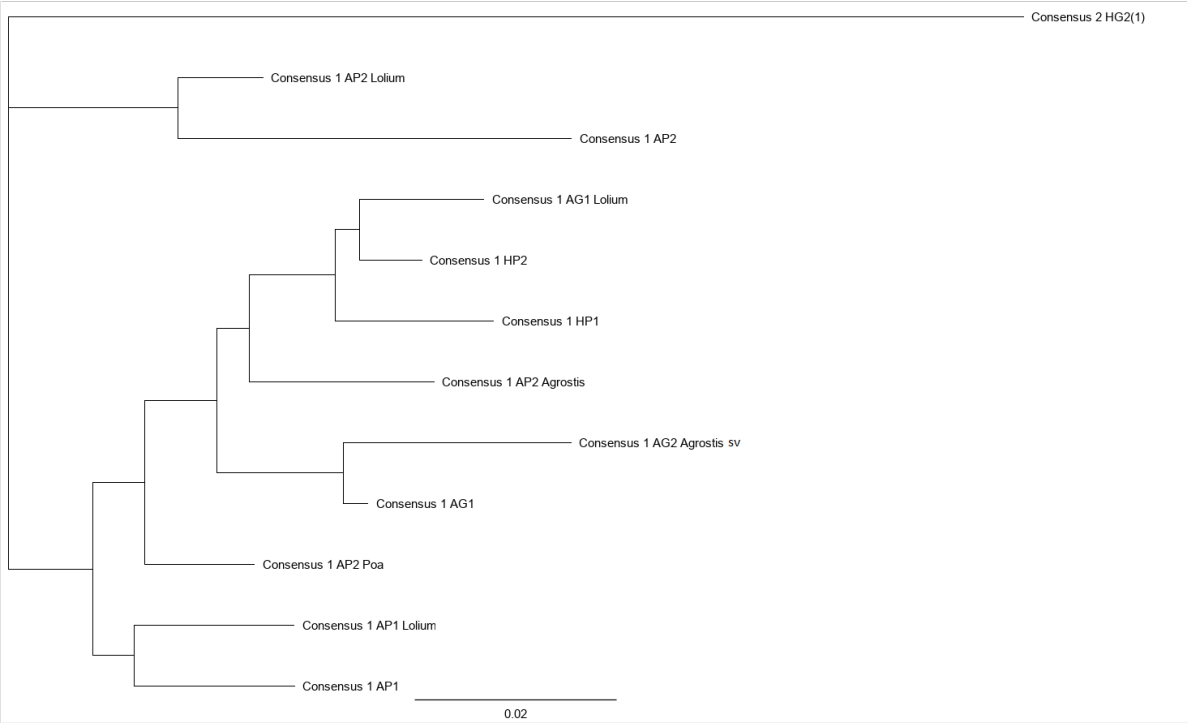



*Figure 49: Phylogenetic tree of consensus sequences of nepovirus RNA2 (coat protein)*

In the Figure 49 corresponding to the phylogenetic tree of the coat protein part of nepoviruses (RNA2), there is a branch with a single individual (Heron grassland year 2) and a branch with two consensus sequences of Antheit pasture year 2 (global ecosystem and *Lolium perenne* L.). In the matrix in Appendix 4 shows that the lowest percentage is found for Heron grassland year 2 (91.6%) and the maximum percentage is observed for Antheit grassland Agrostis year 2 and Antheit grassland year 1 (99.7%). The percentages of identity are high.

#### *4.2.1.2. Nucleotide alignment of consensus sequences and references of nepovirus*

##### 4.2.1.2.1. RNA1

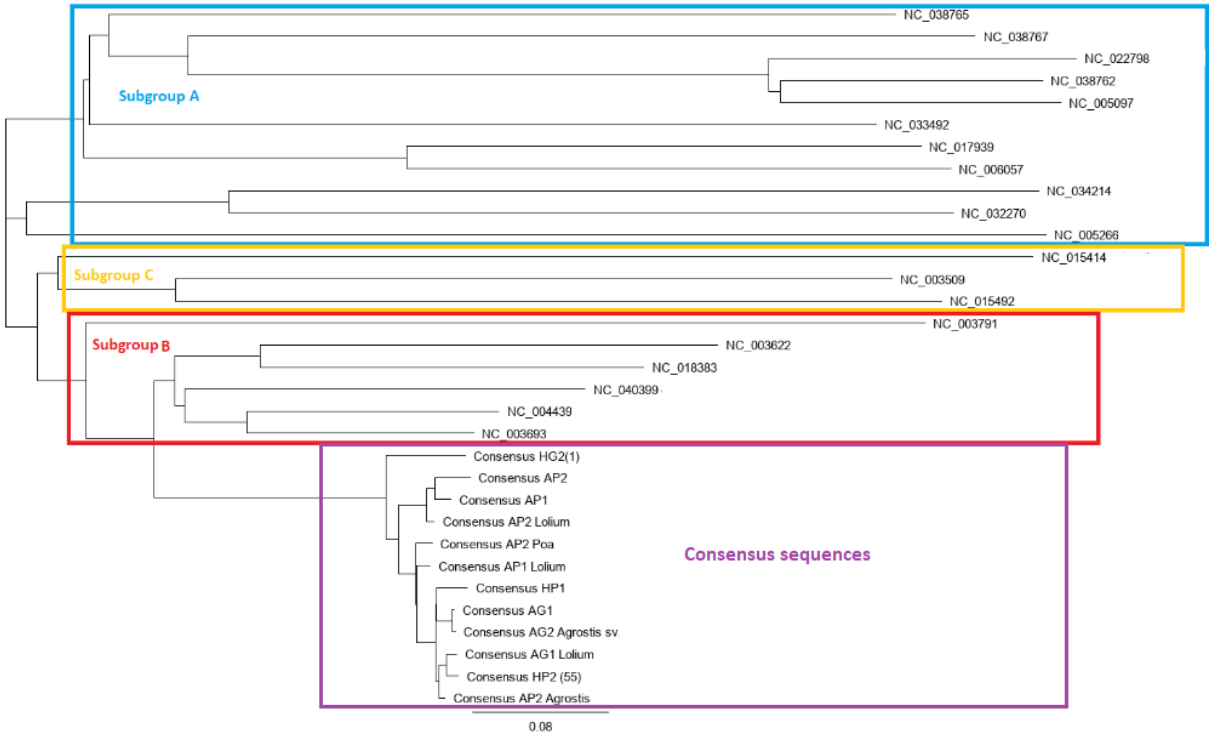



*Figure 50: Phylogenetic tree of consensus sequences of nepovirus RNA1 and references (complete genome). Coloured squares correspond to the three Nepovirus subgroups (blue for group A, orange for group C, red for group B and purple for consensus sequences). (Reference names: see Table 10).*

The phylogenetic tree of nepovirus RNA1 sequences and references (See Figure 50) highlights the three *Nepovirus* subgroups (blue for group A, orange for group C and red for group B). The consensus sequences are grouped at the bottom of the tree. Appendix 5 shows the matrix related to this tree; the percentages of correspondences between the consensus sequences are high and close (minimum 91.6%). The consensus sequences of subgroup B have percentages of identity with the consensus sequences of ecosystems between 48.9 and 70.6%.

#### 4.2.1.2.2. RNA2

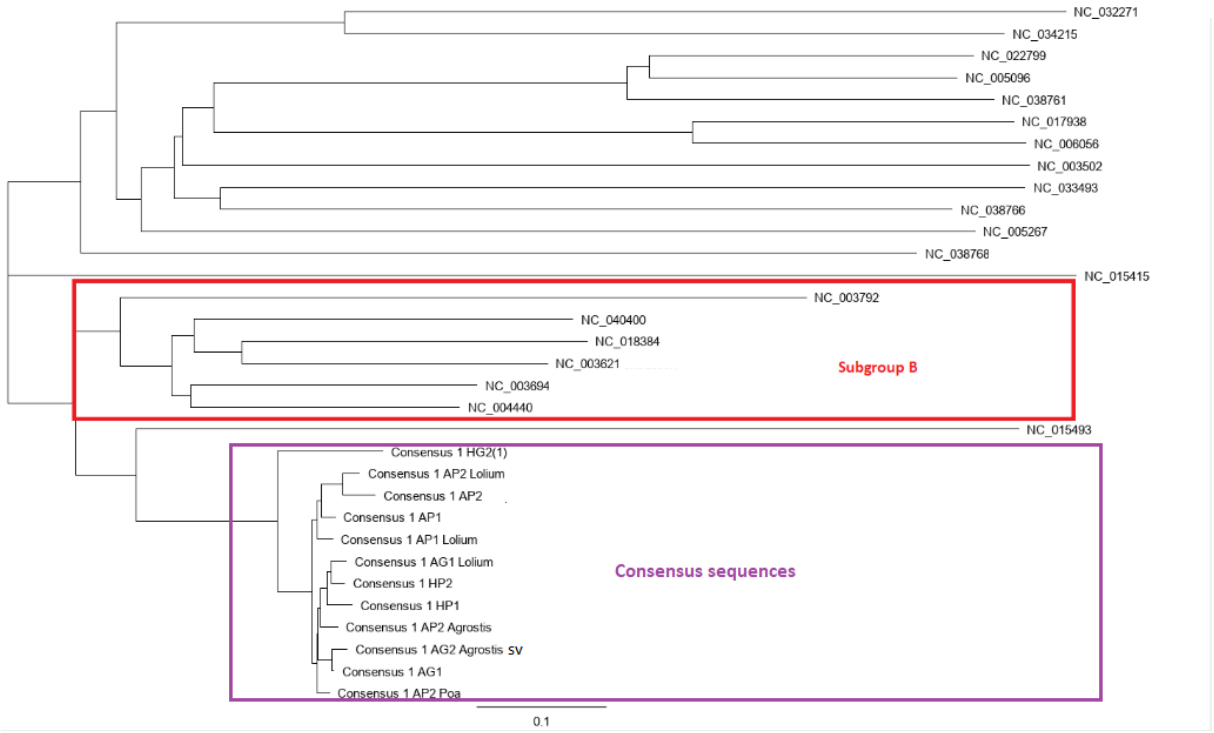



*Figure 51: Phylogenetic tree of consensus sequences of nepovirus RNA2 and references (complete genome). Red box is subgroup B of nepovirus, purple box is consensus sequences. (Reference names: see Table 11).*

In this tree (see Figure 51), only subgroup B is identified because subgroups A and C co-exist in the tree. The findings are the same as for the analysis in Figure 50. The consensus sequences are grouped together. Only the branch with reference NC_015493 (subgroup C) is located between the consensus and subgroup B with 31.2 to 33.6 % of identity with consensus. The matrix (Appendix 6) is similar to the matrix for RNA1. Subgroup B has percentages of identity with consensus sequences between 42.3 and 57.01 percent. The consensus sequences between them have a very high percentage of identity (minimum 85.02%).

#### 4.2.1.2.3. Polymerase – protease (RNA1)

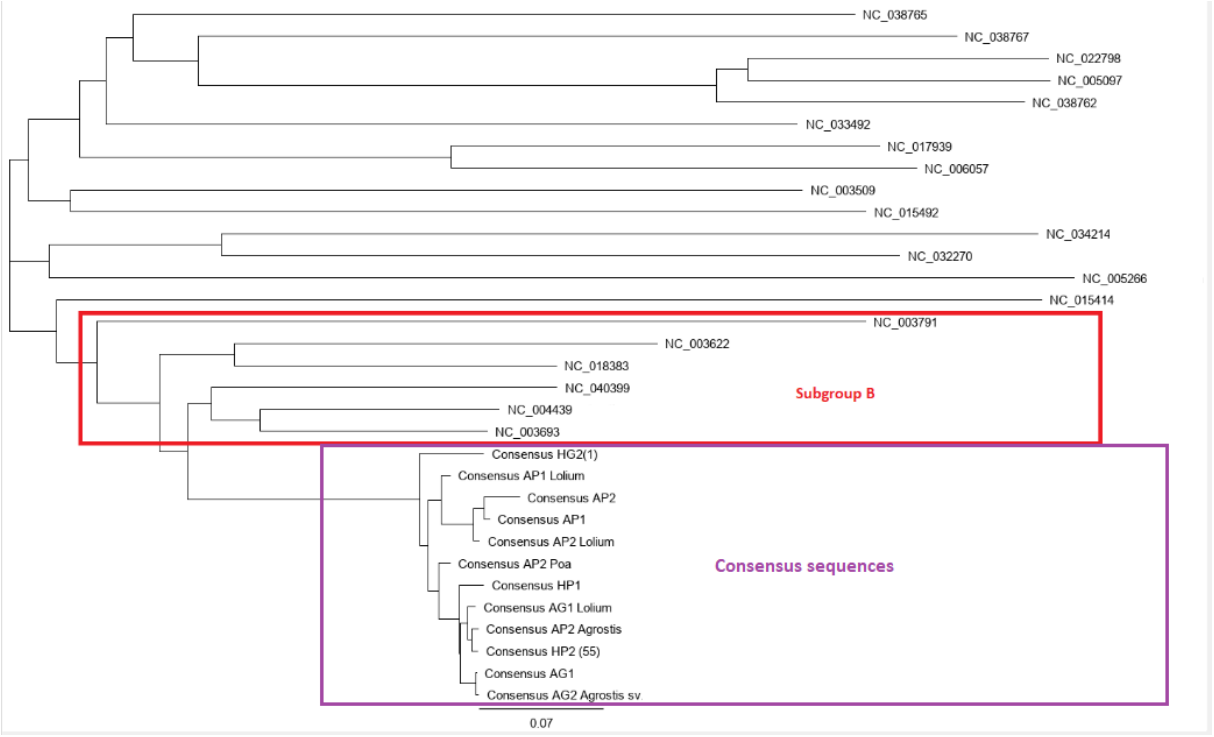



*Figure 52: Phylogenetic tree of consensus sequences of nepovirus RNA1 and references (Polymerase-protease). Red box is subgroup B of nepovirus, purple box is consensus sequences. (Reference names: see Table 10).*

Phylogenetic analysis of the polymerase protease part of the genome also shows that subgroups A and B are mixed. Subgroup C is identified in red in the figure (see Figure 52). Consensual sequences are also grouped together with close and high percentages (minimum 93.4% identity). Subgroup B has a percentage of identity with the different consensual sequences between 53.2 and 73.7 percent (see Appendix 7).

#### 4.2.1.2.4. Coat protein (RNA2)

For analysis of coat protein, the results are very similar to the results obtained for the analysis of polymerase-protease in section 4.2.1.2.3 (see Appendix 8: Figure 68 and Table 25). The consensus sequences are gathered on the same branch. Sub-group B is also identified. It also seems to be close to consensus sequences. The matrix confirms with the percentage of identity of the close consensus sequences (minimum 85.4%). Reference NC_003792 (*Cycas necrotic virus*) of subgroup B has the lowest percentages with consensus sequences and the reference NC_003694 (*Beet ringspot virus*) the highest percentages of identities.

#### *4.2.1.3. Protein alignment of consensus sequences and references of nepovirus*

##### 4.2.1.3.1. Polymerase – protease (CG-GDD) (RNA1)


*Figure 53: Phylogenetic tree of protein alignment of nepovirus RNA1 and references (polymerase-protease). Red box is subgroup B of nepovirus, purple box is consensus sequences. (Reference names: see Table 10)*

Here, the shaft is made for protein alignment. The sequence used for this analysis is within the polymerase protease sequence and lies between the CG and GDD domains (+/-450 bp). This is the system used by the ICTV because the rest of the polymerase protease is not preserved.

The tree (see Figure 53) makes it possible to highlight that the consensus sequences are grouped together and is very close to the reference sequences from subgroup B. The matrix (see appendix 9) shows correlated percentages between 53.9 and 91.1% identity with a maximum percentage resemblance to NC_004439 (*Tomato black ring virus*). The percentages of identity between the consensus sequences are at least 97.9%.

#### 4.2.1.3.2. Coat protein (RNA2)

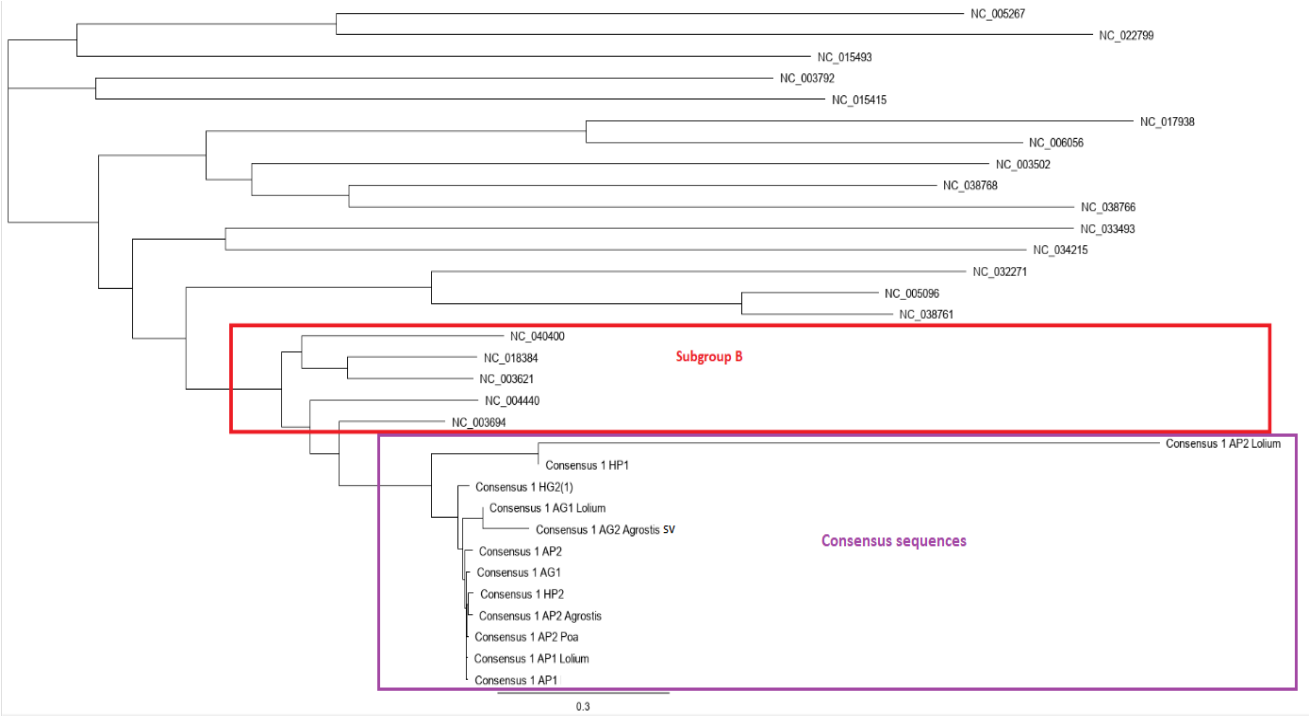



*Figure 54: Phylogenetic tree of protein alignment of nepovirus RNA2 and references (Coat protein). Red box is subgroup B of nepovirus, purple box is consensus sequences. (Reference names: see Table 11).*

On the phylogenetic tree in Figure 54, it is the analysis of the protein sequence of the coat protein in nepoviruses that is analyzed. The complete coat protein is kept for this analysis as well as for the ICTV analysis. The consensus sequences are well grouped except for the Antheit pasture Lolium year 2 sequences, which have only a small percentage of identity with the other consensus sequences (between 17.7 and 32.9%) (see Appendix 10). The other consensus sequences between them have identity percentages of at least 79.2%. The identity percentages of subgroup B with consensus sequences are between 15.2 and 66.9%.

## 4.2.2. Candidate waikavirus

### *4.2.2.1. Nucleotide alignment of the genome consensus sequences*

#### 4.2.2.1.1. Complete genome

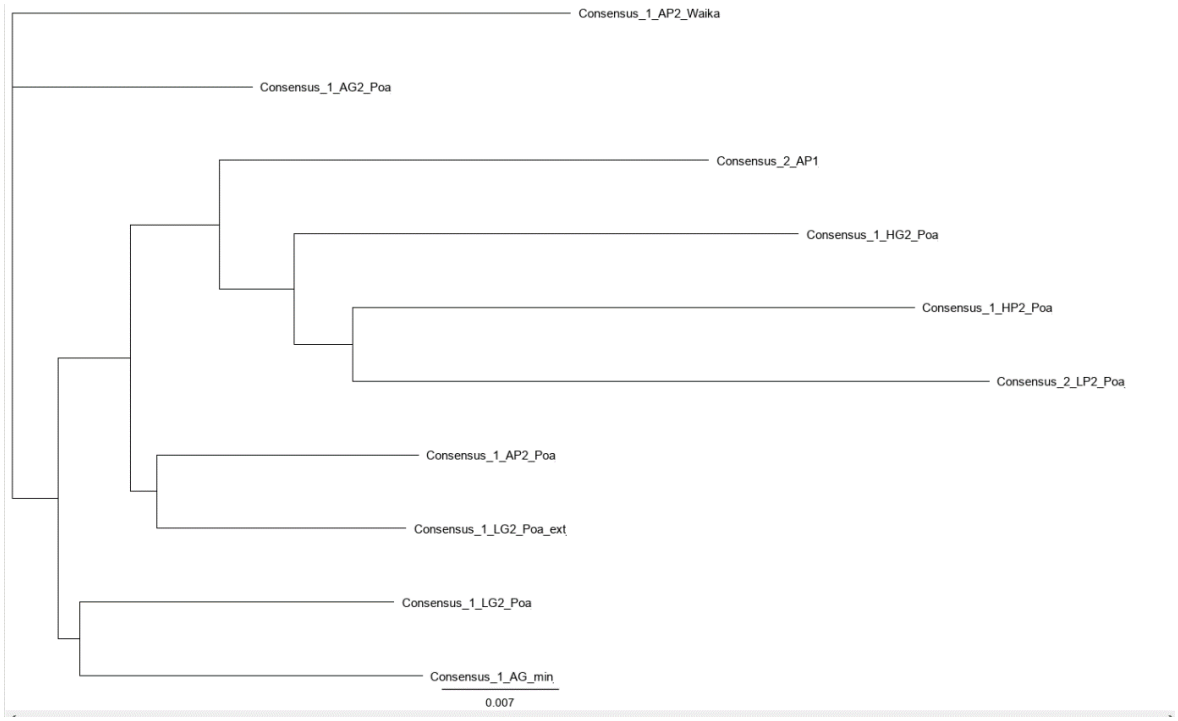



*Figure 55: Phylogenetic tree of consensus sequences of waikavirus (complete genome)*

#### 4.2.2.1.2. Coat protein

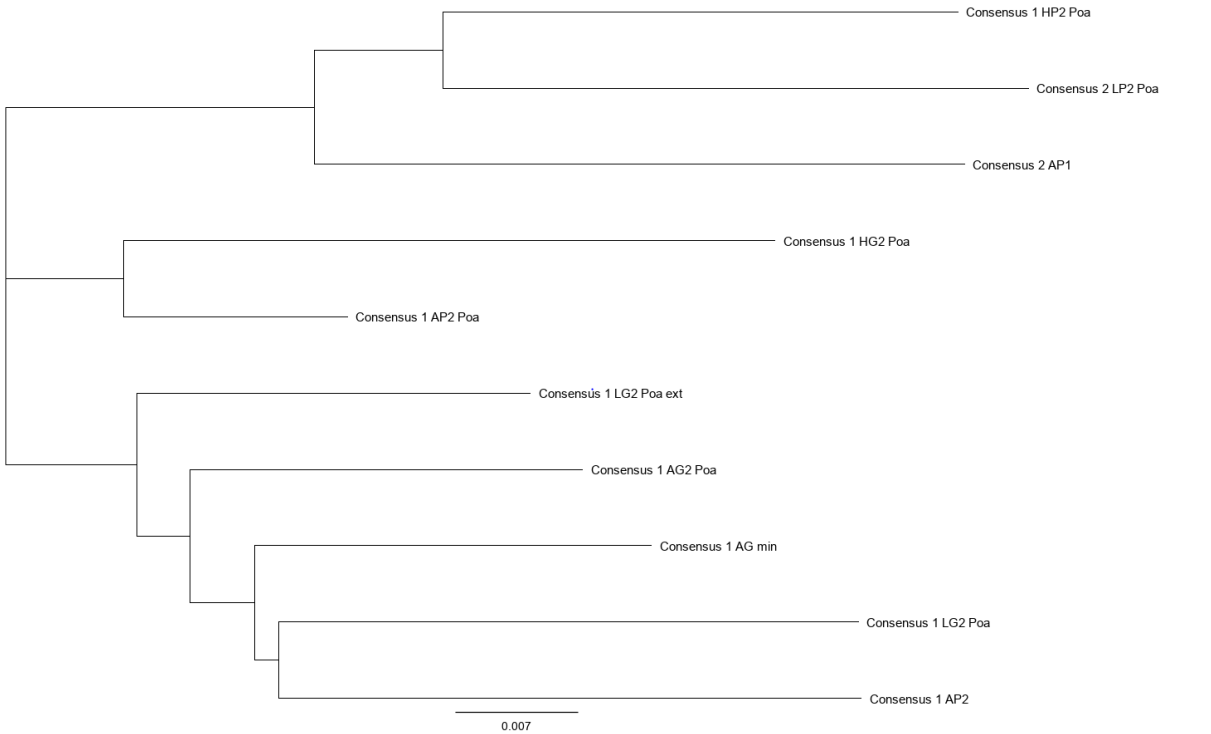



*Figure 56: Phylogenetic tree of consensus sequences of waikavirus (coat protein)*

#### 4.2.2.1.3. Polymerase – protease

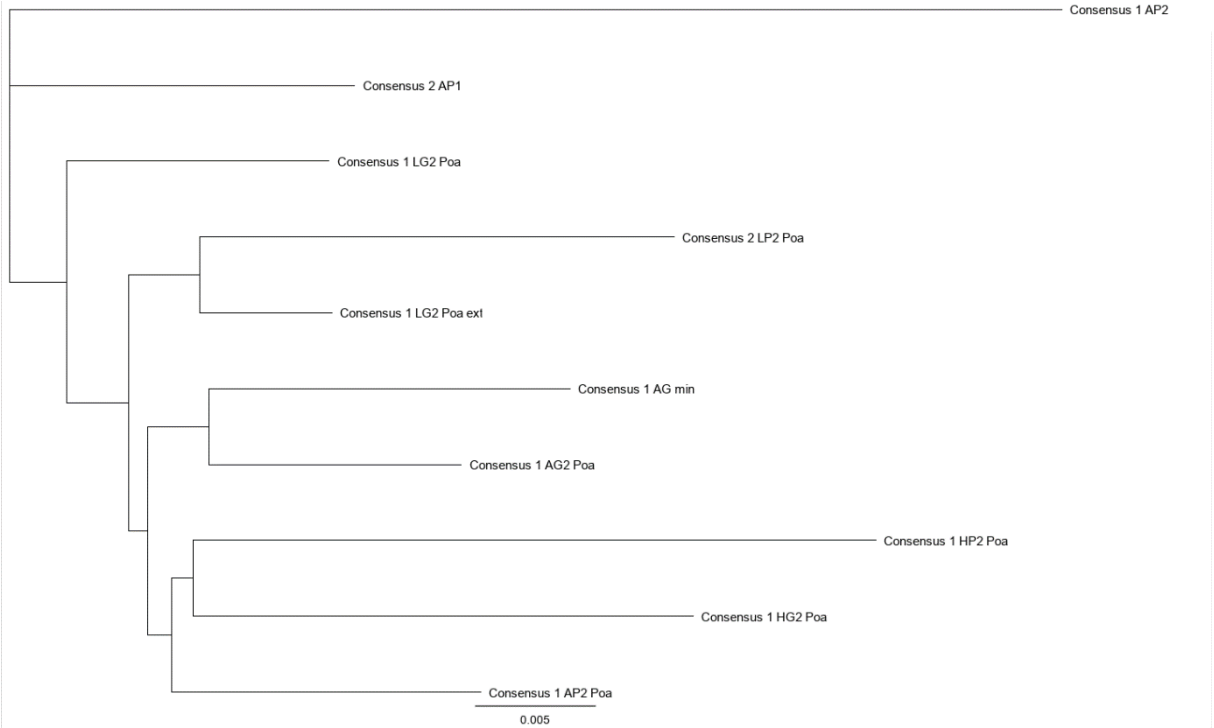



*Figure 57: Phylogenetic tree of consensus sequences of waikavirus (Polymerase-protease)*

Figure 55, Figure 56, Figure 57 show the phylogenetic analysis done on consensus sequences for the whole genome, coat protein zone, and polymerase protease zone, respectively.

The analysis of the entire genome reveals two ecosystems on unique branches Antheit pasture year 2 and Antheit grassland *Poa trivialis* L. year 2 (AP2 and AG2 Poa). The matrix (see appendix 11) shows that the percentages of identity are very close between 87.03% and 96.3%.

The analysis of the coat protein allows highlighting 3 distinct branches all composed of several ecosystems. The percentages of identity vary between 82.659% and 95.347% (see appendix 12).

Finally, the analysis of the polymerase-protease zone reveals two branches with a single ecosystem: Anhteit pasture year 1 and Antheit pasture year 2. The percentages of identity vary between 89.1% and 97.6% (see appendix 13).

#### *4.2.2.2. Nucleotide alignment of consensus sequences and references of waikavirus*

##### 4.2.2.2.1. Complete genome

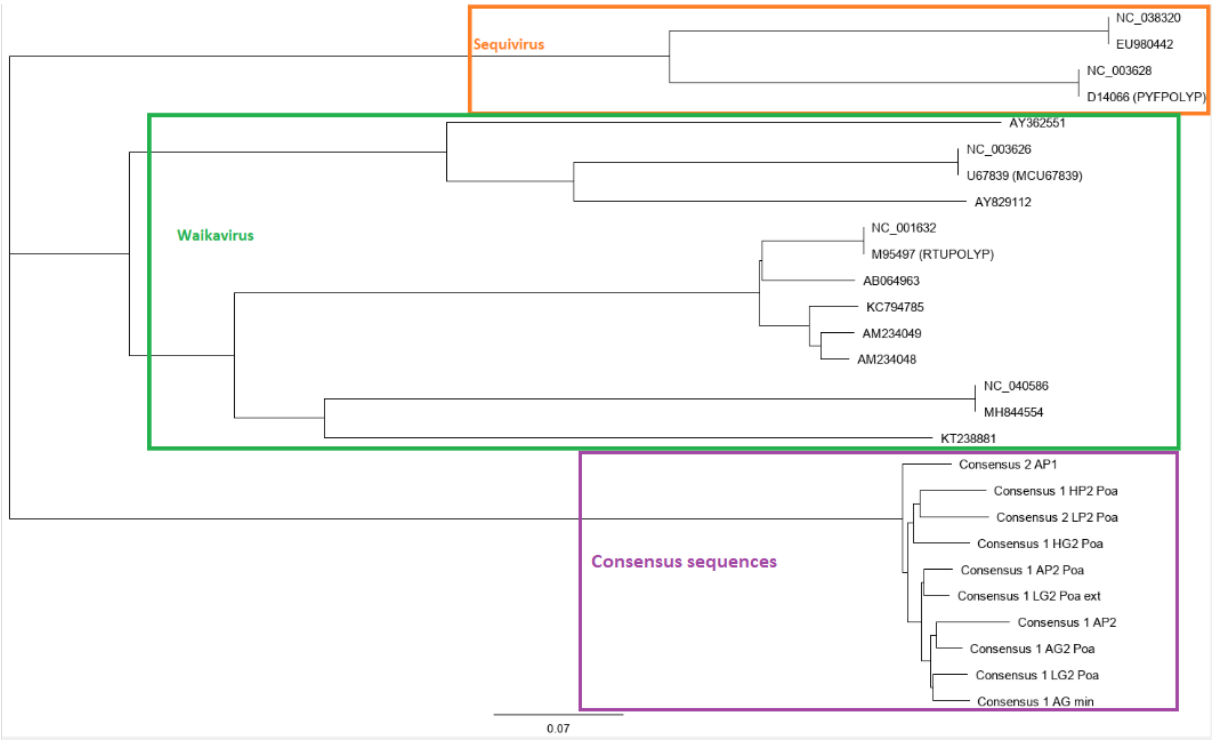



*Figure 58: Phylogenetic tree of consensus sequences of waikavirus and references (complete genome). Coloured squares correspond to Waikavirus, Sequivirus genera and consensus sequences (green for Waikavirus, orange for Sequivirus, purple for consensus sequences). (Reference names: see Table 9).*

Tree in Figure 58 is composed of consensus sequences and references of waikaviruses and sequiviruses. Three main branches are visible: one of them is composed of sequiviruses (orange box), another of waikaviruses (green box) and the last one is composed of ecosystem consensus sequences. Three zones are quite distinct. The matrix in Appendix 14 shows the same areas, with darker areas (which correspond to high identity percentages) for consensus sequences, waikaviruses, and sequiviruses. The identity percentages of the consensus sequences between them are at least 85.9%.

#### 4.2.2.2.2. Coat protein

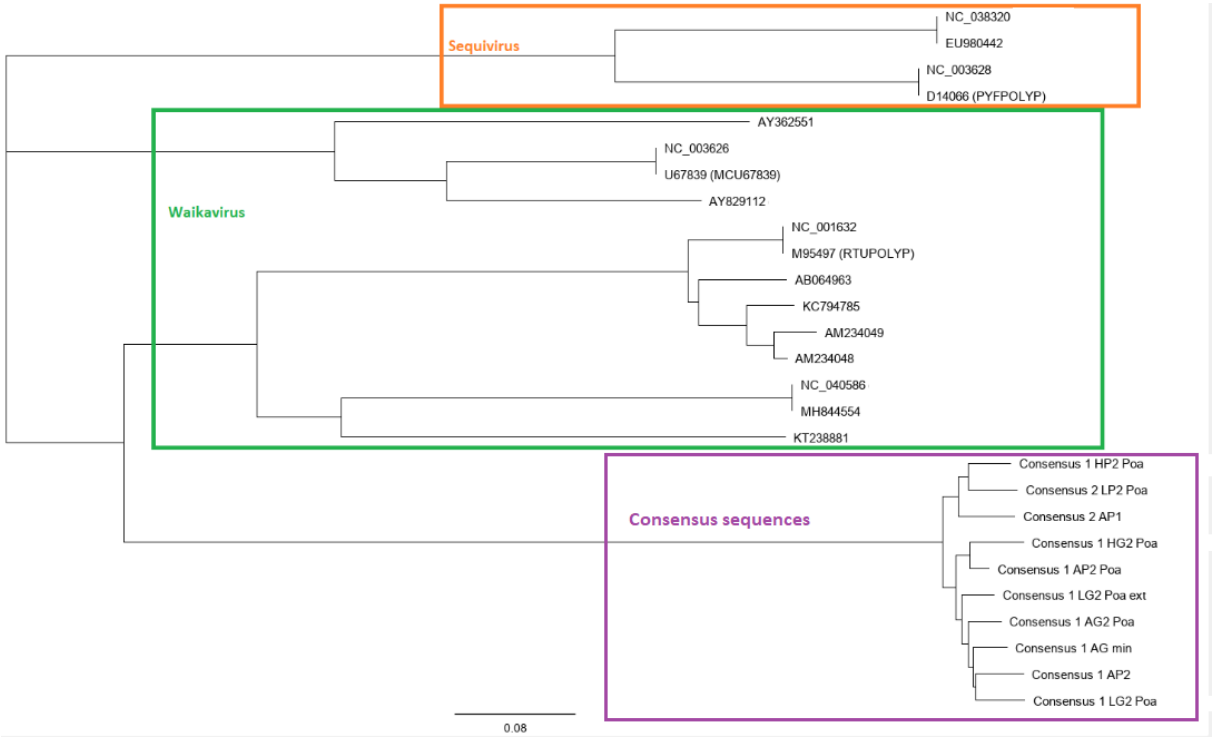



*Figure 59: Phylogenetic tree of consensus sequences of waikavirus and references (coat protein). Coloured squares correspond to Waikavirus, Sequivirus genera and consensus sequences (green for Waikavirus, orange for Sequivirus, purple for consensus sequences). (Reference names: see Table 9).*

For the nucleotide zone that makes up the coat protein divided into 3 parts in waikavirus (see Figure 59), there is a branch with sequiviruses, an area divided into 2 branches for waikaviruses and an area with consensus sequences. The matrix that corresponds to this phylogenetic tree (see Appendix 15) shows 3 areas that correspond to consensus sequences, waikaviruses, and sequiviruses. Waikaviruses have higher identity percentages with consensus sequences than sequiviruses. The identity percentages of the consensus sequences between them are at least 83.2%.

#### 4.2.2.2.3. Polymerase – protease

For analysis of polymerase protease, the results are very similar to the results obtained for the analysis of the complete sequence (see Appendix 16: Figure 69 and Table 33 ).For the genome zone corresponding to polymerase protease, there are also 3 distinct branches, for waikavirus in green, sequivirus in orange and consensus sequences. The matrix also shows these three areas with higher percentages of correspondences between waikaviruses and consensus sequences than between sequiviruses and consensus sequences. The minimum percentage of identity between consensus sequences is 89.1% minimum.

#### *4.2.2.3. Protein alignment of consensus sequences and references of waikavirus*

##### 4.2.2.3.1. Coat protein

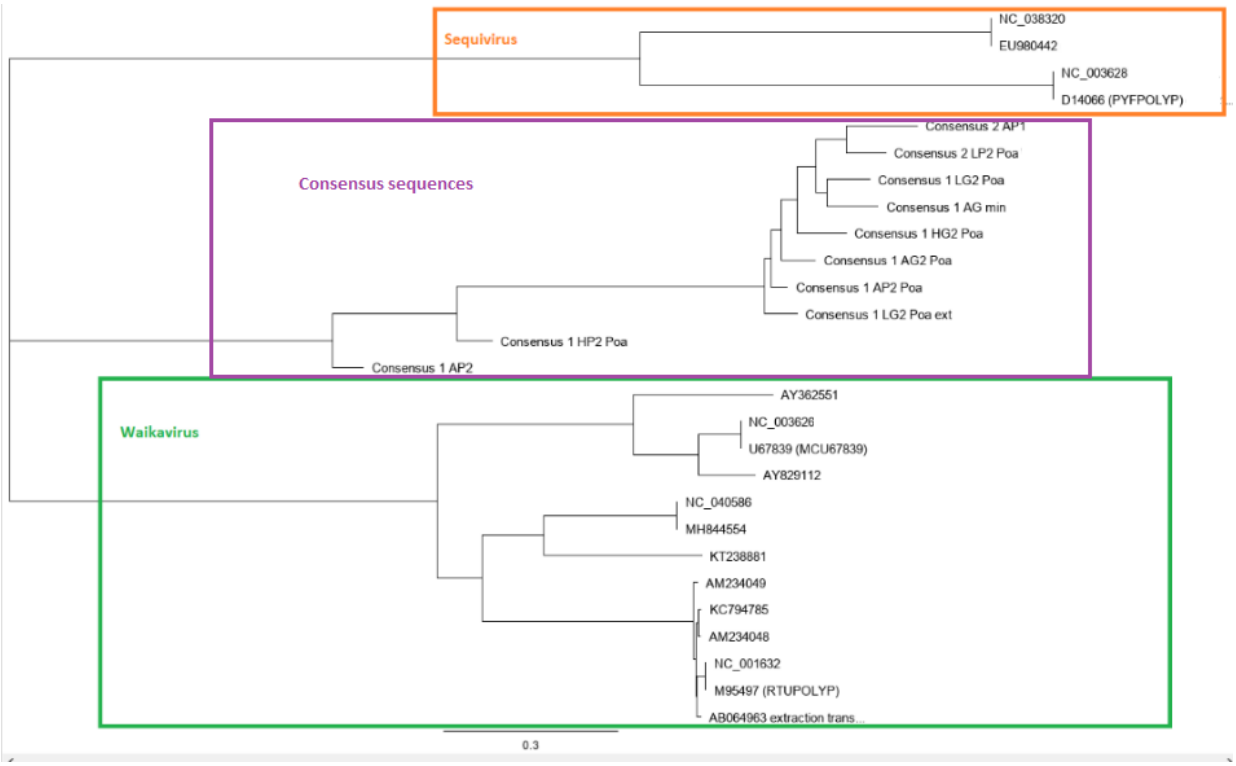



*Figure 60: Phylogenetic tree of protein alignment waikavirus and references (Coat protein). Coloured squares correspond to Waikavirus, Sequivirus genera and consensus sequences (green for Waikavirus, orange for Sequivirus, purple for consensus sequences). (Reference names: see Table 9).*

In this tree (see Figure 60), it is the alignment of protein sequences that are studied. The coat protein is divided into three parts in waikaviruses. The three parts are studied together. In the figure, there are the 3 branches as in the previous trees: one for waikavirus, one for sequivirus and one for consensus sequences. The sequences AP2 (Anthait pasture year 2) and HP2 Poa (Heron pasture *Poa trivialis* L. year 2) are slightly different from the consensus sequences. The corresponding matrix (see Appendix 17) shows the three areas corresponding to the three main branches. For AP2 and HP2 Poa, the percentages of identities with the other consensus sequences are between 33.3 and 51.1% while for the other sequences the percentage is at least 72%. The percentage of identity between consensus sequences and references is low: maximum 17.4%.

#### 4.2.2.3.2. Polymerase – protease (CG-GDD)

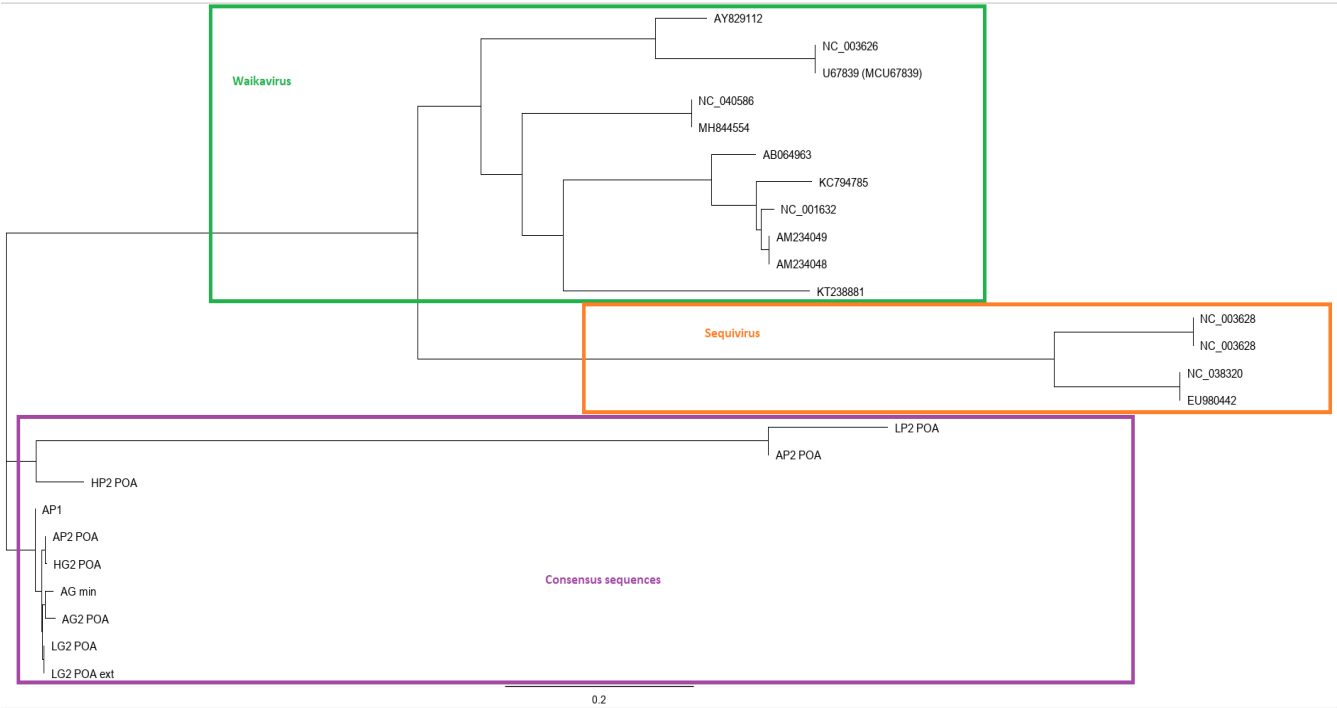



*Figure 61: Phylogenetic tree of protein alignment waikavirus and references (Polymerase-protease). Coloured squares correspond to Waikavirus, Sequivirus genera and consensus sequences (green for Waikavirus, orange for Sequivirus, purple for consensus sequences). (Reference names: see Table 9).*

Here, in Figure 61, it is the protein sequence of the polymerase protease (only the CG-GDD sequence like ICTV). There are two main parts in the trees one part with waikavirus references and sequivirus references (green and orange box) and consensus sequences in the other part. Two consensus sequences are further removed from the other consensus sequences with LP2 Poa (Latinne pasture *Poa trivialis* L. year 2) and AP2 Poa (Antheit pasture *Poa trivialis* L. year 2). The HP2 sequence Poa (Heron pasture *Poa trivialis* L. year 2) is also found on this branch. The matrix (See Appendix 18) shows that the LP2 Poa and AP2 Poa sequences show only between 28.7% and 31.3% identity percentage with the other consensus sequences. The other sequences have at least 80.1% identity with each other. The study was done on 25 sequences instead of 27 because two references did not have the CG-GDD sequence (M95497, AY362551). It can also be seen in the matrix that the areas of correspondence between waikaviruses and sequiviruses are confused.

#### 4.2.2.3.3. Polymerase-protease consensus sequence (QA- STOP codon)

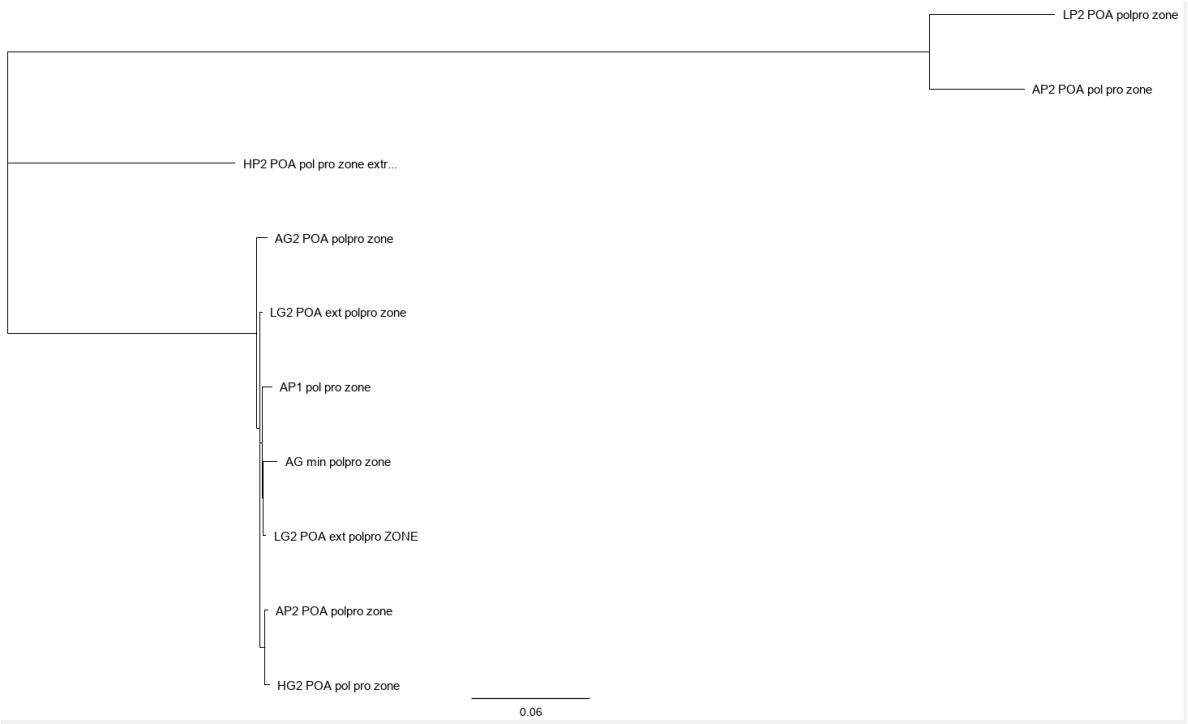



*Figure 62: Phylogenetic tree of protein alignment Waikavirus (Polymerase-protease) (QA-STOP codon)*

Figure 62 shows the phylogenetic tree of the alignment of proteins from the identified sequence in the sequences for polymerase protease between the QA model and the STOP codon (+/- 900 bp). This tree also shows that the LP2 Poa and AP2 Poa sequences are far from the other consensus sequences. The matrix (Appendix 19) confirms this analysis with between 38.9% and 42.9% of correspondence between LP2 Poa and AP2 Poa and the other consensus sequences. There is an 83.3% match between these two sequences. The other sequences have at least 70% identity with each other.

### 4.2.3. *Secoviridea* tree ICTV

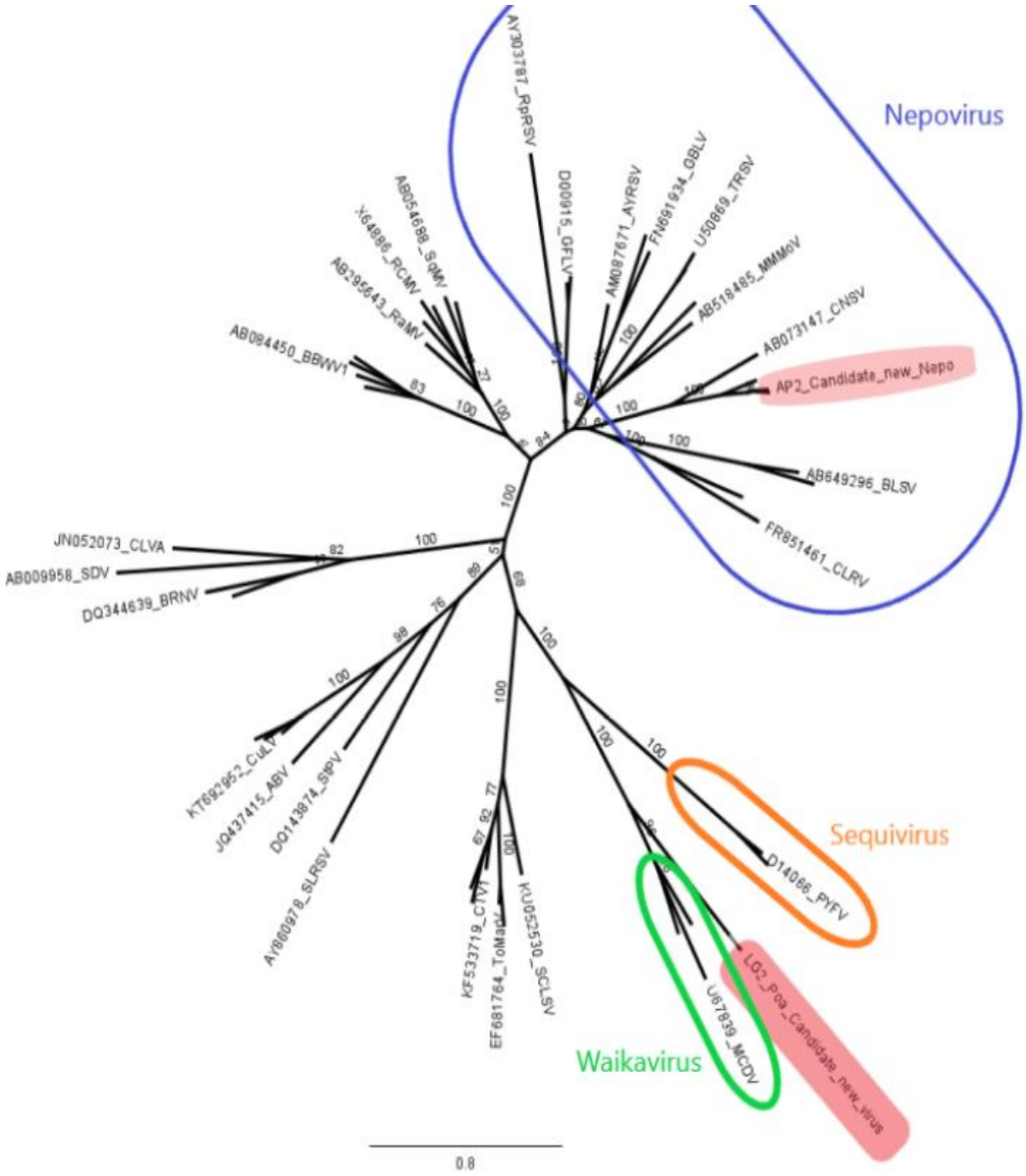



*Figure 63: Phylogenetic tree of Secoviridea with Waikavirus consensus genome. Coloured circles correspond to Waikavirus and Sequivirus and Nepovirus genera (green for Waikavirus, orange for Sequivirus and blue for Nepovirus). (Reference names: see Table 13).*

The tree above (Figure 63) corresponds to the tree used by the ICTV (zone CG-GDD) for the classification of *Secovivridea*. In this tree, the two consensus sequences for nepovirus and waikavirus were injected. LG2 Poa (Latinne grassland *Poa trivialis* L. year 2) for the waikavirus candidate and AP2 (Antheit pasture year 2) for the nepovirus candidate.

For the *Waikavirus* candidate, the sequence is located in a new and well-discussed branch that lies between waikaviruses and sequiviruses. The nepovirus candidate also has a new branch, but it is very close to *Beet ringspot virus* and *Tomato black ring virus*.

### *4.3. White clover mosaic virus*


*Figure 64: Geographical distribution of White clover mosaic virus in Héron pasture (Lolium perenne L.)*


*Figure 65: Geographical distribution of White clover mosaic virus in Héron pasture (Trifolium repens L.)*

The prevalence of *White clover mosaic virus* in the Heron pasture is 98% for white clover and 46% for ryegrass. The only area not infected with the virus for white clovers is area 43, which is also not infected in ryegrass. A zone of highest infection for ryegrass is located between zone 15 and zone 27 (see Figure 64, Figure 65)

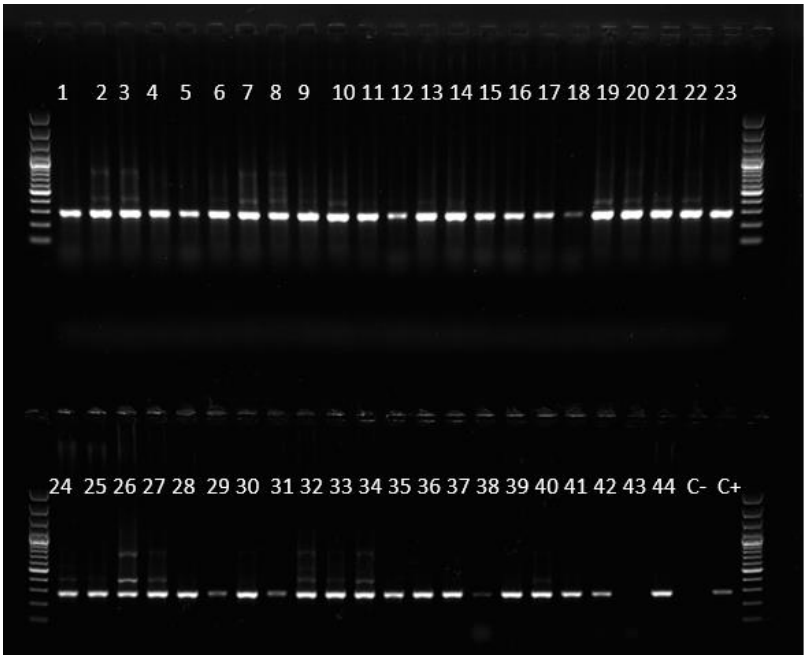



*Figure 66 : PCR gel picture White clover mosaic virus in HP (Héron pasture) (Trifolium repens L.). Sample from 1 to 44 and negative and positive control.*

Figure 66 shows the result of a PCR gel for Héron pasture for white clover mosaic virus in clover. It is thanks to its results that the prevalence has been calculated (presence or absence of white bands at the amplicon size level. Some bands are more apparent than others such as the 29 and 38 sample band for example.

The percentage of identity between the consensus sequence of white clover mosaic virus and the reference sequence of *White clover mosaic virus* (NC_003820) is 94.5% (see Figure 67)

| | Consensus | NC_003820 | LC159489 | LC159490 | LC159488 | AB056720 | AB669182 |
|---|---|---|---|---|---|---|---|
| Consensus | | 94.531% | 86.560% | 86.636% | 86.483% | 86.881% | 86.847% |
| NC_003820 | 94.531% | | 87.819% | 87.904% | 87.853% | 88.075% | 88.092% |
| LC159489 | 86.560% | 87.819% | | 98.460% | 94.164% | 94.951% | 94.968% |
| LC159490 | 86.636% | 87.904% | 98.460% | | 94.215% | 94.985% | 95.003% |
| LC159488 | 86.483% | 87.853% | 94.164% | 94.215% | | 95.396% | 95.448% |
| AB056720 | 86.881% | 88.075% | 94.951% | 94.985% | 95.396% | | 99.812% |
| AB669182 | 86.847% | 88.092% | 94.968% | 95.003% | 95.448% | 99.812% | |

*Figure 67: Matrix of identity percentage for White clover mosaic virus consensus sequences and reference. (Reference names: see Table 12).*

# 5. Discussion

This part is divided into three main sections: the first section concern results for the prevalence and geographical distributions of viruses in the plots studied, the second one examine bioinformatics analyses for the new waikavirus and nepovirus candidates, and the last part consist on analyzing presence of *White clover mosaic virus* in *Lolium perenne* L. in Heron pasture.

## 5.1. Prevalence

Examination of virus prevalence in field, grassland, and pasture in Antheit has highlighted several elements. It is also essential to specify that virus analysis was carried out by PCR with primer sequences based on the consensus sequences of the potentially new species of Waikavirus and Nepovirus. The fact that these viruses were found by PCR, therefore, confirms that the primer was effective and that these potential new viruses were indeed found in the plant communities studied.

First, analysis of BYDV in Antheit field has shown only 6% prevalence of this virus. This is the opposite of what can be found in the literature. Indeed, in fields with low plant diversity, the diversity of viruses present is low, but the infection rate is usually very high [5]. Further investigations, such as analysis of Latinne and Héron fields, should be carried out in this area to understand why such a low prevalence has been found or if this low prevalence is only present in Antheit field.

For the other plant communities, the prevalence of BYDV ranged from 6 % (Antheit Pasture for *Lolium perenne* L.) to 72% (Antheit grassland, global ecosystem). In the literature, high percentages of BYDV infections have been reported as high as 84% in a New Zealand pasture (ryegrass). Pastures from 6 to 30 years old present more than 50% of infections. The age of pasture also has an impact [111]. Here, pastures analyses for BYDV presented from 6 to 56% prevalence, but the ryegrass only shows 6% identity. For grassland, prevalence can reach 59% in some plant species [112]. Here, the prevalence varied from 24 to 72%. Analysis of Heron and Latinne pastures could provide additional informations to determine whether this low infection rate is only present in Antheit or whether it is generalized to other plots.

Virus analysis is complicated because virus concentration in the plant can vary. During the PCR analysis, it was, therefore, possible to have much more evident bands than others, and it could sometimes be difficult to detect the presence of viruses in a very poorly infected plant.

Nepoviruses had the highest prevalence in all plant communities (except for Antheit grassland, global ecosystem) between 90% and 59 % except for the analysis of minor species for which the prevalence is lower. And *Lolium perenne* L. is more infected than *Poa trivialis* L. This can be seen in bioinformatics analyses where the analyses have been carried out in many *Lolium perenne* L. plant communities. Analysis of infection of different plants that make up the global plant communities has shown that only *Dactylis glomerate* L. and *Festuca rubra* L. are not infected. But there were only three individuals of *Dactylis glomerate* L. and only one *Festuca rubra* L., so the results cannot be representative of these two species.

Waikaviruses had a higher prevalence in *Poa trivialis* L. This was the case in Antheit pasture and Antheit grassland and is also confirmed by bioinformatics analyses. Unlike nepovirus, waikavirus infected fewer species in the study of global plant communities. Waikaviruses tend

to attack only a few plant families (rice and maize) [52] while nepoviruses attack many plants in *Poaceae* [2]. This may explain the lower prevalence among waikaviruses.

It is important to note that only two global plant communities have been precisely analysed and only for Antheit. The results are, therefore, valid for this ecosystem but not necessarily for the others. Some species are only present as one or two individuals, and infected plants of these species would have been missed if the prevalence is low. It will take more individuals to know if these viruses are able to infect these species. Same tests will be carried out in two other places (Heron and Latinne) also in a field, grassland and pasture. These tests will confirm the results obtained or qualify them according to the results obtained.

Minor species were less infected by viruses than the other plant communities studied. The percentages of infections were lower.

Examination of eventual co-infections between BYDV, nepoviruses, and waikaviruses are also very interesting. Various analyses show that no BYDV-*Waikavirus* co-infection has been recorded in *Lolium perenne* L. plant communities in the two plant communities studied. This lack of co-infection could be explained by the low rate of BYDV infection in these two plant communities. Indeed, some plants present co-infection with BYDV-*Nepovirus* or *Nepovirus-Waikavirus*. In addition, no cases of BYDV-*Waikavirus* co-infection were detected in ryegrass individuals collected as part of the global plant community study. Case where two viruses cannot remain in the same plant, it is a exclusion phenomenon [113]. But cases of BYDV-*Waikavirus* co-infection have been observed in the Antheit grassland plot explained in the following paragraphs. So BYDV and waikaviruses can coexist. There are no exclusion phenomena in this case.

No case of co-infection with the three viruses has been observed in the same plant communities with *Lolium perenne* L. However, in the Antheit grassland plot as part of the overall plant community study, two cases of triple co-infection were detected in *Lolium perenne* L. So, waikavirus and BYDV can be on the same plant. Prevalence of BYDV being much higher in this plot, this could explain the presence of this co-infection.

Co-infection between BYDV-*Waikavirus* and *Nepovirus-Waikavirus* co-infection was not observed in the Antheit pasture plot (global plant community), but these types of co-infection were present in the Antheit grassland plot (global plant community) with a low prevalence of 6 and 8% respectively. This can be explained by the much higher presence of waikavirus in AG2 (38%) than in AP2 (4.55%). Indeed, the Antheit grassland is more infected by the three viruses than the Antheit pasture.

The highest percentage of co-infection is observed for BYDV-*Nepovirus* co-infection in the Antheit grassland global plant community (50%). It is in this plant community that the highest prevalence of BYDV is also found (76%). Nepoviruses also has a high prevalence in this plot. Which (66%). This could explain this high rate of co-infection.

Percentages of co-infection did not exceed 26% except for the case explained above. The non-infection rate never reaches more than 16% except for Antheit pasture (global plant community) with a value of 31%. It has the lowest values of nepovirus and waikavirus infections (excluding minor species). She has a 29.55% rate of BYDV infection. These low infection rates may explain the relatively high rate of non-infection.

## 5.2. Geographical distribution

Geographical distribution of viruses in plots can be very informative and can help to understand how a virus spreads and interacts in a particular environment. Here, pastures and grasslands are adjacent, and it is therefore very interesting to study the behaviour of viruses in these environments.

### *5.2.1. Nepovirus*

Some areas of the pasture are more infected than others. Indeed, the area between number 1 and number 9 and 31 is mostly infected by nepoviruses as shown on the maps. The border areas at the top of the meadow are also particularly infected, mainly for the pool category *Lolium perenne* L. and *Poa trivialis* L. The area in the recess in the meadow seems to be more spared by infection, especially in the global plant community and *Poa trivialis* L.

Grassland, which is located next to the pasture, also contains areas that are more infected than others. The areas at both ends are highly infected in all three pool categories (1 to 5 and 47 to 50). The central area of the plot is also mainly infected. The other zones are spared (zones 7 to 14 and 21-23).

Harvest made this year, on only 30 samples, but which covered all the grassland, confirms the previous observations. Indeed, a very infected area on the area edge (24 to 28 which corresponds to zone 1 to 5) and a more infected central area. The second border area has not been resampled, but an area is infected in this area.

If we look at maps between two plant communities, it is possible to make some connections between infected areas. In pool category global plant community, an area covered in the pasture (45-56-47) is adjacent to an area covered in the grassland (7-8-12-13). Central area of grassland infected in the three pool categories has a contact zone with a highly infected area of pastures mentioned above (between 1 and 9-31).

Contamination patches (area with more infected plant) that are observable in Antheit pasture and grassland can be explained by the fact that nepoviruses are usually transmitted by nematodes [56]. Indeed, nematodes are found in some areas of grassland and transmit the virus to plants in this area. It also explores the areas of contamination that correspond between the two plots. These two parcels are contiguous, and nematodes are present in both plant communities. This explanation can easily be verified by analysing nematodes contained in the soil of these prairies. Fact that this year's analyses confirm the distribution pattern of virus can also be explained by nematodes in soil that do not move much. Nepoviruses are known to have specific infection areas as their characteristics.

### *5.2.2. Waikavirus*

In the case of waikaviruses, pasture presented low prevalence. However, an area was more infected in the case of *Poa trivialis* L. in the upper pasture area between 15 and 50 L-shaped. Remaining infected areas do not appear to have any particular geographical pattern, or the area is not readily identifiable because the pasture is very poorly infected in the three pool category cases.

As far as the grassland is concerned, an area from 47 to 50 is infected in each pool category. The study of *Poa trivialis* L. shows that the most infected area is between 1 and e 17. The other plant communities do not have a particular pattern of infection. In addition, for *Lolium perenne* L., there are only five infected areas.

Nor does it seem to have any correspondence between pasture and grassland.

Waikaviruses are transmitted by aphids and leafhoppers. Insect transmission of viruses does not present, like nepovirus, particular areas of infection because insects move more quickly. However, the few areas that appear to be more infected could be explained by a significant proliferation of these insects in these areas. It may be interesting to study the proliferation of insects in the areas concerned.

## 5.3. Bioinformatics

Bioinformatics analyses, as part of this study, made it possible to perform precise phylogenetic analyses of two previously identified new virus candidates. These phylogenetic analyses allow us to know if new sequence is really a new virus or if this virus has already been identified. They also make it possible to know his phylogenetic position in genus or family to which he belongs. ICTV has also established criteria to identify new viral species. Some of these criteria require phylogenetic analysis of parts of new genomes. Phylogenetics also plays an important role here.

### *5.3.1. Nepovirus*

Nepoviruses are divided into two RNAs. This was also the case for our candidate nepovirus. Analysis were therefore divided into two parts: RNA1 analysis, which contains polymerase-protease part and RNA2 analysis, which contains the coat protein.

HG2 sequence (Heron grassland year 2) is different from the other sequences, on a single branch. This separation can be explained by presence of gaps in the nucleotide sequences that will vary the percentage of identity. However, percentages of similarities between all sequences remain close between 91.6% and 99.7% for the RNA1 and 84.3% and 97.6% for the RNA2 for the complete sequence.

It was also possible to observe that sequences from Antheit pasture have very high percentages of identities and are gathered in same areas in phylogenetic trees. It would, therefore, seem that sequences from this plant community are distinct from others. This similarity could be explained by the fact that the virus has implanted itself in this plant community and that the virus genome is specific to this plot. The analysis of SNPs of these genomes could provide additional informations on the similarity and differences between the different consensus sequences.

Same findings were possible with analysis of polymerase protease present in RNA1 and coat protein present in RNA2. Consensus sequences have very similar sequences in terms of identity percentages (91.8 to 99.8% for RNA1 and 91.6% to 99.7% for RNA2), and Antheit pasture sequences are also very similar in phylogenetic trees.

These close identity percentages show that there is, therefore, only one virus species in all plant communities studied. However, the percentage does not reach 100% because there may be SNPs between plant communities. It would be very interesting to study these SNPs. Genetic variability can also lead to differences in the genome without impacting protein translation. The study of proteins is, therefore, fundamental. This study will be discussed later.

Then, it is essential to do the same analyses but with reference sequences in different subgroups that constitute *Nepovirus* genus. Indeed, nepoviruses are divided into three subgroups, which are easily identifiable in the analysis of entire nucleotide sequences of RNA1. During the rest

of the analyses, subgroup A and C were not clustered during analyses. Only subgroup B is identified in figures. The fact that the sequences of subgroup A and C are mixed in the different phylogenetic trees was also observed during the analysis of the *Secoviridea* tree. There may be a separation that does not work well during the analysis. The percentages of identity are very close to the references. This may also explain this non-separation between the two subgroups.

Two trees analysing the consensus sequences of RNA1 and RNA2 both show that consensus sequences are in the same tree branch and all have very similar percentages of identity. This confirms the hypothesis that they are coming from the same species which belong to subgroup B. These observations were confirmed by analysis of polymerase-protease and coat protein-nucleic sequences.

Appearance of subgroup B is also confirmed by the length of sequences of two RNAs. Indeed, the consensus sequence RNA1 is 7,001 bp and RNA2 4,205 bp, which corresponds to the length of the Nepovirus subgroup B sequences (7356 for RNA1 and 4662 RNA2 [52]). For subgroup A, the lengths of the two RNAs are 7,342 for RNA1 and 3,774 for RNA2 (*Grapevine fanleaf virus*-F13), and for subgroup C the lengths of the two RNAs are 8,214 for RNA1 and 7,273 for RNA2 ( *Tomato ringspot virus*) [52]. However, it is essential to note that consensus sequences are not entirely complete and that a RACE PCR would have to be performed to obtain complete sequences.

Polymerase-protease and coat protein sequences were translated into proteins for further analysis, including the ICTV criteria for species demarcation. Less than 75% identity is required for coat proteins and less than 80% identity for a retained portion of polymerase protease. These criteria must not be met at the same time. It is essential to work with the CG-GDD sequence (+/- 450 bp) to perform polymerase protease analysis because this is a conserved portion of polymerase protease [52].

For polymerase-protease analysis, some sequences of subgroup B have very high percentages with the consensus sequences. NC_004439 virus (tomato black ring virus) has the highest percentage of identity with consensus sequences between 89.9% and 91.1%. NC_003693 virus (beet ringspot virus) has identity percentages with consensus sequences between 85.7 and 87.2%. These very high identity percentages, therefore, exclude the first ICTV criteria for a new virus species because identity percentages are higher than 80%. Identity percentages of protein sequences are higher than for nucleotide sequence analysis. This difference can easily be explained by the fact that several codons are capable of being translated into the same amino acid. The nucleotide sequences can, therefore diverge but give the same protein sequence. This is also why it is essential to study protein sequences to know if it is really a new species of virus that works with other proteins or a simple variation of the genome of the virus but with silent variations that produce the same proteins.

It is, therefore, essential to study the protein sequence of coat protein to examine another ICTV criterion. This analysis shows several elements. First of all, consensus sequences are well grouped together except AP2 *Lolium perenne* L. consensus sequence (Antheit pasture year 2) because of a significant gap in the sequence. Consensus sequences are also very close to subgroup B, but identity percentages are much lower than for polymerase protease. Tomato black ring virus has identity percentages from 47.1% to 57.2% and for beet ringspot virus from 56.1 to 66.9%. These percentages are lower than the maximum 75% identity proposed by the ICTV.

All these analyses do not confirm whether nepovirus candidate is really a new virus species or not. Indeed, the criteria of the identity of the coat protein is met but not that of the polymerase protease. A similar case has been studied by ITCV. Indeed, beet ringspot virus and tomato black ring virus have very high identity percentages (89%) for pol-pro but much lower for coat protein (62%) [52]. They are distinguished by antigenic reactions and nematode species. As candidate virus is very close to these two viruses, it is essential to continue the investigation of this virus to confirm hypothesis of a new virus. A more in-depth study of the genome of these three viruses could be interesting. Study of nematodes that are being carried out will also provide answers about the vector of this virus. This study will determine which nematodes transmit the virus. Study of seed transmission will also provide essential answers. It would also be interesting to do a study of antigens, which is also a criterion for ICTV. In addition, candidate virus has more or less the same percentages of identities as for these two references. So, the three viruses are very close.

### *5.3.2. Waikavirus*

The analyses were the same as for the new nepovirus candidate.

Fundamental bioinformatics analyses consist of studying consensus sequences of full genome, coat protein and pol-pro at nucleotidic and proteomic levels. These three phylogenetic trees have differences in branch distribution. Indeed, for complete sequences, it is plant communities Antheit pasture year 2 and Antheit grassland *Poa trivialis* L. year 2 that are separated from individual branches. Coat protein has a good sequence distribution with three main branches, and for polymerase, protease is the Antheit pasture year 2 and Antheit pasture year 1 plant community which has individual branches. In polymerase-protease study, consensus sequences of the pool category are grouped together in the same branch with the minor species pool category. These differences between trees can be explained by the fact that some consensus sequences have significant gaps in some places. The most infected sequences are AP1, HP2 Poa, AP2 and LP2 Poa. These gaps are all present during the analysis of the entire sequence but are sometimes absent during the analysis of coat protein or polymerase protease because the gaps are not always found in these areas. This explains the variations in phylogenetic trees.

Identity percentage matrices had shown a very high identity percentages: between 82.7% and 97.8%. These high percentages tend to show that there is only one new candidate waikavirus and different percentages can be explained by the fact that there are gaps in some sequences such as those explained above but also by genetic variations between genomes.

The rest of the analyses were carried out using consensus sequences and references from *Sequivirus* and *Waikavirus* genera database. The use of sequiviruses in addition to waikavirus references is justified by the fact that they are two very close genera and that it is essential to check to which genus new waikavirus candidate virus is the closest.

Analysis of the complete sequences, coat protein region, and polymerase-protease region were performed. The three trees clearly show 3 different zones, the sequivirus zone, the furthest from the consensus sequences, waikavirus zone and finally consensus sequences zone for our sequences. Identity percentage matrices confirm that waikaviruses are closer to consensus sequences than sequiviruses because they have higher identity percentages with waikaviruses than with sequiviruses. In all three cases, we observed three zones corresponding to waikavirus, sequivirus, and consensus sequence. Percentages of identity between consensus sequences themselves are identical to those studied in trees with only consensus sequences.

The analysis of these trees seems to confirm the hypothesis of rapprochement with the waikaviruses. The percentages of identity remain very low (32.8% to 45.3%) suggest that the sequences are very different and that these sequences can be those of a new virus species that have a connection with Waikavirus. The consensus sequences are similar to each other and are gathered in the tree; this fact confirms that there is only one new species that are identical in all plant communities with potentially genetic variations.

To confirm the different observations and hypotheses made with nucleotide sequences, it is essential to study the protein sequences of the coat protein and the CG-GDD sequence of the polymerase protease. Waikaviruses and sequiviruses are part of the *Secoviridea* family-like Nepoviruses. To identify a new virus, the same criteria are used for the analysis of protein sequences of the coat protein and the CG-GDD sequence of polymerase proteases. On the other hand, the coat protein, for waikaviruses, is divided into 3 parts. The three parts are used to perform the analyses.

The coat protein analyses has shown that three branches as for the analysis of nucleotide sequences. For consensus sequences, it is possible to observe two sequences that are further away than the other consensus sequences (AP2 and HP2 Poa). These two sequences are part of the sequences that have many gaps in the coat protein regions and can explain these differences between the consensus sequences. These gaps exist because it was impossible to correctly reconstruct the sequence despite the iterative tests during the mapping of the reads. Consensus sequences have a very low percentage of identity with consensus reference sequences, maximum 17.4%. This percentage is well below the 75% identity allowed for coat protein (ICTV criteria for species demarcation in *Secoviridea* family. This first element also confirms the hypothesis of a new virus.

Analysis of polymerase protease also has shown that three broad areas with sequiviruses, waikaviruses and consensus sequences. The sequences of waikaviruses and sequiviruses are closer than in the other analyses. This can be explained by the fact that the part studied is a very preserved part, and the two types of viruses can be very similar. The analysis also shows 2 separate sequences: LP2 Poa and AP2 Poa. These two sequences have only a maximum of 31.3% identity with the other consensus sequences. This is also the case for HP2 Poa but with a smaller difference (80% identity with the other sequences). This difference can also be explained by the presence of gaps in this area and therefore brings a bias in the analysis. The rest of the consensual sequences have a minimum of 95.3% identity between them. This supports the hypothesis that there is only one new species of *Waikavirus*. The percentage of identity with the other reference sequences is a maximum of 35.3%. It is also a percentage very far from the maximum 80% identity.

The criteria for distinguishing the different genera in the *Secoviridea* family are as follows: the number of coat proteins, the genomic RNAs number, the presence of additional ORFs, the number of protein domains and a new branch on the phylogenetic tree of the conserved area of the polymerase protease of the different virus species of the *Secoviridea* family [52]. The phylogenetic tree in Figure 63 does indeed show a new branch in the tree, but this branch is located between waikaviruses and sequiviruses. It would, therefore, be very interesting to study the other gender demarcation criteria to assess whether the consensus sequence can claim a new gender or not.

These various analyses have shown that the criteria for a new species are well met for waikavirus. It might be interesting to study the gender criteria to determine if it is a new genus or not.

A final analysis was performed with the sequence identified in each plant community. The sequence QA- STOP codon. This analysis was only performed with consensus sequences because references do not have this schema in their genome. The analysis confirmed the analysis of the CG-GDD sequences because the tree schema is the same and the identity percentages are very high minimum 96.628% except for the consensus sequences LP2 Poa, AP2 Poa, HP2 Poa which have significant gaps in the area. The fact that the sequence schema is not found in the references can be explained by the fact that the consensus sequences are derived from a new virus species.

All analyses performed on the consensus sequence of the new waikavirus candidate confirmed the hypothesis that this virus is new. It is closer to waikaviruses than sequiviruses, and the identity percentages between coat protein and polymerase protease protein sequences (CG-GDD) are very low and follow ICTV guidelines. The length of the consensus sequences is 10988 bp which is slightly below the ICTV standards (11700 bp), but the consensus sequences are not complete, the ends are missing. It would also be necessary to do a 3' and 5' RACE PCR to have the complete genomes because the two ends of genomes are missing.

### 5.3.3. *Secoviridea* tree ICTV

The phylogenetic tree available on the ITCV website for *Secoviridae* tree was reconstituted by injecting the consensus sequences of waikavirus and nepovirus (LG2 Poa and AP2). The program parameters were the same except for the number of bootstraps, which is 100 instead of 1000 for computer computing power reasons. However, the tree gives an idea of the tree that could be obtained with the 1000 bootstraps.

For waikavirus candidate, the findings are the same as for the individual analysis. A new branch appears between waikaviruses and sequiviruses. This further legitimizes the presence of both genera in the in-depth analysis of the waikavirus candidate. This new branch is very interesting because it is far from the other viruses and therefore confirms the hypothesis of this new species of virus. But as explained above, this new branch could also correspond to a new kind of virus. It will, therefore, be necessary to thoroughly analyze the genome of the virus to find out if a new genus has been found.

For nepovirus candidate, the analysis is more complicated; the branch is located between the Beet ringspot virus and the tomato black ring virus and very close to both viruses. As indicated by the precise analysis of the CG-GDD sequences of the nepovirus candidate, which also show significant similarities between these sequences. Therefore, they do not help to confirm the presence of a new virus species. Nepovirus needs further analysis like vectors and hosts analysis.

## 5.4. White clover mosaic virus

Analysis of the prevalence of white clover mosaic virus in ryegrass and white clover have shown that the prevalence is very high in this pasture (Heron pasture) because only one of the fifty individuals sampled is negative by RT-PCR. For *Lolium perenne* L. the prevalence is 46%. This virus is transmitted mechanically [73]. In this pasture, plant wounds caused by cows on the plants can explain the transmission of the virus to the ryegrass. It would be interesting to study further in depth the means of transmission of this virus.

Primers were developed using the consensus sequence obtained from the analysis of reads in Heron's pasture. The fact that PCR works show that it is this virus that is present. The bioinformatic analysis of the nucleotide sequence shows a percentage of 94.5% identity with the reference. It also shows that the virus has not had much adaptation to infect the ryegrass.

These analyses have shown that the ryegrass is actually infected with white clover mosaic virus and that therefore this virus has adapted and succeeded in finding a new host.

# 6. Conclusion & perspectives

For prevalence study within plots (global plant communities or plant species), results show a high prevalence of nepoviruses in the ryegrass and a high prevalence of waikaviruses in the *Poa trivialis* L. To develop further and confirm these figures, study of the other plots (Heron and Latinne) is needed. The results of these figures over several years of the harvest would also make it possible to know the behaviour of the viruses over time, to see if the infection rate is maintained or changes according to certain climatic conditions, the intensity of grazing, new vectors, etc. The results of the analyses for BYDV are also very interesting. The analysis of Heron and Latinne will allow us to confirm or qualify the results obtained for Antheit. It is also important to study the different BYDV viruses (PAS, PAV) that could have different impacts on the species present in the environments studied.

Nepovirus candidate is composed of two RNA sequences: the RNA1 which is composed of 7,356 bp and RNA 2 which is composed of 4,662 bp. These sequences are not complete, performing a RACE PCR on these sequences would be interesting to know the exact length of the genome. Nepovirus candidate belongs to the *Nepovirus* subgroup B. It has a high prevalence in the species of *Poaceae*: *Lolium perenne* L. but infects other species of *Poaceae*. This virus also tolerates co-infections with BYDV and waikavirus. Further studies of the virus phylogeny according to ICTV criteria show that the retained portion of polymerase protease (CG-GGD) has a high percentage of identity with beet ringspot and tomato black ring viruses up to 91.2% which is too high to meet ICTV criteria. The coat protein study also shows a high percentage with reference viruses but with only 66.9% identity, which is less than the maximum 75% set by the ICTV. It is therefore impossible to conclude whether or not this virus is a new species. New hosts of the virus can also be studied and can confirm that the Nepovirus studied is indeed a new virus species. The study of virus vectors, know the species of nematodes that transmits it and see if the transmission can be done by the seeds and antigenicity seems essential to confirm whether or not this nepovirus is a new *Nepovirus* species.

Waikavirus candidate is composed of a single long RNA sequence that measures 10,998 bp, but sequence is not complete, performing a RACE PCR, which consists in amplifying the ends of the sequence of interest to obtain the ends of the sequences, on this sequence would be interesting to know the exact length of the genome. It shows a high prevalence in *Poa trivialis* L., which is confirmed by a field study and bioinformatics analyses. It is capable of infecting other species of *Poaceae* and tolerates co-infections with BYDV and nepoviruses. Bioinformatics analyses show that this virus falls within two of the ICTV criteria to be a new virus species: identity percentage of less than 80% for the CG-GDD sequence and identity percentage of less than 75% for the coat protein. A sequence located on the polymerase protease sequence has been discovered QA-codon STOP. This sequence has good results in phylogenetic analysis. It may be interesting to further research this sequence to confirm that it is not in the references. On the ICTV tree, this virus has a new and distinct branch. It would, therefore, be

very interesting to study other criteria ,like new vectors or new hosts, such as hosts to confirm that this virus is indeed a new waikavirus species. Studies of the factors of transmission of this virus may be of interest to confirm whether the sequence is that of a new virus species. A more in-depth analysis of the genome of this sequence to also determine if this sequence is not a new genus of the virus because the branch of the *Secoviridea* family tree is well discarded by the Waikavirus.

The study of *White clover mosaic virus* shows that this virus has adapted well to *Lolium perenne* L. Mechanical transmission by pasture animals is the vector for transmission of the virus. It may be interesting to know if the areas infected by the virus are areas where animals are found more often than the healthy areas of the pasture. The phylogenetic study of this virus deserves further study to really know the percentages of identity between the consensus sequence and the reference to identify the sequences that may have been adapted to infect the ryegrass. There is a 94% identity percentage between the sequence found in *Lolium perenne* L. and the reference sequence, but it would be interesting to study the concerted areas of this virus and the protein sequences to refine the analysis.

To conclude this study, the nepovirus candidate is perhaps a new *Nepovirus* species within subgroup B, but further study (nematodes study and seed transmission) is needed to confirm, and the waikavirus candidate is a new species of the virus according to phylogenetic analyses performed as part of this work. The waikavirus candidate is a new species of the virus, but it will be interesting to study the *Secoviridea* demarcation criteria to determine if it is a new genus or just a new virus species.

# References


[1] P.-Y. Teycheney, “Analyse, fonction et diversité des génomes viraux des plantes et élaboration de stratégies de lutte,” 2008.

[2] H. Lapierre, P. A. Signoret, and Institut national de la recherche agronomique (France), *Viruses and virus diseases of Poaceae (Gramineae)*. Institut national de la recherche agronomique, 2004.

[3] Christopher S. Campbell, “Poaceae - Economic and ecological importance | Britannica.com,” 2008. [Online]. Available: https://www.britannica.com/plant/Poaceae/Economic-and-ecological-importance. [Accessed: 13-Jul-2019].

[4] B. D. HARRISON, “Plant virus ecology: ingredients, interactions and environmental influences,” *Ann. Appl. Biol.*, vol. 99, no. 3, pp. 195–209, Dec. 1981.

[5] P. Bernardo *et al.*, “Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale,” *ISME J.*, vol. 12, no. 1, pp. 173–184, Jan. 2018.

[6] F. Garcia-Arenal, A. Fraile, and J. M. Malpica, “Variation and evolution of plant virus populations,” *Int. Microbiol.*, vol. 6, no. 4, pp. 225–232, Dec. 2003.

[7] R. Hull and J. I. Centre, *Plant virology. Fifth edition.*, ElsevierInc. USA, 2014.

[8] Claude M.Fauquet and John R. Schrock, *Kansas school naturalist.* Kansas School Naturalist, 2006.

[9] B. D. Harrison *et al.*, “Sixteen groups of plant viruses,” *Virology*, vol. 45, no. 2, pp. 356–363, Aug. 1971.

[10] M. J. Roossinck, “Plant Virus Metagenomics: Biodiversity and Ecology,” *Annu. Rev. Genet.*, vol. 46, no. 1, pp. 359–369, Dec. 2012.

[11] International Committee on Taxonomy of Viruses. and A. King, *Virus taxonomy : ninth report of the International Committee on Taxonomy of Viruses*. Elsevier, 2011.

[12] R. C. Gergerich and V. V. Dolja, “Introduction to Plant Viruses, the Invisible Foe.,” *Plant Heal. Instr.*, 2006.

[13] B. Moury, C. Desbiez, M. Jacquemond, and H. Lecoq, “Genetic Diversity of Plant Virus Populations: Towards Hypothesis Testing in Molecular Epidemiology,” *Adv. Virus Res.*, vol. 67, pp. 49–87, Jan. 2006.

[14] M. A. WATSON and F. M. ROBERTS, “A comparative Study of the Transmission of *Hyoscyamus* Virus 3, Potato Virus Y and Cucumber Virus 1 by the Vectors *Myzus persicae* (Sulz.), *M. circumflexus* (Buekton), and *Macrosiphum gei* (Koch).,” *Proc. R. Soc.*, vol. 127, no. 849, 1939.

[15] M. Day and H. Irzykiewicz, “On the Mechanism of Transmission of Non-Persistent Phytopathogenic Viruses by Aphids,” *Aust. J. Biol. Sci.*, vol. 7, no. 3, p. 251, 1954.

[16] F. Escriu, A. Fraile, and F. García-Arenal, “THE EVOLUTION OF VIRULENCE IN A PLANT VIRUS,” *Evolution (N. Y).*, vol. 57, no. 4, pp. 755–765, Apr. 2003.

[17] C. E. Taylor and D. J. F. Brown, “Nematode vectors of plant viruses.,” *Nematode vectors of plant viruses.*, 1997.

[18] J. C. K. NG and K. L. PERRY, “Transmission of plant viruses by aphid vectors,” *Mol. Plant Pathol.*, vol. 5, no. 5, pp. 505–511, Sep. 2004.

[19] N. Crowley, “Studies on the Seed Transmission of Plant Virus Diseases,” *Aust. J. Biol. Sci.*, vol. 10, no. 4, p. 449, 1957.

[20] R. M. LISTER and A. F. MURANT, “Seed-transmission of nematode-borne viruses,” *Ann. Appl. Biol.*, vol. 59, no. 1, pp. 49–62, Feb. 1967.

[21] R. N. Campbell, “FUNGAL TRANSMISSION OF PLANT VIRUSES,” *Annu. Rev. Phytopathol.*, vol. 34, no. 1, pp. 87–108, Sep. 1996.

[22]S. Jarugula, S. R. Charlesworth, F. Qu, and L. R. Stewart, “Soil-borne wheat mosaic virus infectious clone and manipulation for gene-carrying capacity,” *Arch. Virol.*, vol. 161, no. 8, pp. 2291–2297, Aug. 2016.
[23]Kenneth Daniels, “Molecular Biology of Prokaryotes - ppt download,” 2018. [Online]. Available: https://slideplayer.com/slide/13067788/. [Accessed: 12-Jul-2019].
[24]I. Pagán, N. Montes, M. G. Milgroom, and F. García-Arenal, “Vertical Transmission Selects for Reduced Virulence in a Plant Virus and for Increased Resistance in the Host,” *PLoS Pathog.*, vol. 10, no. 7, p. e1004293, Jul. 2014.
[25]C. M. Malmstrom, U. Melcher, and N. A. Bosque-Pérez, “The expanding field of plant virus ecology: Historical foundations, knowledge gaps, and research directions,” *Virus Res.*, vol. 159, no. 2, pp. 84–94, Aug. 2011.
[26]H. R. Prendeville, X. Ye, T. Jack Morris, and D. Pilson, “Virus infections in wild plant populations are both frequent and often unapparent,” *Am. J. Bot.*, vol. 99, no. 6, pp. 1033–1042, Jun. 2012.
[27]I. Cooper and R. A. C. Jones, “Wild Plants and Viruses: Under-Investigated Ecosystems,” *Adv. Virus Res.*, vol. 67, pp. 1–47, Jan. 2006.
[28]A. E. C. Jooste *et al.*, “Identification and distribution of multiple virus infections in Grapevine leafroll diseased vineyards,” *Eur. J. Plant Pathol.*, vol. 142, no. 2, pp. 363–375, Jun. 2015.
[29]M. J. Roossinck, Ed., Plant Virus Evolution. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.
[30]M. J. Roossinck, “Plant Virus Ecology,” *PLoS Pathog.*, vol. 9, no. 5, p. e1003304, May 2013.
[31]R. M. LISTER and A. F. MURANT, “Seed-transmission of nematode-borne viruses,” *Ann. Appl. Biol.*, vol. 59, no. 1, pp. 49–62, Feb. 1967.
[32]A. D. Stewart, J. M. Logsdon, and S. E. Kelley, “AN EMPIRICAL STUDY OF THE EVOLUTION OF VIRULENCE UNDER BOTH HORIZONTAL AND VERTICAL TRANSMISSION,” *Evolution (N. Y).*, vol. 59, no. 4, pp. 730–739, Apr. 2005.
[33]M. J. Roossinck, “MECHANISMS OF PLANTVIRUS EVOLUTION,” *Annu. Rev. Phytopathol.*, vol. 35, no. 1, pp. 191–209, Sep. 1997.
[34]F. A. Murphy *et al.*, *Virus Taxonomy : Classification and Nomenclature of Viruses Sixth Report of the International Committee on Taxonomy of Viruses*. Springer Vienna, 1995.
[35]P. Simmonds *et al.*, “Virus taxonomy in the age of metagenomics,” *Nat. Rev. Microbiol.*, vol. 15, no. 3, pp. 161–168, Mar. 2017.
[36]Bennett B., “PLANTS AS FOOD,” Miami, 2019.
[37]Christopher S. Campbell, “Poaceae - Economic and ecological importance | Britannica.com,” 2008. [Online]. Available: https://www.britannica.com/plant/Poaceae/Economic-and-ecological-importance. [Accessed: 13-Jul-2019].
[38]SIB Swiss Institute of Bioinformatics, “+RNA virus replication/transcription ~ ViralZone page,” 2007. [Online]. Available: https://viralzone.expasy.org/1116. [Accessed: 13-Jul-2019].
[39]Khan academy, “Intro to viruses (article) | Khan Academy,” 2016. [Online]. Available: https://www.khanacademy.org/science/high-school-biology/hs-human-body-systems/hs-the-immune-system/a/intro-to-viruses. [Accessed: 13-Jul-2019].
[40]L. R. Nault, “Arthropod Transmission of Plant Viruses: a New Synthesis,” *Ann. Entomol. Soc. Am.*, vol. 90, no. 5, pp. 521–541, Sep. 1997.
[41]K. F. Harris, “AN INGESTION-EGESTION HYPOTHESIS OF NONCIRCULATIVE VIRUS TRANSMISSION,” *Aphids As Virus Vectors*, pp. 165–220, Jan. 1977.
[42]R. Dietzgen, K. Mann, K. Johnson, R. G. Dietzgen, K. S. Mann, and K. N. Johnson, “Plant Virus–Insect Vector Interactions: Current and Potential Future Research Directions,” *Viruses*, vol. 8, no. 11, p. 303, Nov. 2016.

[43]E. D. Ammar, “Propagative Transmission of Plant and Animal Viruses by Insects: Factors Affecting Vector Specificity and Competence,” Springer, New York, NY, 1994, pp. 289–331.
[44]B. Weischer and D. J. F. Brown, *An introduction to nematodes : general nematology : a student’s textbook*. Pensoft, 2000.
[45]C. W. Bennett, “Seed Transmission of Plant Viruses,” *Adv. Virus Res.*, vol. 14, pp. 221–261, Jan. 1969.
[46]K. S. Sastry, “Mechanism of Seed Transmission,” in *Seed-borne plant virus diseases*, India: Springer India, 2013, pp. 85–100.
[47]C. Tomlin and British Crop Protection Council., *The pesticide manual : a world compendium.* British Crop Protection Council, 2000.
[48]R. T. Roush and B. E. Tabashnik, *Pesticide Resistance in Arthropods*. Springer US, 1991.
[49]SIB Swiss Institute of Bioinformatics, “Nepovirus ~ ViralZone page,” 2014. [Online]. Available: https://viralzone.expasy.org/300. [Accessed: 13-Jul-2019].
[50]J. E. Oliver, E. Vigne, and M. Fuchs, “Genetic structure and molecular variability of Grapevine fanleaf virus populations,” *Virus Res.*, vol. 152, no. 1–2, pp. 30–40, Sep. 2010.
[51]Reference*, “What Is a Non-Enveloped Virus? | Reference.com,” 2017. [Online]. Available: https://www.reference.com/health/non-enveloped-virus-a06b69c9c0f2962c. [Accessed: 29-Jul-2019].
[52]J. R. Thompson *et al.*, “ICTV Virus Taxonomy Profile: Secoviridae,” *J. Gen. Virol.*, vol. 98, no. 4, pp. 529–531, Apr. 2017.
[53]B. or. Virginia Tech Learning Resources Center, Virginia Polytechnic Institute and State University, “tobacco ringspot virus (TRSV) (Nepovirus TRSV ) on tobacco (Nicotiana spp. ) - 5332064,” 2018. [Online]. Available: https://www.forestryimages.org/browse/detail.cfm?imgnum=5332064. [Accessed: 13-Jul-2019].
[54]International Committee on Taxonomy of Viruses, “Genus: Nepovirus - Secoviridae - Picornavirales - International Committee on Taxonomy of Viruses (ICTV),” 2018. [Online]. Available: https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/picornavirales/w/secoviridae/591/genus-nepovirus. [Accessed: 13-Jul-2019].
[55]Rothansted Research, “Notes on Genus : Nepovirus,” 2007. [Online]. Available: http://www.dpvweb.net/notes/showgenus.php?genus=Nepovirus. [Accessed: 13-Jul-2019].
[56]J. Dijkstra and C. P. de Jager, “Virus Transmission by Nematodes,” in *Practical Plant Virology*, Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 128–142.
[57]R. M. LISTER and A. F. MURANT, “Seed-transmission of nematode-borne viruses,” *Ann. Appl. Biol.*, vol. 59, no. 1, pp. 49–62, Feb. 1967.
[58]P. Susi, “Black currant reversion virus, a mite-transmitted nepovirus,” *Mol. Plant Pathol.*, vol. 5, no. 3, pp. 167–173, May 2004.
[59]M. Fuchs, C. Schmitt-Keichinger, and H. Sanfaçon, “A Renaissance in Nepovirus Research Provides New Insights Into Their Molecular Interface With Hosts and Vectors,” *Adv. Virus Res.*, vol. 97, pp. 61–105, Jan. 2017.
[60]M. Digiaro, E. Yahyaoui, G. P. Martelli, and T. Elbeaino, “The sequencing of the complete genome of a Tomato black ring virus (TBRV) and of the RNA2 of three Grapevine chrome mosaic virus (GCMV) isolates from grapevine reveals the possible recombinant origin of GCMV,” *Virus Genes*, vol. 50, no. 1, pp. 165–171, Feb. 2015.
[61]E. Vigne, A. Marmonier, and M. Fuchs, “Multiple interspecies recombination events within RNA2 of Grapevine fanleaf virus and Arabis mosaic virus,” *Arch. Virol.*, vol. 153, no. 9, pp. 1771–1776, Sep. 2008.

[62]M. Walker *et al.*, “Complete genome sequence of three tomato ringspot virus isolates: evidence for reassortment and recombination,” *Arch. Virol.*, vol. 160, no. 2, pp. 543–547, Feb. 2015.
[63]H. Sanfaçon, J. Wellink, O. Le Gall, A. Karasev, R. van der Vlugt, and T. Wetzel, “Secoviridae: a proposed family of plant viruses within the order Picornavirales that combines the families Sequiviridae and Comoviridae, the unassigned genera Cheravirus and Sadwavirus, and the proposed genus Torradovirus,” *Arch. Virol.*, vol. 154, no. 5, pp. 899–907, May 2009.
[64]SIB Swiss Institute of Bioinformatics, “Waikavirus ~ ViralZone page,” 2014. [Online]. Available: https://viralzone.expasy.org/661?outline=all_by_species. [Accessed: 13-Jul-2019].
[65]P. Tennant *et al.*, “Viruses as Pathogens: Plant Viruses,” *Viruses*, pp. 135–156, Jan. 2018.
[66]L. R. S. and P. A. P. Jose L. Zambrano, “Maize Chlorotic Dwarf of Maize | Ohioline,” 2015. [Online]. Available: https://ohioline.osu.edu/factsheet/plpath-cer-08. [Accessed: 13-Jul-2019].
[67]International Committee on Taxonomy of Viruses, “Genus: Waikavirus - Secoviridae - Picornavirales - International Committee on Taxonomy of Viruses (ICTV),” 2018. [Online]. Available: https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/picornavirales/w/secoviridae/587/genus-waikavirus. [Accessed: 13-Jul-2019].
[68]SIB Swiss Institute of Bioinformatics, “Potexvirus ~ ViralZone page,” 2014. [Online]. Available: https://viralzone.expasy.org/272?outline=all_by_species. [Accessed: 14-Jul-2019].
[69]R. L. Forster, M. W. Bevan, S. A. Harbison, and R. C. Gardner, “The complete nucleotide sequence of the potexvirus white clover mosaic virus.,” *Nucleic Acids Res.*, vol. 16, no. 1, pp. 291–303, Jan. 1988.
[70]International Committee on Taxonomy of Virus, “Alphaflexiviridae - Positive Sense RNA Viruses - Positive Sense RNA Viruses (2011) - International Committee on Taxonomy of Viruses (ICTV),” 2018. [Online]. Available: https://talk.ictvonline.org/ictv-reports/ictv_9th_report/positive-sense-rna-viruses-2011/w/posrna_viruses/239/alphaflexiviridae. [Accessed: 14-Jul-2019].
[71]Bercks R., “White clover mosaic virus” 2007. [Online]. Available: http://www.dpvweb.net/dpv/showdpv.php?dpvno=041. [Accessed: 10-Jul-2019].
[72]NGBI - NARO National Agriculture and food Research Organization, “Diseases of Forage Crops,” 2006. [Online]. Available: http://www.naro.affrc.go.jp/org/nilgs/diseases/contents/de34.htm. [Accessed: 14-Jul-2019].
[73]NARO National Agriculture and food Research Organization, “Diseases of Forage Crops,” 2006. [Online]. Available: http://www.naro.affrc.go.jp/org/nilgs/diseases/contents/de34.htm. [Accessed: 14-Jul-2019].
[74]Hampton, R. O. "Seed transmission of white clover mosaic and clover yellow mosaic viruses in red clover." *Phytopathology*. Vol. 53. No. 10. 3340 PILOT KNOB ROAD, ST PAUL, MN 55121: AMER PHYTOPATHOLOGICAL SOC, 1963.
[75]Caradus, J. R., D. R. Woodfield, and A. V. Stewart. "Overview and vision for white clover." Special Publication-Agronomy Society of New Zealand (1996): 1-6.
[76]B. Dudas , D. R. Woodfield , P. M. Tong , M. F. Nicholls , G. R. Cousins ,R. Burgess , D. W. R. White , D. L. Beck , T. J. Lough & R. L. S. Forster “Estimating the agronomic impact of white clover mosaic virus on white clover performance in the NorthIsland of New Zealand”, New Zealand Journal of Agricultural Research, 41:2, 171-178, 1998
[77]C. Y. Park, S.-H. Lee, S. Lim, J. S. Moon, and B.-S. Kim, “First Report of *White clover mosaic virus* on White Clover ( *Trifolium repens* ) in Korea,” *Plant Dis.*, vol. 101, no. 8, pp. 1559–1559, Aug. 2017.

[78]Cambridge University, “SEQUENCING | meaning in the Cambridge English Dictionary,” 2017. [Online]. Available: https://dictionary.cambridge.org/dictionary/english/sequencing. [Accessed: 14-Jul-2019].
[79]F. Sanger, S. Nicklen, and A. R. Coulson, “DNA sequencing with chain-terminating inhibitors,” *Proc. Natl. Acad. Sci.*, vol. 74, no. 12, pp. 5463–5467, Dec. 1977.
[80]bnif snipcademy, “454 Roche Pyrosequencing,” 2017. [Online]. Available: https://binf.snipcademy.com/lessons/ngs-techniques/454-roche-pyrosequencing. [Accessed: 20-Jul-2019].
[81]J. M. Heather and B. Chain, “The sequence of sequencers: The history of sequencing DNA,” *Genomics*, vol. 107, no. 1, pp. 1–8, Jan. 2016.
[82]Brown Stuart M., “Sequencing-by-Synthesis: Explaining the Illumina Sequencing Technology - Bitesize Bio,” 2016. [Online]. Available: https://bitesizebio.com/13546/sequencing-by-synthesis-explaining-the-illumina-sequencing-technology/. [Accessed: 20-Jul-2019].
[83]F. Montel, “Séquençage de l’ADN par nanopores,” *médecine/sciences*, vol. 34, no. 2, pp. 161–165, Feb. 2018.
[84]C. S. Pareek, R. Smoczynski, and A. Tretyn, “Sequencing technologies and genome sequencing,” *J. Appl. Genet.*, vol. 52, no. 4, pp. 413–435, Nov. 2011.
[85]R. Ekblom and J. B. W. Wolf, “A field guide to whole-genome sequencing, assembly and annotation.,” *Evol. Appl.*, vol. 7, no. 9, pp. 1026–42, Nov. 2014.
[86]M. Barba, H. Czosnek, A. Hadidi, M. Barba, H. Czosnek, and A. Hadidi, “Historical Perspective, Development and Applications of Next-Generation Sequencing in Plant Virology,” *Viruses*, vol. 6, no. 1, pp. 106–136, Jan. 2014.
[87]S. Massart, A. Olmos, H. Jijakli, and T. Candresse, “Current impact and future directions of high throughput sequencing in plant virus diagnostics,” *Virus Res.*, vol. 188, pp. 90–96, Aug. 2014.
[88]S. Pabinger *et al.*, “A survey of tools for variant analysis of next-generation genome sequencing data,” *Brief. Bioinform.*, vol. 15, no. 2, pp. 256–278, Mar. 2014.
[89]R. Bao *et al.*, “Review of Current Methods, Applications, and Data Management for the Bioinformatics Analysis of Whole Exome Sequencing,” *Cancer Inform.*, vol. 13s2, p. CIN.S13779, Jan. 2014.
[90]G. Robertson *et al.*, “De novo assembly and analysis of RNA-seq data,” *Nat. Methods*, vol. 7, no. 11, pp. 909–912, Nov. 2010.
[91]A. Bankevich *et al.*, “SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.,” *J. Comput. Biol.*, vol. 19, no. 5, pp. 455–77, May 2012.
[92]D. R. Zerbino, “Using the Velvet de novo assembler for short-read sequencing technologies.,” *Curr. Protoc. Bioinforma.*, vol. Chapter 11, p. Unit 11.5, Sep. 2010.
[93]Gladman S., “De Novo Genome Assembly for Illumina Data - Bioinformatics Documentation,” 2017. [Online]. Available: https://www.melbournebioinformatics.org.au/tutorials/tutorials/assembly/assembly-protocol/. [Accessed: 20-Jul-2019].
[94]M. Ruffalo, T. LaFramboise, and M. Koyuturk, “Comparative analysis of algorithms for next-generation sequencing read alignment,” *Bioinformatics*, vol. 27, no. 20, pp. 2790–2796, Oct. 2011.
[95]NCBI National Center for Biotechnology Information, “BLAST: Basic Local Alignment Search Tool.” [Online]. Available: https://blast.ncbi.nlm.nih.gov/Blast.cgi. [Accessed: 20-Jul-2019].
[96]K.Janton G. gREUB, “PCR en microbiologie : de l’amplification de l’ADN à l’interprétation du résultat - Revue Médicale Suisse,” 2007. [Online]. Available: https://www.revmed.ch/RMS/2007/RMS-106/32181. [Accessed: 14-Jul-2019].

[97]M. Carr Steven, “Principle of RT-PCR,” 2005. [Online]. Available: https://www.mun.ca/biology/scarr/Principle_of_RT-PCR.html. [Accessed: 08-Aug-2019].
[98]D. W. Mount, *Bioinformatics : sequence and genome analysis*. Cold Spring Harbor Laboratory Press, 2004.
[99]Sultan Haque Omar, “Phylogenetics | Britannica.com,” 2016. [Online]. Available: https://www.britannica.com/science/phylogenetics. [Accessed: 11-Aug-2019].
[100] Khan Academy, “Phylogenetic trees | Evolutionary tree (article) | Khan Academy,” 2017. [Online]. Available: https://www.khanacademy.org/science/high-school-biology/hs-evolution/hs-phylogeny/a/phylogenetic-trees. [Accessed: 14-Jul-2019].
[101] D.-F. Feng and R. F. Doolittle, “[23] Progressive alignment and phylogenetic tree construction of protein sequences,” *Methods Enzymol.*, vol. 183, pp. 375–387, Jan. 1990.
[102] J. Felsenstein, “CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP,” *Evolution (N. Y).*, vol. 39, no. 4, pp. 783–791, Jul. 1985.
[103] S. Guindon, F. Lethiec, P. Duroux, and O. Gascuel, “PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference,” *Nucleic Acids Res.*, vol. 33, no. Web Server, pp. W557–W559, Jul. 2005.
[104] Zageno, “A brief guide to the different methods of phylogenetic tree construction | ZAGENO,” 2019. [Online]. Available: https://zageno.com/l/guide-to-different-methods-of-phylogenetic-tree-construction?utm_source=facebook&utm_medium=article&utm_campaign=r_phylogenetic_tree_construction. [Accessed: 17-Jul-2019].
[105] “Antheit,” 1999. [Online]. Available: https://fr.wikipedia.org/wiki/Antheit. [Accessed: 11-Aug-2019].
[106] Celetti Michael, “Sampling Soil and Roots for Plant Parasitic Nematodes,” 2012. [Online]. Available: http://www.omafra.gov.on.ca/english/crops/facts/06-099.htm?fbclid=IwAR1Ia6mEizFlwlDIkayTA_W7lrFRXNxB5_e4ypLVR9qlCgpIAPa30-Np-Pk. [Accessed: 12-Aug-2019].
[107] “Héron (commune),” 2000. [Online]. Available: https://fr.wikipedia.org/wiki/Héron_%28commune%29. [Accessed: 11-Aug-2019].
[108] L. Oñate-Sánchez and J. Vicente-Carbajosa, “DNA-free RNA isolation protocols for Arabidopsis thaliana, including seeds and siliques,” *BMC Res. Notes*, vol. 1, no. 1, p. 93, Oct. 2008.
[109] Thermo Scientific, “T042-TECHNICAL BULLETIN NanoDrop Spectrophotometers, 260/280 and 260/230 Ratios,” 2010. [Online]. Available: http://www.nanodrop.com/Library/T042-NanoDrop-Spectrophotometers-Nucleic-Acid-Purity-Ratios.pdf. [Accessed: 20-Jul-2019].
[110] Baurain D., “PhytoSYSTEMS - Calcul intensif,” 2013. [Online]. Available: http://www.phytosystems.ulg.ac.be/fr/infrastructures/calcul-intensif. [Accessed: 20-Jul-2019].
[111] G. C. M. Latch, “Incidence of barley yellow dwarf virus in ryegrass pastures in New Zealand,” *New Zeal. J. Agric. Res.*, vol. 20, no. 1, pp. 87–89, Feb. 1977.
[112] K. A. Garrett, S. P. Dendy, A. G. Power, G. K. Blaisdell, H. M. Alexander, and J. K. McCarron, “Barley Yellow Dwarf Disease in Natural Populations of Dominant Tallgrass Prairie Species in Kansas,” *Plant Dis.*, vol. 88, no. 5, pp. 574–574, May 2004.
[113] J. SYLLER, “Facilitative and antagonistic interactions between plant viruses in mixed infections,” *Mol. Plant Pathol.*, vol. 13, no. 2, pp. 204–216, Feb. 2012.

# Appendix

## Appendix 1

*Table 18 : Matrix of consensus sequence nepovirus RNA1*

| | Consensus H... | Consensus A... | Consensus A... | Consensus H... | Consensus A... | Consensus A... | Consensus A... | Consensus H... | Consensus A... | Consensus A... | Consensus A... | Consensus A... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Consensus HG2(1) ARN1 ex... | | 92.964% | 92.442% | 91.686% | 91.840% | 91.741% | 91.748% | 92.054% | 91.962% | 91.641% | 92.375% | 92.686% |
| Consensus AP2 Poa ARN1 e... | 92.964% | | 97.777% | 95.668% | 96.277% | 96.587% | 96.625% | 96.633% | 96.936% | 94.174% | 95.880% | 95.004% |
| Consensus AP1 Lolium ARN... | 92.442% | 97.777% | | 95.901% | 96.528% | 96.857% | 96.895% | 96.426% | 97.300% | 94.625% | 96.392% | 95.410% |
| Consensus HP1 ARN1 | 91.686% | 95.668% | 95.901% | | 97.019% | 97.035% | 97.103% | 96.625% | 97.277% | 92.032% | 93.862% | 92.941% |
| Consensus AG1 Lolium extr... | 91.840% | 96.277% | 96.528% | 97.019% | | 97.435% | 97.488% | 98.570% | 98.335% | 92.244% | 94.386% | 93.176% |
| Consensus AG2 Agrostis sv ... | 91.741% | 96.587% | 96.857% | 97.035% | 97.435% | | 99.720% | 97.231% | 98.305% | 92.275% | 94.265% | 93.346% |
| Consensus AG1 ARN1 extra... | 91.748% | 96.625% | 96.895% | 97.103% | 97.488% | 99.720% | | 97.299% | 98.358% | 92.282% | 94.272% | 93.354% |
| Consensus HP2 (55) extraction | 92.054% | 96.633% | 96.426% | 96.625% | 98.570% | 97.231% | 97.299% | | 98.396% | 92.100% | 94.302% | 93.203% |
| Consensus AP2 Agrostis ext... | 91.962% | 96.936% | 97.300% | 97.277% | 98.335% | 98.305% | 98.358% | 98.396% | | 92.691% | 95.138% | 93.687% |
| Consensus AP2 ARN1 extrac... | 91.641% | 94.174% | 94.625% | 92.032% | 92.244% | 92.275% | 92.282% | 92.100% | 92.691% | | 97.132% | 96.789% |
| Consensus AP2 Lolium ARN... | 92.375% | 95.880% | 96.392% | 93.862% | 94.386% | 94.265% | 94.272% | 94.302% | 95.138% | 97.132% | | 97.134% |
| Consensus AP1 ARN1 extrac... | 92.686% | 95.004% | 95.410% | 92.941% | 93.176% | 93.346% | 93.354% | 93.203% | 93.687% | 96.789% | 97.134% | |

## Appendix 2

*Table 19 : Matrix of consensus sequence nepovirus RNA2*

| | Consensus 1 ... | Consensus 2 ... | Consensus 1 ... | Consensus 1 ... | Consensus 1 ... | Consensus 1 ... | Consensus 1 ... | Consensus 1 ... | Consensus 1 ... | Consensus 1 ... | Consensus 1 ... | Consensus 1 ... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Consensus 1 AG1 Lolium Ne... | | 84.274% | 92.787% | 90.918% | 89.247% | 91.973% | 94.209% | 92.328% | 92.152% | 94.054% | 93.387% | 93.766% |
| Consensus 2 HG2(1) Nepo2... | 84.274% | | 86.150% | 86.638% | 87.462% | 86.511% | 86.178% | 87.318% | 87.712% | 87.178% | 87.742% | 88.363% |
| Consensus 1 HP1 Nepo2 ex... | 92.787% | 86.150% | | 92.287% | 91.442% | 93.420% | 95.384% | 93.978% | 94.651% | 95.146% | 95.468% | 95.590% |
| Consensus 1 AP2 Lolium Ne... | 90.918% | 86.638% | 92.287% | | 95.867% | 92.199% | 92.793% | 94.592% | 95.412% | 94.242% | 94.126% | 95.159% |
| Consensus 1 AP2 Nepo2 ex... | 89.247% | 87.462% | 91.442% | 95.867% | | 91.313% | 91.746% | 94.361% | 95.095% | 92.857% | 93.100% | 94.350% |
| Consensus 1 AG2 Agrostis S... | 91.973% | 86.511% | 93.420% | 92.199% | 91.313% | | 93.937% | 94.596% | 94.288% | 95.219% | 97.610% | 95.718% |
| Consensus 1 HP2 Nepo2 ex... | 94.209% | 86.178% | 95.384% | 92.793% | 91.746% | 93.937% | | 94.429% | 95.089% | 96.685% | 95.993% | 96.234% |
| Consensus 1 AP1 Lolium Ne... | 92.328% | 87.318% | 93.978% | 94.592% | 94.361% | 94.596% | 94.429% | | 96.371% | 95.785% | 96.201% | 96.538% |
| Consensus 1 AP1 Nepo2 ext... | 92.152% | 87.712% | 94.651% | 95.412% | 95.095% | 94.288% | 95.089% | 96.371% | | 96.169% | 96.452% | 97.193% |
| Consensus 1 AP2 Agrostis N... | 94.054% | 87.178% | 95.146% | 94.242% | 92.857% | 95.219% | 96.685% | 95.785% | 96.169% | | 97.331% | 97.668% |
| Consensus 1 AG1 Nepo2 ex... | 93.387% | 87.742% | 95.468% | 94.126% | 93.100% | 97.610% | 95.993% | 96.201% | 96.452% | 97.331% | | 97.714% |
| Consensus 1 AP2 Poa Nepo... | 93.766% | 88.363% | 95.590% | 95.159% | 94.350% | 95.718% | 96.234% | 96.538% | 97.193% | 97.668% | 97.714% | |

## Appendix 3

*Table 20 : Matrix of consensus sequences nepovirus RNA1 (Polymerase-protease)*

| | Consensus H... | Consensus A... | Consensus A... | Consensus H... | Consensus A... | Consensus A... | Consensus A... | Consensus A... | Consensus H... | Consensus A... | Consensus A... | Consensus A... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Consensus HG2(1) ARN1 ex... | | 94.191% | 94.303% | 93.595% | 93.547% | 93.579% | 93.515% | 93.434% | 93.482% | 92.742% | 93.692% | 93.724% |
| Consensus AP1 Lolium ARN... | 94.191% | | 98.238% | 95.333% | 95.864% | 95.977% | 95.912% | 96.186% | 95.832% | 95.060% | 96.862% | 96.403% |
| Consensus AP2 Poa ARN1 e... | 94.303% | 98.238% | | 96.524% | 96.749% | 97.055% | 96.991% | 97.023% | 96.846% | 94.062% | 95.792% | 95.639% |
| Consensus HP1 ARN1 extra... | 93.595% | 95.333% | 96.524% | | 97.683% | 97.650% | 97.522% | 97.441% | 97.345% | 91.793% | 93.659% | 93.482% |
| Consensus AG1 Lolium extr... | 93.547% | 95.864% | 96.749% | 97.683% | | 98.101% | 97.972% | 98.648% | 98.906% | 92.115% | 94.271% | 93.740% |
| Consensus AG1 ARN1 extra... | 93.579% | 95.977% | 97.055% | 97.650% | 98.101% | | 99.775% | 98.149% | 97.988% | 91.857% | 93.917% | 93.482% |
| Consensus AG2 Agrostis sv ... | 93.515% | 95.912% | 96.991% | 97.522% | 97.972% | 99.775% | | 98.021% | 97.860% | 91.793% | 93.853% | 93.418% |
| Consensus AP2 Agrostis ext... | 93.434% | 96.186% | 97.023% | 97.441% | 98.648% | 98.149% | 98.021% | | 99.115% | 92.082% | 94.714% | 93.708% |
| Consensus HP2 (55) extract... | 93.482% | 95.832% | 96.846% | 97.345% | 98.906% | 97.988% | 97.860% | 99.115% | | 91.825% | 94.319% | 93.515% |
| Consensus AP2 ARN1 extrac... | 92.742% | 95.060% | 94.062% | 91.793% | 92.115% | 91.857% | 91.793% | 92.082% | 91.825% | | 97.119% | 97.441% |
| Consensus AP2 Lolium ARN... | 93.692% | 96.862% | 95.792% | 93.659% | 94.271% | 93.917% | 93.853% | 94.714% | 94.319% | 97.119% | | 97.876% |
| Consensus AP1 ARN1 extrac... | 93.724% | 96.403% | 95.639% | 93.482% | 93.740% | 93.482% | 93.418% | 93.708% | 93.515% | 97.441% | 97.876% | |

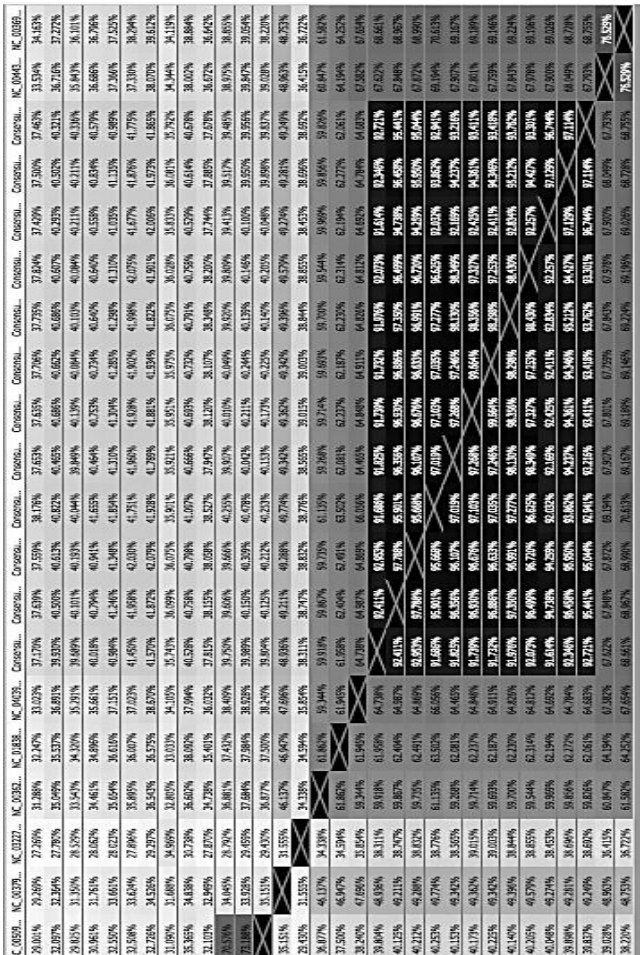
## Appendix 4

*Table 21 : Matrix of consensus sequences nepovirus RNA2 (Coat protein)*

| | Consensus H... | Consensus A... | Consensus A... | Consensus H... | Consensus A... | Consensus A... | Consensus A... | Consensus A... | Consensus H... | Consensus A... | Consensus A... | Consensus A... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Consensus HG2(1) ARN1 ex... | | 92.442% | 92.964% | 91.686% | 91.840% | 91.748% | 91.741% | 91.962% | 92.054% | 91.641% | 92.375% | 92.686% |
| Consensus AP1 Lolium ARN... | 92.442% | | 97.777% | 95.901% | 96.528% | 96.895% | 96.857% | 97.300% | 96.426% | 94.625% | 96.392% | 95.410% |
| Consensus AP2 Poa ARN1 e... | 92.964% | 97.777% | | 95.668% | 96.277% | 96.625% | 96.587% | 96.936% | 96.633% | 94.174% | 95.880% | 95.004% |
| Consensus HP1 ARN1 (reve... | 91.686% | 95.901% | 95.668% | | 97.019% | 97.103% | 97.035% | 97.277% | 96.625% | 92.032% | 93.862% | 92.941% |
| Consensus AG1 Lolium extr... | 91.840% | 96.528% | 96.277% | 97.019% | | 97.488% | 97.435% | 98.335% | 98.570% | 92.244% | 94.386% | 93.176% |
| Consensus AG1 ARN1 extra... | 91.748% | 96.895% | 96.625% | 97.103% | 97.488% | | 99.720% | 98.358% | 97.299% | 92.282% | 94.272% | 93.354% |
| Consensus AG2 Agrostis sv ... | 91.741% | 96.857% | 96.587% | 97.035% | 97.435% | 99.720% | | 98.305% | 97.231% | 92.275% | 94.265% | 93.346% |
| Consensus AP2 Agrostis ext... | 91.962% | 97.300% | 96.936% | 97.277% | 98.335% | 98.358% | 98.305% | | 98.396% | 92.691% | 95.138% | 93.687% |
| Consensus HP2 (55) extract... | 92.054% | 96.426% | 96.633% | 96.625% | 98.570% | 97.299% | 97.231% | 98.396% | | 92.100% | 94.302% | 93.203% |
| Consensus AP2 ARN1 extrac... | 91.641% | 94.625% | 94.174% | 92.032% | 92.244% | 92.282% | 92.275% | 92.691% | 92.100% | | 97.132% | 96.789% |
| Consensus AP2 Lolium ARN... | 92.375% | 96.392% | 95.880% | 93.862% | 94.386% | 94.272% | 94.265% | 95.138% | 94.302% | 97.132% | | 97.134% |
| Consensus AP1 ARN1 extrac... | 92.686% | 95.410% | 95.004% | 92.941% | 93.176% | 93.354% | 93.346% | 93.687% | 93.203% | 96.789% | 97.134% | |

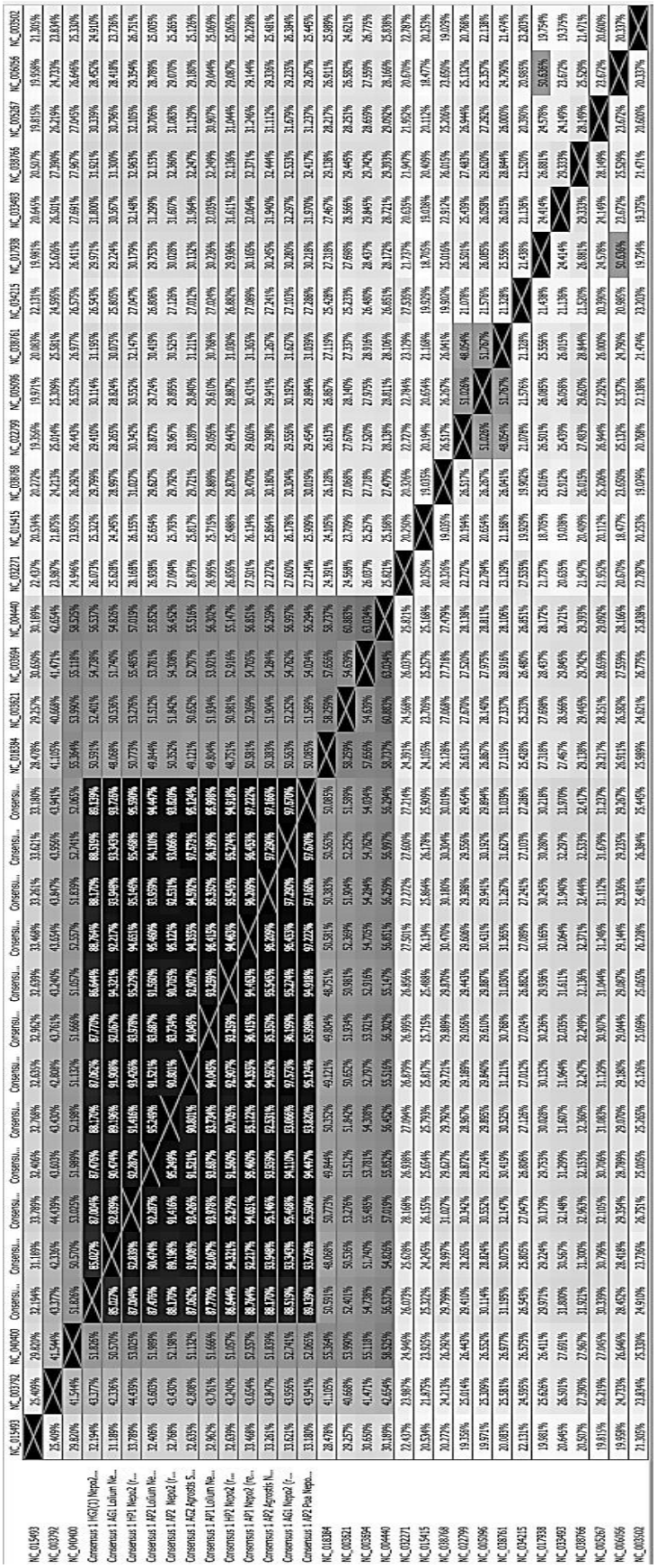
# Appendix 5

*Table 22 : Matrix of consensus sequence and references nepovirus RNA1*



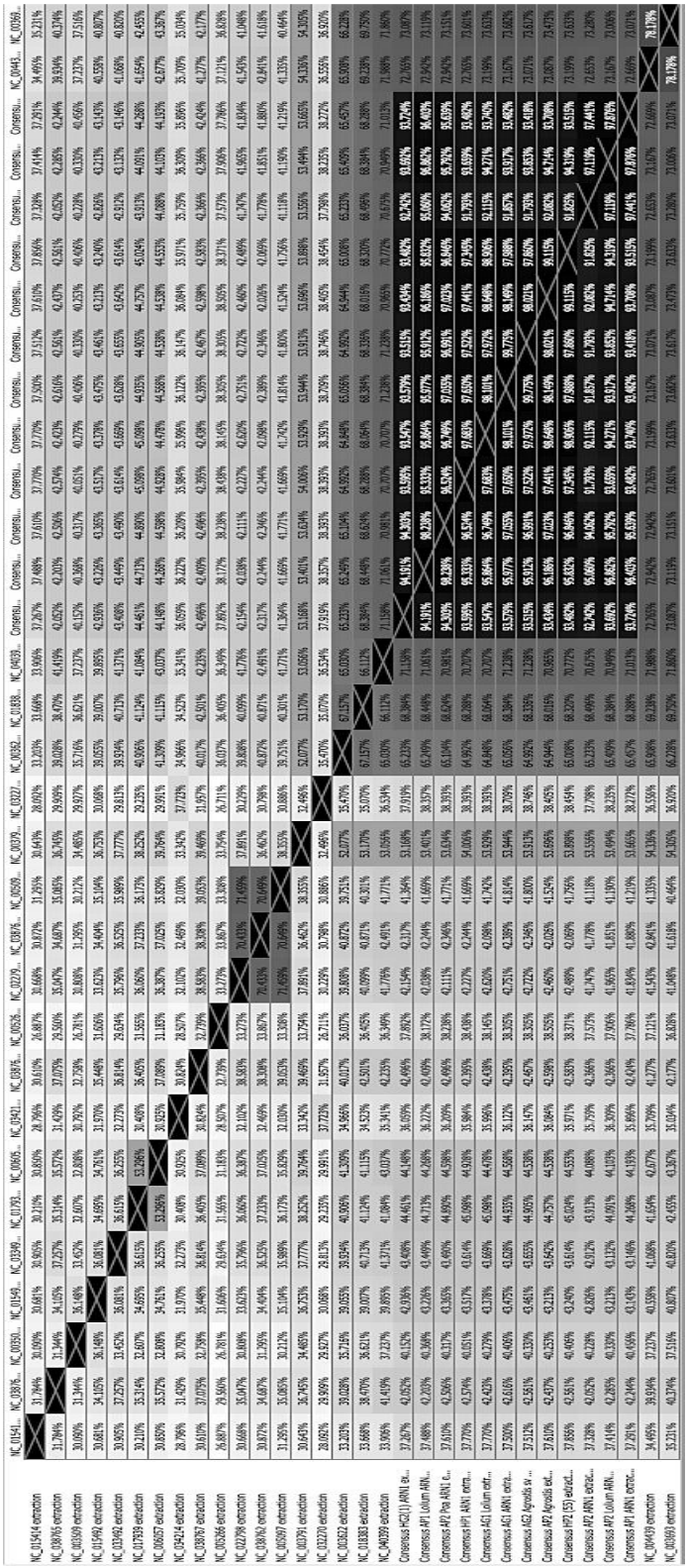
# Appendix 6

*Table 23 : Matrix of consensus sequence and references nepovirus RNA2*



# Appendix 7

*Table 24 : Matrix of consensus sequence and references nepovirus RNA1 (Polymerase-protéase)*



# Appendix 8

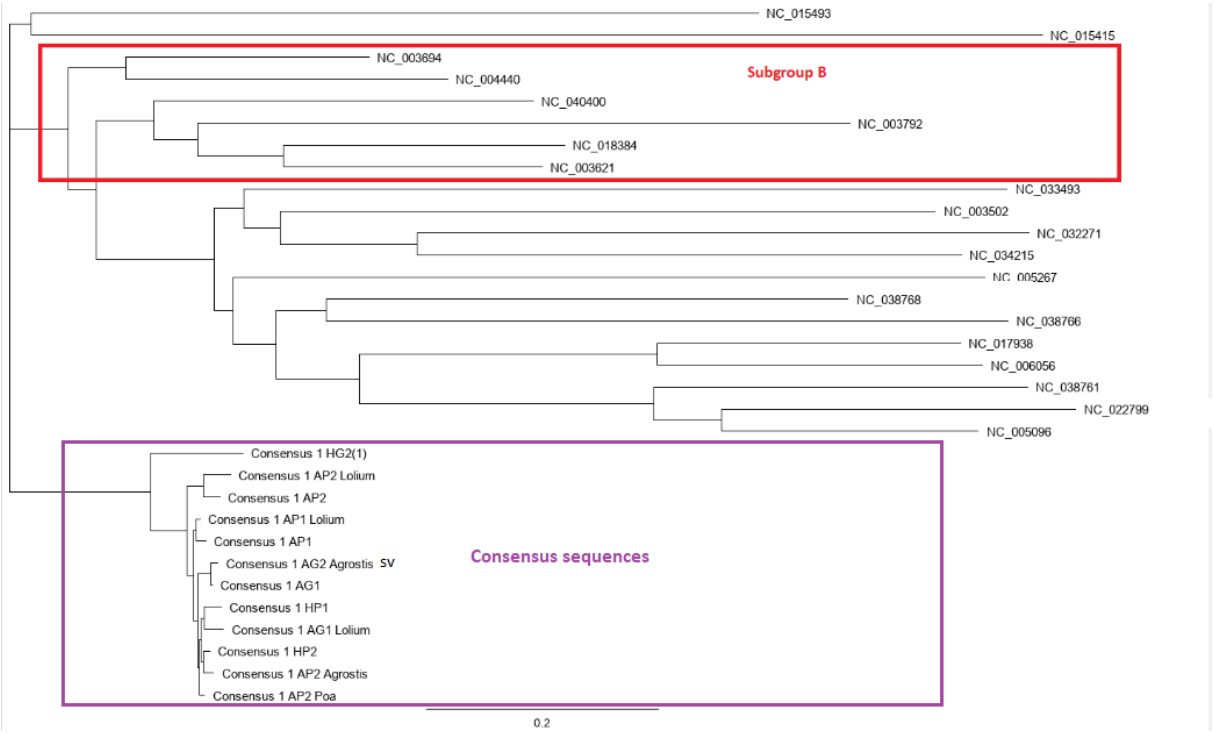
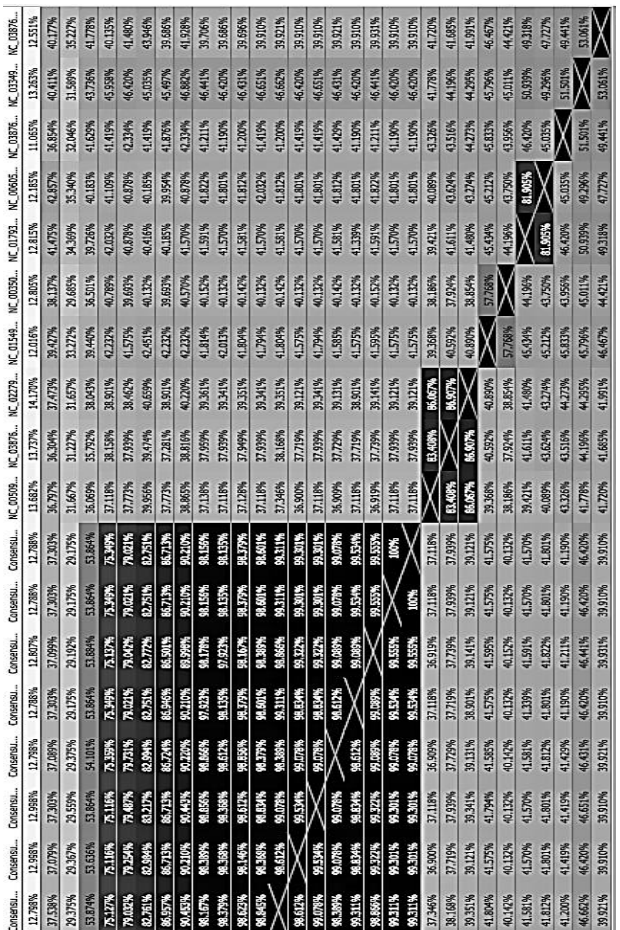



*Figure 68: Phylogenetic tree of consensus sequences of nepovirus RNA2 and references (coat protein). Red box is subgroup B of nepovirus, purple box is consensus sequences.(Reference names: see Table 11).*

*Table 25 : Matrix of consensus sequence and references nepovirus RNA2 (Coat protein)*

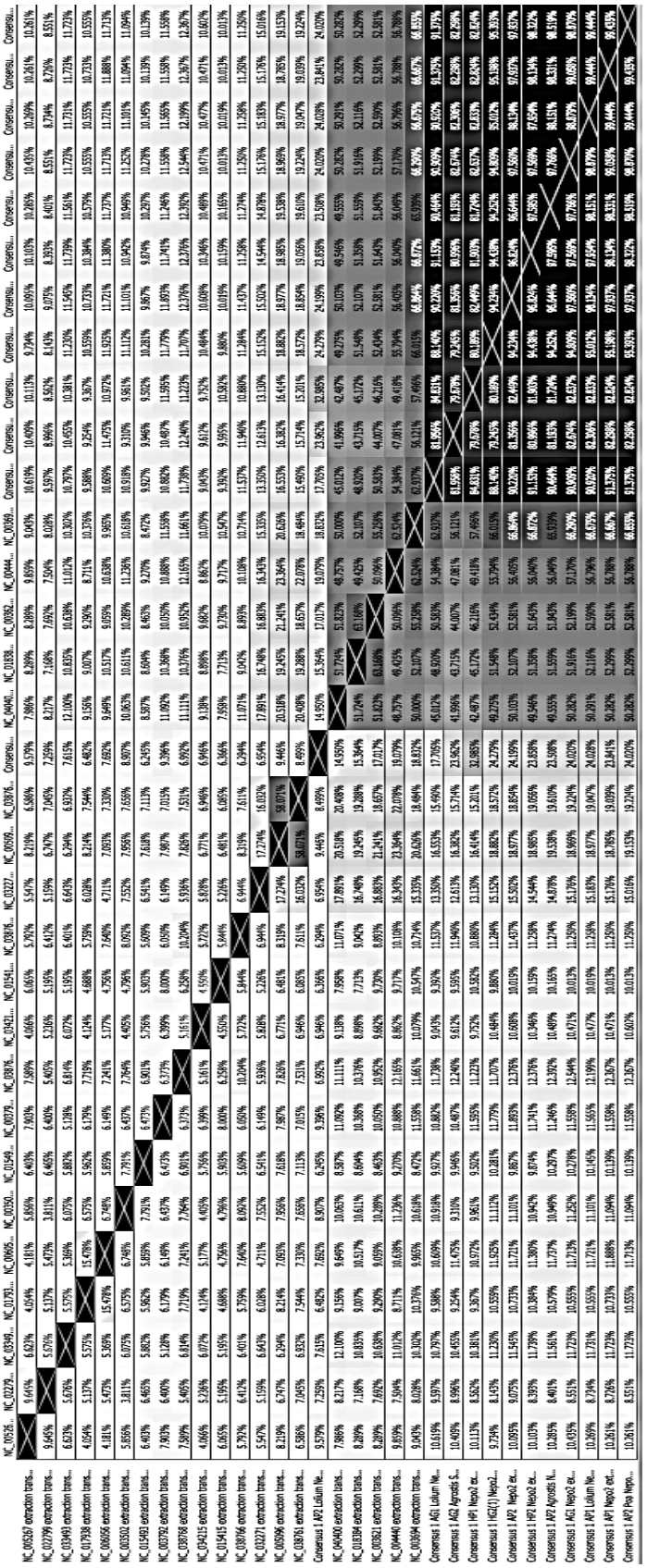
# Appendix 9

*Table 26 : Matrix of protein of consensus sequences and references nepovirus RNA1 (Polymerase-protease CG-GDD)*



# Appendix 10

*Table 27 :Matrix of protein of consensus sequences and references nepovirus RNA2 (Coat protein)*



# Appendix 11

*Table 28 : Matrix of consensus sequence waikavirus*

| | Consensus_1... | Consensus_1... | Consensus_2... | Consensus_2... | Consensus_1... | Consensus_1... | Consensus_1... | Consensus_1... | Consensus_1... | Consensus_1... |
|---|---|---|---|---|---|---|---|---|---|---|
| Consensus_1_HP2_Poa_W... | | 87.034% | 90.458% | 90.615% | 91.286% | 90.902% | 91.040% | 91.980% | 91.315% | 91.565% |
| Consensus_1_AP2_Waika e... | 87.034% | | 89.335% | 87.985% | 89.537% | 92.044% | 91.696% | 91.364% | 92.631% | 91.715% |
| Consensus_2_AP1_Waika | 90.458% | 89.335% | | 91.968% | 92.605% | 93.795% | 93.106% | 93.886% | 93.409% | 93.878% |
| Consensus_2_LP2_Poa_Wa... | 90.615% | 87.985% | 91.968% | | 92.595% | 91.929% | 91.582% | 92.414% | 92.064% | 93.239% |
| Consensus_1_HG2_Poa_W... | 91.286% | 89.537% | 92.605% | 92.595% | | 93.277% | 93.322% | 94.536% | 93.952% | 94.146% |
| Consensus_1_LG2_Poa_W... | 90.902% | 92.044% | 93.795% | 91.929% | 93.277% | | 95.903% | 95.324% | 95.827% | 95.968% |
| Consensus_1_AG_min_Wai... | 91.040% | 91.696% | 93.106% | 91.582% | 93.322% | 95.903% | | 95.246% | 96.118% | 95.780% |
| Consensus_1_AP2_Poa_W... | 91.980% | 91.364% | 93.886% | 92.414% | 94.536% | 95.324% | 95.246% | | 96.075% | 96.594% |
| Consensus_1_AG2_Poa_W... | 91.315% | 92.631% | 93.409% | 92.064% | 93.952% | 95.827% | 96.118% | 96.075% | | 96.346% |
| Consensus_1_LG2_Poa_ext... | 91.565% | 91.715% | 93.878% | 93.239% | 94.146% | 95.968% | 95.780% | 96.594% | 96.346% | |

# Appendix 12

*Table 29 : Matrix of consensus sequence waikavirus (Coat protein)*

| | Consensus 1 ... | Consensus 1 ... | Consensus 2 ... | Consensus 2 ... | Consensus 1 ... | Consensus 1 ... | Consensus 1 ... | Consensus 1 ... | Consensus 1 ... | Consensus 1 ... |
|---|---|---|---|---|---|---|---|---|---|---|
| Consensus 1 HP2 Poa Waika | | 82.659% | 85.132% | 87.617% | 85.944% | 85.670% | 85.579% | 85.974% | 86.430% | 85.944% |
| Consensus 1 AP2 Waika | 82.659% | | 86.743% | 87.339% | 89.270% | 92.505% | 93.185% | 93.212% | 91.662% | 92.451% |
| Consensus 2 AP1 Waika | 85.132% | 86.743% | | 90.685% | 89.166% | 90.916% | 89.560% | 89.532% | 90.102% | 88.990% |
| Consensus 2 LP2 Poa Waika | 87.617% | 87.339% | 90.685% | | 91.522% | 89.944% | 89.537% | 90.479% | 90.810% | 91.599% |
| Consensus 1 HG2 Poa Waika | 85.944% | 89.270% | 89.166% | 91.522% | | 91.421% | 91.802% | 92.999% | 94.857% | 93.152% |
| Consensus 1 LG2 Poa Waika | 85.670% | 92.505% | 90.916% | 89.944% | 91.421% | | 94.272% | 94.043% | 93.585% | 93.737% |
| Consensus 1 AG min Waika | 85.579% | 93.185% | 89.560% | 89.537% | 91.802% | 94.272% | | 95.163% | 93.915% | 94.908% |
| Consensus 1 AG2 Poa Waika | 85.974% | 93.212% | 89.532% | 90.479% | 92.999% | 94.043% | 95.163% | | 95.239% | 95.188% |
| Consensus 1 AP2 Poa Waika | 86.430% | 91.662% | 90.102% | 90.810% | 94.857% | 93.585% | 93.915% | 95.239% | | 95.367% |
| Consensus 1 LG2 Poa ext W... | 85.944% | 92.451% | 88.990% | 91.599% | 93.152% | 93.737% | 94.908% | 95.188% | 95.367% | |

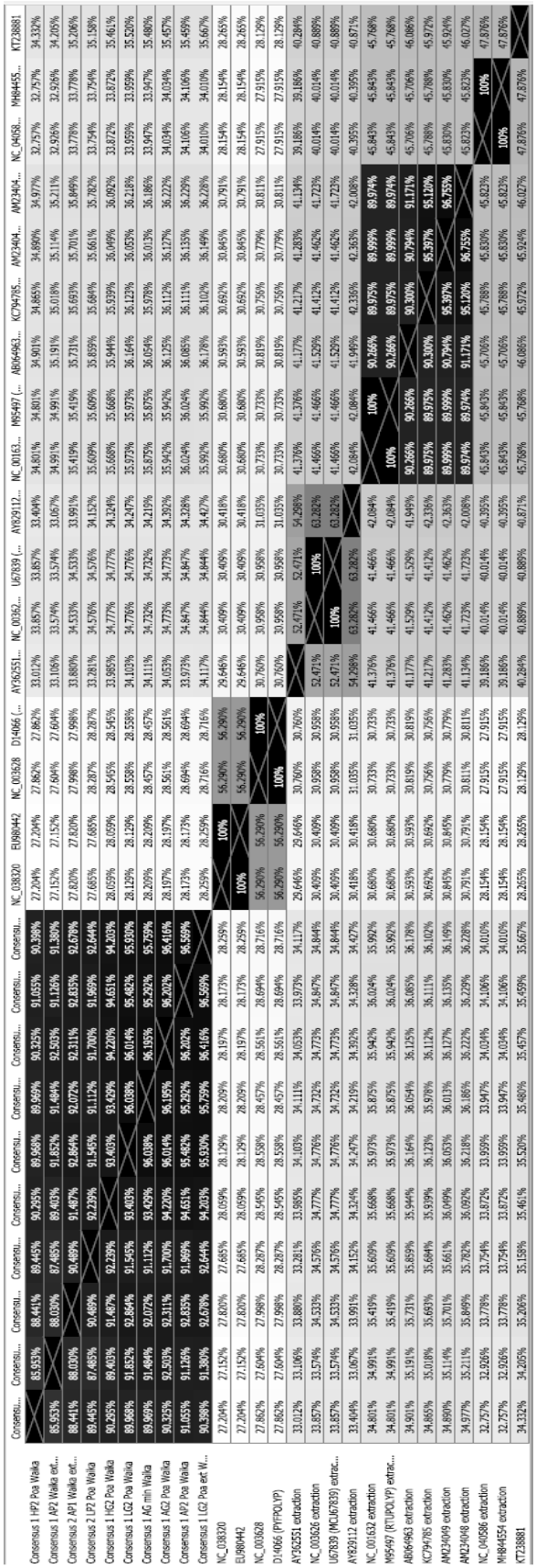
# Appendix 13

*Table 30 : Matrix of consensus sequence waikavirus (Polymerase-protease)*

| | Consensus 1 ... | Consensus 1 ... | Consensus 2 ... | Consensus 2 ... | Consensus 1 ... | Consensus 1 ... | Consensus 1 ... | Consensus 1 ... | Consensus 1 ... | Consensus 1 ... |
|---|---|---|---|---|---|---|---|---|---|---|
| Consensus 1 HP2 Poa Waika | | 89.123% | 92.499% | 92.995% | 93.456% | 93.965% | 93.747% | 93.984% | 93.965% | 94.347% |
| Consensus 1 AP2 Waika | 89.123% | | 92.031% | 91.184% | 91.330% | 92.943% | 91.960% | 92.925% | 92.331% | 92.609% |
| Consensus 2 AP1 Waika | 92.499% | 92.031% | | 94.828% | 95.397% | 96.521% | 95.505% | 96.140% | 96.140% | 96.267% |
| Consensus 2 LP2 Poa Waika | 92.995% | 91.184% | 94.828% | | 95.074% | 96.208% | 95.336% | 95.842% | 95.615% | 97.028% |
| Consensus 1 HG2 Poa Waika | 93.456% | 91.330% | 95.397% | 95.074% | | 96.371% | 95.883% | 96.354% | 96.650% | 96.842% |
| Consensus 1 LG2 Poa Waika | 93.965% | 92.943% | 96.521% | 96.208% | 96.371% | | 96.930% | 97.261% | 97.104% | 97.854% |
| Consensus 1 AG min Waika | 93.747% | 91.960% | 95.505% | 95.336% | 95.883% | 96.930% | | 97.435% | 96.860% | 97.226% |
| Consensus 1 AG2 Poa Waika | 93.984% | 92.925% | 96.140% | 95.842% | 96.354% | 97.261% | 97.435% | | 97.366% | 97.872% |
| Consensus 1 AP2 Poa Waika | 93.965% | 92.331% | 96.140% | 95.615% | 96.650% | 97.104% | 96.860% | 97.366% | | 97.645% |
| Consensus 1 LG2 Poa ext W... | 94.347% | 92.609% | 96.267% | 97.028% | 96.842% | 97.854% | 97.226% | 97.872% | 97.645% | |

# Appendix 14

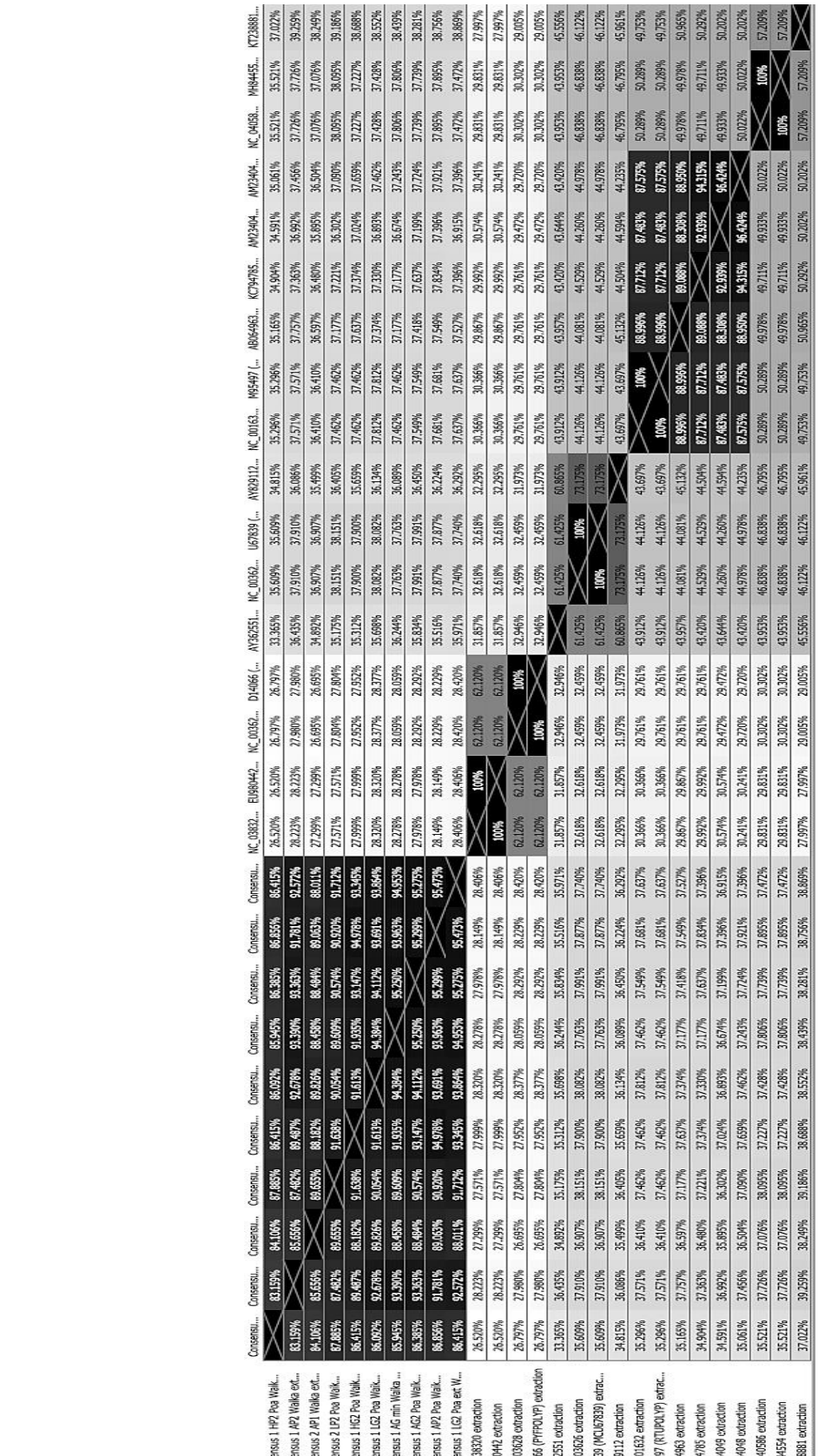
*Table 31 : Matrix of consensus sequence and references waikavirus*



| | Consensus... | Consensus... | Consensus... | Consensus... | Consensus... | Consensus... | Consensus... | Consensus... | Consensus... | NC_038320 | EU980442 | NC_003628 | D14066 (PYPOLYP) | AY362551 extraction | NC_003626 extraction | U67839 (MCU67839) extraction | AY829112 extraction | NC_001632 extraction | M95497 (RTUPOLYP) extraction | AB064963 extraction | KC794785 extraction | AM234949 extraction | AM234948 extraction | NC_040586 extraction | MH844554 extraction | KT238881 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Consensus 1 HP2 Poa Walka | | 85.953% | 88.441% | 89.445% | 90.295% | 89.568% | 89.969% | 90.325% | 91.055% | 90.398% | 27.204% | 27.204% | 27.862% | 33.012% | 33.857% | 33.857% | 33.404% | 34.801% | 34.801% | 34.865% | 34.890% | 34.977% | 32.757% | 34.332% |
| Consensus 1 AP2 Walka ext... | 85.953% | | 88.030% | 87.485% | 89.403% | 91.852% | 91.464% | 92.503% | 91.126% | 91.380% | 27.152% | 27.152% | 27.604% | 33.116% | 33.574% | 33.574% | 33.067% | 34.991% | 34.991% | 35.191% | 35.018% | 35.114% | 35.211% | 32.926% | 32.926% | 34.205% |
| Consensus 2 AP1 Walka ext... | 88.441% | 88.030% | | 90.469% | 91.487% | 92.864% | 92.072% | 92.311% | 92.835% | 92.678% | 27.820% | 27.820% | 27.998% | 33.880% | 34.533% | 34.533% | 33.991% | 35.419% | 35.419% | 35.731% | 35.693% | 35.701% | 35.849% | 33.778% | 33.778% | 35.206% |
| Consensus 2 LP2 Poa Walka | 89.445% | 87.485% | 90.469% | | 92.239% | 91.545% | 91.112% | 91.700% | 91.969% | 92.644% | 27.685% | 27.685% | 28.287% | 33.281% | 34.576% | 34.576% | 34.152% | 35.609% | 35.609% | 35.609% | 35.659% | 35.684% | 35.661% | 35.782% | 33.754% | 35.158% |
| Consensus 1 HG2 Poa Walka | 90.295% | 89.403% | 91.487% | 92.239% | | 93.403% | 93.429% | 94.220% | 94.651% | 94.203% | 28.059% | 28.059% | 28.545% | 33.985% | 34.777% | 34.777% | 34.324% | 35.668% | 35.668% | 35.944% | 35.939% | 36.049% | 36.092% | 33.872% | 33.872% | 35.461% |
| Consensus 1 LG2 Poa Walka | 89.568% | 91.852% | 92.864% | 91.545% | 93.403% | | 96.038% | 96.014% | 95.482% | 95.930% | 28.129% | 28.129% | 28.598% | 34.103% | 34.776% | 34.776% | 34.247% | 35.973% | 35.973% | 36.164% | 36.123% | 36.053% | 36.218% | 33.959% | 33.959% | 35.520% |
| Consensus 1 AG mini Walka | 89.969% | 91.464% | 92.072% | 91.112% | 93.429% | 96.038% | | 96.195% | 95.292% | 95.759% | 28.209% | 28.209% | 28.457% | 34.111% | 34.732% | 34.732% | 34.219% | 35.875% | 35.875% | 36.054% | 35.978% | 36.013% | 36.186% | 33.947% | 33.947% | 35.480% |
| Consensus 1 AG2 Poa Walka | 90.325% | 92.503% | 92.311% | 91.700% | 94.220% | 96.014% | 96.195% | | 96.202% | 96.416% | 28.197% | 28.197% | 28.561% | 34.053% | 34.773% | 34.773% | 34.392% | 35.942% | 35.942% | 36.125% | 36.112% | 36.127% | 36.222% | 34.034% | 34.034% | 35.457% |
| Consensus 1 AP2 Poa Walka | 91.055% | 91.126% | 92.835% | 91.969% | 94.651% | 95.482% | 95.292% | 96.202% | | 96.569% | 28.173% | 28.173% | 28.694% | 33.973% | 34.847% | 34.847% | 34.328% | 36.024% | 36.024% | 36.085% | 36.111% | 36.153% | 36.229% | 34.106% | 34.106% | 35.459% |
| Consensus 1 LG2 Poa ext W... | 90.398% | 91.380% | 92.678% | 92.644% | 94.203% | 95.930% | 95.759% | 96.416% | 96.569% | | 28.259% | 28.259% | 28.716% | 34.117% | 34.844% | 34.844% | 34.427% | 35.992% | 35.992% | 36.178% | 36.102% | 36.149% | 36.228% | 34.010% | 34.010% | 35.667% |
| NC_038320 | 27.204% | 27.152% | 27.820% | 27.685% | 28.059% | 28.129% | 28.209% | 28.197% | 28.173% | 28.259% | | 100% | 56.290% | 29.646% | 30.409% | 30.409% | 30.418% | 30.680% | 30.680% | 30.593% | 30.692% | 30.845% | 30.791% | 28.154% | 28.154% | 28.265% |
| EU980442 | 27.204% | 27.152% | 27.820% | 27.685% | 28.059% | 28.129% | 28.209% | 28.197% | 28.173% | 28.259% | 100% | | 56.290% | 29.646% | 30.409% | 30.409% | 30.418% | 30.680% | 30.680% | 30.593% | 30.692% | 30.845% | 30.791% | 28.154% | 28.154% | 28.265% |
| NC_003628 | 27.862% | 27.604% | 27.998% | 28.287% | 28.545% | 28.558% | 28.457% | 28.561% | 28.694% | 28.716% | 56.290% | 56.290% | | 30.760% | 30.958% | 30.958% | 31.035% | 30.733% | 30.733% | 30.819% | 30.759% | 30.779% | 30.811% | 27.915% | 27.915% | 28.129% |
| D14066 (PYPOLYP) | 27.862% | 27.604% | 27.998% | 28.287% | 28.545% | 28.558% | 28.457% | 28.561% | 28.694% | 28.716% | 56.290% | 56.290% | 100% | | 30.958% | 30.958% | 31.035% | 30.733% | 30.733% | 30.819% | 30.759% | 30.779% | 30.811% | 27.915% | 27.915% | 28.129% |
| AY362551 extraction | 33.012% | 33.106% | 33.880% | 33.281% | 33.985% | 34.103% | 34.111% | 34.053% | 33.973% | 34.117% | 29.646% | 29.646% | 30.760% | | 52.471% | 52.471% | 54.298% | 41.379% | 41.379% | 41.177% | 41.217% | 41.283% | 41.134% | 39.186% | 39.186% | 40.284% |
| NC_003626 extraction | 33.857% | 33.574% | 34.533% | 34.576% | 34.777% | 34.776% | 34.732% | 34.773% | 34.847% | 34.844% | 30.409% | 30.409% | 30.958% | 52.471% | | 100% | 63.282% | 41.466% | 41.466% | 41.466% | 41.462% | 41.462% | 41.723% | 40.014% | 40.014% | 40.889% |
| U67839 (MCU67839) extrac... | 33.857% | 33.574% | 34.533% | 34.576% | 34.777% | 34.776% | 34.732% | 34.773% | 34.847% | 34.844% | 30.409% | 30.409% | 30.958% | 52.471% | 100% | | 63.282% | 41.466% | 41.466% | 41.466% | 41.529% | 41.412% | 41.462% | 40.014% | 40.014% | 40.889% |
| AY829112 extraction | 33.404% | 33.067% | 33.991% | 34.152% | 34.324% | 34.247% | 34.219% | 34.392% | 34.328% | 34.427% | 30.418% | 30.418% | 31.035% | 54.298% | 63.282% | 63.282% | | 42.084% | 42.084% | 42.084% | 41.949% | 42.336% | 42.363% | 40.395% | 40.395% | 40.871% |
| NC_001632 extraction | 34.801% | 34.991% | 35.419% | 35.609% | 35.668% | 35.973% | 35.875% | 35.942% | 36.024% | 35.992% | 30.680% | 30.680% | 30.733% | 41.379% | 41.466% | 41.466% | 42.084% | | 100% | 90.266% | 89.975% | 89.999% | 89.974% | 45.843% | 45.843% | 45.768% |
| M95497 (RTUPOLYP) extrac... | 34.801% | 34.991% | 35.419% | 35.609% | 35.668% | 35.973% | 35.875% | 35.942% | 36.024% | 35.992% | 30.680% | 30.680% | 30.733% | 41.379% | 41.466% | 41.466% | 42.084% | 100% | | 90.266% | 89.975% | 89.999% | 89.974% | 45.843% | 45.843% | 45.768% |
| AB064963 extraction | 34.901% | 35.191% | 35.731% | 35.859% | 35.944% | 36.164% | 36.054% | 36.125% | 36.085% | 36.178% | 30.593% | 30.593% | 30.819% | 41.177% | 41.529% | 41.529% | 41.948% | 90.266% | 90.266% | | 90.300% | 90.794% | 91.171% | 45.706% | 45.706% | 46.086% |
| KC794785 extraction | 34.865% | 35.018% | 35.693% | 35.684% | 35.939% | 36.123% | 35.978% | 36.112% | 36.111% | 36.102% | 30.692% | 30.692% | 30.759% | 41.217% | 41.412% | 41.412% | 42.336% | 89.975% | 89.975% | 90.300% | | 95.397% | 95.120% | 45.788% | 45.788% | 45.972% |
| AM234949 extraction | 34.890% | 35.114% | 35.701% | 35.661% | 36.049% | 36.053% | 36.013% | 36.127% | 36.135% | 36.149% | 30.845% | 30.845% | 30.779% | 41.283% | 41.462% | 41.462% | 42.363% | 89.999% | 89.999% | 90.794% | 95.397% | | 96.755% | 45.830% | 45.830% | 45.924% |
| AM234948 extraction | 34.977% | 35.211% | 35.849% | 35.782% | 36.092% | 36.218% | 36.186% | 36.222% | 36.222% | 36.228% | 30.791% | 30.791% | 30.811% | 41.134% | 41.723% | 41.723% | 42.008% | 89.974% | 89.974% | 91.171% | 95.120% | 96.755% | | 45.823% | 45.823% | 46.027% |
| NC_040586 extraction | 32.757% | 32.926% | 33.778% | 33.754% | 33.872% | 33.959% | 33.947% | 34.034% | 34.106% | 34.010% | 28.154% | 28.154% | 27.915% | 39.186% | 40.014% | 40.014% | 40.395% | 45.843% | 45.843% | 45.706% | 45.788% | 45.830% | 45.823% | 100% | 100% | 47.876% |
| MH844554 extraction | 32.757% | 32.926% | 33.778% | 33.754% | 33.872% | 33.959% | 33.947% | 34.034% | 34.106% | 34.010% | 28.154% | 28.154% | 27.915% | 39.186% | 40.014% | 40.014% | 40.395% | 45.843% | 45.843% | 45.706% | 45.788% | 45.830% | 45.823% | 100% | 100% | 47.876% |
| KT238881 | 34.332% | 34.205% | 35.206% | 35.158% | 35.461% | 35.520% | 35.480% | 35.457% | 35.459% | 35.667% | 28.265% | 28.265% | 28.129% | 40.284% | 40.889% | 40.889% | 40.871% | 45.768% | 45.768% | 46.086% | 45.972% | 45.924% | 46.027% | 47.876% | 47.876% | |

# Appendix 15

*Table 32 : Matrix of consensus sequence and references waikavirus (Coat protein)*



| | Consensu... | Consensu... | Consensu... | Consensu... | Consensu... | Consensu... | Consensu... | Consensu... | NC_03822... | EU980442... | NC_00362... | D14066 (... | AY362551... | NC_00362... | U67839 (... | AY829112... | NC_00163... | M95497 (... | AB064963... | KC794785... | AM234094... | AM234094... | NC_040586... | MH894554... | KT238881... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Consensus 1 HP2 Poa Waik... | | 83.159% | 84.106% | 87.885% | 86.415% | 86.092% | 85.945% | 86.385% | 86.555% | 86.415% | 26.520% | 26.520% | 26.797% | 26.797% | 33.365% | 35.609% | 35.609% | 34.815% | 35.296% | 35.296% | 35.165% | 34.994% | 34.591% | 35.061% | 35.521% | 35.521% | 37.022% |
| Consensus 1 AP2 Waika ext... | 83.159% | | 85.556% | 87.482% | 89.487% | 92.678% | 93.396% | 93.563% | 91.781% | 92.572% | 28.223% | 28.223% | 27.980% | 27.980% | 36.435% | 37.910% | 37.910% | 36.086% | 37.571% | 37.571% | 37.363% | 36.892% | 37.456% | 37.726% | 37.726% | 39.259% |
| Consensus 2 AP1 Waika ext... | 84.106% | 85.556% | | 89.655% | 88.182% | 89.826% | 88.458% | 88.494% | 89.063% | 88.011% | 27.299% | 27.299% | 26.695% | 26.695% | 34.892% | 36.907% | 36.907% | 35.499% | 36.410% | 36.410% | 36.480% | 35.895% | 36.504% | 37.076% | 37.076% | 38.249% |
| Consensus 1 EP2 Poa Waik... | 87.885% | 87.482% | 89.655% | | 91.638% | 90.054% | 89.609% | 90.574% | 90.920% | 91.212% | 27.571% | 27.571% | 27.804% | 27.804% | 35.175% | 38.151% | 38.151% | 36.405% | 37.462% | 37.462% | 37.221% | 36.302% | 37.090% | 38.095% | 38.095% | 39.186% |
| Consensus 1 HG2 Poa Waik... | 86.415% | 89.487% | 88.182% | 91.638% | | 91.613% | 91.935% | 93.147% | 94.578% | 93.345% | 27.999% | 27.999% | 27.952% | 27.952% | 35.312% | 37.500% | 37.500% | 35.659% | 37.462% | 37.462% | 37.637% | 37.374% | 37.024% | 37.659% | 37.227% | 37.227% | 38.688% |
| Consensus 1 LG2 Poa Waik... | 86.092% | 92.678% | 89.826% | 90.054% | 91.613% | | 94.384% | 94.112% | 93.691% | 93.864% | 28.328% | 28.328% | 28.377% | 28.377% | 35.698% | 38.082% | 38.082% | 36.134% | 37.812% | 37.812% | 37.374% | 37.330% | 36.893% | 37.462% | 37.428% | 37.428% | 38.552% |
| Consensus 1 AG min Waika ... | 85.945% | 93.396% | 88.458% | 89.609% | 91.935% | 94.384% | | 95.250% | 93.963% | 94.953% | 28.278% | 28.278% | 28.059% | 28.059% | 36.244% | 37.763% | 37.763% | 36.089% | 37.462% | 37.462% | 37.177% | 37.177% | 36.674% | 37.243% | 37.806% | 37.806% | 38.439% |
| Consensus 1 AG2 Poa Waik... | 86.385% | 93.563% | 88.494% | 90.574% | 93.147% | 94.112% | 95.250% | | 95.299% | 95.275% | 27.978% | 27.978% | 28.292% | 28.292% | 35.834% | 37.991% | 37.991% | 36.452% | 37.549% | 37.549% | 37.418% | 37.637% | 37.199% | 37.724% | 37.739% | 37.739% | 38.281% |
| Consensus 1 AP2 Poa Waik... | 86.555% | 91.781% | 89.063% | 90.920% | 94.578% | 93.691% | 93.963% | 95.299% | | 95.473% | 28.149% | 28.149% | 28.229% | 28.229% | 35.516% | 37.877% | 37.877% | 36.224% | 37.681% | 37.681% | 37.549% | 37.834% | 37.396% | 37.921% | 37.895% | 37.895% | 38.756% |
| Consensus 1 LG2 Poa ext. W... | 86.415% | 92.572% | 88.011% | 91.212% | 93.345% | 93.864% | 94.953% | 95.275% | 95.473% | | 28.406% | 28.406% | 28.420% | 28.420% | 35.971% | 37.740% | 37.740% | 36.292% | 37.637% | 37.637% | 37.396% | 37.396% | 36.915% | 37.396% | 37.472% | 37.472% | 38.869% |
| NC_038320 extraction | 26.520% | 28.223% | 27.299% | 27.571% | 27.999% | 28.320% | 28.278% | 27.978% | 28.149% | 28.406% | | 62.120% | 62.120% | 62.120% | 31.857% | 32.618% | 32.618% | 32.295% | 30.366% | 30.366% | 29.867% | 29.992% | 30.574% | 30.241% | 29.831% | 29.831% | 27.997% |
| EU980442 extraction | 26.520% | 28.223% | 27.299% | 27.571% | 27.999% | 28.320% | 28.278% | 27.978% | 28.149% | 28.406% | 100% | | 62.120% | 62.120% | 31.857% | 32.618% | 32.618% | 32.295% | 30.366% | 30.366% | 29.867% | 29.992% | 30.574% | 30.241% | 29.831% | 29.831% | 27.997% |
| NC_003626 extraction | 26.797% | 27.980% | 26.695% | 27.804% | 27.952% | 28.377% | 28.059% | 28.292% | 28.229% | 28.420% | 62.120% | 62.120% | | 100% | 32.946% | 32.459% | 32.459% | 31.973% | 29.761% | 29.761% | 29.761% | 29.761% | 29.472% | 29.720% | 30.302% | 30.302% | 29.005% |
| D14066 (PYPOLYP) extraction | 26.797% | 27.980% | 26.695% | 27.804% | 27.952% | 28.377% | 28.059% | 28.292% | 28.229% | 28.420% | 62.120% | 62.120% | 100% | | 32.946% | 32.459% | 32.459% | 31.973% | 29.761% | 29.761% | 29.761% | 29.761% | 29.472% | 29.720% | 30.302% | 30.302% | 29.005% |
| AY362551 extraction | 33.365% | 36.435% | 34.892% | 35.175% | 35.312% | 35.698% | 36.244% | 35.834% | 35.516% | 35.971% | 31.857% | 32.946% | 32.946% | | 61.425% | 61.425% | 61.425% | 60.865% | 43.912% | 43.912% | 43.697% | 43.912% | 43.420% | 43.944% | 43.420% | 43.953% | 45.556% |
| NC_003626 extraction | 35.609% | 37.910% | 36.907% | 38.151% | 37.500% | 38.082% | 37.763% | 37.991% | 37.877% | 37.740% | 32.618% | 32.618% | 32.459% | 32.459% | 61.425% | 100% | | 73.175% | 44.126% | 44.126% | 44.081% | 44.529% | 44.260% | 44.978% | 46.838% | 46.838% | 46.122% |
| U67839 (MCU67839) extrac... | 35.609% | 37.910% | 36.907% | 38.151% | 37.500% | 38.082% | 37.763% | 37.991% | 37.877% | 37.740% | 32.618% | 32.618% | 32.459% | 32.459% | 61.425% | 100% | | 73.175% | 44.126% | 44.126% | 44.081% | 44.529% | 44.260% | 44.978% | 46.838% | 46.838% | 46.122% |
| AY829112 extraction | 34.815% | 36.086% | 35.499% | 36.405% | 35.659% | 36.134% | 36.089% | 36.450% | 36.224% | 36.292% | 32.295% | 32.295% | 31.973% | 31.973% | 60.865% | 73.175% | 73.175% | | 43.697% | 43.697% | 43.697% | 45.132% | 44.504% | 44.594% | 44.235% | 46.556% | 46.122% |
| NC_001632 extraction | 35.296% | 37.571% | 36.410% | 37.462% | 37.462% | 37.812% | 37.462% | 37.549% | 37.681% | 37.637% | 30.366% | 30.366% | 29.761% | 29.761% | 43.912% | 44.126% | 44.126% | 43.697% | 100% | 100% | 88.996% | 87.712% | 87.483% | 87.575% | 50.289% | 50.289% | 49.753% |
| M95497 (RTUPOLYP) extrac... | 35.296% | 37.571% | 36.410% | 37.462% | 37.462% | 37.812% | 37.462% | 37.549% | 37.681% | 37.637% | 30.366% | 30.366% | 29.761% | 29.761% | 43.912% | 44.126% | 44.126% | 43.697% | 100% | 100% | 88.996% | 87.712% | 87.483% | 87.575% | 50.289% | 50.289% | 49.753% |
| AB064963 extraction | 35.165% | 37.757% | 36.597% | 37.177% | 37.637% | 37.374% | 37.177% | 37.418% | 37.549% | 37.527% | 29.867% | 29.867% | 29.867% | 29.867% | 44.981% | 45.132% | 44.981% | 44.504% | 88.996% | 88.996% | 100% | 89.088% | 89.308% | 88.950% | 49.978% | 49.978% | 50.965% |
| KC794785 extraction | 34.994% | 37.363% | 36.480% | 37.221% | 37.374% | 37.330% | 37.177% | 37.637% | 37.834% | 37.396% | 29.992% | 29.992% | 29.761% | 29.761% | 43.420% | 44.529% | 44.529% | 44.504% | 87.712% | 87.712% | 89.088% | 100% | 92.939% | 94.315% | 49.711% | 49.711% | 50.292% |
| AM234094 extraction | 34.591% | 36.892% | 35.895% | 36.302% | 37.024% | 36.893% | 36.674% | 37.199% | 37.396% | 36.915% | 30.574% | 30.574% | 29.472% | 29.472% | 43.644% | 44.260% | 44.260% | 44.594% | 87.483% | 87.483% | 88.308% | 92.939% | 100% | 96.424% | 49.933% | 49.933% | 50.202% |
| AM234094 extraction | 35.061% | 37.456% | 36.504% | 37.090% | 37.659% | 37.462% | 37.243% | 37.724% | 37.921% | 37.396% | 30.241% | 30.241% | 29.720% | 29.720% | 43.420% | 44.978% | 44.978% | 44.235% | 87.575% | 87.575% | 88.950% | 94.315% | 96.424% | 100% | 50.022% | 50.022% | 50.202% |
| NC_040586 extraction | 35.521% | 37.726% | 37.076% | 38.095% | 37.227% | 37.428% | 37.806% | 37.739% | 37.895% | 37.472% | 29.831% | 29.831% | 30.302% | 30.302% | 43.953% | 46.838% | 46.838% | 46.795% | 50.289% | 50.289% | 49.978% | 49.711% | 49.933% | 50.022% | 100% | 57.209% | |
| MH894554 extraction | 35.521% | 37.726% | 37.076% | 38.095% | 37.227% | 37.428% | 37.806% | 37.739% | 37.895% | 37.472% | 29.831% | 29.831% | 30.302% | 30.302% | 43.953% | 46.838% | 46.838% | 46.795% | 50.289% | 50.289% | 49.978% | 49.711% | 49.933% | 50.022% | 100% | 57.209% | |
| KT238881 extraction | 37.022% | 39.259% | 38.249% | 39.186% | 38.688% | 38.552% | 38.439% | 38.281% | 38.756% | 38.869% | 27.997% | 27.997% | 29.005% | 29.005% | 45.556% | 46.122% | 46.122% | 45.961% | 49.753% | 49.753% | 50.965% | 50.292% | 50.202% | 50.202% | 57.209% | 57.209% | |

# Appendix 16

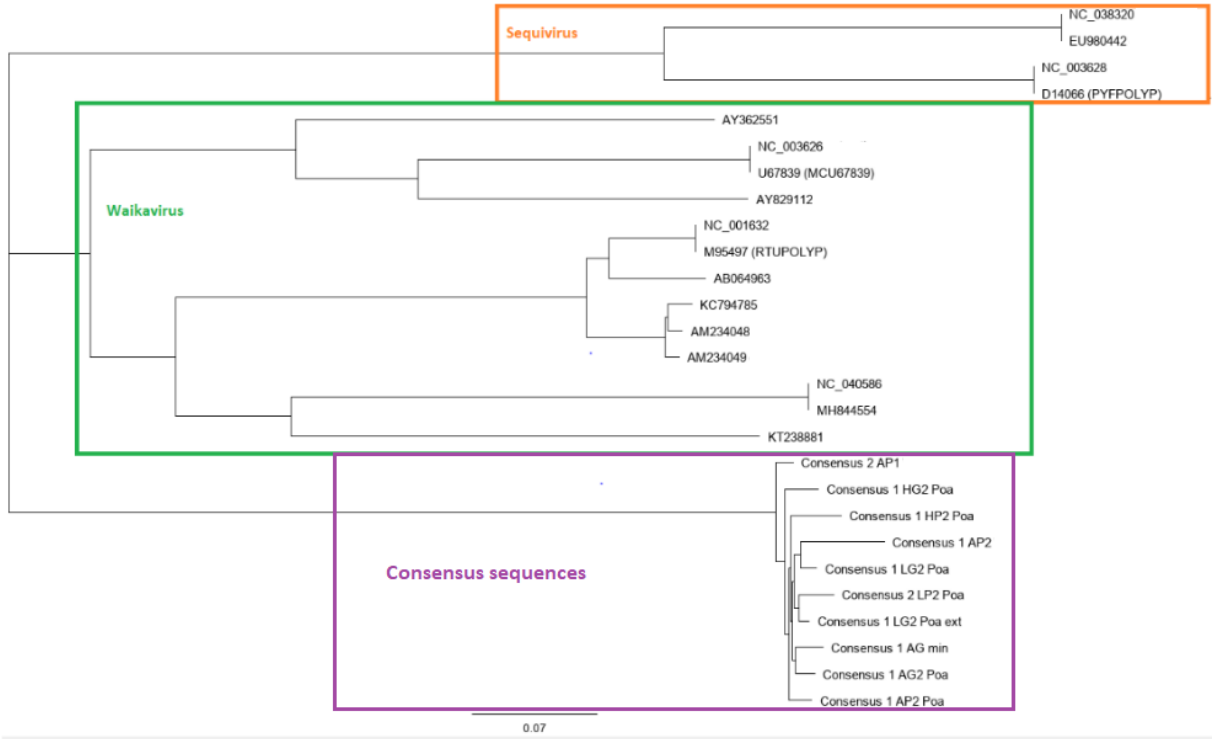


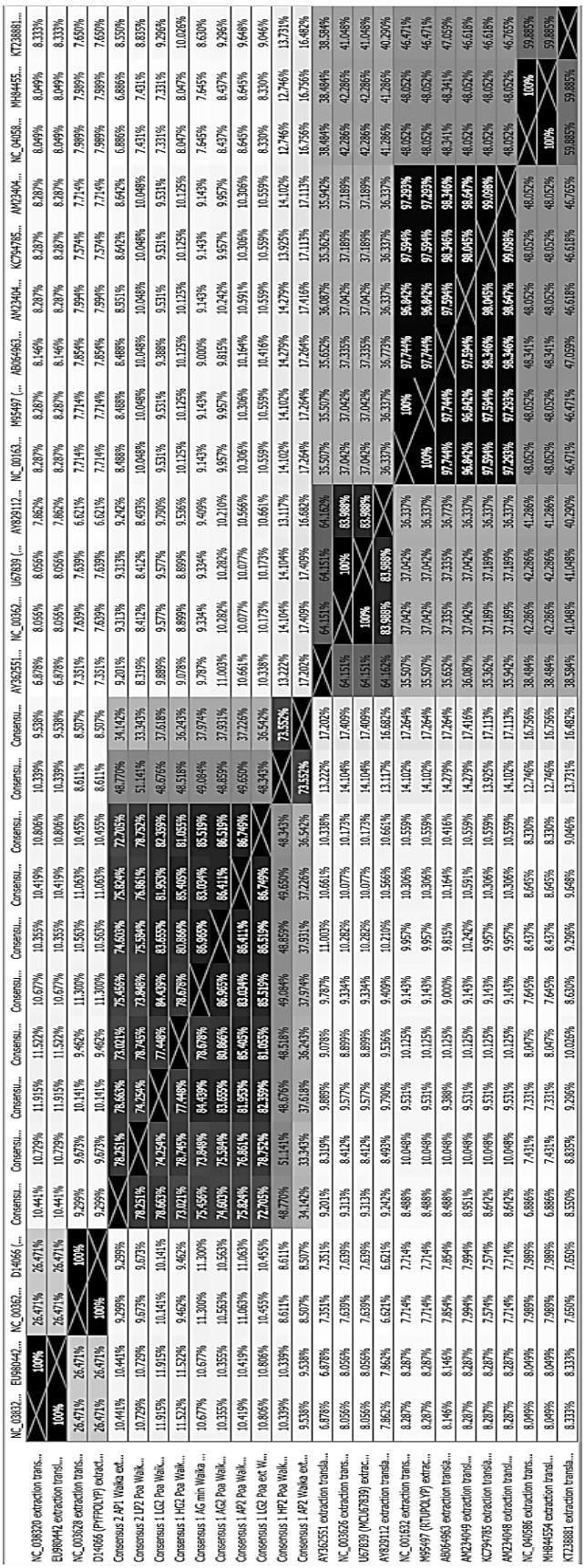

*Figure 69: Phylogenetic tree of consensus sequences of waikavirus and references (polymerase-protease). Coloured squares correspond to Waikavirus, Sequivirus genera and consensus sequences (green for Waikavirus, orange for Sequivirus, purple for consensus sequences). (Reference names: see Table 9).*

*Table 33 : Matrix of consensus sequence and references waikavirus (Polymerase-protease)*

| | Consensu... | Consensu... | Consensu... | Consensu... | Consensu... | Consensu... | Consensu... | Consensu... | Consensu... | NC_03832... | EU880442... | NC_00362... | D140665 (... | AY362551... | NC_00362... | U67839 (... | AY829112... | NC_00163... | M95497 (... | AB064963... | KC794785... | AM234044... | NC_04058... | MH844554... | KT238881... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Consensus 1 HP2 Poa Waik... | 89.122% | 92.499% | 93.495% | 93.495% | 93.495% | 93.747% | 93.906% | 93.965% | 94.347% | 33.274% | 33.274% | 33.841% | 33.841% | 42.048% | 41.529% | 41.529% | 40.876% | 43.553% | 43.553% | 43.432% | 43.532% | 43.754% | 43.772% | 40.685% | 40.685% | 41.495% |
| Consensus 1 AP2 Waika ext... | 89.122% | 92.051% | 91.184% | 91.330% | 92.345% | 91.960% | 92.192% | 92.923% | 92.639% | 31.775% | 31.775% | 32.551% | 32.551% | 40.872% | 40.121% | 40.121% | 40.100% | 42.718% | 42.718% | 42.507% | 42.685% | 43.087% | 43.017% | 39.605% | 39.605% | 39.375% |
| Consensus 2 AP1 Waika ext... | 92.499% | 92.051% | 94.828% | 95.397% | 96.521% | 95.576% | 96.140% | 96.140% | 96.357% | 34.290% | 34.290% | 34.707% | 34.707% | 43.665% | 43.032% | 43.032% | 42.870% | 44.669% | 44.669% | 44.703% | 44.929% | 45.288% | 45.323% | 42.461% | 42.461% | 43.139% |
| Consensus 2 LP2 Poa Waik... | 92.995% | 91.184% | 94.828% | 95.074% | 96.208% | 95.336% | 96.342% | 96.615% | 97.038% | 33.671% | 33.671% | 34.622% | 34.622% | 42.729% | 42.367% | 42.367% | 42.129% | 44.391% | 44.391% | 44.596% | 44.459% | 44.772% | 44.706% | 42.018% | 42.018% | 42.091% |
| Consensus 1 HG2 Poa Waik... | 93.495% | 91.330% | 95.397% | 95.074% | 96.371% | 95.885% | 96.335% | 96.650% | 96.842% | 34.035% | 34.035% | 34.758% | 34.758% | 43.330% | 42.604% | 42.604% | 42.366% | 44.695% | 44.695% | 44.728% | 44.846% | 45.192% | 45.093% | 42.106% | 42.106% | 42.427% |
| Consensus 1 LG2 Poa Waik... | 93.965% | 92.943% | 96.521% | 96.208% | 96.371% | 96.930% | 97.261% | 97.104% | 97.859% | 33.725% | 33.725% | 34.742% | 34.742% | 43.562% | 42.654% | 42.654% | 42.432% | 44.960% | 44.960% | 44.993% | 44.945% | 45.325% | 45.192% | 42.355% | 42.355% | 42.725% |
| Consensus 1 AG min Waika ... | 93.747% | 91.960% | 95.576% | 95.336% | 96.883% | 96.930% | 97.435% | 97.435% | 96.860% | 33.970% | 33.970% | 34.498% | 34.498% | 43.082% | 43.571% | 42.571% | 42.284% | 44.778% | 44.778% | 44.579% | 44.730% | 45.242% | 45.176% | 41.874% | 41.874% | 42.229% |
| Consensus 1 AG2 Poa Waik... | 93.984% | 92.925% | 96.140% | 95.892% | 96.334% | 97.261% | 97.435% | 97.366% | 97.872% | 34.003% | 34.003% | 34.628% | 34.628% | 43.247% | 42.505% | 42.505% | 42.416% | 44.678% | 44.678% | 44.612% | 44.597% | 44.944% | 44.910% | 41.940% | 41.940% | 42.345% |
| Consensus 1 AP2 Poa Waik... | 93.965% | 92.331% | 96.140% | 95.615% | 96.650% | 97.104% | 96.860% | 97.366% | 97.645% | 34.052% | 34.052% | 34.693% | 34.693% | 43.413% | 42.422% | 42.422% | 42.350% | 45.109% | 45.109% | 44.595% | 44.697% | 45.043% | 44.944% | 42.073% | 42.073% | 42.262% |
| Consensus 1 LG2 Poa ext W... | 94.347% | 92.639% | 96.357% | 97.038% | 96.842% | 97.859% | 97.872% | 97.872% | 97.645% | 33.807% | 33.807% | 34.742% | 34.742% | 43.413% | 42.753% | 42.753% | 42.564% | 44.794% | 44.794% | 44.808% | 44.786% | 45.192% | 45.093% | 41.957% | 41.957% | 42.543% |
| NC_038320 extraction | 33.274% | 31.775% | 34.290% | 33.671% | 34.035% | 33.725% | 33.970% | 34.003% | 34.052% | 33.807% | 100% | 62.298% | 62.298% | 37.440% | 37.466% | 37.466% | 37.907% | 37.237% | 37.237% | 37.638% | 37.784% | 37.705% | 37.738% | 37.239% | 37.239% | 37.718% |
| EU880442 extraction | 33.274% | 31.775% | 34.290% | 33.671% | 34.035% | 33.725% | 33.970% | 34.003% | 34.052% | 33.807% | 100% | 62.298% | 62.298% | 37.440% | 37.466% | 37.466% | 37.907% | 37.237% | 37.237% | 37.638% | 37.784% | 37.705% | 37.738% | 37.239% | 37.239% | 37.718% |
| NC_003626 extraction | 33.841% | 32.551% | 34.707% | 34.622% | 34.758% | 34.742% | 34.498% | 34.628% | 34.693% | 34.742% | 62.298% | 62.298% | 100% | 37.778% | 38.018% | 38.018% | 37.644% | 37.023% | 37.023% | 37.124% | 37.425% | 37.425% | 37.559% | 35.954% | 35.954% | 36.053% |
| D140665 (PrFPOLYP) extraction | 33.841% | 32.551% | 34.707% | 34.622% | 34.758% | 34.742% | 34.498% | 34.628% | 34.693% | 34.742% | 62.298% | 100% | 100% | 37.778% | 38.018% | 38.018% | 37.644% | 37.023% | 37.023% | 37.124% | 37.425% | 37.425% | 37.559% | 35.954% | 35.954% | 36.053% |
| AY362551 extraction | 42.048% | 40.872% | 43.665% | 42.729% | 43.330% | 43.562% | 43.082% | 43.247% | 43.413% | 43.413% | 37.440% | 37.440% | 37.778% | 100% | 60.937% | 60.937% | 62.097% | 50.480% | 50.480% | 50.486% | 49.794% | 50.257% | 50.103% | 49.794% | 50.905% | 50.757% |
| NC_003626 extraction | 41.529% | 40.121% | 43.032% | 42.367% | 42.604% | 42.654% | 42.571% | 42.505% | 42.422% | 42.753% | 37.466% | 37.466% | 38.018% | 60.937% | 100% | 100% | 69.436% | 50.587% | 50.587% | 50.465% | 50.276% | 50.500% | 50.500% | 50.314% | 50.314% | 51.501% |
| U67839 (MCU67839) extrac... | 41.529% | 40.121% | 43.032% | 42.367% | 42.604% | 42.654% | 42.571% | 42.505% | 42.422% | 42.753% | 37.466% | 37.466% | 38.018% | 60.937% | 100% | 100% | 69.436% | 50.587% | 50.587% | 50.465% | 50.276% | 50.500% | 50.500% | 50.314% | 50.314% | 51.501% |
| AY829112 extraction | 40.876% | 40.100% | 42.870% | 42.129% | 42.366% | 42.432% | 42.284% | 42.416% | 42.350% | 42.564% | 37.907% | 37.907% | 37.644% | 62.097% | 69.436% | 69.436% | 100% | 51.923% | 51.923% | 50.566% | 52.023% | 51.938% | 51.630% | 51.638% | 51.638% | 50.951% |
| NC_001672 extraction | 43.553% | 42.718% | 44.669% | 44.391% | 44.695% | 44.960% | 44.778% | 44.678% | 45.109% | 44.794% | 37.237% | 37.023% | 37.023% | 50.480% | 50.587% | 50.587% | 51.923% | 100% | 100% | 99.618% | 89.252% | 89.283% | 89.109% | 53.586% | 53.586% | 54.781% |
| M95497 (TCTUPOLYP) extrac... | 43.553% | 42.718% | 44.669% | 44.391% | 44.695% | 44.960% | 44.778% | 45.109% | 44.794% | 37.237% | 37.023% | 37.023% | 37.023% | 50.480% | 50.587% | 50.587% | 51.923% | 100% | 99.618% | 99.618% | 89.252% | 89.283% | 89.109% | 53.586% | 53.586% | 54.781% |
| AB064963 extraction | 43.432% | 42.507% | 44.703% | 44.596% | 44.728% | 44.993% | 44.579% | 44.612% | 44.595% | 44.828% | 37.638% | 37.638% | 37.124% | 49.794% | 50.465% | 50.465% | 50.566% | 90.618% | 90.618% | 100% | 89.860% | 89.353% | 89.609% | 53.138% | 53.138% | 55.022% |
| KC794785 extraction | 43.532% | 42.685% | 44.929% | 44.459% | 44.846% | 44.945% | 44.730% | 44.697% | 44.697% | 44.780% | 37.784% | 37.784% | 37.425% | 50.257% | 50.276% | 50.276% | 52.023% | 89.252% | 89.252% | 89.860% | 100% | 89.353% | 89.609% | 53.568% | 53.568% | 54.969% |
| AM234044 extraction | 43.754% | 43.087% | 45.288% | 44.772% | 45.192% | 45.325% | 45.242% | 44.944% | 45.043% | 45.192% | 37.705% | 37.705% | 37.425% | 50.103% | 50.500% | 50.500% | 51.938% | 89.283% | 89.283% | 89.353% | 89.353% | 100% | 98.419% | 53.690% | 53.690% | 55.091% |
| AM234098 extraction | 43.772% | 43.017% | 45.323% | 44.706% | 45.093% | 45.192% | 45.178% | 44.910% | 44.944% | 45.093% | 37.738% | 37.738% | 37.559% | 49.794% | 50.500% | 50.500% | 51.630% | 89.109% | 89.109% | 89.609% | 97.822% | 98.419% | 100% | 53.655% | 53.655% | 55.195% |
| NC_040586 extraction | 40.685% | 39.605% | 42.981% | 42.018% | 42.106% | 42.355% | 41.874% | 41.940% | 42.073% | 41.957% | 37.239% | 37.239% | 35.954% | 50.905% | 50.314% | 50.314% | 51.638% | 53.586% | 53.586% | 53.138% | 53.568% | 53.690% | 53.655% | 100% | 100% | 58.415% |
| MH844554 extraction | 40.685% | 39.605% | 42.981% | 42.018% | 42.106% | 42.355% | 41.874% | 41.940% | 42.073% | 41.957% | 37.239% | 37.239% | 35.954% | 50.905% | 50.314% | 50.314% | 51.638% | 53.586% | 53.586% | 53.138% | 53.568% | 53.690% | 53.655% | 100% | 100% | 58.415% |
| KT238881 extraction | 41.495% | 39.375% | 43.139% | 42.091% | 42.427% | 42.725% | 42.229% | 42.345% | 42.262% | 42.543% | 37.718% | 37.718% | 36.053% | 50.757% | 51.501% | 51.501% | 50.951% | 54.781% | 54.781% | 55.022% | 54.969% | 55.091% | 55.195% | 58.415% | 58.415% | 100% |

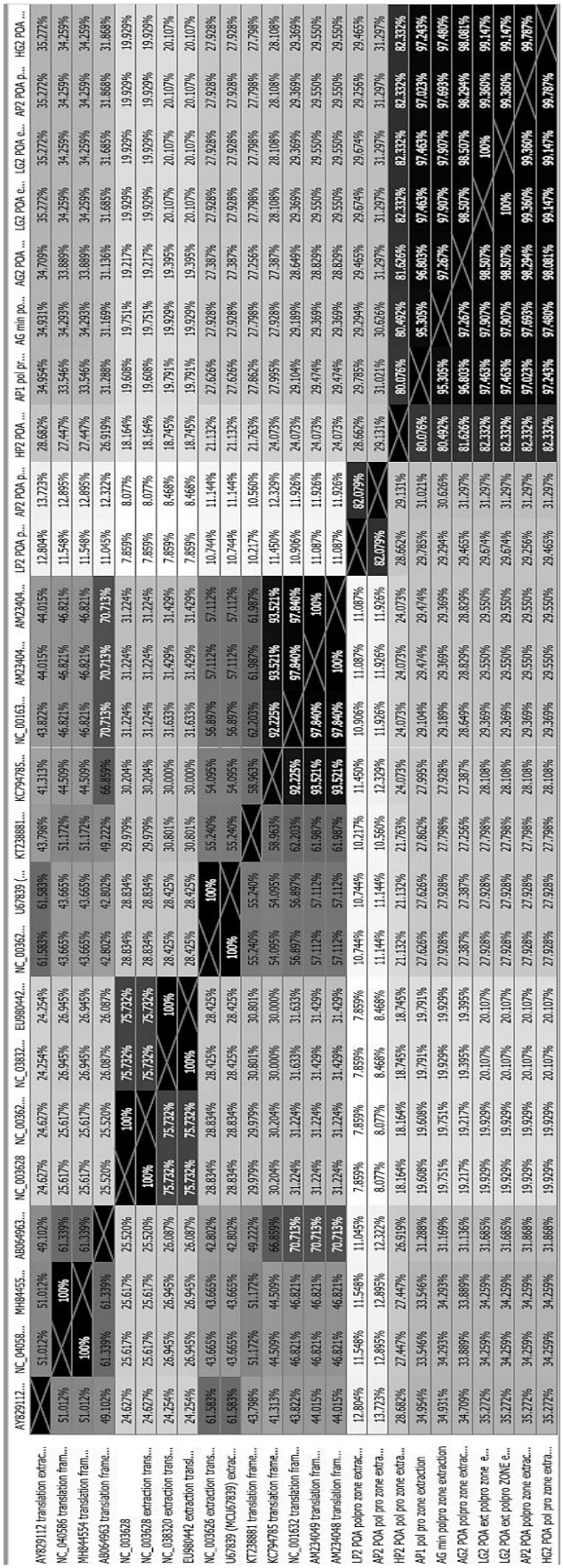
# Appendix 17

*Table 34 : Matrix of protein of consensus sequences and references waikavirus (Coat protein)*



# Appendix 18

*Table 35 : Matrix of protein of consensus sequences and references waikavirus (Polymerase-protéase CG-GDD)*



| | AY829112... | NC_040586... | MH894554... | AB064963... | NC_003628 | NC_00362... | NC_03832... | EU980442... | NC_00362... | U67839 (... | KT238881... | KC794785... | NC_00163... | AM234094... | LP2 POA p... | AP2 POA p... | HP2 POA ... | AP1 pol pr... | AG min po... | AG2 POA ... | LG2 POA e... | LG2 POA e... | AP2 POA p... | HG2 POA ... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AY829112 translation extrac... | | 51.012% | 51.012% | 49.102% | 24.627% | 24.627% | 24.254% | 24.254% | 61.583% | 61.583% | 43.788% | 41.313% | 43.822% | 44.015% | 12.804% | 13.723% | 28.682% | 34.954% | 34.931% | 34.709% | 35.272% | 35.272% | 35.272% | 35.272% |
| NC_040586 translation fram... | 51.012% | | 100% | 61.339% | 25.617% | 25.617% | 26.945% | 26.945% | 43.665% | 43.665% | 44.509% | 44.509% | 46.821% | 46.821% | 11.548% | 12.895% | 27.447% | 33.546% | 34.293% | 33.889% | 34.259% | 34.259% | 34.259% | 34.259% |
| MH894554 translation fram... | 51.012% | 100% | | 61.339% | 25.617% | 25.617% | 26.945% | 26.945% | 43.665% | 43.665% | 51.172% | 44.509% | 46.821% | 46.821% | 11.548% | 12.895% | 27.447% | 33.546% | 34.293% | 33.889% | 34.259% | 34.259% | 34.259% | 34.259% |
| AB064963 translation frame... | 49.102% | 61.339% | 61.339% | | 25.520% | 25.520% | 26.087% | 26.087% | 42.802% | 42.802% | 49.222% | 66.859% | 70.713% | 70.713% | 11.045% | 12.322% | 26.919% | 31.288% | 31.169% | 31.136% | 31.685% | 31.685% | 31.868% | 31.868% |
| NC_003628 | 24.627% | 25.617% | 25.617% | 25.520% | | 100% | 75.732% | 75.732% | 28.834% | 28.834% | 29.979% | 30.204% | 31.224% | 31.224% | 7.859% | 8.077% | 18.164% | 19.608% | 19.751% | 19.217% | 19.929% | 19.929% | 19.929% | 19.929% |
| NC_003628 extraction trans... | 24.627% | 25.617% | 25.617% | 25.520% | 100% | | 75.732% | 75.732% | 28.834% | 28.834% | 29.979% | 30.204% | 31.224% | 31.224% | 7.859% | 8.077% | 18.164% | 19.608% | 19.751% | 19.217% | 19.929% | 19.929% | 19.929% | 19.929% |
| NC_038320 extraction trans... | 24.254% | 26.945% | 26.945% | 26.087% | 75.732% | 75.732% | | 100% | 28.425% | 28.425% | 30.801% | 30.000% | 31.633% | 31.429% | 7.859% | 8.468% | 18.745% | 19.791% | 19.929% | 19.395% | 20.107% | 20.107% | 20.107% | 20.107% |
| EU980442 extraction transl... | 24.254% | 26.945% | 26.945% | 26.087% | 75.732% | 75.732% | 100% | | 28.425% | 28.425% | 30.801% | 30.000% | 31.633% | 31.429% | 7.859% | 8.468% | 18.745% | 19.791% | 19.929% | 19.395% | 20.107% | 20.107% | 20.107% | 20.107% |
| NC_003626 extraction trans... | 61.583% | 43.665% | 43.665% | 42.802% | 28.834% | 28.834% | 28.425% | 28.425% | | 100% | 55.240% | 54.095% | 56.897% | 57.112% | 10.744% | 11.144% | 21.132% | 27.626% | 27.928% | 27.387% | 27.928% | 27.928% | 27.928% | 27.928% |
| U67839 (MCU67839) extrac... | 61.583% | 43.665% | 43.665% | 42.802% | 28.834% | 28.834% | 28.425% | 28.425% | 100% | | 55.240% | 54.095% | 56.897% | 57.112% | 10.744% | 11.144% | 21.132% | 27.626% | 27.928% | 27.387% | 27.928% | 27.928% | 27.928% | 27.928% |
| KT238881 translation frame... | 43.798% | 51.172% | 51.172% | 49.222% | 29.979% | 29.979% | 30.801% | 30.801% | 55.240% | 55.240% | | 58.963% | 62.703% | 61.987% | 10.217% | 10.560% | 21.763% | 27.862% | 27.798% | 27.256% | 27.798% | 27.798% | 27.798% | 27.798% |
| KC794785 translation frame... | 41.313% | 44.509% | 44.509% | 66.859% | 30.204% | 30.204% | 30.000% | 30.000% | 54.095% | 54.095% | 58.963% | | 92.225% | 93.521% | 11.450% | 12.329% | 24.073% | 27.995% | 27.928% | 27.387% | 28.108% | 28.108% | 28.108% | 28.108% |
| NC_001632 translation fram... | 43.822% | 46.821% | 46.821% | 70.713% | 31.224% | 31.224% | 31.633% | 31.633% | 56.897% | 56.897% | 62.203% | 92.225% | | 97.840% | 10.906% | 11.926% | 24.073% | 29.104% | 29.189% | 28.649% | 29.369% | 29.369% | 29.369% | 29.369% |
| AM234049 translation fram... | 44.015% | 46.821% | 46.821% | 70.713% | 31.224% | 31.224% | 31.429% | 31.429% | 57.112% | 57.112% | 61.987% | 93.521% | 97.840% | | 11.087% | 11.926% | 24.073% | 29.474% | 29.369% | 28.829% | 29.550% | 29.550% | 29.550% | 29.550% |
| AM234048 translation fram... | 44.015% | 46.821% | 46.821% | 70.713% | 31.224% | 31.224% | 31.429% | 31.429% | 57.112% | 57.112% | 61.987% | 93.521% | 97.840% | 100% | 11.087% | 11.926% | 24.073% | 29.474% | 29.369% | 28.829% | 29.550% | 29.550% | 29.550% | 29.550% |
| LP2 POA polpro zone extrac... | 12.804% | 11.548% | 11.548% | 11.045% | 7.859% | 7.859% | 7.859% | 7.859% | 10.744% | 10.744% | 10.217% | 11.450% | 10.906% | 11.087% | | 82.079% | 28.662% | 29.785% | 29.394% | 29.465% | 29.674% | 29.674% | 29.256% | 29.465% |
| AP2 POA pol pro zone extra... | 13.723% | 12.895% | 12.895% | 12.322% | 8.077% | 8.077% | 8.468% | 8.468% | 11.144% | 11.144% | 10.560% | 12.329% | 11.926% | 11.926% | 82.079% | | 29.131% | 31.021% | 30.626% | 31.297% | 31.297% | 31.297% | 31.297% | 31.297% |
| HP2 POA pol pro zone extra... | 28.682% | 27.447% | 27.447% | 26.919% | 18.164% | 18.164% | 18.745% | 18.745% | 21.132% | 21.132% | 21.763% | 24.073% | 24.073% | 24.073% | 28.662% | 29.131% | | 80.076% | 80.492% | 81.636% | 82.332% | 82.332% | 82.332% | 82.332% |
| AP1 pol pro zone extraction | 34.954% | 33.546% | 33.546% | 31.288% | 19.608% | 19.608% | 19.791% | 19.791% | 27.626% | 27.626% | 27.862% | 27.995% | 29.104% | 29.474% | 29.785% | 31.021% | 80.076% | | 95.385% | 96.803% | 97.463% | 97.463% | 97.032% | 97.243% |
| AG min polpro zone extraction | 34.931% | 34.293% | 34.293% | 31.169% | 19.751% | 19.751% | 19.929% | 19.929% | 27.928% | 27.928% | 27.798% | 27.928% | 29.189% | 29.369% | 29.294% | 30.626% | 80.492% | 95.385% | | 97.267% | 97.907% | 97.907% | 97.693% | 97.480% |
| AG2 POA polpro zone extrac... | 34.709% | 33.889% | 33.889% | 31.136% | 19.217% | 19.217% | 19.395% | 19.395% | 27.387% | 27.387% | 27.256% | 27.387% | 28.649% | 28.829% | 29.465% | 31.297% | 81.636% | 96.803% | 97.267% | | 98.507% | 98.507% | 98.294% | 98.081% |
| LG2 POA ext polpro zone e... | 35.272% | 34.259% | 34.259% | 31.685% | 19.929% | 19.929% | 20.107% | 20.107% | 27.928% | 27.928% | 27.798% | 28.108% | 29.369% | 29.550% | 29.674% | 31.297% | 82.332% | 97.463% | 97.907% | 98.507% | | 100% | 99.360% | 99.147% |
| LG2 POA ext polpro ZONE e... | 35.272% | 34.259% | 34.259% | 31.685% | 19.929% | 19.929% | 20.107% | 20.107% | 27.928% | 27.928% | 27.798% | 28.108% | 29.369% | 29.550% | 29.674% | 31.297% | 82.332% | 97.463% | 97.907% | 98.507% | 100% | | 99.360% | 99.147% |
| AP2 POA polpro zone extrac... | 35.272% | 34.259% | 34.259% | 31.868% | 19.929% | 19.929% | 20.107% | 20.107% | 27.928% | 27.928% | 27.798% | 28.108% | 29.369% | 29.550% | 29.550% | 31.297% | 82.332% | 97.023% | 97.693% | 98.294% | 99.360% | 99.360% | | 99.787% |
| HG2 POA pol pro zone extra... | 35.272% | 34.259% | 34.259% | 31.868% | 19.929% | 19.929% | 20.107% | 20.107% | 27.928% | 27.928% | 27.798% | 28.108% | 29.369% | 29.550% | 29.465% | 31.297% | 82.332% | 97.240% | 97.480% | 98.081% | 99.147% | 99.147% | 99.787% | |

# Appendix 19

*Table 36 : Matrix of protein of consensus sequences and references waikavirus (Polymerase-protéase QA-STOP codon)*

| | LP2 POA polp... | AP2 POA pol ... | HP2 POA pol ... | AP1 pol pro z... | AG min polpr... | AG2 POA polp... | LG2 POA ext ... | LG2 POA ext ... | AP2 POA polp... | HG2 POA pol ... |
|---|---|---|---|---|---|---|---|---|---|---|
| LP2 POA polpro zone | | 83.256% | 41.136% | 42.333% | 42.580% | 42.798% | 42.918% | 42.798% | 42.450% | 42.798% |
| AP2 POA pol pro zone | 83.256% | | 38.204% | 40.642% | 40.915% | 41.542% | 41.542% | 41.284% | 41.167% | 41.542% |
| HP2 POA pol pro zone extra... | 41.136% | 38.204% | | 69.170% | 70.091% | 70.943% | 71.318% | 71.068% | 70.955% | 71.193% |
| AP1 pol pro zone | 42.333% | 40.642% | 69.170% | | 96.628% | 97.225% | 97.598% | 97.722% | 97.231% | 97.349% |
| AG min polpro zone | 42.580% | 40.915% | 70.091% | 96.628% | | 98.202% | 98.566% | 98.687% | 98.213% | 98.202% |
| AG2 POA polpro zone | 42.798% | 41.542% | 70.943% | 97.225% | 98.202% | | 99.150% | 98.908% | 98.676% | 98.786% |
| LG2 POA ext polpro zone | 42.918% | 41.542% | 71.318% | 97.598% | 98.566% | 99.150% | | 99.757% | 99.283% | 99.393% |
| LG2 POA ext polpro ZONE | 42.798% | 41.284% | 71.068% | 97.722% | 98.687% | 98.908% | 99.757% | | 99.162% | 99.150% |
| AP2 POA polpro zone | 42.450% | 41.167% | 70.955% | 97.231% | 98.213% | 98.676% | 99.283% | 99.162% | | 99.404% |
| HG2 POA pol pro zone | 42.798% | 41.542% | 71.193% | 97.349% | 98.202% | 98.786% | 99.393% | 99.150% | 99.404% | |