AIX-MARSEILLE UNIVERSITÉ

ÉCOLE DOCTORALE SCIENCES DE LA VIE ET DE LA SANTÉ (ED 62)

LABORATOIRE INFORMATION GÉNOMIQUE ET STRUCTURALE

Exploration de la diversité virale dans les échantillons environnementaux

Thèse présentée pour obtenir le grade universitaire de
Docteur en Sciences de la Vie et de la Santé
Discipline : Biologie

Présentée et soutenue publiquement
le 19 Janvier 2017 par :

Elisabeth FABRE

Directeur de thèse :
Dr. Chantal ABERGEL

JURY :

Pr. Christophe ROBAGLIA	Président
Dr. Yves GAUDIN	Rapporteur
Dr. Hervé MOREAU	Rapporteur
Dr. Emmanuelle GUILLOT-COMBE	Examinateur
Dr. Chantal ABERGEL	Directeur de thèse

Remerciements

Je tiens à remercier en premier lieu Jean-Michel Claverie et Chantal Abergel qui m'ont accueillie dans leur laboratoire et m'ont fait confiance toutes ces années. Chantal, merci pour tes conseils, merci d'avoir encadré mon travail tout au long de cette thèse, et de m'avoir permis de prendre du recul pour voir l'image globale des projets, là où je ne voyais que quelques pièces du puzzle.

Je remercie également sincèrement chacun des membres de mon jury, le Pr. Christophe Robaglia, le Dr. Hervé Moreau, le Dr. Yves Gaudin et le Dr. Emmanuelle Guillot-Combe, ma tutrice auprès de la DGA, pour avoir accepté de juger mon travail.

Je remercie bien évidemment l'ensemble des membres du laboratoire que j'ai eu la chance de rencontrer au cours de ces trois années, et qui m'ont apporté leur aide lorsque j'en avais besoin, et notamment Audrey (pour nos journées mémorables de PFGE !), Sandra, Matthieu, Olivier, merci pour votre disponibilité, ça a été un plaisir de travailler avec vous sur les différents projets !

Sébastien, je te remercie tout particulièrement pour ta gentillesse (même si tu grognes souvent, je grogne encore plus fort que toi !) et pour toutes nos discussions sur les séries et le Hellfest, qui me manqueront !

Merci au club très fermé des thésards de l'IGS, dont je fus la doyenne jusqu'à ce jour (après le départ du Dr. Doutre !) : Adrien et Lucie, merci pour tout, ça a été un plaisir de partager l'open-space avec vous. Merci Adrien pour les nombreux gâteaux, tartes, chouquettes, et pour tes blagues bien-sûr ! Merci Lucie pour nos discussions : ta gentillesse et ton originalité me manqueront. Je vous souhaite de réussir, et j'attends avec impatience mon invitation pour vos soutenances. Lucie, je te passe le flambeau, tu es maintenant la doyenne des thésards !

Merci aux membres de la plateforme de microscopie électronique de l'IBDM qui m'ont aidée et formée au cours de ces années : Jean-Paul Chauvin, Aïcha Aouane et Fabrice Richard. Aïcha, reste comme tu es, tu es vraiment une femme exceptionnelle !

Je tiens sincèrement à adresser un merci tout particulier à la dream-team de l'IGS, le duo de choc : Lionel et Jean-Marie ! Merci de m'avoir donné le sourire toutes ces années, vous êtes formidables, ne changez rien !

Un merci vraiment spécial à ces "collègues de travail" qui sont devenues des amies : Julia, Dorothée et Nadège. Merci les filles pour votre soutien et vos encouragements ! Nadège je tiens à t'adresser un merci tout particulier : tu m'as formée dès mon arrivée au laboratoire, tu es devenue une véritable amie, et tu m'as montré qu'il y avait toujours de l'espoir. Je sais que je peux compter sur toi en toute circonstance, tu es toujours là pour moi, et tu sais que la réciproque est vraie ! Namasté !

Un merci également à mes “copines de fac” que je connais depuis 8 ans, déjà ! Victoria, Lætitia et Céline, vous êtes de véritables amies et vous êtes au top : bonne chance pour la suite !

Enfin, je tiens sincèrement à remercier ma famille, sans qui je ne serais pas là aujourd’hui. Leur amour, leur soutien et leurs encouragements ont été d’une grande aide au cours de ces années, et m’ont motivée à poursuivre, quoiqu’il arrive. Vous êtes ce que j’ai de plus cher en ce monde, vous êtes ma priorité absolue et ça ne changera jamais ! Et bien évidemment, un merci tout particulier à Johannick, qui partage ma vie depuis toutes ces années : tu n’as jamais douté de mes capacités à réussir, et tu as toujours su trouver les mots justes pour me faire avancer et grandir. Ton amour, ton sourire, tes encouragements et ton soutien sans failles m’ont réellement motivée à poursuivre dans cette voie, et je t’en remercie vraiment.

“La science est un jeu dont la règle du jeu consiste à trouver quelle est la règle du jeu”
François Cavanna (1923-2014)

Table des matières

Remerciements **iii**

Nomenclature **xiii**

1 Introduction **1**
1.1 Historique des virus . . . 1
1.1.1 Découverte des premiers virus . . . 1
1.1.2 Définition des virus . . . 4
1.2 Classification des virus . . . 7
1.2.1 La classification de Baltimore . . . 8
1.2.2 La classification de l'*International Committee on Taxonomy of Viruses* (ICTV) . . . 14
1.3 Les grands virus à ADN nucléo-cytoplasmiques (NCLDV) . . . 15
1.4 Les virus géants . . . 20
1.4.1 Découverte du premier virus géant . . . 20
1.4.2 Écologie et virus (géants) . . . 25
1.4.3 Recherche des virus géants dans l'environnement . . . 27
1.4.4 Hôtes des virus géants . . . 30
Les Amibes . . . 30
Les Algues . . . 33
Les échantillons cliniques . . . 34
1.4.5 Les différentes familles de virus géants . . . 37
Les *Mimiviridae* . . . 37
Les virophages . . . 40
Les *Marseilleviridae* . . . 44
Les *Pandoraviridae* . . . 46
Les *Pithoviridae* . . . 50
Mollivirus sibericum . . . 54
1.4.6 Évolution et origine des virus géants . . . 57
1.5 Objectifs de cette thèse . . . 63

2 Matériel et méthodes **65**
2.1 Isolement des virus . . . 65

2.1.1 Culture des amibes . . . 65
2.1.2 Traitement des échantillons environnementaux . . . 66
2.1.3 Co-cultures . . . 66
2.2 Purification des virus . . . 67
2.2.1 Purification sur gradient de sucrose . . . 67
2.2.2 Titration et comptage des virus . . . 68
2.2.3 Clonage des virus sur cellule unique . . . 69
2.3 Cycles infectieux . . . 69
2.4 Fluorescence des cellules infectées par Noumeavirus . . . 70
2.5 Microscopie électronique . . . 71
2.5.1 Inclusions en résine . . . 71
2.5.2 Coloration négative . . . 72
2.6 Génomique . . . 72
2.6.1 Extraction génomique . . . 72
2.6.2 Séquençage des génomes . . . 73
2.6.3 Assemblage et annotation des génomes . . . 74
Pandoravirus . . . 74
Noumeavirus . . . 74
2.6.4 Phylogénie . . . 75
Pandoravirus . . . 75
Noumeavirus . . . 75
2.7 Protéomique . . . 75
2.7.1 Extraction de protéines . . . 75
2.7.2 Profils protéiques . . . 76
2.7.3 Protéome de surface des particules (Surfome) . . . 76
2.7.4 Digestion des protéines extraites . . . 76
2.7.5 Analyses Nano-LC-MS/MS . . . 77
2.7.6 Analyses des protéomes . . . 77
2.8 Transcriptome des cellules au cours du cycle infectieux . . . 78
2.8.1 Extraction des ARN . . . 78
2.8.2 Enrichissement en ARN messagers . . . 78
2.9 Études de la composition des parois des Pandoravirus et Mollivirus . . . 79
2.9.1 Incubations en acide orthophosphorique . . . 79
2.9.2 Incubations en cellulase . . . 79
2.9.3 Ouverture des virions et digestions enzymatiques . . . 80
Ouverture des particules virales . . . 80
Digestions enzymatiques . . . 80
2.9.4 Analyse des sucres solubles . . . 81
2.9.5 Marquages au Calcofluor-white . . . 81
2.9.6 Digestions de coupes . . . 82

3 Caractérisation d'un nouveau *Marseilleviridae* et études comparatives 83
3.1 Résultats . . . 83
3.1.1 Découverte d'un nouveau membre de la famille des *Marseilleviridae* 83
Clonage du virus sur cellule unique . . . 85
Titration des clones de Noumeavirus . . . 86
Profils protéiques des clones de Noumeavirus . . . 87
3.1.2 Analyse génomique de Noumeavirus . . . 87
Analyse des protéines impliquées dans la maturation des transcrits 90
3.1.3 Analyse protéomique de Noumeavirus . . . 91
Les protéines de l'hôte partagées par les virions . . . 92
Les protéines virales identifiées dans les virions . . . 94
3.1.4 Étude du cycle infectieux de Noumeavirus . . . 96
Étude du cycle infectieux de Noumeavirus au microscope à épifluorescence . . . 96
Étude du cycle infectieux de Noumeavirus au MET . . . 99
3.2 Discussion . . . 104
3.3 Conclusion . . . 109

4 Caractérisation de nouveaux *Pandoraviridae* et études comparatives 111
4.1 Résultats . . . 111
4.1.1 Découverte d'un nouveau membre de la famille des *Pandoraviridae* 111
Clonage des Pandoravirus sur cellule unique . . . 113
4.1.2 Analyse génomique des Pandoravirus . . . 114
Extraction des génomes . . . 114
Analyses des génomes . . . 115
Détermination du *core-genome* des Pandoravirus . . . 118
4.1.3 Analyse protéomique des Pandoravirus . . . 119
Protéomes des particules purifiées . . . 119
Étude des protéines de surface des particules . . . 121
4.1.4 Comparaison des cycles infectieux des Pandoravirus . . . 124
4.2 Discussion . . . 133
4.3 Conclusion . . . 136

5 Études de la composition des *Pandoraviridae* et *M. sibericum* 139
5.1 Contexte . . . 139
5.2 Résultats et discussion . . . 141
5.3 Conclusion . . . 149

6 Conclusion et Perspectives 151

Bibliographie 153

Annexe A Noumeavirus replication relies on a transient remote control of its host nucleus **167**

Annexe B Noumeavirus : Supplementary Material and Methods **201**

Résumé/Abstract

Table des figures

Chapitre 1

1.1 Histoire de la variole 1
1.2 Histoire de la rage 2
1.3 Systèmes de filtration Chamberland 2
1.4 Histoire du virus de la mosaïque du tabac 3
1.5 Histoire du virus de la fièvre jaune 4
1.6 Différents types de génomes viraux 6
1.7 Différentes formes de capsides virales 6
1.8 Réplication des virus à ADN 8
1.9 Réplication des virus à ARN simple brin à polarité positive 10
1.10 Réplication des virus à ARN simple brin à polarité positive, avec ARN sub-génomiques 11
1.11 Réplication des virus à ARN simple brin à polarité négative 12
1.12 Réplication des rétrovirus et des pararétrovirus 13
1.13 Arbre phylogénétique des familles appartenant au groupe des NCLDV . . 15
1.14 Virion de la variole (Smallpox) observé au microscope électronique à transmission 16
1.15 African Swine Fever Virus (ASFV) observé au microscope électronique . . 17
1.16 Invertebrate Iridescent virus 6 (IIV-6) observé au microscope électronique 18
1.17 *Paramecium bursaria* chlorella virus 1 (PBCV-1) observé au microscope électronique 19
1.18 *Trichoplusia ni* ascovirus 2c (TnAV-2c) observé au microscope électronique 20
1.19 Mimivirus observé au microscope électronique 21
1.20 Couverture du journal American Scientist mettant en scène Mimivirus . . 22
1.21 Stargate de Mimivirus observée au microscope à balayage 23
1.22 Carte du chromosome de Mimivirus 24
1.23 Biomasse et abondance des procaryotes, des protistes et des virus dans les océans 25
1.24 L'algue unicellulaire *Emiliania huxleyi* observée au microscope à balayage 26
1.25 Efflorescence (ou bloom) de l'algue *Emiliania huxleyi* observée par satellite 27
1.26 Carte représentant le parcours effectué lors de la campagne Tara Oceans 28
1.27 *Megavirus chilensis* observé au microscope électronique 29

1.28 L'amibe *Acanthamoeba* observée au microscope électronique 30
1.29 Faustovirus observé en cryo-microscopie électronique 32
1.30 Organe de fructification de l'amibe *Dictyostelium discoideum* 32
1.31 *Cafeteria roenbergensis* virus observé au microscope électronique 34
1.32 Arbre phylogénétique des NCLDV qui comprend les différentes familles de virus géants 37
1.33 Mimivirus et Mégavirus dans la même vacuole observés au microscope électronique 38
1.34 Arbre phylogénétique de la famille des *Mimiviridae* 39
1.35 Représentation schématique du cycle infectieux de Mimivirus 40
1.36 Sputnik à l'intérieur de Mamavirus observé au microscope électronique . 41
1.37 Schéma du cycle de réplication de Sputnik opposé à celui d'un virus satellite 42
1.38 Samba virus et de son virophage observés au microscope électronique . . 43
1.39 Marseillevirus observé au microscope électronique 45
1.40 Arbre phylogénétique de la famille des *Marseilleviridae* 46
1.41 Couverture du journal Science mettant en avant *Pandoravirus salinus* . . 47
1.42 Représentation schématique du cycle infectieux des Pandoravirus dans l'amibe 49
1.43 L'endosymbionte KLaHel observé au microscope électronique 50
1.44 *Pithovirus sibericum* observé au microscope électronique 51
1.45 Représentation schématique du cycle infectieux de *Pithovirus sibericum* dans l'amibe 52
1.46 Parasite KC 5/2 dans l'amibe observé au microscope électronique 53
1.47 Comparaison de l'organisation des génomes de *P. sibericum* et de *P. massiliensis* 54
1.48 *M. sibericum* observé au microscope électronique à transmission et à balayage, et au microscope optique 55
1.49 Représentation schématique du cycle infectieux de *Mollivirus sibericum* dans l'amibe 56
1.50 Arbre phylogénétique de la gamme des hôtes eucaryotes des grands virus à ADN 58
1.51 Redéfinition des virus 61

Chapitre 2

2.1 Cellule de Malassez 68
2.2 Tube Microvette 71

Chapitre 3

3.1 Observation du parasite au MET 84

3.2 Purification du virus 85
3.3 Observation du virus au microscope optique 85
3.4 Calcul du $TCID_{50}$ de Noumeavirus 86
3.5 Profils protéiques des clones de Noumeavirus 87
3.6 Répartition des homologues des ORFs de Noumeavirus 88
3.7 Arbre phylogénétique de la famille des *Marseilleviridae* 89
3.8 Protéines de l'hôte partagées par Noumeavirus et Melbournevirus 93
3.9 Protéines impliquées dans la transcription et la maturation des transcrits 95
3.10 Modifications du noyau des cellules au cours du cycle infectieux de Noumeavirus 97
3.11 Cellules infectées par Noumeavirus observées au microscope à épifluorescence 99
3.12 Étapes du cycle infectieux de Noumeavirus observées au microscope optique 100
3.13 Observation du cycle infectieux de Noumeavirus au MET 101

Chapitre 4

4.1 Particules de Pandoravirus observées au microscope optique et électronique112
4.2 Profils protéiques des clones de *P. neocaledonia* 113
4.3 Nombre de gènes présents dans les génomes des Pandoravirus 116
4.4 Arbre phylogénétique de la famille des *Pandoraviridae* 117
4.5 Histogramme des différentes catégories d'homologues des Pandoravirus . 118
4.6 Courbes du *core-genome* et du *pan-genome* des Pandoravirus 119
4.7 Graphique des protéines identifiées dans les surfomes de *P. quercus* . . . 122
4.8 Particules de Pandoravirus observées au MET après digestion à la trypsine 123
4.9 Cycle de *P. neocaledonia* observé au microscope optique 125
4.10 Étape d'internalisation des Pandoravirus observée au MET 126
4.11 Étape de libération du génome des Pandoravirus observée au MET 127
4.12 Étape de modification du noyau des cellules hôtes au cours du cycle des Pandoravirus observée au MET 128
4.13 Observation d'une cellule infectée par *P. neocaledonia* au microscope à épifluorescence 128
4.14 Étape de synthèse des Pandoravirus observée au MET 130
4.15 Observation d'amas au MET au cours de la synthèse des virions 131
4.16 Cellules infectées par des Pandoravirus en cours de lyse observées au MET 132

Chapitre 5

5.1 Pandoravirus et *M. sibericum* observés au MET 139
5.2 Kystes d'*A. castellanii* observés au MET 140

5.3 Kyste de l'amibe *F. pleistocenica* observé au MET 141
5.4 Transcriptome d'un gène de la cellulose synthase d'*A. castellanii* 142
5.5 Graphiques correspondant aux sucres libérés dans les cellules *A. castellanii* 142
5.6 Graphiques correspondant aux sucres libérés chez les Pandoravirus et Mimivirus 143
5.7 Graphique correspondant aux sucres libérés chez Mollivirus 143
5.8 Particules de Pandoravirus observées au microscope optique 144
5.9 Kystes d'*A. castellanii* marqués au Calcofluor-white et observés au microscope optique 146
5.10 Coupes de kystes d'*A. castellanii* observées au MET 147
5.11 Coupes de *P. neocaledonia* observées au MET 147

Nomenclature

μm	micro mètre
ADN	Acide Désoxyribonucléique
ARN	Acide Ribonucléique
ARNm	ARN messager
ARNt	ARN de transfert
ASFV	African Swine Fever Virus
ASSEMBLE	*Association of European Marine Biological Laboratories*
CeV	*Chrysochromulina ericina* virus
CroV	*Cafeteria roenbergensis* virus
CTAB	Cetyltrimethylammonium bromide
DMSO	diméthylsulfoxyde
dsDNA	*double-strand Deoxyribonucleic acid* ou ADN double-brin
dsRNA	*double-strand Ribonucleic acid* ou ARN double-brin
DTT	Dithiothreitol
EhV	*Emiliania huxleyi* virus
FLA	*Free Living Amoeba* ou amibe sous forme libre
g	gramme
GBM	Giant Blood Marseillevirus
GOS	*Global Ocean Sampling*
HPAEC-PAD	*High Pressure Anion Exchange Chromatography coupled with Pulsed Amperometric Detection*
hpi	heure post-infection
iBAQ	*intensity-Based Absolute Quantification* ou quantification absolue basée sur l'intensité
ICTV	*International Committee on Taxonomy of Viruses*

kb	kilo base
kDa	kilo Daltons
LECA	*Last Eukaryotic Common Ancestor*
LUCA	*Last Universal Common Ancestor*
M	molaire
Mb	Méga base
MCP	*Major Capsid Protein* ou protéine majeure de capside
MET	Microscope Électronique à Transmission
mL	millilitre
mM	millimolaire
MOI	*Multiplicity of Infection*
MORN	*Membrane Occupation and Recognition Nexus*
NCLDV	*Nucleo-Cytoplasmic Large DNA Viruses*
nm	nanomètre
ORF	*Open Reading Frame* ou cadre de lecture ouvert
pb	paires de bases
PBS	*Phosphate Buffered Saline*
PgV	*Phaeocystis globosa* virus
pi	post-infection
PPYG	Proteose-Peptone-Yeast extract-Glucose
RPM	Revolutions Per Minute, ou tours par minute
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	*sodium dodecyl sulfate polyacrylamide gel electrophoresis* ou électrophorèse sur gel de polyacrylamide en présence de dodécylsulfate de sodium
ssDNA	*single-strand Deoxyribonucleic acid* ou ADN simple brin
ssRNA	*single-strand Ribonucleic acid* ou ARN simple brin
$TCID_{50}$	*50% Tissue Culture Infective Dose*
TFA	*Trifluoroacetic acid* ou acide trifluoroacétique
UTR	*Untranslated Transcribed Region* ou région non traduite
VIH	Virus de l'Immunodéficience Humaine
VMT	Virus de la Mosaïque du Tabac ou *Tobacco Mosaic disease Virus*

1 Introduction

1.1 Historique des virus

1.1.1 Découverte des premiers virus

Les maladies infectieuses, telles que la variole ou la rage, sont connues depuis l'Antiquité. Au XVI^{e} siècle, des médecins chinois pratiquaient la variolisation : ils récupéraient le contenu des vésicules de patients atteints par la maladie, qu'ils laissaient sécher à l'air libre, puis inoculaient cette forme atténuée de la maladie à des personnes saines, afin de les immuniser (BERCHE, 2007). Cette technique fut importée en occident au début du XVIII^{e} siècle par Lady Mary Wortley Montagu. A la fin du XVIII^{e} siècle, Edward Jenner, un médecin anglais, utilisa cette technique en inoculant la variole bovine ou "*cowpox*" à des patients, dans le but de les immuniser contre la maladie et nomma cette procédure "vaccination" (RIEDEL, 2005). Ce principe de vaccination était alors employé de façon préventive, sur des personnes saines, afin de les immuniser contre les maladies.

FIGURE 1.1 – Histoire de la variole : à gauche, un patient atteint de variole, au milieu, Lady Mary Wortley Montagu (1689-1762), et à droite, Edward Jenner (1749-1823).

En 1885, Louis Pasteur et son collaborateur Emile Roux, appliquèrent ce principe d'atténuation sur le virus de la rage, non pas pour immuniser la population, mais afin de traiter des personnes atteintes. Pasteur inocule par passages successifs la rage à des lapins, afin de "fixer" le virus, puis récupère les moelles épinières et les laisse sécher à l'air libre. Il réalise ensuite des broyats de moelle, qu'il injecte à un jeune patient, Joseph Meister, qui a été mordu par un chien enragé. Après une dernière injection de contrôle, du virus non atténué, qui est un succès, Pasteur en conclut que le vaccin est efficace, et vaccine de nombreuses personnes ayant été mordues. En 1886, soit moins d'un an plus tard, plus de 300 personnes ont été vaccinées avec succès (BORDENAVE, 2003).

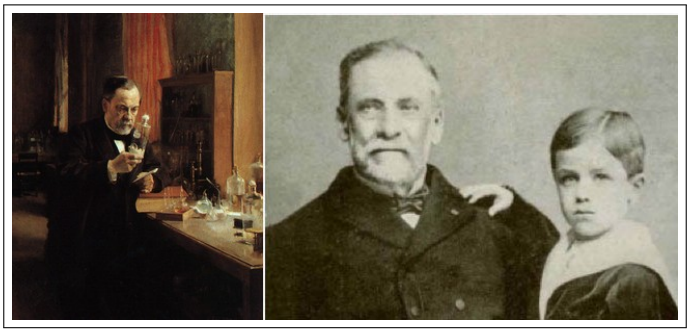


FIGURE 1.2 – Histoire de la rage : à gauche, Louis Pasteur (1822-1895), et à droite, Louis Pasteur et Joseph Meister.

Le microbiologiste français Charles Chamberland, qui travaille avec Louis Pasteur sur la stérilisation, développe en 1884, lors d'une épidémie de fièvre typhoïde, un filtre constitué de pores d'une taille inférieure à celle d'une bactérie, appelé "filtre de Chamberland". Ce filtre, fabriqué à partir d'une bougie de porcelaine poreuse, permet de filtrer les liquides, et notamment l'eau potable, afin d'en retenir les micro-organismes présents, et des les éliminer (CHAMBERLAND, 1884). Ce filtre servira par la suite d'instrument de recherche en microbiologie.

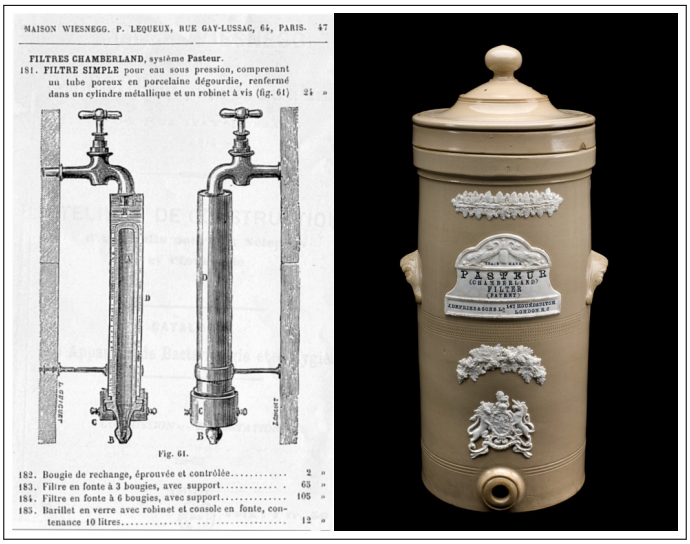


FIGURE 1.3 – Systèmes de filtration Chamberland : à gauche, une illustration de l'intérieur d'un filtre, et à droite, une photographie d'un filtre utilisé pour stériliser l'eau.

En 1876, Adolf Mayer, un microbiologiste, réalise des observations et des expériences sur des plant de tabac touchés par une maladie qui colore les feuilles de façon anormale et réduit la croissance des plantes, qu'il appelle la mosaïque du tabac. Il démontre alors la nature infectieuse de cette maladie, en inoculant à des plantes saines un "jus" réalisé en broyant des plantes malades dans un peu d'eau. Après avoir cherché si cette maladie était due à une bactérie ou un champignon, il en conclut qu'il s'agit probablement d'une bactérie non visible et non cultivable.

En 1892, Dimitri Ivanovsky démontra que cet agent infectieux n'était pas observable au microscope optique, et qu'il était capable de passer à travers le filtre de Chamberland

(donc effectivement plus petit qu'une bactérie) tout en restant toujours infectieux. Il en conclut à son tour qu'il s'agit d'une bactérie plus petite que les pores du filtre de Chamberland ou bien d'une toxine produite par une bactérie.

Ces résultats furent confirmés par le microbiologiste Martinus Beijerinck en 1898, qui nomma cette nouvelle forme d'agent infectieux *«contagium vivum fluidum»* ou fluide vivant contagieux, et qui réintroduit alors le mot *virus*, provenant du latin et signifiant poison ou venin (LECOQ, 2001).

Ce virus de la mosaïque du tabac (VMT ou TMV en anglais) sera utilisé par la suite comme modèle des études réalisées sur les virus : il fut cristallisé par Wendell Stanley en 1935 (STANLEY, 1938) et observé au microscope électronique à transmission en 1939 (KAUSCHE *et al.*, 1939).

FIGURE 1.4 – Histoire du virus de la mosaïque du tabac : en haut à gauche, Adolf Mayer (1843-1942), en haut au milieu, Dimitry Ivanovsky (1864-1920), en haut à droite, Martinus Beijerinck (1851-1931), en bas à gauche, une feuille de tabac infectée par le virus (©INRA), et en bas à droite le virus de la mosaïque du tabac observé au microscope électronique à transmission (SCHOLTHOF, 1997).

Il faudra attendre les années 1900 pour que le premier virus humain, l'agent responsable de la fièvre jaune, soit étudié par Walter Reed, James Carroll, Aristides Agramonte et Jesse Lazear : une fois encore, la description de cet agent infectieux se limite à un microbe, non visible au microscope optique et capable de passer à travers le filtre de Chamberland (BALDACCHINO, 2011).

Max Theiler, virologiste et professeur de médecine à Harvard, contribua à isoler le virus responsable de la fièvre jaune et développa un vaccin efficace en 1937 : le prix Nobel de Physiologie ou Médecine lui fut attribué en 1951 pour ses découvertes (NORRBY, 2007).

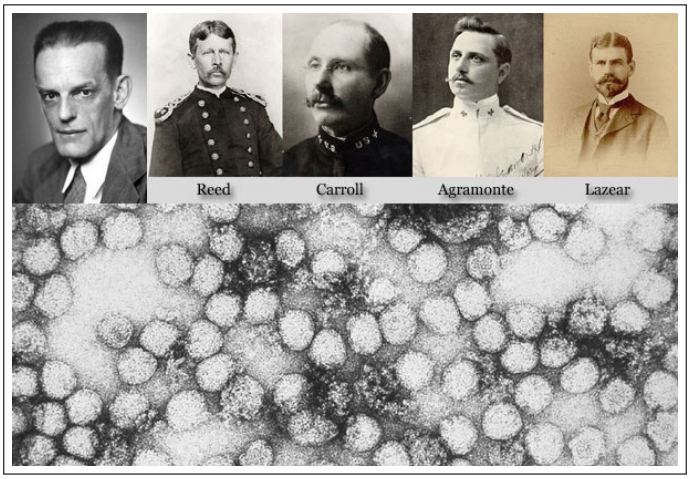


FIGURE 1.5 – Histoire du virus de la fièvre jaune : en haut, Max Theiler (1899-1972), Walter Reed (1851-1902), James Carroll (1854-1907), Aristides Agramonte (1868-1931), et Jesse Lazear (1866- décédé de la fièvre jaune en 1900), et en bas, le virus de la fièvre jaune observé au microscope électronique à transmission (©CDC).

Dans les années 1920, les données sur les virus connus de l'époque ont fait l'objet d'une publication, "*Filterable Viruses*", éditée par Thomas Milton Rivers, bactériologiste et virologiste (RIVERS, 1927). Il donne alors une première définition de ce qu'est un virus, en disant lors d'un congrès de la *Society of American Bacteriology* que *«Les virus semblent être des parasites obligatoires dans le sens où leur réplication est dépendante de cellules vivantes»*.

1.1.2 Définition des virus

Avant les années 1950, la définition de virus se limitait donc à trois observations : les virus sont invisibles au microscope optique, ne sont pas cultivables et passent à travers le filtre de Chamberland.

C'est André Lwoff qui, dans un article publié en 1957, donne une définition plus complète des virus : *«Les virus sont infectieux et potentiellement pathogènes ; ce sont des entités nucléoprotéiques possédant un seul type d'acide nucléique (ARN ou ADN) ; ils sont reproduits par la cellule à partir de leur matériel génétique ; ils sont incapables de croître et de se diviser ; ils sont dépourvus de système de Lipmann»* (LWOFF, 1957).

De cette définition se dégagent quatre critères permettant de différencier les virus des autres micro-organismes :

i. **les virus ne contiennent qu'un seul type d'acide nucléique, ARN ou ADN.**
   En effet, selon Lwoff, les virus ne contiennent qu'un seul type d'acide nucléique, ARN ou ADN, à la différence des autres micro-organismes qui contiennent les deux types d'acides nucléiques.

ii. **les virus ne sont reproduits qu'à partir de leur seul acide nucléique**, contrairement aux autres micro-organismes, reproduits à partir de la somme de tous leurs constituants (noyau, cytoplasme, structures cytoplasmiques, ...).

iii. **les virus sont incapables de croître et de se diviser (division binaire)**.
Les micro-organismes se multiplient en répliquant leur matériel génétique. Cette multiplication va de pair avec la croissance, qui est le résultat de la synthèse du matériel cellulaire, et qui se termine par une division binaire de l'organisme. Selon la définition de Lwoff, les virus sont seulement capables de répliquer leur acide nucléique, car *«il n'y a rien qui ne puisse se diviser»*.

iv. **les virus ne possèdent pas de système de Lipmann**.
Les micro-organismes possèdent un système d'enzymes capables de convertir les nutriments en énergie nécessaire à la synthèse biochimique (Système de Lipmann), absent chez les virus. L'énergie pour la synthèse des virus est alors procurée par la cellule hôte : les virus sont donc des parasites intracellulaires obligatoires.

Au fil du temps et des découvertes, la définition du mot *virus* a évolué et une définition plus simple est proposée par Luis Villarreal en 2005 (VILLARREAL, 2005) : *«Un virus est un parasite génétique qui utilise les systèmes cellulaires pour sa propre réplication»*.

Cette définition ne fait pas référence à l'identité moléculaire du virus, ni aux gènes viraux, ni au rôle de ceux-ci dans le cycle de réplication du virus.

En revanche, cette définition fait référence au *virion*, qui doit être différencié du virus, comme suggéré par Claudiu Bândea en 1983 (BÂNDEA, 1983) : le virion est considéré comme une "spore" ou comme une forme (un véhicule) de propagation du virus.

Il en découle qu'un virion est la forme extra-cellulaire et l'unité structurale du virus, qui est composé :

— **d'un génome**
Ce génome est composé d'ADN ou d'ARN, qui peut être sous plusieurs formes : linéaire ou circulaire, simple ou double brin, à polarité positive ou négative pour les génomes à ARN, segmenté ou non.

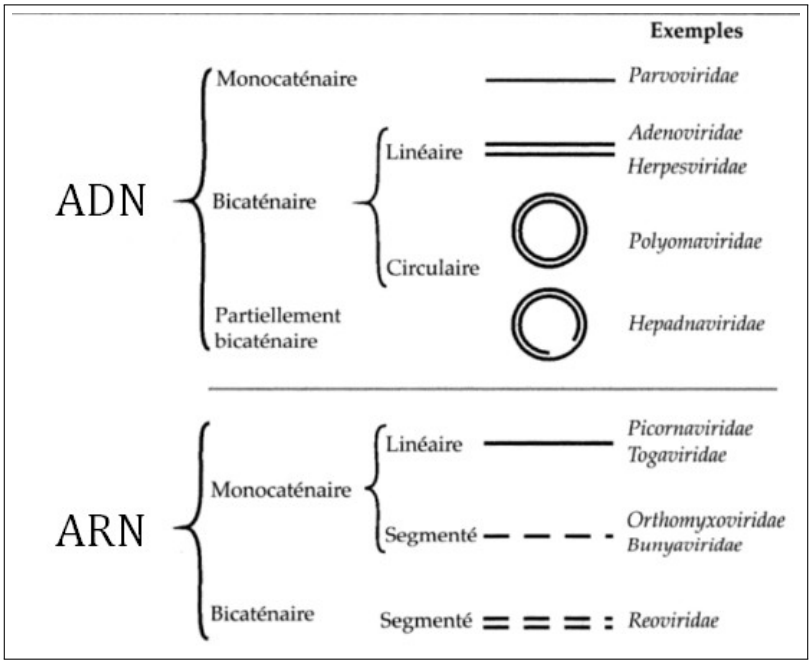


FIGURE 1.6 – Différents types de génomes viraux : ADN ou ARN, monocaténaire ou bicaténaire, segmenté ou non segmenté, linéaire ou circulaire, avec un exemple de virus pour chaque type de génome (MAMMETTE, 2002).

— **d'une capside**

La capside entoure et protège le matériel génétique. Elle est assemblée à partir de la polymérisation d'une ou d'un petit nombre de sous-unités protéiques, ou capsomères, qui forment une certaine symétrie : hélicoïdale (par exemple le virus de la mosaïque du tabac, ou le virus de la rage), icosaédrique (par exemple l'Herpes simplex virus, ou le Papillomavirus humain) ou complexe (par exemple le bactériophage qui possède une "tête" icosaédrique et une "queue" hélicoïdale, ou les poxvirus qui ont une morphologie parallélépipédique)

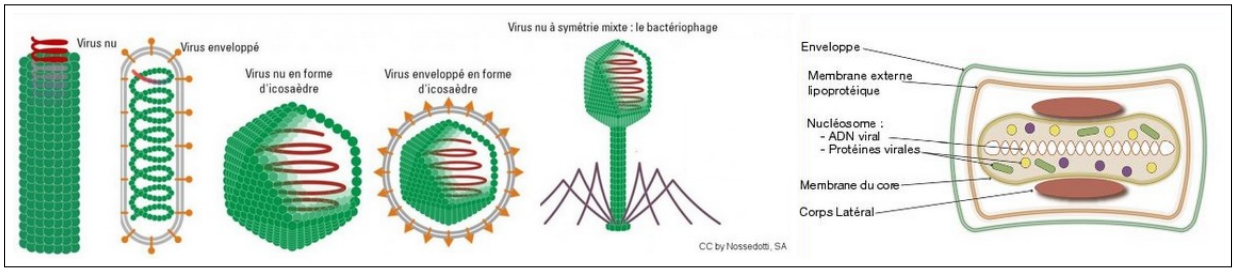


FIGURE 1.7 – Différentes formes de capsides virales : Virus à symétrie hélicoïdale et virus à symétrie hélicoïdale enveloppé, virus à symétrie icosaédrique et virus à symétrie icosaédrique enveloppé, virus à symétrie complexe, comme les bactériophages et les Poxvirus (DURAFFOUR *et al.*, 2008).

— **d'une enveloppe** (éventuellement)
Cette enveloppe est dérivée des membranes cellulaires et présente à sa surface des glycoprotéines, qui peuvent par exemple, faciliter l'attachement du virus à la surface de la cellule hôte, ou sa pénétration dans le cytoplasme de celle-ci (comme par exemple, l'hémagglutinine du virus de la grippe). Cette enveloppe est acquise par bourgeonnement des virus néoformés à travers les membranes cellulaires : la membrane nucléaire, la membrane cytoplasmique ou la membrane du réticulum endoplasmique.
Les virus qui possèdent cette enveloppe sont les virus "enveloppés" et ceux qui en sont dépourvus sont appelés virus "nus" : les virus enveloppés sont toutefois plus fragiles que les virus nus.

Ces trois critères sont essentiels pour établir une classification des virus, ou taxonomie.

1.2 Classification des virus

Historiquement, les virus étaient classés selon les maladies ou les symptômes cliniques qu'ils causaient. Toutefois, un virus peut causer plusieurs symptômes, et un même symptôme peut être causé par plusieurs virus, mais il existe également des virus qui ne semblent avoir aucun pouvoir pathogène. En conséquence, un système de classification basé uniquement sur la pathogénicité des virus est impossible et fut abandonné.

En 1940, Francis Holmes proposa de classer les virus en se basant sur le système de nomenclature binomiale de Linné, et sépare les virus selon trois groupes différents : les *Phaginae* (virus infectant les bactéries), les *Phytophaginae* (virus infectant les plantes) et les *Zoophaginae* (virus infectant les animaux) (HOLMES, 1940).

Cette classification, basée encore une fois sur la pathologie et non pas sur les propriétés et les caractéristiques des virus, a été abandonnée et, en 1962, André Lwoff, Robert Horne et Paul Tournier proposent un nouveau système de classification (LWOFF *et al.*, 1962). Ce système, aussi appelé système L.H.T, se base sur différentes caractéristiques des virus, comme la nature de l'acide nucléique viral (ADN ou ARN), la symétrie de la capside (icosaédrique, hélicoïdale ou complexe), la présence ou non d'enveloppe et enfin, la taille du virion et de la capside virale.

Actuellement, deux types de classification des virus servent de référence : elles ne sont pas incompatibles et peuvent être utilisées de façon complémentaire.

1.2.1 La classification de Baltimore

Proposée en 1971 par David Baltimore, un biologiste ayant obtenu le prix Nobel de Physiologie ou Médecine en 1975, cette classification se base sur le type d'acide nucléique des virions (ADN ou ARN) et sur leur stratégie de réplication (BALTIMORE, 1971).

Elle contient sept catégories, qui sont énoncées ci-dessous. Quelques exemples de virus infectant les eucaryotes viendront illustrer chaque catégorie, ainsi qu'un bref descriptif des stratégies de réplication.

i. **les virus à ADN double-brin (dsDNA).**
Ces virus ont un mécanisme de multiplication des gènes semblable à celui des gènes cellulaires. Leur cycle se compose de deux phases : une phase précoce au cours de laquelle l'ADN va être transcrit en ARN messagers (ARNm) précoces dans le noyau, qui seront ensuite traduits en protéines précoces (protéines non-structurales) dans le cytoplasme. Ces protéines vont par la suite entrer en action au cours de la phase tardive du cycle : elles sont notamment utilisées pour répliquer l'ADN génomique du virus, qui sera ensuite transcrit en ARNm tardifs, à leur tour traduits en protéines tardives (protéines structurales). Ce génome répliqué, ainsi que les protéines tardives servent à construire les virions matures.
Deux familles de virus font exception à cette règle : les Poxvirus, dont la réplication a lieu exclusivement dans le cytoplasme, et les Herpesvirus, dont le cycle comporte trois phases (une phase très précoce, une phase précoce et une phase tardive).

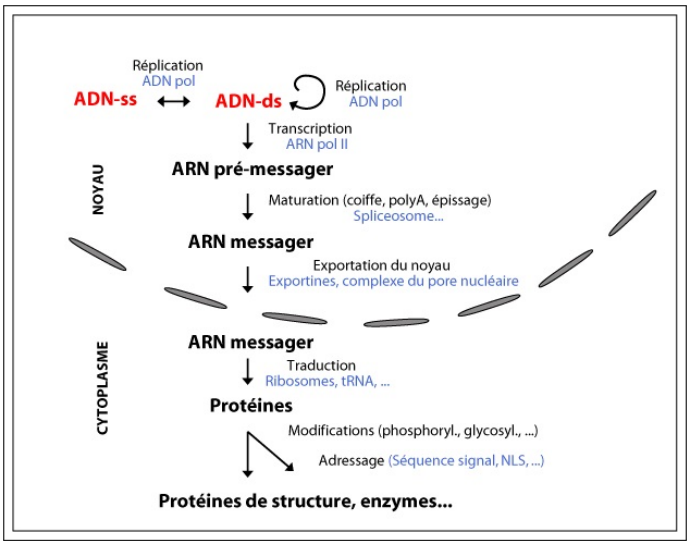


FIGURE 1.8 – Réplication des virus à ADN : l'ADN double-brin du virus est transcrit en ARN messager dans le noyau, qui sera traduit en protéines dans le cytoplasme de la cellule hôte. Pour les virus à ADN simple-brin, l'ADN polymérase de l'hôte synthétise le brin d'ADN complémentaire, avant de commencer un cycle de réplication identique à celui des virus à ADN double-brin (©UCLouvain).

Par exemple : le virus de l'herpès, le virus de la variole, le Papillomavirus

ii. **les virus à ADN simple brin (ssDNA), à polarité positive.**
Pour ces virus (*Erythrovirus*), l'ADN polymérase de la cellule hôte va synthétiser le brin complémentaire du génome viral, afin d'obtenir un génome d'ADN double-brin, qui servira d'une part, de matrice lors de la réplication, et qui sera d'autre part transcrit en ARNm, puis traduit en protéines. Les protéines synthétisées, ainsi que le génome répliqué formeront enfin les nouveaux virions (cf. Figure 1.8.). Certains de ces virus, comme les *Dependoparvovirus*, sont dépendants d'une co-infection avec un autre virus, comme un adénovirus ou un herpèsvirus, pour pouvoir se répliquer.
Par exemple : le Parvovirus B19 (*Erythrovirus*), l'Adeno-associated virus (*Dependoparvovirus*)

iii. **les virus à ARN double brin (dsRNA).**
Cette stratégie de réplication n'est appliquée que par une seule famille de virus, les *Reoviridae*, qui ont un génome segmenté.
Le génome ARN de ces virus étant bicaténaire, le brin d'ARN à polarité positive ne peut pas être directement traduit en protéines par la cellule. L'ARN polymérase virale va tout d'abord transcrire un ARNm à partir de chaque segment du génome. Cet ARNm va sortir de la particule dans le cytoplasme pour être traduit en protéines virales par la cellule, qui formeront notamment la capside du virus. En parallèle, cet ARNm (à polarité positive) va être encapsidé dans les capsides néoformées, puis copié pour la synthèse d'un nouveau génome à ARN bicaténaire.
Par exemple : le Rotavirus

iv. **les virus à ARN simple brin à polarité positive ((+)ssRNA)**
Concernant ces virus, le génome viral peut directement servir d'ARN messager et les enzymes nécessaires à la multiplication n'ont pas besoin d'être apportées par le virus : le génome de ces virus est donc infectieux.
Il existe deux stratégies de multiplication différentes pour ces virus.
Chez les *Flaviviridae* et les *Picornaviridae*, l'ARN génomique est traduit en une seule protéine (polyprotéine), qui sera clivée par la suite par des protéases virales ou cellulaires, donnant à la fois des protéines structurales qui formeront les capsides, mais également des protéines non-structurales qui interviendront dans les mécanismes de réplication et de transcription secondaire. C'est le cas de l'ARN polymérase, qui répliquera le génome viral ARN (+) en ARN à polarité négative (ou anti-génomique), qui lui-même servira de matrice pour la synthèse de nouveaux génomes ARN à polarité positive. Ces génomes ARN (+) nouvellement synthétisés pourront ensuite être encapsidés, afin de former les nouveaux virions, ou recommencer un nouveau cycle de traduction-réplication.

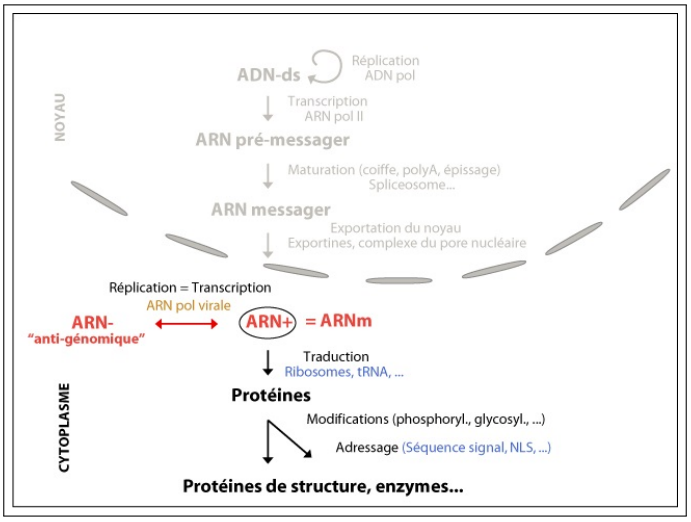


FIGURE 1.9 – Réplication des virus à ARN simple brin à polarité positive : l'ARN viral est traduit en protéines structurales et non-structurales dans le cytoplasme de la cellule hôte. Une de ces protéines, l'ARN polymérase synthétise des copies du génome viral (©UCLouvain).

Par exemple : les virus de la fièvre jaune, de l'Hépatite C et de la dengue (*Flaviviridae*)

Chez les *Astroviridae*, les *Calciviridae*, les *Coronaviridae* et les *Togaviridae*, les protéines non-structurales sont directement traduites à partir de l'extrémité 5' du génome viral (ARN +). L'extrémité 3' du génome viral est copiée en ARN à polarité négative, qui sert de matrice pour la synthèse d'un ARN à polarité positive : cet ARN "sub-génomique" (+) est alors utilisé pour la synthèse des protéines structurales. La réplication de l'ARN viral (+) se poursuit alors : le génome viral est répliqué en ARN à polarité négative, qui sert de matrice pour la synthèse de nouveaux génomes ARN (+), qui seront encapsidés ou qui recommenceront un nouveau cycle.

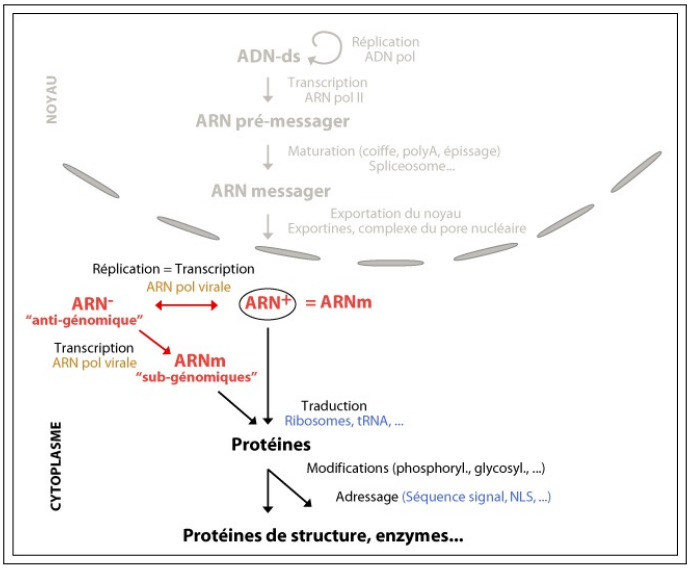


FIGURE 1.10 – Réplication des virus à ARN simple brin à polarité positive, avec ARN sub-génomiques : une des extrémités du génome est directement traduite en protéines dans le cytoplasme de la cellule hôte. L'autre extrémité est copiée, afin de synthétiser d'une part, des ARN "sub-génomiques" qui seront traduits en protéines structurales et d'autre part, de réaliser des copies du génome viral, qui pourront être encapsidées afin de former de nouveaux virions (©UCLouvain).

Par exemple : le virus de la Rubéole (*Togaviridae*)

v. **les virus à ARN simple brin à polarité négative ((-)ssRNA).**
Chez ces virus, le génome ARN (-) segmenté ou non, est copié par l'ARN polymérase virale, pour former un ARN messager (+). Cet ARNm (+) sert ensuite à la synthèse des protéines virales (non-structurales et structurales). Les protéines non-structurales, dont l'ARN polymérase, vont être utilisées à la fois pour copier le génome viral en ARNm (+), qui servira à la synthèse de nouvelles protéines, et en parallèle, en ARN (+), qui servira de matrice pour la synthèse de nouveaux génomes ARN (-). Les génomes nouvellement synthétisés, ainsi que les protéines structurales formeront alors les nouveaux virions.

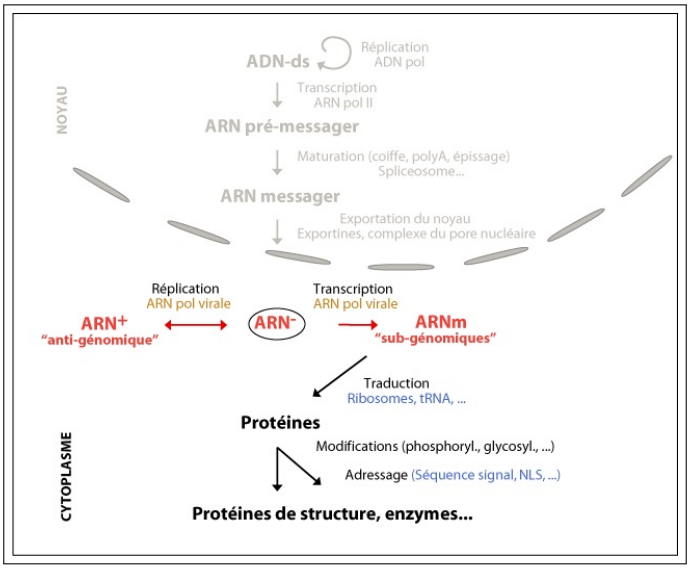


FIGURE 1.11 – Réplication des virus à ARN simple brin à polarité négative : l'ARN génomique viral est copié en ARN messager, qui est traduit en protéines dans le cytoplasme de la cellule. L'ARN polymérase, une de ces protéines, va synthétiser des copies de ce génome qui seront ensuite soit traduites en protéines, soit intégrées dans ces protéines pour former des nouveaux virions (©UCLouvain).

Par exemple : le virus de la grippe, le virus Ebola, le virus de la Rougeole, le virus de la Rage, le virus de la fièvre de Lassa

vi. **les rétrovirus à ARN simple brin (ssRNA-RT), à polarité positive avec une phase ADN intermédiaire dans le cycle de réplication.**
Ces virus possèdent un génome composé de deux copies d'ARN (+) monocaténaires. L'ARN génomique n'est pas immédiatement traduit, mais sert de matrice à la transcriptase inverse virale (rétrotranscriptase), qui synthétise un brin d'ADN complémentaire du génome, dégrade l'ARN génomique et synthétise un deuxième brin d'ADN complémentaire au premier. Cet ADN bicaténaire migre ensuite dans le noyau de la cellule hôte où il est intégré au génome de la cellule, à l'aide d'une intégrase virale, formant ainsi un provirus.
Le provirus peut être transcrit en ARN viral génomique (+) et en ARNm (subgénomique), qui sont exportés dans le cytoplasme pour synthétiser des protéines virales. Les protéines virales nouvellement formées ainsi que l'ADN viral génomique (+) formeront les nouveaux virions.

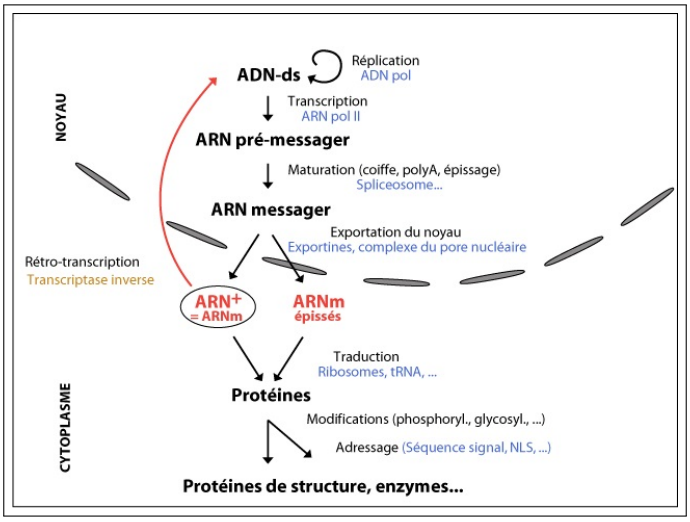


FIGURE 1.12 – Réplication des rétrovirus et des pararétrovirus : concernant les Rétrovirus, la rétrotranscriptase virale va synthétiser un ADN double-brin à partir de l'ARN génomique viral, qui est ensuite transcrit en ARN messager puis traduit en protéines virales. Pour les Pararétrovirus, l'ADN génomique viral, qui est double-brin mais incomplet, est complété par l'ADN polymérase virale, puis est transcrit en ARN messager avant d'être traduit en protéines. Cet ADN est également transcrit en ARN génomique, qui sert de matrice à la rétrotranscriptase, puis à l'ADN polymérase, dans le but d'obtenir des copies du génome viral (ADN double-brin incomplet) (©UCLouvain).

Par exemple : le VIH (Virus de l'Immunodéficience Humaine)

vii. **les pararétrovirus à ADN double brin (dsDNA-RT), avec une phase ARN intermédiaire dans le cycle de vie**

Le génome de ces virus est constitué d'ADN bicaténaire et circulaire, dont le brin interne est incomplet. Une fois dans le noyau de la cellule hôte, l'ADN polymérase virale complète le brin interne de l'ADN génomique, afin de former un ADN bicaténaire complet et super-enroulé. Cet ADN est ensuite transcrit en ARNm, qui seront traduits en protéines virales. En parallèle, cet ADN est également transcrit en ARN (+) génomique, qui sert de matrice à la rétrotranscriptase virale, qui synthétise un brin d'ADN (-). Ce brin d'ADN (-) est alors circularisé et sert de matrice à l'ADN polymérase, qui synthétise un brin d'ADN complémentaire (mais incomplet). Les protéines virales nouvellement synthétisées, ainsi que le génome ADN circulaire et bicaténaire, mais incomplet, forment alors les nouveaux virus (cf. Figure 1.12).

Par exemple : le virus de l'Hépatite B, le virus de la mosaïque du chou-fleur

1.2.2 La classification de l'*International Committee on Taxonomy of Viruses* (ICTV)

Cette classification se base sur la proposition de Lwoff, Horne et Tournier (cf. Section 1.2) (FENNER, 1975).

Les virus sont classés selon une méthode semblable à celle des organismes cellulaires, sur cinq niveaux hiérarchiques :

— Ordre (*-virales*)
— Famille (*-viridae*)
— Sous-famille (*-virinae*)
— Genre (*-virus*)
— Espèce

Par exemple, le virus de la rage se classe dans l'ordre des *Mononegavirales*, dans la famille des *Rhabdoviridae* et dans le genre *Lyssavirus* ; le virus Ebola se classe dans l'ordre des *Mononegavirales*, dans la famille des *Filoviridae* et dans le genre *Ebolavirus*.

En 2015, la classification ICTV comporte sept ordres : les *Caudovirales* (Bactériophages, dsDNA), les *Herpesvirales* (Herpesvirus, dsDNA), les *Ligamenvirales* (virus d'Archées, dsDNA), les *Mononegavirales* (virus Ebola, virus de la Rage, virus de la Rougeole, (-)ssRNA), les *Nidovirales* (Coronavirus,(+)ssRNA), les *Picornavirales* (virus de l'Hépatite A, Rhinovirus, (+)ssRNA), et les *Tymovirales* (virus de la mosaïque du papayer, virus X de la pomme de terre, (+)ssRNA).

Les virus qui sont le sujet de cette thèse sont tous des virus infectant les eucaryotes, possédant un génome ADN double-brin (dsDNA, groupe I dans la classification de Baltimore) et sont classés comme suit selon la classification ICTV :

— Ordre : Non-assigné à l'heure actuelle
— Famille : *Pandoraviridae*, *Marseilleviridae*
— Sous-famille : Aucune
— Genre : *Pandoravirus*, *Marseillevirus*
— Espèce : *Pandoravirus neocaledonia*, Noumeavirus

Ces virus font partie du groupe des NCLDV (*Nucleo-Cytoplasmic Large DNA Viruses*), regroupant plusieurs familles de grands virus à ADN, avec un cycle de réplication cytoplasmique ou nucléo-cytoplasmique.

1.3 Les grands virus à ADN nucléo-cytoplasmiques (NCLDV)

Le groupe des NCLDV (*Nucleo-Cytoplasmic Large DNA Viruses*) ou grands virus à ADN nucléo-cytoplasmiques est composé de plusieurs familles de virus et a été établi grâce à des études de génomique comparative.

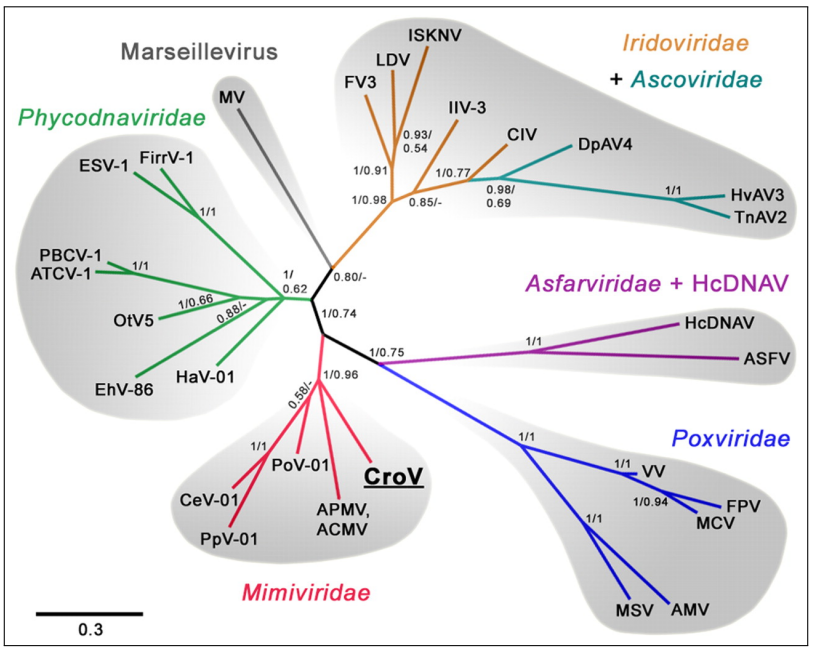


FIGURE 1.13 – Arbre phylogénétique des familles appartenant au groupe des NCLDV, généré à l'aide des séquences de l'ADN polymérase des différents virus (FISCHER *et al.*, 2010).

Les virus appartenant à ce groupe ont tous un grand génome à ADN double-brin linéaire ou circulaire (supérieur à 100 kb), et ont la particularité de n'infecter que des eucaryotes. Ils se répliquent soit exclusivement dans le cytoplasme de la cellule hôte, dans des structures appelées usines virales ; soit ils amorcent leur cycle de réplication par une phase nucléaire avant de le terminer dans le cytoplasme.

Tous les virus de ce groupe sont strictement dépendants de la machinerie de traduction de l'hôte. En revanche, certains de ces virus ont un nombre important de gènes codant pour des protéines impliquées dans les processus de réplication et de réparation de l'ADN (par exemple l'ADN polymérase), dans la formation de la structure des virions, ainsi que certains facteurs de transcription et de traduction.

Ces gènes sont conservés entre les NCLDV et sont appelés "*core genes*" : ils permettent notamment d'étudier les relations phylogénétiques des virus au sein de ce groupe, mais également d'assigner un nouveau virus à l'une des familles qui composent ce groupe.

En 2001, le groupe des NCLDV présenté par Iyer *et al.* (IYER *et al.*, 2001), est alors composé de quatre familles :

i. **Les Poxviridae**

La famille des *Poxviridae* comprend des virus classés en deux sous-familles : les *Chordopoxvirinae* qui infectent les vertébrés (mammifères, reptiles, oiseaux) et les *Entomopoxvirinae* infectant les invertébrés (arthropodes). Le génome des *Poxviridae* est linéaire et sa taille est comprise entre 130 et 360 kb : il est contenu dans une particule de forme ovale de 200 nm de diamètre sur 300 nm de long, généralement enveloppée. Ces virus se multiplient exclusivement dans le cytoplasme de la cellule hôte, grâce à une machinerie de transcription, de maturation des transcrits et de réplication de l'ADN indépendante de la cellule (HUGHES *et al.*, 2010).

FIGURE 1.14 – Virion de la variole (Smallpox) observé au microscope électronique à transmission (©CDC).

Par exemple : le virus de la variole, le Molluscipoxvirus.

ii. **Les Asfarviridae**

La famille des *Asfarviridae* ne comportait qu'un seul membre, l'African Swine Fever virus (ASFV) , et a été nommée d'après ce virus : *African Swine Fever And Related viruses*. Ce virus infecte les porcs domestiques et est vraisemblablement transmis par des tiques, qui agissent comme vecteur. Le génome de ce virus est linéaire et sa taille est comprise entre 170 et 190 kb selon les isolats : il est contenu dans une particule enveloppée, de forme icosaédrique, mesurant de 170 à 200 nm de diamètre. Ce virus se multiplie dans le cytoplasme de la cellule hôte : toutefois, il a été montré que le noyau est nécessaire et qu'il subit des modifications au cours de l'infection (DIXON *et al.*, 2013).

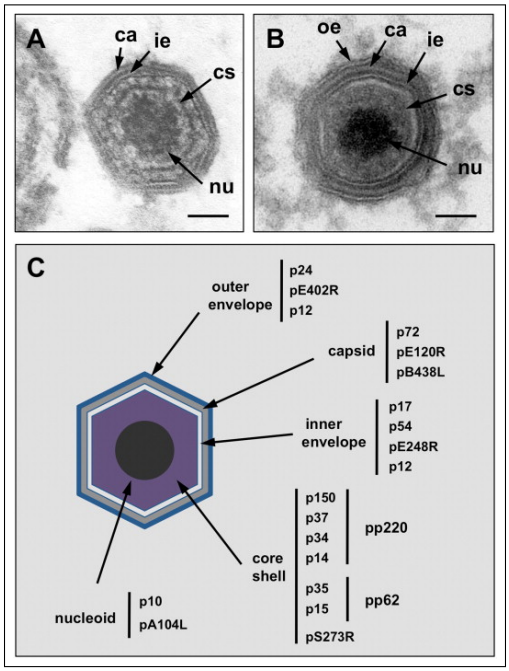


FIGURE 1.15 – African Swine Fever Virus (ASFV) observé au microscope électronique et schéma de la localisation des différentes protéines structurales qui forment la particule. Barres d'échelles : 50 nm (SALAS *et al.*, 2013).

Plus récemment, d'autres virus qui semblent être apparentés à cette famille ont été découverts, comme Faustovirus, un virus infectant les amibes *Vermamoeba vermiformis*, et possédant un génome de 466 kb (RETENO *et al.*, 2015).

iii. **Les Iridoviridae**

La famille des *Iridoviridae* est composée de virus, classés en cinq genres : les *Megalocytivirus*, les *Ranavirus*, les *Iridovirus*, les *Chloriridovirus* et les *Lymphocystivirus*. Ces virus infectent des invertébrés et quelques vertébrés poïkilothermes (ou plus communément appelés "à sang-froid"), comme les poissons et les amphibiens. Le génome de ces virus peut être circulaire ou linéaire avec des séquences terminales redondantes et circulairement permutées, et sa taille est comprise entre 140 et 303 kb. Ce génome est contenu dans des particules de forme icosaédriques (certaines sont enveloppées), qui contiennent une membrane lipidique interne, et mesurent entre 120 et 350 nm de diamètre. Ces virus ont un cycle de réplication nucléo-cytoplasmique : le virus passe dans un premier temps par une phase nucléaire, où son génome est transcrit par l'ARN polymérase de la cellule hôte, puis dans le cytoplasme, où a lieu notamment, l'assemblage des nouveaux virions (WILLIAMS *et al.*, 2005).

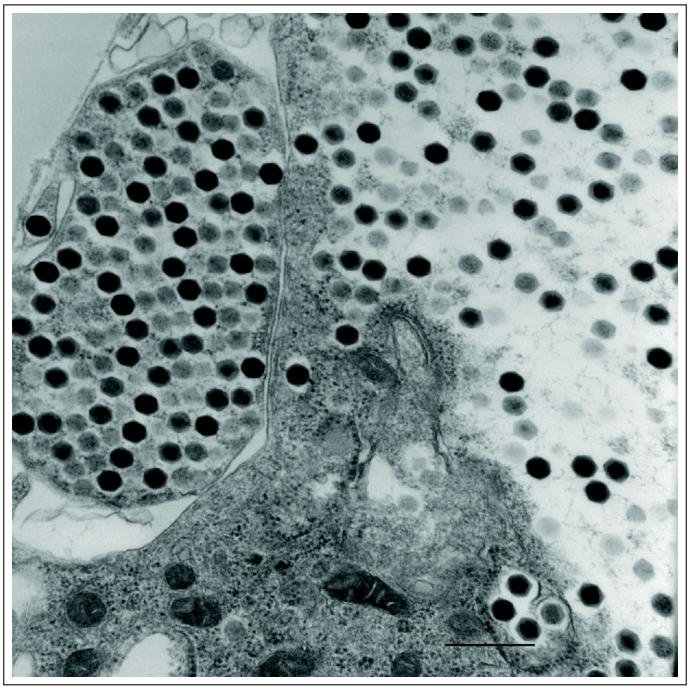


FIGURE 1.16 – Invertebrate Iridescent virus 6 (IIV-6) isolat CrIV (Iridovirus du Criquet) observé au microscope électronique dans le cytoplasme d'une cellule. Barre d'échelle : 500 nm (ADAMO *et al.*, 2014).

Par exemple : l'Invertebrate Iridescent virus 6 (IIV-6) fait partie du genre *Iridovirus*.

iv. **Les Phycodnaviridae**

La famille des *Phycodnaviridae* se divise en 6 genres : les *Chlorovirus*, les *Coccolithovirus*, les *Phaeovirus*, les *Prasinovirus*, les *Prymnesiovirus* et les *Raphidovirus*. Ces virus infectent des algues eucaryotes marines ou d'eau douce : les virus du genre *Chlorovirus* infectent des algues vertes d'eau douce, comme les chlorelles, alors que les virus des cinq autres genres infectent des algues vertes et brunes marines. Le génome de ces virus est linéaire, mesure entre 160 et 560 kb, et est contenu dans des particules icosaédriques, dont la taille est comprise entre 100 et 220 nm de diamètre. Le cycle de réplication de ces virus passe dans un premier temps par une phase nucléaire, avant de se terminer dans le cytoplasme de la cellule hôte (WILSON *et al.*, 2009).

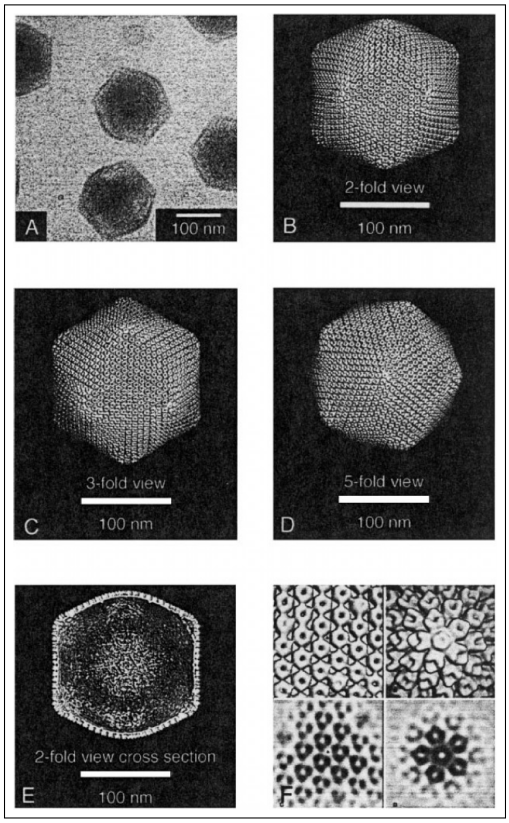


FIGURE 1.17 – *Paramecium bursaria* chlorella virus 1 (PBCV-1) observé au microscope électronique et reconstructions en deux et trois-dimensions de la particule. Barres d'échelle : 100 nm (VAN ETTEN *et al.*, 1999).

Par exemple : *Paramecium bursaria* chlorella virus 1 (PBCV-1) fait partie du genre *Chlorovirus*.

Au fil des années et des découvertes, d'autres familles ont été ajoutées au groupe des NCLDV (YUTIN *et al.*, 2009) :

v. **Les Ascoviridae**

La famille des *Ascoviridae* ne contient qu'un seul genre, les *Ascovirus*. Ces virus infectent des invertébrés et plus particulièrement des insectes faisant partie de la famille des noctuidés (chenilles et papillons de nuit). Le génome de ces virus est circulaire, sa taille est comprise entre 150 et 186 kb, et il est contenu dans des particules de forme ovoïde (*asco* signifie sac en grec). Ces particules mesurent 130 nm de diamètre sur 200 à 240 nm de long. Le cycle de réplication de ces virus est nucléo-cytoplasmique : les premières étapes du cycle s'effectuent dans le noyau de la cellule hôte, avant de se terminer dans le cytoplasme (CHENG *et al.*, 2008).

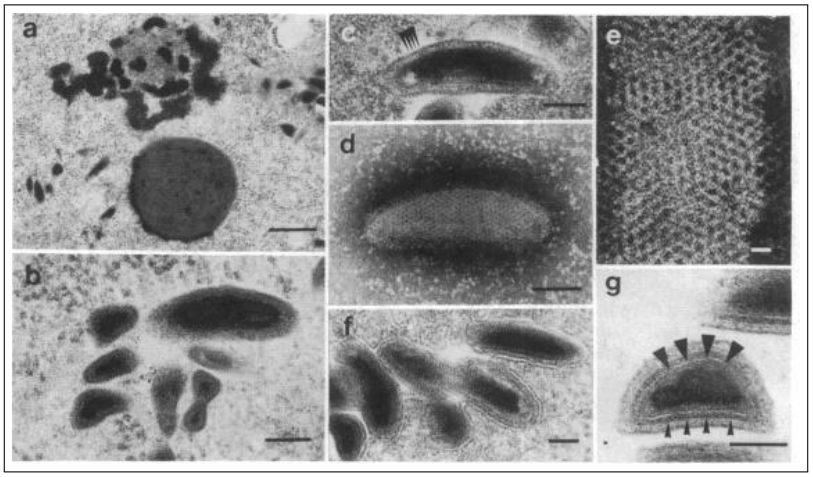


FIGURE 1.18 – *Trichoplusia ni* ascovirus 2c (TnAV-2c) observé au microscope électronique. Barres d'échelles : (a) 500 nm, (b, c, d, f, g) 100 nm, (e) 25 nm (FEDERICI, 1983).

Par exemple : *Spodoptera frugiperda* ascovirus 1a (SfAV-1a), *Trichoplusia ni* ascovirus 2c (TnAV-2c).

Les différentes familles de virus géants, qui appartiennent également au groupe des NCLDV seront décrites en détail dans la section suivante.

La création d'un nouvel ordre viral, les *Megavirales*, qui regroupe toutes les familles de virus appartenant aux NCLDV a été proposée en 2013, mais n'a pas encore été acceptée par l'ICTV à ce jour (COLSON *et al.*, 2013b).

1.4 Les virus géants

Jusqu'à très récemment, les virus étaient considérés comme des parasites, plus petits que des bactéries, capables de passer à travers les pores des filtres de stérilisation (entre 0.2 et 0.3 μm) , et ne pouvaient pas être observés au microscope optique. La découverte des premiers virus "géants" il y a une dizaine d'années a bousculé les idées établies sur la notion de virus, leur origine et leur évolution.

1.4.1 Découverte du premier virus géant

En 1992, un micro-organisme provenant d'une tour de refroidissement d'un hôpital, à Bradford (Angleterre), a été isolé à partir d'amibes du genre *Acanthamoeba* infectées, lors d'une épidémie de pneumonie. Ce micro-organisme, visible au microscope optique, qui ressemblait à une bactérie Gram positive en forme de coque, a été baptisé *Bradfordcoccus*. Il s'est avéré, cependant, que ce micro-organisme n'était pas cultivable comme toutes les autres bactéries, et les différents tests de caractérisation utilisant les amorces universelles bactériennes (sur le gène de l'ARN ribosomique 16S), ont échoué malgré des conditions permissives (RAOULT *et al.*, 2007).

Une dizaine d'années plus tard, en 2003, la véritable nature de ce micro-organisme, qui avait été laissé de côté, a été révélée. En effet, il s'est avéré qu'après observation au microscope électronique à transmission (MET) de la particule et de son cycle de réplication, ainsi que le séquençage partiel de son génome, *Bradfordcoccus* n'était pas une bactérie, mais bien un virus, rebaptisé Mimivirus pour *Mimicking Microbe virus* (ou en français, le virus qui imite les microbes). Mimivirus ou *Acanthamoeba polyphaga* mimivirus est donc le premier virus observable au microscope optique, qui soit retenu par le filtre de Chamberland et constitue le membre fondateur de la famille des *Mimiviridae*. Sa capside icosaédrique gigantesque, mesurant 450 nm de diamètre (700 nm en ajoutant les fibres entourant la particule), ainsi que la dimension démesurée de son génome (1.18 Mb), font de Mimivirus le premier virus "hors-normes" (La Scola *et al.*, 2003 ; Raoult *et al.*, 2004).

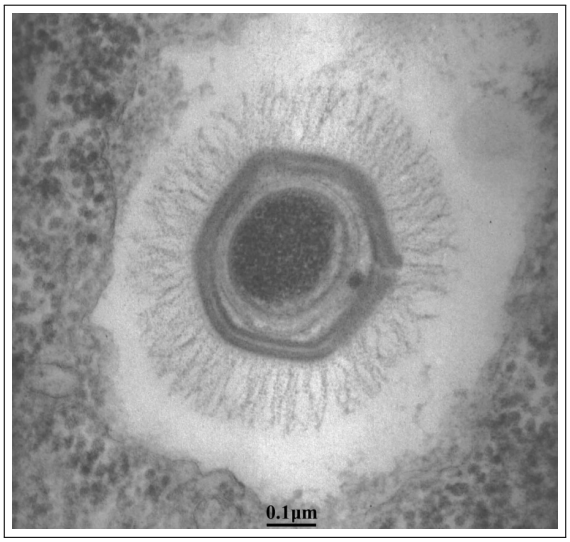


Figure 1.19 – Mimivirus observé au microscope électronique (Claverie *et al.*, 2009).

Ses caractéristiques morphologiques sont surprenantes pour un virus : sa capside est, en terme de volume, treize fois plus grande que celle de *Paramecium bursaria* chlorella virus (PBCV-1, 190 nm de diamètre, Van Etten *et al.*, 1999), et quarante et une fois plus grande que celle de l'Herpes simplex virus (HSV-1, 130 nm de diamètre, Liu *et al.*, 2007), des virus considérés comme "conventionnels" ou "classiques". Il est même trois fois plus grand que l'une des plus petites bactéries connues, *Mycoplasma genitalium*, qui n'est pas observable au microscope optique, et mesure environ 300 nm de diamètre (Daley *et al.*, 2014).

FIGURE 1.20 – Couverture du journal American Scientist mettant en scène des particules de Mimivirus dont la taille est comparée avec celles des virus du rhume et du VIH (©American Scientist).

Mimivirus possède une capside à symétrie icosaédrique qui mesure environ 450 nm de diamètre et est recouverte d'une couche dense de fibres mesurant environ 125 nm (XIAO *et al.*, 2005). Cette capside est composée de plusieurs membranes, dont une entourant le cœur de la particule contenant le génome, ainsi qu'une membrane interne. Cette membrane interne est dérivée de la membrane du réticulum endoplasmique de la cellule hôte : elle est recyclée lors de la formation de nouveaux virions au cours du cycle infectieux (MUTSAFI *et al.*, 2013).

Les fibres qui entourent la capside virale sont essentiellement composées de trois protéines, nommées R135, L725 et L829, dont deux sont glycosylées (R135 et L829), et deux sont de fonction inconnue (L725 et L829) (BOYER *et al.*, 2011). La structure de la troisième protéine, R135, a été résolue récemment (KLOSE *et al.*, 2015). Ces fibres sont protégées par des structures peptidoglycaniques probablement réticulées, ce qui peut expliquer d'une part le fait que Mimivirus réagisse positivement au marquage utilisé pour les bactéries Gram +, et d'autre part, que les amibes soient "attirées" par ces virus qui imitent les structures des polysaccharides bactériens (KLOSE *et al.*, 2010). Enfin, une étude récente a identifié trois oligo(poly)saccharides associés à ces fibres : le rhamnose, le N-acétylglucosamine (GlcNAc) et la viosamine, comme composants majeurs des sucres associés aux fibres de Mimivirus (PIACENTE *et al.*, 2012).

Située sur un des sommets de la particule virale et dépourvue de fibres, une structure à l'architecture étonnante en forme d'étoile, baptisée "*stargate*", constitue un portail, une ouverture, qui permet au génome de Mimivirus de pénétrer dans le cytoplasme de

la cellule hôte lors de l'infection (cf. Figure 1.21). Étonnamment, cette structure n'est pas utilisée comme portail pour charger l'ADN dans les capsides des virions en cours de formation, comme c'est le cas chez les autres virus, qui ne possèdent qu'une seule ouverture dans leur capside (ZAUBERMAN *et al.*, 2008).

Le génome de 1.18 Mb de Mimivirus est entouré d'une enveloppe : cette structure, aussi appelée nucléoïde, mesure environ 300 nm de diamètre. Contrairement aux autres virus, il existe un espace de 30 à 50 nm entre le génome enveloppé et l'intérieur de la capside virale. Enfin, ce nucléoïde n'est pas complètement sphérique : en effet, il y a une dépression concave, située en face du sommet où se trouve la stargate. Au cours du cycle infectieux de Mimivirus, lorsque la stargate s'ouvre, le nucléoïde est relâché dans le cytoplasme de la cellule hôte (cf. Figure 1.21) : ce phénomène marque alors l'initiation de l'infection (XIAO *et al.*, 2009 ; CLAVERIE *et al.*, 2009).

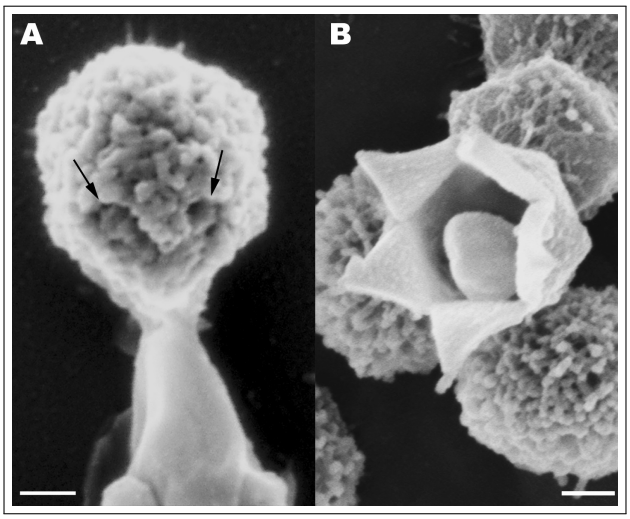


FIGURE 1.21 – Stargate de Mimivirus observée au microscope à balayage : à gauche, une particule fermée où les flèches indiquent les bords de la stargate, et à droite, l'ouverture de la stargate où l'on peut apercevoir le nucléoïde (ou nucléocapside) de Mimivirus. Barres d'échelle : 100 nm. (ZAUBERMAN *et al.*, 2008).

Plus étonnant encore, le génome de Mimivirus est huit fois plus grand que celui de HSV-1 (150 kb), trois fois plus grand que celui de PBCV-1 (330 kb), qui possédait jusqu'à présent l'un des génomes les plus importants pour un virus, et deux fois plus grand que le génome de *Mycoplasma genitalium* (580 kb).

Son génome à ADN double-brin linéaire, riche en A+T (72%), contient un peu plus de 1000 gènes, dont 979 codent pour des protéines (LEGENDRE *et al.*, 2010), ce qui accentue le côté "hors-normes" de Mimivirus, lorsque l'on sait que HSV-1 n'a besoin "que" de 84 gènes pour se répliquer (WIDENER *et al.*, 2014).

Dans le génome de ce virus, plus de 70% des gènes n'ont pas d'homologues dans les bases de données. De plus, certains gènes normalement retrouvés chez les organismes cellulaires, tel que des gènes codant pour des éléments de la machinerie de synthèse

des protéines, ont également été trouvés dans ce génome, ce qui est inhabituel pour un virus.

Ces gènes codent pour des protéines impliquées dans la machinerie de transcription (dont une ARN polymérase, une hélicase, quelques facteurs de transcription, ...), mais aussi de réparation de l'ADN (dont des enzymes de réparation des dommages de l'ADN causés par les rayons UV) et même dans la machinerie de traduction (quatre amino-acyl-ARNt synthétases, des facteurs d'initiation et d'élongation de la traduction), ce qui n'avait jamais été observé chez un virus jusqu'à présent (RAOULT *et al.*, 2004).

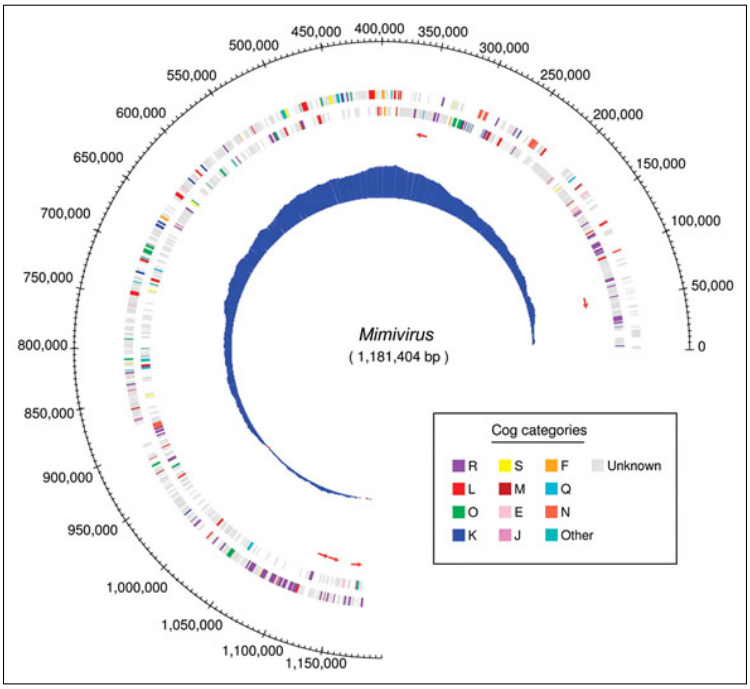


FIGURE 1.22 – Carte du chromosome de Mimivirus où les séquences qui codent pour des protéines sont indiquées sur les deux brins du génome : les séquences sont colorées selon leurs catégories fonctionnelles. Les séquences colorées en gris sont celles dont la fonction est inconnue. Les flèches rouges indiquent l'emplacement et l'orientation des ARN de transfert (ARNt) (RAOULT *et al.*, 2004).

Des études du transcriptome des cellules infectées par Mimivirus ont été réalisées et ont révélé que tous les ARN messagers du virus sont polyadénylés, grâce à une séquence palindromique, qui sert de signal définissant le site de polyadénylation des transcrits. Cette séquence forme une structure en tige-boucle et est située en 3' des transcrits, dans la région intergénique. Ainsi, l'arrêt de la transcription des gènes, ainsi que la poly-adénylation des ARN messagers, obéissent à la "*hairpin-rule*", ou la règle de l'épingle à cheveux (BYRNE *et al.*, 2009), qui a également été observée chez les *Iridoviridae*, les *Marseilleviridae* et les *Pithoviridae*.

La découverte de Mimivirus marque donc un tournant majeur dans le monde de la virologie, remettant en question la définition même du mot "*virus*", et soulevant de nombreuses interrogations, notamment sur les conditions de sa découverte : existe t-il d'autres virus géants, que l'on aurait considéré -à tort- comme des bactéries ? Existe

t-il d'autres "types" de virus géants, plus extraordinaires encore ? Où peut-on les trouver ? Sont-ils ubiquitaires et auraient-ils été occultés pour des raisons historiques, qui auraient biaisé les méthodes utilisées dans l'étude des virus ?

1.4.2 Écologie et virus (géants)

Les virus sont trouvés partout où il y a de la vie, or, il s'avère que les mers et les océans couvrent plus de 70% de la planète : des études ont donc été réalisées dans le but d'estimer l'abondance des virus dans les océans.

On sait, depuis la fin des années 1970, que les virus sont très abondants dans les océans, mais ce n'est que depuis quelques années que cette abondance peut être quantifiée, grâce à des observations en microscopie électronique et au microscope à épifluorescence. Les différentes observations ont permis d'estimer qu'il y a environ 10^7 virus par millilitre d'eau de mer et entre 10^8 et 10^9 virus par cm^3 dans les sédiments marins.

Ces observations ont également montré que les virus sont plus abondants lorsque les prélèvements sont réalisés dans des eaux peu profondes et proches des rivages. En réalité, les virus sont nettement plus abondants là où il y a une forte concentration de bactéries, de plancton et autres micro-organismes ou organismes marins, ce qui est également vrai pour les virus retrouvés dans les sédiments marins (SUTTLE, 2005).

Les virus sont donc les entités biologiques les plus abondantes dans les océans, surpassant de loin les procaryotes et les protistes. En revanche, du fait de leur taille, ils ne représentent qu'environ 5% de la biomasse, se plaçant loin derrière les procaryotes, qui représentent plus de 90% de la biomasse (SUTTLE, 2007).

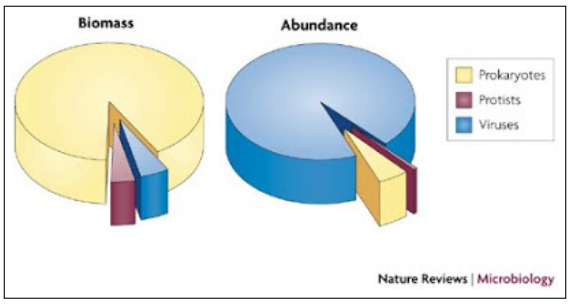


FIGURE 1.23 – Graphiques indiquant la biomasse et l'abondance des procaryotes, des protistes et des virus dans les océans. Les virus représentent environ 94% des particules contenant des acides nucléiques dans les océans, et seulement 5% de la biomasse, largement dominée par les procaryotes (SUTTLE, 2007).

Les virus représentent donc, du fait de leur forte abondance, la plus grande diversité génétique dans les océans et jouent un rôle majeur dans la régulation des écosystèmes marins (SUTTLE, 2005).

Le rôle des virus dans la régulation des écosystèmes marins peut être illustré par l'exemple de l'algue marine unicellulaire : *Emiliania huxleyi*.

Cette algue appartient à l'ordre des *Coccolithophores* : elle est très abondante et est retrouvée dans presque tous les milieux (salinité faible ou au contraire élevée, à des températures oscillant entre 5°C et 30°C). La cellule d'*E. huxleyi* est entourée d'une sorte de carapace, constituée de plusieurs coccolithes, des plaques de calcite à l'architecture complexe.

Cette algue joue un rôle important dans les écosystèmes marins, notamment dans la création de sédiments calcaires (falaises de craie, par exemple) et est également impliquée dans le cycle du carbone (source de dioxyde de carbone, CO_2).

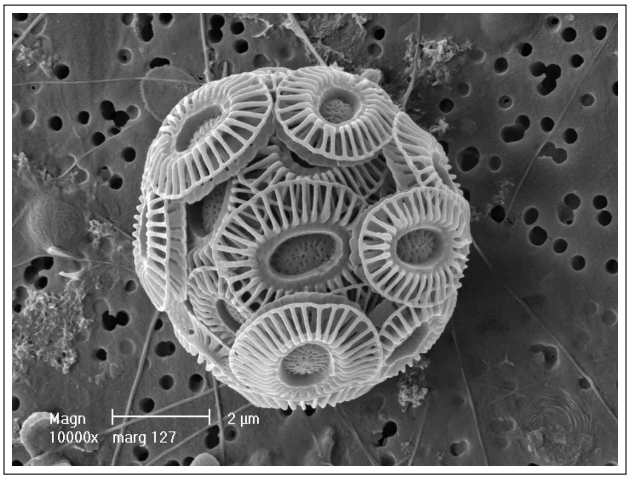


FIGURE 1.24 – L'algue unicellulaire *Emiliania huxleyi* observée au microscope à balayage : les coccolithes, ou plaques de calcite, situées à la surface de la cellule ont une architecture complexe (©Jérémy Young).

De plus, *E. huxleyi* est fortement impliquée dans des efflorescences algales ou "*blooms*" (ou encore efflorescences planctoniques). Ces efflorescences se produisent régulièrement, le plus généralement en été, apparaissent assez rapidement (en quelques jours) et disparaissent tout aussi vite. On retrouve dans ces efflorescences une quantité colossale de phytoplancton estimée à plusieurs millions de cellules par litre, dont *E. huxleyi*, l'un des composants majeur de ce phytoplancton. Ces efflorescences peuvent s'étendre sur de larges surfaces et sont d'ailleurs visibles depuis l'espace (cf. Figure 1.25), notamment grâce aux coccolithes d'*E. huxleyi*, qui réfractent la lumière (WESTBROEK *et al.*, 1989).

La disparition rapide des efflorescences d'*E. huxleyi* est due à un virus : *Emiliania huxleyi* virus (EhV). EhV-86 est un virus dont la capside icosaédrique mesure environ 175 nm de diamètre, qui a été isolé au cours d'une efflorescence dans la Manche. Son génome à ADN double-brin circulaire de plus de 400 kb, le classe parmi les grands virus à ADN (NCLDV), et plus particulièrement dans la famille des *Phycodnaviridae* (WILSON *et al.*, 2005).

Quelques années plus tard, le même phénomène de disparition d'efflorescences algales de *Phaeocystis globosa*, une autre algue unicellulaire, a été observé, impliquant

FIGURE 1.25 – Efflorescence (ou bloom) de l'algue *Emiliania huxleyi* observée par satellite au large de la Grande-Bretagne (©Steve Groom).

cette fois un virus appelé *Phaeocystis globosa* virus, ou PgV. Ce virus possède une capside icosaédrique qui mesure environ 150 nm de diamètre, et un génome à ADN double-brin linéaire de 460 kb. Des analyses phylogénétiques ont montré que PgV-16T n'appartient pas à la famille des *Phycodnaviridae*, mais à celle fondée par Mimivirus (SANTINI *et al.*, 2013).

1.4.3 Recherche des virus géants dans l'environnement

Les virus sont les entités biologiques les plus abondantes sur la planète et constituent un des plus grands réservoirs de la diversité génétique, notamment dans les océans (cf. Section 1.4.2).

Les océans contiennent un mélange complexe de micro-organismes qui ne sont pas, pour la plupart, caractérisés génétiquement ou biochimiquement : c'est donc dans ce but que des expéditions d'échantillonnages du monde marin se sont mises en place.

Entre 2003 et 2004, l'expédition baptisée *Global Ocean Sampling* ou GOS, menée par le scientifique Craig Venter, consiste à analyser des échantillons d'eau de mer prélevés en surface et collectés à bord du navire *Sorcerer II*, afin de réaliser des études métagénomiques du microbiote planctonique marin.

Des fractions de ces échantillons ont été isolées selon leur taille, l'ADN en a été extrait, puis séquencé et enfin, les séquences ont été analysées. Cette première étude, basée sur l'évolution et la nature des variations génétiques des micro-organismes marins, a permis de démontrer qu'il y a une grande diversité dans les communautés microbiennes marines, les bactéries étant majoritaires (RUSCH *et al.*, 2007).

Des analyses supplémentaires sur ces échantillons, ainsi que sur des échantillons prélevés lors d'une précédente expédition (VENTER *et al.*, 2004), ont été réalisées. En se focalisant cette fois sur les virus, ces analyses ont montré qu'il y avait une forte

abondance de séquences virales, la majorité de ces séquences étant représentée par des bactériophages (virus infectant les bactéries). Grâce à la découverte des virus géants, cette étude des séquences virales s'est portée sur la fraction allant de 0,1 μm à 0,8 μm, contrairement aux études métagénomiques précédentes, où l'accent était mis sur les séquences virales obtenues à partir de fractions filtrées sur 0,22 μm (WILLIAMSON *et al.*, 2008). Des séquences homologues de virus inconnus, apparentés à Mimivirus ont d'ailleurs été retrouvées dans ce jeu de séquences virales obtenues à partir de la fraction 0,1 μm-0,8 μm (GHEDIN *et al.*, 2005).

En 2009, une campagne d'échantillonnage menée par un consortium de plus de cent scientifiques venant de tous horizons et baptisée *Tara Oceans* a été lancée, à bord du bateau Tara (cf. Figure 1.26). Cette campagne, ayant pour but la visualisation, la quantification et la caractérisation fonctionnelle de la biodiversité des écosystèmes planctoniques dans tous les océans, a réalisé, contrairement aux autres expéditions, des prélèvements à diverses profondeurs de la colonne d'eau, dans une gamme de taille allant des virus aux larves de poissons.

Les prélèvements ont été collectés, fractionnés en fonction de leur taille, puis observés et quantifiés immédiatement à bord, à l'aide d'une plateforme d'imagerie à haut-débit, avant d'être préservés pour les analyses génomiques (séquençage haut-débit) et morphologiques réalisées une fois à terre (KARSENTI *et al.*, 2011). Encore une fois, des séquences homologues de virus inconnus, car ils n'ont pas encore été isolés, appartenant à la famille de Mimivirus ont été retrouvées dans les séquences provenant de cette expédition (MOZAR *et al.*, 2014).

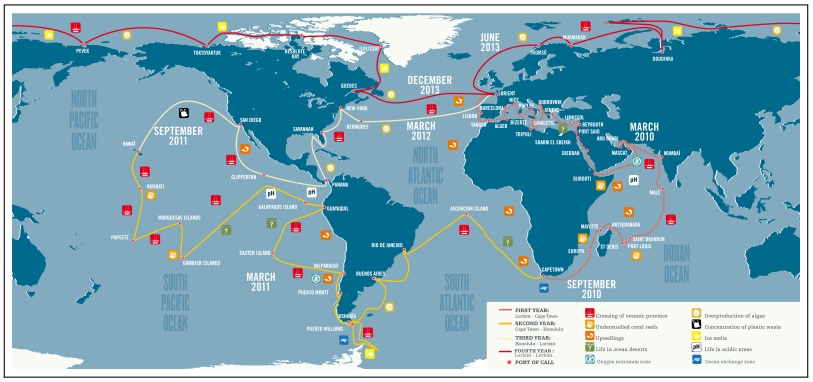


FIGURE 1.26 – Carte représentant le parcours effectué lors de la campagne Tara Oceans, les différents points de prélèvement et leurs caractéristiques ainsi que les ports d'escale (©Tara Oceans).

En 2011, dans le cadre du projet européen ASSEMBLE ("*Association of European Marine Biological Laboratories*"), qui permet notamment aux chercheurs (non océanographes) d'accéder à des stations marines dans le monde entier, une campagne d'échantillonnage a été mise en place.

Cet échantillonnage de l'environnement aquatique, suivi d'une co-culture avec plusieurs espèces d'amibes du genre *Acanthamoeba*, a permis d'isoler un virus plus surprenant encore que Mimivirus.

Megavirus chilensis, ou Mégavirus, a été isolé à partir d'un échantillon d'eau de mer prélevé près des côtes de Las Cruces, au Chili. Ce virus, similaire à Mimivirus, est encore plus imposant que ce dernier, que ce soit en terme de taille de capside ou de génome. Sa capside, plus grande que celle de Mimivirus, mesure environ 520 nm de diamètre, et est également recouverte de fibres, en revanche plus courtes que celles de Mimivirus (environ 80 nm de long). Le génome linéaire à ADN double-brin de 1.2 Mb de Mégavirus, contient 1120 gènes qui codent pour des protéines, dont sept amino-acyl-ARNt synthétases (Mimivirus en possède seulement quatre). La présence de séquences palindromiques, capables de former des tiges-boucles, et définissant les sites de polyadénylation des transcrits, ont également été retrouvées. Environ la moitié des gènes de Mégavirus sont retrouvés chez Mimivirus et présentent, en moyenne, 50% d'identité de séquence au niveau protéique. De plus, 23% des gènes de Mégavirus n'ont pas d'homologues chez Mimivirus et sont donc uniques à Mégavirus, et plus de 80% correspondent à des gènes orphelins (ORFans) (ARSLAN *et al.*, 2011).

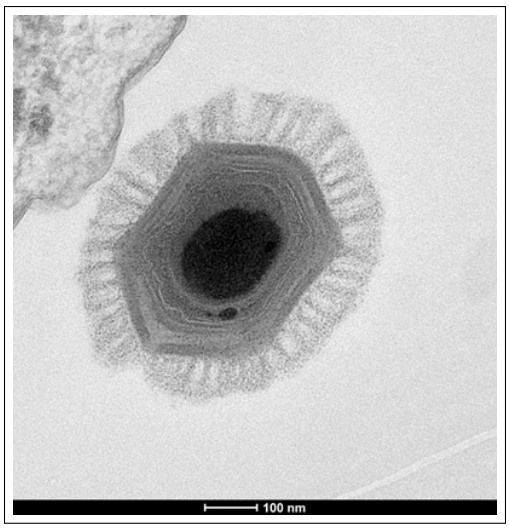


FIGURE 1.27 – *Megavirus chilensis* observé au microscope électronique : sa capside est plus grande que celle de Mimivirus, mais les fibres qui entourent sa particule sont plus courtes (©Chantal Abergel, IGS).

La découverte de ce virus dans un environnement différent de celui où a été isolé Mimivirus (eau de mer *versus* eau douce), encourage la recherche de nouveaux virus dans des environnements variés, en utilisant la co-culture avec les amibes du genre *Acanthamoeba*, qui ont déjà permis d'isoler deux virus géants exceptionnels qui repoussent encore un peu plus loin les limites de la virologie.

1.4.4 Hôtes des virus géants

Les Amibes

Les amibes sont des eucaryotes unicellulaires (protistes) opportunistes, qui sont retrouvées sous forme libre (“*Free Living Amoeba*” ou FLA) de façon ubiquitaire dans l'environnement.

Les amibes du genre *Acanthamoeba* mesurent entre 10 et 25 μm, et possèdent des spicules ou épines à leur surface ainsi que de nombreuses vacuoles dans leur cytoplasme (dont des vacuoles digestives et des lysosomes). Ces cellules ont également des pseudopodes, qui leur permettent d'adhérer sur un substrat, de se mouvoir dans leur milieu et de phagocyter des bactéries.

Ces amibes peuvent se retrouver sous deux formes différentes au cours de leur cycle de vie :

— une forme végétative, le trophozoïte, qui est une forme active, au cours de laquelle la cellule va se nourrir de micro-organismes (le plus souvent des bactéries) et se diviser ;

— une forme dormante, le kyste, dont l'activité métabolique est réduite lorsque les conditions ne sont pas favorables pour la cellule, par exemple en cas de stress, de famine, de températures extrêmes ou en présence de toxines (SIDDIQUI *et al.*, 2012).

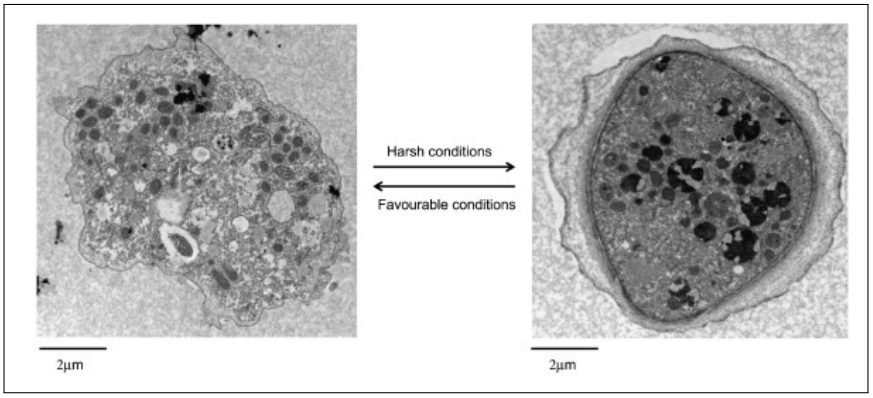


FIGURE 1.28 – L'amibe *Acanthamoeba* observée au microscope électronique, avec les deux étapes de son cycle de vie : lorsque les conditions sont favorables, elle est sous forme de trophozoïte (image de gauche), et lorsque les conditions deviennent défavorables (stress, famine, ...) elle se transforme en kyste, très résistant (image de droite) (SIDDIQUI *et al.*, 2012).

Les kystes sont très résistants : cette résistance est conférée par la présence de deux parois qui entourent la cellule, composées de protéines et de polysaccharides (dont de la cellulose, TOMLINSON *et al.*, 1962).

Ces parois sont séparées par un espace, à l'exception de certains endroits où elles se rejoignent pour former un opercule, au centre des ostioles : ces opercules constituent une “voie de sortie” pour le trophozoïte, lors de l'excystation ou désenkystement (BOWERS *et al.*, 1969).

Les amibes sont ubiquitaires : elles sont trouvées dans une grande diversité d'environnements, comme dans l'eau de mer ou les sédiments marins, dans des étangs ou des lacs d'eau douce, dans des tours de refroidissement, dans l'Antarctique, et même dans l'air (SIDDIQUI *et al.*, 2012). Les amibes sont également très abondantes dans le sol, où elles représentent la communauté la plus importante de prédateurs bactériens (RODRÍGUEZ-ZARAGOZA, 1994).

Les amibes sont retrouvées dans une grande diversité de sols, dans des environnements parfois extrêmes comme les déserts, les sols pollués et même dans des sols situés à haute altitude (Tibet), où elles jouent un rôle majeur dans l'écosystème, en régulant les populations bactériennes et en rendant les éléments nutritifs disponibles, favorisant la croissance des plantes (SIDDIQUI *et al.*, 2012).

Les amibes du genre *Acanthamoeba* sont responsables de pathologies, causant notamment des lésions cutanées, des infections des sinus, des kératites, ou encore des encéphalites granulomateuses. Elles servent également de réservoirs ("*Trojan horse*" ou cheval de Troie) à de nombreux endosymbiontes, comme des virus, des bactéries (pathogènes ou non) et des levures. Ces amibes sont également des hôtes pour les virus géants, dont Mimivirus, isolé dans l'amibe *Acanthamoeba polyphaga* et *Megavirus chilensis*, capable de se répliquer dans *A. griffini, A. polyphaga* et *A. castellanii* (SIDDIQUI *et al.*, 2012 ; LA SCOLA *et al.*, 2003 ; ARSLAN *et al.*, 2011).

D'autres types d'amibes ont également été testés en parallèle afin d'isoler de nouveaux virus. C'est le cas de *Vermamoeba vermiformis*, un protiste sous forme libre (FLA), fréquemment retrouvé dans des environnements associés aux hommes : cette amibe a récemment été isolée à partir d'échantillons d'eau prélevés dans différents hôpitaux (PAGNIER *et al.*, 2015). *V. vermiformis*, qui n'avait jamais été utilisée comme hôte dans la recherche de nouveaux virus, a permis d'isoler Faustovirus, un virus phylogénétiquement proche des *Asfarviridae*, appartenant au groupe des NCLDV (RETENO *et al.*, 2015). Ce virus possède une capside icosaédrique d'environ 240 nm de diamètre : dernièrement, la présence étonnante d'une double capside composant ce virus a été révélée (cf. Figure 1.29 KLOSE *et al.*, 2016). Faustovirus possède un génome à ADN double-brin circulaire de 466 kb, avec 451 gènes codant pour des protéines dont environ les deux tiers n'ont aucun homologue dans les bases de données (RETENO *et al.*, 2015).

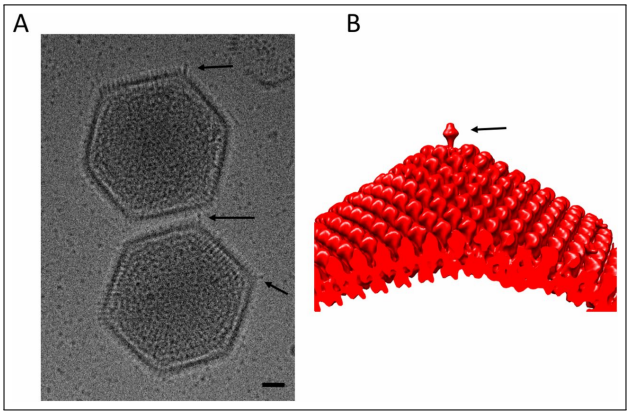


FIGURE 1.29 – Faustovirus observé en cryo-microscopie électronique : à gauche, des particules de Faustovirus où les flèches indiquent les fibres situées à la surface des particules. Barre d'échelle : 50 nm, et à droite, la reconstruction d'une partie de la capside externe de Faustovirus où la fibre située à un sommet de la particule est indiquée par une flèche. (KLOSE *et al.*, 2016).

La recherche de nouveaux virus géants a également été réalisée en utilisant l'amibe *Dictyostelium discoideum*, l'espèce la plus étudiée parmi les amibes dites "sociales". Ces amibes vivent dans le sol ou sur des tapis de feuilles mortes dans les forêts et se nourrissent de bactéries et de levures. Lorsqu'elles se retrouvent en situation de famine, elles forment des agrégats, qui forment une sorte d'organe de fructification, composé d'une tige sur laquelle se trouve une boule d'amibes (cf. Figure 1.30), qui vont se différencier en spores très résistants, afin de pouvoir se disséminer (KESSIN, 2001). L'utilisation de *D. discoideum* comme hôte potentiel de nouveaux virus géants est restée sans succès à ce jour (KHALIL *et al.*, 2016).

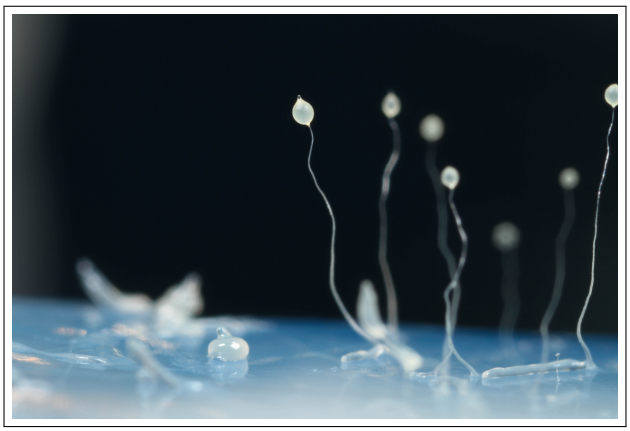


FIGURE 1.30 – Organe de fructification de l'amibe *Dictyostelium discoideum* (©Usman Bashir).

L'utilisation d'un panel de différents protistes, ubiquitaires dans l'environnement et connus pour être des réservoirs de micro-organismes, constitue vraisemblablement une bonne stratégie pour isoler de nouveaux virus géants à partir d'échantillons environnementaux (DORNAS *et al.*, 2015).

Les Algues

La présence de séquences de virus phylogénétiquement proches de Mimivirus dans des échantillons marins, prélevés lors de plusieurs campagnes d'échantillonnage, a déjà été démontrée (GHEDIN *et al.*, 2005), ce qui laisse présager que les grands virus à ADN apparentés à la famille des *Mimiviridae* seraient ubiquitaires dans les océans (MONIER *et al.*, 2008a).

Des études des séquences de Mimivirus ont notamment montré que ce virus était proche, au niveau phylogénétique, des virus appartenant à la famille des *Phycodnaviridae*. En effet, en réalisant la phylogénie de l'ADN polymérase à partir de séquences de Mimivirus et de plusieurs *Phycodnaviridae*, Monier *et al.* ont montré que Mimivirus et certains virus infectant des algues unicellulaires seraient classés dans la famille des *Mimiviridae*. Ils supposent, de plus, que le réservoir naturel de Mimivirus serait peut-être une algue, celles-ci étant fréquemment retrouvées avec des *Acanthamoeba*, dans des environnements impliquant les humains, comme dans des tours de refroidissement par exemple. Les micro-algues constitueraient donc des hôtes potentiels pour de nouvelles espèces de virus appartenant à la famille des *Mimiviridae* (MONIER *et al.*, 2008b).

Un virus géant isolé dans les années 1990, infectant un biflagellé hétérotrophe, *Cafeteria roenbergensis*, de l'ordre des *Bicosoecales*, qui se nourrit de bactéries et de virus, et est répandu dans l'environnement marin (eau de surface, sédiments marins profonds), a été caractérisé en 2010. *Cafeteria roenbergensis* virus ou CroV, est constitué d'une capside icosaédrique d'environ 300 nm de diamètre (cf. Figure 1.31), qui contient un génome à ADN double-brin linéaire d'environ 730 kb, dont 544 gènes codent pour des protéines.

Parmi ces gènes, des gènes codant pour des facteurs de transcription, des enzymes de réparation de l'ADN, des photolyases, des ARNt ou encore pour une isoleucyl-ARNt-synthétase, qui ne sont habituellement pas retrouvés chez des virus, sont présents chez CroV. Ce virus, bien qu'il soit plus petit en terme de taille de capside et de génome, est phylogénétiquement proche de Mimivirus, et ferait partie de la famille des *Mimiviridae*, avec d'autres virus infectant des algues (FISCHER *et al.*, 2010).

C'est également le cas pour *Phaeocystis globosa* virus, ou PgV, un virus qui infecte également les algues, dont la capside et le génome sont plus petits que ceux de Mimivirus, et qui est classé dans la famille des *Mimiviridae* (cf. Section 1.4.2, SANTINI *et al.*, 2013).

Un virus infectant l'algue unicellulaire, *Haptolina ericina* (précédemment *Chrysochromulina ericina*), a été isolé en 1998, à partir d'échantillons marins prélevés à proximité des côtes Norvégiennes. Ce virus, appelé CeV, pour *Chrysochromulina ericina* virus, possède une capside icosaédrique de 160 nm de diamètre et a été initialement classé parmi les *Phycodnaviridae*. Son génome à ADN double-brin linéaire de 473 kb, dont 512 gènes codent pour des protéines, a été étudié récemment : cette étude a mis en évidence

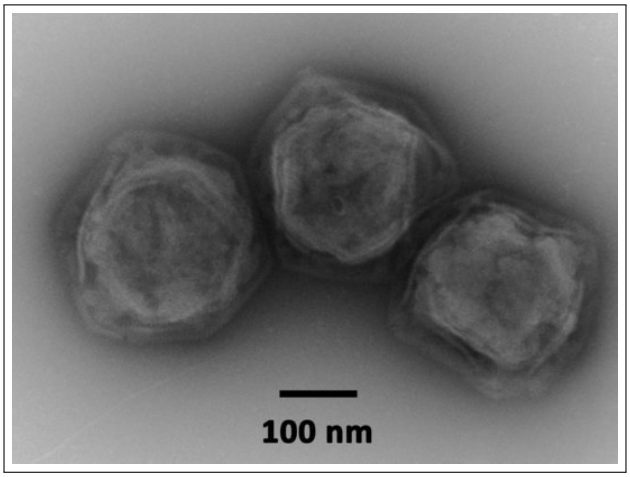


FIGURE 1.31 – *Cafeteria roenbergensis* virus observé au microscope électronique en coloration négative (©U. Mersdorf).

que CeV est phylogénétiquement proche de PgV, et qu'il appartiendrait (comme PgV) à la famille des *Mimiviridae*, formant une "sous-famille" composée de virus plus petits que Mimivirus et Mégavirus, infectant les algues (GALLOT-LAVALLÉE *et al.*, 2015).

A la lumière de ces découvertes, il est évident que les virus géants, ou des virus plus petits apparentés aux *Mimiviridae*, sont ubiquitaires dans l'environnement et infectent une grande diversité d'hôtes. Il devient donc nécessaire de prospecter toute une variété d'échantillons, provenant de différents environnements, et d'élargir la gamme des hôtes potentiels testés.

Les échantillons cliniques

Les virus géants sont majoritairement retrouvés dans des environnements où il y a des amibes, comme dans des sédiments marins, dans de l'eau douce, ou encore dans le sol. Le plus souvent, ces environnements sont en interaction avec les humains, comme, par exemple, des tours de refroidissement, où a été isolé Mimivirus. Il est donc judicieux de tester la présence de tels virus dans tous les environnements qui entourent les hommes, mais aussi, chez les hommes eux-mêmes.

C'est ainsi qu'en 2012, lors d'une étude sur le microbiome intestinal humain, un virus appartenant aux NCLDV a été isolé. Plusieurs échantillons de selles provenant de personnes saines européennes ou sénégalaises ont été mis en co-culture avec des amibes *Acanthamoeba polyphaga* : Senegalvirus est le plus grand virus trouvé dans des intestins humains. Ce virus possède une capside icosaédrique d'environ 194 nm de diamètre, qui embarque un génome à ADN double-brin de 373 kb. Des analyses phylogénétiques ont permis de le classer dans une nouvelle famille au sein des NCLDV : les *Marseilleviridae*. Ainsi, Senegalvirus est le premier virus des *Marseilleviridae* retrouvé chez l'homme (LAGIER *et al.*, 2012).

En 2013, ce sont des échantillons pulmonaires provenant de patients atteints de pneumonie qui ont été testés et mis en co-culture avec les amibes *A. polyphaga*. Cette fois, c'est un Mimivirus (isolat LBA111) qui a été découvert, faisant alors de ce virus le plus grand virus isolé à partir d'un échantillon humain. Ce Mimivirus LBA111 possède une capside icosaédrique d'environ 550 nm de diamètre, recouverte d'une couche dense de fibres et qui contient un génome à ADN double-brin de 1.23 Mb. Des analyses phylogénétiques le classent dans la famille des *Mimiviridae*, où il est proche de *Megavirus chilensis* (SAADI *et al.*, 2013a).

Après la découverte de ces deux virus dans des échantillons humains, des chercheurs ont voulu réitérer l'expérience sur un plus grand nombre d'échantillons. C'est pour cela qu'en 2013, plus de 1600 échantillons de selles provenant de personnes et d'enfants sains et malades (pneumonie ou gastroentérite) ont été mis en co-culture avec les amibes *A. polyphaga*. Sur le grand nombre d'échantillons testés, seul un prélèvement, provenant d'un patient atteint de pneumonie, a permis d'isoler un nouveau virus géant. Shan virus est constitué d'une capside icosaédrique d'environ 640 nm de diamètre, entourée d'une couche dense de fibres. Les analyses phylogénétiques le classent dans la famille des *Mimiviridae*, où il est proche de *Megavirus chilensis* et du Mimivirus LBA111, isolé chez un patient atteint de pneumonie (SAADI *et al.*, 2013b).

La même année, une équipe a recherché la présence de virus géants dans des échantillons de sang provenant de donneurs sains et asymptomatiques. Ces recherches ont permis l'identification d'un nouveau virus, le "Giant Blood Marseillevirus" ou GBM, qui ressemble à un *Marseilleviridae* : il possède une capside icosaédrique d'environ 220 nm de diamètre et un génome de 357 kb.

Les analyses phylogénétiques le classent bien dans la famille des *Marseilleviridae*, comme Senegalvirus, ce qui en fait le deuxième virus de la famille des *Marseilleviridae* à avoir été isolé chez des humains (POPGEORGIEV *et al.*, 2013).

Toujours en 2013, une équipe a recherché la présence de virus géants dans un tout autre type d'hôte : les sangsues appartenant à l'espèce *Hirudo medicinalis*, utilisées de façon médicale, notamment pour extraire du sang ou éviter la coagulation. Ces sangsues sont majoritairement trouvées dans de l'eau douce, et principalement dans les flaques d'eau, où l'on retrouve également certains hôtes des virus géants, comme les amibes. Dans leur habitat naturel, les sangsues sont en interaction constante avec les micro-organismes, qui peuvent coloniser leur tube digestif : *H. medicinalis* a donc un rôle probable de réservoir pour les virus, ceux-ci étant largement présents dans les environnements aquatiques et les sols. Ces sangsues étant utilisées pour des traitements médicaux, le risque de transmission de micro-organismes, potentiellement pathogènes pour l'homme, peut être problématique. Boughalmi *et al.* ont donc isolé des sangsues *Hirudo medicinalis* en Tunisie et en France, les ont disséquées, et leurs parties internes ont été mises en co-culture avec des amibes *A. polyphaga*. C'est ainsi qu'un nouveau virus géant, nommé Hirudovirus, a été isolé à partir du tube digestif d'une sangsue. Il

possède une capside icosaédrique d'environ 520 nm de diamètre, entourée de fibres, contenant un génome à ADN double-brin de 1.18 Mb, dont 922 gènes codent pour des protéines : des analyses phylogénétiques le classent parmi les *Mimiviridae* (BOUGHALMI *et al.*, 2013a).

Ces virus isolés à partir de prélèvements humains ou d'animaux, pourraient également provenir d'amibes infectées. En effet, les amibes sont ubiquitaires et peuvent être retrouvées dans une grande variété d'environnements, comme le sol, les flaques d'eau, ou encore les hôpitaux (SIDDIQUI *et al.*, 2012 ; RODRÍGUEZ-ZARAGOZA, 1994). Ces amibes infectées auraient donc pu parasiter les hommes, ou encore coloniser les animaux, et auraient libéré les particules virales isolées par la suite.

De nouvelles études sur la proportion des infections de ces virus sur les humains sont en cours et semblent former un nouveau domaine dans la virologie médicale. Il semblerait, d'après quelques études, que certains virus appartenant à la famille des *Mimiviridae* soient reliés à plusieurs cas de pneumonies, et que des virus appartenant à la famille des *Marseilleviridae* colonisent les hommes de façon asymptomatique (COLSON *et al.*, 2013a). Toutefois, un *Marseilleviridae* a été isolé récemment chez une patiente atteinte d'un lymphome (AHERFI *et al.*, 2016b). Cependant, l'effet pathologique de ces virus sur les humains est sujet à controverse : des études systématiques, qui ont été réalisées afin de rechercher la présence de ces virus dans des patients, se sont révélées négatives (ZHANG *et al.*, 2016 ; SAUVAGE *et al.*, 2014 ; GOODMAN, 2013 ; KUTIKHIN *et al.*, 2014 ; PHAN *et al.*, 2015 ; DARE *et al.*, 2008 ; VANSPAUWEN *et al.*, 2013). Le rôle de ces virus en tant que pathogènes humains reste donc à démontrer.

L'étude des virus géants est un domaine émergent en virologie. Ces virus sont capables d'infecter différents types d'eucaryotes, comme les amibes ou les algues, et sont retrouvés dans une grande variété d'environnements, mais aussi chez l'homme : il est donc nécessaire d'élargir les champs de recherche. Le criblage des échantillons environnementaux et cliniques à haut-débit permet d'augmenter, d'une part, le nombre d'hôtes testés, ainsi que le nombre d'échantillons, afin d'isoler un plus grand nombre de virus géants de façon plus rapide (KHALIL *et al.*, 2016).

La découverte de nouveaux virus géants, ainsi que leur caractérisation, permet d'agrandir des familles de virus déjà connues, comme les *Mimiviridae*, ou d'en découvrir de nouvelles au sein des grands virus à ADN ou NCLDV.

1.4.5 Les différentes familles de virus géants

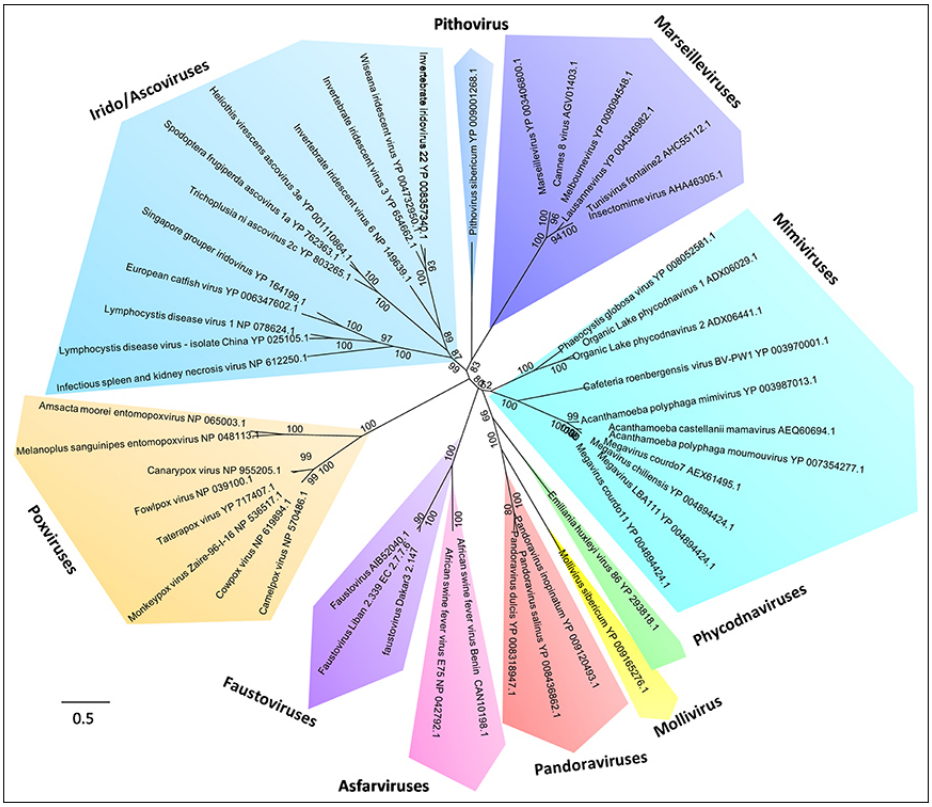


FIGURE 1.32 – Arbre phylogénétique des NCLDV qui comprend les différentes familles de virus géants récemment découvertes, comme les Pandoravirus, Pithovirus et Mollivirus, généré à partir des séquences de l'ARN polymérase de ces virus (AHERFI *et al.*, 2016a).

Les *Mimiviridae*

La famille des *Mimiviridae* a été établie en 2003, après la découverte du premier virus géant, Mimivirus, un virus infectant les amibes (LA SCOLA *et al.*, 2003).

De nombreux autres virus présentant les mêmes caractéristiques que Mimivirus ont été découverts depuis, et ont été isolés à partir d'échantillons provenant d'environnements différents : sol, eau de mer, eau douce (rivières, tours de refroidissement), compost, solution de lavage pour des lentilles de contact, selles et échantillons pulmonaires humains, et même dans une sangsue. Ces échantillons, prélevés dans différents pays comme la France, l'Angleterre, le Chili, le Brésil et la Tunisie, suggèrent que ces virus sont ubiquitaires dans l'environnement (AHERFI *et al.*, 2016a).

Les virus appartenant à la famille des *Mimiviridae* possèdent des capsides icosaédriques qui mesurent entre 370 et 750 nm de diamètre (en comptant les fibres), et sont recouvertes d'une couche de fibres, dont la taille peut varier en fonction des virus : par exemple, les fibres de *Megavirus chilensis* mesurent 80 nm, alors que celles de Mimivirus mesurent 125 nm.

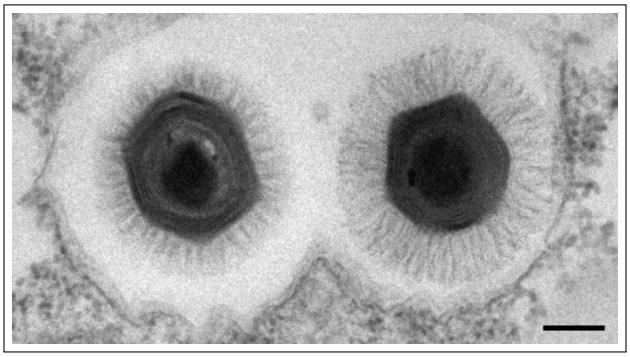


FIGURE 1.33 – Mimivirus et Mégavirus dans la même vacuole (co-infection) observés au microscope électronique : à gauche, Mégavirus dont les fibres sont plus courtes que celles de Mimivirus, situé à droite. Barre d'échelle : 200 nm (ARSLAN *et al.*, 2011).

Les génomes de ces virus, à ADN double-brin, ont une taille comprise entre 1.02 et 1.26 Mb, sont riches en A+T (entre 72 et 75%) et possèdent entre 930 et 1120 gènes qui codent pour des protéines. Ces virus contiennent des gènes codant pour des protéines impliquées dans la transcription, la réparation de l'ADN, la synthèse de nucléotides, ou encore le métabolisme des polysaccharides et des lipides. De plus, ces virus codent pour des gènes impliqués notamment dans le métabolisme des acides aminés, jamais trouvés chez des virus auparavant (cf. Section 1.4.1, AHERFI *et al.*, 2016a).

Les analyses phylogénétiques des *Mimiviridae*, séparent cette famille en deux sous-groupes : le premier sous-groupe est composé des virus infectant les amibes, comme Mimivirus. Le deuxième regroupe des virus proches du premier sous-groupe, plus petits en taille de capside et de génome, et qui infectent des unicellulaires marins, comme CroV, PgV ou CeV (cf. Section 1.4.4).

Ce premier sous-groupe est divisé en trois lignées : A, B, et C (cf. Figures 1.32 et 1.34). La lignée A comprend des virus comme, par exemple, Mimivirus (LA SCOLA *et al.*, 2003), Mamavirus (LA SCOLA *et al.*, 2008), Hirudovirus, isolé à partir d'une sangsue (BOUGHALMI *et al.*, 2013a), Lentille virus qui a été isolé à partir d'une solution de lavage pour les lentilles de contact (COHEN *et al.*, 2011) et Samba virus, récemment isolé au Brésil (CAMPOS *et al.*, 2014).

La lignée B ne comprend que quelques virus, dont Moumouvirus et Monve virus, isolés à partir de tours de refroidissement en France, Istres virus et Cassis49 virus, isolés à partir d'échantillons de sol (PAGNIER *et al.*, 2013), et Goulette virus, isolé à partir d'eau de mer en Tunisie (BOUGHALMI *et al.*, 2013b).

La lignée C comprend des virus comme *Megavirus chilensis* (ARSLAN *et al.*, 2011), LBA111 isolé à partir d'un échantillon pulmonaire d'un patient atteint de pneumonie (SAADI *et al.*, 2013a), Shan virus, isolé à partir de selles d'un patient ayant une pneumonie (SAADI *et al.*, 2013b), et Mont1 virus, isolé à partir d'un échantillon de sol prélevé dans une montagne en Tunisie (GAIA *et al.*, 2014).

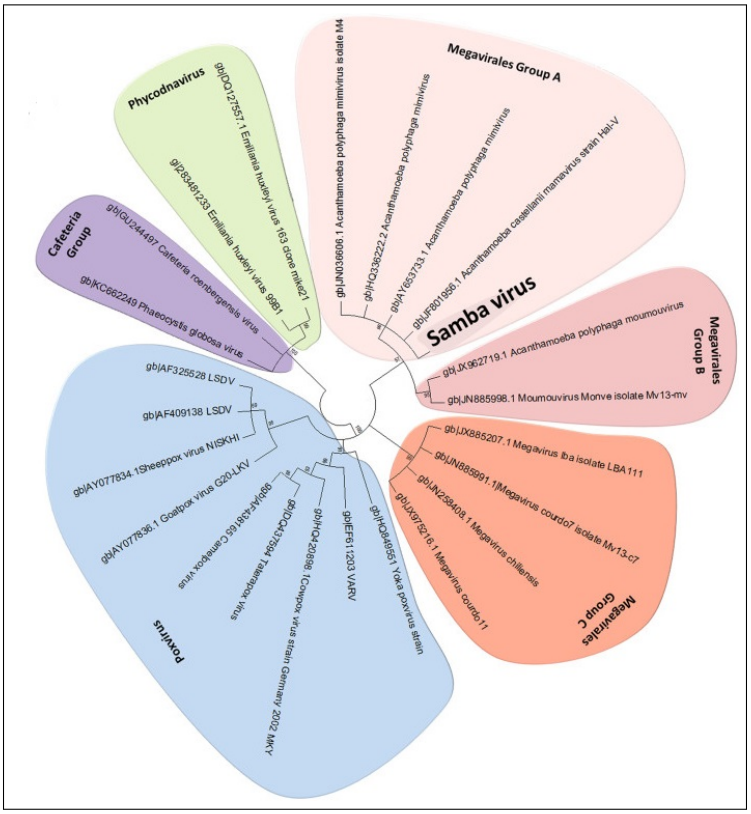


FIGURE 1.34 – Arbre phylogénétique de la famille des *Mimiviridae* généré à l'aide des séquences de la ribonucléotide réductase de ces virus et d'autres virus appartenant au groupe des NCLDV (CAMPOS *et al.*, 2014).

Le cycle de réplication des *Mimiviridae* se produit de la même façon pour les virus appartenant à cette famille (cf. Figure 1.35). Premièrement, la particule virale va être internalisée dans la cellule par phagocytose : l'amibe va déployer ses pseudopodes et entourer le virus, qui va se retrouver dans un phagosome. Les fibres protéoglycaniques entourant la particule, qui imitent les polysaccharides capsulaires des bactéries, jouent probablement un rôle d'appât pour l'amibe, qui croyant trouver une bactérie dans le milieu, sa "nourriture", va l'ingérer. Une fois "piégé" dans la vacuole, qui se trouve dans le cytoplasme de l'amibe, le virus, en réponse à un signal de nature inconnue, va ouvrir sa particule. L'ouverture se fait par le biais de la stargate : une fois ouverte, la membrane interne de la particule va fusionner avec la membrane de la vacuole, créant alors une zone d'échange entre l'intérieur du virus et le cytoplasme de la cellule hôte. Le matériel génomique du virus et les protéines associées, contenus dans le nucléoïde qui apparaît dense aux électrons en microscopie électronique, vont alors être libérés dans le cytoplasme de l'amibe : cette étape constitue l'initiation de l'infection (CLAVERIE *et al.*, 2009 ; ABERGEL *et al.*, 2015).

La transcription précoce des gènes viraux, ainsi que la réplication de l'ADN viral, vont être initiés par ce nucléoïde (ou "*core*") : les ARN messagers nouvellement synthétisés sont alors libérés dans le cytoplasme, où ils pourront être traduits en protéines à l'aide de la machinerie de traduction de l'hôte (ribosomes). La réplication de l'ADN viral a lieu

autour de ce nucléoïde, où l'usine de réplication virale va se mettre en place (FRIDMANN-SIRKIS *et al.*, 2016).

Les virus en cours de formation émergent de cette usine, où leur membrane interne est synthétisée à partir des membranes du réticulum endoplasmique de la cellule hôte (MUTSAFI *et al.*, 2013 ; KUZNETSOV *et al.*, 2013).

L'empaquetage du génome viral s'effectue lorsque les capsides sont formées, par le biais d'une ouverture située sur une face de l'icosaèdre, différente de la stargate (ZAUBERMAN *et al.*, 2008).

Enfin, l'assemblage de la couche de fibres en périphérie des capsides virales constitue la dernière étape de maturation des nouveaux virions (KUZNETSOV *et al.*, 2013).

La lyse cellulaire permet finalement la libération d'environ 1000 particules virales, disséminées dans le milieu extra-cellulaire (CLAVERIE *et al.*, 2009 ; ABERGEL *et al.*, 2015).

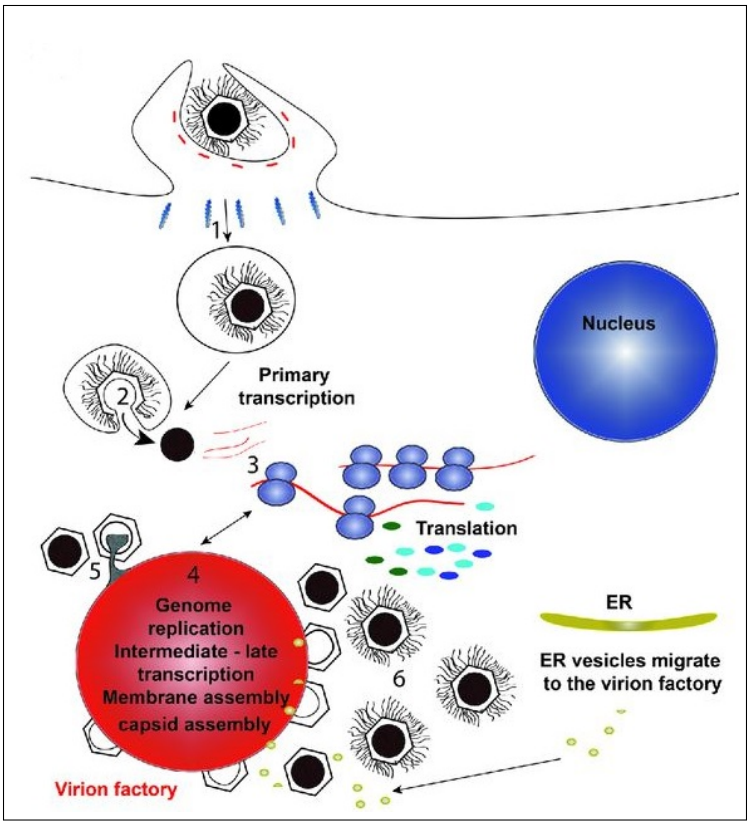


FIGURE 1.35 – Représentation schématique du cycle infectieux de Mimivirus dans l'amibe (ABERGEL *et al.*, 2015).

Les virophages

Avec la découverte de ces virus géants, qui appartiennent à la famille des *Mimiviridae*, des virus d'un nouveau genre ont été isolés.

Le premier de ces virus, isolé en 2008 à partir d'un échantillon prélevé dans une tour de refroidissement à Paris, en même temps que Mamavirus, a été nommé Sputnik.

Ce virus possède une capside icosaédrique d'environ 50 nm de diamètre, qui contient un génome à ADN double-brin circulaire de 18 kb, avec 21 gènes codant pour des protéines, dont 13 n'ont pas d'homologues dans les bases de données. Ce qui est surprenant, c'est que les cellules doivent être préalablement infectées par un virus de la famille des *Mimiviridae*, comme Mimivirus ou Mamavirus, pour que Sputnik puisse se multiplier. La production de Sputnik s'effectue dans l'usine virale, établie par Mimivirus ou Mamavirus au cours de leur cycle infectieux. Cette production s'accompagne de phénotypes visibles sur les virus géants : il y a plus de Mamavirus ou Mimivirus malformés, il y a moins de virus infectieux, et enfin, la lyse des cellules est diminuée après 24 heures post-infection. Il s'agit du premier cas publié, où un virus parasite et utilise l'usine virale d'un autre virus pour se multiplier et se propager : Sputnik est donc le membre fondateur d'un nouveau groupe de virus, aussi connu sous le nom de "virophage", par analogie avec les bactériophages, les virus parasitant les bactéries (La Scola *et al.*, 2008).

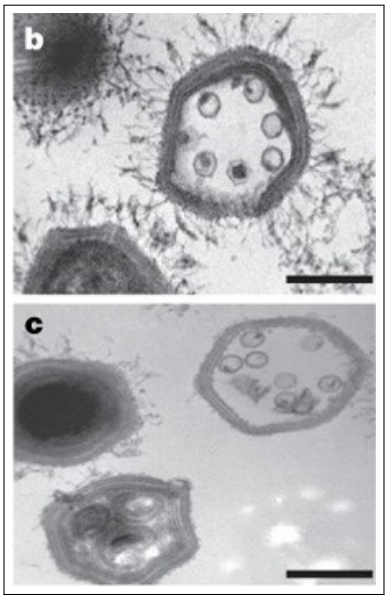


Figure 1.36 – Sputnik à l'intérieur de Mamavirus observé au microscope électronique : en haut, le virophage est visible dans le virion de Mamavirus, et en bas, l'image montre des particules de Mamavirus défectueuses. Barres d'échelle : 200 nm (La Scola *et al.*, 2008).

Plusieurs arguments suggèrent que Sputnik serait le premier virus représentant le groupe des virophages, et non pas un virus satellite (cf. Figure 1.37). En effet, la réplication de Sputnik s'effectue dans les usines virales de Mimivirus ou Mamavirus, indépendamment de la machinerie de transcription ou de réplication d'ADN de l'hôte cellulaire, alors que la réplication d'un virus satellite repose à la fois sur le virus auxiliaire et la cellule hôte. Sputnik est donc un parasite du virus géant, et non pas de l'amibe *Acanthamoeba*. De plus, des structures qui ressemblent aux structures en forme de tige-boucle de Mimivirus sont retrouvées dans le génome de Sputnik, suggérant que les gènes de ce virus sont transcrits par la machinerie de transcription codée par Mimivirus, qui se

trouve dans les usines virales, et sont sous le contrôle d'un promoteur tardif (CLAVERIE *et al.*, 2009).

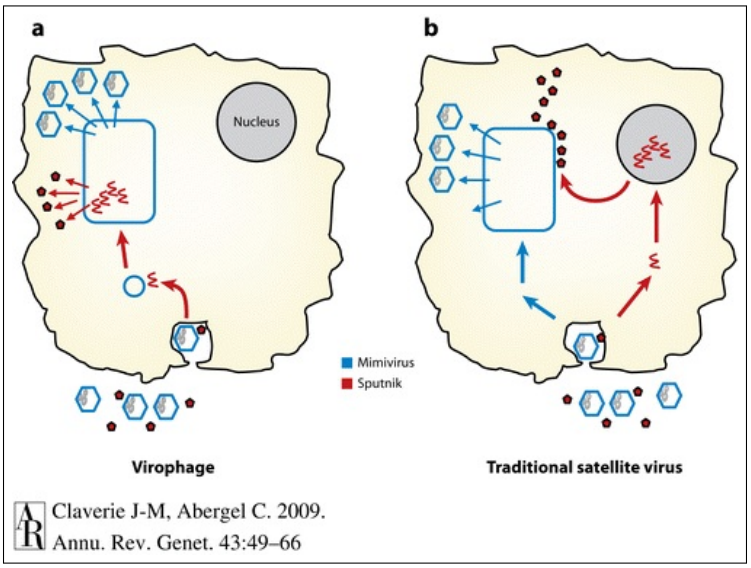


FIGURE 1.37 – Schéma du cycle de réplication de Sputnik opposé à celui d'un virus satellite : (a) la réplication et la transcription de Sputnik s'effectuent entièrement dans l'usine virale des *Mimiviridae* et (b) dans le cas d'un virus satellite, la réplication et la génération de transcrits s'effectuent dans le noyau de la cellule (CLAVERIE *et al.*, 2009).

En 2011, deux autres virophages ont été isolés conjointement avec leur virus hôte : Mavirus, le virophage de CroV, qui fait partie du deuxième sous-groupe de la famille des *Mimiviridae* (FISCHER *et al.*, 2011), et Sputnik2, isolé avec Lentille virus dans une solution de lavage pour lentilles de contact (COHEN *et al.*, 2011).

En 2013, un nouveau virophage a également été découvert chez PgV, un virus infectant les algues. Ce virophage a été appelé PgVV, et sa séquence génomique a été retrouvée dans le génome de PgV. Il semblerait que PgVV n'ait jamais été observé au cours du cycle infectieux de son virus hôte, et que son génome ne contienne pas de gène codant pour une protéine de capside, ou pour une protéine qui pourrait ressembler à une protéine de capside. PgVV ne pourrait donc pas exister sous forme "libre", mais existerait sous forme d'épisome ou serait intégré aléatoirement dans le génome de son hôte, sous la forme d'un provirus (SANTINI *et al.*, 2013).

Un an plus tard, un nouveau virophage, isolé en même temps que son virus hôte a été décrit : le Rio Negro Virophage, ou RNV, a été isolé avec un *Mimiviridae* de la lignée A, Samba virus, à partir d'un échantillon d'eau prélevé dans une rivière au Brésil (CAMPOS *et al.*, 2014).

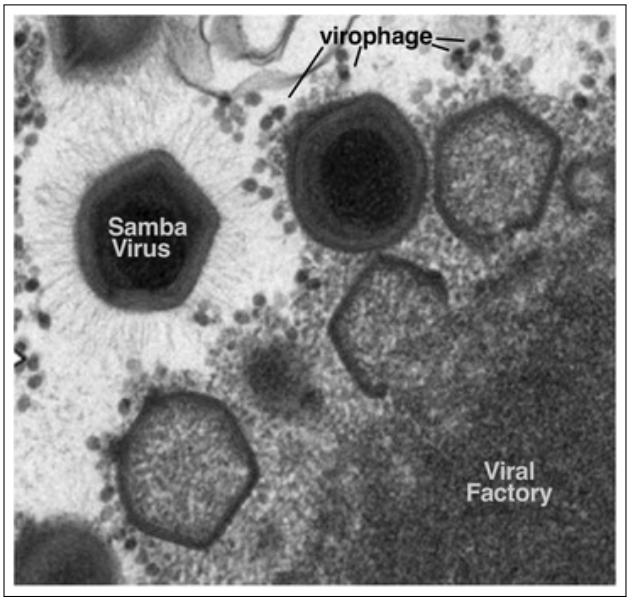


FIGURE 1.38 – Samba virus et de son virophage observés au microscope électronique : les virions sont en cours de synthèse en périphérie de l'usine à virions dans le cytoplasme de la cellule hôte, tout comme le virophage (KATZOURAKIS *et al.*, 2014).

Les virophages découverts jusqu'alors ont été isolés conjointement avec seulement des virus de la lignée A, ou des virus du deuxième sous-groupe des *Mimiviridae*. Ils possèdent une capside icosaédrique d'environ 50 nm de diamètre et ont un génome à ADN double-brin, d'environ 18-19 kb, qui contient une vingtaine de gènes codant pour des protéines.

Toutefois, en 2014, le premier virophage infectant un virus de la lignée C des *Mimiviridae* a été isolé, à partir d'un échantillon de sol prélevé en Tunisie. Ce virus, appelé Zamilon, possède une capside icosaédrique d'environ 50 nm de diamètre, et un génome à ADN double-brin de 17 kb, qui contient 20 gènes codant pour des protéines. Contrairement aux autres virophages, Zamilon ne semble pas affecter son virus hôte : en effet, il n'y a pas d'augmentation du nombre de virus anormaux ou mal-formés, et le cycle viral de l'hôte dans la cellule se passe sans anomalie détectable. D'autres études et observations sont évidemment nécessaires, afin de déterminer l'effet de ce virophage sur les *Mimiviridae*. Sputnik est capable de se répliquer lors du cycle infectieux des *Mimiviridae* des trois lignées (GAIA *et al.*, 2013) : en revanche, Zamilon est incapable de se répliquer avec les virus appartenant à la lignée A (GAIA *et al.*, 2014).

Une hypothèse, sur le fait que Zamilon est incapable d'infecter les virus de la lignée A, a été soulevée. Levasseur *et al.* suggèrent que les *Mimiviridae* de la lignée A portent un système un défense qui ressemble au système CRISPR-Cas ("*Clustered Regularly Interspaced Short Palindromic Repeat*"), présent chez les bactéries et les archées, et qui leur confère une immunité face aux éléments génétiques reconnus comme des intrus. En étudiant les génomes des virus résistants de la lignée A, ils ont repéré une séquence répétée appartenant à Zamilon, insérée dans un "opéron", appelé MIMIVIRE ("*Mimivirus Virophage Resistance Element*"). En réprimant cette séquence répétée, ainsi que les

gènes du système MIMIVIRE, la susceptibilité des virus appartenant à la lignée A face à une infection par Zamilon est restaurée (LEVASSEUR *et al.*, 2016b).

Jean-Michel Claverie et Chantal Abergel, proposent un modèle alternatif qui permettrait d'expliquer cette résistance : ils suggèrent que des interactions de type protéine-protéine pourraient être impliquées dans l'inhibition de la réplication de Zamilon dans les virus de la lignée A. En analysant la séquence de Zamilon insérée dans un gène de Mimivirus (R349), ils repèrent un motif peptidique, composé des acides aminés Aspartate-Asparagine-Glutamate-Serine (ou "DNES") dans ces deux séquences. Aucun homologue de cette séquence n'a pu être retrouvé chez Sputnik, ni chez les virus appartenant aux lignées B et C des *Mimiviridae*. Cette séquence d'acides aminés DNES est prédite pour former une boucle flexible et hydrophile, située à la surface de la protéine : la présence de cette boucle à la fois chez Zamilon et chez un *Mimiviridae* pourrait interférer négativement lors de l'infection, et réprimer l'infection du virophage (CLAVERIE *et al.*, 2016a).

Des études supplémentaires sont toutefois nécessaires pour comprendre le mode d'infection de ces virophages, ainsi que les systèmes qui permettent aux virus hôtes d'échapper à cette infection.

Les *Marseilleviridae*

La famille des *Marseilleviridae* est composée de grands virus à ADN, appartenant au groupe des NCLDV, et infectant les amibes du genre *Acanthamoeba*.

Le membre fondateur de cette famille, Marseillevirus, a été isolé en 2007 en co-culture avec des amibes *Acanthamoeba castellanii*, à partir d'un échantillon d'eau prélevé dans une tour de refroidissement à Paris. Marseillevirus possède une capside icosaédrique d'environ 250 nm de diamètre, empaquetant un génome à ADN double-brin circulaire de 368 kb, dont 457 gènes ont été prédits comme codant pour des protéines (BOYER *et al.*, 2009).

Au fil des ans, de nouveaux virus appartenant à cette famille ont été isolés dans des échantillons variés : eau, sol, insecte, selles et sang humains.

La famille des *Marseilleviridae* semble se composer de cinq lignées :

— la lignée A contient Marseillevirus, Cannes8 virus, Fontaine Saint-Charles virus, tous isolés à partir d'échantillons d'eau, Tokyovirus et Melbournevirus, isolés à partir de sédiments, Senegalvirus, isolé à partir d'un échantillon de selles humaines, et Giant Blood Marseillevirus, trouvé dans des échantillons de sang, provenant de donneurs sains (AHERFI *et al.*, 2014b ; DOUTRE *et al.*, 2014 ; TAKEMURA, 2016 ; POPGEORGIEV *et al.*, 2013) ;

— la lignée B regroupe Lausannevirus et Port-Miou virus, isolés à partir d'échantillons d'eau (THOMAS *et al.*, 2011 ; DOUTRE *et al.*, 2015) ;

— la lignée C rassemble Tunisvirus, isolé à partir d'un échantillon d'eau prélevé à Tunis, et Insectomime virus, isolé à partir du tube digestif d'une larve d'insecte (diptère *Eristalis tenax*) prélevée à Tunis, dans un réservoir d'eau stagnante (AHERFI *et al.*, 2014a ; BOUGHALMI *et al.*, 2013a) ;
— la lignée D contient Brazilian marseillevirus, isolé à partir d'un échantillon d'eau prélevé dans une station d'épuration au Brésil (DORNAS *et al.*, 2015 ; DORNAS *et al.*, 2016) ;
— la lignée E contient Golden marseillevirus, isolé à partir d'un échantillon d'eau de mer contenu dans une moule *Limnoperna fortunei* au Brésil (DOS SANTOS *et al.*, 2016).

Ces virus possèdent tous une capside icosaédrique mesurant entre 190 et 250 nm de diamètre dépourvue de fibres et présentent des génomes à ADN double-brin qui peuvent être circulaires ou linéaires avec des séquences terminales répétées et circulairement permutées, dont la taille est comprise entre 346 et 386 kb.

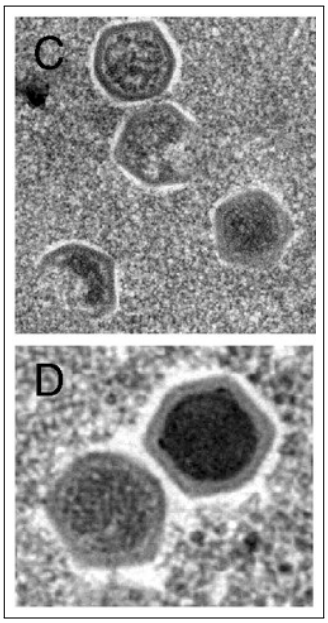


FIGURE 1.39 – Marseillevirus observé au microscope électronique : en haut, des virions en cours de synthèse, et en bas, un virion mature (à droite) se trouve à côté d'un virion en cours de formation (à gauche) (BOYER *et al.*, 2009).

Ces génomes ont un pourcentage en G+C compris entre 43 et 45% et possèdent entre 455 et 486 gènes prédits comme codant pour des protéines. Parmi les familles de protéines identifiées, on retrouve des protéines contenant des domaines répétés de type MORN ("*Membrane Occupation and Recognition Nexus*"), qui favorisent les interactions de type membrane-membrane ou membrane-cytosquelette.

Étonnamment, des protéines homologues à des histones ont été retrouvées dans les génomes de certains *Marseilleviridae* : elles pourraient être impliquées dans la compaction de l'ADN viral avant son empaquetage dans la capside.

Contrairement aux virus de la famille des *Mimiviridae*, les *Marseilleviridae* ne possèdent pas, à ce jour, d'ARNt ou d'amino-acyl-ARNt synthétases, et aucun virophage associé n'a été retrouvé (AHERFI *et al.*, 2014b).

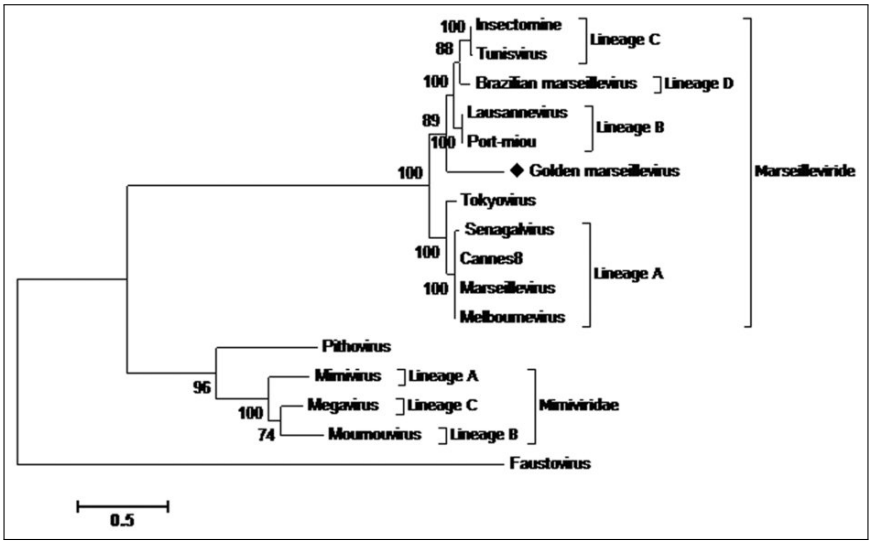


FIGURE 1.40 – Arbre phylogénétique de la famille des *Marseilleviridae* généré à partir de cinq arbres phylogénétiques, avec les séquences de l'ADN polymérase, de la protéine majeure de capside, et de trois hélicases (DOS SANTOS *et al.*, 2016).

Enfin, la détection d'un *Marseilleviridae* dans des ganglions lymphatiques d'une patiente atteinte d'un lymphome a été rapportée cette année (AHERFI *et al.*, 2016b).

Ce n'est pas la première fois qu'un virus appartenant à cette famille est détecté dans des échantillons prélevés sur des humains : Senegalvirus a été isolé à partir de selles d'un homme en bonne santé, et Giant Blood Marseillevirus a été retrouvé dans le sang de donneurs sains.

La question de la pathogénicité potentielle de ces virus sur les hommes est soulevée, et ouvre un nouveau champ de recherche pour des agents infectieux, encore inconnus, qui seraient retrouvés chez les hommes, en étudiant leur virome (AHERFI *et al.*, 2016b).

Toutefois, les infections des *Marseilleviridae* chez les hommes est sujet à controverse, et reste encore à démontrer, notamment par l'isolement de virus supplémentaires à partir d'échantillons cliniques humains (PHAN *et al.*, 2015 ; SAUVAGE *et al.*, 2014 ; GOODMAN, 2013 ; KUTIKHIN *et al.*, 2014).

La découverte de Noumeavirus à partir d'un échantillon environnemental, ainsi que des analyses comparatives de génomique, protéomique, et de son cycle infectieux seront détaillées dans le chapitre consacré aux *Marseilleviridae* (cf. Chapitre 3).

Les *Pandoraviridae*

Deux virus d'un nouveau genre, qui ne ressemblent à rien de connu dans le monde viral ont été décrits en 2013, et ont été appelés Pandoravirus. *Pandoravirus salinus* a été

isolé à partir d'un échantillon de sédiments marins, prélevé au Chili, et *Pandoravirus dulcis* a été isolé à partir d'un échantillon de boue, prélevé dans un étang d'eau douce en Australie, en co-culture avec l'amibe *Acanthamoeba castellanii*.

Ces deux virus possèdent une morphologie identique, unique dans le monde viral. En effet, ces virus ne possèdent pas de capside à symétrie hélicoïdale ou icosaédrique : les particules sont ovoïdes, et mesurent entre 0.8 et 1.2 μm de longueur sur 0.5 μm de diamètre. Ces particules sont composées d'un tégument, formé de plusieurs couches, dont une couche intermédiaire dense aux électrons et formée d'un réseau de fibres parallèles, ainsi qu'une membrane interne, qui délimite l'intérieur de la particule. Les virions possèdent un pore apical, scellé par une sorte de bouchon diffus. Étonnamment, aucune structure visible ne semble délimiter le génome, situé dans l'intérieur diffus de la particule.

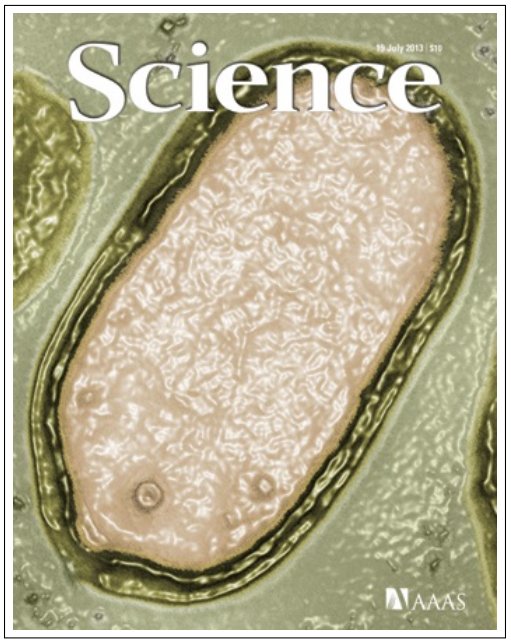


FIGURE 1.41 – Couverture du journal Science mettant en avant *Pandoravirus salinus*, en Juillet 2013 (©Science).

Le génome de ces virus est composé d'ADN double-brin, mesure 2.77 Mb pour *P. salinus* et 1.9 Mb pour *P. dulcis* et est riche en G+C (60%). L'alignement de ces deux génomes montre une colinéarité presque parfaite, seulement interrompue par quatre segments uniques à *P. salinus*, démontrant que le contenu génomique de *P. dulcis* constitue un sous-ensemble de celui de *P. salinus*. Ces génomes contiennent entre 1 502 (*P. dulcis*) et 2 556 (*P. salinus*) gènes qui codent pour des protéines, dont plus de 90% n'ont pas d'homologues dans les bases de données. L'analyse protéomique des particules indique que *P. salinus* se compose de 266 protéines, dont 56 correspondent à des protéines de l'amibe *A. castellanii* : parmi les 210 protéines "propres" à *P. salinus*, 80% n'ont aucun homologue dans les bases de données. La quasi totalité des protéines présentes dans

les particules de *P. salinus* (93%) ont un homologue retrouvé chez *P. dulcis*, démontrant que ces deux virus distincts appartiennent bien à la même famille de virus, les *Pandoraviridae*.

Les analyses phylogénétiques classent les Pandoravirus dans le groupe des NCLDV, dans une nouvelle famille : les *Pandoraviridae*.

On retrouve, dans les génomes des Pandoravirus, des gènes codant pour une ADN polymérase, une enzyme de coiffe pour les ARNm, plusieurs sous-unités de l'ARN polymérase, des facteurs de transcription, des hélicases, des kinases, des phosphatases, mais aussi deux amino-acyl ARNt synthétases et trois ARNt. En revanche, les Pandoravirus ne possèdent pas tous les gènes codant pour des protéines impliquées dans la machinerie de réplication et de réparation de l'ADN, comme des ligases, des topoisomérases, ou une "*DNA sliding clamp*", une protéine qui se lie à l'ADN polymérase et qui l'empêche de se dissocier du brin d'ADN matrice lorsqu'elle est en cours de synthèse d'ADN.

De plus, les Pandoravirus ne possèdent pas de gène codant pour une "*Major Capsid Protein*" (MCP), ou protéine majeure de capside, conservée chez tous les virus icosaédriques à ADN double-brin infectant les eucaryotes. Les *Poxviridae* qui ne possèdent pas non plus de capside icosaédrique, codent pour une MCP, qui n'est pas embarquée dans la particule, mais protéolysée lors de la maturation des virions (Bahar *et al.*, 2011).

Étonnamment, environ 9% des gènes de *P. salinus* qui ont un homologue dans les bases de données (soit 16 gènes sur 186), contiennent un ou plusieurs introns, qui ne ressemblent pas à des introns autocatalytiques ("*self-splicing*") et sont délimités par les séquences 5'-GT et 3'-AG. Ces introns semblent être épissés par la machinerie cellulaire de l'hôte *Acanthamoeba*, ce qui suggère qu'une partie du génome de ces virus est transcrite dans le noyau de la cellule hôte.

L'absence d'une machinerie de transcription codée par les Pandoravirus dans les particules virales, ainsi que la présence d'introns épissables, confirment que le noyau de la cellule hôte est impliqué dans le cycle de réplication des ces virus, contrairement aux *Mimiviridae* qui transportent leur machinerie de transcription dans leurs virions, leur permettant ainsi de se répliquer sans passer par le noyau cellulaire (Philippe *et al.*, 2013 ; Abergel *et al.*, 2015).

Le cycle infectieux des Pandoravirus (cf. Figure 1.42) commence par l'internalisation des particules dans le cytoplasme des cellules hôtes, à l'intérieur de phagosomes. Le virus, par un procédé encore inconnu, perd son bouchon, ce qui permet alors à sa membrane interne de fusionner avec la membrane de la vacuole, créant une continuité entre l'intérieur de la particule et le cytoplasme de l'hôte. Cette continuité permet un échange entre l'hôte et son parasite, le génome pouvant être libéré dans le cytoplasme et initier sa réplication. Le noyau de la cellule hôte va ensuite se désorganiser progressivement, et va ensuite disparaître. Une zone se forme alors dans le cytoplasme, où toutes les structures sub-cellulaires, comme les mitochondries, sont exclues : c'est

l'usine virale. La formation des nouveaux virions est étonnante : l'intérieur, ainsi que l'enveloppe entourant la particule, sont assemblés simultanément, comme s'ils étaient "tricotés". Des particules à différents stades de maturité sont observées dans l'usine virale, dans le cytoplasme de l'hôte. Des particules matures peuvent être libérées dans le milieu extra-cellulaire par exocytose, mais la fin du cycle est marquée par la lyse des cellules remplies de particules, libérant environ une centaine de virions dans le milieu (ABERGEL *et al.*, 2015).

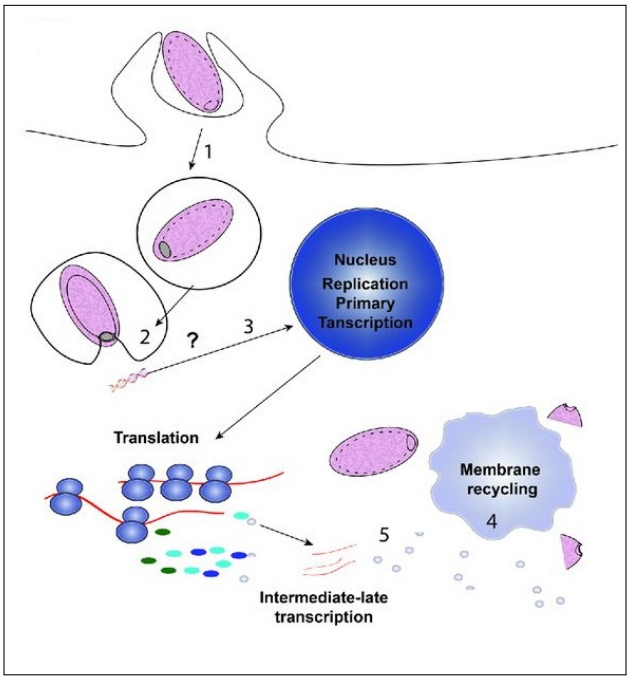


FIGURE 1.42 – Représentation schématique du cycle infectieux des Pandoravirus dans l'amibe (ABERGEL *et al.*, 2015).

Pandoravirus salinus et *P. dulcis* ressemblent à un endosymbionte d'*Acanthamoeba*, décrit quelques années plus tôt, en 2008, qui n'avait toutefois pas été caractérisé. Cet endosymbionte, appelé KLaHel, a été isolé à partir d'un échantillon prélevé sur une patiente atteinte de kératite amibienne, et son observation au microscope optique a révélé une particule ovoïde d'environ 1 μm de long, qui possède un pore apical (cf. Figure 1.43). KLaHel se réplique dans les amibes du genre *Acanthamoeba*, et n'est pas cultivable dans les différents milieux nutritifs testés, adaptés à la croissance bactérienne et fongique (SCHEID *et al.*, 2010).

Ce n'est qu'après la publication des Pandoravirus, ainsi qu'avec des analyses supplémentaires de cet endosymbionte, que KLaHel a été caractérisé comme un virus géant, qui appartiendrait à la famille des *Pandoraviridae* (SCHEID *et al.*, 2014), et le séquençage de son génome complet a été réalisé récemment (ANTWERPEN *et al.*, 2015).

Ce nouveau membre de la famille des *Pandoraviridae* (cf. Figure 1.32), nommé *Pandoravirus inopinatum*, possède un génome à ADN double brin de 2.2 Mb, riche en G+C (60%), qui contient 1 902 gènes prédits comme codant pour des protéines. Ce génome est identique à 85% au génome de *P. salinus* et à 89% à celui de *P. dulcis*, et environ

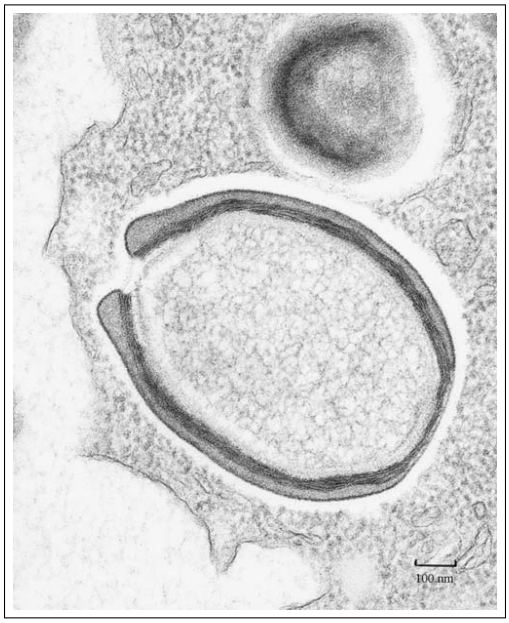


FIGURE 1.43 – L'endosymbionte KLaHel observé au microscope électronique, qui a une morphologie typique des Pandoravirus (SCHEID *et al.*, 2010).

73% de ses gènes n'ont pas d'homologues autres que les gènes de Pandoravirus dans les bases de données de séquences (ORFans) (SCHEID, 2016).

La découverte de nouveaux Pandoravirus à partir d'échantillons environnementaux, ainsi que des analyses comparatives de génomique, de protéomique, et de leurs cycles infectieux seront détaillées dans le chapitre consacré aux *Pandoraviridae* (cf. Chapitre 4).

Les *Pithoviridae*

Récemment, un nouveau virus géant infectant *A. castellanii*, nommé *Pithovirus sibericum*, a été découvert à partir d'un échantillon prélevé dans le pergélisol sibérien, un sol qui reste toujours gelé, en Russie. Cet échantillon a été prélevé à une profondeur de 30 mètres dans le pergélisol, ce qui correspond à un "âge" estimé entre 34 000 et 37 000 ans selon une datation au carbone 14 (ou radiocarbone).

Les particules de *P. sibericum* présentent une morphologie ovoïde, et sont facilement visibles au microscope optique, tout comme les Pandoravirus. Étonnamment, ces particules sont encore plus imposantes que celles des Pandoravirus, considérés comme des géants parmi les géants, et peuvent mesurer jusqu'à 1.5 μm de long, pour un diamètre d'environ 0.5 μm. Les virions sont également entourés d'un tégument épais et dense aux électrons : l'enveloppe entourant les particules est étrangement constituée d'une couche de bandes parallèles, d'environ 60 nm d'épaisseur. Une membrane interne délimite l'intérieur des particules, qui apparaît diffus, et où aucune structure n'est observée. *P. sibericum* possède aussi un pore apical fermé par une sorte de bouchon de structure

particulière : il est constitué de fibres parallèles, formant une structure hexagonale en "nid d'abeille" lorsqu'il est observé en coupe transversale.

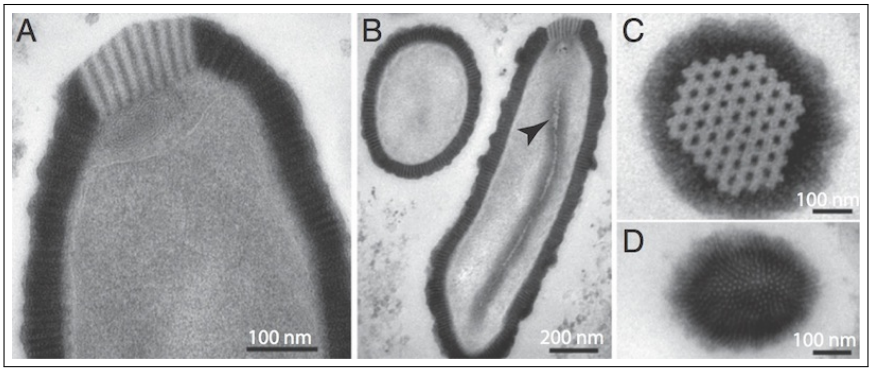


FIGURE 1.44 – *Pithovirus sibericum* observé au microscope électronique : (A) Apex de la particule montrant la structure particulière du tégument et du bouchon, (B) Particules de Pithovirus présentant une structure tubulaire au centre du compartiment interne, (C) Pithovirus vu du dessus montrant la structure en nid d'abeille de son bouchon, (D) Particule vue du dessous montrant l'organisation striée du tégument (LEGENDRE *et al.*, 2014).

P. sibericum possède un génome à ADN double-brin circulaire de 610 kb, riche en A+T (64%), ce qui est étonnamment petit étant donné la taille des particules. Un total de 467 gènes ont été prédits comme codant pour des protéines : *P. sibericum* code pour sa propre machinerie de transcription complète, ainsi que pour des protéines impliquées dans la réparation et la réplication de l'ADN et dans la synthèse des nucléotides. De plus, la présence d'un intron autocatalytique, capable de s'épisser de façon autonome sans passer par la machinerie de l'hôte, suggère que ce virus est capable de se répliquer exclusivement dans le cytoplasme de la cellule hôte.

Environ les deux tiers des protéines de *P. sibericum* n'ont pas d'homologues dans les bases de données et aucun gène codant pour une machinerie de traduction, ou pour une protéine de capside ("*Major Capsid Protein*") n'ont pu être retrouvés dans son génome. Enfin, comme chez les *Mimiviridae*, environ 70% des transcrits possèdent une séquence palindromique en 3', capable de former une structure en tige-boucle, et qui peut agir comme un signal de terminaison de la transcription et/ou comme un signal de polyadénylation.

Le cycle infectieux de *Pithovirus sibericum* (cf. Figure 1.45) est similaire à celui des Pandoravirus, à l'exception de l'étape de synthèse des particules. Celles-ci sont retrouvées dans un premier temps sous une forme rectangulaire avec un tégument de 25 nm d'épaisseur, avant leur maturation en particules ovoïdes constituées d'un tégument épais de 60 nm. Ces particules matures peuvent être libérées dans le milieu extra-cellulaire par exocytose, mais la fin du cycle est marquée par la lyse des cellules remplies de particules, libérant environ une centaine de virions dans le milieu (ABERGEL *et al.*, 2015).

Des analyses phylogénétiques montrent que *Pithovirus sibericum* fait partie du groupe des NCLDV et qu'il est plus proche des virus appartenant à la famille des *Marseilleviridae*

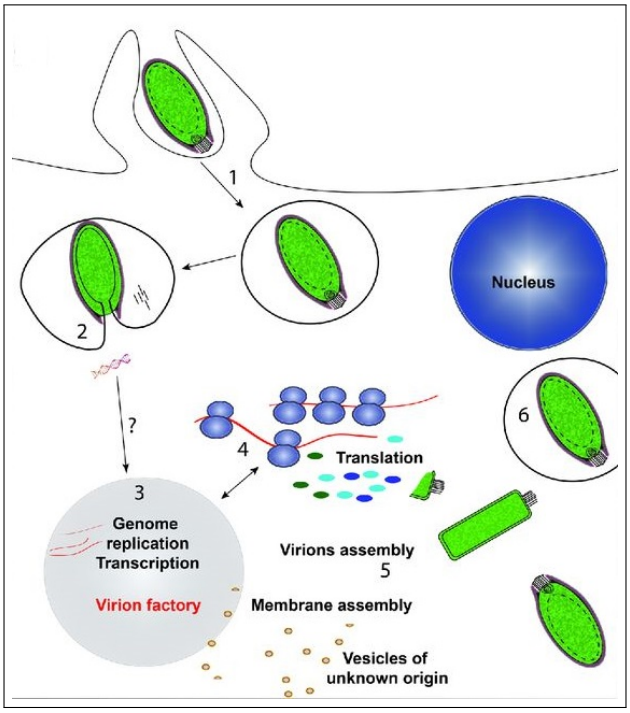


FIGURE 1.45 – Représentation schématique du cycle infectieux de *Pithovirus sibericum* dans l'amibe (ABERGEL *et al.*, 2015).

et des *Iridoviridae*, que des Pandoravirus qui ont une morphologie comparable. Cependant, *P. sibericum* constitue le membre fondateur d'une nouvelle famille de virus géants, les *Pithoviridae* (cf. Figure 1.32).

Jusqu'à ce jour, *Pithovirus sibericum* est le plus ancien des grands virus à ADN infectant les eucaryotes à avoir été isolé : cela soulève des questions sur les micro-organismes potentiellement pathogènes qui peuvent être réactivés, suite à une fonte du pergélisol due au réchauffement climatique ou à la prospection des richesses de ce sol (LEGENDRE *et al.*, 2014 ; ABERGEL *et al.*, 2015).

Comme pour les Pandoravirus, *Pithovirus sibericum* ressemble à un parasite isolé précédemment, mais qui n'avait pas été caractérisé comme un virus. L'endocytoparasite KC 5/2 a été isolé à partir d'un échantillon d'eau potable provenant d'un réservoir, lors de la recherche d'amibes. Des amibes du genre *Acanthamoeba* ont été trouvées dans cet échantillon, dont certaines étaient infectées par des parasites de morphologie ovoïde, mesurant entre 1 et 1.5 μm , et visibles au microscope optique. Ces parasites se multiplient dans le cytoplasme des amibes, et leur observation en microscopie électronique n'a pas permis de distinguer des événements de division. De plus, ces parasites ne semblent pas capables de croître *in vitro* sur les différents milieux nutritifs testés, utilisés en routine pour les bactéries, les fungi ou les eucaryotes, et il a été suggéré que ce parasite KC 5/2 appartienne au domaine des archées (HOFFMANN *et al.*, 1998).

Quelques années plus tard, de nouvelles études ont été menées sur ce parasite, dont une étude de sa réplication dans les amibes *Acanthamoeba* : il semblerait que plusieurs espèces d'*Acanthamoeba* soient susceptibles à une infection par ce parasite. Cette étude

révèle la formation d'une zone assez homogène dans le cytoplasme de l'amibe lors d'une infection avec ce parasite (cf. Figure 1.46) : en effet, cette zone ne correspond plus au cytoplasme "naturel" de l'hôte, d'aspect granuleux, que l'on retrouve chez les amibes non-infectées. Dans les amibes infectées, cette zone homogène est entourée de particules ovoïdes, qui mesurent entre 1 et 1.2 μm de long, qui contrairement aux bactéries, ne contiennent pas de structures évoquant des ribosomes ou à un nucléoïde. L'interaction observée entre ce parasite et son hôte amibien n'est pas comparable aux autres relations hôte-parasite décrites précédemment pour les bactéries, les fungi ou encore les eucaryotes, et les auteurs en concluent que ce parasite ne peut être classé dans aucun des ordres existants (MICHEL *et al.*, 2003).

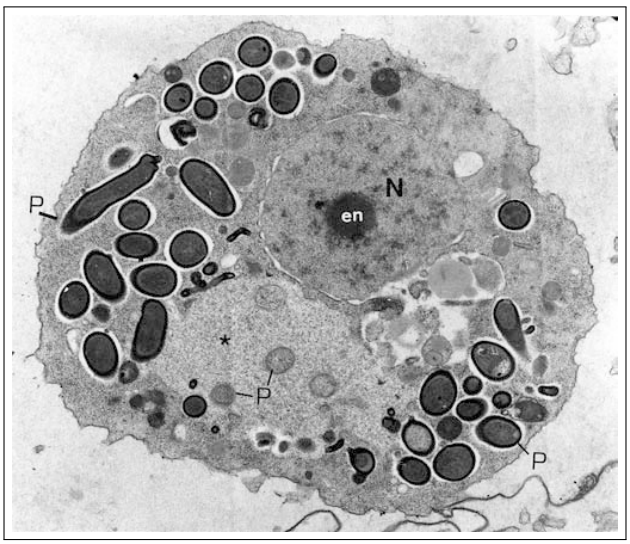


FIGURE 1.46 – Parasite KC 5/2 dans l'amibe observé au microscope électronique où (P) indique le parasite, (N) le noyau, (en) le nucléole et (*) la zone homogène du cytoplasme de la cellule (MICHEL *et al.*, 2003).

Enfin, un nouveau membre de la famille des *Pithoviridae* a été isolé récemment, à partir d'un échantillon d'eaux usées, prélevé en France, placé en co-culture avec des amibes *Acanthamoeba castellanii*. Ce virus, appelé *Pithovirus massiliensis*, possède une morphologie ovoïde mesurant environ 1.4 μm de long, est visible au microscope optique, et ses caractéristiques morphologiques sont identiques à celles de *P. sibericum*.

Son génome à ADN double-brin de 683 kb est riche en A+T (65%), et contient 520 gènes prédits comme codant pour des protéines. L'étude comparative des génomes qui a été réalisée sur ce Pithovirus moderne et sur *P. sibericum* a montré qu'une grande majorité des gènes de *P. sibericum* sont retrouvés dans le génome de *P. massiliensis* (437/467). De plus, l'arrangement des génomes est globalement similaire, ce qui démontre une certaine stabilité (cf. Figure 1.47). Enfin, le contenu génomique semble être conservé entre ces deux virus, même après plusieurs milliers d'années, ce qui suggère une pression sélective sur les gènes conservés, et que l'évolution des Pithovirus serait lente (LEVASSEUR *et al.*, 2016a).

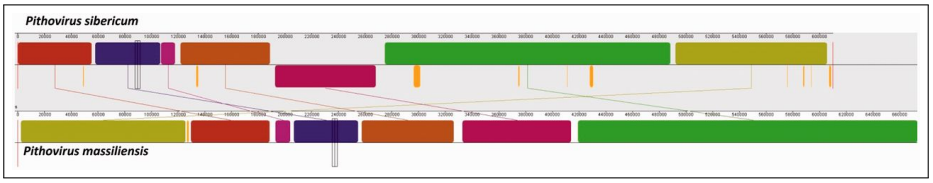


FIGURE 1.47 – Comparaison de l'organisation des génomes de *P. sibericum* et de *P. massiliensis* (LEVASSEUR *et al.*, 2016a).

Il est toutefois nécessaire d'isoler plus de membres de cette famille, pour pouvoir les étudier et comparer leurs génomes, afin d'évaluer la vitesse d'évolution de ces virus : l'étude du génome du parasite KC 5/2, contemporain de *P. massiliensis*, pourrait notamment apporter quelques éclaircissements sur ce point.

Mollivirus sibericum

Étonnamment, *Pithovirus sibericum* n'a pas été le seul virus géant isolé à partir du pergélisol : *Mollivirus sibericum* a été découvert à partir de ce même échantillon, cette fois encore, en co-culture avec l'amibe *A. castellanii*. Ce nouveau virus ne ressemble pourtant pas à son acolyte sibérien, ses virions étant d'apparence sphérique, et mesurant 500 à 600 nm de diamètre.

M. sibericum est visible au microscope optique, et les virions sont entourés de deux à quatre anneaux, espacés de quelques nanomètres, correspondant probablement à des fibres de différentes longueurs. Le tégument qui entoure la particule est composé de deux couches, de structures et densités différentes : la couche externe est constituée de bandes espacées, tangentes à la surface de la particule, et la couche interne apparaît dense aux électrons, et est composée d'un réseau de fibres qui ressemble à celui des Pandoravirus. La particule possède également une ouverture, formant une dépression circulaire d'environ 160 à 200 nm de diamètre et qui apparaît dépourvue de fibres. *M. sibericum* possède également une membrane qui délimite l'intérieur diffus de la particule, où aucune structure n'est observable.

Le génome de *M. sibericum* est constitué d'une molécule d'ADN double-brin linéaire de 651 kb, riche en G+C (60%), qui possède 523 gènes codant pour des protéines, dont 64% n'ont pas d'homologues dans les bases de données. Contrairement à *Pithovirus sibericum*, ce virus ne possède pas de séquences répétées dans son génome, et environ 4% des gènes qui codent pour des protéines possèdent de courts introns épissables, entourés des séquences 5'-GT et 3'-AG. Ces gènes qui possèdent des introns correspondent à des protéines retrouvées dans les particules, lors de l'analyse du protéome des particules virales, ce qui suggère que la machinerie d'épissage des introns de l'hôte reste fonctionnelle tout au long du cycle infectieux.

Le génome contient des gènes codant pour des enzymes impliquées dans le métabolisme de l'ADN (ADN polymérase, primase, hélicase, exonucléase, méthylase, recombinase), ainsi que des composants de la machinerie de transcription, comme quatre

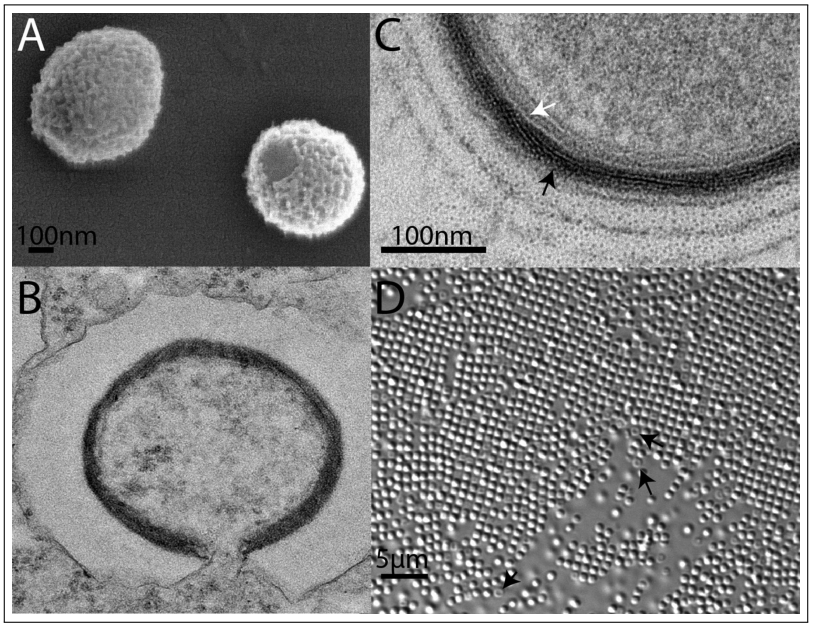


FIGURE 1.48 – *M. sibericum* observé au microscope électronique à transmission et à balayage, et au microscope optique : (A) Particules observées au microscope à balayage montrant la structure de l'apex dépourvu de fibres, (B) Image d'une particule dont la membrane interne a fusionné avec celle de la vacuole, (C) Image montrant le tégument entourant la particule composé d'un réseau de fibres (flèche noire) et d'une membrane interne (flèche blanche) et (D) Particules observées au microscope optique dont certaines présentent une dépression à l'apex (flèches) (LEGENDRE *et al.*, 2015).

sous-unités de l'ARN polymérase, ou un facteur d'élongation de la transcription.

L'étude du protéome des particules révèle que 230 protéines sont embarquées dans le virion, dont 54% n'ont pas d'homologues dans les bases de données. Des composants de la machinerie de transcription, codés par le virion, sont absents dans le protéome du virion, suggérant que le cycle de réplication de *M. sibericum* doit passer, au moins pour l'étape précoce, par le noyau de la cellule hôte. Étonnamment, une protéine homologue à la protéine de capside ("Major Capsid Protein") est retrouvée chez *M. sibericum* : cette protéine est habituellement retrouvée chez les virus à ADN qui possèdent une capside présentant une symétrie icosaédrique, ce qui n'est pas le cas de ce virus.

Le cycle de réplication de *M. sibericum* commence comme pour les autres virus géants (cf. Figure 1.49) : la particule virale est internalisée dans des vacuoles, situées dans le cytoplasme de la cellule hôte. L'ouverture de la particule n'a jamais été visualisée, mais elle est suivie par la fusion de la membrane interne du virus avec la membrane de la vacuole, conduisant à une zone d'échange entre le virus et le cytoplasme de l'hôte (cf. Figure 1.48).

Le génome de *M. sibericum* est libéré dans le cytoplasme, et sa migration dans le noyau de la cellule hôte a été suivie par microscopie à fluorescence. Le noyau de la cellule hôte va ensuite se désorganiser progressivement, la membrane nucléaire subissant de nombreuses invaginations : finalement, le noyau est déformé mais ne disparaît pas totalement.

La synthèse de nouveaux virions s'effectue à la périphérie de ce noyau déformé, et les organelles de la cellule hôte se retrouvent exclues de cette zone, qui se remplit d'un réseau de structures fibrillaires, semblant contribuer à la formation des particules virales. Les nouveaux virions sont formés par la synthèse simultanée du tégument et de l'intérieur de la particule, l'ouverture située à l'apex de la particule semblant se former en dernier, à l'opposé des Pandoravirus et de Pithovirus.

Enfin, des particules à différents stades de maturité sont retrouvées dans le cytoplasme entre 6 et 8 heures post-infection, dont certaines sont situées dans des vacuoles, suggérant une libération des particules matures par exocytose. La lyse des cellules, remplies de nouvelles particules, libère entre 200 et 300 virions dans le milieu extracellulaire, et marque la fin du cycle infectieux.

La stratégie de réplication nucléocytoplasmique de *M. sibericum* est différente de celle des *Mimiviridae* et des *Pithoviridae*, qui se répliquent entièrement dans le cytoplasme de la cellule hôte.

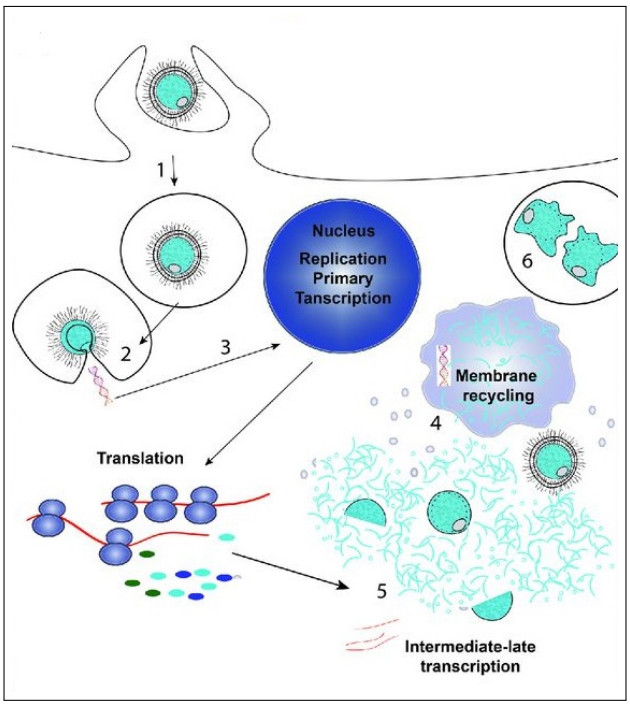


FIGURE 1.49 – Représentation schématique du cycle infectieux de *Mollivirus sibericum* dans l'amibe (ABERGEL *et al.*, 2015).

Les analyses phylogénétiques montrent que *Mollivirus sibericum* présente une affinité pour la famille des *Pandoraviridae* (cf. Figure 1.32), bien qu'il soit également le prototype d'une nouvelle famille de virus géants, les *Molliviridae* (LEGENDRE *et al.*, 2015 ; ABERGEL *et al.*, 2015).

Il faudra toutefois isoler d'autres virus comme *Mollivirus sibericum* afin de pouvoir les étudier, les comparer et déterminer leur place au sein des autres familles de virus géants.

La découverte de deux virus géants différents dans un échantillon de pergélisol sibérien vieux de 30 000 ans montre que ces virus ne sont pas une “nouveauté”, et que leur existence n’est pas aussi récente que ce que l’on pouvait penser. Cette découverte soulève de nombreuses questions sur l’origine de ces virus, ainsi que sur leur évolution.

1.4.6 Évolution et origine des virus géants

La découverte du premier virus géant, Mimivirus, a soulevé de nombreuses interrogations et de nombreux débats sur la définition de ce qu’est un virus, quelle est son origine évolutive, et de façon plus générale, qu’est-ce que le “vivant” ?

Plusieurs scénarios pour tenter d’expliquer l’origine évolutive des virus ont été proposés.

Après la découverte des premiers virus géants (Mimivirus et Mégavirus), des études comparatives ont été réalisées avec d’autres grands virus à ADN (CroV, CeV, ...) qui infectent des hôtes différents : les analyses phylogénétiques de l’ADN polymérase ainsi que celles des protéines de capsides, ont permis de montrer que ces virus appartenaient au même groupe.

Ce résultat suggère que ces grands génomes à ADN semblent être un trait phylogénétique hérité d’un ancêtre commun.

De plus, l’affinité phylogénétique entre ces virus est renforcée par le fait que leurs génomes codent pour leur propre machinerie de réplication et de transcription de l’ADN (ainsi que pour une enzyme de réparation des mésappariements de l’ADN), suggérant, encore une fois, la possibilité d’un ancêtre commun qui aurait probablement un génome encore plus grand que celui de Mimivirus ou de Mégavirus.

Ces virus, qui infectent une large gamme d’hôtes eucaryotes, pourraient avoir un ancêtre commun qui aurait précédé la divergence principale au sein des eucaryotes, c’est à dire, qui aurait existé avant (et peut-être même infecté) LECA (“*Last Eukaryotic Common Ancestor*”) ou le dernier ancêtre commun eucaryote.

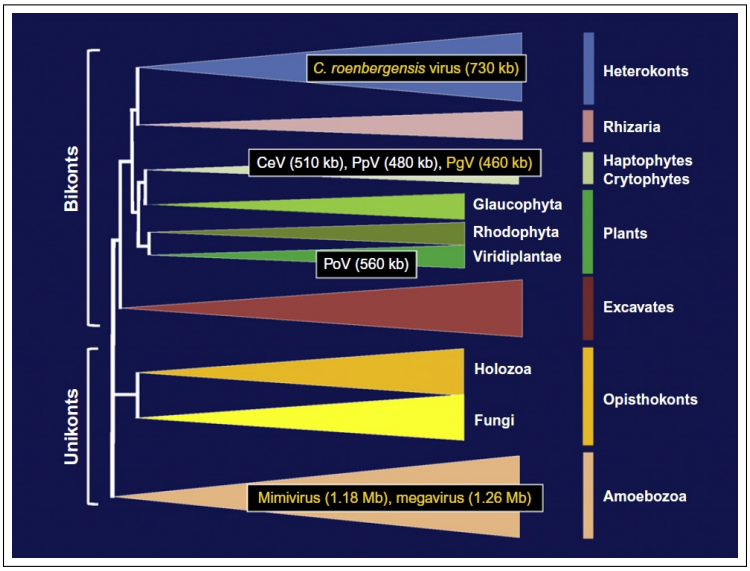


FIGURE 1.50 – Arbre phylogénétique de la gamme des hôtes eucaryotes des grands virus à ADN, avec les virus correspondants (CLAVERIE *et al.*, 2013).

Ces études permettent de supposer que tous les virus à ADN double-brin ont pour origine le même ancêtre commun, et deux hypothèses différentes sont formulées pour expliquer la grande variété d'hôtes infectés par ces virus.

La première suppose que tous ces virus (à ADN double-brin) étaient "petits" à l'origine, et que certains d'entre eux seraient devenus des grands virus à ADN (ou des virus géants) par l'acquisition de certains gènes provenant de leur hôte (transfert latéral de gènes). La deuxième hypothèse suppose que tous les virus à ADN étaient grands à l'origine, et que certains d'entre eux seraient devenus "petits" en perdant certains de leurs gènes.

Cette première hypothèse peut poser quelques problèmes, notamment dans le cas des virus géants, dont certaines protéines encodées par les génomes viraux présentent moins de 30% d'identité avec leurs homologues cellulaires : si ces gènes ont été acquis à partir d'un hôte ou d'un autre organisme cellulaire, pourquoi y a-t-il si peu de traces de ces organismes dans les bases de données ? Une explication peut être que les virus évoluent plus rapidement que les organismes cellulaires et que les séquences qui sont dérivées des cellules ne seraient plus phylogénétiquement détectables. Toutefois, beaucoup de protéines codées par ces fameux gènes montrent toujours une similarité de séquence significative à travers les différentes lignées de virus à ADN, qui ont divergé au cours du temps. De plus, les protéines des virus géants qui montrent une similarité significative avec leurs homologues cellulaires, sont tellement divergentes que leur origine apparaît à la racine de la branche des eucaryotes (ou même avant). Enfin, les transferts latéraux de gènes récents représentent moins de 10% des plus grands génomes viraux, et la plupart des gènes des virus géants n'ont pas d'homologues cellulaires.

La deuxième hypothèse, celle de la perte de gènes au cours du temps pour certains virus (ou évolution réductive), semble donc favorisée.

Ce processus de réduction génomique est irréversible : en effet, un organisme qui perd certaines de ses fonctions devient donc un parasite obligatoire, et ne peut pas redevenir autonome.

Cette perte de gènes ou d'une fonction donnée peut se produire de façon indépendante, à différents moments, dans différentes lignées. Par exemple, la perte de la machinerie de transcription est probablement survenue plusieurs fois au cours de l'évolution, et de façon indépendante (lignée-spécifique), à partir du génome à ADN (géant) de l'ancêtre commun des *Megaviridae*.

En résumé, ce modèle d'évolution réductive, utilisé par des parasites intracellulaires obligatoires (bactériens ou eucaryotes), semble être le plus parcimonieux. Ce modèle émet l'hypothèse que les NCLDV, qui sont tous des parasites intracellulaires obligatoires, seraient le résultat d'un processus de réduction du génome irréversible, et que les grands virus à ADN que l'on connaît aujourd'hui auraient hérité leurs génomes d'un micro-organisme possédant plus de gènes et de fonctions. La nature de ce micro-organisme ancestral reste encore un mystère à ce jour : cet ancêtre était probablement capable de synthétiser lui-même ses propres protéines, laissant imaginer qu'il était peut-être une cellule, et que son existence remonte possiblement avant la séparation des trois domaines du vivant (Claverie *et al.*, 2013).

D'un autre côté, Patrick Forterre décrit l'impact des récentes découvertes sur le débat concernant le statut vivant/non-vivant des virus, il propose une définition du "vivant" et présente le concept de "virocell".

Les virus utilisent les mêmes macromolécules (protéines et acides nucléiques) que les organismes cellulaires pour la reproduction et l'expression de leur information génétique, ce qui indique que les cellules et les virus s'insèrent dans le même processus que l'on appelle la "vie".

Le cycle réplicatif des virus possède deux phases caractéristiques : une phase extracellulaire, au cours de laquelle le génome viral est inactif dans la particule virale, aussi appelée virion, et une phase intracellulaire, appelée virus, dans laquelle le génome peut soit rester silencieux (sous forme de chromosome libre ou intégré dans le génome cellulaire), soit être activement exprimé et répliqué dans la cellule.

Les trois domaines cellulaires (eucaryotes, bactéries et archées) sont associés avec trois ensembles différents de familles de virus (eukaryovirus, bacteriophages et archaeovirus) : des analyses comparatives montrent que certaines protéines clés, encodées par des virus qui appartiennent aux différents ensembles, sont évolutivement apparentées, et permettent alors de séparer les virus à ADN double-brin en deux lignées.

La première lignée, appelée Adenovirus/PRD1 contient plusieurs familles de virus d'eucaryotes, dont les *Adenoviridae* et les NCLDV, le virus PRD1 infectant les bactéries (*Tectiviridae*) et les *Turriviridae*, qui infectent les archées.

La deuxième lignée, appelée HK97, regroupe les bactériophages, les virus d'archées et les *Herpesviridae* qui infectent les eucaryotes.

Tous les membres de la même lignée partagent des protéines majeures de capside et des "*packaging ATPases*". En revanche, ces deux protéines caractéristiques d'une lignée ne sont pas homologues à celles de l'autre lignée, confirmant alors que les virus constituent un groupe polyphylétique : c'est à dire que les virions et les mécanismes employés pour leur formation et pour leur dissémination, ne sont pas hérités d'un ancêtre commun, mais seraient "inventés" plusieurs fois indépendamment.

Ces traits homologues chez les virus infectant les différents domaines du vivant ne seraient toutefois pas issus d'une évolution convergente selon l'auteur. En effet, d'une part, les similarités observées entre les structures des protéines et les séquences des virus d'une même lignée, mais qui infectent des hôtes des différents domaines, sont trop grandes pour pouvoir être expliquées par une convergence. D'autre part, les repliements de protéines complètement différents, peuvent aboutir à des structures de capside similaires. Il en déduit que les repliements similaires des protéines majeures de capsides de virus infectant les différents domaines du vivant, fournissent une preuve du lien évolutif entre eux, plutôt qu'une convergence pour produire une structure particulière (capside icosaédrique).

L'auteur propose alors le concept de "virocell" pour expliquer l'évolution des virus. Pour cela, il se concentre sur le stade actif du cycle de reproduction des virus, c'est à dire, le moment où de nouvelles protéines virales pourraient apparaître durant le processus de réplication et/ou de recombinaison. Il suggère que cela pourrait expliquer le grand nombre de gènes présents dans les génomes viraux, et qui n'ont pas d'homologues cellulaires.

Une virocell serait un nouveau type d'entité cellulaire produit par transformation progressive d'une cellule par l'information génétique virale, au cours du processus d'infection. Ce phénomène est marqué par l'émergence d'un métabolisme complètement différent de la cellule, avec une fonction différente, qui fournit une autonomie au virus, supportant la définition du virus comme un organisme qui produit des virions. Ce concept suggère que le virus est aussi un organisme cellulaire au cours de l'étape virocell de son "cycle de vie". En résumé, une virocell est donc une cellule infectée, qui ne se divise plus et qui produit des virions, et à l'inverse, une "ribocell" est une cellule qui encode des ribosomes et qui se divise.

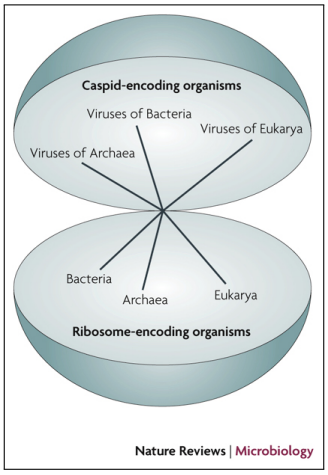


FIGURE 1.51 – Redéfinition des virus où les virus possèdent une capside mais pas de ribosomes et à l'inverse, où les organismes cellulaires possèdent des ribosomes mais pas de capside (RAOULT *et al.*, 2008).

Dans ce concept de virocell, il existe également une troisième forme de cellule, la "ribovirocell" : il s'agit d'une cellule infectée qui produit des virions mais qui est toujours capable de se diviser.

En revanche, ce concept de virocell n'est pas applicable pour les organismes intracellulaires qui codent eux-mêmes pour des ribosomes : par exemple, les bactéries qui "vivent" dans les cellules eucaryotes. Il est toutefois important de distinguer un virus d'une virocell : le terme virus décrirait alors un processus biologique et intégrerait tous les aspects du cycle de reproduction viral (le virion, la virocell et les formes génomiques intégrées).

L'auteur propose alors une définition du "vivant" : un organisme, une cellule, un virus, un plasmide ou une protéine peuvent être définis comme vivants parce que ces entités biologiques sont des "individus", c'est à dire qu'ils sont séparables, comptables et ont des limites spatiales bien définies.

Au contraire, les entités biologiques qui sont "particulières", ne peuvent pas être définies comme vivantes, comme un gène par exemple. La "vie" est donc le mode d'existence d'individus biologiques vivants, et toutes les entités biologiques sont vivantes, à partir du moment où elles sont impliquées activement dans un processus vivant (FORTERRE, 2016).

Avec la découverte d'autres virus géants comme les Pandoravirus, les Pithovirus ou encore *Mollivirus sibericum*, le scénario d'évolution par la réduction du génome de certains virus, s'étoffe un peu plus. Une gradation dans la notion de parasitisme absolu est décrite et divisée en quatre niveaux, des virus les plus "autonomes" aux virus les plus dépendants de leur hôte :

— Le premier niveau regroupe les virus dont le cycle de réplication est entièrement cytoplasmique, qui codent pour leur propre machinerie de transcription et qui

peuvent initier leur cycle de réplication sans l'aide du noyau de la cellule hôte (comme les *Mimiviridae* et les *Pithoviridae*) ;

— Le deuxième niveau correspond aux virus qui codent pour leur propre machinerie de transcription, mais ne l'embarquent pas dans leurs particules et dont l'initiation du cycle infectieux se fait à l'aide du noyau de l'hôte (comme les *Pandoraviridae* ou *Mollivirus sibericum*) ;

— Le troisième niveau comprend des virus qui ne codent pas pour leur propre machinerie de transcription (ou qui codent pour une machinerie incomplète), et sont obligés de passer par un stade intra-nucléaire complet. Beaucoup de ces virus codent d'ailleurs toujours pour leur propre machinerie de réplication de l'ADN, alors que celle de l'hôte est disponible dans le noyau où ils sont en train de faire leur transcription ;

— Le dernier niveau correspond aux virus qui ne codent pas pour leur machinerie de réplication d'ADN et pour qui le noyau de la cellule hôte est indispensable.

Cette hiérarchie dans le parasitisme absolu peut s'expliquer dans le cadre d'une évolution réductive, où les virus initialement complexes se retrouveraient piégés dans un mode de vie intracellulaire et perdraient continuellement leurs gènes et leurs fonctions, à des vitesses variables selon diverses contraintes écologiques et physiologiques. Dans ce contexte, les lignées qui évoluent le plus lentement donneraient les virus les plus complexes aujourd'hui, et à l'inverse, les virus les plus simples seraient issus de lignées qui ont évolué rapidement.

Ce scénario d'évolution réductive est cohérent avec la diversité en taille et en complexité des génomes des diverses familles de virus à ADN infectant les eucaryotes. En effet, ces familles pourraient provenir de voies alternatives dans cette évolution réductive, qui auraient été initiées par la perte de ribosomes fonctionnels, dont l'absence est partagée par tous les virus. Ensuite, des fonctions fondamentales ont pu être perdues dans certaines lignées de virus, comme la transcription ou la réplication d'ADN : ces pertes successives ont pu conduire dans un premier temps, à des virus cytoplasmiques hôte-dépendants, puis à des virus qui se répliquent dans le noyau de la cellule hôte.

Ce scénario ne semble toutefois pas compatible dans le cas des virus géants : les familles de ces virus sont très différentes les unes des autres, autant qu'elles le sont des organismes cellulaires.

Une nouvelle hypothèse est donc avancée : plutôt qu'il n'y ait qu'un seul ancêtre commun pour toutes les familles de virus à ADN, peut-être que ces diverses lignées de virus ont des origines différentes, et proviennent de plusieurs types d'ancêtres communs : les NCLDV constitueraient donc un groupe polyphylétique.

Comme évoqué précédemment, ces ancêtres étaient possiblement de nature cellulaire, et leur existence remonterait avant la séparation des trois domaines du vivant. Il semblerait que ces proto-cellules ancestrales étaient autrefois en compétition avec la

proto-cellule qui aurait par la suite donné LUCA, ou le dernier ancêtre commun universel des eucaryotes, des bactéries et des archées. A la fin de cette compétition, les proto-cellules "perdantes" auraient alors été obligées de parasiter les lignées cellulaires gagnantes afin de survivre, donnant alors lieu à la diversité de virus à ADN que l'on connaît aujourd'hui.

Les auteurs proposent également une nouvelle définition du mot "virus", dans laquelle il n'y a pas distinction entre les types de génomes, les structures des particules, leur taille ou leur morphologie, et qui serait moins ambiguë. Un virus serait donc toute entité biologique dont la molécule d'acide nucléique est répliquée par un système de macromolécules qu'il n'encode pas totalement (parasitisme absolu), qui est disséminée à l'aide d'une structure métaboliquement inerte et qui ne requiert pas d'énergie pour être maintenue (CLAVERIE *et al.*, 2016b).

1.5 Objectifs de cette thèse

La découverte, il y a un peu plus d'une dizaine d'années, d'un nouveau genre de virus, les virus géants, a révolutionné le monde de la biologie et plus particulièrement le domaine de la virologie. Mimivirus, le premier virus géant et fondateur de la famille des *Mimiviridae* a flouté la limite entre le monde cellulaire et le monde viral. Les découvertes successives de nouveaux virus géants au cours de ces dernières années, ont bousculé les définitions traditionnelles du mot "virus" et de ce qu'est le "vivant", et ont soulevé de nombreuses questions sur l'origine et l'évolution des virus. Différentes hypothèses ont été proposées dans le but d'améliorer notre compréhension sur l'origine de ces virus, ainsi que sur leur évolution, et le débat animé à propos de leur nature (vivant/non-vivant) est plus que jamais d'actualité.

La découverte et l'identification de nouveaux virus géants apportent des éléments supplémentaires pour la compréhension du rôle de ces virus au niveau écologique, par exemple leur rôle dans la régulation des populations d'algues (efflorescences), dans l'évolution du domaine viral, et des domaines du vivant. Paradoxalement, la découverte de nouveaux virus, soulève de nombreuses questions : avec plus des deux tiers de leurs gènes qui ne ressemblent à rien de connu, quelle est leur origine ?

Le but de ce travail de thèse était d'isoler et de caractériser de nouveaux virus géants, à partir d'échantillons environnementaux prélevés sur différents continents et mis en co-culture avec deux types d'amibes en parallèle : *Acanthamoeba castellanii* et *Dictyostelium discoideum*.

Pour y parvenir, j'ai réparti mon travail de recherche en plusieurs axes :

— l'isolement de nouveaux virus à partir d'échantillons environnementaux, co-cultivés en parallèle sur les amibes *A. castellanii* et *D. discoideum* ;

— la caractérisation complète de ces nouveaux virus à l'aide de différentes techniques, comme l'étude de leurs cycles infectieux, le séquençage de leurs génomes, des analyses du protéome des virions et des cellules infectées, ainsi que l'étude du transcriptome des cellules infectées au cours du temps ;
— l'étude de la composition des téguments de ces virus géants, afin de comprendre leur nature ;
— l'étude comparative entre les différents membres d'une même famille, comme les *Pandoraviridae* et les *Marseilleviridae*.

Ce mémoire de thèse est composé de trois chapitres, précédés d'une partie Matériel et méthodes décrivant les différentes méthodes utilisées :

i. "Caractérisation d'un nouveau *Marseilleviridae* et études comparatives", décrit le nouveau membre de la famille des *Marseilleviridae* que j'ai isolé, sa caractérisation et l'étude comparative de ce virus avec un autre membre de cette famille, également isolé au laboratoire (cf. Chapitre 3) ;
ii. "Caractérisation de nouveaux *Pandoraviridae* et études comparatives", décrit le nouveau Pandoravirus que j'ai isolé, sa caractérisation et l'étude comparative des différents Pandoravirus du laboratoire grâce à différentes approches (génomique, transcriptomique, protéomique, et comparaison de leurs cycles infectieux au microscope électronique à transmission, cf. Chapitre 4) ;
iii. "Études de la composition des *Pandoraviridae* et *Mollivirus sibericum*", décrit l'étude initiée récemment sur la composition des structures lamellaires retrouvées dans les téguments de ces virus (cf. Chapitre 5).

2 Matériel et méthodes

2.1 Isolement des virus

2.1.1 Culture des amibes

Les amibes du genre *Acanthamoeba* sont utilisées comme organisme modèle au laboratoire, et plus particulièrement l'amibe *Acanthamoeba castellanii*. Les amibes *A. castellanii* (Douglas, souche Neff ATCC 30010™) ont été cultivées dans un milieu axénique appelé "PPYG" ou Proteose-Peptone-Yeast extract-Glucose dont la préparation s'effectue en deux temps : le milieu basal est composé de 15 g de proteose-peptone et d'1 g d'extrait de levure dans 900 mL d'eau ultra-pure. Une fois ce mélange stérilisé par autoclave, sont ajoutés 4 mM de $MgSO_4$, 0,4 mM de $CaCl_2$, 0,05 mM de $Fe(NH_4)_2(SO_4)_2$, 2,5 mM de Na_2HPO_4, 2,5 mM de KH_2PO_4, et 100 mM de glucose. Le volume est complété à 1 L avec de l'eau ultra-pure, le pH est ajusté à 6,5 et le mélange est filtré sur 0,22 μm.

Récemment, des amibes du genre *Dictyostelium discoideum* ont été introduites au laboratoire, afin d'isoler des virus capables d'infecter ces amibes. Les amibes *D. discoideum* (souches DH1 et HMX44A, fournies aimablement par Pierre Golstein, CIML) ont été cultivées dans un milieu axénique riche appelé HL5, qui est composé de 14,3 g/L de bactopeptone, 7,15 g/L d'extrait de levure, 9 g/L de maltose, 0,93 g/L de Na_2HPO_4 et 0,49 g/L de KH_2PO_4 dans de l'eau ultra-pure. Le milieu est ensuite stérilisé par autoclave avant utilisation.

Les cultures d'amibes ont été maintenues en milieu axénique, en mono-couches adhérentes dans des flasques horizontales de taille variable (25 cm^2, 75 cm^2 ou 175 cm^2), à l'obscurité et à une température de 32°C (*A. castellanii*) ou 22°C (*D. discoideum*). Ces cellules ont été régulièrement réensemencées, tous les deux jours environ : ces "passages" consistent à gratter les amibes adhérentes à confluence [1], à l'aide d'un râteau stérile et à les compter sur une cellule de Malassez.

Ces cellules ont ensuite été ensemencées en faible quantité, entre 1 000 et 5 000 cellules par cm^2, dans du milieu de culture frais, et placées à 32°C ou 22°C pour qu'elles se multiplient à nouveau jusqu'à confluence. Ces passages ont été effectués sous un poste de sécurité microbiologique de type II, et les cellules ont été maintenues jusqu'à 35 passages environ : au-delà, une souche de faible passage, stockée à -80°C, est décongelée et réensemencée dans du milieu frais.

1. Les cellules sont à confluence lorsqu'elles forment une mono-couche qui recouvre toute la surface de la flasque dans laquelle elles sont cultivées.

Les échantillons environnementaux étant riches en bactéries et fungi, les cellules utilisées pour les expériences de co-culture avec ces échantillons ont été préalablement adaptées à des concentrations croissantes en antifongique, comme l'amphotéricine B (ou Fungizone, Gibco) à 2,5 μg/mL.

2.1.2 Traitement des échantillons environnementaux

Les échantillons environnementaux (boue) qui ont été testés au cours de cette thèse ont été prélevés dans des étangs à Nouméa en Nouvelle-Calédonie, puis ils ont été prétraités avant d'être placés en co-culture avec les amibes.

Ce traitement consiste en premier lieu, à placer une partie de ces échantillons fraîchement prélevés en milieu riz à une concentration finale de 4% et de conserver ce mélange à l'obscurité pendant au moins trois semaines. Ce milieu riz, riche en amidon, va permettre aux organismes hétérotrophes (hôtes potentiels des virus géants) de se multiplier, augmentant alors la probabilité d'isoler des virus. Le milieu riz a été préparé en ajoutant 40 grains de riz à un litre d'eau ultra-pure, puis a été stérilisé par autoclave, et stocké à 4°C.

Après trois semaines d'incubation, une partie des échantillons a été diluée en PBS 1X (ou "*Phosphate Buffered Saline*", 8 g/L de NaCl, 0,4 g/L de KCl, 1,44 g/L de Na_2HPO_4 et 0,24 g/L de KH_2PO_4). Une combinaison d'antibiotiques et un antifongique ont été ajoutés aux échantillons : ampicilline à 100 μg/mL, kanamycine à 100 μg/mL, chloramphénicol à 100 μg/mL, tétracycline à 100 μg/mL, pénicilline-streptomycine (Gibco) à 100 μg/mL et Fungizone (Gibco) à 25 μg/mL. Après deux jours d'incubation à 4°C sur un agitateur rotatif, le mélange a été centrifugé à 800 x *g* pendant 5 minutes : le surnageant a été mis de côté et le culot a été resuspendu dans un petit volume de PBS.

2.1.3 Co-cultures

Des amibes (*A. castellanii* et *D. discoideum*) adaptées à 2,5 μg/mL de Fungizone ont été ensemencées en plaques multipuits à 10 000 cellules/cm^2 avec le milieu approprié pour chaque souche (PPYG ou HL5), de la Fungizone à 2,5 μg/mL final et des antibiotiques à 100 μg/mL. Les plaques multipuits ont été incubées environ une heure à 32°C ou à 22°C pour que les cellules puissent adhérer dans le fond des puits. Les échantillons ont ensuite été inoculés dans les cellules : 100 μL de surnageant et 30 μL de culot ont été ajoutés dans chaque puits de la plaque (en duplicat), en laissant des puits témoins (cellules non-infectées), puis les plaques ont été incubées à la température adéquate pour chaque espèce.

Les plaques ont été suivies régulièrement au microscope optique, et des "repiquages" ont eu lieu tous les deux jours. Ces repiquages consistent à récupérer le surnageant d'infection de chaque puits et à le placer en co-culture avec des amibes et du milieu

frais, de la Fungizone et une nouvelle combinaison d'antibiotiques, qui est changée fréquemment.

Lorsqu'un phénomène de lyse cellulaire a pu être observé, l'infection a été amplifiée sur une plus grande quantité de cellules, et une partie de ces cellules a été fixée dans un mélange de PBS et de Glutaraldéhyde (EMS) à 2,5% final, afin d'être observée entre lame et lamelle au microscope optique (Zeiss Axio Observer, objectif x63 avec Optovar x1,6). Des flasques supplémentaires contenant des cellules à confluence ont ensuite été infectées, afin de produire, de purifier et de caractériser le virus isolé.

2.2 Purification des virus

2.2.1 Purification sur gradient de sucrose

Cinquante flasques contenant les amibes infectées ont été incubées deux jours à 32°C et régulièrement observées au microscope optique afin de contrôler le phénotype infectieux.

Après deux jours d'incubation, les flasques ont été grattées à l'aide d'un râteau stérile afin de décoller les cellules restantes, puis le contenu de ces flasques a été centrifugé à 500 x *g* entre 5 à 10 minutes pour culoter les débris et les cellules. Le surnageant contenant les virus a ensuite été centrifugé à 6 800 x *g* de 30 minutes à une heure. Le culot a ensuite été resuspendu en PBS puis centrifugé à 6 800 x *g* pendant 30 à 45 minutes, afin de se débarrasser de débris légers éventuels, de résidus du milieu de culture, ...Cette étape de lavage a été répétée une fois, puis le culot viral a été repris dans 5 mL de PBS avant d'être déposé sur un gradient discontinu de sucrose.

Ce gradient a été préparé en superposant des couches de sucrose à différentes concentrations, en commençant par la couche la plus concentrée. En fonction des virus, ces concentrations sont variables : elles peuvent aller de 20 à 50% (poids/volume) de sucrose (Noumeavirus), et de 50 à 80% de sucrose (Pandoravirus). Les 5 mL de suspension virale ont été déposés sur un gradient de 45 mL, qui sera centrifugé à 5 000 x *g* pendant 30 à 45 minutes. Après centrifugation, l'anneau viral, situé dans une des couches de sucrose a été récupéré, puis lavé deux à trois fois en PBS (centrifugation à 6 800 x *g* pendant 30 à 45 minutes).

Le culot contenant les particules a ensuite été resuspendu dans 5 à 10 mL de PBS, puis a été contrôlé entre lame et lamelle au microscope optique (Zeiss Axio Observer, objectif x63 avec Optovar x1,6). Le virus a ensuite été titré avant d'être stocké à 4°C en PBS et à -80°C en présence de DMSO 7,5% (diméthylsulfoxyde, un agent cryoprotecteur).

2.2.2 Titration et comptage des virus

La titration des virus peut être effectuée en deux parties : lorsque le virus est visible au microscope optique, il est d'abord "compté" sur cellule de Malassez, avant d'être titré par dilutions sériées ($TCID_{50}$).

Les virus observables au microscope optique (comme les Pandoravirus) peuvent être comptés : un mélange de 5 μL de virus dilué en PBS ($1/100^{e}$ ou $1/1000^{e}$), 35 μL de glutaraldéhyde 2%, et 10 μL de bleu trypan a été déposé sur une cellule de Malassez. Les virus marqués en bleu et présents sur trois carreaux de la cellule ont été comptés. Le nombre de virus par mL a pu être déduit par le calcul suivant : moyenne des virus par carreau x facteur de dilution x 10 (dilution du bleu trypan) x 100 (100 carreaux = 1 μL) x 1 000.

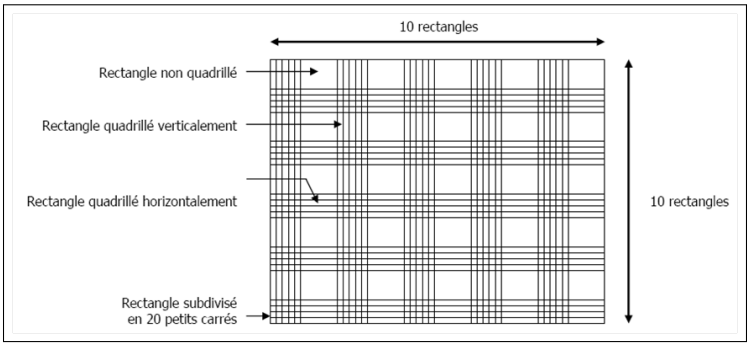


FIGURE 2.1 – Schéma d'une cellule de Malassez ou hémocytomètre, utilisée pour compter les cellules et les virus (©Amel Boumendjel-Messarah).

Les virus sont également titrés, en utilisant la méthode du $TCID_{50}$, ou "*50% Tissue Culture Infective Dose*" : cette méthode permet de déterminer la quantité de particules virales infectieuses nécessaires pour lyser ou infecter 50% des cellules hôtes. Pour cela, des amibes ont été ensemencées à 5 000 cellules par puits dans des plaques multipuits (96 puits) avec du milieu, puis incubées une heure à 32°C pour que les cellules adhèrent.

Parallèlement, des dilutions sériées de virus allant de 10^{-1} à 10^{-15}, ont été réalisées dans le milieu de culture utilisé pour les cellules. Les cellules ont été ensuite infectées avec les dilutions virales : 50 μL de chaque dilution ont été inoculés dans dix puits.

Les plaques ont été observées tous les jours au microscope optique, afin de contrôler l'avancée de l'infection et la lyse éventuelle des cellules. Après 5 à 7 jours d'incubation, les puits où les cellules avaient un phénotype infectieux ainsi que les puits où les cellules étaient lysées ont été notés pour chaque dilution, et ces résultats ont été utilisés pour calculer le $TCID_{50}$. Le $TCID_{50}$ est calculé selon la méthode de Reed-Muench (REED *et al.*, 1938) en utilisant une feuille de calcul adaptée par Brett Lindenbach (LINDENBACH, 2009) et permet d'estimer la concentration de particules virales infectieuses dans le milieu.

2.2.3 Clonage des virus sur cellule unique

Des amibes *A. castellanii* ont été ensemencées à 70 000 cellules/cm^2 dans une plaque multipuits, puis infectées par les virus à une MOI ("*Multiplicity of Infection*") de 50 [2]. La plaque a ensuite été incubée à 32°C pendant une heure.

Après incubation, le surnageant a été retiré du puits et remplacé par du milieu frais. Ce lavage sera répété 50 fois, afin d'éliminer tous les virus qui seraient encore présents dans le milieu et qui n'auraient pas infecté les cellules.

Les cellules adhérentes de ce puits ont ensuite été délicatement décollées, puis diluées en cascade avec du milieu frais dans trois autres puits. Les cellules de la dernière dilution ont été prélevées et des gouttes de 0,5 μL ont été déposées dans les autres puits de la plaque. Ces puits ont ensuite été observés au microscope optique, afin de vérifier la présence d'une seule cellule par goutte : les puits contenant une cellule unique sont marqués, et 500 cellules fraîches y ont été ajoutées. Les puits où les gouttes ne contiennent pas de cellules servent de témoin : 500 cellules fraîches y ont également été ajoutées. La plaque a ensuite été incubée à 32°C et observée tous les jours au microscope optique, afin de contrôler l'état des cellules et des témoins.

Dans le cas où les puits témoins montrent un phénotype infectieux, cela signifie que les lavages réalisés au début pour éliminer les virus en suspension dans le milieu n'étaient pas suffisants : il faut donc recommencer l'opération, en réalisant des lavages supplémentaires.

Les puits contenant une cellule unique au départ, où l'on observe une lyse des cellules ou un phénotype infectieux ont été récupérés, l'infection a été amplifiée dans un plus grand volume et les virus purifiés comme décrit précédemment.

2.3 Cycles infectieux

Des flasques de 25cm^2 contenant des cellules *A. castellanii*, ensemencées à 60 000 cellules/cm^2, ont été infectées par les virus à une MOI de 50, puis incubées à 32°C. Certaines flasques n'ont pas été infectées, afin de servir de témoin.

Après 30 minutes d'infection, les flasques ont été "lavées" deux fois avec du milieu frais, afin d'éliminer les virus en suspension qui n'auraient pas infecté de cellules. Une flasque a ensuite été observée au microscope optique après ces lavages, puis fixée : un volume équivalent de glutaraldéhyde à une concentration de 5% en PBS a été ajouté dans la flasque afin d'atteindre une concentration finale de 2,5%. Cette flasque a été incubée à température ambiante entre 20 minutes et 1 heure, puis les cellules ont été décollées avec un râteau stérile et centrifugées à 5 000 x *g* pendant 20 minutes. Enfin, le culot a été resuspendu dans 1 mL de glutaraldéhyde à une concentration de 2,5% et stocké à 4°C.

2. Une MOI de 50 représente 50 particules virales pour une cellule.

Cette opération a été répétée pour toutes les autres flasques à différents temps (1 heure post-infection (pi), 2 heures pi, ...) jusqu'à la fin du cycle, ainsi que pour les témoins (cellules non-infectées).

Les cellules fixées seront ensuite incluses en résine, en vue de leur observation au microscope électronique.

2.4 Fluorescence des cellules infectées par Noumeavirus

Deux méthyltransférases de *Mollivirus sibericum* qui contiennent un signal de localisation nucléaire (Ml135) ou nucléolaire (Ml216), ont été clonées dans le plasmide d'expression amibien pGAPDH-GFP, afin de produire des protéines marquées à la GFP en C-terminal (BATEMAN, 2010), qui ciblent le noyau ou le nucléole des cellules. Une construction identique a également été réalisée avec la protéine nucléaire SUMO d'*A. castellanii* [3].

Les cellules *A. castellanii* ont été transfectées avec ces vecteurs, en utilisant un protocole de précipitation au phosphate de calcium, habituellement utilisé sur les amibes *D. discoideum* (GAUDET *et al.*, 2007). Les cellules transformées ont été sélectionnées dans un premier temps avec 20 μg/mL de néomycine (G418) : cette sélection a ensuite été progressivement augmentée jusqu'à 60 μg/mL en l'espace de quelques semaines.

Les amibes *A. castellanii* transfectées ont été ensemencées en plaque 12 puits, et infectées par Noumeavirus à une MOI de 50, sauf un puits correspondant au témoin (cellules non-infectées). Après 30 minutes, 1 heure, 2h et 4 hpi, les cellules ont été décollées des puits à l'aide d'un râteau stérile, et centrifugées à 500 x *g* pendant 5 minutes. Les culots ont ensuite été repris et fixés avec 3,7% de formaldéhyde dilué en PBS, pendant 30 minutes à température ambiante.

Ces cellules fixées ont ensuite été lavées en PBS, puis 4 μL ont été mélangés avec 4 μL de milieu de montage contenant du DAPI [4] (Vectashield-DAPI). La fluorescence a été observée sur un microscope à épifluorescence (Zeiss Axio Observer), en utilisant un objectif x 63 associé à un Optovar x 1,6.

3. La protéine SUMO ou "*Small Ubiquitin-like Modifier*", est une protéine retrouvée chez tous les eucaryotes, qui est impliquée dans un mécanisme de modification post-traductionnelle (SUMOylation). Cette protéine est également impliquée dans des activités au sein du noyau de la cellule.

4. Le DAPI ou 4',6-diamidino-2-phénylindole est une molécule qui se lie aux bases A et T de l'ADN et qui émet une fluorescence bleue (450-490 nm) lorsqu'elle absorbe la lumière UV. Cela permet donc de détecter l'ADN dans les cellules (comme le noyau et les mitochondries).

2.5 Microscopie électronique

2.5.1 Inclusions en résine

Les échantillons ont été centrifugés à 5 000 x *g* pendant 5 minutes, puis les culots ont été repris dans du PBS et de nouveau centrifugés à 5 000 x *g* pendant 5 minutes. Les culots ont été maintenus à 42°C dans un bain-marie. Parallèlement, des Microvettes dont le tube externe est rempli d'eau, sont maintenues à 42°C (Sarstedt, cf. Figure 2.2).

FIGURE 2.2 – Tube Microvette dans lequel l'échantillon en agarose est déposé puis centrifugé pour les inclusions (©Sarstedt).

Les échantillons ont ensuite été repris dans 250 μL d'agarose low-melting (dissout à 2% en PBS) en surfusion, puis déposés dans les Microvettes, et centrifugés à 10 000 x *g* pendant 3 minutes. Les Microvettes ont été incubées dans de la glace pour solidifier les échantillons, qui sont enrobés dans l'agarose à l'extrémité des tubes. Après 20 minutes d'incubation, les Microvettes ont été découpées afin d'extraire les échantillons, qui ont été incubés en PBS 3 x 5 minutes à température ambiante.

Chaque échantillon a ensuite été incubé dans 1% de Tétroxyde d'osmium (OsO_4) en PBS pendant deux heures à température ambiante et à l'abri de la lumière, puis lavé 4 x 10 minutes en eau ultra-pure. Enfin, les échantillons ont été stockés en eau toute la nuit à 4°C.

Le lendemain, les échantillons ont été incubés dans 1% d'acétate d'uranyle en eau pendant deux heures à température ambiante et à l'abri de la lumière. Ils ont ensuite été déshydratés progressivement, dans des concentrations croissantes d'acétone, à température ambiante : 15 minutes en acétone 50%, puis 70%, 90%, et enfin 2 x 15 minutes en acétone 100% (sec).

Parallèlement, de la résine Epon a été préparée en mélangeant 22 mL de EMbed 812 (EMS) avec 7,5 mL de DDSA (Dodecenyl Succinic Anhydride, EMS) et 15 mL de NMA (Methyl-5-Norbornene-2,3-Dicarboxylic Anhydride, EMS) : ce mélange a été homogénéisé par agitation douce, pour éviter la formation de bulles. Lorsque ce mélange est homogène, 0,7 mL du durcisseur DMP-30 (2,4,6-Tri(dimethylaminomethyl) phenol, EMS) ont été ajoutés, et ce mélange a été homogénéisé encore une fois.

Les échantillons ont ensuite été incubés dans des mélanges de résine et d'acétone 100% (sec). Dans un premier temps, il ont été incubés dans un mélange composé d'1/4 de résine et de 3/4 d'acétone pendant 1 heure à température ambiante, puis dans un mélange de résine et d'acétone à parts égales pendant 3 heures à température ambiante, et enfin, dans un mélange de 3/4 de résine et d'1/4 d'acétone toute la nuit à température ambiante.

Le lendemain, les échantillons ont été incubés dans de la résine pure pendant 4 heures, puis la résine a été renouvelée pour 3 heures d'incubation supplémentaires. Enfin, ils ont été placés dans un moule en silicone (EMS) rempli de résine, puis incubés à 60°C pendant 5 jours pour que la résine polymérise.

Après incubation, chaque échantillon qui est contenu dans un bloc de résine dure, a été coupé à l'aide d'un ultra-microtome (Leica) et d'un couteau de diamant. Les coupes ultra-fines (70 nm d'épaisseur) obtenues, ont été déposées sur une grille de cuivre recouverte d'un film de formvar [5]. Cette grille a ensuite été contrastée avec du citrate de plomb (méthode de Reynold's) pendant 2 minutes à température ambiante, puis rincée 5 fois en eau ultra-pure, avant d'être observée au microscope électronique à transmission (MET, Tecnai G2, 200 kV).

2.5.2 Coloration négative

Dix microlitres d'échantillon fixé en glutaraldéhyde ont été déposés sur une grille de cuivre recouverte d'un film de formvar, puis incubés pendant 2 minutes à température ambiante. La grille a ensuite été rincée avec de l'acétate d'uranyle en eau, puis incubée 1 minute à température ambiante avec une goutte d'acétate d'uranyle en eau à une concentration de 2%. Enfin, la grille a été séchée sur papier filtre avant d'être observée au MET (Tecnai G2, 200 kV).

2.6 Génomique

2.6.1 Extraction génomique

Deux protocoles ont été utilisés pour extraire l'ADN des virus : l'ADN de Noumeavirus a été extrait à partir de 2 x 10^9 particules, à l'aide du kit PureLink™ Genomic DNA mini kit (Invitrogen), en suivant le protocole spécifique aux bactéries Gram négatives.

En revanche, l'ADN de *Pandoravirus neocaledonia* a été extrait en suivant un protocole utilisant du CTAB (Cetyltrimethylammonium bromide). Une quantité de 2 x 10^{10} particules virales a été centrifugée à 16 000 x *g* pendant 3 minutes et le culot a été repris dans 300 μL d'eau ultra-pure. Cinq cent microlitres de tampon contenant du CTAB

5. Le formvar ou "*polyvinyl formal*" forme un film fin qui recouvre la grille de cuivre et qui est transparent : il permet de maintenir les coupes ultra-fines en place sur la grille.

(20 g/L CTAB, 1,4 M NaCl, 100 mM Tris-HCl, 20 mM Na_2EDTA, pH 8,0), 20 mg/mL de Protéinase K et 3 μL de DTT 1 M ont été ajoutés, et l'échantillon a été incubé à 65°C pendant 1 heure 30. Puis, 20 mg/mL de RNase A ont été rajoutés, et l'échantillon a été incubé à 65°C pendant 10 minutes.

Après incubation, 500 μL de chloroforme ont été ajoutés, l'échantillon a été vortexé pendant 30 secondes, puis centrifugé à 16 000 x *g* pendant 10 minutes à 4°C. La phase aqueuse située dans la partie supérieure du tube a été prélevée et centrifugée à 16 000 x *g* pendant 5 minutes à 4°C. Le surnageant a ensuite été mélangé avec un volume équivalent de chloroforme, vortexé et centrifugé à 16 000 x *g* pendant 5 minutes à 4°C. La phase aqueuse a de nouveau été récupérée et 2 volumes d'un tampon de précipitation (5 g/L CTAB, 40 mM NaCl, pH 8,0) ont été ajoutés dans le tube, qui a été inversé délicatement plusieurs fois, jusqu'à obtenir un précipité.

L'échantillon a été incubé 1 heure à température ambiante avant d'être centrifugé à 16 000 x *g* pendant 5 minutes. Le culot obtenu a ensuite été délicatement dissout dans 350 μL de NaCl à une concentration de 1,2 M, puis 350 μL de chloroforme y ont été ajoutés. L'échantillon a été vortexé très brièvement avant centrifugation à 16 000 x *g* pendant 10 minutes à 4°C. La phase aqueuse a été récupérée, et 0,6 volumes d'isopropanol ont été ajoutés. Le tube a été inversé plusieurs fois jusqu'à ce que l'ADNg précipite, avant d'être centrifugé à 16 000 x *g* pendant 10 minutes à température ambiante.

Le culot a ensuite été délicatement lavé avec 500 μL d'éthanol à 70% en retournant plusieurs fois le tube, qui a été centrifugé à 16 000 x *g* pendant 10 minutes à température ambiante. Enfin, le culot d'ADN a été séché à l'air libre, avant d'être resuspendu dans 20 μL d'eau ultra-pure.

2.6.2 Séquençage des génomes

Deux microgrammes d'ADN purifié de Noumeavirus et 3 μg d'ADN de *Pandoravirus neocaledonia* ont été séquencés par GATC Biotech AG (Allemagne), en utilisant la technique de séquençage à haut-débit Illumina HiSeq 2000, générant des banques de séquences "*paired-end*" [6] de 2 x 100 pb.

De plus, 10 μg d'ADN purifié de *P. neocaledonia* ont été séquencés par GATC Biotech AG (Allemagne), en utilisant la technique de séquençage à haut-débit PacBio, qui permet d'obtenir des séquences de plus de 4 000 paires de bases.

Enfin, 5 μg d'ADN purifié de *P. quercus* et *P. celtis* ont été séquencés par l'Université de Lausanne, en utilisant également la technique PacBio.

6. Les séquences "*paired-end*" sont obtenues par le séquençage des deux extrémités de fragments courts.

2.6.3 Assemblage et annotation des génomes

Pandoravirus

Le séquençage du génome de *P. neocaledonia* a généré 38 500 séquences de 5 500 nucléotides en moyenne. Ces séquences ont ensuite été assemblées en utilisant le logiciel HGAP (ENGLISH *et al.*, 2012), ce qui a donné 3 contigs [7] de 2 Mb, 13 kb et 4 kb. Ces contigs ont été assemblés à l'aide de Minimus-2 (SOMMER *et al.*, 2007), ce qui a généré un seul contig de 2 003 191 nucléotides, avec un contenu en G+C de 61%. Le séquençage du génome de *P. quercus* a généré 72 468 séquences d'une longueur moyenne de 14 612 nucléotides. L'assemblage a également été effectué avec HGAP, ce qui a produit un seul contig de 2 077 388 nucléotides, avec un contenu en G+C de 61%. Enfin, le séquençage du génome de *P. celtis* a généré 71 879 séquences de 13 925 nucléotides de long en moyenne. L'assemblage avec les mêmes outils a produit 17 contigs de plus de 20 kb, et un long contig de 1 337 994 nucléotides. Toutefois, ce grand contig présentait des chutes de couverture, qui sont caractéristiques des problèmes d'assemblage, que nous n'avons pas réussi à résoudre.

Les outils qui ont été utilisés pour l'annotation des gènes ont été précédemment décrits chez PHILIPPE *et al.*, 2013.

Noumeavirus

Le séquençage du génome de Noumeavirus a généré des séquences de 101 nucléotides de long. Le logiciel idba_ud v1.1.1 (PENG *et al.*, 2012) a été utilisé pour assembler 99,15% des données en 18 contigs, dont un de 376 295 nucléotides correspondant au contig le plus long. Les 88 nucléotides aux extrémités 5' et 3' de ce contig étaient identiques et auraient pu être échangés, ce qui suggère un génome circulaire. La partie redondante a été retirée, donnant alors un contig de 376 207 nucléotides. Le logiciel bowtie 0.12.7 (LANGMEAD *et al.*, 2009) a été utilisé pour aligner toutes les paires à l'assemblage, donnant un taux d'erreur de 3%.

Afin de faciliter l'identification des gènes orthologues, des alignements de génomes entiers ont été réalisés par paires, entre Noumeavirus et ses plus proches parents, en utilisant LAST (KIEŁBASA *et al.*, 2011). Des palindromes (ou des "*hairpins*" potentiels) ont été recherchés dans la séquence à l'aide de rnamotif (OKONECHNIKOV *et al.*, 2012).

Le logiciel Genemark 4.29 (BESEMER *et al.*, 2001) a été utilisé pour prédire les gènes codant pour des protéines à partir du génome de Noumeavirus, et l'annotation fonctionnelle des 452 protéines prédites a été réalisée en combinant différentes méthodes. Une recherche Blast (NCBI RESOURCE COORDINATORS, 2016) a été réalisée contre la

7. Un contig est une séquence obtenue par chevauchement de séquences courtes (ou "*reads*") qui ont été générées par le séquenceur.

base de données de protéines NR (non-redondante) au NCBI, et la recherche de domaines a été effectuée contre différentes bases de données comme Pfam 28.0 (FINN *et al.*, 2016) à l'aide d'interproscan 5.14-53 (JONES *et al.*, 2014). Une recherche de domaines conservés ("CD-search", MARCHLER-BAUER *et al.*, 2015) a été réalisée contre la base de données du NCBI, et les protéines potentiellement transmembranaires ont été prédites à l'aide de Phobius (KÄLL *et al.*, 2004). Les peptides signaux nucléaires ont été prédits avec différents logiciels comme NucPred (BRAMEIER *et al.*, 2007) ou NOD (SCOTT *et al.*, 2011). Les résultats ont été compilés manuellement, et les fonctions les plus représentatives ont été assignées pour chaque séquence. La présence d'introns a été prédite en utilisant le logiciel Exonerate (SLATER *et al.*, 2005), et a été validée par RT-PCR.

2.6.4 Phylogénie

Pandoravirus

La phylogénie des Pandoravirus a été déterminée par la concaténation des régions codantes (nucléotidiques) orthologues entre les différents virus. Les protéines des différents virus ont été "clusterisées" (ou regroupées) par MCL (L. LI *et al.*, 2003), puis les groupes présentant strictement un gène par virus ont été extraits. L'alignement multiple de ces séquences a été produit à l'aide du programme mcoffee (WALLACE *et al.*, 2006), et les séquences nucléotidiques codantes ont été alignées à l'aide de cet alignement peptidique. La concaténation de ces séquences a permis d'obtenir un "super-alignement" des gènes orthologues des Pandoravirus, qui a servi à calculer un arbre phylogénétique avec le logiciel PhyML (GUINDON *et al.*, 2010) et un modèle de substitution JC69. L'enracinement de l'arbre a été effectué à l'aide des séquences de gènes orthologues présents chez *Mollivirus sibericum*, qui est le NCLDV le plus proche des Pandoravirus.

Noumeavirus

La phylogénie de Noumeavirus a été déterminée comme décrit précédemment (cf. paragraphe précédent), en utilisant l'Iridovirus Invertebrate iridescent virus 3, qui est proche des *Marseilleviridae*, comme groupe externe pour enraciner l'arbre.

2.7 Protéomique

2.7.1 Extraction de protéines

L'extraction de protéines a été réalisée sur 2 x 10^9 particules purifiées : les virus ont été centrifugés à 10 000 x *g* et le culot a été resuspendu en tampon Laemmli, composé

de 40 mM de Tris-HCl pH 7,5, de 2% de SDS et de 60 mM de DTT. Ce mélange a ensuite été chauffé à 95°C pendant 10 minutes.

2.7.2 Profils protéiques

Les lysats protéiques ont été dosés à l'aide du kit Pierce Coomassie 660 nm (Bradford, Thermo Fisher Scientific, selon les recommandations du fournisseur) : 75 μL de réactif ont été ajoutés à 5 μL de lysat, puis dosés au NanoDrop (Thermo Fisher Scientific). Cinq microgrammes de protéines ont été analysés par électrophorèse SDS-PAGE sur un gel d'acrylamide 12,5% avec un marqueur de poids moléculaire en parallèle ("unstained page ruler" Thermo Fisher Scientific). La migration a été effectuée à 180 V pendant 45 minutes, puis le gel a été coloré au bleu de Coomassie (Bio-Rad).

2.7.3 Protéome de surface des particules (Surfome)

Ce protocole a été utilisé pour analyser les protéines à la surface des particules virales purifiées après une digestion ménagée à la trypsine. Pour cela, deux fois 2 x 10^9 particules ont été traitées en parallèle : une partie a été utilisée comme témoin (non-traitée à la trypsine, tube "contrôle") et l'autre partie a été digérée à la trypsine. Les particules virales ont été centrifugées à 16 000 x *g* pendant 3 minutes, puis les culots ont été resuspendus dans 200 μL de tampon de digestion (50 mM Tris-HCl pH7,5, 150 mM NaCl et 5 mM $CaCl_2$) contenant 4,5 μg de trypsine modifiée (*sequencing grade*, Promega) pour le tube "surfome". Ces tubes ont été incubés à 37°C pendant 30 minutes, puis ont été centrifugés à 16 000 x *g* pendant 3 minutes. Les culots obtenus (les particules virales décapées) ont été fixés en glutaraldéhyde 2,5% pour être observés ultérieurement au MET.

Les surnageants ont été centrifugés encore une fois à 16 000 x *g* pendant 3 minutes, afin de culoter les éventuelles particules virales qui seraient restées en suspension, puis ces surnageants ont été récupérés. Un microgramme de trypsine a été ajouté dans chaque surnageant, et les tubes ont été incubés à 37°C toute la nuit. Le lendemain, de l'acide trifluoroacétique (TFA), utilisé comme agent d'appariement d'ion en chromatographie liquide, a été ajouté à une concentration finale de 0,1%, et les tubes ont été stockés à -80°C, avant d'être analysés.

2.7.4 Digestion des protéines extraites

Les protéines ont été concentrées à la surface d'un gel Nu-PAGE 4-12% (Invitrogen) avant d'être colorées au bleu de Coomassie R-250. Les bandes de gel ont été coupées et récupérées avant d'être lavées 6 fois 15 minutes dans 25 mM de NH_4HCO_3 et dans 25 mM de NH_4HCO_3 contenant 50% d'acétonitrile. Elles ont ensuite été déshydratées dans 100% d'acétonitrile, incubées à 53°C pendant 45 minutes en présence de 10 mM de

DTT dans 25 mM de NH_4HCO_3, puis incubées pendant 35 minutes à l'abri de la lumière avec 55 mM d'iodoacetamide dans 25 mM de NH_4HCO_3.

La réaction a été arrêtée en ajoutant 10 mM de DTT dans 25 mM de NH_4HCO_3, et en mélangeant le tout pendant 10 minutes. Les échantillons ont ensuite été lavés dans 25 mM de NH_4HCO_3 avant d'être déshydratés avec 100% d'acétonitrile. De la trypsine modifiée (*sequencing grade*, Promega) diluée dans 25 mM de NH_4HCO_3 a été ajoutée aux échantillons, qui ont été incubés toute la nuit à 37°C. Les peptides ont ensuite été extraits du gel en 3 étapes de 15 minutes dans 30 μL d'acétonitrile à 50%, 30 μL d'acide formique à 5%, et enfin, 30 μL d'acétonitrile à 100%. Le surnageant a ensuite été placé sous vide pour être séché.

2.7.5 Analyses Nano-LC-MS/MS

Les peptides extraits et séchés ont été resuspendus dans 5% d'acétonitrile et 0,1% de TFA, et analysés par nano-LC-MS/MS en ligne (Ultimate 3000, Dionex et LTQ-Orbitrap Velos pro, Thermo Fisher Scientific). Les peptides ont été échantillonnés sur une pré-colonne PepMap C18 de 300 μm x 5 mm, puis séparés sur une colonne C18 de 75 μm x 250 mm (PepMap, Dionex). La méthode nano-LC consiste en un gradient de 120 minutes (pour les analyses des particules virales et de surfome) ou 240 minutes (pour les analyses des cellules infectées), à un débit de 300 nL/min, allant de 5% à 37% d'acétonitrile dans 0,1% d'acide formique pendant 114 (ou 234) minutes, avant d'atteindre 72% d'acétonitrile dans 0,1% d'acide formique pour les 6 minutes restantes.

Les données de MS et de MS/MS ont ensuite été acquises à l'aide de Xcalibur (Thermo Fisher Scientific). La tension de pulvérisation a été réglée sur 1,4 kV et le capillaire chauffé à 200°C. Les données ont été acquises sur Orbitrap avec une résolution de 60 000, après avoir accumulé 106 ions, et les 20 ions les plus intenses ont été fragmentés par dissociation induite par collision dans le LQT (ou "*Linear Quadropole Trap*" est un piège à ions linéaire), après accumulation de 104 ions. Les peptides et les protéines ont été identifiés et quantifiés en utilisant le logiciel MaxQuant (version 1.5.3.30, Cox *et al.*, 2008).

2.7.6 Analyses des protéomes

Les spectres obtenus ont été recherchés dans les bases de données de séquences protéiques de Noumeavirus, Melbournevirus, *A. castellanii* et les contaminants les plus fréquemment observés. La longueur minimale des peptides a été fixée à 7 acides aminés, et les valeurs de quantification absolue basée sur l'intensité (iBAQ, Schwanhäusser *et al.*, 2011) ont été calculées à partir des intensités MS des peptides. Seules les protéines contenant un minimum de 2 peptides ont été considérées pour la suite des analyses.

Ces analyses ont été réalisées de la même façon pour les Pandoravirus.

2.8 Transcriptome des cellules au cours du cycle infectieux

Des cultures d'*A. castellanii* à confluence (environ 100 000 cellules/cm^2) en flasques de 175 cm^2 ont été infectées à une MOI de 100 et incubées à 32°C.

Après une heure d'incubation, les flasques ont été lavées deux fois en PPYG afin d'éliminer les virus non internalisés, puis incubées de nouveau à 32°C. Parallèlement, à 1 hpi, 4 flasques ont été observées au microscope optique et les cellules ont été grattées à l'aide d'un râteau stérile :

— 5 mL ont été fixés avec 5 mL de glutaraldéhyde à 5% en PBS pendant 1 heure à température ambiante, et centrifugés à 5 000 x *g* pendant 10 minutes. Le culot a ensuite été repris dans 1 mL de glutaraldéhyde à 2,5% en PBS, puis stocké à 4°C pour des observations au MET ;

— 2 mL ont été centrifugés à 16 000 x *g* pendant 3 minutes, puis le culot a été stocké à -80°C avant d'être analysé en protéomique (protéome cinétique des cellules infectées) ;

— 50 μL ont été prélevés, et les cellules ont été comptées sur cellule de Malassez ;

— 72 mL ont été centrifugés à 5 000 x *g* pendant 10 minutes, puis le culot a été resuspendu dans 4 mL de tampon de lyse RLT (kit RNeasy Midi, Qiagen) contenant 1% de β-mercaptoéthanol, et congelé à -80°C pour des extractions d'ARN.

Ce protocole a été répété toutes les heures jusqu'à 15 heures pi pour les flasques infectées, et également pour le témoin (cellules non-infectées).

2.8.1 Extraction des ARN

Les ARN totaux présents dans les cellules infectées (et le témoin) ont été extraits à l'aide du kit RNeasy Midi (Qiagen), en suivant le protocole indiqué pour les cellules animales. Trois élutions ont été réalisées : la première dans 500 μL et les deux autres dans 400 μL d'eau RNase-free. Ces ARN totaux ont été dosés au NanoDrop, puis stockés à -80°C, avant l'étape d'enrichissement en ARN messagers.

2.8.2 Enrichissement en ARN messagers

L'enrichissement en ARNm a été réalisé à l'aide du kit Dynabeads® mRNA Purification (Life Technologies), à partir de 100 μg d'ARN totaux, et l'élution a été faite dans 17 μL de 10 mM de Tris-HCl à pH 7,5. Les ARN messagers ont ensuite été dosés au NanoDrop, puis stockés à -80°C avant d'être envoyés à l'Université de Lausanne. Des librairies d'ADN complémentaires (ADNc) ont été produites à partir de ces ARNm, puis séquencées.

2.9 Études de la composition des parois des Pandoravirus et Mollivirus

2.9.1 Incubations en acide orthophosphorique

Des amibes *A. castellanii* sous forme de trophozoïtes et de kystes, ainsi que des virus (Pandoravirus, Mollivirus et Mimivirus) ont été incubés avec de l'acide orthophosphorique (H_3PO_4), un agent "gonflant" permettant notamment d'augmenter l'accessibilité de la cellulose aux cellulases.

Pour cela, des cellules *A. castellanii* ont préalablement été enkystées dans un milieu spécifique : après avoir été ensemencées dans une flasque de 175 cm^2 à confluence, le milieu PPYG a été retiré et remplacé par du milieu d'enkystement (20 mM Tris-HCl, 100 mM KCl, 8 mM $MgSO_4$, 0,4 mM $CaCl_2$, 1 mM $NaHCO_3$, pH 9). Après 2 semaines passées dans ce milieu d'enkystement, les kystes matures encore adhérents ont été décollés de la flasque à l'aide d'un râteau stérile, avant d'être centrifugés à 10 000 x *g* pendant 5 minutes, et lavés en PBS.

Une flasque de 175 cm^2 d'*A. castellanii*, sous forme de trophozoïtes, a été grattée puis les cellules ont été centrifugées à 10 000 x *g* pendant 5 minutes, avant d'être lavées en PBS.

Enfin, 1 x 10^{10} particules virales (Pandoravirus, Mollivirus et Mimivirus) ont été centrifugées à 10 000 x *g* pendant 5 minutes puis lavées en PBS.

Les culots obtenus après lavages (trophozoïtes, kystes et virus) ont été resuspendus dans 1 mL d'acide orthophosphorique à 85%, puis incubés pendant 24 heures à 4°C, avec de fréquentes agitations vigoureuses.

2.9.2 Incubations en cellulase

Les cellules et virus en acide orthophosphorique ont ensuite été dialysés deux fois en eau ultra-pure pendant 3 heures à température ambiante puis 16 heures à 4°C. Du Tris-Maleate à pH 6 ainsi que de l'azide de sodium ont été ajoutés à des concentrations finales respectives de 20 mM et 0,01%. Ensuite, 250 μL d'échantillon ont été incubés, soit sans enzyme (témoin), soit en présence de cellulase Cel9P [8] à une concentration finale de 4,3 μM, pendant 24 heures à 37°C sous agitation modérée (350 rpm). Les échantillons ont ensuite été refroidis dans de la glace, puis centrifugés à 10 000 x *g* pendant 10 minutes à 4°C. Enfin, 200 μL de surnageant ont été mélangés avec 50 μL de NaOH à 0,5 M.

8. La cellulase Cel9P est obtenue à partir de *Ruminiclostridium cellulolyticum* (RAVACHOL *et al.*, 2014).

2.9.3 Ouverture des virions et digestions enzymatiques

Ouverture des particules virales

Les particules virales de Pandoravirus et de Mollivirus (1 x 10^{10}) ont été partiellement lysées dans un tampon, habituellement utilisé en protéomique, composé de 40 mM de Tris-HCl pH 7,5, de 2% de SDS et de 60 mM de DTT. Après 30 minutes d'incubation à température ambiante, l'ouverture de ces particules a été contrôlée au microscope optique : cette ouverture est caractérisée par un affaissement des virions, leur donnant un aspect "troué". Les parois virales isolées ont ensuite été lavées plusieurs fois en eau ultra-pure, afin d'éliminer toute trace de SDS et de DTT, avant d'être incubées avec de la cellulase (cf. 2.9.2).

D'un autre côté, des particules virales purifiées de Mimivirus, de Pandoravirus et de *Mollivirus sibericum* ont été cassées selon différentes conditions, afin de rendre la cellulose qui serait présente à l'intérieur des virions accessible aux enzymes.

Pour cela, les particules de Mollivirus et Pandoravirus ont été cassées à l'aide de billes d'acier inoxydable de 0,8 mm de diamètre. Trois tubes contenant chacun 2 x 10^{10} particules virales dans 1 mL d'eau ultra-pure avec 50 billes ont été passés dans un FastPrep®-24 (MP Biomedicals) : cette machine sert notamment à homogénéiser et lyser les échantillons à haute-vitesse. Les particules ont été cassées en réalisant 12 cycles de 15 accélérations (6,5 m/s pendant 20 secondes), et incubées dans la glace entre chaque cycle. Après une dernière incubation dans la glace, les échantillons ont été récupérés et centrifugés à 16 000 x *g* pendant 5 minutes.

En revanche, les particules de Mimivirus, qui ne peuvent pas être ouvertes à l'aide de billes, ont été traitées chimiquement. Trois tubes contenant chacun 2 x 10^{10} particules virales ont été placés dans 40 mM de Tris-HCl pH 7,5, en présence de 0,05 mg/mL de trypsine, 10 mM de DTT, et 1 mM de $CaCl_2$, puis incubés pendant 1 heure à 37°C. Après incubation, 0,001% de triton a été ajouté, et les échantillons ont été incubés à 37°C pendant 1 heure, puis centrifugés à 16 000 x *g* pendant 5 minutes, et les culots ont été lavés 3 fois en eau ultra-pure.

Digestions enzymatiques

Les culots ont été repris dans 1 mL d'eau ultra-pure, puis divisés en 6 tubes, centrifugés à 16 000 x *g* pendant 5 minutes. Ils ont ensuite été resuspendus dans 250 μL des solutions suivantes :

— 20 mM de tampon citrate à pH 5,2, 1 mM de $CaCl_2$ et 0,01% de NaN_3 (azide de sodium), qui constitue le témoin sans enzymes (tampon d'incubation) ;

— le même tampon d'incubation contenant 10% de Viscozyme® (complexe de plusieurs enzymes, qui contient de l'arabanase, de la cellulase, de la β-glucanase, de l'hemicellulase, et de la xylanase, et dont le produit de dégradation final est le

glucose, Sigma), préalablement dialysé dans une unité d'ultra-filtration (Concentrateur Vivaspin 500 en PES (polyéthersulfone) avec un seuil de coupure de 10 000 Daltons, Sartorius) afin de retirer le glucose déjà présent dans la préparation enzymatique ;

— le tampon d'incubation contenant 10% de Celluclast®(cellulase provenant du fungi *Trichoderma reesei*, qui catalyse la dégradation de la cellulose en glucose, cellobiose, et en polymères de glucose plus élevés, Sigma), également dialysé ;
— le tampon d'incubation contenant 10% de Viscozyme® brut (non-dialysé) ;
— le tampon d'incubation contenant 10% de Celluclast® brut (non-dialysé) ;
— le tampon d'incubation contenant 4 nM de cellulase.

Ces mélanges ont ensuite été incubés pendant 5 jours à 37°C pour les échantillons en Viscozyme®, à 50°C pour les échantillons en Celluclast®, et à 37°C pendant 24h pour ceux en cellulase.

2.9.4 Analyse des sucres solubles

Les sucres solubles libérés par l'action des enzymes ont été analysés par chromatographie à échange d'ions haute pression ou HPAEC-PAD ("*High Pressure Anion Exchange Chromatography coupled with Pulsed Amperometric Detection*") sur un Dionex (ICS 3000, Sunnyvale, Californie). Vingt-cinq microlitres d'échantillon ont été injectés sur une colonne analytique Dionex CarboPac PA1 (4 x 250 mm), qui est précédée d'une pré-colonne (4 x 50 mm).

Les sucres ont été élués avec deux tampons, les éluants A (0,1 M NaOH) et B (0,5 M acétate de sodium et 0,1 M NaOH), par une procédure divisée en plusieurs étapes : une séparation isocratique (la composition du mélange des solvants constituant la phase éluante est gardée constante, 5 minutes, 95% de tampon A avec 5% de tampon B), un gradient de séparation (8 minutes, 10 à 37% de tampon B), un lavage de la colonne (2 minutes, 99% de tampon B) et une équilibration de la colonne (2,5 minutes, 95% de tampon A et 5% de tampon B). Le débit constant a été réglé à 1 mL/min. Des échantillons contenant du glucose, du cellobiose (Sigma), du cellotriose, du cellotetraose, du cellopentaose et du cellohexaose (Megazyme), à des concentrations connues (de 5 à 100 μM) ont été injectés, et utilisés pour identifier et quantifier les sucres libérés.

2.9.5 Marquages au Calcofluor-white

Cinquante microlitres d'une suspension de kystes ou de trophozoïtes d'*A. castellanii*, ou de particules virales de Mimivirus, Pandoravirus ou Mollivirus, ont été déposés sur une lame gélatinée, la gélatine permettant une meilleure adhésion de l'échantillon sur la lame. Après 10 minutes d'incubation à température ambiante (le temps que les cellules ou les particules adhèrent à la lame), ces 50 μL ont été retirés et la lame a été rincée

avec 100 μL de PBS. Ce PBS a ensuite été remplacé par 5 μL de Calcofluor-white à une concentration finale de 0,01%. Après 5 minutes d'incubation à l'abri de la lumière, la lame a été recouverte d'une lamelle et observée au microscope à épifluorescence.

2.9.6 Digestions de coupes

Des inclusions en résine Epon de particules virales purifiées (Pandoravirus et Mollivirus), ainsi que de kystes d'*A. castellanii* ont été coupées à l'aide d'un ultra-microtome, afin d'obtenir des coupes de 150 nm d'épaisseur. Ces coupes ont ensuite été incubées avec différentes enzymes dans diverses conditions : 1 mg/mL de cellulase à pH 6, 10% de Celluclast® à pH 5,2, 10% de Viscozyme® à pH 5,2, ou avec du tampon d'incubation seul (témoin). Ces coupes ont été incubées à 37°C (Viscozyme®, cellulase, et tampon) ou 50°C (Celluclast®, et tampon) pendant 24h, 48h ou 1 semaine.

Après incubation, les coupes ont été rincées en eau ultra-pure pendant 2 x 5 minutes, contrastées au citrate de plomb pendant 2 minutes à température ambiante, puis rincées 5 x 1 minute en eau ultra-pure, avant d'être observées au MET.

Les "surnageants" d'incubation (les tampons avec ou sans enzymes) qui ont servi à l'incubation des coupes ont été récupérés et les sucres qui ont été libérés ont été analysés.

3 Caractérisation d'un nouveau *Marseilleviridae* et études comparatives

Au cours de ce travail, j'ai cherché la présence de virus géants qui infectent les amibes dans des échantillons environnementaux. Je suis parvenue à isoler plusieurs virus, dont Noumeavirus, appartenant à la famille des *Marseilleviridae*, et je l'ai caractérisé grâce à plusieurs approches (génomique, protéomique et étude de son cycle infectieux au microscope électronique à transmission), en collaboration avec des chercheurs et des ingénieurs du laboratoire.

Cette partie de mon mémoire de thèse fait l'objet d'une publication scientifique soumise chez Nature Communications : Fabre E., Jeudy S., Santini S., Legendre M., Trauchessec M., Couté Y., Claverie J-M., Abergel C., "*Noumeavirus replication relies on a transient remote control of its host nucleus*" (cf. Annexes A et B).

J'ai retranscrit cet article dans ce chapitre, avec des résultats supplémentaires.

3.1 Résultats

3.1.1 Découverte d'un nouveau membre de la famille des *Marseilleviridae*

Un échantillon boueux provenant de Nouvelle-Calédonie, prélevé dans un étang proche de l'aéroport de Nouméa, a été placé en milieu riz, afin d'enrichir la fraction virale de l'échantillon. Après une incubation de 3 semaines à l'obscurité, cet échantillon a été traité avec une combinaison d'antibiotiques et un antifongique, puis a été placé en co-culture avec des amibes *Acanthamoeba castellanii* et *Dictyostelium discoideum* (DH1 et HMX44A). Le phénotype de ces cellules a ensuite été contrôlé tous les jours au microscope optique, afin d'identifier une possible lyse ou infection (cf. 2.1 Chapitre 2).

Une lyse des cellules *A. castellanii* a été observée, sans que l'agent responsable ne puisse être observé au microscope optique. Cet agent cytolytique a pu être amplifié sur des cellules fraîches, indiquant qu'il s'agit d'un parasite, et non d'une toxine présente dans l'échantillon environnemental. Au microscope optique (objectif x63 avec Optovar x1,6), à fort grossissement, on devine des particules avec une morphologie arrondie.

Des cellules infectées et fixées ont été incluses en résine, et des coupes ultra-fines ont été observées au microscope électronique à transmission (MET, cf. 2.5 Chapitre 2). Ces observations ont révélé la présence de particules mesurant environ 200 nm

de diamètre et présentant une capside dont la symétrie icosaédrique est typiquement virale (cf. Figure 3.1). Ces particules semblent être dépourvues de fibres, et possèdent une couche externe qui apparaît translucide, contrairement au compartiment interne, dense aux électrons.

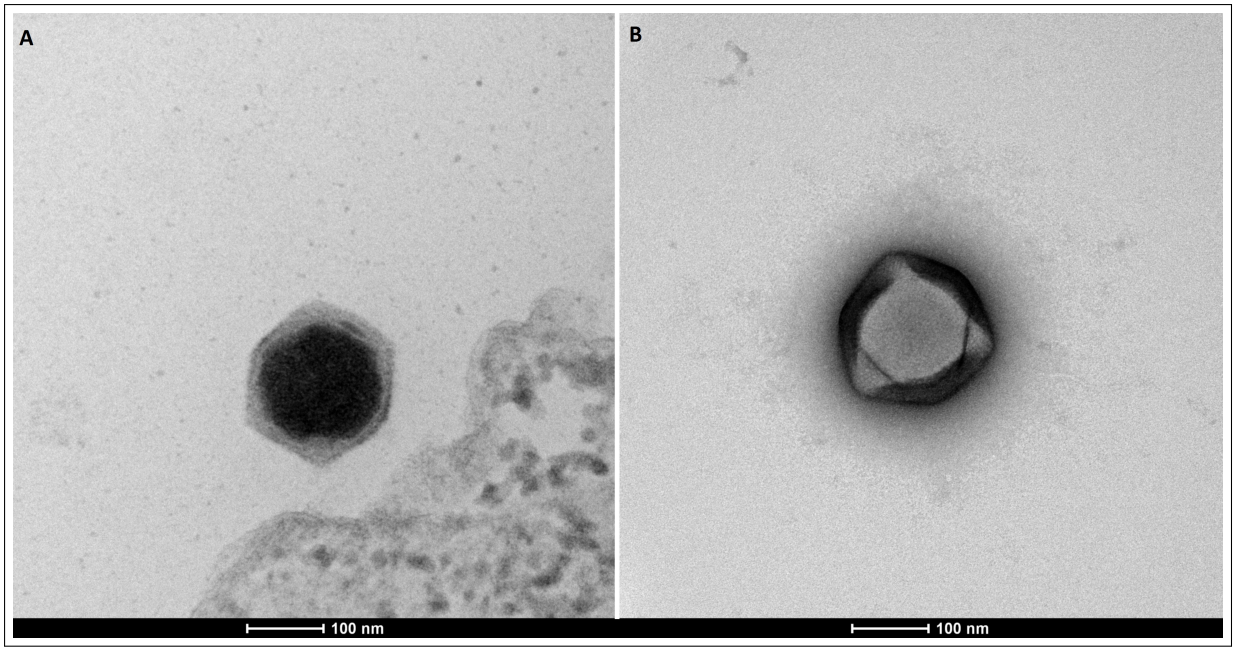


FIGURE 3.1 – Observation du parasite au MET. (A) Coupe ultra-fine d'*A. castellanii* infectée par le parasite : la particule possède une capside à symétrie icosaédrique dépourvue de fibres, son compartiment interne est dense aux électrons et sa couche externe apparaît translucide. (B) Observation en coloration négative du parasite.

En revanche, aucune lyse ou phénotype infectieux n'a pu être observé dans les deux souches d'amibes *D. discoideum* testées, malgré des repiquages successifs. Une infection de ces souches a également été réalisée avec le virus purifié : aucun phénotype infectieux ou lyse cellulaire n'ont été observés.

Ce virus a ensuite été produit en grande quantité sur des cultures fraîches d'*A. castellanii* puis purifié sur gradient de sucrose (cf. 2.2 Chapitre 2). A l'issue de la centrifugation du gradient, un disque blanchâtre contenant les virions a pu être observé dans la couche de sucrose correspondant à 30% (cf. Figure 3.2).

FIGURE 3.2 – Purification du virus. Sur l'image de gauche, le virus dilué en PBS est déposé sur le gradient de sucrose. Sur l'image de droite, après centrifugation du gradient, le disque viral peut être observé dans la couche correspondant à 30% de sucrose.

Les virions ont été récupérés, lavés et contrôlés au microscope optique, révélant des particules à peine visibles mais néanmoins homogènes (cf. Figure 3.3).

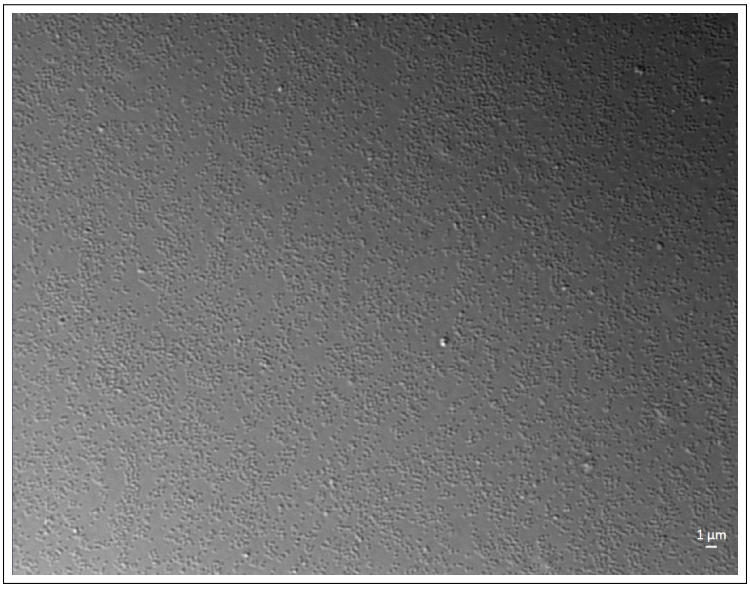


FIGURE 3.3 – Observation du virus au microscope optique (objectif x63 avec Optovar x1,6) : les particules apparaissent arrondies.

Clonage du virus sur cellule unique

Le virus a ensuite été cloné sur cellule unique : des amibes *A. castellanii* ont été infectées avec du virus purifié (non-titré) dans une plaque multipluits. Après 1 heure d'incubation et 50 lavages, les cellules ont été diluées puis réparties dans les autres puits, à raison de 1 cellule par puits, et les puits contenant des gouttes où il n'y avait aucune cellule ont servi de témoins. Des cellules fraîches ont été ajoutées dans tous les puits, et le phénotype a été contrôlé quotidiennement au microscope optique (cf. 2.2

Chapitre 2). Après 3 jours d'incubation, les cellules des puits témoins étaient lysées, indiquant qu'il subsistait des virus dans les puits, et que le lavage était insuffisant.

Cette étape de clonage a donc été recommencée, en réalisant cette fois 100 lavages du puits. Après 5 jours d'incubation à 32°C, les cellules des puits témoins ne présentaient pas de phénotype infectieux, validant ainsi le protocole : deux clones ont donc été sélectionnés, puis purifiés.

Les clones de ce virus ont ensuite été titrés puis caractérisé à l'aide de différentes approches (génomique, protéomique et étude du cycle infectieux) : il s'agit d'un *Marseilleviridae* qui a été appelé Noumeavirus, en accord avec la nomenclature de cette famille.

Titration des clones de Noumeavirus

Les clones de Noumeavirus ont été titrés par dilutions sériées ($TCID_{50}$) selon la méthode de Reed-Muench (REED *et al.*, 1938) (cf. 2.2 Chapitre 2). Ce titre infectieux, ou $TCID_{50}$ correspond au nombre de particules nécessaires pour infecter ou lyser 50% des cellules : ici, les clones de Noumeavirus ont un $TCID_{50}$/mL de 1,75 x 10^{10} particules infectieuses par mL (cf. Figure 3.4).

Reed & Muench Calculator

Created November 20, 2004 by Brett D. Lindenbach, PhD
Modified January 27, 2008

1. Enter the starting dilution	initial dilution:	0,0000001					
2. Enter the dilution factor	dilution factor:	10					
3. Enter the volume tested per well	ml:	0,15					
This is your calculated dilution series:		1,00E-07	1,00E-08	1,00E-09	1,00E-10	1,00E-11	1,00E-12
4. Enter the total # of wells examined per dilution	total wells:	10	10	10	10	10	10
5. Enter the # of positive wells for each dilution	positive wells:	10	10	7	2	0	0
These values are calculated automatically:	negative wells:	0	0	3	8	10	10
	cum pos:	29	19	9	2	0	0
	cum neg:	0	0	3	11	21	31
	% infected:	100,00	100,00	75,00	15,38	0,00	0,00
	prop dist:	0	0	0,419354839	-2,25	0	0
6. This is your TCID50	TCID50:			3,81E-10			
7. This is your TCID50/ml	TCID50/ml:			1,75E+10			

FIGURE 3.4 – Calcul du $TCID_{50}$ de Noumeavirus à l'aide d'une feuille de calcul, où le rapport entre les puits infectés et non-infectés est indiqué (LINDENBACH, 2009).

Profils protéiques des clones de Noumeavirus

Nous avons ensuite étudié le contenu en protéines de chaque clone de Noumeavirus. Pour cela, ces clones ont été lysés, afin d'extraire leurs protéines, et le profil protéique de chaque clone a été déterminé sur un gel d'acrylamide, où nous avons déposé 5 μg de protéines (cf 2.7 Chapitre 2).

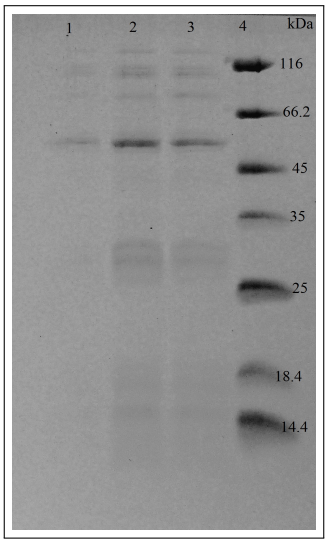


FIGURE 3.5 – Profils protéiques des clones de Noumeavirus obtenus à partir de 5 μg de protéines déposées sur un gel d'acrylamide. La première piste correspond à Noumeavirus non-cloné (1), les deux suivantes sont celles des clones obtenus sur cellule unique (2 et 3), et le marqueur de poids moléculaire est représenté à droite (4).

Ces profils sont strictement identiques et nous permettent d'identifier une bande majoritaire située aux alentours de 50 kDa (cf. Figure 3.5). Un seul de ces clones a donc été sélectionné pour les étapes de caractérisation suivantes.

3.1.2 Analyse génomique de Noumeavirus

L'ADN génomique de Noumeavirus a été extrait et purifié, et 2 μg ont été envoyés à séquencer (cf 2.6 Chapitre 2). Cet ADN a ensuite été utilisé pour générer des séquences "*paired-end*" [1] de 2 x 100 paires de bases, en utilisant la plate-forme de séquençage à haut-débit HiSeq.

Le génome à ADN double-brin de Noumeavirus a été assemblé en un seul contig [2] d'une taille de 376 207 paires de bases, avec un contenu en G+C de 43%.

Au total, 452 ORFs [3] ont été prédits comme codant pour des protéines, d'une taille comprise entre 50 et 1525 acides aminés, avec une taille moyenne de 249 résidus.

1. Les séquences "*paired-end*" sont obtenues par le séquençage des deux extrémités de fragments courts, d'une taille inférieure à 1 kb.

2. Un contig est une séquence obtenue par chevauchement de séquences courtes (ou "*reads*") qui ont été générées par le séquenceur.

3. Un ORF ou "*Open Reading Frame*", correspond à un cadre de lecture ouvert, qui est délimité par deux codons Stop et qui peut coder pour une protéine.

Les régions codant pour des protéines représentent 90% du génome de Noumeavirus, et sont séparées par de courtes régions intergéniques, de 88,2 nucléotides de long en moyenne.

Le génome de Noumeavirus a été assemblé comme une séquence circulaire : toutefois, il pourrait s'agir d'une séquence linéaire, où chaque extrémité serait redondante et circulairement permutée. C'est aussi le cas pour les autres représentants des *Marseilleviridae* dont les génomes ont été entièrement séquencés, ou pour les *Iridoviridae* (WILLIAMS *et al.*, 2005), qui présentent des affinités phylogénétiques avec les *Marseilleviridae* (BOYER *et al.*, 2009 ; THOMAS *et al.*, 2011).

Parmi les 452 protéines prédites chez Noumeavirus, seulement 9 protéines (2%) correspondent à des ORFans [4] et 438 protéines ont un homologue proche chez les *Marseilleviridae* (cf. Figure 3.6) : Noumeavirus est donc un nouveau membre de la famille des *Marseilleviridae*.

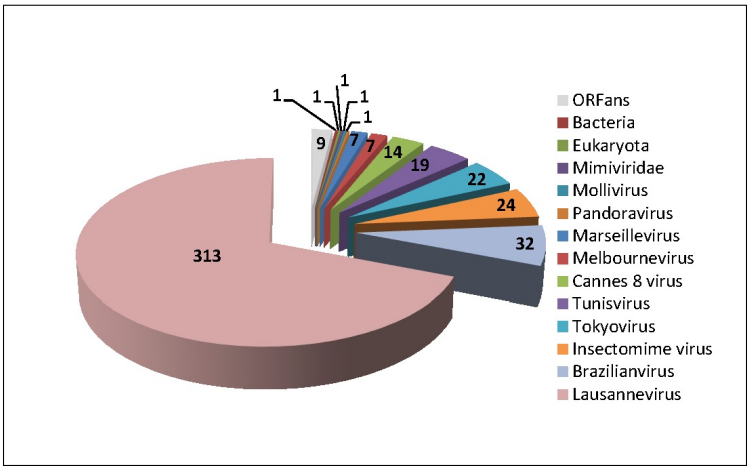


FIGURE 3.6 – Répartition des homologues des ORFs de Noumeavirus, déterminés à l'aide de BlastP contre la base de données NR (NCBI RESOURCE COORDINATORS, 2016). Les ORFs de Noumeavirus ont une majorité d'homologues chez Lausannevirus.

Il a été proposé récemment que la famille des *Marseilleviridae* soit divisée en 5 lignées (DOS SANTOS *et al.*, 2016), où Marseillevirus est le représentant de la lignée A (BOYER *et al.*, 2009), Lausannevirus de la lignée B (THOMAS *et al.*, 2011), Tunisvirus de la lignée C (AHERFI *et al.*, 2014a), *Brazilian marseillevirus* de la lignée D (DORNAS *et al.*, 2016) et enfin, le plus récent, *Golden marseillevirus*, qui a été isolé à partir d'eau de mer contenue dans une moule, est le fondateur de la lignée E (DOS SANTOS *et al.*, 2016) (cf. Figure 3.7).

4. Les ORFans sont des gènes qui n'ont pas d'homologues détectés dans les bases de données de séquences.

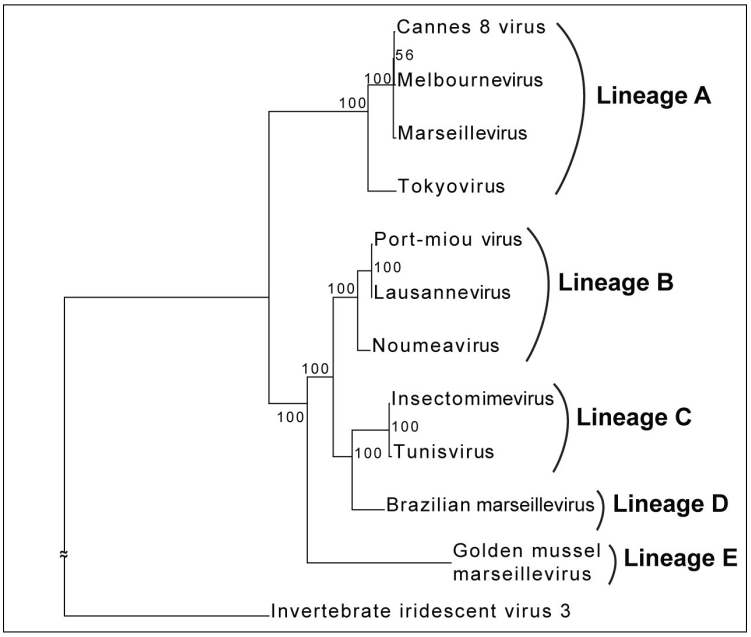


FIGURE 3.7 – Arbre phylogénétique de la famille des *Marseilleviridae*. L'arbre a été obtenu à partir de la concaténation des alignement des 21 gènes orthologues des *Marseilleviridae*, qui sont partagés avec l'Iridovirus Invertebrate iridescent virus 3.

Les analyses du génome, ainsi que des analyses phylogénétiques, ont permis de montrer que Noumeavirus appartient à la lignée B des *Marseilleviridae*, et est proche de Lausannevirus (THOMAS *et al.*, 2011) et de Port-Miou virus (DOUTRE *et al.*, 2015) (cf. Figure 3.7). Noumeavirus partage 374 protéines prédites avec Lausannevirus, avec une identité moyenne de 77%, et 331 protéines avec Melbournevirus (DOUTRE *et al.*, 2014), avec une identité moyenne de seulement 55% (cf. Table S4 Annexe B).

Melbournevirus [5] (lignée A) et Noumeavirus (lignée B) étant les virus les plus distants au sein de cette famille, et ayant été isolés au laboratoire, des analyses comparatives ont été réalisées sur ces deux virus.

Ces analyses ont permis d'observer que deux ORFs adjacents de Noumeavirus montraient une similarité assez forte avec les séquences situées dans les domaines N-terminal (pour l'un) et C-terminal (pour l'autre) de la grande sous-unité de l'ARN polymérase ADN dépendante (RPB1). Ces deux ORFs soulèvent donc l'interrogation suivante : l'homologue de la protéine RPB1 chez Noumeavirus est-il divisé en plusieurs gènes ou est-il interrompu par un intron ?

Pour répondre à cette question, la séquence du transcrit correspondant à RPB1, recouvrant toute la région qui a été prédite comme étant intergénique et qui sépare les deux domaines, a été amplifiée. Le produit PCR qui a été séquencé a permis de déterminer que ce gène contient en effet un intron, dont les bornes ont été précisément délimitées. Les sites caractéristiques d'un intron épissable [6] n'ont pas été identifiés dans

5. Melbournevirus a été isolé en 2012 à partir d'un échantillon environnemental prélevé en Australie, qui a également permis d'isoler *Pandoravirus dulcis* (PHILIPPE *et al.*, 2013).

6. Les introns épissables ou "*spliceosomal intron*" sont excisés à l'aide du spliceosome de la cellule.

cette séquence, qui a donc été analysée à l'aide du serveur web rFAM (NAWROCKI *et al.*, 2015). Les résultats de cette analyse ont montré que cette séquence correspondait à un intron auto-catalytique [7] du groupe I.

A l'exception de Port-Miou virus (DOUTRE *et al.*, 2015), l'intron auto-catalytique du gène RPB1 de Noumeavirus (NMV11) est conservé dans tous les génomes des virus de la famille des *Marseilleviridae* qui ont été entièrement séquencés (il s'agit du gène MEL47 chez Melbournevirus, avec une e-value de 1,5 e^{-8}).

Un autre gène possiblement séparé par un intron a également été étudié : il code pour une protéine de liaison à l'origine de réplication (ORBP ou "*origin of replication binding protein*") pour laquelle 4 paralogues ont été trouvés dans le génome de Noumeavirus. Il a été démontré expérimentalement qu'un ARN messager chevauchait deux ORFs adjacents, séparés par un codon Stop. Cette structure, qui a été annotée NMV287-288, pourrait correspondre à un pseudogène récent, qui serait toujours transcrit.

Huit régions génomiques contenant des introns épissables putatifs ont été prédites en utilisant le logiciel Exonerate (SLATER *et al.*, 2005). Parmi ces 8 régions, 3 ont été amplifiées et séquencées (NMV125, NMV252 et NMV436), mais aucun intron n'a pu être identifié dans ces régions. Ces résultats sont cohérents avec l'hypothèse selon laquelle les gènes de Noumeavirus ne sont pas transcrits dans le noyau de l'hôte.

Noumeavirus possède des gènes codant pour des protéines de type histone, comme tous les membres de la famille des *Marseilleviridae* (BOYER *et al.*, 2009 ; THOMAS *et al.*, 2011 ; AHERFI *et al.*, 2014a ; AHERFI *et al.*, 2014b ; DOUTRE *et al.*, 2014 ; DORNAS *et al.*, 2016 ; DOUTRE *et al.*, 2015 ; ERIVES, 2015). Ces protéines pourraient jouer un rôle dans la compaction et la stabilité de l'ADN viral dans les particules. Cette hypothèse est également soutenue par la présence d'autres gènes, codant pour une ADN topoisomérase II, qui est notamment impliquée dans la compaction de la chromatine. L'origine ainsi que l'histoire évolutive de ces protéines reste toutefois incertaine (ERIVES, 2015).

Analyse des protéines impliquées dans la maturation des transcrits

Deux protéines impliquées dans la maturation des transcrits de Mimivirus (PRIET *et al.*, 2015) (et possiblement de Pithovirus, LEGENDRE *et al.*, 2014) présentent des homologues chez tous les *Marseilleviridae*. Il s'agit de l'enzyme de coiffe des ARN messagers, qui ajoute la coiffe en 5' et la méthyle en N7 (BENARROCH *et al.*, 2008) (NMV139 et MEL282), et de la RNase III (LEGENDRE *et al.*, 2010 ; BYRNE *et al.*, 2009 ; PRIET *et al.*, 2015) (NMV212 et MEL219), reconnaissant la structure palindromique située en 3' des ARNm de Mimivirus, avant leur clivage et leur polyadénylation. Cette observation a soulevé la question suivante : Noumeavirus, et les *Marseilleviridae* en général, obéissent-ils à la règle de l'épingle à cheveux (ou "*hairpin-rule*") utilisée par les *Mimiviridae* pour l'arrêt de la transcription et la polyadénylation des transcrits (BYRNE *et al.*, 2009) ?

7. Les introns auto-catalytiques ou "*self-splicing intron*" sont capables de s'exciser seuls.

L'analyse des génomes a permis de révéler la présence de 226 palindromes chez Noumeavirus, et 245 chez Melbournevirus. Chez Noumeavirus, 75 de ces palindromes ont été retrouvés dans des régions codantes du génome (qui représentent 90% du génome), et 151 sont retrouvés entre les ORFs (10% du génome). Cette distribution, qui n'est pas uniforme, est très significative ($p<0.0001$, test exact de Fisher), et a également été retrouvée chez Melbournevirus, qui possède 87 palindromes dans des régions codantes (89,8% du génome) et 158 entre les ORFs (10,2% du génome) (cf. Figure S4 Annexe B). Ces distributions suggèrent que ces palindromes pourraient fonctionner comme des signaux de terminaison et/ou de polyadénylation des ARN messagers.

3.1.3 Analyse protéomique de Noumeavirus

Nous avons extrait les protéines de Noumeavirus et Melbournevirus en parallèle, puis les lysats protéiques ont été envoyés en protéomique, où les spectres obtenus ont été comparés avec les ORFs prédits des *Marseilleviridae*. Nous avons ensuite analysé ces protéomes, puis nous les avons comparés entre eux (cf 2.7 Chapitre 2).

Ces analyses ont permis d'identifier 283 protéines chez Noumeavirus (dont 128 sont virales et 155 proviennent de l'hôte), et 196 protéines chez Melbournevirus (dont 132 sont virales et 64 proviennent de l'hôte), classées selon leurs abondances estimées. La différence entre le nombre de protéines identifiées chez ces deux virus est principalement due au fait que les particules de Noumeavirus possèdent beaucoup plus de protéines provenant de l'hôte *A. castellanii* (54,8%) que celles de Melbournevirus (32,65%). En revanche, le nombre de protéines virales est équivalent pour chaque virus (132 *versus* 128). Beaucoup de protéines de l'hôte étant détectées en faible abondance, leur présence est probablement due à des conditions différentes de purification ou de stockage, plutôt qu'à des variations biologiques dans la composition des virions (cf. Table S1 Annexe B).

Les protéines de l'hôte détectées uniquement dans un virus (103 protéines chez Noumeavirus et 12 chez Melbournevirus) correspondent probablement à des protéines associées aux virions de Noumeavirus, qui n'a pas été stocké en DMSO, un agent cryoprotecteur suspecté de "décaper" légèrement les particules. Ces protéines pourraient également être des "*by-standers*", empaquetés dans les virions de façon aléatoire à moins d'une copie par virion, et qui seraient dénués de fonctions biologiques. Cette hypothèse est supportée par le fait que la plupart des protéines codées par les virus sont partagées par les deux virions (89% pour Noumeavirus et 86,4% pour Melbournevirus, cf. Table S1 Annexe B).

Les protéines de l'hôte partagées par les virions

Les analyses de protéomique ont permis d'identifier 52 protéines d'*A. castellanii* partagées par les virions de Noumeavirus et Melbournevirus. Les protéines les plus abondantes sont une porine mitochondriale et l'actine 1, également abondantes dans le protéome des amibes non-infectées (cf. Figure 3.8). Les autres protéines de l'hôte partagées par les deux virions (à l'exception de la "*Cuivre/Zinc superoxide dismutase*"), sont situées en-dessous des protéines en 100^{e} position, et sont détectées dans des virions non-apparentés infectant les amibes *Acanthamoeba*, ce qui pourrait refléter leur abondance élevée dans les cellules non-infectées (cf. Figure 3.8).

Cependant, il est possible que certaines d'entre elles puissent participer au processus d'infection, comme par exemple, le trafic intra-cytoplasmique de composants viraux à travers la réorganisation du cytosquelette de la cellule hôte (tubulines, fascines, ...).

Deux d'entre elles sont des protéines de stress oxydatif : la "*Cu/Zn superoxide dismutase*", également partagée par d'autres virus, et la "*Mn/Fe superoxide dismutase*", uniquement retrouvée dans les virions de Noumeavirus et de Melbournevirus. Ces protéines pourraient permettre aux virions de résister au stress oxydant lors de leur passage dans les vacuoles de l'amibe.

Vingt et une protéines de l'hôte sont uniquement partagées par les virions de Noumeavirus et Melbournevirus : deux d'entre elles correspondent à des protéines abondantes, classées 4^{e} ("*Glycéraldéhyde-3-phosphate déhydrogénase*") et 18^{e} ("*hypothetical, protein binding domain*") dans le protéome des cellules non-infectées (cf. Figure 3.8). La présence étonnante de ces protéines à des rangs élevés pourrait être liée à une spécificité du processus d'infection des *Marseilleviridae*.

De nombreuses petites GTPases (des homologues "*6 RAS-like*"), ainsi que des protéines du cytosquelette (alpha et bêta tubulines, calponines) et une kinase, pourraient être impliquées dans le trafic rapide de composants viraux et cellulaires, que ce soit vers ou à partir du noyau de l'hôte (cf. Figure 3.8).

Enfin, parmi les protéines de l'hôte uniquement identifiées dans les protéomes de Noumeavirus et de Melbournevirus, une protéine composée d'un tandem de domaines histone est retrouvée en faible abondance (1621^{e}) dans le protéome des cellules non-infectées (cf. Figure 3.8). Il est intéressant de noter qu'un homologue de cette protéine est codé par les virus de la famille des *Iridoviridae*, et que cette protéine est essentielle à la réplication virale. Une protéine telle que celle-ci est normalement confinée dans le noyau des cellules : sa détection dans les particules des *Marseilleviridae* suggère que l'usine virale cytoplasmique, où sont assemblés les virions, serait infiltrée par des composants nucléaires.

Gene	Protein name	Real rank in Noumeavirus	Real rank in Melbournevirus	Host relative rank in Noumeavirus	Host relative rank in Melbournevirus	Rank in host (NI)	Other viruses
gunplaced-2126338-2127783	Actin 1	63	78	2	2	1	+
g422-496887-497476	fascin subfamily protein	124	119	13	11	2	+
g1525-75307-76608	Euk Porin protein	45	65	1	1	3	+
g1185-112487-113838	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic-like	229	172	47	37	4	0
g414-411909-412468	C2 domain containing protein	127	109	15	7	5	+
g696-284312-285231	LIM domain containing protein	118	145	10	24	9	+
g527-633449-637592	Cobalaminin dependent methionine synthase	247	195	50	52	10	0
g1390-216047-218303	High molecular weight heat shock protein	204	191	41	50	12	+
g1317-556358-557497	Hypothetical, protein binding	248	177	51	41	18	0
g218-68723-70868	Corornin3 putative	147	179	23	43	42	+
gunplaced-1973089-1973871	Copper/zinc superoxide dismutase	167	83	30	3	54	+
g254-213866-215830	Peptidase M16 inactive domain containing protein	212	188	43	47	72	+
g726-224146-225044	Wnt-10b	109	117	6	10	92	+
g1359-61676-62848	hypothetical protein	103	125	4	13	129	+
g757-590645-592959	C2 domain containing protein	123	127	12	14	138	+
g1390-714707-716810	Alpha tubulin, partial	230	132	48	17	140	0
g531-97521-99028	Actin related protein 2/3 complex, 41kDa subunit	179	178	33	42	152	+
g631-162107-163748	Tubulin beta chain	216	141	46	22	164	0
g376-450681-452588	peroxidase	114	106	9	5	174	+
gunplaced-1669823-1671094	Mn/Fe superoxide dismutase	214	166	45	32	188	0
cox1/2	cox1/2	192	192	38	51	220	+
g1406-111173-112358	hypothetical protein	158	175	28	39	224	0
g42-42086-43890	Rac-like protein	129	173	16	38	269	0
g446-42975-43598	hypothetical protein	132	114	17	8	286	+
g1369-497445-500306	Gelation factor, putative	209	169	42	35	290	0
g284-678893-681196	PH domain containing protein	190	161	37	30	308	+
g376-448106-449758	Peroxidase	187	149	35	26	309	+
g1369-729197-730236	Ras-like protein 1	135	133	18	18	331	0
g817-77076-78510	C8 sterol isomerase	102	107	3	6	333	+
g611-72236-73493	exported protein putative	111	115	7	9	338	0
g631-243047-244467	Rab GTPase	142	168	20	34	356	0
g185-111308-112608	autophagy-related protein 27	201	135	40	20	370	+
g397-418028-419548	Ras GTPase Rap1b	154	164	27	31	374	0
g1317-154778-156138	Rho family small GTPase Rac3	143	131	21	16	401	0
g1185-1867882-1870710	aspartic proteinase	113	147	8	25	475	+
g414-40155-41168	hypothetical protein	151	182	26	46	484	0
g527-461746-463738	Ras GTPase	180	155	34	27	487	+
g680-132237-134308	Protein kinase C4, partial	173	189	32	48	521	0
g454-198036-200147	EF-hand containing protein	160	120	29	12	524	+
g1007-410536-413510	Predicted protein	232	170	49	36	570	+
g935-165792-170359	Peroxidase	126	139	14	21	789	+
g1185-1806178-1808434	Protein phosphatase 1 catalytic subunit alpha, putative	194	134	39	19	837	+
g1007-257770-259328	hypotheticalprotein	150	180	25	44	865	+
g254-27747-28588	hypothetical protein	105	103	5	4	1080	+
g1317-272306-275600	Sec1 family protein, partial	260	190	52	49	1143	+
g1007-388106-392098	hypothetical protein	120	159	11	29	1391	0
g253-184876-188554	Hydrolase activity	136	181	19	45	1555	+
g1185-98157-99558	Histone cluster 1, H2bb, putative	213	158	44	28	1621	0
g1406-789283-793494	Glutamate carboxypeptidase	169	176	31	40	1783	0
g631-805994-806720	g1406-111173-112358	148	128	24	15	2601	+
g185-637578-643589	hypothetical protein	144	167	22	33	3020	+
gunplaced-452590-455256	Ef1a	188	142	36	23	0	0

FIGURE 3.8 – Protéines de l'hôte partagées par Noumeavirus et Melbournevirus : leur classement dans les protéomes des virions est indiqué dans les colonnes "real rank", leur classement parmi les protéines de l'hôte retrouvées dans ces protéomes est retrouvé dans les colonnes "relative rank", leur classement dans le protéome des cellules *A. castellanii* non-infectées est indiqué dans la colonne "NI", et leur présence/absence dans des virus n'appartenant pas à la famille des *Marseilleviridae* est signalée dans la colonne "other viruses".

Les protéines virales identifiées dans les virions

Les analyses comparatives des protéomes des virions ont permis de montrer que parmi les 114 protéines codées par les virions, et partagées par Noumeavirus et Melbournevirus, 63% (soit 72 protéines) correspondent à des protéines de fonction inconnue ("*hypothetical proteins*"), dont 20 possèdent des domaines transmembranaires (cf. Table S1 Annexe B). La corrélation entre leurs classements respectifs met en évidence une forte conservation dans la composition des virions des *Marseilleviridae* (R^2=0.8641, p-value « $1e^{-15}$, cf. Figure 1 Annexe A).

Parmi les 14 protéines virales absentes des virions de Melbournevirus, 12 sont situées en-dessous des protéines classées 100^e dans le protéome des particules de Noumeavirus. Ces protéines auraient pu ne pas être détectées lors des analyses de protéomique, notamment à cause de l'absence de réplicats biologiques. Les deux autres sont classées 25^e et 42^e dans le protéome de Noumeavirus et correspondent à des protéines qui ne sont pas codées par Melbournevirus. Une d'entre elles fait partie d'une famille de grandes protéines, pour laquelle des paralogues ont été retrouvés dans les virions de Melbournevirus.

De même, les 18 protéines absentes du protéome des virions de Noumeavirus sont toutes classées en-dessous de la 110^e protéine dans le protéome des virions de Melbournevirus et 5 d'entre elles ne sont pas codées par le génome de Noumeavirus.

Ces analyses comparatives ont permis de découvrir des résultats surprenants. Tout d'abord, la protéine majeure de capside (ou "*Major Capsid Protein*"), très concentrée dans les particules virales, ne semble pas être la protéine la plus abondante, contrairement aux autres virus présentant une capside de symétrie icosaédrique (Arslan *et al.*, 2011 ; Dunigan *et al.*, 2012 ; Fischer *et al.*, 2014 ; W. Li *et al.*, 2014). Cette protéine est seulement classée en 3^e position dans le protéome de Noumeavirus (NMV116), et en 4^e position dans celui de Melbournevirus (MEL305).

Étonnamment, la protéine la plus abondante (NMV189 et MEL236), est une protéine de fonction inconnue d'environ 150 acides aminés, qui est très conservée et unique aux *Marseilleviridae*.

La seconde protéine la plus abondante dans le protéome de Noumeavirus (NMV78) est une "*histone H3-like*" classée 3^e chez Melbournevirus (MEL368). Deux autres "*histone-like*" codées par les virus sont également abondantes dans les protéomes (NMV79 et NMV419 ; MEL247 et MEL149). Le nombre de ces protéines retrouvées dans les virions (3 sont codées par les virions, 1 par l'hôte) suggère fortement qu'elles soient impliquées dans l'empaquetage de l'ADN viral ; ce type de protéine a également été identifié dans les protéomes de certains virions de la famille des *Iridoviridae*, qui ont une forte affinité phylogénétique avec les *Marseilleviridae* (W. Li *et al.*, 2014).

Les virions partagent également 9 protéases, 2 lipases et 5 protéines impliquées dans le stress oxydatif, toutes codées par les virions (comme des péroxydases ou des

thiorédoxines), qui leur permettraient de résister à l'environnement agressif du phagosome, en complément de celles codées par l'hôte et embarquées dans les virions (cf. Paragraphe 3.1.3).

Enfin, 3 glycosyl-transférases, également codées par les virions, pourraient être impliquées dans le trafic intracellulaire des protéines virales ou de l'hôte en fonction de leur glycosylation.

Ces analyses ont également permis de démontrer que l'enzyme de coiffe des ARN messagers ("*mRNA capping enzyme*"), classée 93^{e} chez Noumeavirus et 80^{e} chez Melbournevirus, ainsi que la RNase III (56^{e} chez Noumeavirus et 51^{e} chez Melbournevirus) sont présentes dans les virions (cf. Figure 3.9). Leur présence suggère que la maturation des transcrits viraux doit être réalisée par des protéines codées par les virions dès le début du cycle infectieux, comme chez les virus ayant un cycle exclusivement cytoplasmique (PRIET *et al.*, 2015).

Gene name	Mimivirus				Pithovirus sibericum				Noumeavirus				Melbournevirus			
	Gene #	Theoretical	Identified Peptides	Rank	Gene #	Theoretical	Identified Peptides	Rank	Gene #	Theoretical	Identified Peptides	Rank	Gene #	Theoretical	Identified Peptides	Rank
Rpb1	R501	77	50	65	pv_366 -368	51	35	68	NMV_011	70	noID		MEL_047	72	noID	
Rpb2	L244	60	30	80	pv_242	61	40	78	NMV_436	55	noID		MEL_022	63	noID	
Rpb5	L235	12	7	54	pv_31	11	8	72	NMV_046	14	noID		MEL_342	9	noID	
polyA pol	R341	30	14	87	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
RNAse III	R343	28	noID		pv_75	20	15	63	NMV_212	18	10	56	MEL_219	16	8	51
mRNA capping enzyme	R382	72	41	83	pv_454	52	38	52	NMV_139	45	28	93	MEL_282	46	28	80

FIGURE 3.9 – Protéines impliquées dans la transcription et la maturation des transcrits. Le tableau représente les peptides identifiés théoriquement ("theoretical", prédits à l'aide d'ExPASy) et expérimentalement ("identified peptides") et leur classement dans les protéomes des virions de virus cytoplasmiques (Mimivirus et *P. sibericum*) *versus* Noumeavirus et Melbournevirus. NA : le gène correspondant n'existe pas ; noID : les peptides n'ont pas été identifiés.

Enfin, il est intéressant de noter que dans le protéome des virions suivant la règle de l'épingle à cheveux ("*hairpin-rule*"), les protéines de maturation des transcrits ont habituellement une abondance équivalente à celle de l'appareil de transcription (cf. Figure 3.9). Dans cette étude, les protéines de maturation des transcrits sont présentes avec des abondances comparables à celles de Mimivirus et de Pithovirus : la machinerie de transcription de Noumeavirus et de Melbournevirus aurait donc due être identifiée.

3.1.4 Étude du cycle infectieux de Noumeavirus

Étude du cycle infectieux de Noumeavirus au microscope à épifluorescence

Un des résultats les plus inattendus lors de l'étude comparative des protéomes de Noumeavirus et de Melbournevirus, a été l'absence flagrante d'une ARN polymérase, qui soit codée par les virions ou par l'hôte, alors que la machinerie de transcription a toujours pu être identifiée chez les autres virus ayant un cycle exclusivement cytoplasmique (cf. Figure 3.9, ARSLAN *et al.*, 2011 ; FISCHER *et al.*, 2014 ; DOELLINGER *et al.*, 2015).

Cette observation suggère que bien que les *Marseilleviridae* semblent avoir un cycle infectieux cytoplasmique, ils auraient en réalité besoin de la machinerie de transcription de l'hôte pour générer leurs transcrits précoces. Cette machinerie de transcription est normalement confinée dans le noyau de la cellule hôte : Noumeavirus et Melbournevirus passeraient donc par une phase nucléaire au cours de leur cycle infectieux pour générer ces transcrits.

Ce passage peut se faire de deux façons : le génome viral peut être transféré au noyau de l'hôte, et l'ARNm viral correspondant serait ensuite transféré dans le cytoplasme pour produire la machinerie de transcription. D'un autre côté, les fonctions nucléaires requises pourraient être recrutées à l'usine virale dans le cytoplasme. Les modifications apportées au noyau de la cellule hôte au cours du cycle infectieux ont été étudiées, en suivant l'infection des amibes *A. castellanii*, entre 30 minutes et 4 hpi (cf 2.4 Chapitre 2).

Ces amibes produisent des méthyltransférases de Mollivirus qui sont nucléaires (Ml135-GFP, 62 kDa) ou nucléolaires (Ml216-GFP, 58 kDa), mais aussi, la protéine nucléaire SUMO [8] (GFP-SUMO, 39 kDa) de l'amibe *A. castellanii*, toutes fusionnées à la GFP [9] (cf 2.4 Chapitre 2). Elles ont été utilisées afin de suivre l'évolution du noyau et du nucléole au cours de l'infection, et le marquage DAPI nous a permis de suivre l'ADN de l'hôte, ainsi que l'apparition de l'usine virale dans le cytoplasme.

Les observations au microscope à épifluorescence montrent que les cellules non-infectées présentent toutes une fluorescence de la GFP ainsi qu'un marquage DAPI, qui sont situés exclusivement dans le noyau (cf. B Figure 3.10).

Les cycles infectieux de Noumeavirus et de Melbournevirus présentent quelques différences . En effet, les cellules infectées par Noumeavirus se décollent après 1 hpi et forment des agrégats visibles à l'œil nu à partir de 2 hpi, alors que celles qui sont infectées par Melbournevirus restent adhérentes jusqu'à la fin du cycle.

8. La protéine SUMO ou "*Small Ubiquitin-like Modifier*", est retrouvée chez tous les eucaryotes, et impliquée dans un mécanisme de modification post-traductionnelle (SUMOylation). Cette protéine est également impliquée dans des activités au sein du noyau de la cellule.

9. La GFP ou "*Green Fluorescent Protein*" émet une fluorescence verte, et est fréquemment utilisée comme rapporteur : lorsqu'elle est fusionnée avec une protéine d'intérêt, elle permet de suivre directement la localisation de cette protéine dans la cellule.

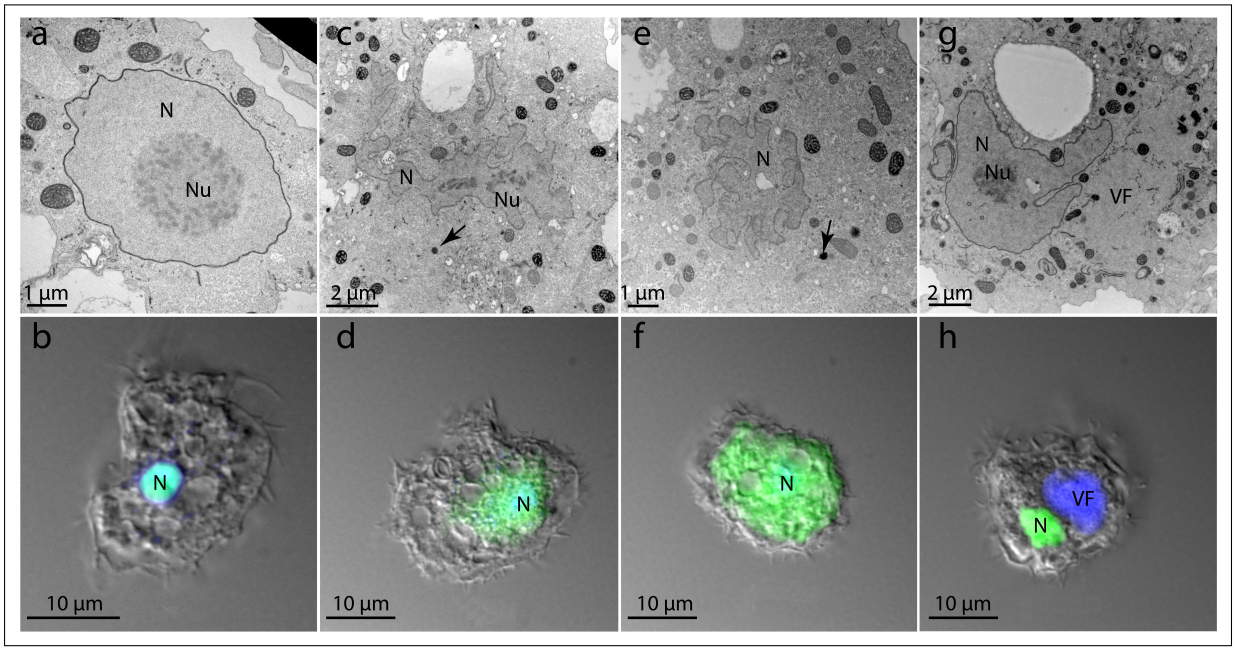


FIGURE 3.10 – Modifications du noyau des cellules au cours du cycle infectieux de Noumeavirus, observées au MET (A, C, E, G) et au microscope à épifluorescence (B, D, F, H, superposition contraste de phase, DAPI, GFP). (A) et (B) Noyau (N) des cellules non-infectées : le nucléole (Nu) est dense et compact et les marquages DAPI et GFP sont localisés dans le noyau. (C) et (D) Après 30 minutes pi, le noyau est drastiquement modifié, et sa membrane subit de nombreuses invaginations. Le nucléole apparaît désorganisé et moins dense aux électrons. Le marquage DAPI reste localisé au noyau, alors que la fluorescence GFP se diffuse dans le cytoplasme. Des nucléoïdes denses aux électrons ont été observés (flèche noire). (E) et (F) A 1 hpi, le noyau est toujours déformé, et toujours marqué au DAPI. La fluorescence GFP n'est plus localisée au noyau et remplit le cytoplasme. Des nucléoïdes denses aux électrons ont également été observés (flèche noire). (G) et (H) Après 4 hpi, le noyau a retrouvé sa forme sphérique, avec un nucléole dense et compact, et coexiste avec les usines virales (VF). La fluorescence GFP est de retour dans le noyau, et le marquage DAPI est intense dans les usines virales.

Cependant, une diffusion de la fluorescence des protéines nucléaires et nucléolaires marquées à la GFP a pu être observée dans les cellules infectées par les deux virus à partir de 30 minutes pi (cf. D Figure 3.10, B et F Figure S6 Annexe B).

Après 1 hpi, la fluorescence de la protéine nucléaire SUMO, ainsi que celle des deux méthyltransférases, n'est plus localisée au niveau du noyau (ou du nucléole) mais a diffusé dans le cytoplasme, jusqu'à le remplir complètement (cf. F Figure 3.10, C et G Figure S6 Annexe B). En revanche, le marquage DAPI reste localisé au niveau du noyau de la cellule, ce qui indique que l'ADN de l'hôte est toujours situé dans le noyau (BOYER *et al.*, 2009). Ce marquage est similaire à celui des cellules non-infectées, mais il est toutefois moins homogène.

Après 2 hpi, des foci fortement marqués au DAPI ont pu être observés dans le cytoplasme des cellules infectées : ils correspondent probablement à l'ADN contenu dans les nucléoïdes viraux (cf. B et E Figure 3.11). Au cours de cette étape, qui correspond au développement des usines virales précoces, ces foci semblent être également concentrés en fluorescence GFP, ce qui suggère que des protéines nucléaires sont recrutées à

l'emplacement de l'ADN viral (cf. C et F Figure 3.11). Ces observations suggèrent que même si l'intégrité globale du noyau cellulaire est préservée tout au long du cycle, les modifications qui y sont apportées au cours de l'étape précoce ont permis aux petites protéines nucléaires de diffuser librement dans le cytoplasme (GFP-SUMO), et aux protéines plus grandes (ml135-GFP ou ml216-GFP) d'être activement recrutées, et que ces protéines sont concentrées dans les usines virales en cours de développement.

Étonnamment, la fluorescence de la GFP a pu être de nouveau observée au niveau du noyau et du nucléole des cellules infectées entre 2 hpi (cf. I Figure 3.11) et 4 hpi (cf. H Figure 3.10, D et H Figure S6 Annexe B), suggérant que ce recrutement actif, ou cette diffusion passive, sont des phénomènes transitoires au cours du cycle infectieux. Le marquage DAPI, en revanche, devient de plus en plus fort au niveau des usines installées dans le cytoplasme des cellules, où la réplication de l'ADN viral se met en place (cf. H Figure 3.10, H Figure 3.11, D et H Figure S6 Annexe B). Après 2 hpi, lorsque la fluorescence GFP n'est pas observée dans le noyau, elle est observée uniquement dans les usines virales (cf. F et I Figure 3.11), en co-localisation avec le marquage DAPI (cf. E et H Figure 3.11), ce qui suggère un recrutement spécifique des protéines nucléaires nécessaires à la réplication de l'ADN viral dans les usines. Ces protéines peuvent remplir des fonctions essentielles, comme la transcription tardive des ARN messagers, ou encore la polyadénylation des transcrits viraux, les *Marseilleviridae* ne possédant pas de polyA polymérase qui soit codée par les virions. Une fois exprimée, la machinerie de transcription codée par les virions peut remplacer celle de l'hôte (à l'exception de la polyA polymérase).

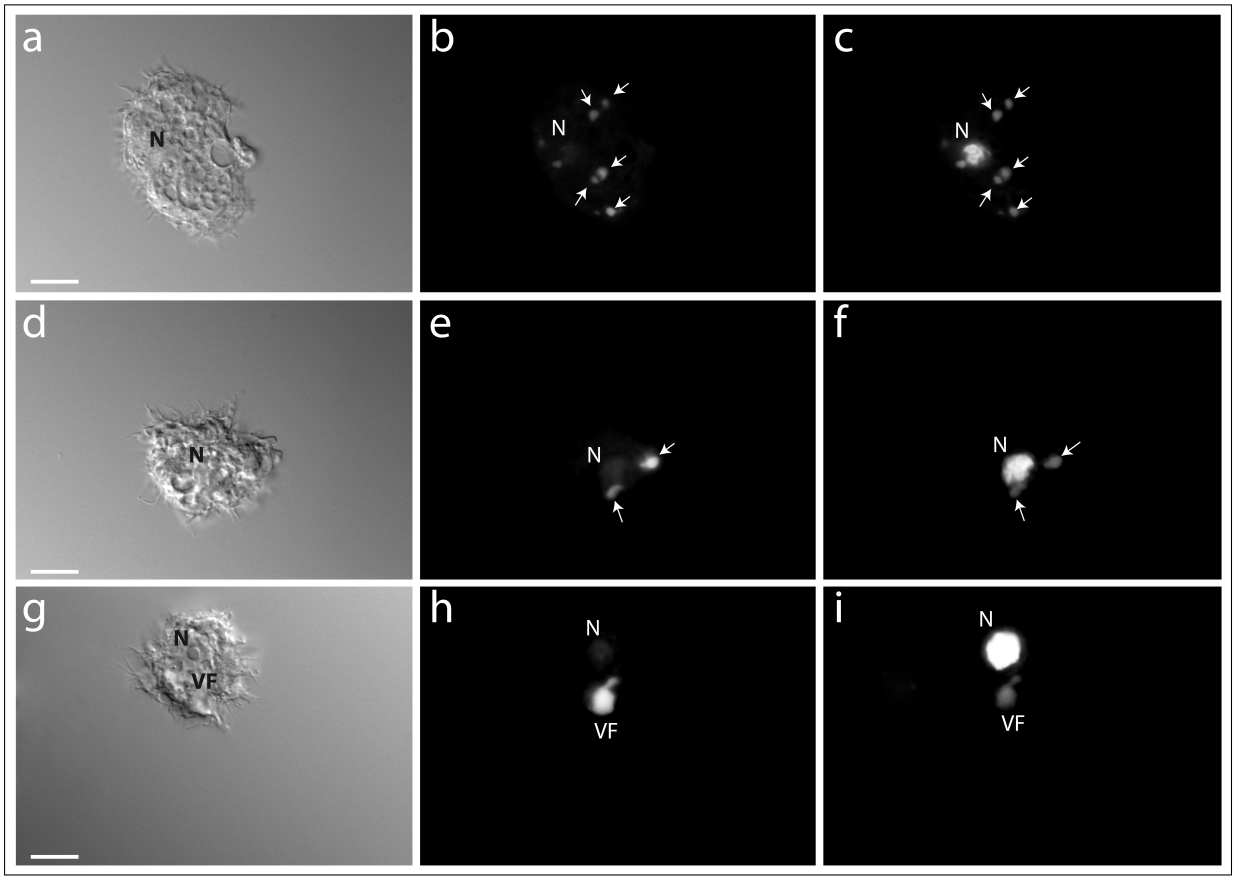


FIGURE 3.11 – Cellules infectées par Noumeavirus observées au microscope à épifluorescence (objectif x63) entre 2 et 3 hpi. A, D et G, les cellules sont observées par contraste interférentiel (DIC) ; B, E et H, observations au DAPI et C, F et I, observations de la GFP. Barres d'échelle : 10 μm. A, B et C, les cellules sont observées à 2 hpi ; D, E et F, à 2h30 pi, et G, H et I, à 3 hpi. Les nucléoïdes (B, flèches blanches) ainsi que les usines virales en cours de développement (E, flèches blanches et H, VF) marqués au DAPI produisent une forte fluorescence, qui masque le marquage DAPI du noyau (N). La fluorescence GFP est co-localisée avec les nucléoïdes viraux (C, flèches blanches) et les usines virales en développement (F, flèches blanches). Par contraste avec la forte fluorescence du noyau, les usines virales matures sont à peine marquées (I). Les mêmes phénotypes ont été observés dans les cellules infectées par Melbournevirus.

Étude du cycle infectieux de Noumeavirus au MET

En parallèle, le cycle infectieux d'*A. castellanii* infectées par Noumeavirus a été étudié par MET. Elles ont ensuite été fixées à différents temps, puis observées de deux façons : une partie des cellules a été marquée au DAPI puis observée au microscope optique à épifluorescence, et le reste a été inclus en résine avant d'être observé au MET (cf 2.3 et 2.5 Chapitre 2).

Lorsque l'on observe le déroulement de ce cycle infectieux au microscope optique, on peut remarquer qu'après 1 hpi, et contrairement au témoin (cf. A Figure 3.12), les cellules se décollent du fond de la flasque et n'adhèrent plus au support (cf. B Figure 3.12). Ces cellules décollées forment ensuite des agrégats visibles à l'œil nu, aux alentours de 2 hpi, qui s'agrandissent au cours du temps (cf. C Figure 3.12). Enfin, vers la

fin du cycle infectieux (après 7 hpi) environ 80% des cellules sont lysées : les cellules restantes en revanche, ont perdu leur forme de trophozoïte allongé et se sont arrondies (cf. D Figure 3.12).

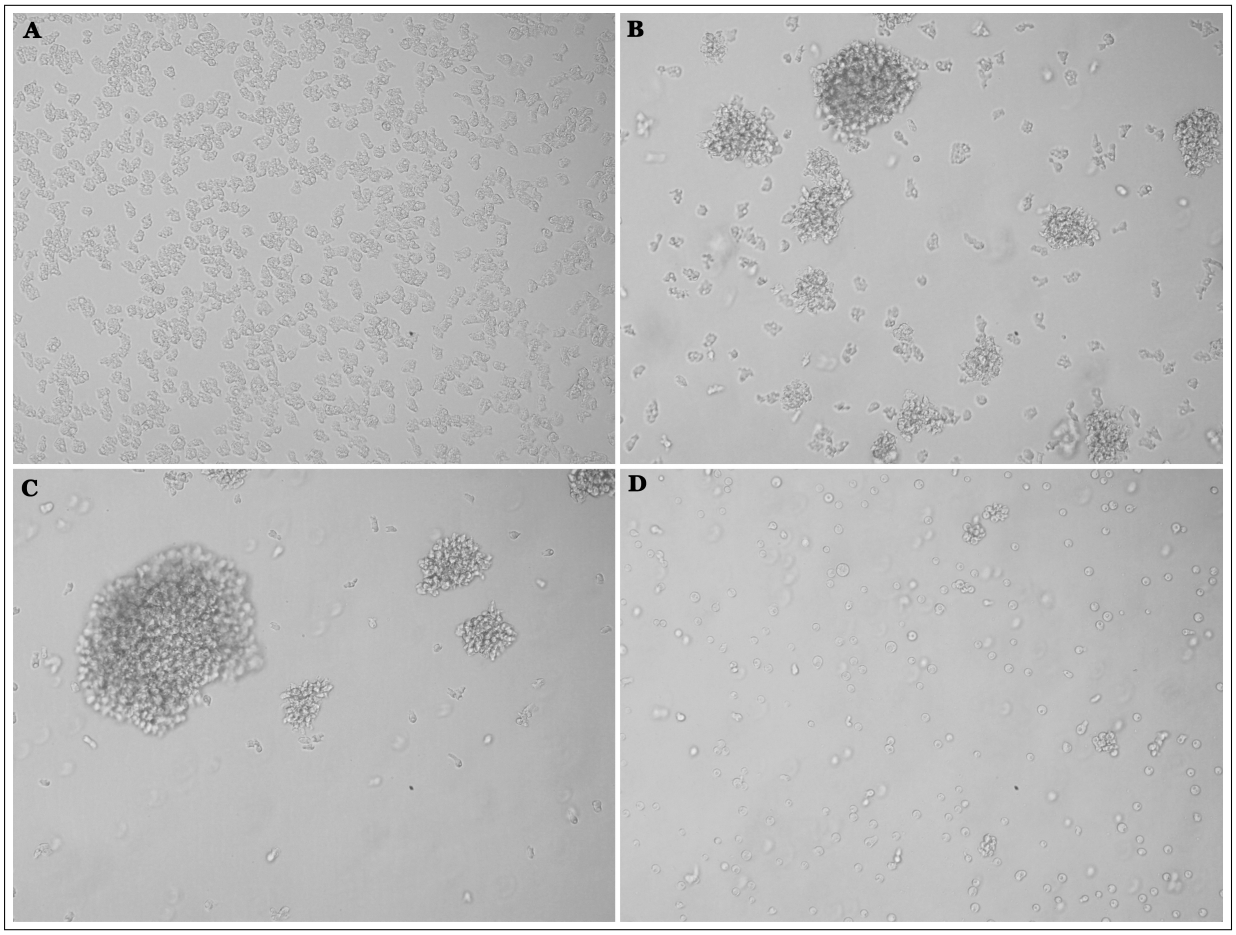


FIGURE 3.12 – Étapes du cycle infectieux de Noumeavirus observées au microscope optique (objectif x20). (A) Témoin : les trophozoïtes sont adhérents au support. (B) Après 1 hpi, les cellules se décollent et forment des amas. (C) A partir de 2 hpi, les amas deviennent plus gros, et sont visibles à l'œil nu. (D) Après 7 hpi, une partie des cellules est lysée ; celles qui restent sont arrondies.

Lorsque ce cycle infectieux est observé au MET, il est similaire à celui des autres virus de la famille des *Marseilleviridae*, et se déroule en plusieurs étapes :

i. **L'internalisation**
   A partir de 10 minutes pi, des particules de Noumeavirus peuvent être observées dans des vacuoles, situées dans le cytoplasme des cellules (cf. A Figure 3.13). Au même moment, des particules sont encore observées dans le milieu extracellulaire, et d'autres se retrouvent entourées par les pseudopodes de la cellule : ces particules sont en cours d'internalisation.

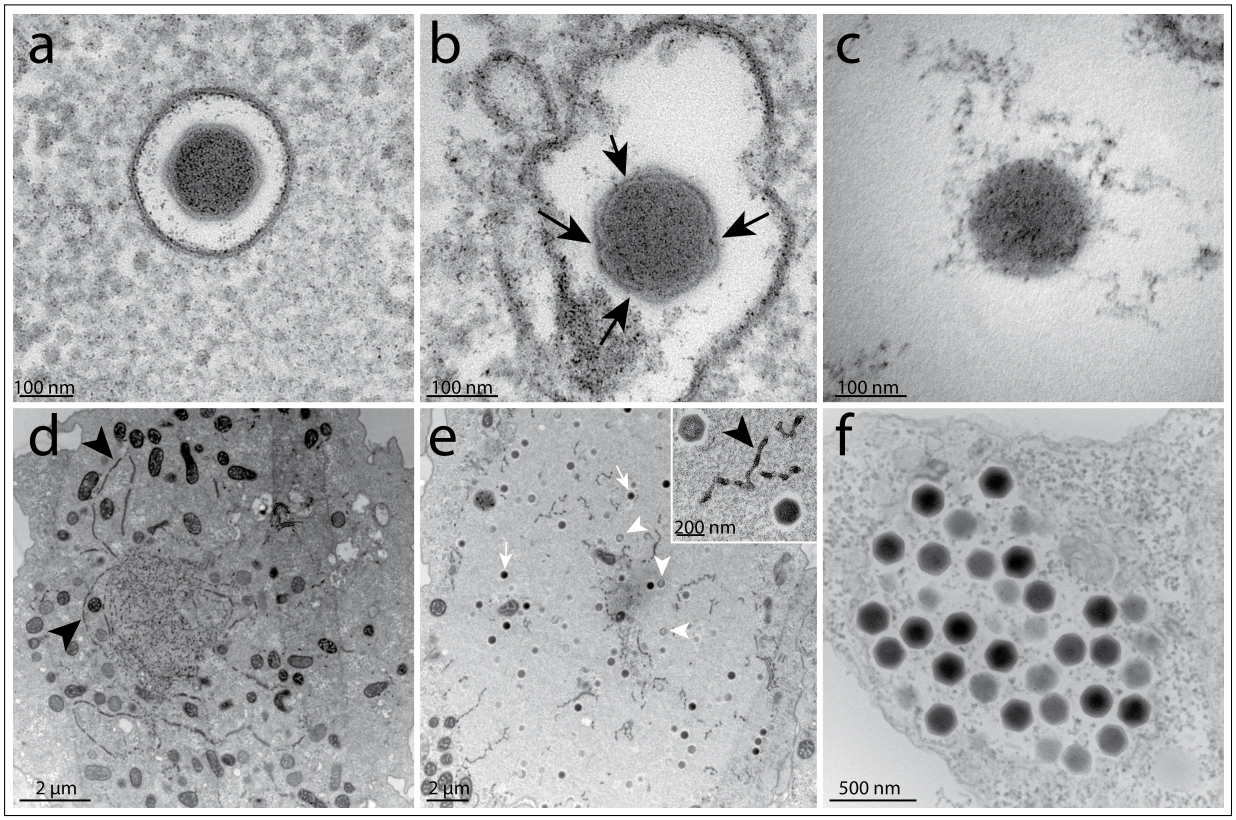


FIGURE 3.13 – Observation du cycle infectieux de Noumeavirus au MET. (A) Virion dans une vacuole après 10 minutes pi. (B) Après 20 minutes pi, les virions présentent des trous dans leur couche externe translucide (flèches) et commencent à devenir sphériques. (C) A partir de 30 minutes pi, les virions ont perdu leur couche externe et apparaissent comme des nucléoïdes denses aux électrons, qui peuvent être observés dans des vacuoles et dans le cytoplasme de la cellule. (D) Après 1 hpi, des structures tubulaires denses aux électrons sont observées dans le cytoplasme des cellules infectées (flèches). (E) A 4 hpi, les usines virales se mettent en place dans le cytoplasme des cellules à côté du noyau, et les organites sont repoussés en périphérie. Les virions en cours de synthèse sont assemblés : les virions matures denses aux électrons (flèches blanches), et les virions immatures (têtes de flèches), sont dispersés dans la même usine virale. L'encart montre des virions matures et immatures dans l'usine virale, en même temps que des structures tubulaires (flèche). (F) A partir de 7 hpi, les virions neo-synthétisés sont rassemblés dans des vacuoles dans le cytoplasme de la cellule, avant d'être libérés dans le milieu extra-cellulaire.

ii. **La libération du matériel génomique viral**

A partir de 20 minutes pi, certaines particules contenues dans les vacuoles, semblent perdre leur couche externe translucide : celle-ci paraît se dissoudre progressivement, et apparaît même trouée par endroits (cf. B Figure 3.13). Ces observations ont montré que certaines particules semblaient même s'arrondir aux alentours de 30 minutes pi (cf. C Figure 3.13).

Des nucléoïdes denses aux électrons, qui proviendraient de l'intérieur des particules virales, sont observés dans des vacuoles, qui ne semblent plus être complètement fermées (cf. C Figure 3.13). En effet, des composants cytoplasmiques, comme des ribosomes, peuvent être observés à l'intérieur de ces vacuoles. Ce phénomène

pourrait être impliqué dans la libération du matériel génomique viral dans le cytoplasme de la cellule. Les particules seraient progressivement dissoutes dans les vacuoles, et leur nucléoïde serait alors libéré dans le cytoplasme, après ouverture de ces vacuoles. Ces nucléoïdes, qui contiennent le génome viral, pourraient correspondre aux foci marqués au DAPI dans le cytoplasme des cellules aux alentours de 2 hpi (cf. B et E Figure 3.11), et qui semblent se regrouper pour former l'usine virale mature (cf. H Figure 3.10, H Figure 3.11, D et H Figure S6 annexe B). De plus, des sphères denses aux électrons, similaires à ces nucléoïdes, ont pu être observées en périphérie des noyaux des cellules hôtes (cf. C et E Figure 3.10).
Aucun événement de fusion n'a pu être observé entre la membrane de la vacuole et celle du virion, suggérant un mode d'infection (la pénétration du génome viral dans le cytoplasme de la cellule hôte) différent des autres virus, comme par exemple les *Mimiviridae*, qui réalisent cette étape par le biais de leur stargate (ZAUBERMAN *et al.*, 2008).

iii. **Les modifications apportées aux noyaux des cellules hôtes**
Entre 20 et 30 minutes pi, des changements drastiques de la morphologie du noyau et du nucléole de la cellule hôte sont observés (cf. C et D Figure 3.10). En effet, le noyau semble perdre progressivement son apparence sphérique, et sa membrane subit de nombreuses invaginations. De plus, le nucléole, habituellement dense aux électrons et bien organisé, perd en densité et se désorganise au cours du temps (cf. C, D, E et F Figure 3.10).
Après 1 hpi, des structures tubulaires, probablement membranaires, sont observées dans le cytoplasme des cellules (cf. D Figure 3.13), à côté de ces noyaux déformés (cf. E Figure 3.10).
Peu de temps après, le noyau cellulaire subit encore des réorganisations. Il semblerait en effet, que le noyau et le nucléole retrouvent leur intégrité : le noyau reprend une forme arrondie et le nucléole est de nouveau dense aux électrons (cf. G et H Figure 3.10).
A l'exception du noyau et des structures tubulaires, aucun autre phénomène n'a pu être observé dans la cellule hôte pendant ce temps.

iv. **Les usines virales et la synthèse de virions**
A partir de 3 hpi, les usines virales commencent à croître dans le cytoplasme des cellules jusqu'à la fin du cycle (cf. G et H Figure 3.10, E et H Figure 3.11, et E Figure 3.13). Ces usines peuvent parfois occuper jusqu'à un tiers de la cellule, et sont fortement marquées au DAPI, jusqu'à éclipser totalement le noyau (cf. H Figure 3.11).
De plus, aucune membrane visible ne délimite ces usines, qui ne semblent pas parfaitement structurées.

Malgré cette absence de limites bien définies, les organites de la cellule n'ont jamais été observés à l'intérieur de ces usines. Les mitochondries, le réticulum endoplasmique, l'appareil de Golgi, et les vacuoles sont repoussés en périphérie de ces usines virales, et procurent probablement l'énergie nécessaire à la synthèse des virions. Le noyau cellulaire, qui retrouve progressivement sa morphologie originelle, est lui aussi situé en périphérie de ces usines.

On retrouve dans les usines virales de Noumeavirus des virions à tous les stades de maturation. Des virions en cours de synthèse peuvent être observés à côté de virions matures dont le compartiment interne apparaît dense aux électrons (cf. E Figure 3.13). La synthèse de ces virions dans les usines virales est totalement différente de celle observée chez les *Mimiviridae*, où les virions matures bourgeonnent en périphérie des usines, et ne sont pas observés à l'intérieur (MUTSAFI *et al.*, 2013). Cette synthèse des virions de Noumeavirus débute par la capside : au cours de cette étape, les particules sont formées d'une couche translucide, qui possède une symétrie icosaédrique. L'intérieur de ces particules, également translucide, semble vide, et est ensuite rempli par les nucléoïdes denses aux électrons, ce qui est caractéristique des virions matures (cf. E Figure 3.13).

Les structures tubulaires observées précédemment dans les cellules infectées restent visibles, et sont situées dans les usines virales tout au long de la synthèse des virions (cf. D et E Figure 3.13).

v. **La libération des virions dans le milieu**

A partir de 7 hpi, de nombreux virions matures peuvent être observés dans de grandes vacuoles (cf. F Figure 3.13) localisées en bordure des usines virales, et proches de la périphérie des cellules, suggérant une libération de ces nouveaux virions par exocytose (THOMAS *et al.*, 2011 ; DOUTRE *et al.*, 2014 ; DORNAS *et al.*, 2016 ; ARANTES *et al.*, 2016). Toutefois, la production virale peut être assez conséquente, et les virions synthétisés peuvent remplir le cytoplasme des cellules.

La lyse de ces cellules va alors conduire à la libération de toutes les particules dans le milieu extra-cellulaire.

3.2 Discussion

Les *Marseilleviridae* sont des grands virus à ADN qui utilisent des virions à capside icosaédrique pour propager leurs génomes d'environ 300 kb, riches en A+T. Les génomes de ces virus entièrement séquencés, sont tous assemblés comme des séquences circulaires. Cependant, ces génomes pourraient également être linéaires, et auraient leurs extrémités redondantes et circulairement permutées. Cette caractéristique a également été observée chez les *Iridoviridae* (WILLIAMS *et al.*, 2005), qui ont une forte affinité phylogénétique avec les *Marseilleviridae*.

Ces virus codent tous pour leur propre machinerie de transcription, et par analogie avec les autres virus icosaédriques qui infectent les amibes *Acanthamoeba*, comme les *Mimiviridae*, il a été supposé que ces virus possèdent un cycle infectieux exclusivement cytoplasmique (THOMAS *et al.*, 2011 ; DOUTRE *et al.*, 2014). Les premières observations en microscopie électronique à transmission semblaient supporter cette idée : les noyaux de l'hôte apparaissent intacts jusqu'à la fin du cycle et co-existent avec les grandes usines virales dans le cytoplasme.

Des analyses comparatives des protéomes ont été réalisées sur Noumeavirus et Melbournevirus, deux virus isolés au laboratoire les plus éloignés dans la famille des *Marseilleviridae*. Noumeavirus a été isolé au cours de cette thèse, à partir d'échantillons environnementaux prélevés en Nouvelle-Calédonie, et Melbournevirus a été isolé en 2012, à partir de l'échantillon prélevé en Australie qui a notamment permis de découvrir *Pandoravirus dulcis*.

Cette étude a été réalisée afin de déterminer le niveau de conservation des compositions des particules, et d'obtenir quelques indications sur les processus infectieux caractéristiques de cette famille virale.

Les protéomes des deux virions sont similaires : deux tiers des protéines sont codées par les virions, et un tiers des protéines sont codées par l'hôte.

Un des résultats les plus étonnants, et important de surcroît, est l'absence d'une machinerie de transcription à l'intérieur des particules virales, suggérant que les *Marseilleviridae* dépendent du noyau de l'hôte, au moins pendant la phase précoce de leur cycle infectieux.

Ce résultat a donc été l'élément déclencheur d'une étude plus minutieuse du cycle infectieux, à la fois au MET et en fluorescence, en se focalisant notamment sur le noyau des cellules infectées, pendant les premières heures du cycle.

Des cellules produisant des protéines nucléaires et nucléolaires, couplées à la GFP, ont permis de déterminer qu'après 30 minutes pi, les protéines normalement confinées dans le noyau étaient exportées dans le cytoplasme des cellules, alors que l'ADN de l'hôte restait localisé au noyau.

Les observations de MET ont confirmé que cette perméabilisation coïncidait avec des changements morphologiques du noyau. En effet, les membranes nucléaires subissent de nombreuses invaginations, déformant complètement les noyaux, et les nucléoles, habituellement denses aux électrons, le sont de moins en moins.

Contrairement aux Pandoravirus qui ont un cycle infectieux nucléo-cytoplasmique, au cours duquel le noyau des cellules disparaît progressivement (PHILIPPE *et al.*, 2013), ici, les modifications semblent transitoires. En effet, le noyau subit de grands changements morphologiques au cours du cycle infectieux, mais ne disparaît pas. De plus, ces modifications semblent réversibles : après 2 hpi, la fluorescence GFP est de nouveau localisée au noyau, dont la forme et la structure semblent être revenues à la normale. Ces observations sont identiques pour Noumeavirus et Melbournevirus.

D'autre part, les caractéristiques du cycle infectieux de Noumeavirus, et des *Marseilleviridae* en général, sont étonnantes : aucun événement de fusion entre la membrane de la vacuole et celle du virus n'a jamais été observé, comme c'est le cas chez d'autres virus. Chez les *Mimiviridae* par exemple, la stargate s'ouvre et la membrane interne du virus fusionne avec celle de la vacuole, ce qui permet la libération du génome viral dans le cytoplasme de l'hôte (ZAUBERMAN *et al.*, 2008). Un phénomène similaire a été observé chez les *Pandoraviridae*, où le génome est délivré dans le cytoplasme par l'intermédiaire de leur pore apical (PHILIPPE *et al.*, 2013).

Nous en déduisons donc que les *Marseilleviridae* ont un mode d'infection différent des aux autres virus. Ici, il semblerait que la structure translucide qui entoure la particule soit dégradée, et que le nucléoïde, contenant le génome viral, soit libéré dans le cytoplasme après ouverture de la vacuole. Des structures évoquant des nucléoïdes ont pu être observées à des temps précoces du cycle, et semblent à l'origine des usines virales dans le cytoplasme des cellules hôtes.

Une autre caractéristique des *Marseilleviridae* est la présence de structures tubulaires dans les cellules infectées, en périphérie des noyaux déformés dans un premier temps, puis dans les usines virales en cours de production par la suite. Ces structures peuvent éventuellement correspondre à des membranes de la cellule qui seraient recyclées par le virus, afin de synthétiser de nouveaux virions, comme c'est le cas chez les *Mimiviridae* (MUTSAFI *et al.*, 2013), ou pour former des vacuoles d'exocytose.

Enfin, les virions matures sont regroupés à la fin du cycle dans de grandes vacuoles, en périphérie des cellules, suggérant une libération de ces virions par exocytose. Ces vacuoles remplies de virus ont pu être aperçues dans le milieu extra-cellulaire, et pourraient éventuellement pénétrer dans les cellules par phagocytose (ARANTES *et al.*, 2016). En revanche, les virions isolés, qui mesurent environ 200 nm de diamètre, semblent être trop petits pour être phagocytés, et seraient alors internalisés dans les cellules par un autre mécanisme d'endocytose, comme la macro-pinocytose par exemple.

Les résultats obtenus au cours de cette étude peuvent être interprétés de la façon suivante : les protéines normalement localisées et confinées dans le noyau des cellules hôtes sont activement recrutées au niveau des usines virales en cours de développement. Néanmoins, ce recrutement actif est transitoire, et se réalise uniquement au cours de la phase précoce du cycle infectieux. Parmi les protéines recrutées, il pourrait y avoir la machinerie de transcription, requise pour synthétiser les transcrits viraux précoces, qui est absente dans les virions.

Contrairement aux virus ayant un cycle infectieux nucléo-cytoplasmique, il n'y a aucune preuve que les *Marseilleviridae* transfèrent leur génome dans le noyau des cellules hôtes, ou même que ce noyau soit détérioré. En effet, le marquage DAPI des cellules infectées est toujours localisé au noyau, suggérant que l'ADN de l'hôte demeure dans le noyau tout au long du cycle infectieux (Boyer *et al.*, 2009).

De plus, aucun gène de Noumeavirus n'est interrompu par un intron épissable, contrairement aux virus qui possèdent une "vraie" phase intranucléaire, comme les *Pandoraviridae* (Philippe *et al.*, 2013) ou Mollivirus (Legendre *et al.*, 2015), ce qui est cohérent avec nos observations précédentes.

Le scénario que nous proposons à partir de ces résultats et de nos observations est le suivant : les gènes des *Marseilleviridae* sont initialement transcrits dans l'usine virale située dans le cytoplasme, à l'aide des ARN polymérases de l'hôte, avant que la machinerie codée par les virions ne prenne le relais plus tardivement au cours du cycle. La relation entre les *Marseilleviridae* et leur hôte correspondrait donc à un stade évolutif intermédiaire entre les virus nucléo-cytoplasmiques, comme les *Pandoraviridae* et les virus exclusivement cytoplasmiques, comme les *Mimiviridae*.

Ce scénario est supporté par le fait que les *Marseilleviridae* obéissent à la "*hairpin-rule*", similaire à celle des *Mimiviridae* (Byrne *et al.*, 2009) pour terminer leurs transcrits, ce qui est également le cas des *Iridoviridae* (Williams *et al.*, 2005) qui présentent une forte affinité phylogénétique avec les *Marseilleviridae*.

Ainsi, les *Marseilleviridae* embarquent dans leurs particules les protéines impliquées dans la maturation des transcrits, codées par les virions. Ces protéines sont alors libérées dans le cytoplasme des cellules en même temps que le génome viral, mais ne présentent aucun signal de localisation au niveau du noyau de l'hôte. Par conséquence, la transcription des gènes précoces des *Marseilleviridae* doit avoir lieu dans les usines virales, à l'aide de l'ARN polymérase codée par l'hôte et des protéines de maturation des transcrits codées par les virions. Étant donné que les virions ne codent pas pour une polyA polymérase, une partie de la maturation des transcrits doit être réalisée à l'aide de celle codée par l'hôte.

En se plaçant dans un modèle de réduction génomique, qui est le scénario évolutif que nous préférons pour les grands virus à ADN (Arslan *et al.*, 2011 ; Claverie *et al.*,

2013), nous nous attendions à l'existence d'une gamme de cycles de réplication intermédiaires entre les virus exclusivement cytoplasmiques et nucléo-cytoplasmiques. Ce scénario est le plus parcimonieux et a été favorisé par rapport à un scénario d'acquisition de gènes : la machinerie de transcription reposant sur un grand nombre de gènes, de multiples événements d'acquisition par HGT (ou transfert horizontal de gènes) auraient été nécessaires.

Suivant ce modèle, l'ancêtre viral hypothétique était capable de se répliquer indépendamment du noyau de l'hôte, codait pour ses propres machineries de transcription et de réplication de l'ADN, et devait libérer un complexe transcriptionnel fonctionnel (ou des ARNm) dans le cytoplasme pour initier l'infection.

A partir de cet ancêtre, différents degrés de perte de gènes auraient alors conduits à une variété de grands virus à ADN dépendants du noyau de l'hôte que l'on connaît aujourd'hui, comme par exemple ceux qui ne possèdent pas de machinerie de transcription codée par les virus (*Chloroviridae*, *Herpesviridae*).

Les *Marseilleviridae* seraient donc au commencement de leur voyage évolutif : ils codent toujours pour leur propre appareil de transcription, mais ont perdu la capacité de l'utiliser pour l'étape précoce de leur cycle infectieux. Les nombreuses GTPases, à la fois codées par les virus et l'hôte, et chargées dans les particules, auraient pu jouer un rôle clé pour permettre cette transition, notamment en modulant le trafic intracellulaire et en déclenchant l'échange de protéines entre le noyau de l'hôte et l'usine virale naissante.

Une telle étape de transition évolutive illustre bien la hiérarchie de la dépendance des virus à ADN vis-à-vis de leur hôte, à travers une perte de gènes compensée par l'établissement progressif d'une connexion entre le noyau et l'usine virale. L'étape évolutive suivante pourrait concerner la capacité des virus à transporter leur génome à ADN dans le noyau. Cette étape serait ensuite suivie par une perte complète de la machinerie de transcription codée par les virions, ainsi que par l'acquisition progressive d'introns, ce qui permettrait d'intégrer totalement les ARNm dérivés des virus dans la machinerie d'export des transcrits du noyau de la cellule hôte.

La capacité des usines virales des *Marseilleviridae* à mimer le noyau de façon transitoire, et à détourner les protéines nucléaires essentielles à la transcription précoce des gènes, constitue un argument en faveur de la théorie d'une "eucaryogénèse virale", une hypothèse qui propose qu'un grand virus à ADN aurait pu être à l'origine du noyau (Bell, 2001 ; Claverie, 2006 ; Takemura, 2001 ; Forterre *et al.*, 2016). En effet, les usines virales ainsi que le noyau ont en commun le fait de confiner la réplication de l'ADN et la transcription dans des compartiments, où les organites qui produisent de l'énergie sont exclus (comme les mitochondries par exemple). Ils pourraient donc être considérés comme des "parasites" du cytoplasme de la cellule, dont ils dépendent pour synthétiser les protéines qu'ils codent.

Les caractéristiques uniques du cycle de réplication des *Marseilleviridae* d'aujourd'hui dans des *Acanthamoeba* modernes, pourraient donc constituer un témoignage de la contribution des virus à ADN à l'émergence du noyau dans des lignées cellulaires ancestrales.

3.3 Conclusion

Cette analyse comparative du protéome des particules purifiées de Noumeavirus et de Melbournevirus nous a conduit à deux conclusions majeures.

Premièrement, et contrairement aux virus à ADN qui possèdent une capside à symétrie icosaédrique (ARSLAN *et al.*, 2011 ; DUNIGAN *et al.*, 2012 ; FISCHER *et al.*, 2014 ; W. LI *et al.*, 2014), la protéine majeure de capside (ou "*Major Capsid Protein*") n'est pas la protéine la plus abondante dans les protéomes des virus étudiés ici. En effet, la protéine la plus abondante dans ces protéomes est composée de 150 acides aminés, et est fortement conservée chez tous les *Marseilleviridae*, avec un pourcentage d'identité supérieur à 77%. Cette protéine sera l'objet d'études, qui permettront notamment de déterminer sa localisation dans les virions, ainsi que son rôle probable au cours du cycle infectieux.

Enfin, depuis la découverte du premier *Marseilleviridae* en 2009 (BOYER *et al.*, 2009), et des autres virus de cette famille les années suivantes, nous pensions que leur cycle infectieux était exclusivement cytoplasmique. Or, l'analyse des protéomes montre qu'il n'y a pas de machinerie de transcription qui soit embarquée dans les virions, qui est pourtant nécessaire à l'initiation du cycle infectieux sans l'aide des fonctions nucléaires de l'hôte.

Cette surprenante découverte nous indique qu'il est impossible de déterminer le type de cycle infectieux d'un virus en se basant uniquement sur des données de génomique et de transcriptomique. Une analyse du protéome des particules, comme celle que nous avons réalisée ici, pourrait donc faciliter la distinction entre des virus qui ont un cycle nucléo-cytoplasmique et ceux dont le cycle est exclusivement cytoplasmique.

L'observation minutieuse du cycle infectieux au microscope électronique à transmission, s'est également révélée utile pour découvrir des indices nous permettant de déterminer le type de cycle de Noumeavirus. Ces observations nous ont notamment permis de nous focaliser sur le noyau des cellules infectées, qui subissent de nombreux changements morphologiques au cours de ce cycle.

De nombreuses questions ont été soulevées lors de ces observations, qui restent sans réponse claire pour l'instant, et qui mériteraient d'être étudiées. Par exemple, quelle est la véritable nature des structures tubulaires apparaissant dans les usines virales au cours de la synthèse des virions ? Quel est le véritable mécanisme d'internalisation des virions dans les cellules (phagocytose, macro-pinocytose ou endocytose médiée par des récepteurs) ? Comment le génome viral est-il libéré dans le cytoplasme de l'hôte ? Quels sont les mécanismes impliqués dans la dissolution de la capside virale ?

L'utilisation de protéines fluorescentes ciblant le noyau de la cellule hôte a été un outil complémentaire pour déterminer le type de cycle de ces *Marseilleviridae*, et a révélé la présence d'un niveau supplémentaire dans la dépendance des virus vis-à-vis de

leur hôte. Ce niveau constitue un intermédiaire entre les grands virus à ADN qui passent par une vraie phase intranucléaire, comme les *Herpesviridae*, les virus qui détériorent complètement le noyau, comme les *Pandoraviridae*, et les virus qui se répliquent entièrement dans le cytoplasme des cellules, comme les *Mimiviridae*.

Les *Marseilleviridae* nous renseignent donc sur un nouveau mode d'interaction entre les virus et leur hôte, où les usines virales qui se développent dans le cytoplasme des cellules infectées miment certains aspects du noyau, afin de détourner les protéines nucléaires nécessaires à l'initiation de la transcription du génome viral.

4 Caractérisation de nouveaux *Pandoraviridae* et études comparatives

Cette partie est consacrée à un nouveau Pandoravirus que j'ai isolé au cours de cette thèse, à partir d'un échantillon environnemental qui a été prélevé en Nouvelle-Calédonie : ce virus qui infecte les amibes, a été nommé *Pandoravirus neocaledonia*.

Deux autres Pandoravirus ont également été isolés au laboratoire, à partir d'échantillons de terre qui ont été prélevés à Marseille.

J'ai caractérisé ces virus à l'aide de plusieurs approches : génomique, protéomique, transcriptomique et étude de leur cycle infectieux au microscope électronique à transmission. Ces analyses comparatives auxquelles j'ai contribué, ont été réalisées en collaboration avec plusieurs membres du laboratoire, et sont toujours en cours.

4.1 Résultats

4.1.1 Découverte d'un nouveau membre de la famille des *Pandoraviridae*

Un échantillon environnemental d'aspect boueux, prélevé dans un étang à Nouméa, en Nouvelle-Calédonie, a été dilué dans du milieu riz, puis a été incubé à l'obscurité. Après incubation, cet échantillon a été traité avec des antibiotiques et un antifongique, puis a été placé en co-culture avec des amibes *Acanthamoeba castellanii* et *Dictyostelium discoideum* (DH1 et HMX44A). Les cellules ont été observées quotidiennement au microscope optique, le contenu de chaque puits a été récupéré, et ensemencé dans des puits contenant des amibes ainsi que du milieu frais, complété d'antibiotiques et d'antifongique (cf. 2.1 Chapitre 2).

Une lyse des amibes *A. castellanii*, ainsi que la présence de nombreuses particules dans le milieu extra-cellulaire, ont pu être observées après quelques jours. Le parasite responsable a été observé au microscope optique, et sa particule, de l'ordre du micromètre, apparaît ovoïde (cf. Figure 4.1, image de gauche). Cette infection a été amplifiée avec des amibes et du milieu frais, puis les cellules infectées ont été fixées et inclues en résine Epon, avant d'être observées au microscope électronique à transmission (MET).

Ce parasite, qui possède une morphologie typique des Pandoravirus (PHILIPPE *et al.*, 2013), a été nommé *Pandoravirus neocaledonia*. Le virion est ovoïde et mesure environ 1 μm de long sur 500 nm de diamètre. Il possède un pore situé à l'apex de la particule,

large d'environ 120 nm, et scellé par un bouchon diffus. Cette particule est constituée d'une enveloppe, d'environ 70 nm d'épaisseur, qui est composée de plusieurs couches : une couche externe de densité moyenne (environ 25 nm), une couche intermédiaire, qui est dense aux électrons et qui est constituée d'un réseau de fibres qui forment des structures lamellaires (environ 25 nm), et d'une couche interne, qui est peu dense (environ 20 nm). Enfin, l'intérieur de la particule, qui est délimité par une membrane lipidique, est peu dense aux électrons : il apparaît homogène et assez diffus, sans aucune structure visible (cf. Figure 4.1, image de droite). Toutefois, selon certains plans de coupe, une sphère d'environ 50 nm de diamètre très dense aux électrons, peut être observée (PHILIPPE *et al.*, 2013).

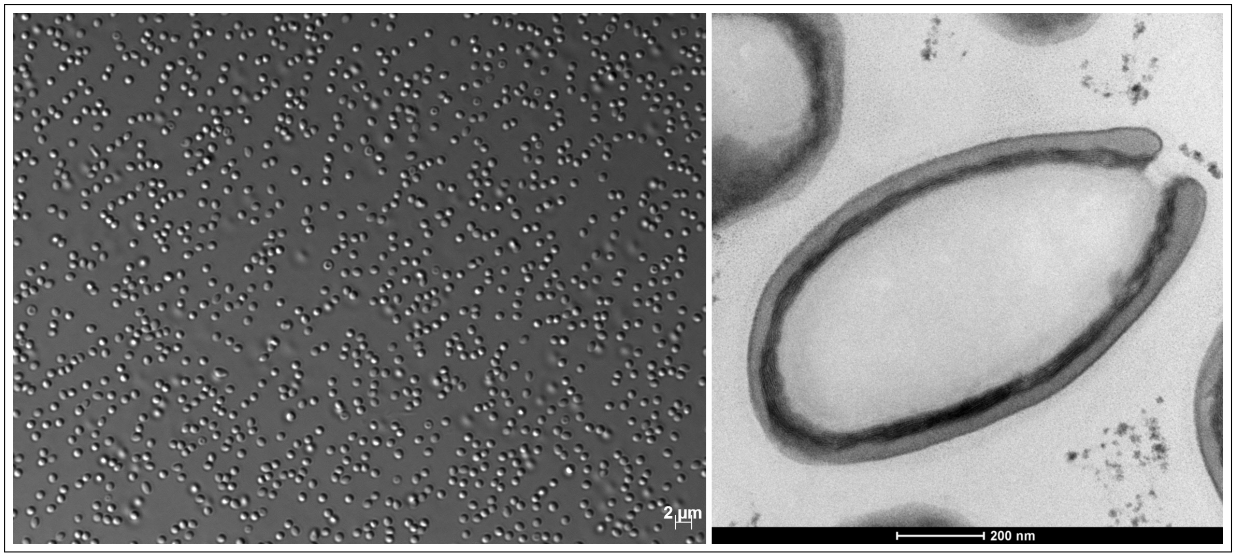


FIGURE 4.1 – Particules de *Pandoravirus neocaledonia* observées au microscope optique et électronique. Image de gauche, particules observées au microscope optique (objectif x63 avec Optovar x1,6). Image de droite, particule observée au MET.

En revanche, après de nombreux repiquages successifs, aucun phénotype infectieux ou lyse n'a pu être observé sur les souches d'amibes *D. discoideum* testées. Une fois *P. neocaledonia* purifié, des tests d'infection ont de nouveau été réalisés sur ces amibes, et sont restés sans succès.

Deux autres virus, possédant des caractéristiques morphologiques identiques ainsi qu'un phénotype infectieux similaire, ont été isolés à partir d'échantillons de terre, qui ont été prélevés à Marseille. Ils ont été nommés *Pandoravirus quercus* et *Pandoravirus celtis*. Ces virus ont également été purifiés, puis clonés, avant d'être caractérisés et comparés aux autres membres de cette famille.

Ces virus ont été produits en grande quantité sur des cultures fraîches d'*A. castellanii*, puis purifiés sur gradient de sucrose (cf. 2.2 Chapitre 2). A l'issue de la centrifugation des gradients, un disque blanchâtre contenant les virions a été observé dans la couche

de sucrose correspondant à 60%. Les virions ont ensuite été récupérés, lavés et contrôlés au microscope optique (objectif x63 avec Optovar x1,6, cf. Figure 4.1, image de gauche).

Clonage des Pandoravirus sur cellule unique

Ces Pandoravirus ont ensuite été clonés sur cellule unique : des amibes *A. castellanii* ont été infectées avec du virus purifié (non-titré) dans une plaque multipuits. Après 1 heure d'incubation et 50 lavages, les cellules ont été diluées puis réparties dans les autres puits à raison de 1 cellule par puits, et les puits ne contenant pas de cellules ont servis de témoins. Des cellules fraîches ont été ajoutées dans tous les puits et le phénotype a été contrôlé au microscope optique quotidiennement (cf. 2.2 Chapitre 2). Après 5 jours d'incubation, toutes les cellules étaient lysées ou présentaient un phénotype infectieux, à l'exception des puits témoins.

Des clones ont été sélectionnés, purifiés, puis comptés à l'aide d'une cellule de Malassez avant d'être titrés par dilutions sériées selon la méthode de Reed-Muench (Reed *et al.*, 1938) (cf. 2.2 Chapitre 2). Le comptage de ces clones indique un ordre de 10^{10} particules par mL, et la titration par dilutions limites indique un $TCID_{50}$/mL d'un ordre de 10^{8} particules par mL.

Ces clones ont ensuite été lysés, afin d'extraire leurs protéines et d'établir leur profil protéique (cf. 2.7 Chapitre 2).

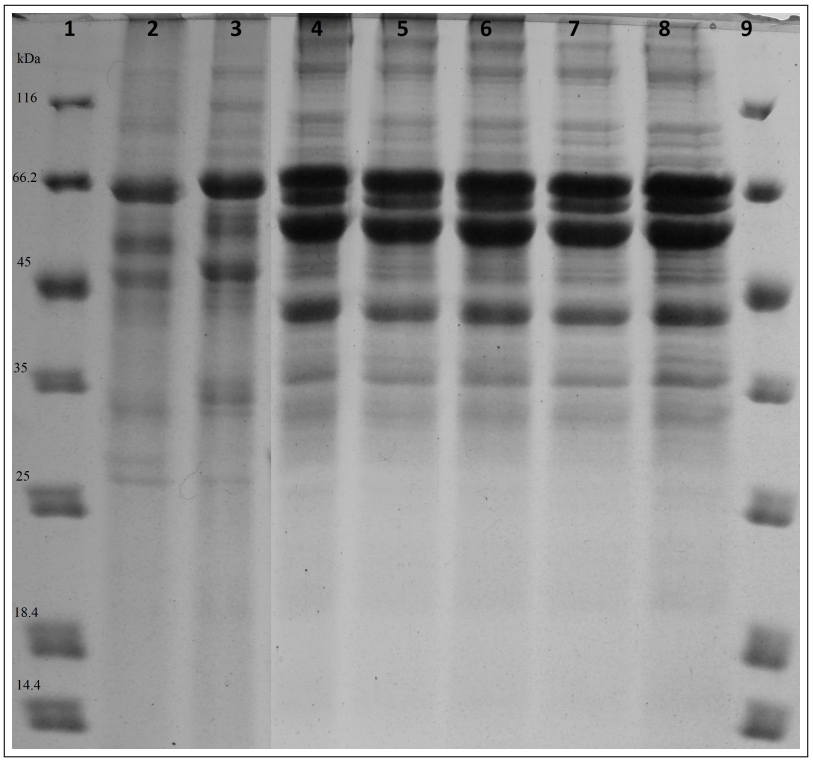


Figure 4.2 – Profils protéiques des clones de *P. neocaledonia*. Les pistes 1 et 9 correspondent au marqueur de poids moléculaire, la piste 2 au profil de *P. dulcis*, la piste 3 à celui de *P. salinus*, et les pistes 4 à 8 aux profils de *P. neocaledonia* et de ses clones.

L'observation des gels indique que les profils des clones apparaissent strictement identiques entre eux (cf. Figure 4.2, pistes 4 à 8), et qu'ils sont différents de ceux de *P. dulcis* et de *P. salinus* (cf. Figure 4.2), les premiers Pandoravirus qui ont été découverts au laboratoire (PHILIPPE *et al.*, 2013). Une protéine majoritaire de ces clones (environ 66 kDa) semble être similaire à celle des deux autres Pandoravirus.

Un clone de chaque Pandoravirus a donc été sélectionné pour les étapes de caractérisation et les études comparatives suivantes.

4.1.2 Analyse génomique des Pandoravirus

Extraction des génomes

L'ADN de chacun de ces trois Pandoravirus a été extrait et purifié (cf. 2.6 Chapitre 2). Cette méthode d'extraction n'a toutefois pas permis de récupérer le génome de *P. neocaledonia* avec une quantité et un taux de pureté qui soient satisfaisants pour le séquençage à haut-débit. Différents protocoles, ainsi que différentes quantités de virus ont été testés : la technique qui a été utilisée est une extraction à l'aide de CTAB (Cetyltrimethylammonium bromide), à partir de 2 x 10^{10} particules virales.

Environ 15 μg d'ADN de *P. neocaledonia* purifié a été séquencé à l'aide de deux méthodes différentes : la première technique a permis de générer des séquences "*paired-end*" [1] de 2 x 100 paires de bases, en utilisant la plate-forme de séquençage à haut-débit HiSeq. La deuxième technique de séquençage à haut-débit qui a été utilisée, est le séquençage de molécule unique en temps réel, ou PacBio, qui permet d'obtenir des séquences supérieures à 4 000 paires de bases.

Le génome à ADN double-brin de *P. neocaledonia* a été assemblé à partir des données de séquençage PacBio en 3 contigs [2] de 2 Mb, 13 kb et 4 kb. Ces contigs ont ensuite été assemblés en un seul contig de 2 003 191 paires de bases. Enfin, cet assemblage a été vérifié et validé à l'aide des séquences qui ont été obtenues à partir du séquençage HiSeq.

En parallèle, 5 μg d'ADN purifié de *P. celtis* et de *P. quercus* ont été séquencés selon la technique de séquençage à haut-débit PacBio. Concernant le génome de *P. quercus*, 72 468 séquences d'une longueur moyenne de 14 612 nucléotides ont été générées, et l'assemblage a produit un seul contig de 2 077 388 nucléotides.

En revanche, à cause d'un problème lors de l'assemblage du génome de *P. celtis*, celui-ci n'a pas été pris en compte pour la suite des analyses, qui ont donc été recentrées sur *P. neocaledonia* et *P. quercus* (qui possède 97% d'identité avec *P. celtis*), ainsi que sur

1. Les séquences "*paired-end*" sont obtenues par le séquençage des deux extrémités de fragments courts, d'une taille inférieure à 1 kb.

2. Un contig est une séquence obtenue par chevauchement de séquences courtes (ou "*reads*") qui ont été générées par le séquenceur.

les Pandoravirus qui ont déjà été publiés : *P. dulcis* et *P. salinus* (PHILIPPE *et al.*, 2013), et *P. inopinatum* (SCHEID, 2016).

Ces génomes ont ensuite été étudiés à l'aide de plusieurs approches, qui sont toujours en cours d'analyse, afin d'évaluer leur complexité : sont-ils plus complexes que les génomes de *P. salinus* et *P. dulcis* ? Ces génomes sont-ils conservés entre tous les Pandoravirus ? Peut-on déterminer le "*core-genome*" des Pandoravirus ?

Analyses des génomes

Pandoravirus neocaledonia et *P. quercus* possèdent un génome à ADN double-brin de 2 Mb, avec un contenu en G+C de 61% : au total, 1 794 (*P. neocaledonia*) et 1 830 (*P. quercus*) ORFs [3] ont été prédits comme codant pour des protéines. A titre de comparaison, *P. salinus* et *P. dulcis* possèdent un génome à ADN double-brin de 2,4 Mb et 1,9 Mb, avec un contenu en G+C de 61,7 et 63,7%, respectivement. Un total de 2 556 (*P. salinus*) et 1 502 (*P. dulcis*) ORFs codant pour des protéines ont été prédits (PHILIPPE *et al.*, 2013). Les codons Stop étant riches en A+T, ils sont donc moins fréquents dans ces génomes riches en G+C, ce qui pourrait expliquer le nombre de gènes prédits.

Les génomes de *P. salinus* et de *P. dulcis* possèdent des blocs de séquences répétées à une de leurs extrémités, d'environ 50 kb pour *P. salinus*, et 20 kb pour *P. dulcis*. En revanche, celui de *P. neocaledonia* en contient deux : chaque bloc est différent, mesure environ 5 kb, et est situé à chacune des extrémités de ce génome. Enfin, aucun bloc de séquences répétées n'a pu être identifié dans le génome de *P. quercus*.

En parallèle, des analyses de transcriptome ont été réalisées : des amibes *A. castellanii* infectées par les différents Pandoravirus ont été prélevées toutes les heures au cours des cycles infectieux. Chaque prélèvement a été analysé par trois approches différentes : les cellules infectées ont été observées au MET, les protéines de ces cellules infectées ont été envoyées en protéomique afin d'analyser les protéomes cinétiques des cellules infectées, et les librairies d'ADN complémentaires (ADNc) ont été produites à partir des ARN messagers avant d'être séquencées (cf. 2.8 Chapitre 2). Enfin, une étude de protéomique a été réalisée sur les particules purifiées de chacun des Pandoravirus (cf. 2.7 Chapitre 2).

Les données de génomique ont par la suite été ré-annotées à l'aide d'une approche combinée utilisant les données de conservation entre les différents Pandoravirus, les transcriptomes orientés, et les données de protéomiques obtenues sur les particules purifiées. Nous avons dans un premier temps analysé le niveau d'expression des gènes au cours de l'infection, pour lesquels les protéines ont été identifiées dans les protéomes des particules. Ces niveaux nous ont servi à établir un seuil, afin de rechercher des gènes

3. Un ORF ou "*Open Reading Frame*", correspond à un cadre de lecture ouvert, qui est délimité par deux codons Stop et qui peut coder pour une protéine.

potentiels dont les niveaux de transcription sont équivalents ou supérieurs à ce seuil. Nous avons étudié en parallèle les niveaux de transcription des zones intergéniques, où nous avons identifié des gènes potentiels (codants ou non-codants) satisfaisant ce niveau minimal de transcription. Enfin, ces approches ont été combinées aux données de conservation des gènes, ce qui nous a permis de réduire le nombre de gènes que nous avions précédemment prédit. En effet, 1 794 et 1 830 ORFs avaient été prédits comme codant pour des protéines dans les génomes de *P. neocaledonia* et *P. quercus*, respectivement. Après ré-annotation de ces génomes, nous n'en avons retenu que 1 081 (*P. neocaledonia*) et 1 185 (*P. quercus*) comme codant pour des protéines (cf. Figure 4.3).

Les génomes de *P. salinus* et *P. dulcis* ont également été ré-annotés, en suivant cette approche. Un total de 2 556 et 1 487 gènes codants avait été prédit pour *P. salinus* et *P. dulcis*, respectivement (PHILIPPE *et al.*, 2013). Après analyse des génomes, 1 430 (*P. salinus*) et 1 071 (*P. dulcis*) ont été retenus comme codants, soit une réduction de 44 et 28% du nombre de gènes prédits (cf. Figure 4.3).

Ces analyses nous ont également permis de déterminer que 12 à 20% des gènes des Pandoravirus ne sont pas codants (cf. Figure 4.3). Ces transcrits non-codants ont pu être identifiés grâce à la présence d'un fort signal de transcription sans ORF correspondant qui soit significatif[4]. De plus, des ARN de transfert (ARNt), qui sont des gènes liés à la traduction, ont été retrouvés chez tous les Pandoravirus (cf. Figure 4.3) : par exemple, *P. salinus* possède 3 ARNt ($ARNt^{Pro}$, $ARNt^{Met}$ et $ARNt^{Trp}$, PHILIPPE *et al.*, 2013).

FIGURE 4.3 – Nombre de gènes présents dans les génomes des Pandoravirus

Gènes	*P. neocaledonia*	*P. quercus*	*P. salinus*	*P. dulcis*
Total	1332	1343	1648	1346
Codants	1081	1185	1430	1071
ARNt	3	1	3	1
Non-codants	248	157	215	274

Les régions qui codent pour des protéines représentent entre 61 et 68% des génomes des Pandoravirus, et sont séparées par de longues régions intergéniques de 650 nucléotides de long en moyenne, qui correspondraient à des UTRs[5] plus longs que ceux classiquement rencontrés dans les autres virus géants.

Enfin, la présence d'introns qui avait été assez inattendue dans les gènes de *P. salinus* (PHILIPPE *et al.*, 2013), a été recherchée dans les gènes de tous les Pandoravirus. En moyenne, 7% des gènes de *P. salinus* (80), *P. dulcis* (92), *P. neocaledonia* (82), et *P. quercus* (84) contiennent au moins un intron dans leur partie codante. Ces introns sont

4. Un ORF qui n'est pas significatif signifie qu'il est trop petit ou qu'il possède des codons Stop dans toutes les phases de lecture.

5. Les UTRs ("*Untranslated Transcribed Region*") sont des régions retrouvées aux extrémités 5' et 3' des ARN messagers : elles sont transcrites mais ne sont pas traduites en protéines. Ces régions sont notamment impliquées dans la régulation de l'expression des gènes.

tous des introns épissables, et sont délimités par les séquences 5'-GT et 3'-AG : ils doivent donc passer par le splicéosome de la cellule hôte pour pouvoir être excisés.

Récemment, un nouveau Pandoravirus a été isolé au laboratoire, à partir d'un échantillon qui a été prélevé en Australie. Son nom provisoire est *Pandoravirus D1*, il est morphologiquement identique aux autres Pandoravirus, et son génome a été séquencé.

Des analyses phylogénétiques ont donc été réalisées à partir de la concaténation des régions codantes orthologues entre tous les Pandoravirus (cf. Figure 4.4) : *P. salinus* et *P. dulcis* (PHILIPPE *et al.*, 2013), *P. neocaledonia*, *P. quercus*, *P. D1* et *P. inopinatum* (SCHEID, 2016). Ces analyses ont permis de montrer que *P. neocaledonia*, *P. quercus* et *P. D1* appartiennent bien à la famille des Pandoravirus, et que *P. quercus* semble être plus proche de *P. inopinatum*. En revanche, *P. neocaledonia* et *P. D1*, qui sont proches (76% d'identité) ne s'intègrent pas dans la lignée formée par les 4 autres Pandoravirus, et semblent fonder une nouvelle lignée au sein de cette famille (cf. Figure 4.4).

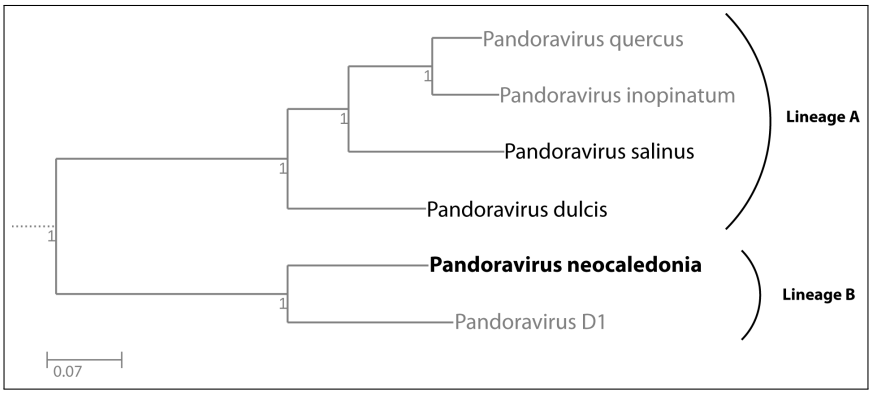


FIGURE 4.4 – Arbre phylogénétique de la famille des *Pandoraviridae*, qui se divise en deux lignées. Les noms marqués en noir correspondent aux deux premiers Pandoravirus et celui marqué en gras correspond à *P. neocaledonia*, le membre fondateur de la lignée B. L'arbre a été réalisé à partir des régions nucléotidiques codantes orthologues entre les virus, qui ont été alignées et concaténées. L'enracinement de l'arbre a été réalisé à partir des séquences orthologues de *Mollivirus sibericum*, le NCLDV le plus proche des Pandoravirus.

Ces analyses nous ont également permis de déterminer qu'environ 70% des ORFs prédits chez les Pandoravirus n'avaient pas d'homologues dans les bases de données (ORFans, cf. Figure 4.5). Parmi les 30% d'ORFs dont les homologues ont été identifiés, la majorité présente un homologue eucaryote, et 2 à 3% sont retrouvés chez *A. castellanii* (cf. Figure 4.5). Ces proportions semblent être similaires pour tous les Pandoravirus étudiés (cf. Figure 4.5), et suggèrent que les événements de transferts de gènes (ou HGT) entre ces virus et les amibes *A. castellanii* sont peu nombreux.

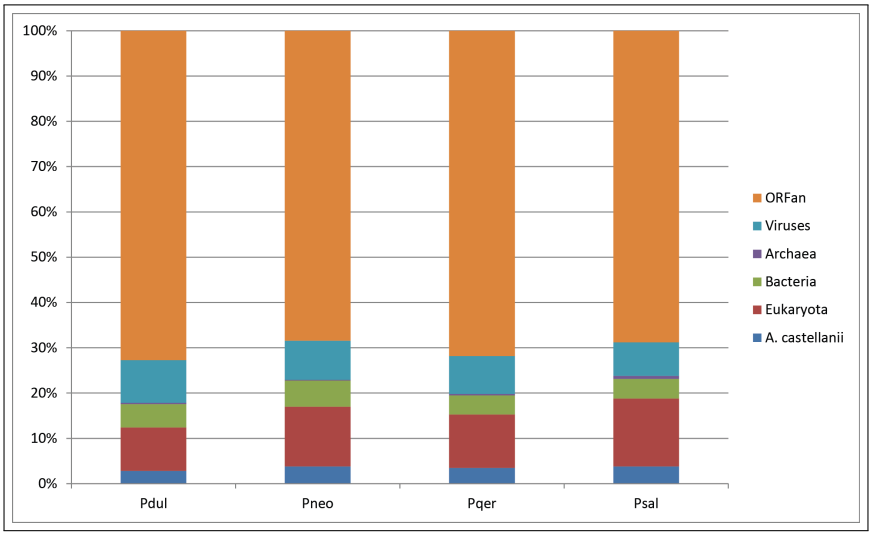


FIGURE 4.5 – Histogramme des différentes catégories d'homologues des Pandoravirus. La recherche d'homologues a été réalisée à l'aide de Blast contre la base de données NR (NCBI RESOURCE COORDINATORS, 2016) pour tous les ORFs des Pandoravirus. En orange, les ORFs qui n'ont pas d'homologues dans les bases de données (ORFans) ; en bleu clair, les homologues retrouvés chez les virus ; en violet, ceux retrouvés chez les archées ; en vert, ceux retrouvés chez les bactéries ; en rouge, ceux retrouvés chez les eucaryotes ; et en bleu foncé, les homologues qui ont été retrouvés chez *A. castellanii*.

Suite à l'analyse de ces gènes, nous nous sommes intéressés au "*core-genome*" de ces virus : quels sont les gènes indispensables aux Pandoravirus ? Ce nombre de gènes sera t-il réduit après la découverte de nouveaux membres de la famille des *Pandoraviridae* ?

Détermination du *core-genome* des Pandoravirus

Nous avons donc commencé à identifier tous les gènes qui sont partagés et indispensables aux Pandoravirus et ceux qui sont spécifiques à chaque virus. Pour cela, des "clusters" de gènes ont été établis : chaque cluster regroupe des gènes qui partagent une séquence commune (et qui ont éventuellement une fonction commune). Les clusters qui contiennent un orthologue partagé par chaque virus sont ajoutés au *core-genome*, et ceux qui sont uniques à chaque virus constituent le "génome accessoire". Cette analyse a été réalisée à partir des génomes des 6 Pandoravirus dont nous disposons : les 5 qui ont été isolés au laboratoire, et *P. inopinatum* (SCHEID, 2016).

Cette analyse nous a permis de déterminer que le nombre de clusters composant le *core-genome* diminue jusqu'à atteindre un plateau : à partir du 8^e génome de Pandoravirus, ce nombre de gènes communs ne devrait plus évoluer (cf. Figure 4.6). Ces *core genes* représentent environ 70% des gènes des Pandoravirus, et nous nous attendons à retrouver des gènes qui sont notamment nécessaires à l'infection et d'autres qui sont impliqués dans la morphologie des particules, qui sont identiques.

A l'inverse, lors de l'ajout de nouveaux génomes, le nombre de clusters constituant le *pan-genome* augmentera (cf. Figure 4.6). Chaque nouveau génome apportera un certain

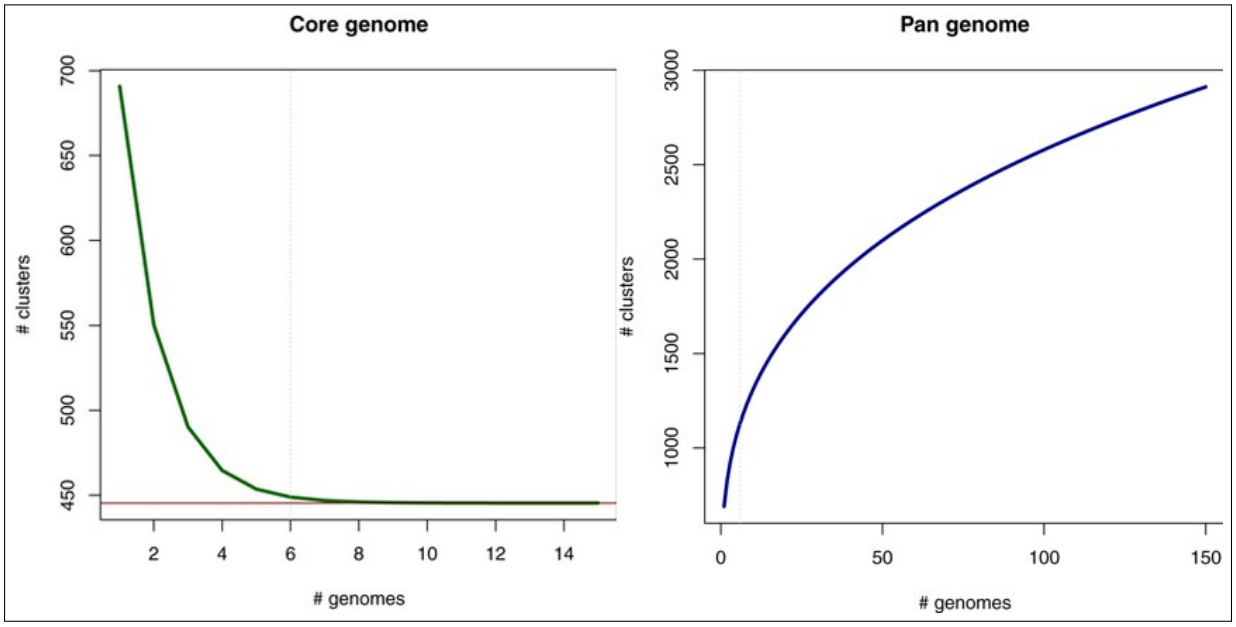


FIGURE 4.6 – Courbes du *core-genome* et du *pan-genome* des Pandoravirus, où le nombre de clusters identifiés en fonction du nombre de génomes est indiqué. Le nombre de clusters composant le *core-genome* diminue et semble se stabiliser autour de 430, à partir de 8 génomes. En revanche, le nombre de clusters composant le *pan-genome* tend à augmenter lors de l'ajout de nouveaux génomes, démontrant qu'il faut encore plus de génomes de Pandoravirus pour arriver à saturation.

nombre de gènes spécifiques qui nous permettra d'évaluer la diversité de ces génomes qui n'est pas encore bien définie. Il faudra toutefois encore plus de génomes de Pandoravirus pour arriver à saturation.

4.1.3 Analyse protéomique des Pandoravirus

Protéomes des particules purifiées

Nous avons étudié les protéines qui sont retrouvées dans les particules virales, et nous avons comparé les différents protéomes obtenus entre les Pandoravirus, afin d'identifier les protéines présentes dans les virions qui sont partagées par tous les virus, et celles qui leurs sont propres.

L'extraction des protéines a été réalisée sur les 4 Pandoravirus en parallèle, en triplicat pour chacun des virus. Les lysats protéiques ont ensuite été dosés et envoyés en protéomique [6] (cf. 2.7 Chapitre 2). Les spectres obtenus ont ensuite été comparés avec les ORFs prédits des Pandoravirus : ceux qui ont été utilisés ici pour l'identification, proviennent des génomes qui ont été ré-annotés.

Ces analyses ont permis d'identifier 267 protéines chez *P. neocaledonia*, 302 chez *P. quercus*, 334 chez *P. salinus* et 317 chez *P. dulcis*. Environ 85% de ces protéines sont

6. Les peptides obtenus après digestion des protéines sont séparés sur une colonne chromatographique en fonction de leur hydrophobicité : ils sont ensuite ionisés, fragmentés et analysés par spectrométrie de masse.

partagées par tous les virus, et constituent le *core-proteome* : les protéines restantes, qui sont donc spécifiques à chaque virus, devront être étudiées. L'analyse du protéome de *P. salinus* a montré qu'environ 60 de ces protéines correspondaient aux protéines de l'amibe *A. castellanii* (PHILIPPE *et al.*, 2013), ce qui est également le cas pour les autres Pandoravirus. Ces protéines ne semblent pas être fortement abondantes : en effet, les protéines les plus abondantes d'*A. castellanii* sont classées 44^{e} dans le protéome de *P. neocaledonia*, 91^{e} chez *P. quercus*, 100^{e} chez *P. dulcis* et 109^{e} chez *P. salinus*.

De plus, la grande majorité des protéines qui composent les Pandoravirus sont de fonction inconnue ("*hypothetical protein*"), et 80% n'ont pas d'homologues dans les bases de données de séquences.

Parmi les protéines les plus abondantes dans les protéomes des particules qui ont un homologue dans les bases de données, on retrouve les sous-unités gamma et tau de l'ADN polymérase, un facteur d'inhibition de l'initiation de la transcription (de la famille YjgF), du collagène triple-hélice, une "*thioredoxin like*" et une disulfure isomérase, qui sont toutes les deux des oxydoréductases. Certaines de ces protéines sont retrouvées à des niveaux d'abondance équivalents dans le protéome de chaque Pandoravirus, comme la protéine qui a pour homologue une "*thioredoxin-like*", classée 4^{e} dans tous les protéomes. D'autres protéines en revanche, sont retrouvées à des niveaux d'abondance différents, comme les sous-unités gamma et tau de l'ADN polymérase qui sont classées 2^{e} chez *P. neocaledonia*, 19^{e} chez *P. quercus*, 40^{e} chez *P. dulcis* et 34^{e} chez *P. salinus*.

Les trois protéines les plus abondantes de ces protéomes ont un classement identique chez tous les Pandoravirus, à l'exception de *P. quercus*. En effet, la protéine la plus abondante de son protéome est classée 3^{e} chez les autres, et la protéine qui est la plus abondante chez les autres Pandoravirus est classée 2^{e} dans son protéome. Ces protéines, qui sont homologues entre les virions, ne possèdent toutefois pas d'homologues dans les bases de données : elles sont de fonction inconnue pour le moment, et devront être étudiées.

Les études précédentes ont montré que *P. salinus* et *P. dulcis* codent pour 4 sous-unités de l'ARN polymérase, pour une enzyme de coiffe des ARNm ainsi que pour des facteurs de transcription, mais que la machinerie de transcription codée par ces virus n'est pas retrouvée dans les virions matures. Ils codent également pour une ADN polymérase, qui est retrouvée dans les protéomes des virions : en revanche, d'autres composés essentiels à la réplication de l'ADN en sont absents (ligase, topoisomérases, ...). Ces résultats ont suggéré que les fonctions de l'hôte étaient nécessaires pour initier la transcription des gènes, ainsi que la réplication des génomes des Pandoravirus, et que le cycle infectieux de ces virus était nucléo-cytoplasmique, ce qui a été confirmé par les observations au MET (PHILIPPE *et al.*, 2013 ; ABERGEL *et al.*, 2015). L'étude comparative des protéomes que nous avons menée sur tous les Pandoravirus confirme ces résultats,

et nous a permis de renforcer l'idée que ces virus appartiennent bien à la même famille, et qu'ils suivent les mêmes règles.

Cette étude des protéomes, qui est encore en cours, démontre que la composition des virions est globalement similaire, mais qu'il existe tout de même des différences qui expliquent les profils protéiques distincts. Enfin, aucun gène codant pour une "*Major Capsid Protein*", ou protéine majeure de capside n'a pu être identifié chez les Pandoravirus. L'absence de cette protéine est étonnante, car elle est habituellement retrouvée chez tous les virus à ADN double-brin qui infectent les eucaryotes, dont les virions sont généralement icosaédriques.

Les protéines qui composent l'enveloppe de ces Pandoravirus n'ont pas encore été identifiées : c'est donc dans ce but que des études sur la composition des particules ont été initiées. Nous nous sommes intéressés dans un premier temps à l'identification des protéines qui sont situées à la surface de ces particules : ces protéines sont-elles structurales et contribuent-elles à former l'enveloppe des virus ? Quelles sont les protéines qui sont toujours retrouvées en surface ? Quelles sont celles qui ne sont jamais libérées ? Ces dernières pourraient être impliquées dans la compaction et la protection de l'ADN viral, ou encore être nécessaires à l'initiation du processus infectieux.

Étude des protéines de surface des particules

Une étude de "surfome" ou protéome de surface des particules a donc été réalisée en triplicats et en parallèle sur *P. neocaledonia* et *P. quercus*. Nous avons mené cette étude sur deux virus pour identifier les protéines qui sont communes et celles qui sont caractéristiques de chaque virus, afin d'extraire toutes les informations possibles sur leur fonction (cf. 2.7 Chapitre 2).

Les premiers résultats que nous avons obtenus montrent que les niveaux de dégradation des particules ne sont pas équivalents : en effet, les données du surfome de *P. neocaledonia* sont parallèles à celles qui ont été obtenues lors de l'étude de son protéome, ce qui suggère que la digestion "partielle" de ces particules a été plus importante que celle de *P. quercus*. De plus, des résultats différents ont été obtenus pour les triplicats (cf. Figure 4.7), ce qui nous confirme que ce traitement n'était pas équivalent pour chaque préparation et que le protocole que nous avons utilisé devra être optimisé. Nous avons toutefois pu récupérer des informations sur les protéines de surface les plus abondantes ainsi que sur celles qui sont contenues à l'intérieur des particules et qui ne sont jamais libérées.

Le graphique (cf. Figure 4.7) représente les valeurs d'abondance des protéines qui ont été identifiées dans le surfome des particules, qui sont représentées par des boîtes à moustaches ("*box-plot*") : les "moustaches" indiquent les valeurs maximales et minimales qui ont été obtenues pour chaque protéine, ce qui illustre bien le fait que les

valeurs ne sont pas équivalentes pour chaque triplicat. Les courbes (cf. Figure 4.7) représentent les valeurs d'abondance minimales et maximales qui ont été obtenues pour les triplicats des protéines identifiées dans le protéome des particules (c'est à dire les particules non-traitées, "Contrôle").

Nous pouvons déduire d'après ce graphique que les protéines qui sont situées au-dessus des courbes sont plus abondantes dans le surfome que dans le protéome : ces protéines seraient donc situées à la surface des particules. Nous avons ensuite analysé les protéines les plus abondantes de ces surfomes. Cette analyse a montré que la protéine la plus abondante du surfome de *P. quercus* est une protéine structurale ("*collagen triple-helix*"), qui est classée en 6^e position dans son protéome, et qui est également retrouvée dans le surfome de *P. neocaledonia*, où elle est classée 2^e. Parmi les 15 protéines les plus abondantes qui sont orthologues entre les deux virus, plus de la moitié sont de fonction inconnue ("*hypothetical protein*") et n'ont pas d'homologues dans les bases de données. Ces protéines sont abondantes à la fois dans les surfomes et dans les protéomes et pourraient probablement correspondre à des protéines structurales qui composent l'enveloppe des Pandoravirus.

A l'inverse, les protéines situées en-dessous des courbes sont peu abondantes dans le surfome, et seraient probablement des protéines contenues dans les particules, qui sont retrouvées ici à cause du protocole que nous avons utilisé. Parmi celles-ci nous avons retrouvé de nombreuses protéases, kinases, phosphatases, ainsi que des protéines de fonction inconnue, qui pourraient par exemple être impliquées dans la compaction de l'ADN, ou encore être nécessaires lors de l'infection.

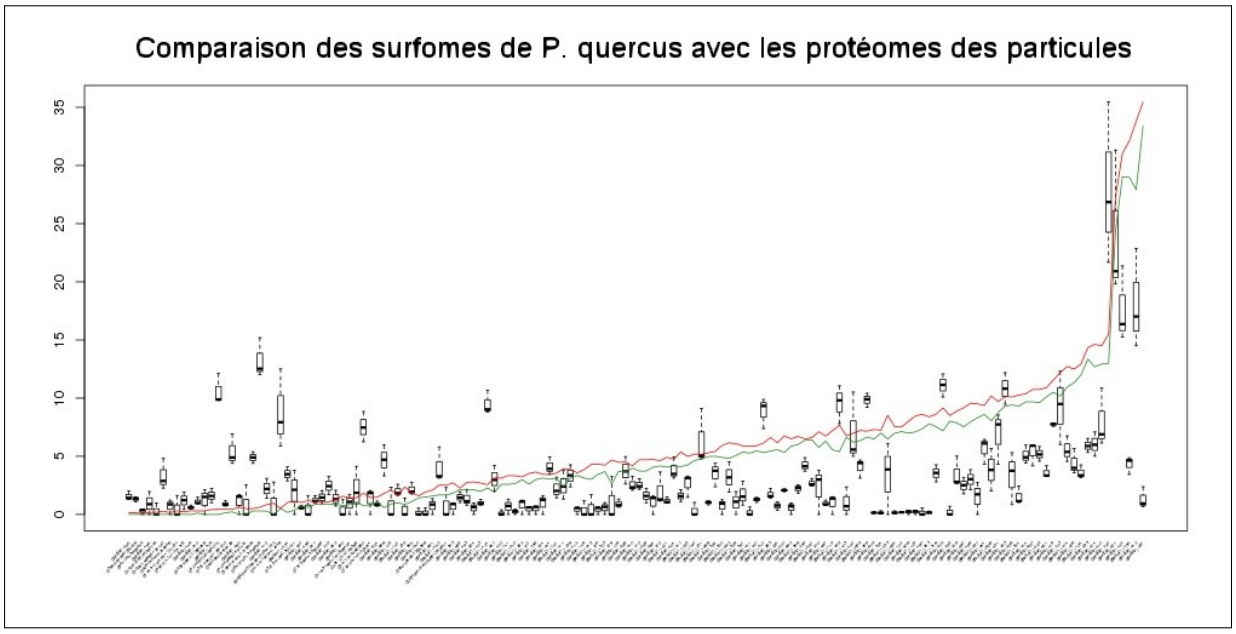


FIGURE 4.7 – Graphique des protéines identifiées dans les surfomes des triplicats de *P. quercus*, où les protéines sont représentées en fonction de leur niveau d'abondance. Les courbes représentent le minimum (vert) et le maximum (rouge) des triplicats du protéome des particules, et les boîtes à moustaches ("*box-plot*") représentent les triplicats du surfome.

Une grande proportion des protéines de l'amibe *A. castellanii* ont pu être retrouvées parmi les protéines les plus abondantes dans ces deux analyses de surfome. Quelques unes de ces protéines ont également été retrouvées en faible abondance dans les protéomes des particules virales. Toutefois, certaines de ces protéines ont été retrouvées uniquement dans les surfomes, et pourraient correspondre à des "*by-standers*" qui seraient accrochées aux particules, et qui auraient été décrochées lors de cette digestion partielle. Comme nous l'avons récemment démontré chez Noumeavirus (cf. Chapitre 3), les protéines d'*A. castellanii* ont pu être décrochées des particules à l'aide d'un traitement au DMSO [7]. Cette étude du surfome des Pandoravirus sera recommencée, et pourra être précédée par ce traitement afin d'éliminer ces protéines "*by-standers*".

L'observation au MET des particules qui ont été partiellement digérées, n'est en revanche pas très informative (cf. Figure 4.8). En effet, aucune différence claire n'a pu être observée lorsque l'on compare l'aspect des particules qui ont été traitées à la trypsine par rapport à celles non-traitées ("contrôle"). Les particules virales, ainsi que leurs enveloppes semblent être intactes, et les observations des triplicats de chaque Pandoravirus qui a été traité sont identiques (cf. Figure 4.8). Ces observations suggèrent que les protéines qui ont été décrochées de la surface des particules ne sont peut-être pas visibles au MET : ces particules pourraient donc être observées au cryo-microscope électronique, qui possède une résolution supérieure.

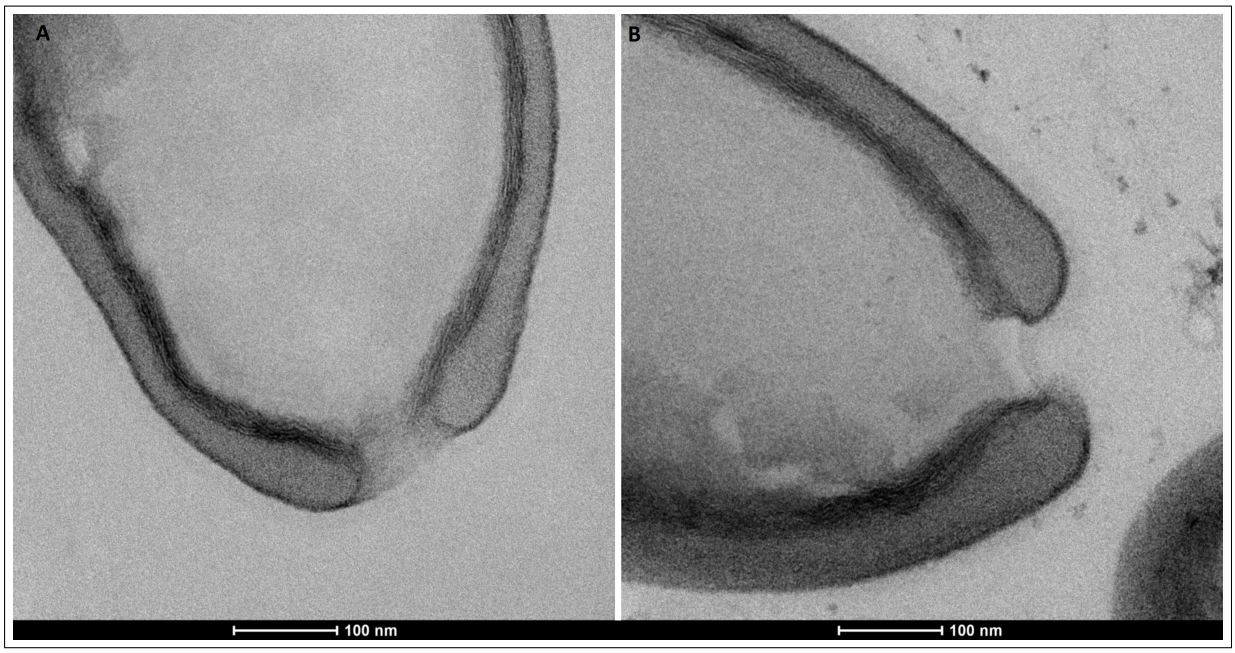


FIGURE 4.8 – Particules de Pandoravirus observées au MET après digestion à la trypsine. (A) Contrôle : l'enveloppe des particules est intacte et formée de trois couches. (B) Particule observée après digestion partielle à la trypsine : l'enveloppe apparaît intacte et identique au contrôle.

7. Le DMSO ou diméthylsulfoxyde est un solvant qui peut être utilisé comme agent cryoprotecteur, à une concentration finale de 7,5%.

4.1.4 Comparaison des cycles infectieux des Pandoravirus

Les cycles infectieux de ces trois Pandoravirus (*P. neocaledonia*, *P. quercus* et *P. celtis*) ont été étudiés, puis comparés entre eux, ainsi qu'avec ceux de *P. dulcis* et *P. salinus*. Pour cela, des amibes *A. castellanii* ont été infectées par *P. neocaledonia*, *P. quercus* ou *P. celtis*. Ces cellules ont été régulièrement observées au microscope optique, puis ont été fixées à différents temps. Deux types d'observations ont alors été réalisées : une partie des cellules a été marquée au DAPI avant d'être observée au microscope à épifluorescence, et le reste des cellules a été inclus en résine Epon avant d'être observé au MET (cf. 2.3 et 2.5 Chapitre 2).

Lorsque l'on observe le déroulement de ces cycles infectieux au microscope optique, on peut remarquer qu'après 3 hpi, les cellules se divisent en deux populations, ce qui n'est pas le cas des cellules non-infectées (cf. A et B Figure 4.9). Une partie des cellules se décolle du fond de la flasque et elles s'arrondissent progressivement ; l'autre partie est constituée de cellules qui restent adhérentes et qui gardent leur forme de trophozoïte.

En revanche, la membrane externe de ces cellules devient de plus en plus difficile à observer : ces cellules ne sont plus parfaitement délimitées et palissent au cours du temps, jusqu'à avoir un aspect fantomatique (cf. B Figure 4.9). Elles disparaissent ensuite progressivement au cours du cycle, au profit des cellules rondes qui perdent leur adhérence (cf. C Figure 4.9).

Ce phénotype se poursuit jusqu'à 8 hpi, où des particules commencent à être observées dans le milieu extra-cellulaire, ce qui suggère une libération des virions par exocytose (cf. C Figure 4.9). Le nombre de cellules commence véritablement à diminuer à partir de 10 hpi : les virions neo-synthétisés remplissent les cellules jusqu'à ce qu'elles lysent, ce qui entraîne leur libération dans le milieu extra-cellulaire, et marque la fin du cycle (cf. D Figure 4.9).

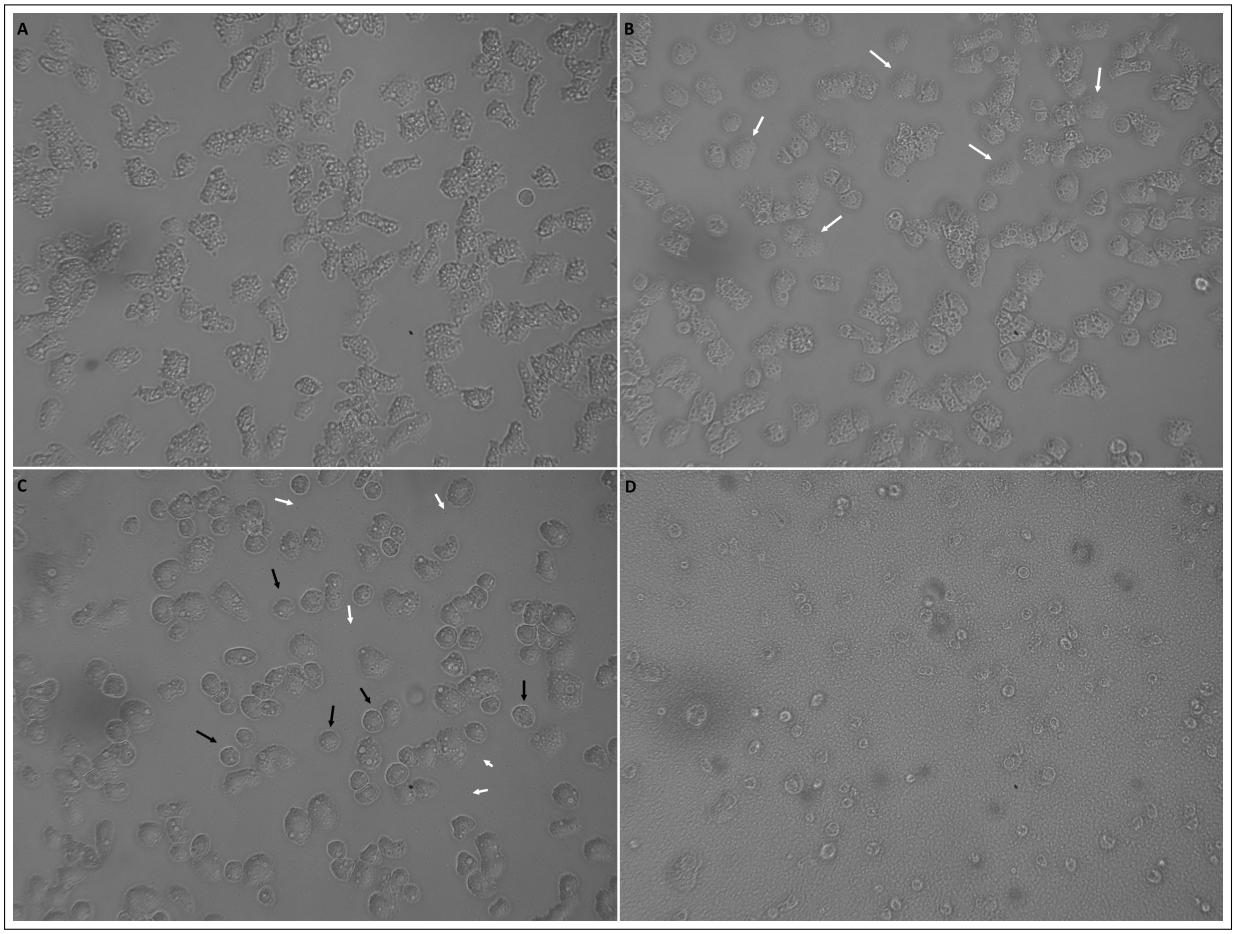


FIGURE 4.9 – Cycle de *P. neocaledonia* observé au microscope optique (objectif x40). (A) Témoin (cellules non-infectées). (B) Cellules infectées observées à 3 hpi, avec des cellules "fantômes" indiquées par des flèches blanches. (C) Cellules infectées observées à 8 hpi : la proportion de cellules arrondies augmente (flèches noires) et les premières particules virales sont observées dans le milieu extra-cellulaire (flèches blanches). (D) Cellules infectées observées à 25 hpi : il reste encore quelques cellules rondes et le milieu est rempli de particules virales.

Ces cycles infectieux ont ensuite été observés au MET. Ces observations nous ont permis d'identifier plusieurs étapes, qui sont identiques pour tous les virus :

i. **L'internalisation**

Dès 30 minutes pi, des particules de Pandoravirus sont observées dans des vacuoles, qui sont situées dans le cytoplasme des cellules (cf. B Figure 4.10). Quelques particules peuvent être encore observées à l'extérieur des cellules, ce qui reste assez rare.

Des observations ont également montré que certaines particules étaient entourées par les pseudopodes de la cellule, ce qui indique qu'elles sont en cours d'internalisation (cf. A Figure 4.10).

Cette étape d'internalisation, qui est réalisée par un mécanisme de phagocytose (PHILIPPE *et al.*, 2013), marque la première étape du cycle infectieux des Pandoravirus.

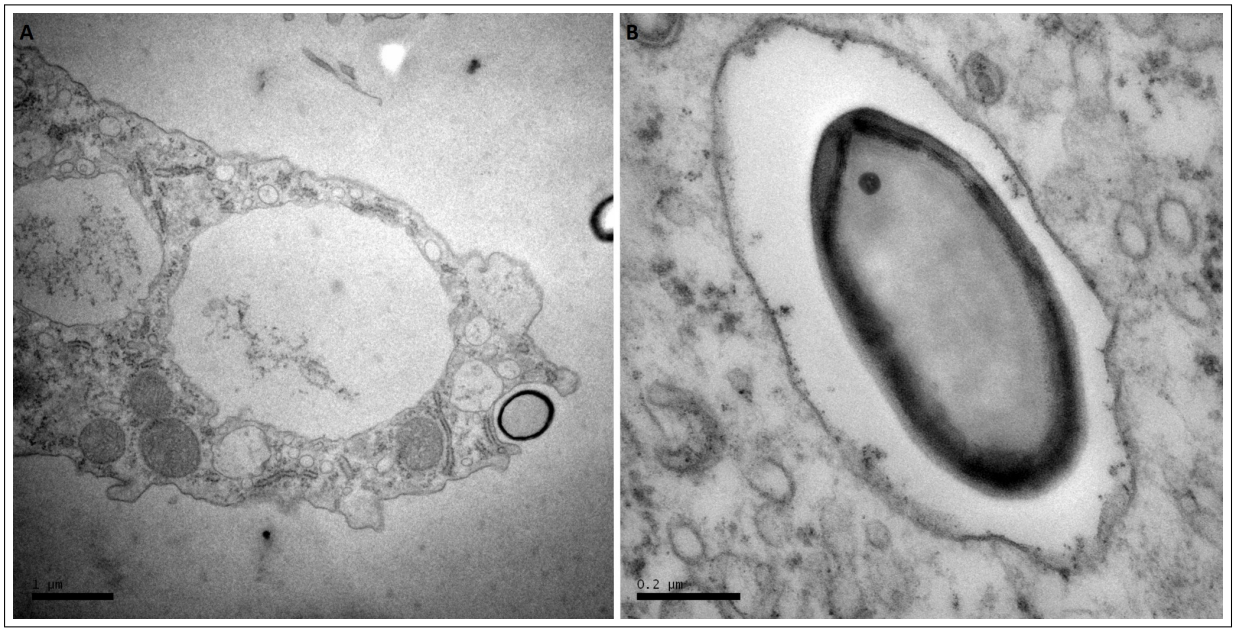


FIGURE 4.10 – Étape d'internalisation des Pandoravirus observée au MET. (A) *P. dulcis* en cours d'internalisation : la particule est entourée par les pseudopodes de la cellule. (B) Particule de *P. salinus* internalisée et contenue dans une vacuole à l'intérieur du cytoplasme de la cellule.

ii. **La libération du matériel génomique viral**

A partir de 30 minutes pi, les particules qui sont contenues dans les vacuoles perdent leur bouchon situé au niveau du pore apical, selon un mécanisme qui n'est pas encore connu.

Une fois ce bouchon retiré, la membrane lipidique qui délimite l'intérieur diffus de la particule fusionne avec la membrane de la vacuole. Cette fusion crée une continuité entre la particule et la cellule, ce qui permet au virion de délivrer son matériel génomique dans le cytoplasme : cette étape correspond à l'initiation de l'infection (cf. Figure 4.11).

En revanche, aucune structure correspondant au génome du virion n'a pu être observée dans le cytoplasme après sa libération. Cette étape pourrait constituer une sorte de "phase d'éclipse", au cours de laquelle le génome ne peut pas être suivi dans le cytoplasme (cf. Figure 4.11), contrairement aux *Mimiviridae* (ABERGEL *et al.*, 2015).

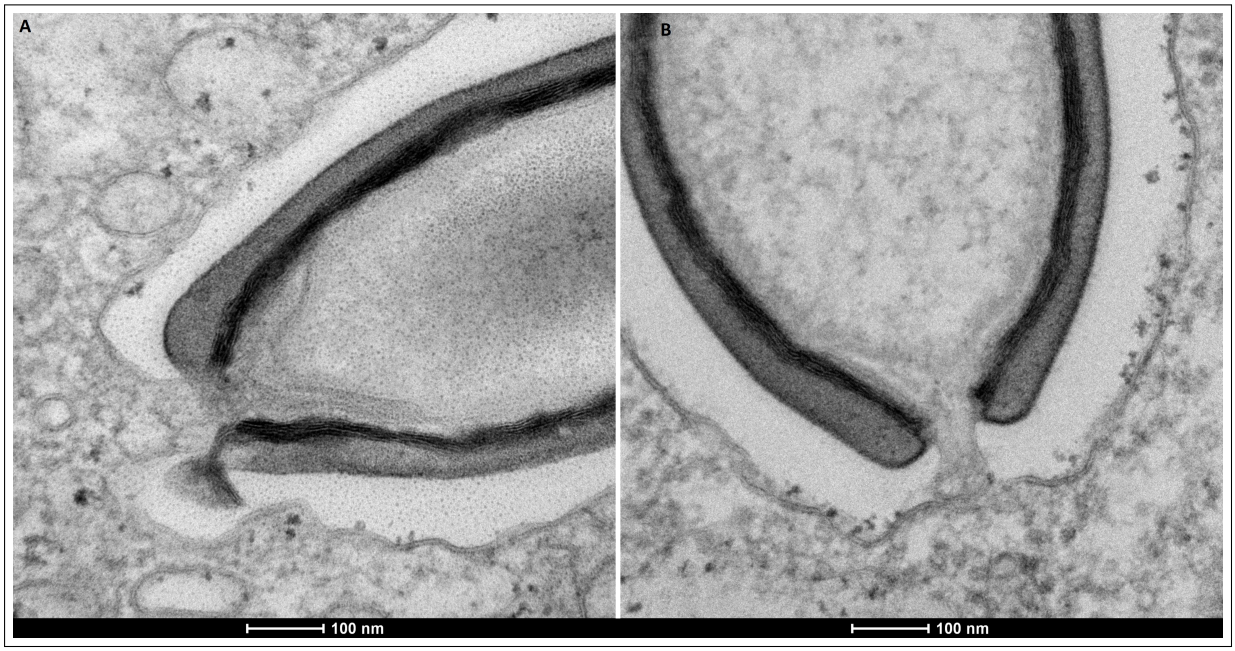


FIGURE 4.11 – Étape de libération du génome des Pandoravirus observée au MET. (A) La membrane interne de *P. neocaledonia* a fusionné avec celle de la vacuole, et le matériel génétique du virion peut être déversé dans le cytoplasme. (B) Un phénomène identique est observé au cours du cycle infectieux de *P. quercus*.

iii. **Les modifications apportées aux noyaux des cellules hôtes**

A partir de 5 hpi, les noyaux des cellules infectées sont drastiquement modifiés : la membrane nucléaire subit de nombreuses invaginations, qui peuvent être plus ou moins marquées selon les virions.

En effet, la membrane nucléaire des cellules qui sont infectées par *P. dulcis* ne présente que très rarement des invaginations, contrairement à celle des cellules infectées par *P. salinus*, qui est complètement déformée (PHILIPPE *et al.*, 2013). Le même phénomène a également été observé chez *P. neocaledonia* (cf. B Figure 4.12), *P. celtis* et *P. quercus* (cf. A Figure 4.12).

De plus, le nucléole des cellules infectées est également altéré (cf. Figure 4.12). Dans les cellules non-infectées, il apparaît dense aux électrons et bien organisé : en revanche, lors d'une infection par un Pandoravirus, celui-ci se désorganise progressivement et devient de moins en moins dense (cf. Figure 4.12).

Ces modifications apportées aux noyaux des cellules hôtes au cours du cycle se font progressivement, qu'il s'agisse des membranes ou des nucléoles, jusqu'à leur complète disparition. La région qui contenait l'ancien noyau paraît assez homogène, diffuse et peu dense, contrairement au reste du cytoplasme, qui a un aspect granuleux, et qui contient de nombreux organites (cf. Figure 4.13). Cette zone est particulièrement visible au microscope à épifluorescence par marquage au DAPI

après 6 hpi pour *P. neocaledonia* (cf. Figure 4.13) et 7 hpi pour *P. celtis* et *P. quercus*. Cette disparition du noyau confirme que les Pandoravirus passent par une phase nucléaire au cours de leur cycle infectieux (PHILIPPE *et al.*, 2013).

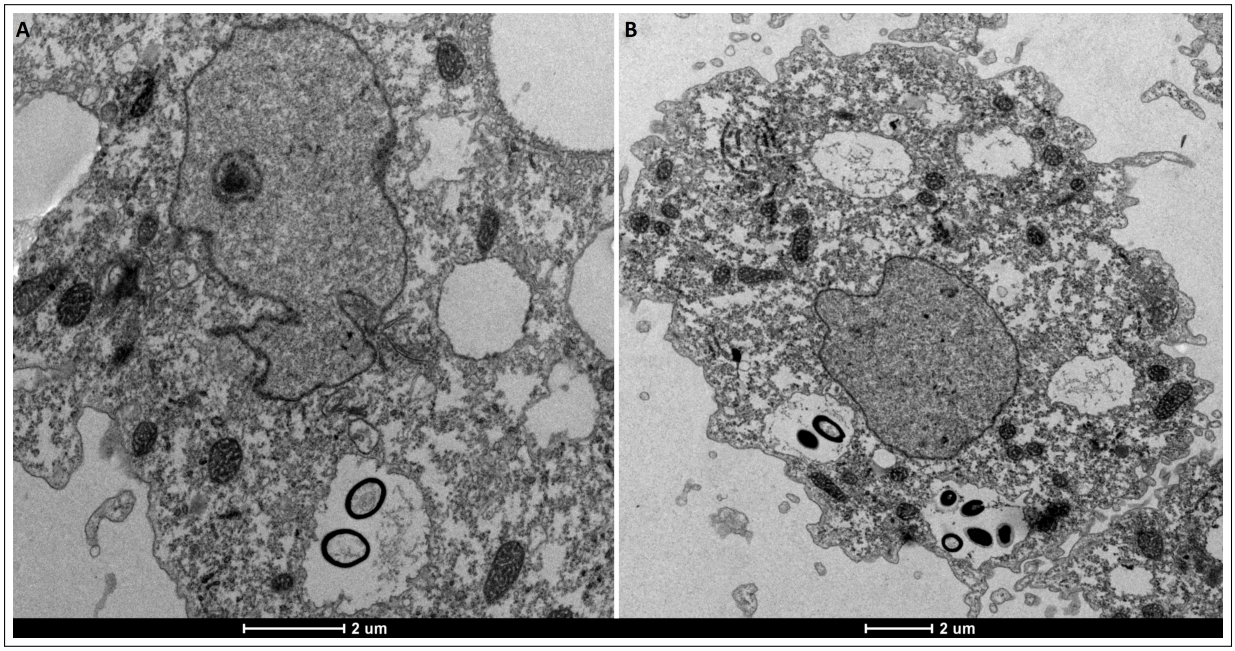


FIGURE 4.12 – Étape de modification du noyau des cellules hôtes au cours du cycle des Pandoravirus observée au MET. (A) Cellule infectée par *P. quercus* : la membrane nucléaire est déformée et le nucléole a été altéré. Des particules virales vides sont observées dans une vacuole de la cellule. (B) Cellule infectée par *P. neocaledonia* : la membrane nucléaire est partiellement déformée et le nucléole n'est plus visible. Des particules vides sont également observées dans des vacuoles.

Au cours de ces nombreuses modifications de la morphologie du noyau, des particules de Pandoravirus qui sont vides, et en cours de dégradation, peuvent être observées dans les vacuoles (cf. Figure 4.12).

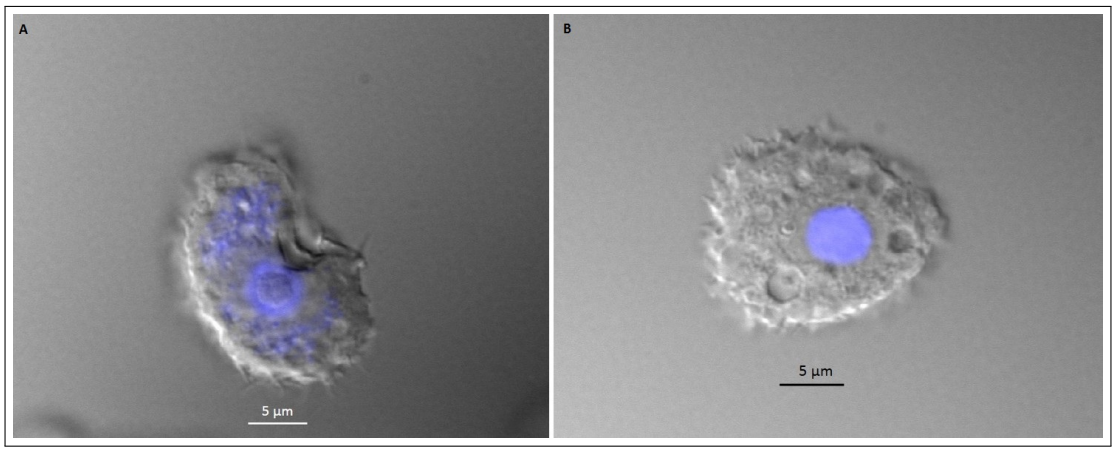


FIGURE 4.13 – Observation d'une cellule infectée par *P. neocaledonia* au microscope à épifluorescence (objectif x63 et Optovar x1,6). (A) Témoin (cellule non-infectée) : le marquage DAPI est localisé au niveau du noyau et des mitochondries de la cellule. (B) Cellule infectée par *P. neocaledonia* observée après 6 hpi : le marquage DAPI, de forte intensité, est localisé dans une zone homogène, qui correspond à l'ancien noyau de la cellule.

iv. **Les usines virales et la synthèse de virions**

Des virions en cours de synthèse ont pu être observés en périphérie de cette région assez diffuse et peu dense, qui correspond à l'ancien noyau (cf. A Figure 4.14). La synthèse des virions, qui est identique pour tous les Pandoravirus, a lieu à différents temps du cycle : à partir de 8 hpi pour *P. neocaledonia*, à 9 hpi pour *P. dulcis*, à 10 hpi pour *P. celtis* et *P. quercus*, et enfin, à partir de 12 hpi pour *P. salinus*.

Les observations de ces cycles au MET ont permis de montrer que le compartiment interne des virions, qui est diffus et homogène, ainsi que leur enveloppe composée de 3 couches, sont synthétisés simultanément dans le cytoplasme des cellules (cf. C et D Figure 4.14).

Cette synthèse est initiée par le pore situé à l'apex des particules, et il semblerait que le bouchon qui ferme ce pore soit également assemblé en même temps. Ces structures denses aux électrons, peuvent être observées dans le cytoplasme des cellules infectées. Cette synthèse progressive, qui fait penser à du "tricotage" (PHILIPPE *et al.*, 2013), prend fin lorsque les virions sont matures et que les génomes ont été chargés dans les particules. De plus, une zone d'exclusion du cytoplasme a pu être observée pour tous les virions en cours de synthèse, à l'endroit où se forme l'enveloppe qui entoure la particule (cf. C et D Figure 4.14).

Par ailleurs, des structures ont pu être observées au cours de cette synthèse dans la zone diffuse, précédemment occupée par le noyau. Ces structures sont retrouvées sous plusieurs formes : certaines sont tubulaires, comme cela a déjà été observé au cours du cycle de Noumeavirus, et d'autres sont sous forme de ronds (cf. B Figure 4.14). Ces structures denses aux électrons, font penser à des membranes qui pourraient être utilisées par les virions pour leur synthèse.

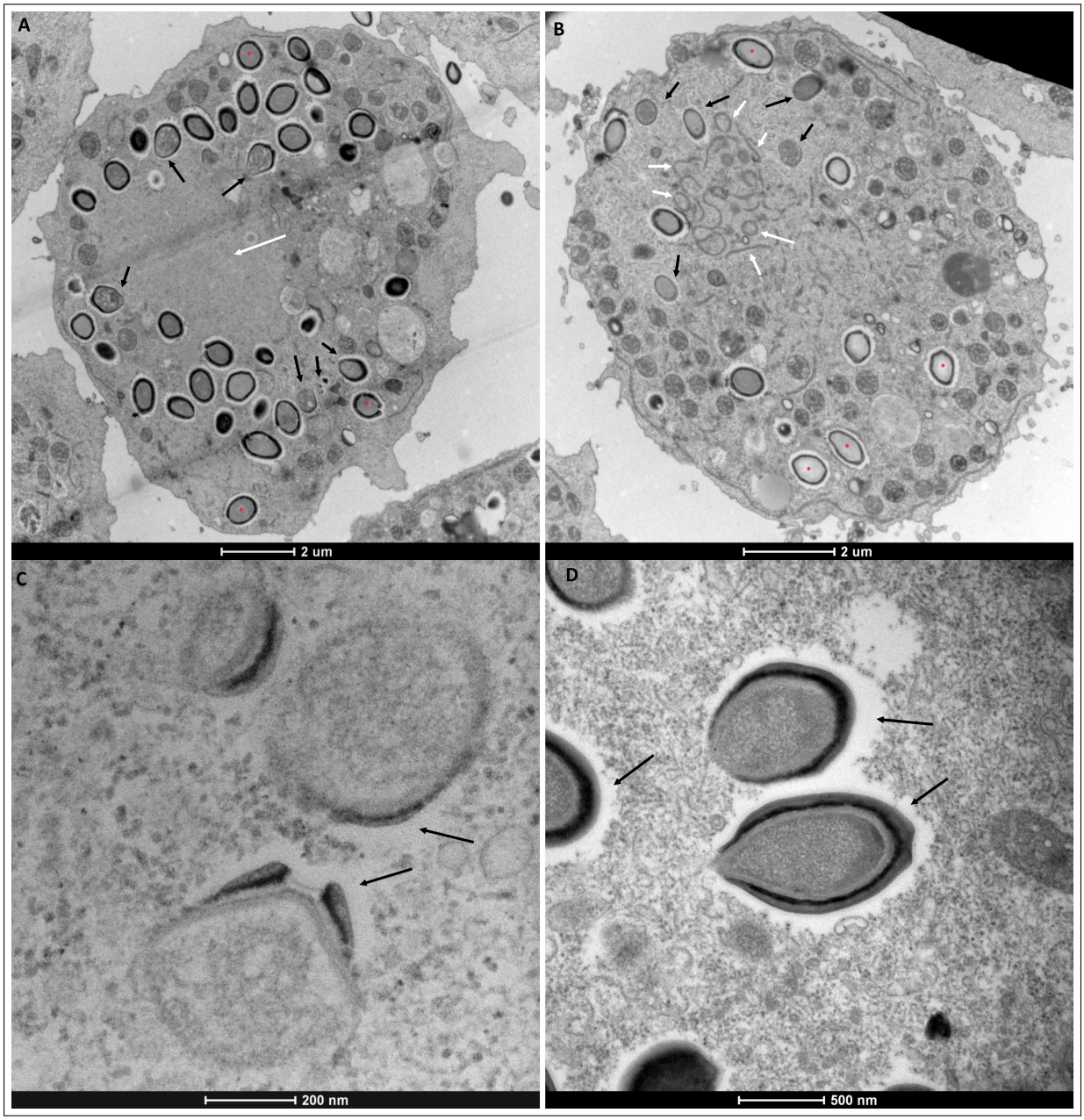


FIGURE 4.14 – Étape de synthèse des Pandoravirus observée au MET. (A) Cellule infectée par *P. neocaledonia* : les virions sont en cours de synthèse (flèches noires) en périphérie d'une zone homogène (flèche blanche), qui correspond à l'ancien noyau. Les particules matures sont contenues dans des vacuoles (astérisques rouges) (B) Cellule infectée par *P. neocaledonia* : des structures tubulaires et en forme de rond sont observées (flèches blanches) à côté des virions en cours de synthèse (flèches noires). Les particules matures sont contenues dans des vacuoles (astérisques rouges). (C) Particules de *P. neocaledonia* en cours de synthèse : des zones d'exclusion du cytoplasme sont observées autour des virions (flèches noires). (D) Particules de *P. quercus* en cours de synthèse : des zones d'exclusion du cytoplasme (flèches noires) ainsi que des virgules de membrane sont observées autour des virions.

D'autres structures, en forme de virgules, moins denses aux électrons, ont également pu être observées, cette fois en périphérie des virions (cf. D Figure 4.14). Leur épaisseur ainsi que leur densité font penser à des membranes lipidiques, qui ressemblent à celles qui délimitent l'intérieur des particules virales.

La présence de petits amas anguleux et denses aux électrons, a été fréquemment observée au cours de ces cycles infectieux : ces amas sont situés dans le cytoplasme des cellules, où les virions sont en cours de production (cf. A Figure 4.15). Lorsqu'ils sont observés à fort grossissement, ces amas sont constitués de structures lamellaires, similaires à celles de la couche intermédiaire des enveloppes virales (cf. B Figure 4.15). Ces amas pourraient donc correspondre à des fragments d'enveloppes, qui participeraient à l'assemblage des virions.

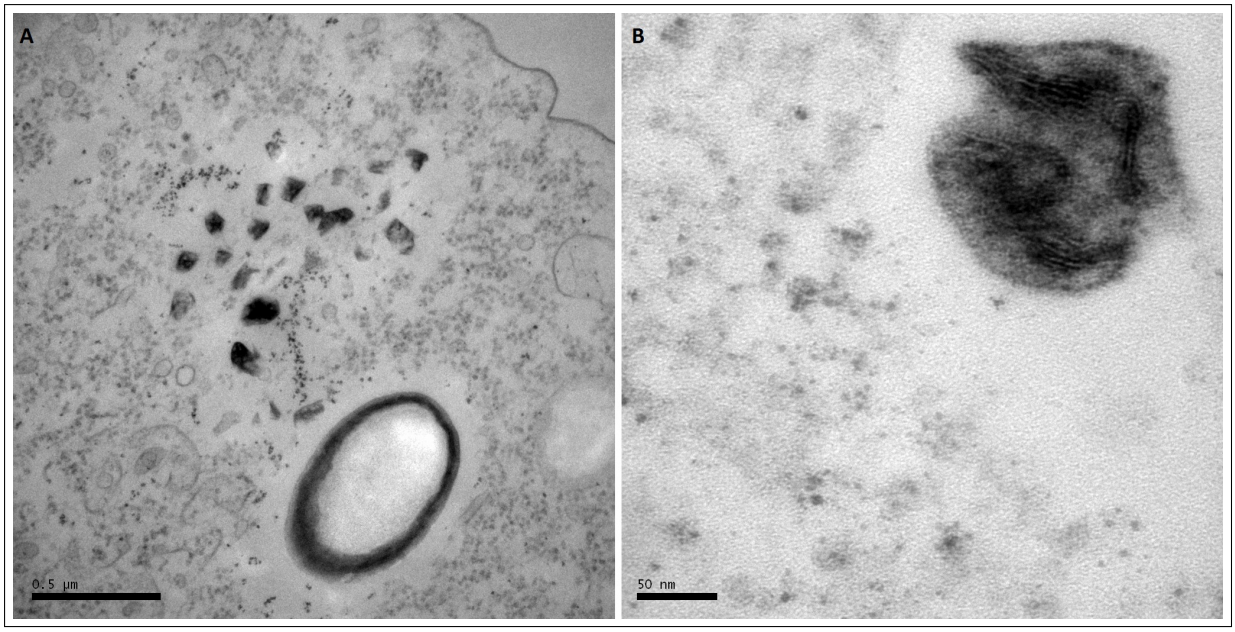


FIGURE 4.15 – Observation d'amas au MET au cours de la synthèse des virions. (A) Amas regroupés dans le cytoplasme à côté d'un virion de *P. dulcis* en fin de synthèse. (B) Agrandissement d'un de ces amas, qui est constitué de structures lamellaires similaires à celle de la couche intermédiaire de l'enveloppe des Pandoravirus.

v. **La libération des virions dans le milieu**

Enfin, après l'étape de synthèse des virions, les virions matures peuvent être observés dans des vacuoles situées dans le cytoplasme, et qui sont proches de la périphérie des cellules (cf. A et B Figure 4.14). L'observation de virions dans ces vacuoles en fin de cycle, suggère une libération des nouvelles particules par exocytose dans le milieu extra-cellulaire (PHILIPPE *et al.*, 2013), ce qui a pu être confirmé par nos observations au microscope optique.

La production virale au cours de ces cycles infectieux est assez importante, et les virions peuvent remplir le cytoplasme des cellules infectées : celles-ci vont lyser, conduisant alors à la libération de toutes les particules dans le milieu extra-cellulaire (cf. Figure 4.16).

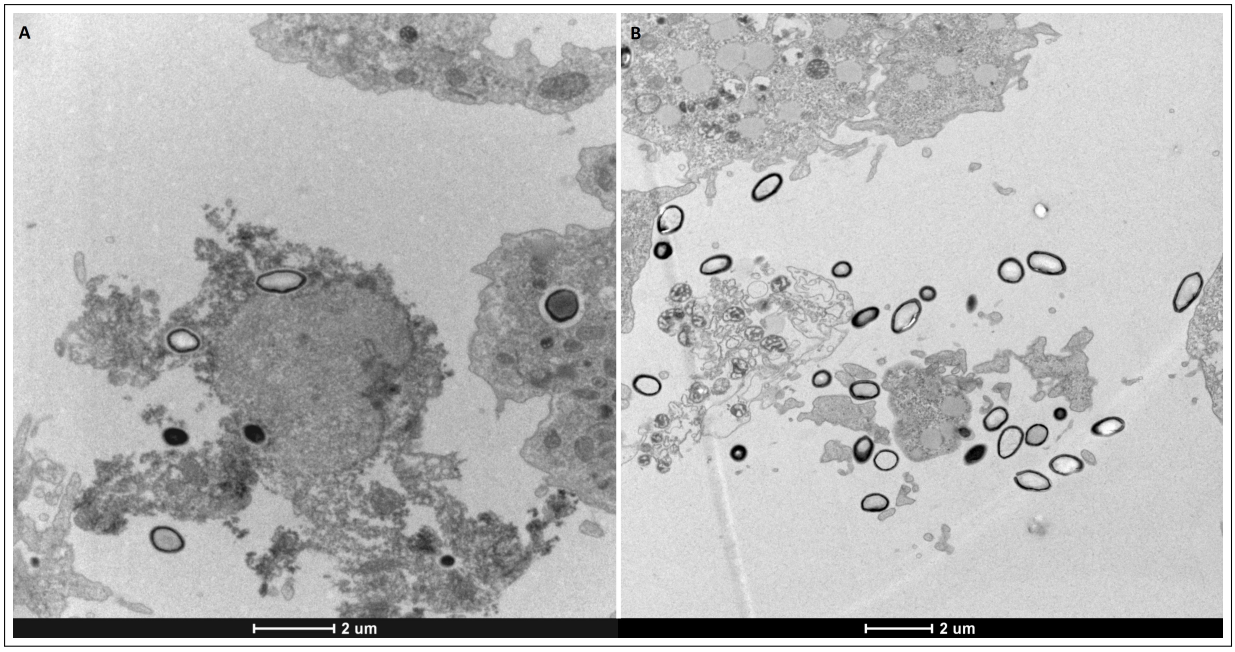


FIGURE 4.16 – Cellules infectées par des Pandoravirus en cours de lyse observées au MET. (A) Cellule en cours de lyse libérant des particules de *P. quercus* dans le milieu extra-cellulaire. (B) Cellule en cours de lyse libérant des particules de *P. neocaledonia* dans le milieu extra-cellulaire.

4.2 Discussion

Les Pandoravirus, qui sont des virus uniques dans le monde viral, possèdent une morphologie étonnante : leur enveloppe ovoïde d'1 μm de long est composée de 3 couches, et l'intérieur de leur particule semble homogène. Ce compartiment interne diffus ne possède pas de structures visibles, à l'exception d'une sphère d'environ 50 nm de diamètre, qui est dense aux électrons : celle-ci n'est toutefois pas observée régulièrement dans les virions, notamment à cause des différents plans de coupe. En revanche, cette sphère ne semble pas correspondre au génome colossal de ces virions (2 Mb) qui aurait été compacté (PHILIPPE *et al.*, 2013) : sa nature reste inconnue à ce jour, et mérite d'être étudiée. Des expériences d'immunomarquage à l'aide de billes d'or qui sont observables au MET, pourraient éventuellement nous renseigner sur la localisation du génome dans les particules.

L'analyse des génomes de ces virus a été combinée avec les analyses des protéomes des particules purifiées et des transcriptomes des cellules infectées au cours du temps. Ces approches combinées nous ont notamment permis de prédire avec certitude un certain nombre de gènes correspondant à un peu plus de la moitié des ORFs possibles. L'analyse de la structure des transcrits montre que les régions codantes des génomes des Pandoravirus (environ 65%) sont séparées par de longues régions intergéniques de 650 nucléotides en moyenne. Nous avons réduit le nombre de gènes codants afin de ne pas surinterpréter les données, le nombre de gènes initialement prédits pouvant se trouver pollué par des ORFs ne correspondant pas à de vraies protéines, en raison du fort contenu en G+C du génome des Pandoravirus. En effet, les codons Stop étant riches en A+T, ils sont moins fréquents. D'autre part, les gènes présentent de très longs UTR ("*Untranslated Transcribed Region*"), qui sont des régions transcrites mais non traduites en protéines, retrouvées aux extrémités 5' et 3' des ARN messagers. Ces régions sont notamment impliquées dans la régulation de l'expression des gènes, et dans la stabilité des ARN messagers. La présence de ces longues séquences pourrait permettre aux transcrits viraux d'être traduits en protéines plus efficacement, et plus rapidement.

Toutefois, notre annotation de ces gènes codants a peut-être été trop conservative, et certains des gènes que nous n'avons pas retenus pourraient correspondre à des protéines réelles. Afin de vérifier si les gènes qui ont été identifiés dans les génomes sont vraiment codants, le protéome cinétique des cellules infectées devra être analysé. Par la suite, une analyse de Ribo-seq (ou "*Ribosome profiling*", soit le profilage des ribosomes), pourra être envisagée. Grâce à cette technique, les ribosomes qui sont actifs dans les cellules infectées à un temps donné du cycle infectieux peuvent être identifiés, ce qui nous permettrait de déterminer les transcrits qui ont été traduits en protéines (INGOLIA, 2014).

Ces analyses comparatives des génomes des différents Pandoravirus nous ont également permis d'établir une phylogénie de la famille des *Pandoraviridae*. Nous avons déterminé grâce à cette étude que cette famille semblait être constituée de deux lignées. La première lignée est composée de *P. salinus*, *P. dulcis*, *P. quercus* et *P. inopinatum*, et la deuxième est composée de deux virus qui ont été isolés récemment : *P. neocaledonia* et *P. D1*. La découverte de nouveaux Pandoravirus nous permettra de déterminer si ceux-ci s'intègrent bien dans une de ces deux lignées, ou s'ils seront les membres fondateurs d'autres lignées, comme chez les *Marseilleviridae* (Dos Santos *et al.*, 2016). Elle permettra, de plus, d'obtenir des indices supplémentaires sur l'origine et l'histoire évolutive de ces virus.

L'étude comparative des protéomes des particules purifiées des Pandoravirus, nous a permis de déterminer que ces virus ont une majorité de protéines en commun, qui ont des niveaux d'abondance comparables, mais également quelques protéines qui leur sont propres, souvent en abondance moindre.

En complément, l'étude du surfome (ou des protéines de surface) de ces particules, nous a permis de voir que l'une des protéines les plus abondantes est une protéine structurale ("*collagen triple-helix*"), qui est probablement impliquée dans la composition de l'enveloppe de ces virus. Toutefois les traitements n'étaient pas équivalents et cette étude mériterait d'être recommencée et optimisée : le temps de digestion devrait être réduit, afin de ne récupérer que les protéines de surface. Par ailleurs, des particules qui étaient ouvertes, partiellement dégradées, ou qui n'étaient pas suffisamment matures, ont pu être digérées lors de cette analyse : des protéines situées à l'intérieur des virions ont alors été libérées dans le milieu, ce qui a contribué à modifier nos résultats.

Enfin, des études sur la composition en protéines des parois des Pandoravirus sont en cours d'analyse : pour cela, les particules virales ont été vidées de leur contenu, et les protéomes de ces parois isolées pourront être étudiés, et comparés aux protéomes des particules entières.

Le cycle infectieux de ces virus est complètement différent de celui des *Mimiviridae* (Claverie *et al.*, 2006), et passe par une phase nucléaire, comme les *Phycodnaviridae* (Wilson *et al.*, 2009).

Les virions semblent pénétrer dans les cellules par un mécanisme de phagocytose : cela mériterait toutefois d'être vérifié, en utilisant par exemple des inhibiteurs de la phagocytose, puis en vérifiant la présence de virions à l'intérieur des cellules.

Les virions fusionnent ensuite leur membrane interne avec celle de la vacuole dans laquelle ils sont contenus, afin de libérer leur génome dans le cytoplasme des cellules hôtes. Ce génome n'est cependant pas visible, que ce soit dans la particule, ou dans le cytoplasme, et ne peut donc pas être suivi dans la cellule une fois libéré. Un marquage fluorescent de ces génomes pourrait toutefois nous renseigner sur le chemin qu'ils parcourent dans les cellules infectées, comme cela a déjà été fait chez *Mollivirus sibericum*

(LEGENDRE *et al.*, 2015), ou encore en utilisant le suivi de la fluorescence d'une seule molécule, couplé à une imagerie à très haute résolution ("*single-molecule tracking and super-resolution imaging*", GAHLMANN *et al.*, 2014).

De même, l'étape de chargement du génome lors de la synthèse des virions reste à résoudre : ce génome est-il chargé au fur et à mesure lors de la synthèse ou est-il empaqueté à la fin comme c'est le cas chez les *Mimiviridae* (ZAUBERMAN *et al.*, 2008) ?

Enfin, la libération des particules virales dans le milieu marque la fin de ce cycle. Cette libération semble se faire par un mécanisme d'exocytose, où les virions matures sont dans des vacuoles, avant d'être relâchés dans le milieu extra-cellulaire. Toutefois, après une quinzaine d'heures post-infection, la grande majorité des cellules est lysée, libérant alors tous les virions dans le milieu, même ceux qui ne sont pas encore matures. Ce phénomène pourrait expliquer la différence importante que nous avons observée lors du comptage et de la titration de ces particules : beaucoup de virions sont présents dans le milieu, mais ils ne sont pas tous infectieux, probablement en raison de leur manque de maturité.

Le déroulement des cycles infectieux des différents Pandoravirus qui ont été étudiés au cours de cette analyse comparative est similaire pour tous les virus. Toutefois, certains semblent être plus rapides que d'autres, ce qui pourrait être expliqué par une différence dans le nombre de particules qui ont été utilisées, ou encore par un manque de synchronicité lors de l'infection des cellules.

4.3 Conclusion

La découverte des virus géants a été un tournant majeur dans le domaine de la virologie, et la découverte des Pandoravirus quelques années plus tard, en a encore repoussé les limites.

Pandoravirus salinus et *P. dulcis* (PHILIPPE *et al.*, 2013) ont été les premiers Pandoravirus a avoir été décrits et caractérisés comme des virus. Toutefois, un parasite possédant la même morphologie avait déjà été décrit quelques années plus tôt, mais sa nature virale n'avait jamais été révélée, jusqu'à récemment (SCHEID *et al.*, 2014). Ce nouveau membre de la famille des *Pandoraviridae* a été nommé *P. inopinatum* et son génome a été séquencé il y a peu (SCHEID, 2016).

La découverte ainsi que la caractérisation de nouveaux Pandoravirus, qui ont été réalisées au cours de cette thèse, nous ont permis de réaliser des analyses comparatives entre tous les membres de cette famille, qui s'agrandit au fil du temps. Ces études, toujours en cours pour certaines, nous ont donné des indices à propos de ce qui caractérise ces virus, afin de pouvoir répondre aux nombreuses questions qui ont été soulevées par cette découverte, dont : quelle est la place des Pandoravirus au sein du monde viral ?

Les Pandoravirus sont des virus complexes, qui possèdent de grands génomes, avec un grand nombre de gènes. Notre étude des transcriptomes des cellules infectées, combinée aux analyses des génomes et des protéomes de ces virus, nous ont permis de réduire le nombre de gènes codants qui avaient été prédits au sein de ces génomes. Toutefois, notre annotation est très conservative et nous avons peut-être éliminé des gènes qui correspondent à des protéines réelles. Des analyses expérimentales pourront être réalisées pour identifier et quantifier les gènes codants et non-codants qui sont présents dans ces génomes, afin de valider notre annotation.

Par ailleurs, la majorité des gènes (70%) ne possèdent pas d'homologues dans les bases de données de séquences, et ces gènes ne ressemblent à rien de connu, ce qui rend leur étude d'autant plus compliquée. Cependant, grâce aux découvertes récentes de nouveaux Pandoravirus, une ébauche de cette famille se dessine, et il semblerait que la famille des *Pandoraviridae* soit divisée en 2 lignées.

Notre analyse comparative des génomes, protéomes, transcriptomes et des cycles infectieux nous a mené à la conclusion suivante : ces virus qui sont morphologiquement identiques, possèdent tout de même des différences, que ce soit au niveau de leur composition ou du déroulement de leur cycle réplicatif, qui nous permettent de les discriminer. Cette étude soulève toutefois des questions supplémentaires : d'où viennent ces gènes ? Constituent-ils de nouvelles voies métaboliques que nous n'avons pas encore identifiées ? Comment évoluent ces génomes ?

La découverte de nouveaux Pandoravirus nous permettra d'obtenir des réponses supplémentaires, notamment sur la phylogénie de cette famille : ces nouveaux virus

s'intégreront-ils dans une des lignées existantes ou en formeront-ils une nouvelle comme chez les *Marseilleviridae* (DOS SANTOS *et al.*, 2016) ? La découverte de nouveaux virus permettra t-elle de diminuer la proportion d'ORFans dans les génomes ? Découvrira t-on des génomes encore plus complexes ?

La découverte de cette famille virale est récente, et il reste encore beaucoup de travail à faire, afin de répondre aux questions que nous nous posons. De nombreuses études ont été initiées et sont toujours en cours d'analyse : les résultats obtenus nous fourniront certainement des pistes supplémentaires pour améliorer notre compréhension de ces virus hors-normes, pour déterminer leur origine, ainsi que leur histoire évolutive (CLAVERIE *et al.*, 2016b).

5 Études de la composition des *Pandoraviridae* et *M. sibericum*

5.1 Contexte

La découverte des Pandoravirus il y a quelques années a révolutionné le domaine de la virologie : ces virions ont une morphologie inédite, et possèdent des caractéristiques étonnantes dans le monde viral. C'est également le cas de *Mollivirus sibericum*, isolé quelques années plus tard dans un échantillon de pergélisol sibérien vieux de 30 000 ans (LEGENDRE *et al.*, 2015).

Ces virus, qui appartiennent à des familles différentes, partagent toutefois une structure commune dans leur enveloppe. Cette structure est dense aux électrons et est située dans la couche intermédiaire des enveloppes de ces particules. Elle est constituée d'un réseau de fibres, qui forment des structures lamellaires parallèles entre elles (cf. Figure 5.1). Nous avons donc étudié la composition de cette structure inhabituelle.

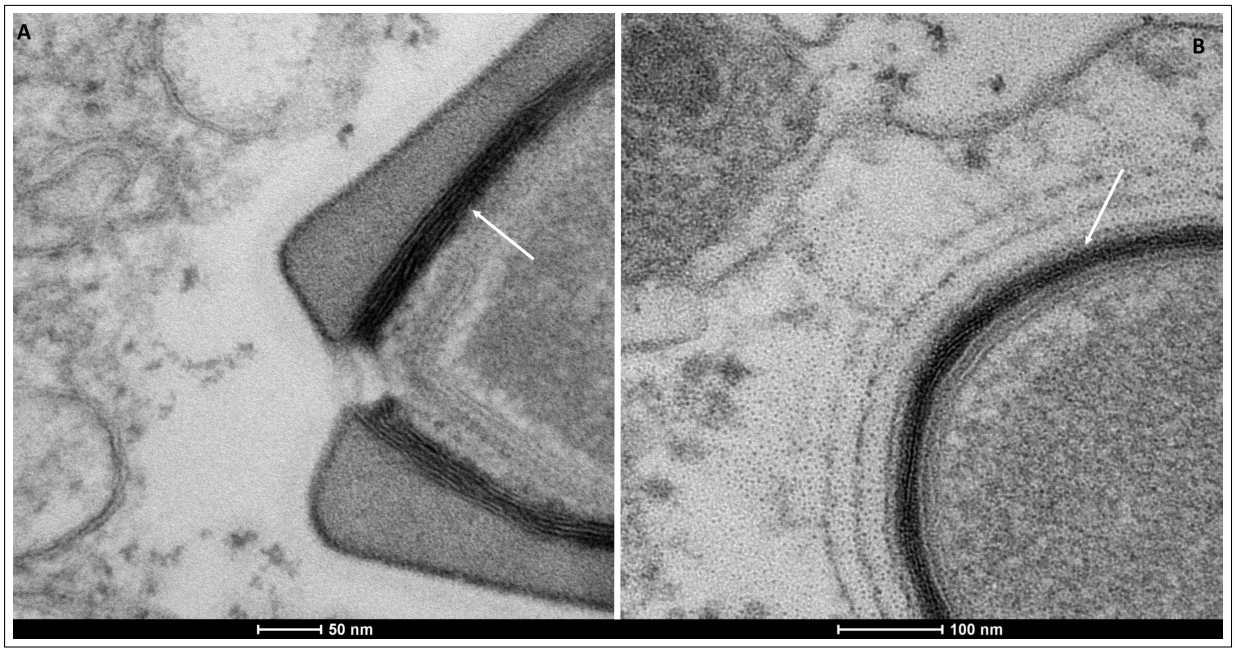


FIGURE 5.1 – Pandoravirus et *M. sibericum* observés au MET. (A) Pore apical d'un Pandoravirus, avec son enveloppe caractéristique composée de 3 couches, dont une avec des structures lamellaires (flèche blanche). (B) Enveloppe de *Mollivirus sibericum* avec des structures lamellaires similaires à celles de Pandoravirus (flèche blanche).

Parallèlement, l'hôte dans lequel ces virus ont été isolés, l'amibe *Acanthamoeba castellanii* peut être retrouvée sous deux formes : une forme active (le trophozoïte), et une

forme stationnaire (le kyste). Lorsque les conditions sont défavorables pour la cellule, c'est à dire en cas de stress, d'environnement toxique ou pauvre en nutriments, la cellule active alors son mécanisme de protection et s'enkyste (BOWERS *et al.*, 1969 ; SIDDIQUI *et al.*, 2012). Ce kyste est composé de deux parois : une paroi interne, l'endocyste, et une paroi externe, l'ectocyste, formées de réseaux de fibres plus ou moins denses (cf. Figure 5.2). Les fibres qui constituent l'endocyste (cf. Figure 5.2) sont essentiellement composées de cellulose (TOMLINSON *et al.*, 1962), et ressemblent aux structures observées dans les parois des Pandoravirus et de Mollivirus.

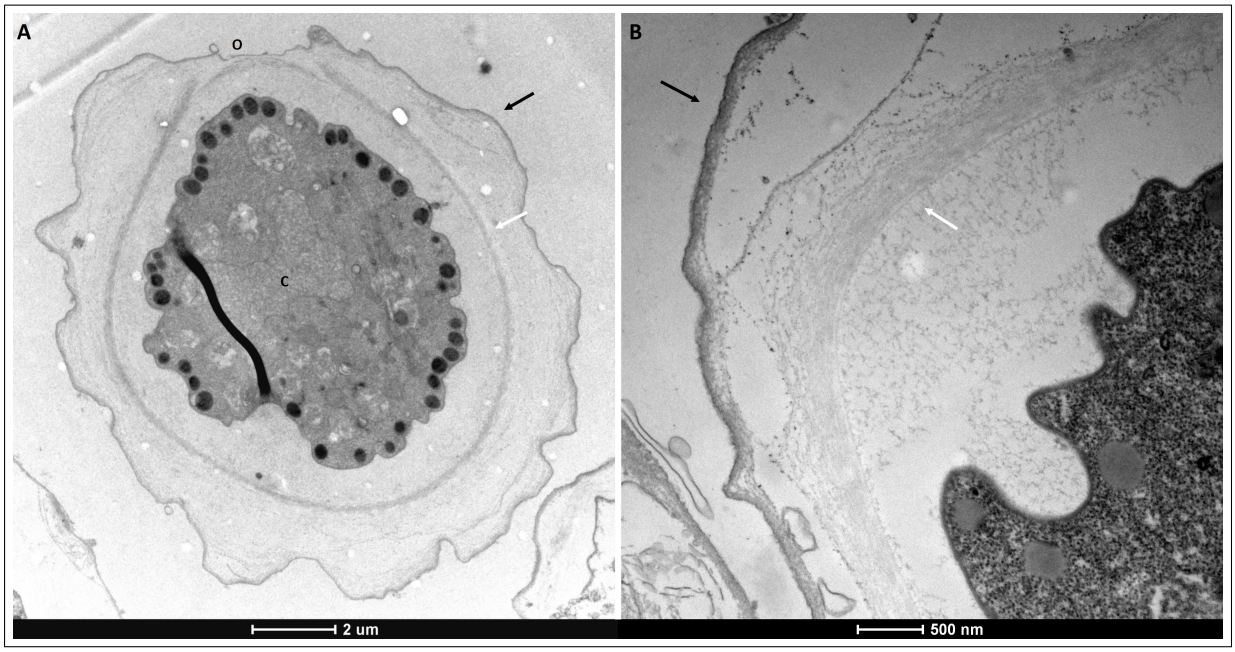


FIGURE 5.2 – Kystes d'*A. castellanii* observés au MET. (A) Le kyste est une cellule (c) entourée de 2 parois, l'endocyste (flèche blanche) et l'ectocyste (flèche noire), et qui possède un ostiole (o). (B) Agrandissement des parois du kyste, constituées d'un réseau de fibres (endocyste, flèche blanche ; ectocyste, flèche noire).

Étrangement, les kystes de l'amibe *Flamella pleistocenica*, qui a été isolée à partir d'un échantillon de pergélisol sibérien vieux de 30 000 ans, possèdent un réseau de fibres très compact (cf. Figure 5.3), qui rappelle les structures lamellaires des virions et dont la composition reste encore à déterminer (SHMAKOVA *et al.*, 2016).

Les réseaux de fibres des kystes d'amibes (cf. Figure 5.2 et 5.3), qui ressemblent étonnamment aux structures lamellaires des Pandoravirus et de Mollivirus (cf. Figure 5.1), sont principalement composés de cellulose : nous nous sommes donc demandé si les parois de ces virus pouvaient également contenir de la cellulose. Pour cela, une étude sur la composition de ces structures, à l'aide de différentes approches, a été initiée récemment en collaboration avec Henri-Pierre Fierobe, un chercheur de l'IMM, spécialiste de la cellulose.

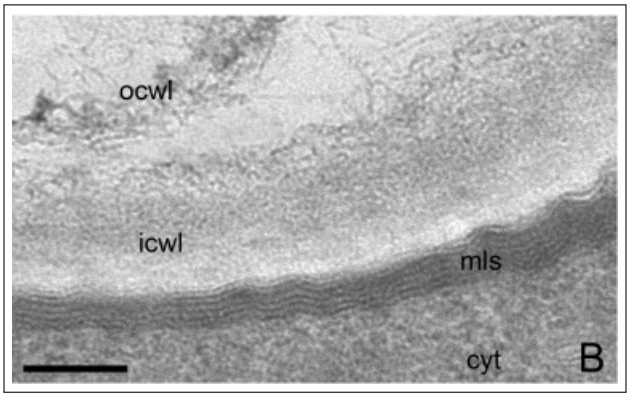


FIGURE 5.3 – Kyste de l'amibe *F. pleistocenica* observé au MET, qui est constitué de plusieurs couches (ectocyste, ocwl ; endocyste, icwl), dont une avec des structures lamellaires (mls), située en périphérie du cytoplasme (cyt). Barre d'échelle : 100 nm (SHMAKOVA *et al.*, 2016).

5.2 Résultats et discussion

Dans un premier temps, nous avons étudié le transcriptome des amibes *A. castellanii* infectées par ces virions, en nous focalisant sur les trois gènes qui codent pour des celluloses synthases chez l'amibe. Étonnamment, lorsque les amibes sont infectées par un Pandoravirus ou Mollivirus, ces gènes sont exprimés, et cette activation semble même être plus marquée aux temps du cycle qui correspondent à l'étape de synthèse de ces nouveaux virions (cf. Figure 5.4).

En revanche, chez Mimivirus ou *Megavirus chilensis*, dans lesquels ces structures lamellaires n'ont jamais été observées, ces gènes sont réprimés (cf. Figure 5.4). Cette inhibition des celluloses synthases amibiennes par les *Mimiviridae*, empêcherait donc ces amibes de s'enkyster au cours de l'infection, afin que les virions puissent être synthétisés. L'activation de ces gènes par les Pandoravirus et Mollivirus est toutefois surprenante : pourquoi les virions ne bloqueraient-ils pas l'enkystement ?

Nous avons recherché en parallèle la présence de celluloses synthases dans les protéomes des virions des Pandoravirus et de *Mollivirus sibericum*. Cependant aucune de ces protéines n'a pu être identifiée, suggérant que si de la cellulose était effectivement retrouvée dans les particules, celle-ci proviendrait de l'amibe et non des virions.

Nous avons donc réalisé une première étude, au cours de laquelle nous avons incubé des trophozoïtes (contrôle négatif) et des kystes d'*A. castellanii*, ainsi que des particules de Pandoravirus, de Mollivirus et de Mimivirus (contrôle négatif), en acide orthophosphorique (85%). L'acide orthophosphorique (H_3PO_4) est un agent "gonflant" qui permet notamment d'augmenter l'accessibilité de la cellulose aux cellulases. Après incubation, ces échantillons ont été dialysés, incubés en présence de cellulase, et les sucres solubles qui ont été libérés ont été analysés sur Dionex [1] (cf. 2.9 Chapitre 2).

1. Cette technique de chromatographie liquide à haute performance échangeuse d'anions est couplée à un système de détection ampérométrique pulsé : c'est une technique qualitative et quantitative, qui permet d'analyser directement les polysaccharides d'un échantillon.

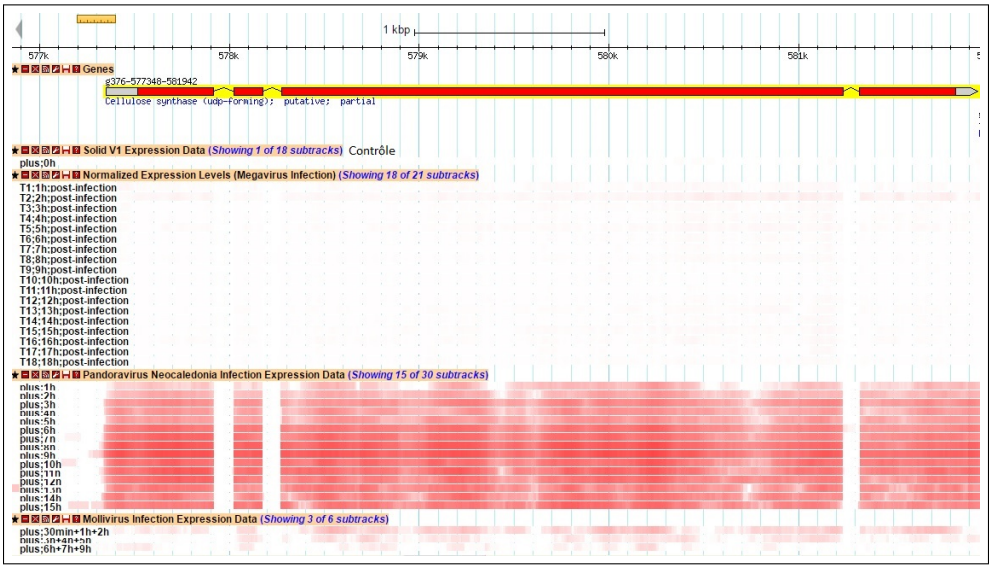


FIGURE 5.4 – Transcriptome d'un gène de la cellulose synthase d'*A. castellanii* avec la piste correspondant aux cellules non-infectées (Contrôle, pas d'expression), les pistes correspondant aux cellules infectées par *Megavirus chilensis* (pas d'expression), les pistes correspondant aux cellules infectées par *Pandoravirus neocaledonia* (expression maximale entre 8 et 9 hpi), et les pistes correspondant aux cellules infectées par *Mollivirus sibericum* (expression entre 30 minutes et 2 hpi), à différents temps du cycle.

Ces analyses ont permis de montrer que des grandes quantités de glucose et de cellobiose (un polysaccharide issu de la dégradation de la cellulose, constitué de deux molécules de glucose reliées par une liaison de type $\beta(1 \rightarrow 4)$) avaient été libérées dans l'échantillon correspondant aux kystes (cf. Figure 5.5).

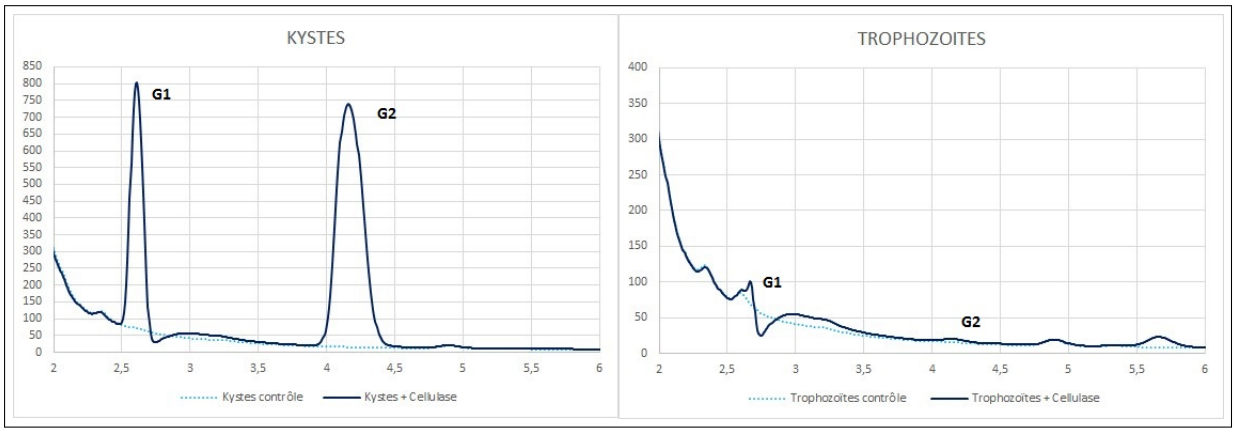


FIGURE 5.5 – Graphiques correspondant aux sucres libérés dans les cellules *A. castellanii*. A gauche, le graphique correspondant au glucose (G1) et au cellobiose (G2) libérés chez les kystes (courbe bleu foncé) par rapport au témoin sans enzyme (courbe bleu clair). A droite, le graphique correspondant au glucose (G1) et au cellobiose (G2) libérés chez les trophozoïtes (courbe bleu foncé) par rapport au témoin sans enzyme (courbe bleu clair).

En revanche, dans les échantillons de Mimivirus (cf. Figure 5.6) et de trophozoïtes (cf. Figure 5.5), seule une faible quantité de glucose a pu être détectée. Ces résultats

semblent donc valider notre hypothèse, ainsi que le choix de Mimivirus comme contrôle négatif.

Les mêmes expériences réalisées sur Pandoravirus (cf. Figure 5.6) et Mollivirus (cf. Figure 5.7) montrent la présence (en faible quantité par rapport aux kystes) de glucose et de cellobiose, ce qui va dans le sens de notre hypothèse.

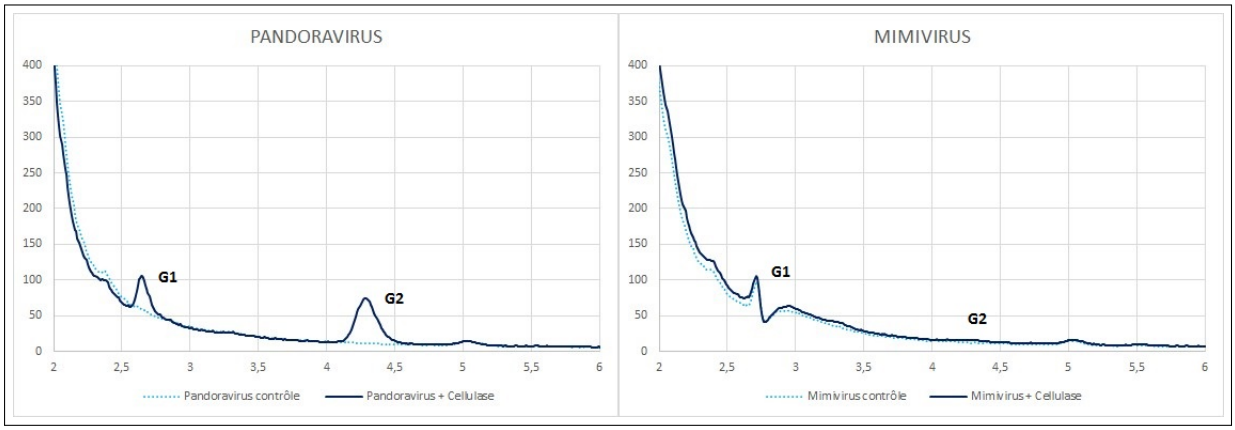


FIGURE 5.6 – Graphiques correspondant aux sucres libérés dans les Pandoravirus et Mimivirus. A gauche, le graphique correspondant au glucose (G1) et au cellobiose (G2) libérés chez les Pandoravirus (courbe bleu foncé) par rapport au témoin sans enzyme (courbe bleu clair). A droite, le graphique correspondant au glucose (G1) et au cellobiose (G2) libérés chez Mimivirus (courbe bleu foncé) par rapport au témoin sans enzyme (courbe bleu clair).

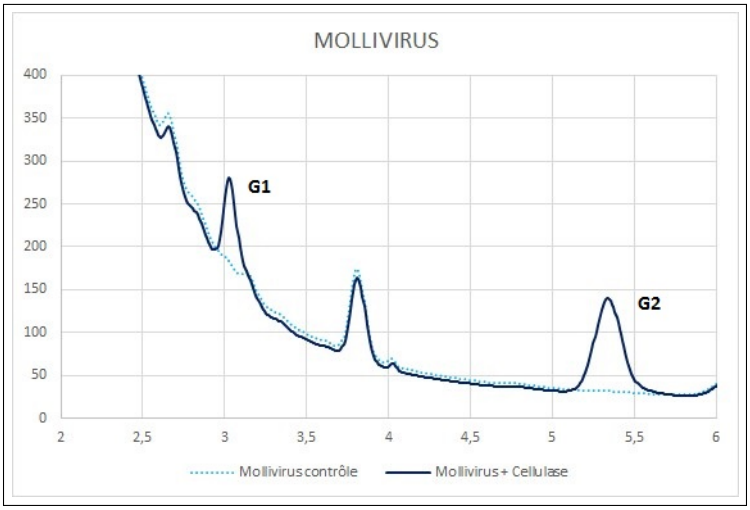


FIGURE 5.7 – Graphique correspondant aux sucres libérés chez Mollivirus : le graphique correspond au glucose (G1) et au cellobiose (G2) libérés chez Mollivirus (courbe bleu foncé) par rapport au témoin sans enzyme (courbe bleu clair). Le décalage des courbes, ainsi que la présence d'un pic supplémentaire entre les pics G1 et G2 (identique pour les deux échantillons), sont dus à des traces d'acide orthophosphorique dans l'échantillon, qui n'a pas été suffisamment dialysé.

Afin de vérifier si la cellulose est bien présente dans les parois des Pandoravirus et de Mollivirus comme nous le pensons, celles-ci ont été isolées par une lyse partielle des

particules (cf. 2.9 Chapitre 2). Cette lyse a été contrôlée au microscope optique : un affaissement des particules, qui paraissent également moins denses a pu être observé pour tous les virus (cf. Figure 5.8). Ces parois isolées ont été lavées, incubées en présence de cellulase puis analysées sur Dionex (cf. 2.9 Chapitre 2).

Toutefois, ces analyses n'ont pas permis d'identifier du glucose ou du cellobiose dans les échantillons. Ceci peut être dû à la présence de SDS et de DTT qui seraient restés dans ces échantillons malgré les lavages des particules, et qui auraient inhibé l'enzyme, ou alors ce traitement a éliminé les composants comprenant du glucose et du cellobiose. Cela peut également s'expliquer par le fait que la cellulose qui est probablement contenue dans les parois est protégée, et n'est donc pas suffisamment accessible à la cellulase.

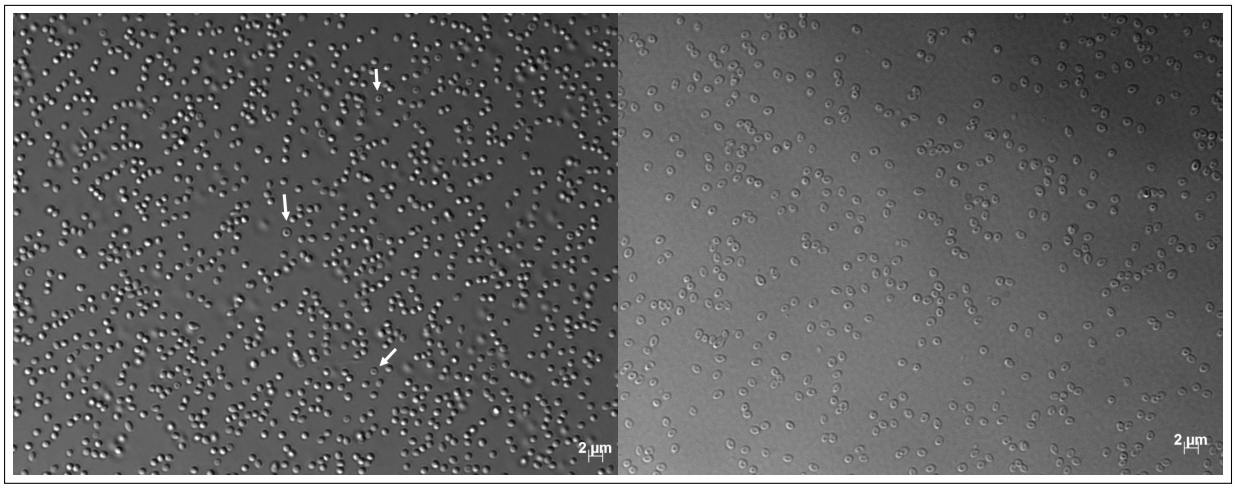


FIGURE 5.8 – Particules de Pandoravirus observées au microscope optique (objectif x63 et Optovar x1,6). A gauche, des particules entières de Pandoravirus purifié : les flèches blanches indiquent des particules vides, qui ont un aspect affaissé. A droite, des parois isolées de Pandoravirus après traitement : toutes les particules ont le même aspect affaissé, caractéristiques des particules vides.

Nous avons donc testé une autre méthode, réalisée en triplicat pour chaque virus. Les particules virales de Pandoravirus et de Mollivirus ont été cassées à l'aide de billes en acier inoxydable, et celles de Mimivirus (contrôle négatif), trop résistantes, ont été ouvertes avec un tampon composé de trypsine, de DTT et de triton. Ces particules cassées ont ensuite été incubées dans différentes conditions : en tampon d'incubation (témoin), en présence de Viscozyme®, ou de Celluclast®, ou de cellulase (cf. 2.9 Chapitre 2).

Après incubation et analyse des sucres libérés sur Dionex, les résultats sont étonnants : en ce qui concerne les particules incubées avec les mélanges d'enzymes Viscozyme® ou Celluclast®, aucune trace de cellobiose n'a pu être détectée. Une faible quantité de glucose a toutefois pu être observée, mais celle-ci peut être due au glucose déjà présent dans ces préparations enzymatiques. Ces enzymes avaient été préalablement dialysées afin d'éliminer tout le glucose, sans que cela n'inhibe leur activité : il semblerait cependant que ces dialyses n'aient pas été suffisantes.

Enfin, pour les particules incubées avec de la cellulase, du glucose et quelques traces de cellobiose ont pu être détectés chez Pandoravirus et Mollivirus : ce résultat n'a toutefois pas pu être reproduit dans tous les triplicats. Le cellobiose a pu être quantifié dans les 3 échantillons de Pandoravirus, à des concentrations différentes (0,091 μM, 0,218 μM et 0,588 μM), malgré un nombre de particules utilisées équivalent. En revanche, cette libération de cellobiose n'a pu être observée que sur un seul échantillon de Mollivirus parmi les 3, avec une concentration de 0,16 μM.

En ce qui concerne les triplicats de Mimivirus (contrôle négatif), aucune trace de glucose ou de cellobiose n'a pu être identifiée.

Ces résultats suggèrent la présence de cellulose dans les particules de Pandoravirus et de Mollivirus en faible quantité. Cette quantité de cellulose peut être expliquée par différentes hypothèses : peut-être y a t-il tout simplement peu de cellulose présente dans les particules virales, ou peut-être que cette cellulose n'est pas suffisamment accessible aux enzymes. Si la cellulose retrouvée dans les analyses de sucres est bien présente dans les parois des particules comme nous le pensons, celle-ci serait sous une forme très compacte, et serait donc probablement protégée par une gaine ou une membrane. Les cellulases auraient alors des difficultés à passer à travers cette protection, et/ou à dégrader cette cellulose très compacte. La technique de casse des particules que nous avons utilisée ici devra également être optimisée : les particules devront être encore plus abîmées pour que la cellulose soit accessible.

Afin de localiser la cellulose dans les particules virales, nous avons utilisé deux techniques de visualisation : la microscopie à épifluorescence et la microscopie électronique à transmission.

Dans un premier temps, les kystes et les trophozoïtes (contrôle négatif) d'*A. castellanii* ont été marqués avec du Calcofluor-white, un marqueur fluorescent qui se lie aux structures contenant de la cellulose (cf. 2.9 Chapitre 2). Après observation au microscope à épifluorescence, seule la périphérie des kystes est fortement marquée (cf. Figure 5.9).

Ce marquage a donc été réalisé sur les particules virales de Pandoravirus, Mollivirus et Mimivirus. Après observation, toutes les particules étaient fortement marquées, même celles de Mimivirus, notre contrôle négatif.

De plus, à cette résolution, le marquage des particules de Pandoravirus et de Mollivirus ne semble pas être spécifique, et le fort marquage des particules de Mimivirus peut être dû aux fibres qui entourent les particules, et qui auraient retenu le marqueur.

Par ailleurs, le Calcofluor-white se lie également aux structures qui contiennent de la chitine : le fort marquage des particules pourrait donc être dû à la présence de chitine dans celles-ci. Afin de valider ces hypothèses, la fluorescence des particules devrait être observée à l'aide d'un microscope qui possède une meilleure résolution, et où il serait

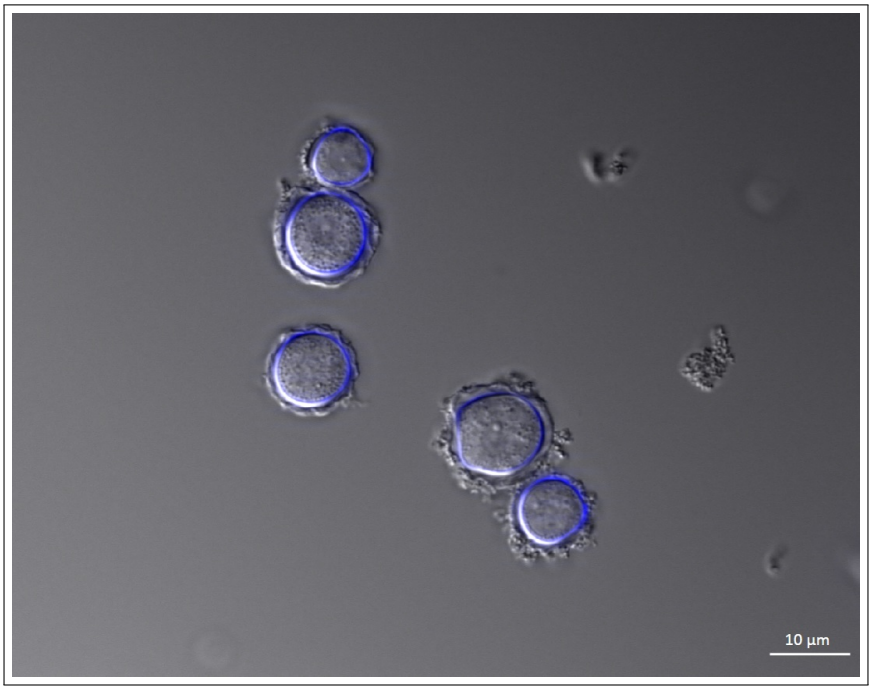


FIGURE 5.9 – Kystes d'*A. castellanii* marqués au Calcofluor-white et observés au microscope optique (objectif x63). Le marquage au Calcofluor-white est situé à l'intérieur des kystes, en périphérie, indiquant la présence de cellulose dans les parois.

possible de détecter dans quel plan se situe cette fluorescence, comme un microscope confocal ou un microscope 3D SR (super-résolution).

La deuxième technique utilisée pour localiser la cellulose dans les particules virales est la microscopie électronique à transmission. Pour cela, des kystes d' *A. castellanii* et des particules virales purifiées (Pandoravirus et Mollivirus) ont été inclus en résine Epon puis coupés. Ces coupes de 150 nm d'épaisseur, ont été incubées avec de la cellulase, du Viscozyme®, du Celluclast®, ou seulement du tampon (contrôle négatif) avant d'être observées au MET (cf. 2.9 Chapitre 2).

Les observations des coupes de kystes montrent que les réseaux de fibres composant la paroi interne ont été dégradés après incubation avec les enzymes, contrairement aux coupes "contrôle", où ces réseaux sont clairement visibles (cf. Figure 5.10).

En ce qui concerne les coupes de virus en revanche, après de multiples tentatives au cours desquelles les conditions d'incubation ont été variées, aucune dégradation par rapport aux coupes "contrôle" n'a pu être observée. En effet, les particules ne semblent pas être affectées, et la paroi constituée de structures lamellaires ne semble pas être digérée ou dégradée, même après 1 semaine d'incubation (cf. Figure 5.11).

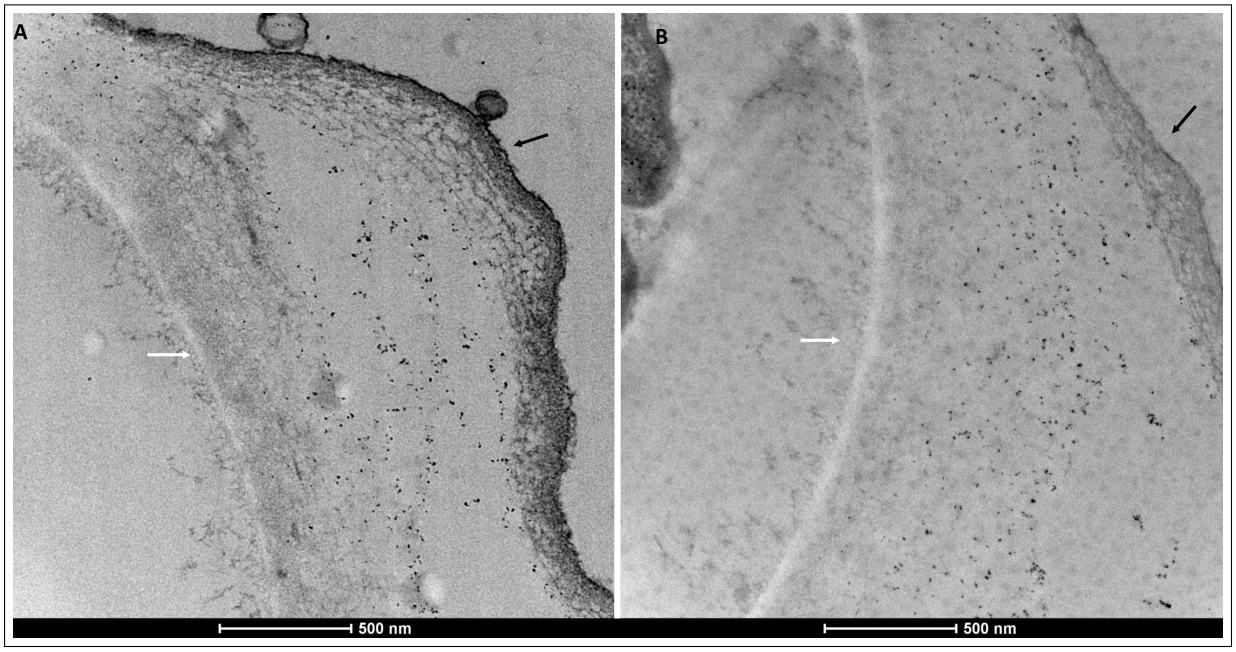


FIGURE 5.10 – Coupes de kystes d'*A. castellanii* observées au MET avant et après traitement à la cellulase 48 heures à 37°C. (A) Contrôle : les réseaux de fibres composant l'endocyste (flèche blanche) et l'ectocyste (flèche noire) sont bien visibles. (B) Après digestion des coupes avec de la cellulase, les réseaux de fibres de l'endocyste (flèche blanche) ne sont quasiment plus visibles, alors que ceux de l'ectocyste (flèche noire) peuvent toujours être observés.

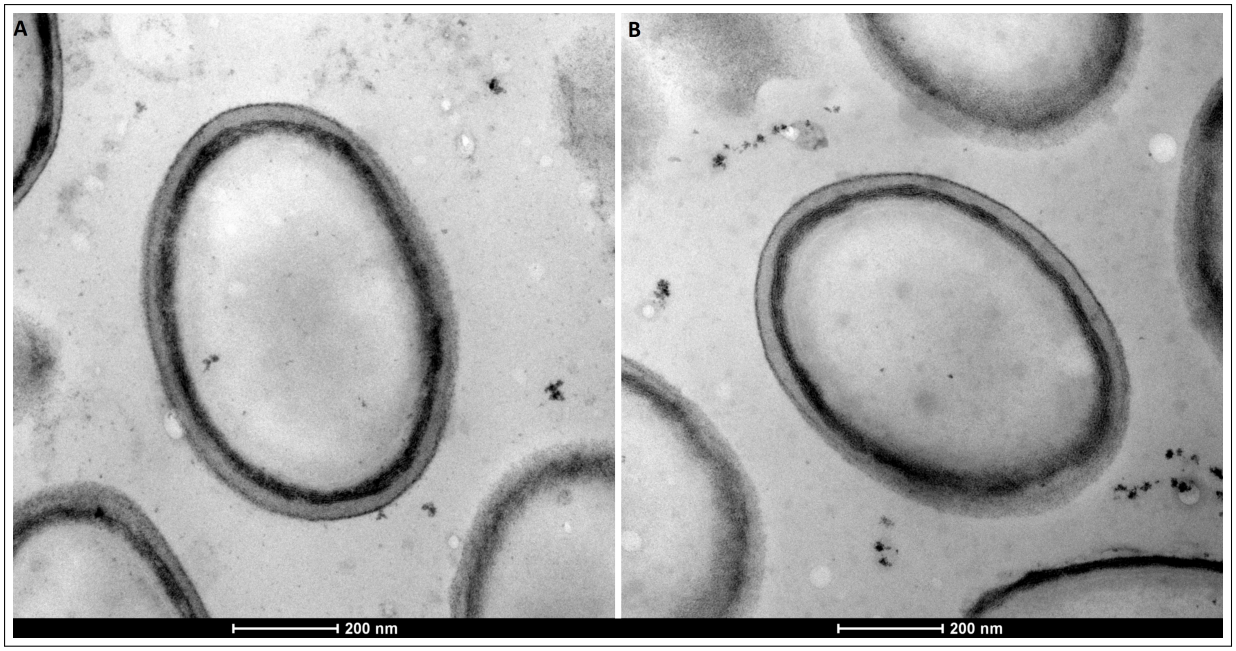


FIGURE 5.11 – Coupes de *P. neocaledonia* observées au MET avant et après traitement à la cellulase 48 heures à 37°C. (A) Contrôle : les structures lamellaires composant la couche intermédiaire de l'enveloppe de la particule sont visibles. (B) Après digestion des coupes avec de la cellulase, ces structures lamellaires peuvent toujours être observées, et les particules virales semblent être identiques au contrôle.

Parallèlement, les tampons dans lesquels ces coupes ont été incubées ont été récupérés et analysés sur Dionex, afin de vérifier si des sucres avaient été libérés au cours de ces incubations. Toutefois, aucune trace de glucose ou de cellobiose n'a pu être observée dans ces échantillons. Ces résultats sont donc peu concluants en ce qui concerne les particules virales : les protocoles utilisés devront être optimisés, en variant encore les conditions d'incubation, ou le type de résine utilisée, ou même en ajoutant peut-être des traitements supplémentaires.

Il reste encore une dernière étude à réaliser concernant la localisation de cette cellulose. En effet, l'utilisation de billes d'or qui seraient couplées à de la cellulase, nous permettrait d'identifier précisément où est située cette cellulose dans les particules. Lors de l'incubation de ces coupes, la cellulase se lierait à la cellulose qui est présente dans les échantillons, et les billes d'or pourraient être visualisées au MET. Cette étude nous permettrait donc de vérifier que les structures lamellaires des parois de ces virus sont bien composées de cellulose, comme nous le pensons.

5.3 Conclusion

Cette étude a été menée afin d'identifier la composition de la structure unique des parois des Pandoravirus (PHILIPPE *et al.*, 2013) et de *Mollivirus sibericum*, un virus isolé à partir d'un échantillon de pergélisol vieux de 30 000 ans (LEGENDRE *et al.*, 2015). Leur ressemblance étonnante avec des structures retrouvées dans des kystes d'amibes (BOWERS *et al.*, 1969 ; SHMAKOVA *et al.*, 2016), essentiellement composés de cellulose (TOMLINSON *et al.*, 1962), nous a incité à rechercher la présence de cellulose dans les virions.

Différentes approches ont été réalisées, et les résultats obtenus sont assez encourageants pour la suite.

En effet, du glucose, ainsi que du cellobiose, produits caractéristiques de la dégradation de la cellulose, ont pu être retrouvés chez les Pandoravirus et Mollivirus. Ces résultats suggèrent qu'il y a effectivement de la cellulose dans ces virions. Toutefois, le manque de reproductibilité entre les expériences et les différences de concentrations qui ont été retrouvées dans les mêmes échantillons, montrent que le protocole utilisé peut encore être optimisé.

Cette cellulose, probablement contenue dans les parois des virions, semble très compacte, et est certainement protégée. Notre technique de casse des parois virales, qui est restée assez modérée, pourrait donc être plus drastique, afin d'améliorer l'accessibilité de cette cellulose à la cellulase, pour pouvoir en libérer une quantité plus importante. D'un autre côté, ces structures lamellaires pourraient ne pas être constituées uniquement de cellulose, et en contenir seulement en faible quantité, ce qui expliquerait les concentrations que nous retrouvons lors de nos analyses.

Notre étude nous a permis de déterminer qu'il y a probablement de la cellulose dans les virions, mais sa localisation reste encore à déterminer. Les marquages fluorescents avec du Calcofluor-white, un marqueur de la cellulose et de la chitine, fonctionnent sur les kystes, mais restent encore à optimiser sur les particules virales. L'utilisation d'un microscope confocal nous permettrait d'observer ces marquages avec une meilleure résolution, et dans tous les plans : cela pourrait donc nous donner une indication sur la localisation de cette cellulose dans les particules.

En parallèle, des cellules infectées pourraient également être marquées lors de l'étape de synthèse des virions : en effet, les membranes qui composent les particules semblent provenir de membranes cellulaires recyclées. Celles-ci pourraient contenir des celluloses synthases, nécessaires à l'enkystement de ces amibes, qui seraient activées au cours de l'infection, comme nous l'avons vu dans les transcriptomes des virions. L'activation de ces protéines entraînerait alors la synthèse de la cellulose en périphérie des particules, formant ainsi les structures lamellaires, caractéristiques des Pandoravirus et de Mollivirus. Ces observations nous permettraient donc de vérifier que la cellulose est bien

impliquée dans ce mécanisme de synthèse des virions, ce qui participerait à cette impression de “tricotage” des particules (PHILIPPE *et al.*, 2013 ; LEGENDRE *et al.*, 2015).

De plus, l'utilisation de billes d'or couplées à la cellulase, nous permettrait d'obtenir une localisation plus précise de cette cellulose dans les particules lors de leur observation au MET. Cette étude permettrait de répondre aux questions : cette cellulose est-elle vraiment située dans l'enveloppe des virions, ou dans le compartiment interne des particules, qui apparaît assez diffus et peu dense aux électrons ?

Par ailleurs, l'absence de celluloses synthases chez les virions, et leur activation chez l'amibe au cours de l'infection, nous laissent penser que les Pandoravirus et Mollivirus détournent la machinerie de leur hôte pour synthétiser leurs particules. Si la présence de cellulose est confirmée dans les parois de ces particules, cela signifierait que les virions utilisent les gènes de l'hôte pour fabriquer le véhicule qui leur permet de propager leur génome, ce qui est inédit dans le monde viral. Habituellement, les virions codent eux-mêmes les gènes qui composent leur capside, comme par exemple la “*Major Capsid Protein*” (ou protéine majeure de capside), retrouvée chez tous les virus à ADN infectant les eucaryotes (à l'exception des Pandoravirus).

Le fait que ces virions ne codent pas les gènes nécessaires à la formation de leur particule, et qu'ils détournent la machinerie de la cellule pour le faire, soulève de nombreuses questions sur leur place au sein du monde viral, sur leur origine, ainsi que sur leur histoire évolutive.

6 Conclusion et Perspectives

Depuis la découverte du premier virus, et jusque dans les années 2000, les virus étaient considérés comme des parasites, qui ne pouvaient pas être observés au microscope optique, et qui étaient capables de passer à travers le filtre de Chamberland.

La découverte de Mimivirus, ainsi que sa caractérisation en tant que virus géant en 2003 a complètement bouleversé cette vision : Mimivirus peut être observé au microscope optique et est retenu par le filtre de Chamberland. De nombreux autres virus géants ont été découverts depuis, démontrant alors que les méthodes utilisées pour isoler les virus, comme la filtration des échantillons sur 0,22 μm, ne permettaient pas d'isoler des grands virus à ADN (ou des virus géants), et avaient complètement retardé la découverte de ces virus extraordinaires.

Cette découverte a donc remis en question la définition même du mot "*virus*", et a soulevé de nombreuses interrogations sur l'origine et l'histoire évolutive de ces virus.

Au cours de ces dernières années, différentes familles de virus géants (ou de grands virus à ADN) ont pu être identifiées. Ces virus ont été isolés à partir de tout type d'échantillon (boue, terre, sédiments marins et d'eau douce, pergélisol, ...), et dans le monde entier (Europe, Afrique, Amérique du Sud, Australie, Japon, Nouvelle-Calédonie, ...), ce qui démontre qu'ils sont ubiquitaires, et qu'il en reste beaucoup à découvrir.

Ces virus infectent des amibes (*A. castellanii*, *A. polyphaga*, *V. vermiformis*), mais également des algues unicellulaires (*E. huxleyi*, *C. roenbergensis*, *P. globosa*, ...), suggérant que la gamme d'hôtes qui est testée afin d'isoler de nouveaux virus, pourrait être élargie à d'autres genres d'amibes ou d'algues.

Certains virus ont également été isolés à partir d'échantillons prélevés sur des patients (malades ou sains), mais leur infectivité n'a pas encore été formellement démontrée : ces patients auraient pu être porteurs d'amibes infectées, qui auraient libérés les virions isolés lors des prélèvements. Des études complémentaires sont nécessaires afin de déterminer la pathogénicité de ces virus envers les hommes.

Au cours de ce travail de thèse, j'ai isolé deux virus qui appartiennent à deux familles différentes : le premier, Noumeavirus, fait partie de la famille des *Marseilleviridae* (Aherfi *et al.*, 2014b). Sa capside icosaédrique de 200 nm de diamètre le situe à la frontière entre les grands virus à ADN et les virus géants. Les études que nous avons menées sur ce virus nous ont permis de montrer que la protéine la plus abondante dans ses particules n'est pas la protéine majeure de capside comme nous le pensions. Il s'agit d'une protéine qui est très conservée chez les *Marseilleviridae*, et dont la nature et la

fonction restent encore à découvrir. De plus, les analyses du protéome des particules, ainsi que des observations du cycle infectieux au MET et au microscope à épifluorescence, nous ont permis de constater que pour initier la transcription des gènes précoces, un recrutement transitoire des protéines nucléaires auprès de l'usine virale en cours de développement est nécessaire.

Le deuxième virus isolé en parallèle au cours de cette thèse, appartient à la famille des *Pandoraviridae*, et a été nommé *Pandoravirus neocaledonia*. Sa caractérisation nous a notamment permis de constater que ce virus était le plus éloigné des autres membres de cette famille. Après la découverte de nouveaux Pandoravirus, nous avons initié une étude comparative de tous ces virus, qui nous a notamment permis d'établir une phylogénie : il semblerait que *P. neocaledonia* et *P. D1* forment une deuxième lignée au sein de la famille des *Pandoraviridae*.

Nous avons également réalisé une étude du transcriptome orienté de ces virus, qui nous a permis de réduire le nombre de gènes codants et de déterminer que ces gènes possèdent de longs UTR et sont séparés par de longues régions intergéniques. Enfin, ces virus qui possèdent une morphologie identique, ont tout de même des différences au niveau de leur composition en protéines ou encore de leurs cycles infectieux, qui nous permettent de les différencier. L'étude qui a été amorcée sur leur enveloppe, qui serait composée de cellulose synthétisée par les protéines de la cellule hôte, bousculerait encore un dogme de la virologie : les Pandoravirus et Mollivirus seraient les premiers virus à détourner la machinerie cellulaire pour fabriquer leur "capside".

Nos études sur ces virus nous ont permis d'obtenir des indices, grâce auxquels nous pouvons établir des scénarios sur leur origine et leur histoire évolutive (Claverie *et al.*, 2016b), et actualiser la définition du "vivant" (Forterre *et al.*, 2016).

L'étude des virus géants est un champ émergent dans le domaine de la virologie, où il reste encore beaucoup de choses à découvrir. Les nombreuses études menées sur ces virus sont difficiles, notamment en raison d'un manque d'outils génétiques, ce qui demande plus de temps et représente un vrai défi. Plusieurs analyses par différentes approches ont été initiées au cours de cette thèse (génomique, transcriptomique, protéomique, fluorescence, MET), et restent encore à finaliser, afin de pouvoir répondre aux nombreuses questions qui ont été soulevées par ces découvertes exceptionnelles.

Bibliographie

ABERGEL, C., LEGENDRE, M. et CLAVERIE, J.-M. (2015). « The rapidly expanding universe of giant viruses : Mimivirus, Pandoravirus, Pithovirus and Mollivirus ». *FEMS microbiology reviews* 39.6, p. 779–796.

ADAMO, S. A. *et al.* (2014). « A viral aphrodisiac in the cricket Gryllus texensis ». *Journal of Experimental Biology* 217.11, p. 1970–1976.

AHERFI, S. *et al.* (2014a). « Complete genome sequence of Tunisvirus, a new member of the proposed family Marseilleviridae ». *Archives of Virology* 159.9, p. 2349–2358.

AHERFI, S. *et al.* (2014b). « The expanding family Marseilleviridae ». *Virology* 466-467, p. 27–37.

AHERFI, S. *et al.* (2016a). « Giant Viruses of Amoebas : An Update ». *Frontiers in Microbiology* 7, p. 349.

AHERFI, S. *et al.* (2016b). « Marseillevirus in lymphoma : a giant in the lymph node ». *The Lancet. Infectious Diseases*.

ANTWERPEN, M. H. *et al.* (2015). « Whole-genome sequencing of a pandoravirus isolated from keratitis-inducing acanthamoeba ». *Genome Announcements* 3.2.

ARANTES, T. S. *et al.* (2016). « The Large Marseillevirus Explores Different Entry Pathways by Forming Giant Infectious Vesicles ». *Journal of Virology* 90.11, p. 5246–5255.

ARSLAN, D. *et al.* (2011). « Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae ». *Proceedings of the National Academy of Sciences of the United States of America* 108.42, p. 17486–17491.

BAHAR, M. W. *et al.* (2011). « Insights into the Evolution of a Complex Virus from the Crystal Structure of Vaccinia Virus D13 ». *Structure(London, England :1993)* 19.7-12, p. 1011–1020.

BALDACCHINO, F. (2011). *Histoire de l'épidémiologie de la fièvre jaune : Synthèse scientifique*. Editions universitaires europeennes EUE.

BALTIMORE, D. (1971). « Expression of animal virus genomes. » *Bacteriological Reviews* 35.3, p. 235–241.

BÂNDEA, C. I. (1983). « A new theory on the origin and the nature of viruses ». *Journal of Theoretical Biology* 105.4, p. 591–602.

BATEMAN, E. (2010). « Expression plasmids and production of EGFP in stably transfected Acanthamoeba ». *Protein Expression and Purification* 70.1, p. 95–100.

BELL, P. J. (2001). « Viral eukaryogenesis : was the ancestor of the nucleus a complex DNA virus ? » *Journal of Molecular Evolution* 53.3, p. 251–256.

BENARROCH, D., SMITH, P. et SHUMAN, S. (2008). « Characterization of a trifunctional mimivirus mRNA capping enzyme and crystal structure of the RNA triphosphatase domain ». *Structure (London, England : 1993)* 16.4, p. 501–512.

BERCHE, P. (2007). *Une histoire des microbes*. John Libbey Eurotext.

BESEMER, J., LOMSADZE, A. et BORODOVSKY, M. (2001). « GeneMarkS : a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions ». *Nucleic Acids Research* 29.12, p. 2607–2618.

BORDENAVE, G. (2003). « Louis Pasteur (1822–1895) ». *Microbes and Infection* 5.6, p. 553–560.

BOUGHALMI, M. *et al.* (2013a). « First isolation of a giant virus from wild Hirudo medicinalis leech : Mimiviridae isolation in Hirudo medicinalis ». *Viruses* 5.12, p. 2920–2930.

BOUGHALMI, M. *et al.* (2013b). « High-throughput isolation of giant viruses of the Mimiviridae and Marseilleviridae families in the Tunisian environment ». *Environmental Microbiology* 15.7, p. 2000–2007.

BOWERS, B. et KORN, E. D. (1969). « The fine structure of Acanthamoeba castellanii (Neff strain). II. Encystment ». *The Journal of Cell Biology* 41.3, p. 786–805.

BOYER, M. *et al.* (2009). « Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms ». *Proceedings of the National Academy of Sciences of the United States of America* 106.51, p. 21848–21853.

BOYER, M. *et al.* (2011). « Mimivirus shows dramatic genome reduction after intraamoebal culture ». *Proceedings of the National Academy of Sciences of the United States of America* 108.25, p. 10296–10301.

BRAMEIER, M., KRINGS, A. et MACCALLUM, R. M. (2007). « NucPred–predicting nuclear localization of proteins ». *Bioinformatics (Oxford, England)* 23.9, p. 1159–1160.

BYRNE, D. *et al.* (2009). « The polyadenylation site of Mimivirus transcripts obeys a stringent 'hairpin rule' ». *Genome Research* 19.7, p. 1233–1242.

CAMPOS, R. K. *et al.* (2014). « Samba virus : a novel mimivirus from a giant rain forest, the Brazilian Amazon ». *Virology Journal* 11, p. 95.

CHAMBERLAND, C. (1884). « Sur un filtre donnant de l'eau physiologiquement pure ». *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, p. 247–248.

CHENG, X.-W. *et al.* (2008). « Ascovirus and its evolution ». *Virologica Sinica* 22.2, p. 137.

CLAVERIE, J.-M. (2006). « Viruses take center stage in cellular evolution ». *Genome Biology* 7.6, p. 110.

CLAVERIE, J.-M. et ABERGEL, C. (2009). « Mimivirus and its virophage ». *Annual Review of Genetics* 43, p. 49–66.

CLAVERIE, J.-M. et ABERGEL, C. (2013). « Open questions about giant viruses ». *Advances in Virus Research* 85, p. 25–56.

CLAVERIE, J.-M. et ABERGEL, C. (2016a). « CRISPR-Cas-like system in giant viruses : why MIMIVIRE is not likely to be an adaptive immune system ». *Virologica Sinica* 31.3, p. 193–196.

CLAVERIE, J.-M. et ABERGEL, C. (2016b). « Giant viruses : The difficult breaking of multiple epistemological barriers ». *Studies in History and Philosophy of Biological and Biomedical Sciences* 59, p. 89–99.

CLAVERIE, J.-M. *et al.* (2006). « Mimivirus and the emerging concept of "giant" virus ». *Virus Research* 117.1, p. 133–144.

COHEN, G. *et al.* (2011). « Ameba-associated Keratitis, France ». *Emerging Infectious Diseases* 17.7, p. 1306–1308.

COLSON, P., LA SCOLA, B. et RAOULT, D. (2013a). « Giant viruses of amoebae as potential human pathogens ». *Intervirology* 56.6, p. 376–385.

COLSON, P. *et al.* (2013b). « "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses ». *Archives of Virology* 158.12, p. 2517–2521.

COX, J. et MANN, M. (2008). « MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification ». *Nature Biotechnology* 26.12, p. 1367–1372.

DALEY, G. M. *et al.* (2014). « Mycoplasma genitalium : a review ». *International journal of STD & AIDS* 25.7, p. 475–487.

DARE, R. K., CHITTAGANPITCH, M. et ERDMAN, D. D. (2008). « Screening pneumonia patients for mimivirus ». *Emerging Infectious Diseases* 14.3, p. 465–467.

DIXON, L. K. *et al.* (2013). « African swine fever virus replication and genomics ». *Virus Research* 173.1, p. 3–14.

DOELLINGER, J., SCHAADE, L. et NITSCHE, A. (2015). « Comparison of the Cowpox Virus and Vaccinia Virus Mature Virion Proteome : Analysis of the Species- and Strain-Specific Proteome ». *PloS One* 10.11, e0141527.

DORNAS, F. P. *et al.* (2015). « Isolation of new Brazilian giant viruses from environmental samples using a panel of protozoa ». *Frontiers in Microbiology* 6, p. 1086.

DORNAS, F. P. *et al.* (2016). « A Brazilian Marseillevirus Is the Founding Member of a Lineage in Family Marseilleviridae ». *Viruses* 8.3, p. 76.

DOS SANTOS, R. N. *et al.* (2016). « A new marseillevirus isolated in Southern Brazil from Limnoperna fortunei ». *Scientific Reports* 6, p. 35237.

DOUTRE, G. *et al.* (2014). « Genome analysis of the first Marseilleviridae representative from Australia indicates that most of its genes contribute to virus fitness ». *Journal of Virology* 88.24, p. 14340–14349.

DOUTRE, G. *et al.* (2015). « Complete Genome Sequence of a New Member of the Marseilleviridae Recovered from the Brackish Submarine Spring in the Cassis Port-Miou Calanque, France ». *Genome Announcements* 3.6.

DUNIGAN, D. D. *et al.* (2012). « Paramecium bursaria chlorella virus 1 proteome reveals novel architectural and regulatory features of a giant virus ». *Journal of Virology* 86.16, p. 8821–8834.

DURAFFOUR, S., ANDREI, G. et SNOECK, R. (2008). « Nouvelles armes thérapeutiques contre les poxvirus ». *Virologie* 12.5, p. 323–338.

ENGLISH, A. C. *et al.* (2012). « Mind the gap : upgrading genomes with Pacific Biosciences RS long-read sequencing technology ». *PloS One* 7.11, e47768.

ERIVES, A. J. (2015). « Eukaryotic core histone diversification in light of the histone doublet and DNA topo II genes of Marseilleviridae ». *bioRxiv*, p. 022236.

FEDERICI, B. A. (1983). « Enveloped double-stranded DNA insect virus with novel structure and cytopathology ». *Proceedings of the National Academy of Sciences of the United States of America* 80.24, p. 7664–7668.

FENNER, F. (1975). « The Classification and Nomenclature of Viruses ». *Intervirology* 6.1, p. 1–12.

FINN, R. D. *et al.* (2016). « The Pfam protein families database : towards a more sustainable future ». *Nucleic Acids Research* 44.D1, p. D279–285.

FISCHER, M. G. et SUTTLE, C. A. (2011). « A virophage at the origin of large DNA transposons ». *Science (New York, N.Y.)* 332.6026, p. 231–234.

FISCHER, M. G. *et al.* (2010). « Giant virus with a remarkable complement of genes infects marine zooplankton ». *Proceedings of the National Academy of Sciences of the United States of America* 107.45, p. 19508–19513.

FISCHER, M. G. *et al.* (2014). « The virion of Cafeteria roenbergensis virus (CroV) contains a complex suite of proteins for transcription and DNA repair ». *Virology* 466-467, p. 82–94.

FORTERRE, P. (2016). « To be or not to be alive : How recent discoveries challenge the traditional definitions of viruses and life ». *Studies in History and Philosophy of Biological and Biomedical Sciences* 59, p. 100–108.

FORTERRE, P. et GAÏA, M. (2016). « Giant viruses and the origin of modern eukaryotes ». *Current Opinion in Microbiology* 31, p. 44–49.

FRIDMANN-SIRKIS, Y. *et al.* (2016). « Efficiency in Complexity : Composition and Dynamic Nature of Mimivirus Replication Factories ». *Journal of Virology*.

GAHLMANN, A. et MOERNER, W. E. (2014). « Exploring bacterial cell biology with single-molecule tracking and super-resolution imaging ». *Nature Reviews Microbiology* 12.1, p. 9–22.

GAIA, M. *et al.* (2013). « Broad spectrum of mimiviridae virophage allows its isolation using a mimivirus reporter ». *PloS One* 8.4, e61912.

GAIA, M. *et al.* (2014). « Zamilon, a novel virophage with Mimiviridae host specificity ». *PloS One* 9.4, e94923.

GALLOT-LAVALLÉE, L. *et al.* (2015). « The 474-Kilobase-Pair Complete Genome Sequence of CeV-01B, a Virus Infecting Haptolina (Chrysochromulina) ericina (Prymnesiophyceae) ». *Genome Announcements* 3.6.

GAUDET, P. *et al.* (2007). « Transformation of Dictyostelium discoideum with plasmid DNA ». *Nature Protocols* 2.6, p. 1317–1324.

GHEDIN, E. et CLAVERIE, J.-M. (2005). « Mimivirus relatives in the Sargasso sea ». *Virology Journal* 2, p. 62.

GOODMAN, J. L. (2013). « Marseillevirus, blood safety, and the human virome ». *The Journal of Infectious Diseases* 208.7, p. 1039–1041.

GUINDON, S. *et al.* (2010). « New algorithms and methods to estimate maximum-likelihood phylogenies : assessing the performance of PhyML 3.0 ». *Systematic Biology* 59.3, p. 307–321.

HOFFMANN, R. *et al.* (1998). « Archaea like endocytobiotic organisms isolated from Acanthamoeba SP (GR II) ». *ResearchGate* 12.3, p. 185–188.

HOLMES, F. O. (1940). « Handbook of Phytopathogenic Viruses ». *Annals of the Entomological Society of America* 33.1, p. 226–226.

HUGHES, A. L., IRAUSQUIN, S. et FRIEDMAN, R. (2010). « The Evolutionary Biology of Poxviruses ». *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 10.1, p. 50.

INGOLIA, N. T. (2014). « Ribosome profiling : new views of translation, from single codons to genome scale ». *Nature Reviews. Genetics* 15.3, p. 205–213.

IYER, L. M., ARAVIND, L. et KOONIN, E. V. (2001). « Common Origin of Four Diverse Families of Large Eukaryotic DNA Viruses ». *Journal of Virology* 75.23, p. 11720–11734.

JONES, P. *et al.* (2014). « InterProScan 5 : genome-scale protein function classification ». *Bioinformatics (Oxford, England)* 30.9, p. 1236–1240.

KÄLL, L., KROGH, A. et SONNHAMMER, E. L. L. (2004). « A combined transmembrane topology and signal peptide prediction method ». *Journal of Molecular Biology* 338.5, p. 1027–1036.

KARSENTI, E. *et al.* (2011). « A holistic approach to marine eco-systems biology ». *PLoS biology* 9.10, e1001177.

KATZOURAKIS, A. et ASWAD, A. (2014). « The origins of giant viruses, virophages and their relatives in host genomes ». *BMC Biology* 12, p. 51.

KAUSCHE, G. A., PFANKUCH, E. et RUSKA, H. (1939). « Die Sichtbarmachung von pflanzlichem Virus im Übermikroskop ». *Naturwissenschaften* 27.18, p. 292–299.

KESSIN, R. H. (2001). *Dictyostelium : Evolution, Cell Biology, and the Development of Multicellularity*. Cambridge University Press.

KHALIL, J. Y. B. *et al.* (2016). « High-Throughput Isolation of Giant Viruses in Liquid Medium Using Automated Flow Cytometry and Fluorescence Staining ». *Frontiers in Microbiology* 7, p. 26.

KIEŁBASA, S. M. *et al.* (2011). « Adaptive seeds tame genomic sequence comparison ». *Genome Research* 21.3, p. 487–493.

KLOSE, T. *et al.* (2010). « The Three-Dimensional Structure of Mimivirus ». *Intervirology* 53.5, p. 268–273.

KLOSE, T. *et al.* (2015). « A Mimivirus Enzyme that Participates in Viral Entry ». *Structure (London, England : 1993)* 23.6, p. 1058–1065.

KLOSE, T. *et al.* (2016). « Structure of faustovirus, a large dsDNA virus ». *Proceedings of the National Academy of Sciences of the United States of America* 113.22, p. 6206–6211.

KUTIKHIN, A. G., YUZHALIN, A. E. et BRUSINA, E. B. (2014). « Mimiviridae, Marseilleviridae, and virophages as emerging human pathogens causing healthcare-associated infections ». *GMS hygiene and infection control* 9.2, Doc16.

KUZNETSOV, Y. G. *et al.* (2013). « Morphogenesis of mimivirus and its viral factories : an atomic force microscopy study of infected cells ». *Journal of Virology* 87.20, p. 11200–11213.

LA SCOLA, B. *et al.* (2003). « A giant virus in amoebae ». *Science (New York, N.Y.)* 299.5615, p. 2033.

LA SCOLA, B. *et al.* (2008). « The virophage as a unique parasite of the giant mimivirus ». *Nature* 455.7209, p. 100–104.

LAGIER, J. .-.-C. *et al.* (2012). « Microbial culturomics : paradigm shift in the human gut microbiome study ». *Clinical Microbiology and Infection* 18.12, p. 1185–1193.

LANGMEAD, B. *et al.* (2009). « Ultrafast and memory-efficient alignment of short DNA sequences to the human genome ». *Genome Biology* 10.3, R25.

LECOQ, H. (2001). « Discovery of the first virus, the tobacco mosaic virus : 1892 or 1898 ? » *Comptes rendus de l'Académie des sciences. Série III, Sciences de la vie* 324.10, p. 929–933.

LEGENDRE, M. *et al.* (2010). « mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus ». *Genome Research* 20.5, p. 664–674.

LEGENDRE, M. *et al.* (2014). « Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology ». *Proceedings of the National Academy of Sciences of the United States of America* 111.11, p. 4274–4279.

LEGENDRE, M. *et al.* (2015). « In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting Acanthamoeba ». *Proceedings of the National Academy of Sciences of the United States of America* 112.38, E5327–5335.

LEVASSEUR, A. *et al.* (2016a). « Comparison of a Modern and Fossil Pithovirus Reveals Its Genetic Conservation and Evolution ». *Genome Biology and Evolution* 8.8, p. 2333–2339.

LEVASSEUR, A. *et al.* (2016b). « MIMIVIRE is a defence system in mimivirus that confers resistance to virophage ». *Nature* 531.7593, p. 249–252.

LI, L., STOECKERT, C. J. et ROOS, D. S. (2003). « OrthoMCL : identification of ortholog groups for eukaryotic genomes ». *Genome Research* 13.9, p. 2178–2189.

LI, W. *et al.* (2014). « Virion-associated viral proteins of a Chinese giant salamander (Andrias davidianus) iridovirus (genus Ranavirus) and functional study of the major capsid protein (MCP) ». *Veterinary Microbiology* 172.1-2, p. 129–139.

LINDENBACH, B. D. (2009). « Measuring HCV infectivity produced in cell culture and in vivo ». *Methods in Molecular Biology (Clifton, N.J.)* 510, p. 329–336.

LIU, F. et ZHOU, Z. H. (2007). « Comparative virion structures of human herpesviruses ». *Human Herpesviruses : Biology, Therapy, and Immunoprophylaxis*. Sous la dir. d'A. ARVIN *et al.* Cambridge : Cambridge University Press.

LWOFF, A. (1957). « The concept of virus ». *Journal of General Microbiology* 17.2, p. 239–253.

LWOFF, A., HORNE, R. et TOURNIER, P. (1962). « A System of Viruses ». *Cold Spring Harbor Symposia on Quantitative Biology* 27, p. 51–55.

MAMMETTE, A. (2002). *Virologie médicale*. Presses Universitaires Lyon.

MARCHLER-BAUER, A. *et al.* (2015). « CDD : NCBI's conserved domain database ». *Nucleic Acids Research* 43.Database issue, p. D222–226.

MICHEL, R. *et al.* (2003). « Endocytobiont KC5/2 induces transformation into sol-like cytoplasm of its host Acanthamoeba sp. as substrate for its own development ». *Parasitology Research* 90.1, p. 52–56.

MONIER, A., CLAVERIE, J.-M. et OGATA, H. (2008a). « Taxonomic distribution of large DNA viruses in the sea ». *Genome Biology* 9.7, R106.

MONIER, A. *et al.* (2008b). « Marine mimivirus relatives are probably large algal viruses ». *Virology Journal* 5, p. 12.

MOZAR, M. et CLAVERIE, J.-M. (2014). « Expanding the Mimiviridae family using asparagine synthase as a sequence bait ». *Virology* 466-467, p. 112–122.

MUTSAFI, Y. *et al.* (2013). « Membrane assembly during the infection cycle of the giant Mimivirus ». *PLoS pathogens* 9.5, e1003367.

NAWROCKI, E. P. *et al.* (2015). « Rfam 12.0 : updates to the RNA families database ». *Nucleic Acids Research* 43.Database issue, p. D130–137.

NCBI RESOURCE COORDINATORS (2016). « Database resources of the National Center for Biotechnology Information ». *Nucleic Acids Research* 44.D1, p. D7–19.

NORRBY, E. (2007). « Yellow fever and Max Theiler : the only Nobel Prize for a virus vaccine ». *The Journal of Experimental Medicine* 204.12, p. 2779–2784.

OKONECHNIKOV, K. *et al.* (2012). « Unipro UGENE : a unified bioinformatics toolkit ». *Bioinformatics (Oxford, England)* 28.8, p. 1166–1167.

PAGNIER, I. *et al.* (2013). « A decade of improvements in Mimiviridae and Marseilleviridae isolation from amoeba ». *Intervirology* 56.6, p. 354–363.

PAGNIER, I. *et al.* (2015). « Isolation of Vermamoeba vermiformis and associated bacteria in hospital water ». *Microbial Pathogenesis* 80, p. 14–20.

PENG, Y. *et al.* (2012). « IDBA-UD : a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth ». *Bioinformatics (Oxford, England)* 28.11, p. 1420–1428.

PHAN, T. G. *et al.* (2015). « Absence of giant blood Marseille-like virus DNA detection by polymerase chain reaction in plasma from healthy US blood donors and serum from multiply transfused patients from Cameroon ». *Transfusion* 55.6, p. 1256–1262.

PHILIPPE, N. *et al.* (2013). « Pandoraviruses : amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes ». *Science (New York, N.Y.)* 341.6143, p. 281–286.

PIACENTE, F. *et al.* (2012). « Giant DNA Virus Mimivirus Encodes Pathway for Biosynthesis of Unusual Sugar 4-Amino-4,6-dideoxy-d-glucose (Viosamine) ». *The Journal of Biological Chemistry* 287.5, p. 3009–3018.

POPGEORGIEV, N. *et al.* (2013). « Marseillevirus-like virus recovered from blood donated by asymptomatic humans ». *The Journal of Infectious Diseases* 208.7, p. 1042–1050.

PRIET, S. *et al.* (2015). « mRNA maturation in giant viruses : variation on a theme ». *Nucleic Acids Research* 43.7, p. 3776–3788.

RAOULT, D. et FORTERRE, P. (2008). « Redefining viruses : lessons from Mimivirus ». *Nature Reviews. Microbiology* 6.4, p. 315–319.

RAOULT, D., LA SCOLA, B. et BIRTLES, R. (2007). « The discovery and characterization of Mimivirus, the largest known virus and putative pneumonia agent ». *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America* 45.1, p. 95–102.

RAOULT, D. *et al.* (2004). « The 1.2-megabase genome sequence of Mimivirus ». *Science (New York, N.Y.)* 306.5700, p. 1344–1350.

RAVACHOL, J. *et al.* (2014). « Characterization of all family-9 glycoside hydrolases synthesized by the cellulosome-producing bacterium Clostridium cellulolyticum ». *The Journal of Biological Chemistry* 289.11, p. 7335–7348.

REED, L. J. et MUENCH, H. (1938). « A Simple Method of Estimating Fifty Per Cent Endpoints, » *American Journal of Epidemiology* 27.3, p. 493–497.

RETENO, D. G. *et al.* (2015). « Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae ». *Journal of Virology* 89.13, p. 6585–6594.

RIEDEL, S. (2005). « Edward Jenner and the history of smallpox and vaccination ». *Proceedings (Baylor University. Medical Center)* 18.1, p. 21–25.

RIVERS, T. M. (1927). « Filterable viruses a critical review ». *Journal of Bacteriology* 14.4, p. 217–258.

RODRÍGUEZ-ZARAGOZA, S. (1994). « Ecology of free-living amoebae ». *Critical Reviews in Microbiology* 20.3, p. 225–241.

RUSCH, D. B. *et al.* (2007). « The Sorcerer II Global Ocean Sampling expedition : northwest Atlantic through eastern tropical Pacific ». *PLoS biology* 5.3, e77.

SAADI, H. *et al.* (2013a). « First isolation of Mimivirus in a patient with pneumonia ». *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America* 57.4, e127–134.

SAADI, H. *et al.* (2013b). « Shan virus : a new mimivirus isolated from the stool of a Tunisian patient with pneumonia ». *Intervirology* 56.6, p. 424–429.

SALAS, M. L. et ANDRÉS, G. (2013). « African swine fever virus morphogenesis ». *Virus Research*. African swine fever virus 173.1, p. 29–41.

SANTINI, S. *et al.* (2013). « Genome of Phaeocystis globosa virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes ». *Proceedings of the National Academy of Sciences of the United States of America* 110.26, p. 10800–10805.

SAUVAGE, V. *et al.* (2014). « No evidence of Marseillevirus-like virus presence in blood donors and recipients of multiple blood transfusions ». *The Journal of Infectious Diseases* 210.12, p. 2017–2018.

SCHEID, P. (2016). « A strange endocytobiont revealed as largest virus ». *Current Opinion in Microbiology* 31, p. 58–62.

SCHEID, P., BALCZUN, C. et SCHAUB, G. A. (2014). « Some secrets are revealed : parasitic keratitis amoebae as vectors of the scarcely described pandoraviruses to humans ». *Parasitology Research* 113.10, p. 3759–3764.

SCHEID, P., HAURÖDER, B. et MICHEL, R. (2010). « Investigations of an extraordinary endocytobiont in Acanthamoeba sp. : development and replication ». *Parasitology Research* 106.6, p. 1371–1377.

SCHOLTHOF, K.-B. (1997). « Tobacco mosaic virus. » *Plant Health Instructor*.

SCHWANHÄUSSER, B. *et al.* (2011). « Global quantification of mammalian gene expression control ». *Nature* 473.7347, p. 337–342.

SCOTT, M. S., TROSHIN, P. V. et BARTON, G. J. (2011). « NoD : a Nucleolar localization sequence detector for eukaryotic and viral proteins ». *BMC bioinformatics* 12, p. 317.

SHMAKOVA, L., BONDARENKO, N. et SMIRNOV, A. (2016). « Viable Species of Flamella (Amoebozoa : Variosea) Isolated from Ancient Arctic Permafrost Sediments ». *Protist* 167.1, p. 13–30.

SIDDIQUI, R. et KHAN, N. A. (2012). « Biology and pathogenesis of Acanthamoeba ». *Parasites & Vectors* 5, p. 6.

SLATER, G. S. C. et BIRNEY, E. (2005). « Automated generation of heuristics for biological sequence comparison ». *BMC bioinformatics* 6, p. 31.

SOMMER, D. D. *et al.* (2007). « Minimus : a fast, lightweight genome assembler ». *BMC bioinformatics* 8, p. 64.

STANLEY, W. M. (1938). « Isolation and Properties of Tobacco Mosaic and Other Virus Proteins ». *Bulletin of the New York Academy of Medicine* 14.7, p. 398–428.

SUTTLE, C. A. (2005). « Viruses in the sea ». *Nature* 437.7057, p. 356–361.

SUTTLE, C. A. (2007). « Marine viruses–major players in the global ecosystem ». *Nature Reviews. Microbiology* 5.10, p. 801–812.

TAKEMURA, M. (2001). « Poxviruses and the origin of the eukaryotic nucleus ». *Journal of Molecular Evolution* 52.5, p. 419–425.

TAKEMURA, M. (2016). « Draft Genome Sequence of Tokyovirus, a Member of the Family Marseilleviridae Isolated from the Arakawa River of Tokyo, Japan ». *Genome Announcements* 4.3.

THOMAS, V. *et al.* (2011). « Lausannevirus, a giant amoebal virus encoding histone doublets ». *Environmental Microbiology* 13.6, p. 1454–1466.

TOMLINSON, G. et JONES, E. A. (1962). « Isolation of cellulose from the cyst wall of a soil amoeba ». *Biochimica et Biophysica Acta* 63.1, p. 194–200.

VAN ETTEN, J. L. et MEINTS, R. H. (1999). « Giant Viruses Infecting Algae ». *Annual Review of Microbiology* 53.1, p. 447–494.

VANSPAUWEN, M. J. *et al.* (2013). « Mimivirus is not a frequent cause of ventilator-associated pneumonia in critically ill patients ». *Journal of Medical Virology* 85.10, p. 1836–1841.

VENTER, J. C. *et al.* (2004). « Environmental genome shotgun sequencing of the Sargasso Sea ». *Science (New York, N.Y.)* 304.5667, p. 66–74.

VILLARREAL, L. P. (2005). *Viruses and the Evolution of Life*. American Society of Microbiology.

WALLACE, I. M. *et al.* (2006). « M-Coffee : combining multiple sequence alignment methods with T-Coffee ». *Nucleic Acids Research* 34.6, p. 1692–1699.

WESTBROEK, P., YOUNG, J. R. et LINSCHOOTEN, K. (1989). « Coccolith Production (Biomineralization) in the Marine Alga Emiliania huxleyi ». *The Journal of Protozoology* 36.4, p. 368–373.

WIDENER, R. W. et WHITLEY, R. J. (2014). « Herpes simplex virus ». *Handbook of Clinical Neurology* 123, p. 251–263.

WILLIAMS, T., BARBOSA-SOLOMIEU, V. et CHINCHAR, V. G. (2005). « A decade of advances in iridovirus research ». *Advances in Virus Research* 65, p. 173–248.

WILLIAMSON, S. J. *et al.* (2008). « The Sorcerer II Global Ocean Sampling Expedition : metagenomic characterization of viruses within aquatic microbial samples ». *PloS One* 3.1, e1456.

WILSON, W. H., VAN ETTEN, J. L. et ALLEN, M. J. (2009). « The Phycodnaviridae : The Story of How Tiny Giants Rule the World ». *Current topics in microbiology and immunology* 328, p. 1–42.

WILSON, W. H. *et al.* (2005). « Complete genome sequence and lytic phase transcription profile of a Coccolithovirus ». *Science (New York, N.Y.)* 309.5737, p. 1090–1092.

XIAO, C. *et al.* (2005). « Cryo-electron microscopy of the giant Mimivirus ». *Journal of Molecular Biology* 353.3, p. 493–496.

XIAO, C. *et al.* (2009). « Structural studies of the giant mimivirus ». *PLoS biology* 7.4, e92.

YUTIN, N. *et al.* (2009). « Eukaryotic large nucleo-cytoplasmic DNA viruses : clusters of orthologous genes and reconstruction of viral genome evolution ». *Virology Journal* 6, p. 223.

ZAUBERMAN, N. *et al.* (2008). « Distinct DNA exit and packaging portals in the virus Acanthamoeba polyphaga mimivirus ». *PLoS biology* 6.5, e114.

ZHANG, X.-A. *et al.* (2016). « Lack of Mimivirus Detection in Patients with Respiratory Disease, China ». *Emerging Infectious Diseases* 22.11, p. 2011–2012.

Annexes

A Noumeavirus replication relies on a transient remote control of its host nucleus

Noumeavirus replication relies on a transient remote control of the host nucleus

Elisabeth Fabre[a, 1], Sandra Jeudy[a, 1], Sébastien Santini[a], Matthieu Legendre[a], Mathieu Trauchessec[b,c,d], Yohann Couté[b,c,d], Jean-Michel Claverie[a, e*], Chantal Abergel[a*]

[a]Aix–Marseille Univ, Centre National de la Recherche Scientifique, Information Génomique & Structurale, Unité Mixte de Recherche 7256 (Institut de Microbiologie de la Méditerranée, FR3479), 13288 Marseille Cedex 9, France

[b]Université Grenoble Alpes, BIG-BGE, F-38000 Grenoble, France

[c]Commissariat à l'Energie Atomique, BIG-BGE, F-38000 Grenoble, France

[d]INSERM, BGE, F-38000 Grenoble, France

[e]Assistance Publique - Hôpitaux de Marseille, 13385 Marseille, France

[1]E.F. and S.J contributed equally to this work.

Abstract

Acanthamoeba are infected by a remarkable diversity of large dsDNA viruses the infectious cycles of which have been characterized using genomics, transcriptomics and electron microscopy. Given their gene content and the persistence of the host nucleus throughout their infectious cycle, the Marseilleviridae were initially assumed to fully replicate in the cytoplasm. Unexpectedly, we found that their virions did not incorporate the virus-encoded transcription machinery, making their replication nucleus-dependent. However, instead of delivering their DNA to the nucleus, the Marseilleviridae initiate their replication by transiently recruiting the nuclear transcription machinery to their cytoplasmic viral factory. The nucleus recovers its integrity after becoming leaky at an early stage. Such intermediary host-virus cross-talk is consistent with a progressive transition from fully cytoplasmic to increasingly nucleus-dependent viruses through reductive evolution. This work highlights the importance of virion proteomic analyses to complement genome sequencing in the elucidation of the replication scheme of large dsDNA viruses.

INTRODUCTION

Since the discovery of Marseillevirus[1], the Marseilleviridae family of large DNA viruses infecting Acanthamoeba has been steadily expanding with now more than ten members[2–9]. They all exhibit 200 nm diameter icosahedral virions enclosing dsDNA genomes in the 340-400 kb range. Members of the family were isolated from distant locations all around the world and seem to distribute among 5 lineages comprising Marseillevirus, Lausannevirus, Tunisvirus, Brazilian virus and Golden mussel marseillevirus relatives[4,8,10]. The presence of Marseilleviridae has been reported in several human patients, but their link to specific diseases remains elusive[11–19]. Initial TEM observations of Marseilleviridae infectious cycles revealed that their particles were assembled at the periphery of the virion factory located next to the intact cell nucleus. The mature virions are then gathered in vacuoles suggesting their release via exocytosis[2–4,6,18,20,21]. This global picture, together with the presence of a well conserved transcription machinery encoded in all Marseilleviridae genomes, led to the assumption that their infectious cycle was exclusively cytoplasmic. Following the characterization of a new member of the Marseilleviridae family isolated from an environmental sample collected near Noumea in New-Caledonia, we compared the genomes and the virion proteomes of Noumeavirus with that of Melbournevirus, its most distant relative within the family. Unexpectedly, the virus-encoded transcription machinery is absent from their virions, implying that they must initiate their replication using the cellular transcription machinery, normally confined in nucleus. Detailed TEM and fluorescence microscopy studies were focused on the host nucleus in search for evidences of viral DNA transfer during the early stage of Noumeavirus infectious cycle. Our results revealed a transient recruitment of the nucleus that could not have been suspected without the proteomic analysis of the virions.

RESULTS

Isolation of Noumeavirus

A muddy sample of fresh water collected in a pond near Noumea airport (New-Caledonia) treated with antibiotics triggered a lytic infection phenotype on a culture of *A. castellanii* cells. No visible growing parasite could be detected by light microscopy. The virus was cloned and amplified by isolating a single infected *A. castellanii* cell to which were added fresh Acanthamoeba cells. The observation of an asynchronous infected cells culture by electron microscopy revealed intracellular icosahedral particles 200 nm in diameter, sometimes gathered in large vacuoles as observed on Acanthamoeba cells infected by members of the Marseilleviridae.

Noumeavirus genome

The double-stranded DNA genome sequence of Noumeavirus was assembled into a single contig of 376,207 bp in length with a G+C content of 43%. A total of 452 ORFs were predicted to encode proteins ranging from 50 to 1525 amino acids for an average length of 249 residues. Protein coding regions account for 90% of the genome and are separated by short intergenic regions (88.2 nucleotides long in average). As for all entirely sequenced Marseilleviridae representatives, Noumeavirus genome was assembled as a circular sequence that might also correspond to a terminally redundant linear circularly permutated DNA molecule, as described for Iridoviruses[22], the closest relative of the Marseilleviridae family[1,2]. Out of the 452 predicted Noumeavirus proteins, 438 proteins have their closest homolog within the Marseilleviridae. Noumeavirus is thus clearly a new member of the Marseilleviridae family with only 9 ORFans (2%) among its predicted proteins (Fig. S3). It has been proposed to partition the Marseilleviridae in 5 lineages[5] comprising Marseillevirus (A), Lausannevirus (B), Tunisvirus (C), Brazilianvirus (D) and Golden mussel marseillevirus (E)[10] as respective prototypes. Noumeavirus belongs to the B lineage with Port-Miou virus[7] and Lausannevirus[2] (Fig. S2) as its closest relatives. Noumeavirus shares 374 homologous proteins with

Lausannevirus (with 77% identical residues in average, Table S4) and 331 with Melbournevirus (55 % identical on average).

As two adjacent Noumeavirus ORFs exhibited a strong similarity with the N-terminal (for one) and C-terminal (for the other) domain of the large DNA-directed RNA polymerase subunit (RPB1), we suspected the presence of an intron and its boundaries were precisely mapped using RT PCR. As they did not correspond to donor and acceptor sites associated to a spliceosomal intron, we analyzed the intervening sequence on the rFAM webserver[23] which identified it as a self-splicing Group-I intron (E-value: 4.8 e^{-10}). Excepted for Port-Miou virus, the RPB1 gene (Noumeavirus: NMV_011) is found interrupted by a self-splicing intron in all fully sequenced Marseilleviridae (Melbournevirus MEL_047, E-value: 1.5 e^{-8}). A second putative case of split gene was investigated, concerning a predicted origin of replication binding protein of which 4 paralogs are found in the Noumeavirus genome. We experimentally demonstrated the existence of transcript entirely encompassing the two adjacent ORFs separated by a STOP codon. This structure, now annotated NMV_287-288, may correspond to a recent pseudogenization event. Eight additional genomic regions were predicted to contain spliceosomal introns using Exonerate[24]. Three of them were amplified and sequenced (NMV_125, NMV_252 and NMV_436) but none of them were found to contain introns. This finding is consistent with the hypothesis that the Noumeavirus genes are not transcribed within the host nucleus.

Like all other members of the Marseilleviridae family[1–8,25], Noumeavirus possesses histone-like proteins that were suggested to play a role in the compaction and stability of the viral DNA in the particles[2]. The presence of an additional gene encoding a DNA topoisomerase II, known to be involved in chromatin compaction, is consistent with this hypothesis. However, the origin of these proteins and their evolutionary history remain unclear[25].

Many Marseilleviridae transcripts terminate according to the Mimiviridae "hairpin" rule

Two proteins involved in the maturation of Mimivirus (and probably Pithovirus[26]) transcripts[27] have homologs in all Marseilleviridae. A mRNA capping enzyme performing the 5'-addition and N7

methylation of the cap[28] (NMV_139 and MEL_282) and a RNAse III (NMV_212 and MEL_219)[27,29,30] recognizing the palindromic structure at the 3'-end of Mimivirus mRNAs prior to their cleavage and polyadenylation. This prompted us to investigate whether the Mimivirus "hairpin rule" for mRNA termination[29] was at least partially obeyed in Marseilleviridae. The genomes analysis of Noumeavirus and Melbournevirus revealed a total of 226 and 245 palindromes respectively (Fig. S4). In Noumeavirus, 75 of them (33.2%) are found in coding regions (90% of the genome) and 151 (66.8%) between ORFs (10% of the genome). Such a non-uniform distribution is highly significant ($p<0.0001$, Fisher exact test [75, 151; 204, 22]). A similar highly significant bias was found in Melbournevirus with 87 palindromes found in coding regions (89.8% of the genome) and 158 between ORFs (10.2% of the genome) ($p<0.0001$, Fisher exact test [87, 158; 220, 25]). These uneven distributions suggest that the palindromes might play a role in the termination/polyadenylation of mRNAs. This was experimentally validated for the NMV_1 transcript, the polyadenylation of which is precisely located within the predicted 3'end hairpin (Fig. S5).

Proteomic analyses of purified Melbournevirus vs. Noumeavirus particles

We used Melbournevirus (Lineage A) and Noumeavirus (Lineage B), the two most divergent Marseilleviridae representatives at our disposal in the laboratory, to perform a comparative analysis of their respective virion proteomes. The particles of the previously described large and giant DNA viruses[26,31–39] have all been found to contain a large number of proteins. This is also the case for the Marseilleviridae. We identified respectively 283 (128 viral, 155 host-derived) and 196 (132 viral, 64 host-derived) proteins in the virions of Noumeavirus and Melbournevirus and ranked them according to their estimated abundances[40]. The two viruses thus appear to mostly differ by the larger proportion of *A. castellanii* proteins (i.e. host-derived) associated to Noumeavirus (54.8%) than to those associated to Melbournevirus particles (32.65%). In contrast we detected very similar numbers of virally-encoded proteins in the two virion proteomes (Table S1). As most host-derived proteins are detected at very low abundance, the discrepancy is most likely due to a difference in the purification

and storage of the virions rather than to biologically meaningful variations. We interpreted the host-derived proteins detected in only one of the viruses (103 in Noumeavirus, 12 in Melbournevirus) as bystanders most likely devoid of biological functions. The presence/absence of these low abundance proteins is either due to a difference in virion storage protocols (not stored in DMSO for Noumeavirus) or their random incorporation in low copy numbers. This view is comforted by the fact that, in contrast, most of the detected virus-encoded proteins are shared by the two virions (89% for Noumeavirus and 86.4% for Melbournevirus) (Table S1).

Host-encoded proteins shared by the virions

Among the 52 host-encoded proteins shared between Noumeavirus and Melbournevirus virions, the most abundant are a mitochondrial porin and Actin 1, also abundant in the uninfected amoeba proteome and in the particles of unrelated Acanthamoeba-infecting viruses (Table S2). Except for the Copper/zinc superoxide dismutase (ranked 83rd in Melbournevirus), the other shared host proteins are ranked below 100th and are also detected in the particles of unrelated Acanthamoeba-infecting viruses, which might reflect their high abundance in non-infected Acanthamoeba cells (Table S2). Nevertheless, the possibility remains that some of them could participate to the infection process such as the intracytoplasmic trafficking of viral components through the reorganization of the host cytoskeleton[41] (actophorin, fascin, coronin, tubulins, etc).

Twenty one host proteins are uniquely shared by the Noumeavirus and Melbournevirus virions (Table S2). Two correspond to abundant proteins ranked 4th (Glyceraldehyde-3-phosphate dehydrogenase) and 18rd (hypothetical, protein binding domain) in proteomes from uninfected cells (Table S2). Their unique presence might thus be specific to the Marseilleviridae infection process. We noticed a strong bias for small GTPase (6 RAS-like homologs) that could be involved in the rapid trafficking of viral and cellular components to and from the host nucleus, together with cytoskeleton proteins (alpha and beta tubulin, calponin) and one kinase (Table S2). Among the host proteins only identified in Melbournevirus and Noumeavirus virions, one is made of tandem histone domains and was found in low abundance in the proteome of uninfected *A. castellanii* cells (ranked 1621st).

Interestingly, a homologue of this protein is encoded by the Iridoviridae and essential for viral replication[42]. As such protein is normally confined to the nucleus, its detection in Marseilleviridae particles suggests that nuclear components may infiltrate the cytoplasmic viral factory where virions are assembled.

Virus-encoded proteins found in the virions

Out of the 114 virus-encoded proteins common to Noumeavirus and Melbournevirus particles, 72 (63%) correspond to hypothetical proteins, including 20 transmembrane domains containing proteins (Table S1). The correlation between their respective ranks highlights a strong conservation of the composition of the Marseilleviridae virions (Fig. 1, R^2=0.8641, p-value << 1e-15). Among the 14 viral proteins absent from Melbournevirus virions, 12 have ranking below 100th in Noumeavirus particles and could have been left undetected by the proteomic analyses due to the absence of biological replicates. The last two discrepant proteins (ranked 25th and 42th in Noumeavirus) correspond to proteins not encoded by Melbournevirus, one of them being a member of a large protein family for which paralogs are present in Melbournevirus virions. Similarly, the 18 proteins absent from Noumeavirus virions all have ranking >110th in Melbournevirus virions. Five of them are not encoded by the Noumeavirus genome.

The analysis of the Marseilleviridae particle proteomes led to several unexpected results.

First, we found that the Major Capsid Protein, usually the most abundant in other icosahedral viruses particles[32,33,35,36,43–47], is only ranked third in Noumeavirus (NMV_116) and fourth in Melbournevirus (MEL_305). Instead, the most abundant protein (NMV_189 or MEL_236) is a hypothetical protein of about 150 residues, without homolog outside of the Marseillevidae but highly conserved (albeit not always annotated) in all fully sequenced genomes of the family. The second most abundant protein in Noumeavirus (NMV_078), third in Melbournevirus (MEL_368), is a histone H3-like protein. Two other virally-encoded histone-like proteins are also abundant in the Noumeavirus and Melbournevirus proteomes (NMV_079, MEL_247 and NMV_419 and MEL_149). The number of histone-like proteins in the virions (3 virus- and 1 host-encoded), strongly suggests their involvement

178 in viral DNA packaging, especially since the Marseilleviridae do not encode a *bona fide* Major Core
179 protein. Interestingly, a histone-like protein was also identified in the virion proteomes of
180 Iridoviridae, the family of viruses for which the Marseilleviridae have the highest phylogenetic
181 affinity[36,43,45,47,48].
182 Finally, the virions share 9 proteases, two lipases and five virus-encoded oxidative stress proteins
183 (peroxidases, thioredoxins) that could complement the packaged host-encoded ones in resisting the
184 aggressive environment of the phagosome. Three virus-encoded glycosyl transferases might be
185 involved in the intracellular trafficking of viral or host proteins through their glycosylation.
186 We also noticed that both the mRNA capping enzyme and the RNAse III (Table S3) are present in the
187 virions, suggesting that viral transcripts need to be maturated by virus-encoded proteins from the
188 very beginning of the infectious cycle, as for cytoplasmic viruses[27]. It is worth noticing that transcript
189 maturation proteins are usually as abundant as the transcription apparatus itself in the particles of
190 cytoplasmic viruses obeying the hairpin rule (Table S3). As transcript maturation proteins are found
191 in the Noumeavirus and Melbournevirus virions at an abundance level comparable to that of
192 Pithovirus and Mimivirus, their transcription machinery, if present, should have also been detected.
193

194 **The Marseilleviridae require the host transcription machinery to initiate infection**

195 The most unexpected results of our proteomic analysis of Melbournevirus and Noumeavirus virions is
196 the conspicuous absence of the virus-encoded (or host-encoded) RNA polymerases, while the
197 transcription machinery was always easily identified in other cytoplasmic viruses[32,35,38,39,44] (Table S3).
198 This immediately suggests that, even though their replication cycle appears to entirely take place in
199 the cytoplasm, the Marseilleviridae must rely on the host transcription machinery to generate their
200 earliest transcripts.
201 As the host machinery is normally confined in the nucleus, this process could either involve a
202 translocation of the viral genome into the nucleus (i.e. *via* a true intranuclear phase of the replication
203 cycle), a transfer of the corresponding mRNA from the virion to the cytoplasm to produce the

transcription machinery, or a recruitment of the needed nuclear functions to the cytoplasmic viral factory. We investigated the possible presence of mRNAs associated to the transcription machinery in the two Marseilleviridae virions (Supplementary material). While the mRNA encoding for the most abundant protein (NMV_189 and MEL_236, control for the experiment) was detected in both virions, we solely obtained a PCR product for the Melbournevirus virion RPB1 mRNA and none for the other polymerase subunits or the host encoded PolyA polymerase. As for MEL_236, the RPB1 mRNA most likely corresponds to a bystander as less than one copy per virion was detected (Supplementary Material). We thus performed a detailed analysis of the two viruses' infectious cycles focusing on the status of the host nucleus.

The nature of the cross-talk between the viruses and the host nucleus was investigated by following the infection from 30 min to 4h pi in *A. castellanii* cells expressing either the GFP-tagged nucleolar (ML216-GFP, MW 58kDa) or nuclear (ML135-GFP, MW 62kDa) Mollivirus methyltransferases, or the *A. castellanii* nuclear GFP-tagged SUMO protein (GFP-SUMO, 39kDa). The non-infected control cells all exhibited DAPI staining and GFP fluorescence localized in the nucleus. Noumeavirus and Melbournevirus replicative cycles exhibit some differences when observed by light microscopy. Noumeavirus-infected cells quickly lose their adherence (1 h pi) and form aggregates visible to the naked eye (2 h pi) while Melbournevirus-infected cells remain adherent until the end. However, when the cells were infected by one or the other virus, we noticed a spreading of the nucleolar and nuclear GFP fluorescence 30min pi (Fig. 3d, Fig. S6b, f). One hour pi, the nuclear ML135-GFP (Fig. 3f), GFP-SUMO (Fig. S6g) as well as the nucleolar ML216-GFP proteins (Fig. S6c) were no longer confined into the nucleus, but appeared to have diffused out entirely, completely filling the cytoplasm. In contrast, the DAPI staining remained into the nucleus indicating the persistence of the host DNA, as previously reported[1], even if the labelling was not as homogenous as in uninfected *A. castellanii* cells. Interestingly, around 2h pi, strong DAPI stained foci, probably corresponding to the viral nucleoid DNA (Fig. 3b, e), were observed in the cytoplasm during the infection of *A. castellanii* by Noumeavirus or Melbournevirus. During this phase corresponding to the development of the early

viral factory, these foci appear to concentrate the GFP-fluorescence (Fig. 3c, f), suggesting the recruitment of nuclear proteins around the viral DNA. These observations suggest that while the global integrity of the host nucleus is preserved, modifications triggered at the early stage of infection enhances the free diffusion of small nuclear proteins such as the GFP-SUMO into the cytoplasm as well as the active recruitment of the Mollivirus larger GFP-tagged proteins that concentrate in the developing viral factories. Intriguingly, these virus-induced modifications appear transient, as the GFP fluorescence is back into the nuclei between 2h pi (Fig. 3i) and 4h pi (Fig. 2h, Fig. S6d, h). The DAPI staining becomes stronger in the viral factories, where viral DNA replication is actively taking place (Fig. 2h, Fig. 3h, Fig. S6d, h). After 2h pi, the GFP fluorescence when visible outside the nucleus appears exclusively localized in the viral factory (Fig. 3f, i) and co-localized with the DAPI staining (Fig. 3e, h) suggesting that the nuclear proteins are specifically driven into this compartment. These proteins could perform essential functions such as mRNA early transcription as well as the polyadenylation of the viral transcripts since the Marseilleviridae lack a virus-encoded PolyA polymerase. Once expressed, the viral transcription machinery can take over the cellular one, except for the PolyA polymerase.

The Noumeavirus replication cycle.

Noumeavirus replicative cycle is very similar to that of Melbournevirus and other Marseilleviridae when observed by Transmission Electronic Microscopy (TEM). The status of the cell nucleus during the entire infectious cycle was precisely assessed, from the initial virus-cell contact until the release of neo-synthesized virions 7h pi. As previously described for other Marseilleviridae, free Noumeavirus icosahedral particles are ~200 nm in diameter, with a translucent external layer contrasting with their denser internal compartment. Ten minutes pi, Acanthamoeba cells are seen engulfing viral particles at their periphery while others exhibit virions already captured in small vacuoles (Fig. 4a). The first stage of the infection thus corresponds to the internalization of Noumeavirus particles *via* phagocytosis[21] or endocytosis. After 20 min, some of the virions in the

vacuoles appear to have lost their external translucent shell as if it has been digested (Fig. 4b). In many infected cells, the nuclei have lost their spherical appearance, form circumvolutions and no longer exhibit a dense well-organized nucleolus (Fig. 2c). These changes become even more dramatic after 30 min. Quasi spherical nucleoid corresponding to the electron-dense core of the virions (Fig. 4c) can be recognized in vacuole, some of which open and being filled with cytoplasmic components such as ribosomes. Similar electron-dense spheres are also seen in the periphery of deformed nuclei (Fig. 2c-e). These images suggest that once in the vacuole, the icosahedral capsids are dissolved and the virion core transferred to the cytoplasm through an opening of the vacuole membrane. The DNA containing cores could correspond to the intense DAPI stained foci scattered in infected cells around 2h pi (Fig. 3b, e) that seem to ultimately gather to form the mature viral factory (Fig. 2h, Fig. 3.h, Fig. S6d, h). This process is thus different from the one used by all other described Acanthameoba-infecting viruses which involves the opening of a specific portal followed by a virus-phagosome membrane fusion, funneling the viral core into the cytoplasm.

After 1h, many electron-dense tubular structures (possibly ruptured membranes) appear in infected cells (Fig. 3d) the nucleus of which is still deformed (Fig. 2e). Then, after 2 hours pi, the nuclei of infected cells regain their initial spherical shape and electron dense nucleolus (Fig. 2g). Faces of capsids start to bud in the viral factories with cell organelles crowding at their periphery. As the infectious cycle progresses, more spherical nuclei with dense nucleolus are seen to coexist with large viral factories. New virions are first assembled as empty looking particles made of the translucent symmetrical layer and then filled with the electron-dense core characteristic of mature virions (Fig. 4e). Mature and immature virions can be seen side-by-side in the same viral factory. Tubular structures remain visible in the viral factories (Fig. 4e). As for other Marseilleviridae, the newly-synthesized virions are ultimately gathered into intracytoplasmic vacuoles (7h pi), suggesting an exocytosis-like mode of dissemination[2,6,8,21] (Fig. 4f).

DISCUSSION

The Marseilleviridae are large DNA viruses using icosahedral virions to propagate their genomes. They all encode a complete transcription apparatus and by analogy with other icosahedral viruses infecting Acanthamoeba such as the Mimiviridae, it was assumed that they were cytoplasmic viruses[2,6]. This notion was supported by initial TEM studies of the infectious cycle, during which all visible replication events appeared to happen in the cytoplasm where large viral factories coexisted next to the host nuclei remaining intact until the end.

We performed a proteomic comparison of Noumeavirus, a new member of the Marseilleviridae family and it's most distant relative Melbournevirus, to determine to which extent the viral particle compositions were conserved and could possibly provide some insights on infectious processes characteristic of the family. Our first finding was that the virions proteomes were strikingly similar and composed of 2/3rd of virus-encoded proteins and 1/3rd of host-derived proteins.

An important and unexpected result was that the transcription machinery was absent from the two virions proteomes, immediately suggesting that the Marseilleviridae had to rely on the host nucleus, at least during the earliest phase of infection. This prompted us to carefully scrutinize infected cells by electronic and fluorescence microscopy, specially focusing on the first hours of the infectious cycle. Taking advantage of *A. castellanii* cells expressing three different nucleus-targeted fluorescent proteins (one nucleolar protein and two nuclear proteins of different sizes), we showed that 30 min after initiating an infection by Noumeavirus, these GFP-fused proteins normally localized in the nucleus were exported to the cell cytoplasm, while the cellular DNA remained confined in the nucleus. TEM studies indicated that a deformation of the cell nucleus (also observed by others)[1] and a fading of the normally electron-dense nucleolus were co-occurring with this apparent permeabilization. In contrast with what has been described for Acanthamoeba cells infected by a *bona fide* nucleocytoplasmic virus such as Pandoravirus, the phenomenon of nucleus disruption does not end up with its complete disappearance[34]. It is reversible and after 2 hours, the GFP signal appears again confined to the nucleus, the shape and structure of which also revert to normal. The same phenomenon was observed during Melbournevirus infection cycle.

Supported by fluorescence microscopy observations of infected cells, our interpretation is that proteins normally targeted to the nucleus are actively but transiently recruited to the developing cytoplasmic viral factory at the initial stage of the infection cycle. These proteins would indeed include the transcription machinery required to synthesize the earliest viral transcripts. In contrast to *bona fide* nuclear viruses, there is no evidence that the Marseilleviridae ever transfer their genomes into the host nucleus or that the nuclear compartment is ever broken, as indicated by the persistence of the DAPI-stained DNA in the nucleus (Fig. 2, Fig. 3, Fig. S6)[1].

Consistent with the above scenario, we did not identify a single Noumeavirus gene interrupted by a spliceosomal intron, in contrast with viruses undergoing a true intranuclear phase, such as Pandoraviruses or Mollivirus[34,37]. Our view is that the Noumeavirus genes are initially transcribed in the cytoplasmic factory by the host DNA-directed RNA polymerases, until the neo-translated virus-encoded RNA polymerase can take over for the later stages of the infectious cycle. This virus-host relationship would thus correspond to a transitional evolutionary stage between that of the nucleocytoplasmic viruses and that of the purely cytoplasmic viruses, such as the Mimiviridae. The unexpected detection of RPB1 mRNA in the Melbournevirus virions can have two interpretations. Either this molecule is a simple bystander devoid of biological function as suggested by our experimental results or the Melbournevirus RPB1 is actually translated from this mRNA. The latter scenario would suggest yet another intermediate evolutionary step from cytoplasmic to nuclear viruses. It would consist of a virus loading its virus-encoded transcription machinery as transcripts rather than as virion-loaded functional proteins.

In the framework of the genome reduction model, our preferred evolutionary scenario for large DNA viruses[32,49], the existence of intermediates between purely cytoplasmic and nucleus-dependent replication cycles is expected. We propose that the hypothetical viral ancestor was able to replicate totally independently from the host nucleus. Thus, in addition to encode its own complete DNA transcription machinery, it was able to deliver it as a functional transcriptional complex (or as mRNAs) in the cytoplasm to bootstrap the infection. Various degrees of gene loss from this

hypothetical ancestor could lead to the variety of nucleus-dependent large DNA viruses known today, such as those not encoding a transcription machinery (e.g. Chloroviruses, Herpesviruses). The Marseilleviridae appear at the brink of this evolutionary journey, still encoding their transcription apparatus, but having lost the ability to use it for the earliest phase of the infection. The many virally and host encoded GTPases loaded in the Marseilleviridae particles could have been key in allowing this transition by modulating intracellular trafficking and protein exchange between the host nucleus and the nascent viral factory. Such a transitional evolutionary stage illustrates how the gradual establishment of a cross-talk between the nucleus and the virion factory could compensate for the loss of seemingly “essential” viral core genes. The next evolutionary step, following the acquisition of the capacity to transport the viral DNA into the nucleus, would open the way to a complete loss of a virus-encoded DNA processing (transcription then replication) enzymes and to a total reliance on the cell machinery for the generation and processing of the viral transcripts.

The capacity of the Marseilleviridae viral factory to transiently mimic the nucleus and highjack nuclear proteins essential for early transcription might also be linked to the "viral eukaryogenesis" theory, a hypothesis proposing that a large DNA virus might be at the origin of the nucleus[50–53]. Viral factories and nuclei have in common to both confine replication and transcription in a well-defined compartment from which energy producing organelles are excluded. They both appear as obligate "parasites" of the cellular cytoplasm from which they depend to synthetize the proteins they encode. The unique features of today's Marseilleviridae replication cycle in modern Acanthamoeba might constitute a testimony from a distant past about the contribution of DNA viruses to the emergence of the nucleus in ancestral cell lineages.

CONCLUSION

The comparative analysis of the virion proteomes of two distant Marseilleviridae led to two major findings.

1) At odd with icosahedral DNA viruses[32,33,35,36,43–45,47,48,54,55], the Major Capsid Protein was not found to be the most abundant in the virion proteomes. Instead it is a 150-residues long protein of unknown function, specific and highly conserved in all Marseilleviridae (>77% identity). Its role in the particle remains to be determined.

2) While it was assumed that the Marseilleviridae were entirely cytoplasmic viruses, the virion proteomes lack the transcription machinery necessary to initiate the infectious cycle without the help of host nuclear functions. This unexpected finding illustrates the impossibility to infer the type of infectious cycle of a (new) virus from sole genomic and transcriptomic data. The analysis of the virion proteome could thus help discriminating between purely cytoplasmic and nucleocytoplasmic viruses. The use of fluorescent proteins targeted to the host nucleus unveiled an intermediate level of viral dependency toward the host, in between large DNA viruses undergoing a truly intranuclear phase (e.g. herpesviridae), viruses totally disrupting the nucleus (e.g. Pandoravirus), and viruses entirely replicating within the cytoplasm (e.g. Mimiviridae). The Marseilleviridae now document a new mode of virus-host cross-talk whereby the viral factory transiently mimics the nucleus in order to highjack nuclear proteins required to initiate the transcription of the viral genome.

MATERIALS AND METHODS

Virus isolation

Noumeavirus was isolated from a muddy fresh water sample collected in a pond near Noumea Airport (New-Caledonia, Lat: 22°16'29.50"S, Long: 166°28'11.61"E). After mixing the mud and the water, 1.5 mL were recovered and 150 µL of pure Fungizone (25 µg/mL final) were added to the sample which was vortexed and incubated overnight at 4°C on a stirring wheel. After decantation, the supernatant was recovered and centrifuged at 800 × g for 5 min. *Acanthamoeba castellanii* (Douglas) Neff (ATCC 30010TM) cells adapted to Fungizone (2.5 µg/mL) were inoculated with 100 µL of the supernatant and with the pellet resuspended in 50 µL of PBS buffer. The cells were cultured at 32°C in microplates with 1 mL of proteose peptone–yeast extract–glucose (PPYG) medium (2%

proteose peptone, 0.1% yeast extract, 2.5 mM KH_2PO_4, 2.5 mM Na_2HPO_4, 0.4 mM $CaCl_2$, 4 mM $MgSO_4(H_2O)_7$, 50 µM $Fe(NH_4)_2(SO_4)_2$, 100 mM glucose pH 6.5) supplemented with antibiotics [ampicillin, 100 µg/mL, and penicillin–streptomycin, 100 µg/mL (Gibco); Fungizone, 2.5 µg/mL (Life Technologies)] and monitored for cell death.

Virus purification

The wells showing evidence of cell death were recovered, centrifuged for 5 min at 500 × g to remove the cellular debris and used to infect four T-75 tissue-culture flasks plated with fresh Acanthamoeba cells. After completion of the infectious cycle, the cultures were recovered, centrifuged for 5 min at 500 × g to remove the cellular debris, and the virus was pelleted by centrifugation at 6,800 × g for 45 min. The viral pellet was then resuspended, washed twice in PBS buffer, layered on a discontinuous sucrose gradient [20%/30%/40%/50% (wt/vol)], and centrifuged at 5,000 × g for 45 min. The virus produced a white iridescent disk, which was recovered, washed twice in PBS buffer, and stored at 4 °C.

Virus cloning

A. castellanii cells (70,000/cm^2) were seeded in one well of a 12-well culture plate with 1 mL of PPYG. After allowing enough time for cell adhesion, viruses were added to the well at a multiplicity of infection (MOI) of 50. After 1 h, the well was washed several times with 1 mL of PPYG to remove the excess of viruses. The cells were then recovered by gently scrapping the well, and a serial dilution was performed in the next three wells by mixing 200 µL of the previous well with 500 µL of PPYG. Drops of 0.5 µL of the last dilution were recovered and observed by light microscopy to verify that wells contained a single cell. The 0.5-µL droplets were then distributed in each well of a 24-well culture plate. Thousand uninfected *A. castellanii* cells in 500 µL of PPYG were added to the wells seeded with a single cell and monitored for cell death. The corresponding viral clones were recovered

and amplified prior purification, DNA extraction, proteome analysis, and cell cycle characterization by electron microscopy.

Infectious cycle observations by TEM

Eleven T25 flasks were plated with 100,000 *A. castellanii* cells/cm^2 in PPYG medium. Ten of them were infected by Noumeavirus (MOI 50, based on titration measurements) to ensure a synchronous infectious cycle. The cells from each flask were recovered at different time post infection (pi): t_0 (right after the cells encountered the viruses), 10 min pi and 20 min pi. The excess of virions was removed from the seven other flasks 30 min pi and cells were recovered from each of the seven flasks at 30 min, 45 min, 1h, 3h, 3h30, 4h and 7h pi. The *A. castellanii*-infected cell cultures were fixed by adding an equal volume of PBS buffer with 5% glutaraldehyde and incubated for 1 hour at room temperature. Cells were recovered and pelleted for 20 min at 5,000 × g. The pellet was resuspended in 1mL of PBS with 2.5% glutaraldehyde, incubated at least 1 h at 4 °C, and washed twice in PBS buffer prior coating in agarose and embedding in Epon resin. Each pellet was mixed with 2% low melting agarose and centrifuged to obtain small flanges of approximately 1 mm^3 containing the sample coated with agarose. These samples were then prepared using the OTO (osmium-thiocarbohydrazide-osmium) method: 1 h fixation in 2% osmium tetroxide with 1.5% potassium ferrocyanide, 20 minutes in 1% thiocarbohydrazide, 30 minutes in 2% osmium tetroxide, overnight incubation in 1% uranyl acetate, 30 minutes in lead aspartate, dehydration in increasing ethanol concentrations (50%, 70%, 90%, and 100% ethanol), and embedding in Epon-812. Ultrathin sections of 90 nm were observed using a FEI Tecnai G2 operating at 200 kV.

Genome sequencing, assembly and annotation

Noumeavirus genomic DNA was recovered from 2×10^9 purified particles using the Purelink™ Genomic DNA Extraction Mini Kit (Life Technologies) according to the manufacturer recommendations. Two micrograms of purified viral DNA were sent to Lausanne Genomic

Technologies Facility (Lausanne, Switzerland) where they were used to generate paired-end sequence (2x100 bp) reads using the HiSeq platform. We retained 28,912,450 paired reads (overall Q > 32). We used idba_ud v1.1.1[56] to assemble this dataset into a 376,295 bp longest contig. The 88nt at 5' and 3' ends of this contig were identical and could be exchanged with one another suggesting a circular genome. This was confirmed by a PCR (data not shown). The redundant part was then removed leading to the longest contig of 376,207nt. We used bowtie 0.12.7[57] to align all pairs to the assembly allowing a 3% error rate. 99.95% of the mapped reads were aligned to the longest contig with 0.1% of mismatch and a mean coverage of 3,547. This genome is almost entirely colinear to the published genomes of Port-Miou[7] (349,275 nt, 410 predicted proteins) and Lausannevirus[2] (346,754 nt, 444 predicted proteins). As these genomes are circular (or linear, circularly permutated and terminally redundant), sequence numbering is arbitrary. To facilitate the identification of orthologous genes, pairwise whole genome alignments were generated between Noumeavirus and its closest relatives using LAST[58]. The sequence was screened for palindromes (potential hairpin) using rnamotif[59] as described elsewhere[26]. Genemark 4.29[60] was used to *ab initio* predict protein coding genes from the Noumeavirus genome sequence. The functional annotation of the 452 predicted proteins was achieved by combining different methods. Blast[61] search was performed versus the whole non-redundant protein database at NCBI[61]. A domain search was performed against Pfam[62] 28.0, TIGRFAM[63] 15.0, SMART[64] 6.2, ProDom[65] 2006.1, PANTHER[66] 9.0, Prosite[67] 20.113, Hamap[68] 201502.04 using interproscan[69] 5.14-53. A CD-search[70] was achieved against the conserved domain database at NCBI. Potential (trans)membrane proteins were predicted using Phobius[71]. Nuclear signal peptides were predicted by NucPred[72], PredictNLS[73], NLStradamus[74] and Nucleolar localization sequence detector[75] (NOD). The results were manually compiled and the most representative function was assigned to each sequence. Seven genes initially overlooked in the initial annotation of Melbournevirus were identified through the comparison of the Noumeavirus and Melbournevirus virions proteomes and their mapping to the genomic data.

Prediction of putative exon-intron gene structures

Alignments revealing discontinuities between reference genes and their Noumeavirus and Melbournevirus homologs were specifically scrutinized in search for possible donor (GT) and acceptor (AG) splice sites flanking potential introns. We also used Exonerate[24] with the "protein2genome" model to map UniProtKB/Swiss-Prot protein sequences to systematically investigate intron-containing genes in the Noumeavirus genome. We then extracted genomic regions covered by at least 25 aligned translated genes, among which at least 90 % were predicted to contain an intron at that position. We identified 8 candidate introns and experimentally tested three of them using RT-PCR.

Intron and polyadenylation site ("hairpin rule") validations

RNA was extracted using the RNeasy Mini kit (QIAGEN, Venlo, Netherlands) according to the manufacturer's protocol. Briefly, cells were resuspended in the provided buffer and disrupted by -80°C freezing and thawing, and shaken vigorously. Total RNA was eluted with 30µL of RNase free molecular biology grade water (5 PRIME). Total RNA was quantified on the Nanodrop spectrophotometer (Thermo Scientific). Poly(A) enrichment was performed (Life Technologies, Dynabeads oligodT_{25}) and first-strand cDNA poly(A) synthesis was performed with the PrimeScript Reverse Transcriptase (Clontech Laboratories) using an oligo$(dT)_{24}$ primer and then treated with RNase H (New England Biolabs). PCR reactions were performed using 1 units of Phusion DNA Polymerase (Thermo Scientific) in a 50µL final volume.

- 3' and 5' specific primers were designed to cover the region predicted to separate the two RPB1-like domains and PCR products were sequenced.

- 3' and 5' primers were designed to cover the predicted spliceosomal introns of 3 Noumeavirus genes (NMV_125, NMV_252 and NMV_436). The PCR products obtained on the genomic DNA and on the cDNAs were found to migrate at the same location on an agarose gel.

- To validate the polyadenylation site (i.e. the hairpin rule), we generated amplicons using 5' gene-specific primers and an oligo(dT)$_{24}$ primer. PCR products were sequenced. Structure prediction of the hairpin was performed using the RNAfold server[76]. All PCR programs were designed using manufacturer's recommendations.

Proteome analyses

The same number of virions (10^9) for Noumeavirus (stored at 4°C) and Melbournevirus (stored at −80 °C in 7.5% DMSO) were lysed in 250 µL of Tris-HCl 50 mM, SDS 2% and DTT 60 mM pH 7.5 before protein quantification. 5 µg of proteins were resuspended in gel loading buffer (100 mM Tris-HCl pH 6.8, SDS 2%, glycerol 4%, β-mercaptoethanol 5% and traces of bromophenol blue), heated for 10 min at 95°C before being analysed as previously described[37]. Briefly, proteins were stacked in the top of a 4–12% NuPAGE gel (Invitrogen) before R-250 Coomassie blue staining and in-gel digestion with trypsin (sequencing grade, Promega). The equivalent of 1µg of resulting peptides were analysed in duplicate by online nanoLC-MS/MS (Ultimate 3000, Dionex and LTQ-Orbitrap Velos Pro, Thermo Scientific) using a 120-min gradient. Peptides and proteins were identified and quantified using MaxQuant software[77] (version 1.5.3.30). Spectra were searched against the Noumeavirus or Melbournevirus protein sequence databases, *A. castellanii* protein sequence database, and the frequently observed contaminants database embedded in MaxQuant. Minimum peptide length was set to 7 aa. Maximum false discovery rates were set to 0.01 at peptide and protein levels. Intensity-based absolute quantification (iBAQ)[78] values were calculated from MS intensities of unique+razor peptides. Only proteins identified by a minimum of 2 peptides were further considered. iBAQ values from both analytical replicates were column-wise normalized and the protein ranking in particle was deducted from to the sum of the two normalized iBAQ values per protein. Given the correlation between the two virion proteomes we considered that the use of biological replicates was superfluous.

The ranking of *A. castellanii* proteins in uninfected cells was computed from a proteome analysis previously performed[37]. Uninfected cells were grown in the exact same conditions as cells infected with Noumeavirus and Melbournevirus (growth medium, temperature, volume and number of cells). To compare the presence and abundancy of the proteins involved in transcription between the proteomes of exclusively cytoplasmic viruses and Marseilleviridae, we reanalysed published data[26] obtained on *Pithovirus sibericum* and Mimivirus virions using the method described in this study.

Fluorescence study of infected *A. castellanii* cells

Two Mollivirus Methyltransferases containing either a nucleolar or a nuclear localization signal (NoLS, ML216 or NLS, ML135) were cloned into the pGAPDH-GFP amoebal expression plasmid[79] to yield a C-terminally GFP-tagged protein targeted to the nucleolus or the nucleus. The *A. castellanii* SUMO protein fused with a C-terminal GFP tag was used to monitor the passive diffusion of small proteins out of the nucleus.

Acanthamoeba castellanii cells were transfected with the transfection vectors using the calcium phosphate precipitation protocol used for *Dictyostelium discoideum* as described[80]. Selection of transformed cells was initially performed at 20µg/mL Neomycin and increased up to 60 µg/mL within a couple of weeks.

Transfected *A. castellanii* cells were grown in a 12-well plate and infected with Noumeavirus at a MOI of 50 except for the negative control. At 30min, 1h, 2h and 4h post-infection, cells were scraped, centrifuged for 5 min at 500g and the pellet was resuspended and fixed with PBS containing 3.7% formaldehyde for 30min at room temperature. After one wash with PBS buffer, 4µl of the cells were mixed with 4µl of VECTASHIELD mounting medium with DAPI and the fluorescence was observed using a Zeiss Axio Observer Z1 inverted microscope using a 63x objective lens associated with a 1.6x Optovar for DAPI or GFP fluorescence recording.

References

1. Boyer, M. *et al.* Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 21848–21853 (2009).
2. Thomas, V. *et al.* Lausannevirus, a giant amoebal virus encoding histone doublets. *Environ. Microbiol.* **13,** 1454–1466 (2011).
3. Aherfi, S. *et al.* Complete genome sequence of Cannes 8 virus, a new member of the proposed family 'Marseilleviridae'. *Virus Genes* **47,** 550–555 (2013).
4. Aherfi, S., La Scola, B., Pagnier, I., Raoult, D. & Colson, P. The expanding family Marseilleviridae. *Virology* **466–467,** 27–37 (2014).
5. Aherfi, S. *et al.* Complete genome sequence of Tunisvirus, a new member of the proposed family Marseilleviridae. *Arch. Virol.* **159,** 2349–2358 (2014).
6. Doutre, G., Philippe, N., Abergel, C. & Claverie, J.-M. Genome analysis of the first Marseilleviridae representative from Australia indicates that most of its genes contribute to virus fitness. *J. Virol.* **88,** 14340–14349 (2014).
7. Doutre, G. *et al.* Complete Genome Sequence of a New Member of the Marseilleviridae Recovered from the Brackish Submarine Spring in the Cassis Port-Miou Calanque, France. *Genome Announc.* **3,** (2015).
8. Dornas, F. *et al.* A Brazilian Marseillevirus Is the Founding Member of a Lineage in Family Marseilleviridae. *Viruses* **8,** 76 (2016).
9. Takemura, M. Draft Genome Sequence of Tokyovirus, a Member of the Family Marseilleviridae Isolated from the Arakawa River of Tokyo, Japan. *Genome Announc.* **4,** (2016).
10. Dos Santos, R. N. *et al.* A new marseillevirus isolated in Southern Brazil from Limnoperna fortunei. *Sci. Rep.* **6,** 35237 (2016).
11. Popgeorgiev, N. *et al.* Marseillevirus-like virus recovered from blood donated by asymptomatic humans. *J. Infect. Dis.* **208,** 1042–1050 (2013).
12. Popgeorgiev, N., Michel, G., Lepidi, H., Raoult, D. & Desnues, C. Marseillevirus adenitis in an 11-month-old child. *J. Clin. Microbiol.* **51,** 4102–4105 (2013).

13. Popgeorgiev, N. *et al.* Marseillevirus prevalence in multitransfused patients suggests blood transmission. *J. Clin. Virol. Off. Publ. Pan Am. Soc. Clin. Virol.* **58,** 722–725 (2013).
14. Goodman, J. L. Marseillevirus, blood safety, and the human virome. *J. Infect. Dis.* **208,** 1039–1041 (2013).
15. Boughalmi, M. *et al.* First isolation of a Marseillevirus in the Diptera Syrphidae Eristalis tenax. *Intervirology* **56,** 386–394 (2013).
16. Sauvage, V. *et al.* No evidence of Marseillevirus-like virus presence in blood donors and recipients of multiple blood transfusions. *J. Infect. Dis.* **210,** 2017–2018 (2014).
17. Kutikhin, A. G., Yuzhalin, A. E. & Brusina, E. B. Mimiviridae, Marseilleviridae, and virophages as emerging human pathogens causing healthcare-associated infections. *GMS Hyg. Infect. Control* **9,** Doc16 (2014).
18. Aherfi, S. *et al.* Marseillevirus in lymphoma: a giant in the lymph node. *Lancet Infect. Dis.* (2016). doi:10.1016/S1473-3099(16)30051-2
19. Pozzetto, B. & Garraud, O. [New viral risks in blood transfusion by 2016]. *Transfus. Clin. Biol. J. Société Fr. Transfus. Sang.* **23,** 20–27 (2016).
20. Colson, P. *et al.* 'Marseilleviridae', a new family of giant viruses infecting amoebae. *Arch. Virol.* **158,** 915–920 (2013).
21. Arantes, T. S. *et al.* The Large Marseillevirus Explores Different Entry Pathways by Forming Giant Infectious Vesicles. *J. Virol.* **90,** 5246–5255 (2016).
22. Jakob, N. J., Müller, K., Bahr, U. & Darai, G. Analysis of the first complete DNA sequence of an invertebrate iridovirus: coding strategy of the genome of Chilo iridescent virus. *Virology* **286,** 182–196 (2001).
23. Nawrocki, E. P. *et al.* Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* **43,** D130–D137 (2015).
24. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6,** 31 (2005).

590 25. Erives, A. J. *Eukaryotic core histone diversification in light of the histone doublet and DNA topo II*
591 *genes of Marseilleviridae.* bioRxiv 022236; doi: http://dx.doi.org/10.1101/022236 (2015).
592 26. Legendre, M. *et al.* Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses
593 with a pandoravirus morphology. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 4274–4279 (2014).
594 27. Priet, S., Lartigue, A., Debart, F., Claverie, J.-M. & Abergel, C. mRNA maturation in giant viruses:
595 variation on a theme. *Nucleic Acids Res.* **43,** 3776–3788 (2015).
596 28. Benarroch, D., Smith, P. & Shuman, S. Characterization of a trifunctional mimivirus mRNA
597 capping enzyme and crystal structure of the RNA triphosphatase domain. *Struct. Lond. Engl.*
598 *1993* **16,** 501–12 (2008).
599 29. Byrne, D. *et al.* The polyadenylation site of Mimivirus transcripts obeys a stringent 'hairpin rule'.
600 *Genome Res.* **19,** 1233–1242 (2009).
601 30. Legendre, M. *et al.* mRNA deep sequencing reveals 75 new genes and a complex transcriptional
602 landscape in Mimivirus. *Genome Res.* **20,** 664–674 (2010).
603 31. Chung, C.-S. *et al.* Vaccinia virus proteome: identification of proteins in vaccinia virus intracellular
604 mature virion particles. *J. Virol.* **80,** 2127–2140 (2006).
605 32. Arslan, D., Legendre, M., Seltzer, V., Abergel, C. & Claverie, J.-M. Distant Mimivirus relative with a
606 larger genome highlights the fundamental features of Megaviridae. *Proc. Natl. Acad. Sci. U. S. A.*
607 **108,** 17486–17491 (2011).
608 33. Dunigan, D. D. *et al.* Paramecium bursaria chlorella virus 1 proteome reveals novel architectural
609 and regulatory features of a giant virus. *J. Virol.* **86,** 8821–8834 (2012).
610 34. Philippe, N. *et al.* Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of
611 parasitic eukaryotes. *Science* **341,** 281–286 (2013).
612 35. Fischer, M. G., Kelly, I., Foster, L. J. & Suttle, C. A. The virion of Cafeteria roenbergensis virus
613 (CroV) contains a complex suite of proteins for transcription and DNA repair. *Virology* **466–467,**
614 82–94 (2014).

615 36. Li, W. *et al.* Virion-associated viral proteins of a Chinese giant salamander (Andrias davidianus)
616 iridovirus (genus Ranavirus) and functional study of the major capsid protein (MCP). *Vet.*
617 *Microbiol.* **172,** 129–139 (2014).
618 37. Legendre, M. *et al.* In-depth study of *Mollivirus sibericum* , a new 30,000-y-old giant virus
619 infecting *Acanthamoeba*. *Proc. Natl. Acad. Sci.* **112,** E5327–E5335 (2015).
620 38. Doellinger, J., Schaade, L. & Nitsche, A. Comparison of the Cowpox Virus and Vaccinia Virus
621 Mature Virion Proteome: Analysis of the Species- and Strain-Specific Proteome. *PloS One* **10,**
622 e0141527 (2015).
623 39. Reteno, D. G. *et al.* Faustovirus, an asfarvirus-related new lineage of giant viruses infecting
624 amoebae. *J. Virol.* **89,** 6585–6594 (2015).
625 40. Shin, J.-B. *et al.* Molecular architecture of the chick vestibular hair bundle. *Nat. Neurosci.* **16,**
626 365–374 (2013).
627 41. Greber, U. F. & Fassati, A. Nuclear import of viral DNA genomes. *Traffic Cph. Den.* **4,** 136–143
628 (2003).
629 42. Tran, B. N. *et al.* Novel histone H3 binding protein ORF158L from the Singapore grouper
630 iridovirus. *J. Virol.* **85,** 9159–9166 (2011).
631 43. Song, W. J. *et al.* Functional genomics analysis of Singapore grouper iridovirus: complete
632 sequence determination and proteomic analysis. *J. Virol.* **78,** 12576–12590 (2004).
633 44. Renesto, P. *et al.* Mimivirus giant particles incorporate a large fraction of anonymous and unique
634 gene products. *J. Virol.* **80,** 11678–11685 (2006).
635 45. Song, W., Lin, Q., Joshi, S. B., Lim, T. K. & Hew, C.-L. Proteomic studies of the Singapore grouper
636 iridovirus. *Mol. Cell. Proteomics MCP* **5,** 256–264 (2006).
637 46. Wang, F., Bi, X., Chen, L. M. & Hew, C.-L. ORF018R, a highly abundant virion protein from
638 Singapore grouper iridovirus, is involved in serine/threonine phosphorylation and virion
639 assembly. *J. Gen. Virol.* **89,** 1169–1178 (2008).

47. Wong, C. K., Young, V. L., Kleffmann, T. & Ward, V. K. Genomic and proteomic analysis of invertebrate iridovirus type 9. *J. Virol.* **85,** 7900–7911 (2011).
48. Ince, I. A., Boeren, S. A., van Oers, M. M., Vervoort, J. J. M. & Vlak, J. M. Proteomic analysis of Chilo iridescent virus. *Virology* **405,** 253–258 (2010).
49. Claverie, J.-M. & Abergel, C. Open questions about giant viruses. *Adv. Virus Res.* **85,** 25–56 (2013).
50. Bell, P. J. Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? *J. Mol. Evol.* **53,** 251–256 (2001).
51. Takemura, M. Poxviruses and the origin of the eukaryotic nucleus. *J. Mol. Evol.* **52,** 419–425 (2001).
52. Claverie, J.-M. Viruses take center stage in cellular evolution. *Genome Biol.* **7,** 110 (2006).
53. Forterre, P. & Gaïa, M. Giant viruses and the origin of modern eukaryotes. *Curr. Opin. Microbiol.* **31,** 44–49 (2016).
54. Dong, C.-F. *et al.* Global landscape of structural proteins of infectious spleen and kidney necrosis virus. *J. Virol.* **85,** 2869–2877 (2011).
55. Shuang, F. *et al.* Virions proteins of an RSIV-type megalocytivirus from spotted knifejaw Oplegnathus punctatus (SKIV-ZJ07). *Virology* **437,** 89–99 (2013).
56. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinforma. Oxf. Engl.* **28,** 1420–1428 (2012).
57. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
58. Kiełbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21,** 487–493 (2011).
59. Okonechnikov, K., Golosova, O., Fursov, M. & the UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28,** 1166–1167 (2012).

60. Besemer, J., Lomsadze, A. & Borodovsky, M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **29,** 2607–2618 (2001).
61. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **44,** D7–D19 (2016).
62. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44,** D279-285 (2016).
63. Haft, D. H. *et al.* TIGRFAMs and Genome Properties in 2013. *Nucleic Acids Res.* **41,** D387-395 (2013).
64. Letunic, I., Doerks, T. & Bork, P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* **40,** D302-305 (2012).
65. Bru, C. *et al.* The ProDom database of protein domain families: more emphasis on 3D. *Nucleic Acids Res.* **33,** D212-215 (2005).
66. Mi, H., Muruganujan, A. & Thomas, P. D. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* **41,** D377-386 (2013).
67. Sigrist, C. J. A. *et al.* New and continuing developments at PROSITE. *Nucleic Acids Res.* **41,** D344-347 (2013).
68. Pedruzzi, I. *et al.* HAMAP in 2013, new developments in the protein family classification and annotation system. *Nucleic Acids Res.* **41,** D584-589 (2013).
69. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinforma. Oxf. Engl.* **30,** 1236–1240 (2014).
70. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43,** D222-226 (2015).
71. Käll, L., Krogh, A. & Sonnhammer, E. L. L. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338,** 1027–1036 (2004).

692 72. Brameier, M., Krings, A. & MacCallum, R. M. NucPred--predicting nuclear localization of proteins.
693 *Bioinforma. Oxf. Engl.* **23,** 1159–1160 (2007).
694 73. Kaján, L. *et al.* Cloud prediction of protein structure and function with PredictProtein for Debian.
695 *BioMed Res. Int.* **2013,** 398968 (2013).
696 74. Nguyen Ba, A. N., Pogoutse, A., Provart, N. & Moses, A. M. NLStradamus: a simple Hidden
697 Markov Model for nuclear localization signal prediction. *BMC Bioinformatics* **10,** 202 (2009).
698 75. Scott, M. S., Troshin, P. V. & Barton, G. J. NoD: a Nucleolar localization sequence detector for
699 eukaryotic and viral proteins. *BMC Bioinformatics* **12,** 317 (2011).
700 76. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R. & Hofacker, I. L. The Vienna RNA Websuite.
701 *Nucleic Acids Res.* **36,** W70–W74 (2008).
702 77. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-
703 range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–
704 1372 (2008).
705 78. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature*
706 **473,** 337–342 (2011).
707 79. Bateman, E. Expression plasmids and production of EGFP in stably transfected Acanthamoeba.
708 *Protein Expr. Purif.* **70,** 95–100 (2010).
709 80. Gaudet, P., Pilcher, K. E., Fey, P. & Chisholm, R. L. Transformation of Dictyostelium discoideum
710 with plasmid DNA. *Nat. Protoc.* **2,** 1317–1324 (2007).
711

712 **Acknowledgments**

713 We are grateful to Dr. Laurent Bordez who collected the sample from which Noumeavirus was
714 isolated and to Pr. E. Bateman for his kind gift of the original pGAPDH-GFP vector. We thank Sylvain
715 Poquet for his help on Melbournevirus fluorescence studies, Dr. J.-P. Chauvin, Dr. A. Kosta, F. Richard
716 and A. Aouane for their expert assistance on the imaging platforms and J.-M Alempic and L. Bertaux

for their help all along the project. YC and MT thank the support of the bottom-up platform and informatics group of EDyP.

Author contributions: S.J., Y.C., J-M.C and C.A. conceived and designed the research; E.F., S.J., S.S., M. L., M.T., Y.C., J-M.C., and C.A. performed research; E.F., S.S., M.L. and M.T. contributed analytic tools; S.J., Y.C., M.T., E.F., J-M.C., and C.A. analysed data, S.J. J-M.C. and C.A. wrote the paper.

Competing Financial Interests

This work was partially supported by the French National Research Agency ANR-14-CE14-0023-01, by a DGA-MRIS scholarship and by the Provence-Alpes-Côte-d'Azur région (2010 12125). Proteomic experiments were partly supported by the ProFi grant (ANR-10-INBS-08-01).

Materials & Correspondence

*to whom correspondence should be addressed. Email: chantal.abergel@igs.cnrs-mrs.fr or jean-michel.claverie@univ-amu.fr

Data availability

The Noumeavirus genome sequence was deposited in GenBank (accession no. KX066233).The mass spectrometry proteomics data was deposited to the ProteomeXchange Consortium (proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD003910.

Figure Legends

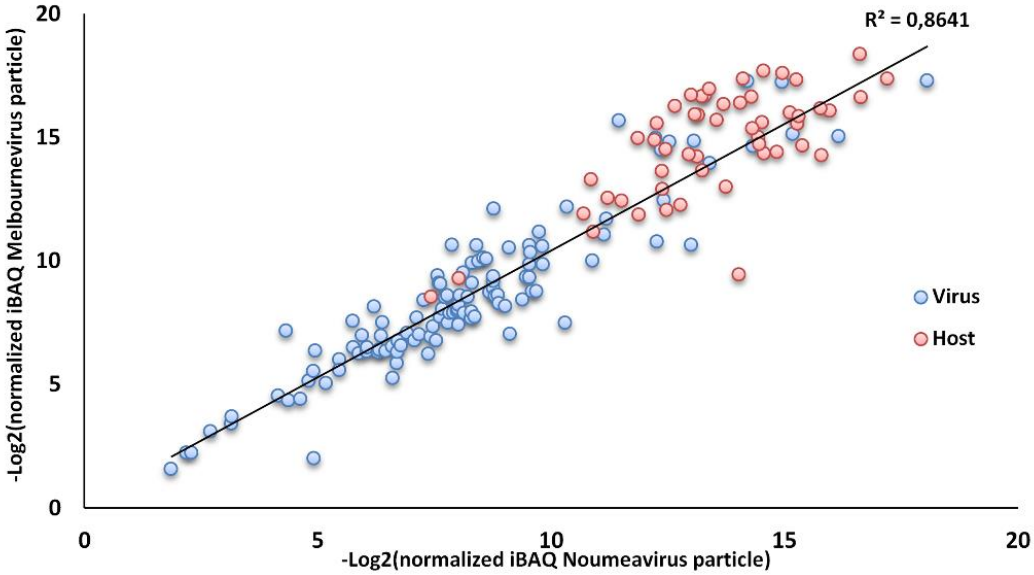


Figure 1: Proteomic analyses of Noumeavirus and Melbournevirus virion compositions.

Comparison of the abundances of host encoded (red) and virally encoded orthologous proteins (blue) in Noumeavirus and Melbournevirus. Higher values correspond to lower proteins abundances.

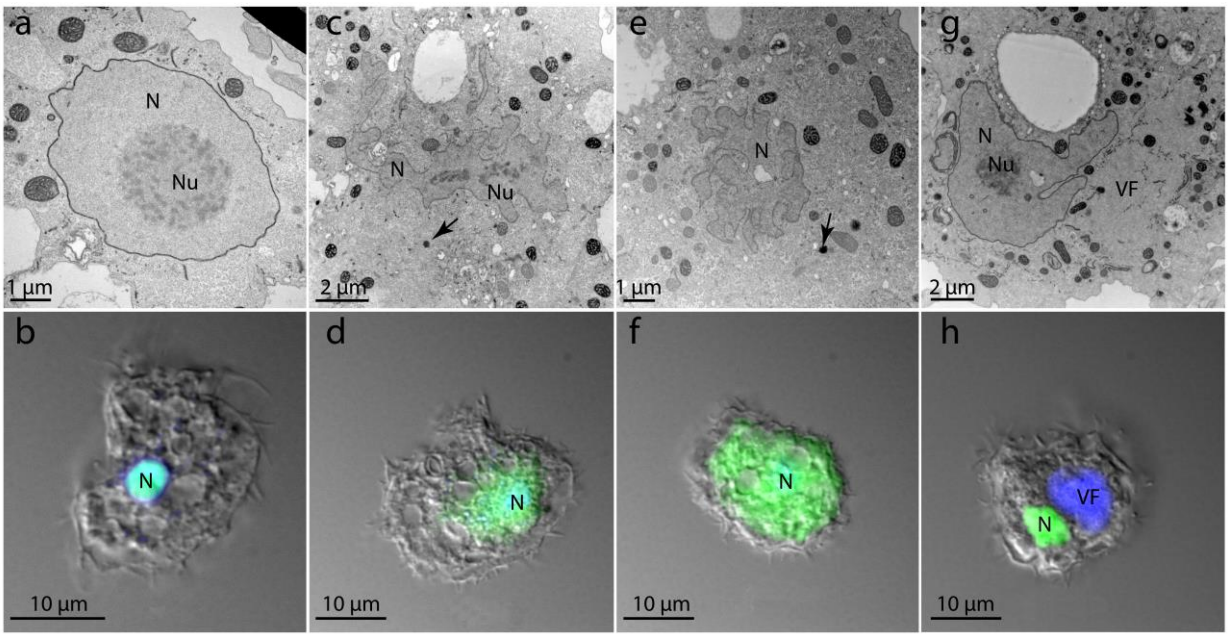
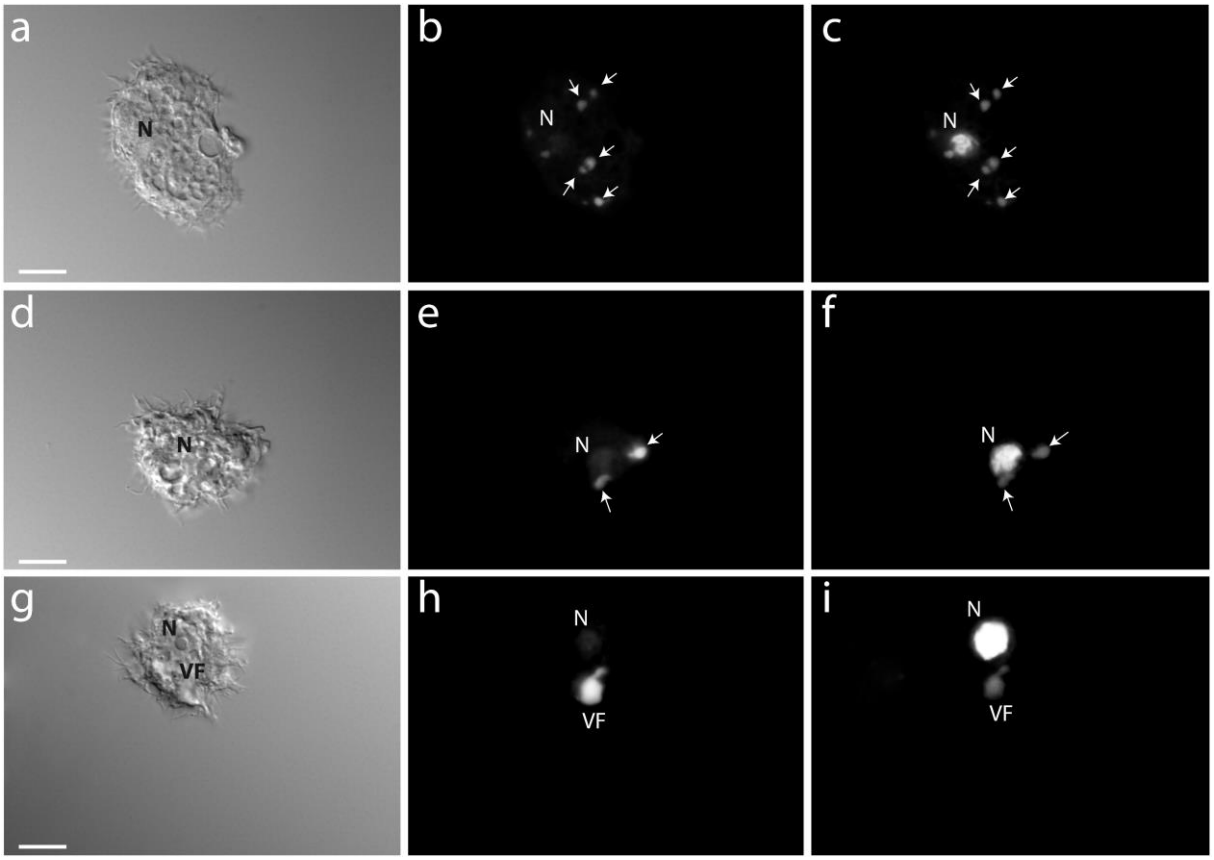


Figure 2: TEM (A, C, E, G) and Fluorescence study (B, D, F, H) of *A. castellanii* cells expressing ML135-GFP at the nucleus: a, b] Nucleus (N) in non-infected cells. The nucleolus (Nu) is electron-dense and compact. DAPI and GFP staining are located in the nucleus. c,d] 30 minutes pi by

744 Noumeavirus, the nucleus is drastically modified with many circumvolutions. The nucleolus appears
745 disorganized and less electron-dense. DAPI staining remains at the nucleus, the nuclear GFP
746 fluorescence is spreading in the cytoplasm. e,f] 1 h pi, the nucleus is still deformed. The GFP staining
747 is no longer located at the nucleus and fills the cytoplasm. Electron dense nucleoids can be seen in
748 the cytoplasm (c, e, black arrows, Fig 3, white arrows). The nucleus is still DAPI-stained. g,h] 4 h pi,
749 the nucleus recovered its spherical shape with electron-dense and compact nucleolus and coexists
750 with viral factories (VF). The GFP fluorescence is back into the nucleus; DAPI staining is intense in the
751 viral factories.



752

753 **Figure 3: Observation of A.castellanii cells infected by Noumeavirus between 2 and 3h pi.** a, d, g] cells
754 were observed by interference contrast (DIC, scale bar 10 µm). DAPI-stained early nucleoids (b, white
755 arrows) and developing viral factories (VF, e, h) produce very intense fluorescence masking the DAPI
756 staining at the nucleus (N). GFP staining co-localizes with the viral nucleoids (c, white arrows) and the
757 early VF (f, white arrows). (i) In contrast with the strongly fluorescent nucleus, the mature VF is barely
758 GFP stained. The same phenotypes were observed in cells infected by Melbournevirus.

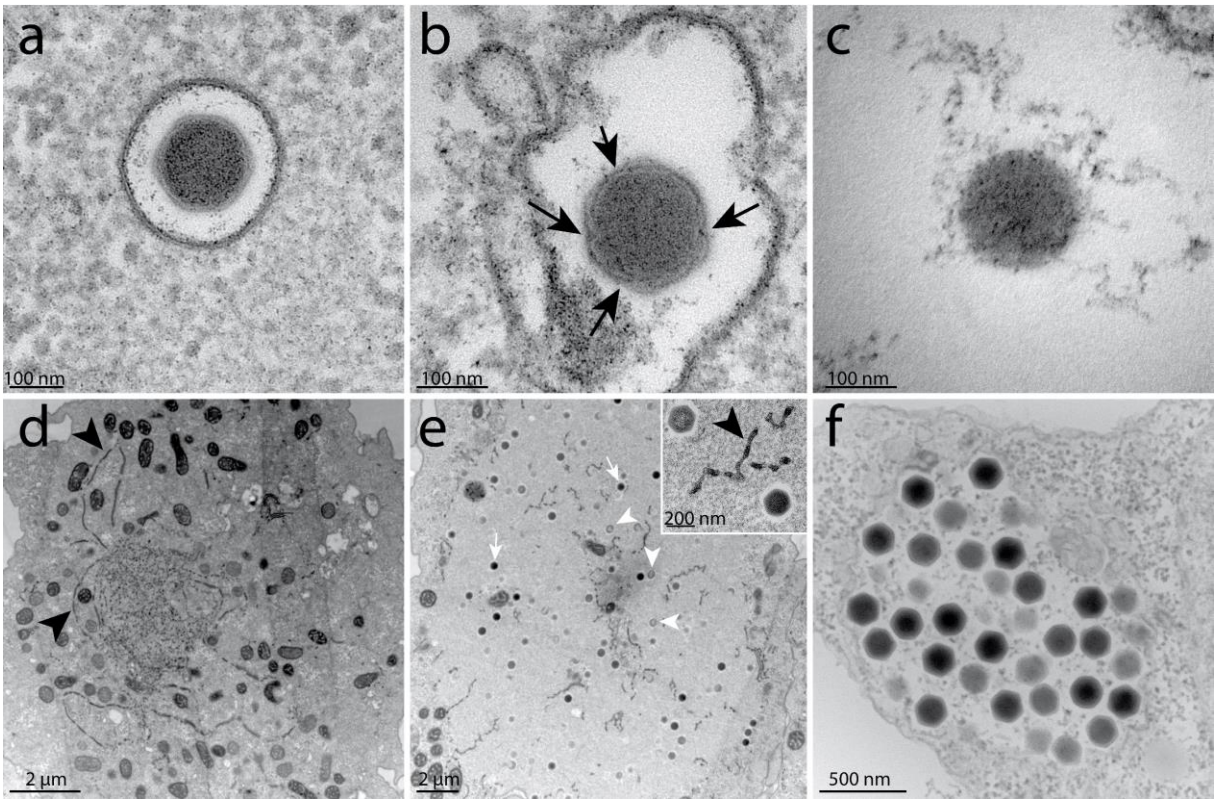


Figure 4: Ultrathin-section TEM imaging of Noumeavirus-infected Acanthamoeba cells. a] 10 minutes pi, virions have been engulfed and are in vacuoles. b] 20 minutes pi, virions in vacuoles exhibit holes (black arrows) in their external translucent shell and appear more spherical. c] 30 minutes pi virions have entirely lost their external shell and appear as spherical electron-dense nucleoids which can be seen in vacuoles as well as in the cell cytoplasm (Fig. 2c & e, black arrows). d] 1 h pi infected cells. Electron-dense tubular structures (black arrowheads) appear in the cell cytoplasm. e] 4 h pi viral factories (VF) settle in the cell cytoplasm next to the nucleus and the cell organelles are pushed at their periphery. The new virions are assembled and the electron-dense mature (white arrowheads) and immature (white arrows) virions are scattered in the same VF. Inset: immature and mature virions inside the VF along with tubular structures (arrowheads). f] 7 h pi, the newly-synthesized virions are gathered into vacuoles inside the cell cytoplasm before being released in the extracellular environment.

B Noumeavirus : Supplementary Material and Methods

Noumeavirus replication relies on a transient remote control of the host nucleus

Elisabeth Fabre[a,1], Sandra Jeudy[a,1], Sébastien Santini[a], Matthieu Legendre[a], Mathieu Trauchessec[b,c,d], Yohann Couté[b,c,d], Jean-Michel Claverie[a, e,*], Chantal Abergel[a,*]

Supplementary Material & Methods

Phylogenetic analyses

Noumeavirus was positionned within the Marseilleviridae family using the orthologous DNA polymerase protein (DNApol) sequences from Brazilian marseillevirus (YP_009238893), Cannes 8 virus (AGV01694), Lausannevirus (YP_004347308), Insectomime virus (AHA45970), Melbournevirus (AIT54904), Port-miou virus (ALH07009), Tunisvirus fontaine2 (AHC54969) and Tokyovirus (NC_030230). Since the DNApol sequences of Marseillevirus (NC_013756) and Golden mussel marseillevirus (NC_031465.1) were not annotated, they were recovered using the "protein2genome" option of Exonerate[1]. The Iridovirus Invertebrate iridescent virus 3 (YP_654692) was used as outgroup. We then performed a multiple alignment using Mafft[2] with the L-INS-i parameter. The best model for phylogenetic tree reconstruction was chosen using Protest[3] (i.e. LG+G) and RaxML[4] was finally used to build the tree.

We next used OrthoMCL[5] with standard parameters and predicted proteins with a sequence length > 100 amino acids to generate clusters of Marseilleviridae homologs. We retained clusters containing 1:1 orthologs of each virus (n=21), aligned them, concatenated the alignments and computed the tree as for the above phylogenetic tree.

RNA extraction, mRNA enrichment and RT-PCR of virion-associated mRNAs

To investigate whether mRNAs could be packaged in the viral particles, we extracted RNA from purified virions and used Noumeavirus and Melbournevirus infected *A. castellanii* cells as a positive control. Three different quantities of Noumeavirus and Melbournevirus purified particles (5.10^9, 5.10^{10} and 5.10^{11}) were treated with 20µg/mL DNase-free RNase (Life Technologies) for 10 minutes at room temperature. The viral pellets were washed 3 times in PBS and incubated with 1U/µL of RNasin ribonuclease inhibitor (Promega) for one hour at 37°C. After 3 additional washes in PBS, the pellets were resuspended in RLT buffer supplemented with B-mercaptoethanol. For the the positive controls, infections with Noumeavirus and Melbournevirus were performed in 7 T25 flasks containing 100 000 *A. castellanii* cells per cm^2 at MOI 50. The flasks were collected every hour from 1h to 7h pi. All cells pellets were pooled and resuspended in RLT buffer supplemented with B-mercaptoethanol.

RNAs were extracted using the RNeasy mini kit (Qiagen) according to the manufacturer protocol. One step of RNase-free DNase digestion was performed on-column using the RNase-Free DNase Set (Qiagen) and total RNAs were eluted in RNase-free water. Excepted for 4µL, total RNAs were used for mRNA enrichment (Life Technologies, Dynabeads oligodT25). The 4 µL total RNAs and purified mRNAs were reverse transcribed using an oligodT(25) primer and the SMARTScribe Reverse Transcriptase (Clontech Laboratories) and then treated with RNase H. Noumeavirus and Melbourne virions were checked for the presence of viral (NMV_11 and MEL_47) and cellular RPB1 transcripts, for the RPB2 subunit of Melbournevirus (MEL_22), for the transcripts of the most abundant proteins in the virions (NMV_189 and MEL_236), the major capsid proteins (NMV_116 and MEL_305) and for the cellular polyA polymerase mRNA. PCR products were analyzed on 1.2% agarose gel and sent for sequencing.

To estimate if the amplicons corresponding to the RPB1 mRNA in Melbournevirus virions were compatible with one molecule of mRNA per virion, we added a quantity of commercial GFP mRNAs (OZ Biosciences) corresponding to one molecule per virion in the virions lysates using 10^9, 5.10^9 and 5.10^{10} viral particles. We then performed the mRNA purification, reverse transcription and PCR amplification as described above.

Results

After RNA purification using from $5x10^9$ to $5x10^{11}$ particles of the two virions, the RNA were quantified using the Qubit fluorometric quantification (Thermofisher). The corresponding signal was much lower than the DNA signal (3 to 10 times) despite multiple treatments with DNAse. As the viruses RPB1 genes contain a self-splicing intron and the host RPB1 contains 24 introns, we focused on their transcripts to investigate the presence of mRNA encoding for the virus and host transcription machinery in the two Marseilleviridae virions. No amplicons were obtained for the host RPB1 and PAP mRNAs on total RNA and mRNA extracted from the two virions, while clear amplicons were obtained from the infected cells. Amplicons of the virus-encoded RPB1 mRNA were obtained in infected cells and from the mRNA extracted from Melbournevirus when using $>10^{10}$ virions (Figure S1B). In contrast, none were obtained with Noumeavirus or when using $5x10^9$ Melbournevirus virions. The mRNA corresponding to the virus-encoded RPB2 was not detected in Melbournevirus virions. As a control for possible bystanders, the mRNA of the most abundant protein (NMV_189 and MEL_236) was detected in both virions, while the mRNA of the Major Capsid Protein was only detected in Melbournevirus, suggesting that, as for the RPB1 mRNA, they most likely corresponded to bystanders at less than one copy per virion (Figure S1C). This was confirmed by the use of the commercial GFP mRNA which produced an intense band on agarose gel revealed by BET with 1 molecule per virion even when using 10^9 particles, a quantity for which RPB1 was never amplified

from Noumeavirus or Melbournevirus particles (Figure S1A). This argues for the bystander nature of the mRNAs identified in the virions.

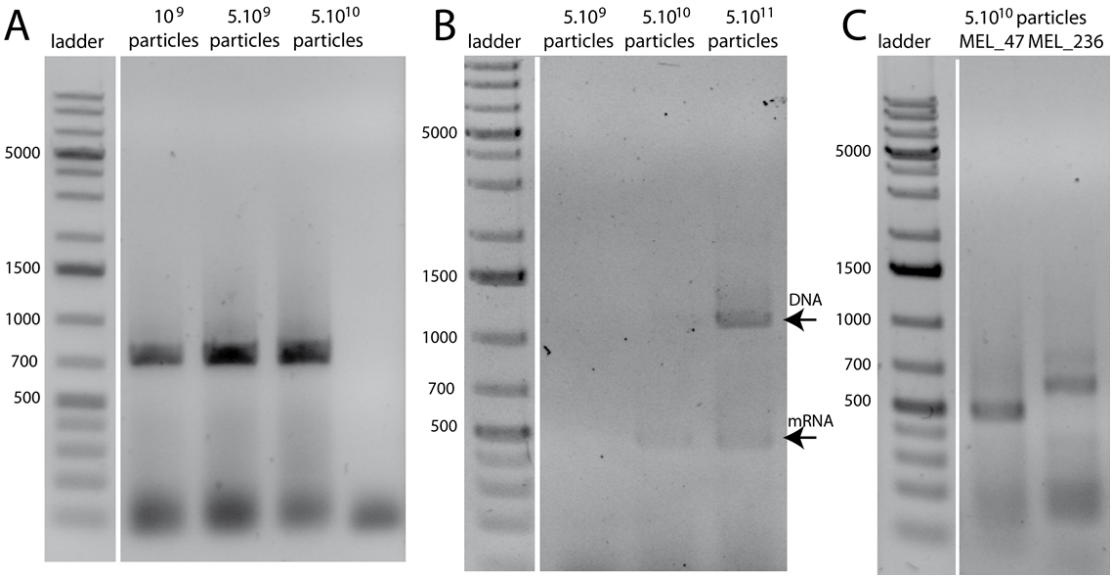


Figure S1: PCR gels

A) PCR product of the reverse transcribed GFP mRNA added to Melbournevirus virions extracts at the level of 1 molecule of mRNA per virion. The last lane corresponds to the not reverse transcribed GFP mRNA used as a negative control. B) PCR products of the reverse transcribed RPB1 (MEL_47) mRNA after RNA extraction and mRNA purification. The PCR product is only visible when using $>5.10^9$ viral particles. For 5.10^{11} particles the PCR product of the RPB1 DNA is more intense than the one of the mRNA. C) PCR products of the Melbournevirus reverse transcribed RPB1 and of the most abundant protein (MEL236).

Comparison of Noumeavirus with Lausannevirus and Melbournevirus

The best blast reciprocal hit of all protein sequences from Noumeavirus were computed against all protein sequences from Melbournevirus and Lausannevirus with an e-value threshold of $1e^{-5}$. Each resulting pair of sequence was then aligned using Needleman-Wunch algorithm from the EMBOSS suite to compute the percentage of identity between the two sequences (Table S4).

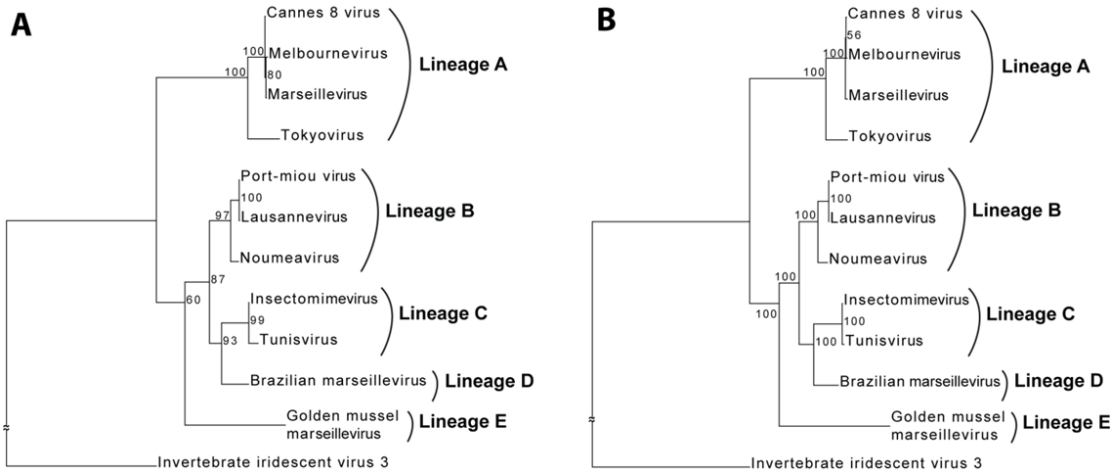


Figure S2: Phylogenetic analyses of Marseilleviridae

A) Maximum likelihood phylogenetic tree of the DNA polymerase of Marseilleviridae build using RaxML [4]. Invertebrate iridescent virus 3 DNA was used as the outgroup. Numbers at the nodes indicate percent bootstrap values. B) Maximum likelihood phylogenetic tree based on the concatenated alignment of the 21 orthologous genes shared by Marseilleviridae and Iridovirus Invertebrate iridescent virus 3.

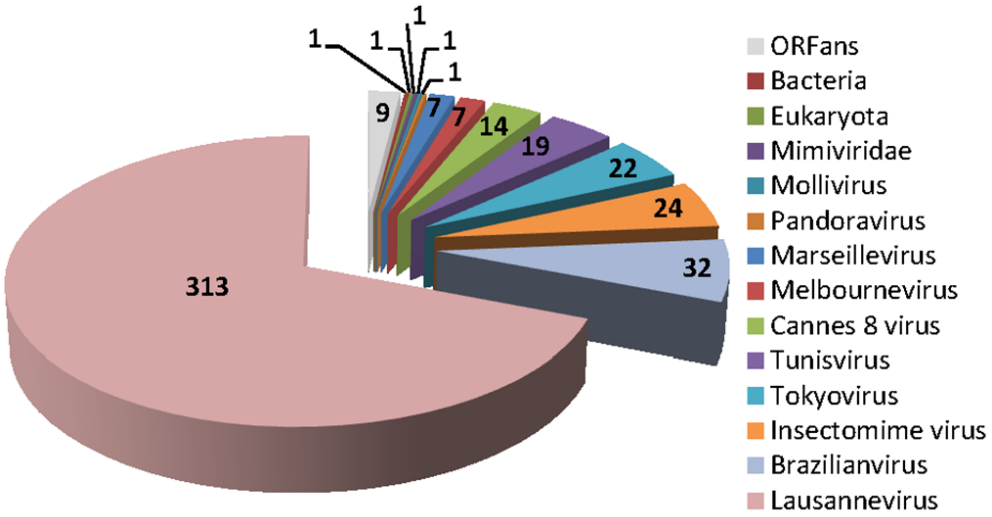


Figure S3: Distribution of the best-matching homologs to Noumeavirus predicted proteins. Best-matching homologous proteins were determined using BLASTP (E value < 10^{-5}) against the nonredundant (NR) database at the National Center for Biotechnology Information.

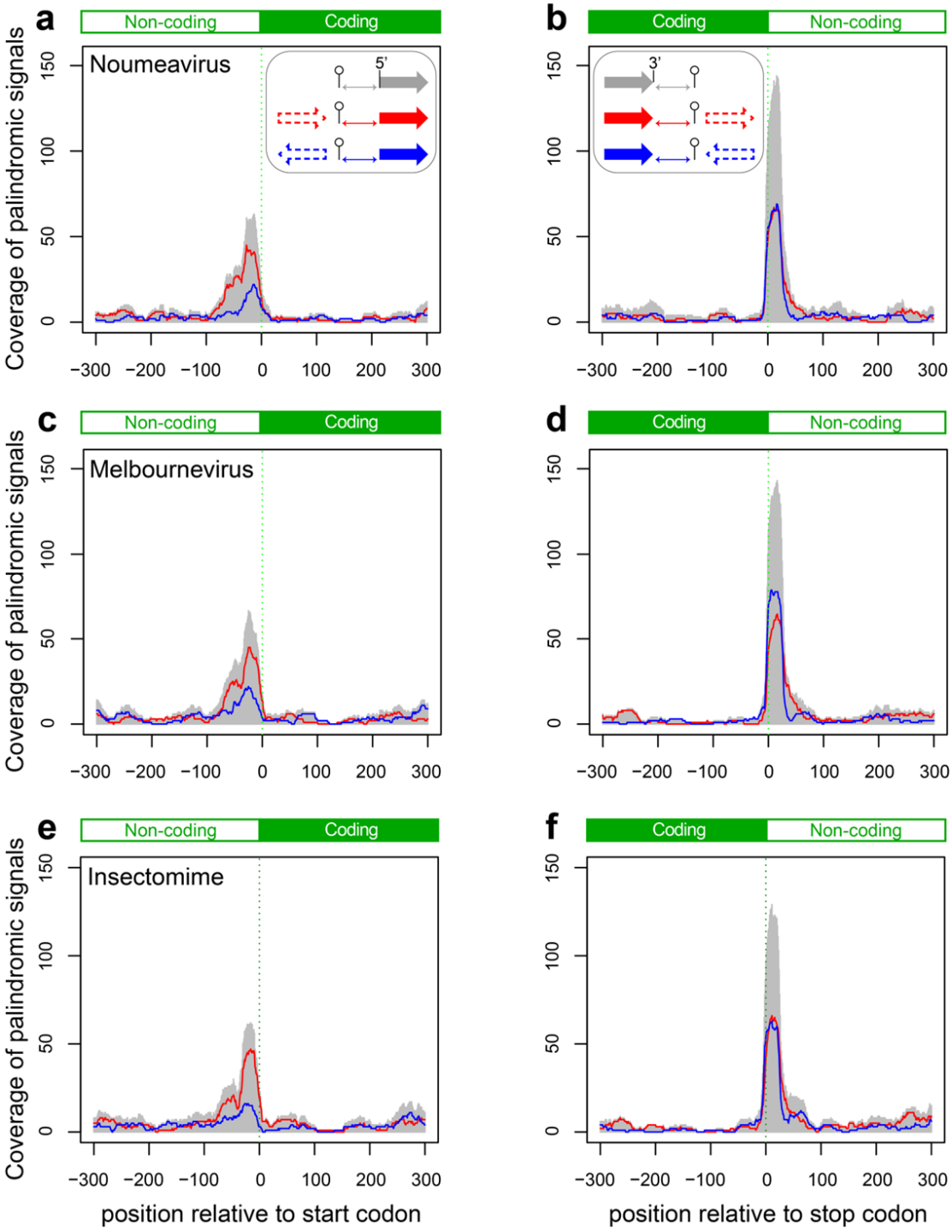


Figure S4: Distribution of palindromic sequences in Noumeavirus, Melbournevirus and Insectomime virus.

A, C and E: distributions of palindromic motifs observed at each position relative to the start codon (gray curve and Inset arrow). The signal was deconvoluted in two parts depending on whether the

gene upstream of each start codon was in the same orientation (red curve and Inset arrows) or in the reverse orientation (blue curve and Inset arrows).

B, D and F distributions of palindromic motifs observed at each position relative to the stop codon (gray curve and Inset arrow). The signal was deconvoluted in two parts depending on whether the gene downstream of each stop codon is in the same orientation (red curve and inset arrows) or in the reverse orientation (blue curve and Inset arrows). The enrichment seen at the start appears mostly due to palindromic signals associated with the upstream gene. Regardless of the downstream gene, the observed enrichment of palindromic signals is highly correlated to the presence of an upstream stop codon.

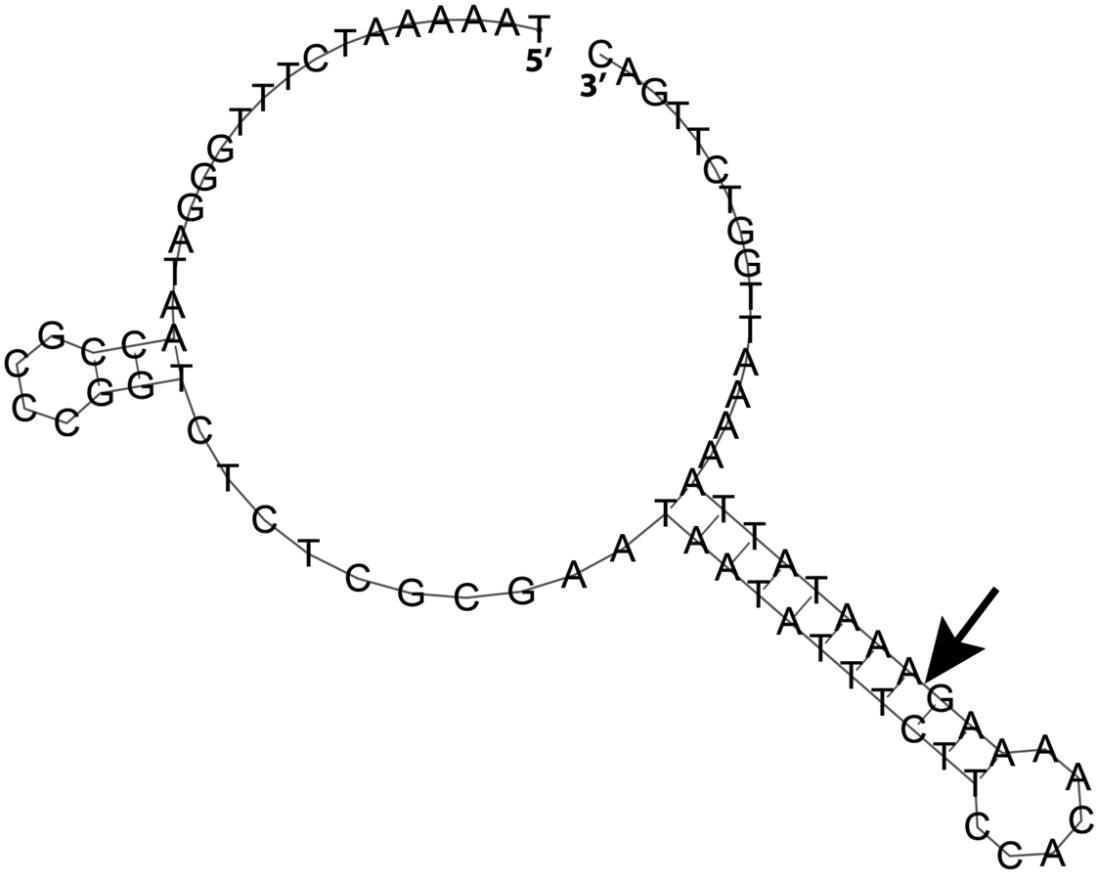


Figure S5: 3' UTR Hairpin sequence and structure as produced by RNAfold. The cleavage site of the NMV_1 is indicated by an arrow.

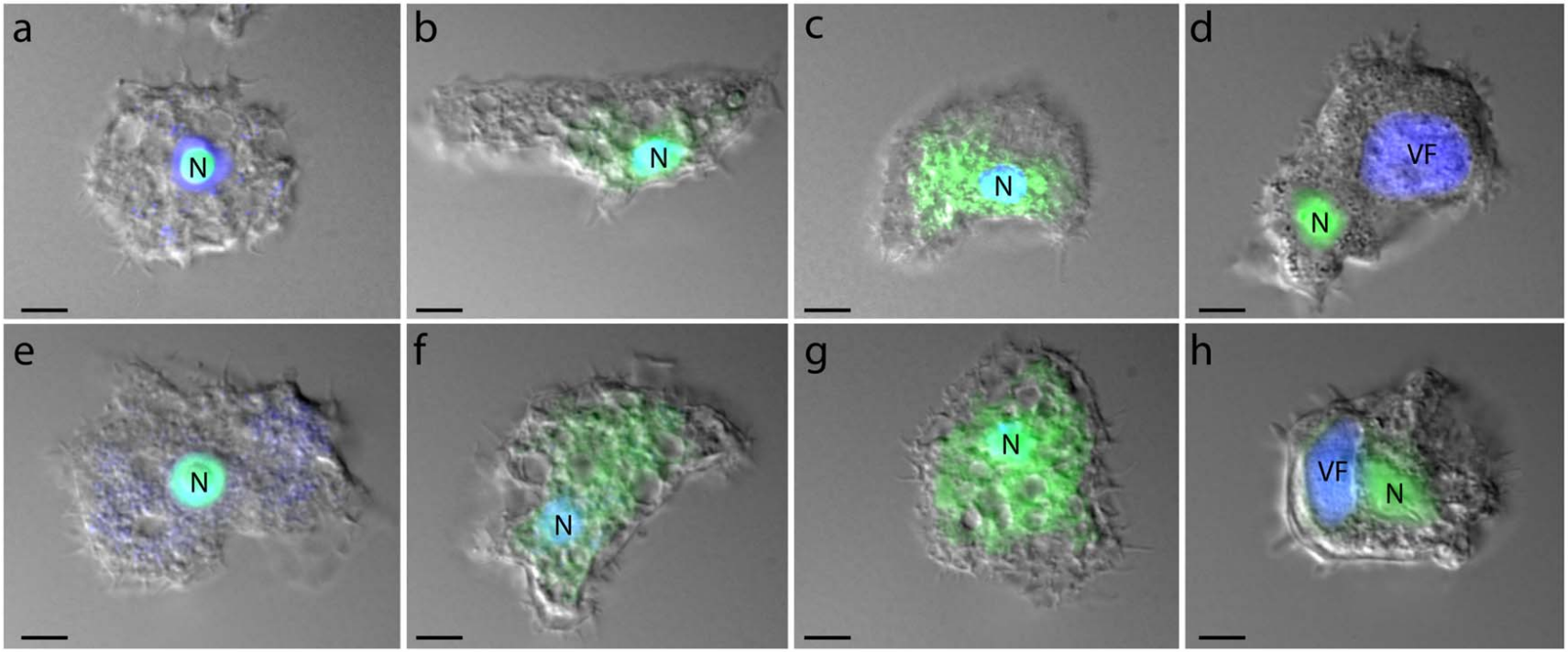


Figure S6: Fluorescence study of *A. castellanii* cell nucleus during Noumeavirus infection: Nucleus (N) in a] non-infected and infected cells expressing the nucleolar ML216-GFP at b] 30 min pi, c] 1h pi and d] 4 h pi. e] Non-infected cells and infected cells expressing the nuclear GFP-SUMO at f] 30 min pi, g] 1h pi and h] 4 h pi. DAPI staining remains at the nucleus all along the infection but the intense staining of the late VF makes it barely visible at the nucleus 4h pi (h). The GFP staining located in the nucleus of uninfected cells progressively spreads out of the nucleus (b, f) before filling the cytoplasm (c, g). After 4h, the GFP florescence is back into the nucleus and the DAPI staining is intense in the viral factories (d, h). The same observations were obtained on cells infected by Melbournevirus.

126 **Table S1 : Comparison of Noumeavirus and Melbournevirus virion proteomes**

Origin	Protein Ids NMV	Protein name	Equivalent MEL	Final rank in NMV	Final rank MEL
Noumeavirus	NMV_189	hypothetical protein	MEL_236	1	1
Noumeavirus	NMV_078	histone H3-like protein	MEL_368	2	3
Noumeavirus	NMV_116	major capsid protein	MEL_305	3	4
Noumeavirus	NMV_149	hypothetical protein	MEL_274	4	5
Noumeavirus	NMV_171	Papain-likecysteine protease	MEL_369	5	6
Noumeavirus	NMV_079	histone H2B/H2A fusion protein	MEL_247	6	7
Noumeavirus	NMV_185	hypothetical protein	MEL_241	7	10
Noumeavirus	NMV_333	hypothetical protein	MEL_089	8	39
Noumeavirus	NMV_141	hypothetical protein	MEL_280	9	8
Noumeavirus	NMV_192	hypothetical protein	MEL_234	10	9
Noumeavirus	NMV_225	thioredoxin	MEL_207	11	12
Noumeavirus	NMV_230	hypothetical protein	MEL_202	12	14
Noumeavirus	NMV_045	hypothetical protein	MEL_342b	13	2
Noumeavirus	NMV_238	hypothetical protein	MEL_196	14	24
Noumeavirus	NMV_066	hypothetical protein	MEL_358	15	11
Noumeavirus	NMV_349	hypothetical protein	MEL_065	16	17
Noumeavirus	NMV_163	hypothetical protein	MEL_260	17	15
Noumeavirus	NMV_198	membrane protein	MEL_223b	18	45
Noumeavirus	NMV_068	AAA-familyATPase	MEL_360	19	26
Noumeavirus	NMV_296	hypothetical protein	MEL_097	20	19
Noumeavirus	NMV_058	galactose binding lectin domain containing protein	MEL_352	21	35
Noumeavirus	NMV_060	Transmembrane domain containing protein	MEL_354	22	22
Noumeavirus	NMV_183	membrane protein	MEL_244	23	27
Noumeavirus	NMV_350	hypothetical protein	MEL_061	24	57
Noumeavirus	NMV_355	Transmembrane domain containing protein	N/A	25	N/A
Noumeavirus	NMV_244	ATP-dependenthelicase	MEL_181	26	21
Noumeavirus	NMV_128	thioredoxin	MEL_294	27	25

Noumeavirus	NMV_093	hypothetical protein	MEL_331	28	34
Noumeavirus	NMV_204	hypothetical protein	MEL_231	29	44
Noumeavirus	NMV_074	hypothetical protein	MEL_365b	30	23
Noumeavirus	NMV_248	hypothetical protein	MEL_185	31	28
Noumeavirus	NMV_439	hypothetical protein	MEL_020	32	13
Noumeavirus	NMV_232	Transmembrane domain containing protein	MEL_200	33	16
Noumeavirus	NMV_138	Transmembrane domain containing protein	MEL_282b	34	20
Noumeavirus	NMV_157	Transmembrane domain containing protein	MEL_269	35	30
Noumeavirus	NMV_148	hypothetical protein	MEL_275	36	29
Noumeavirus	NMV_419	histone H2A domain-containing protein	MEL_149	37	38
Noumeavirus	NMV_196	Transmembrane domain containing protein	MEL_228	38	31
Noumeavirus	NMV_218	hypothetical protein	MEL_213b	39	47
Noumeavirus	NMV_221	Transmembrane domain containing protein	MEL_211	40	36
Noumeavirus	NMV_081	hypothetical protein	MEL_371	41	61
Noumeavirus	NMV_441	hypothetical protein	N/A	42	N/A
Noumeavirus	NMV_004	peptidase	MEL_025	43	18
Noumeavirus	NMV_014	peptidase	MEL_041	44	33
A. castellanii	g1525-75307-76608	Euk Porin protein		45	65
Noumeavirus	NMV_214	helicase	MEL_216	46	40
Noumeavirus	NMV_275	lipase	MEL_168	47	32
Noumeavirus	NMV_367	hypothetical protein	MEL_385b	48	82
Noumeavirus	NMV_080	hypothetical protein	MEL_370	49	76
Noumeavirus	NMV_164	Papain-like cysteine protease	MEL_255	50	49
Noumeavirus	NMV_336	amine oxidase	MEL_085	51	74
Noumeavirus	NMV_032	hypothetical protein	MEL_052	52	55
Noumeavirus	NMV_124	ankyrin repeat-containing protein	MEL_292	53	63
Noumeavirus	NMV_166	hypothetical protein	MEL_256	54	67
Noumeavirus	NMV_120	Transmembrane domain containing protein	MEL_301	55	43
Noumeavirus	NMV_212	RibonucleaseIII	MEL_219	56	51
Noumeavirus	NMV_162	SWIB-domain-containing protein	MEL_261	57	98

Noumeavirus	NMV_144	disulfide oxidoreductase	MEL_278	58	52
Noumeavirus	NMV_038	hypothetical protein	MEL_055	59	54
Noumeavirus	NMV_047	hypothetical protein	MEL_341	60	56
Noumeavirus	NMV_006*	peptidase	MEL_040*	61	59
Noumeavirus	NMV_191	serine/threonine protein kinase	MEL_235	62	41
A. castellanii	gunplaced-2126338-2127783	Actin 1		63	78
Noumeavirus	NMV_173	metallopeptidaseWLM	MEL_249	64	68
Noumeavirus	NMV_424	glycosyltransferase	MEL_146	65	84
Noumeavirus	NMV_365	hypothetical protein	MEL_387	66	50
Noumeavirus	NMV_346	hypothetical protein	MEL_066	67	64
Noumeavirus	NMV_132	hypothetical protein	MEL_297	68	46
Noumeavirus	NMV_246	Transmembrane domain containing protein	MEL_183	69	53
Noumeavirus	NMV_005	peptidase	MEL_024	70	87
Noumeavirus	NMV_181	Transmembrane domain containing protein	MEL_245	71	75
Noumeavirus	NMV_216	hypothetical protein	MEL_215	72	48
Noumeavirus	NMV_206	hypothetical protein	MEL_231c	73	95
Noumeavirus	NMV_177	hypothetical protein	MEL_252	74	88
Noumeavirus	NMV_140	thioredoxin	MEL_281	75	91
Noumeavirus	NMV_119	membrane protein	MEL_300	76	90
Noumeavirus	NMV_136	ankyrin repeat-containing protein	MEL_284	77	70
Noumeavirus	NMV_207	hypothetical protein	MEL_232	78	73
Noumeavirus	NMV_170	Transmembrane domain containing protein	MEL_246	79	77
Noumeavirus	NMV_263	Transmembrane domain containing protein	MEL_173	80	81
Noumeavirus	NMV_210	ubiquitin	MEL_221c	81	112
Noumeavirus	NMV_259	hypothetical protein	MEL_177	82	66
Noumeavirus	NMV_176	membraneprotein	MEL_251b	83	69
Noumeavirus	NMV_147	Transmembrane domain containing protein	MEL_276	84	60
Noumeavirus	NMV_361	glycosyltransferase domain containing protein	MEL_398	85	58
Noumeavirus	NMV_389	dual specificity phosphatase	MEL_109	86	93
Noumeavirus	NMV_092	serine/threonine protein kinase	MEL_380	87	37

Noumeavirus	NMV_037	peptidase	MEL_056	88	62
Noumeavirus	NMV_297	hypothetical protein	MEL_098	89	80
Noumeavirus	NMV_359	hypothetical protein	MEL_397	90	96
Noumeavirus	NMV_139	mRNA capping enzyme	MEL_282	91	79
Noumeavirus	NMV_130	glycosyltransferase	MEL_290	92	86
Noumeavirus	NMV_156	Transmembrane domain containing protein	MEL_268	93	92
Noumeavirus	NMV_226	hypothetical protein	MEL_206	94	72
Noumeavirus	NMV_217	hypothetical protein	MEL_214	95	71
Noumeavirus	NMV_154	Transmembrane domain containing protein	MEL_266	96	102
Noumeavirus	NMV_234	class3 lipase	MEL_199	97	94
Noumeavirus	NMV_083	Transmembrane domain containing protein	MEL_373	98	85
Noumeavirus	NMV_347	hypothetical protein	MEL_63b	99	N/A
Noumeavirus	NMV_042	hypothetical protein	MEL_338	100	42
Noumeavirus	NMV_035	hypothetical protein	MEL_133	101	113
A. castellanii	g817-77076-78510	C8 sterol isomerase		102	107
A. castellanii	g1359-61676-62848	hypothetical protein		103	125
Noumeavirus	NMV_126	inactivated thioredoxin/glutaredoxin	MEL_296	104	89
A. castellanii	g254-27747-28588	hypothetical protein		105	103
Noumeavirus	NMV_182	hypothetical protein	MEL_243	106	N/A
Noumeavirus	NMV_385	hypothetical protein	MEL_096	107	100
Noumeavirus	NMV_180	Transmembrane domain containing protein	MEL_259	108	105
A. castellanii	g726-224146-225044	Wnt-10b		109	117
Noumeavirus	NMV_241	hypothetical protein	MEL_193	110	163
A. castellanii	g611-72236-73493	exported protein putative		111	115
Noumeavirus	NMV_007	peptidase	N/A	112	N/A
A. castellanii	g1185-1867882-1870710	aspartic proteinase		113	147
A. castellanii	g376-450681-452588	peroxidase		114	106
A. castellanii	g254-319025-319938	actophorin		115	N/A
A. castellanii	g626-142077-143389	adenine nucleotide translocator		116	N/A
A. castellanii	g641-748251-749192	C-type lysozyme		117	N/A

A. castellanii	g696-284312-285231	LIM domain containing protein		118	145
Noumeavirus	NMV_249	hypothetical protein	MEL_186	119	146
A. castellanii	g1007-388106-392098	hypothetical protein		120	159
Noumeavirus	NMV_348	hypothetical protein	MEL_064	121	99
Noumeavirus	NMV_115	hypothetical protein	MEL_306	122	138
A. castellanii	g757-590645-592959	C2 domain containing protein		123	127
A. castellanii	g422-496887-497476	fascin subfamily protein		124	119
Noumeavirus	NMV_082	hypothetical protein	MEL_372	125	116
A. castellanii	g935-165792-170359	Peroxidase		126	139
A. castellanii	g414-411909-412468	C2 domain containing protein		127	109
Noumeavirus	NMV_269	EBmodulecontainingprotein	MEL_172	128	143
A. castellanii	g42-42086-43890	Rac-like protein		129	173
A. castellanii	g527-541096-542131	Cyclophylin		130	N/A
Noumeavirus	NMV_388	hypotheticalprotein	MEL_108	131	N/A
A. castellanii	g446-42975-43598	hypothetical protein		132	114
A. castellanii	g726-530322-533705	hypothetical protein		133	N/A
A. castellanii	g763-171894-173621	Rab 1B, Ras oncogene family		134	N/A
A. castellanii	g1369-729197-730236	Ras-like protein 1		135	133
A. castellanii	g253-184876-188554	Hydrolase activity		136	181
Noumeavirus	NMV_085	hypothetical protein	MEL_374	137	97
Noumeavirus	NMV_070	hypothetical protein	MEL_361b	138	N/A
A. castellanii	g640-633777-634798	Rho family small GTPase Rac3		139	N/A
Noumeavirus	NMV_363	Transmembrane domain containing protein	MEL_095b	140	N/A
Noumeavirus	NMV_236	hypothetical protein	MEL_197	141	144
A. castellanii	g631-243047-244467	Rab GTPase		142	168
A. castellanii	g1317-154778-156138	Rho family small GTPase Rac3		143	131
A. castellanii	g185-637578-643589	hypothetical protein		144	167
A. castellanii	g631-28616-31737	PA domain containing protein		145	N/A
A. castellanii	g696-243559-244965	Ras family protein		146	N/A
A. castellanii	g218-68723-70868	Corornin3 putative		147	179

A. castellanii	g631-805994-806720	Chorismate mutase subfamily protein		148	128
A. castellanii	g490-123167-124218	hypothetical protein		149	N/A
A. castellanii	g1007-257770-259328	hypotheticalprotein		150	180
A. castellanii	g414-40155-41168	hypothetical protein		151	182
Noumeavirus	NMV_161	hypothetical protein	MEL_262	152	130
Noumeavirus	NMV_127	hypothetical protein	MEL_295	153	N/A
A. castellanii	g397-418028-419548	Ras GTPase Rap1b		154	164
A. castellanii	g433-73037-74438	Inosineuridine preferring nucleoside hydrolase family protein		155	N/A
A. castellanii	g39-686-2452	EGF-like domain containing protein		156	N/A
A. castellanii	g254-323982-326008	Cysteine proteinase precursor		157	N/A
A. castellanii	g1406-111173-112358	hypothetical protein		158	175
A. castellanii	g1185-1161386-1166670	Amidohydrolase domain containing protein		159	N/A
A. castellanii	g454-198036-200147	EF-hand containing protein		160	120
A. castellanii	g935-470651-471228	Cystatin putative		161	N/A
A. castellanii	g548-81823-84908	I/LWEQ domain containing protein		162	N/A
A. castellanii	g817-118331-118888	hypothetical protein		163	N/A
A. castellanii	g609-257397-258401	hypothetical protein		164	N/A
A. castellanii	g763-133835-134874	Glycosyl hydrolase family 25 subfamily protein		165	N/A
A. castellanii	g376-325048-327718	PhoPQactivated pathogenicity-related protein-like protein; putative		166	N/A
A. castellanii	gunplaced-1973089-1973871	Copper/zinc superoxide dismutase		167	83
A. castellanii	g593-129866-130530	Sequence specific DNA binding		168	N/A
A. castellanii	g1406-789283-793494	Glutamate carboxypeptidase		169	176
A. castellanii	g1449-175373-177058	Ras subfamily protein		170	N/A
A. castellanii	g726-297776-299069	Cysteine proteinase 3, partial		171	N/A
A. castellanii	g360-448580-450093	lmbE family protein		172	N/A
A. castellanii	g680-132237-134308	Protein kinase C4, partial		173	189
A. castellanii	g726-22077-25022	Alkaline phosphatase family subfamily protein		174	N/A
A. castellanii	g593-240006-240660	Peptidylprolyl isomerase FKB12, putative		175	N/A
Noumeavirus	NMV_387	serine/threonine protein kinase	MEL_107	176	186

Noumeavirus	NMV_223	hypothetical protein	MEL_209	177	N/A
A. castellanii	g631-541984-543885	Peptidase S8 and S53 subtilisin kexin sedolisin, putative		178	N/A
A. castellanii	g531-97521-99028	Actin related protein 2/3 complex, 41kDa subunit		179	178
A. castellanii	g527-461746-463738	Ras GTPase		180	155
A. castellanii	gunplaced-1594092-1595819	Factin-capping protein subunit alpha-2; putative		181	N/A
A. castellanii	g307-171526-174660	Glycosyl hydrolase family 20; catalytic domain containing protein		182	N/A
Noumeavirus	NMV_113	zinc ribbon-containing protein	MEL_309	183	140
A. castellanii	gunplaced-519067-520320	hypothetical protein		184	N/A
A. castellanii	g1295-400157-403348	Fasciclin domain containing protein		185	N/A
Noumeavirus	NMV_335	hypothetical protein	MEL_087	186	N/A
A. castellanii	g376-448106-449758	Peroxidase		187	149
A. castellanii	gunplaced-452590-455256	Ef1a		188	142
A. castellanii	g1295-157226-158056	hypothetical protein		189	N/A
A. castellanii	g284-678893-681196	PH domain containing protein		190	161
A. castellanii	g1317-24676-26730	Prokumamolisin, putative		191	N/A
A. castellanii	cox1/2	cox1/2		192	192
A. castellanii	g535-133406-135131	ATP:L-Methionine S-Adenosyltransferase		193	N/A
A. castellanii	g1185-1806178-1808434	Protein phosphatase 1 catalytic subunit alpha, putative		194	134
A. castellanii	g763-137286-138388	Glycosyl hydrolase family 25 subfamily protein		195	N/A
A. castellanii	g414-25246-26688	Sar1 family small GTPase		196	N/A
A. castellanii	g406-2586-7412	Lysosomal alphamannosidase		197	N/A
A. castellanii	g297-276196-278530	Prokumamolisin, activation domain containing protein		198	N/A
A. castellanii	g536-73396-74613	hypothetical protein		199	N/A
A. castellanii	g1196-29473-31845	amineoxidase, Flavine containing superfamily protein		200	N/A
A. castellanii	g185-111308-112608	autophagy-related protein 27		201	135
A. castellanii	g1185-1240877-1242718	PH domain containing protein		202	N/A
Noumeavirus	NMV_242	hypothetical protein	MEL_192	203	185
A. castellanii	g1390-216047-218303	High molecular weight heat shock protein		204	191

A. castellanii	g1306-148886-150120	Vacuolar proton ATPase, putative		205	N/A
A. castellanii	g739-222522-223876	Fibrillarin, putative		206	N/A
A. castellanii	g734-49529-51643	Peptidase M16 family protein		207	N/A
A. castellanii	g1449-118768-120377	Actin related protein 2/3 complex		208	N/A
A. castellanii	g1369-497445-500306	Gelation factor, putative		209	169
A. castellanii	gunplaced-2251673-2253071	ARP2/3 complex 34 kDa subunit, putative		210	N/A
Noumeavirus	NMV_400	zinc finger protein	MEL_347	211	152
A. castellanii	g254-213866-215830	Peptidase M16 inactive domain containing protein		212	188
A. castellanii	g1185-98157-99558	Histone cluster 1, H2bb, putative		213	158
A. castellanii	gunplaced-1669823-1671094	Mn/Fe superoxide dismutase		214	166
A. castellanii	g1185-1855544-1856934	Mitochondrial phosphate transporter, putative		215	N/A
A. castellanii	g631-162107-163748	Tubulin beta chain		216	141
A. castellanii	g527-210936-212510	Ras-related protein Rab-21		217	N/A
A. castellanii	g171-459415-462379	amino acid transmembrane transport		218	N/A
Noumeavirus	NMV_229	hypothetical protein	MEL_203	219	N/A
A. castellanii	g360-68713-70325	RNA recognition motif domain containing protein		220	N/A
A. castellanii	g433-288-2349	Carbohydrate metabolic process		221	N/A
A. castellanii	g185-569957-572411	phospholipid binding		222	N/A
A. castellanii	g544-146276-147708	Factin-capping protein subunit beta; putative		223	N/A
A. castellanii	g509-307446-310070	Serine carboxypeptidase (CPB1), putative		224	N/A
A. castellanii	g677-37001-39098	Actin related protein 3, putative		225	N/A
A. castellanii	g531-32516-35280	V ATPase subunit A, putative		226	N/A
A. castellanii	g171-371925-373194	RSNARE, VAMP71-family		227	N/A
A. castellanii	g935-641463-642298	Cytochrome C Oxydase		228	N/A
A. castellanii	g1185-112487-113838	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic-like		229	172
A. castellanii	g1390-714707-716810	Alpha tubulin, partial		230	132
A. castellanii	g993-461397-463268	Protein P80, putative		231	N/A
A. castellanii	g1007-410536-413510	Predicted protein		232	170
A. castellanii	g452-22276-25291	ATP synthase, putative		233	N/A

A. castellanii	g218-115547-116688	Peroxiredoxin 2, putative		234	N/A
A. castellanii	g993-399293-400878	RAB11B protein, putative		235	N/A
Noumeavirus	NMV_075	hypothetical protein	MEL_366	236	N/A
Noumeavirus	NMV_312	ATPase	MEL_141	237	150
A. castellanii	g757-61267-67282	Amidohydrolase domain containing protein		238	N/A
A. castellanii	g481-83586-84700	Actin related protein ARPC3, putative		239	N/A
A. castellanii	g757-460169-461398	START domain containing protein		240	N/A
A. castellanii	g327-32846-34973	Malate dehydrogenase		241	N/A
A. castellanii	g490-271761-274518	Eukaryotic translation elongation factor 2, putative		242	N/A
A. castellanii	g397-224837-226378	Rab7 GTPase		243	N/A
A. castellanii	g1067-140826-143107	Glycoside hydrolase family protein		244	N/A
A. castellanii	g1185-686457-688604	Cell cycle control protein		245	N/A
A. castellanii	g1271-599999-602983	Acyloxyacyl hydrolase		246	N/A
A. castellanii	g527-633449-637592	Cobalaminin dependent methionine synthase		247	195
A. castellanii	g1317-556358-557497	protein domain specific binding		248	177
A. castellanii	g171-452759-454367	Rab4, putative		249	N/A
A. castellanii	g935-656616-658717	Plastin 3 (T isoform), putative		250	N/A
A. castellanii	g1377-88396-89795	Synaptobrevin protein		251	N/A
A. castellanii	g527-274667-278548	Gamaglutamyl transferase		252	N/A
A. castellanii	g448-516809-518949	Coenzyme binding		253	N/A
A. castellanii	g376-14236-15730	hypothetical protein		254	N/A
A. castellanii	g1359-44666-49600	Oxidation-reduction process		255	N/A
A. castellanii	g1369-642385-649286	P53 inducible protein		256	N/A
A. castellanii	g1007-235917-237588	Soluble NSF attachment protein alpha isoform		257	N/A
A. castellanii	g557-252166-256085	Glycosyl hydrolase family 65 central catalytic domain containing protein		258	N/A
A. castellanii	g422-858104-863968	Amidohydrolase domain containing protein		259	N/A
A. castellanii	g1317-272306-275600	Sec1 family protein, partial		260	190
A. castellanii	g527-226157-232618	Myosin-I binding protein Acan125		261	N/A
A. castellanii	atp1	ATP11 protein		262	N/A

A. castellanii	g535-204667-205958	Ras family GTPase		263	N/A
A. castellanii	g40-21764-24998	BnaC01g40190D		264	N/A
A. castellanii	gunplaced-469113-471317	Ribosomal protein S32, putative		265	N/A
A. castellanii	g40-189273-197081	RecName: Full		266	N/A
A. castellanii	g311-83679-90488	Phospholipid translocating P-type ATPase family protein		267	N/A
A. castellanii	g1185-932177-934308	Vacuolar ATPase subunit DVA41		268	N/A
A. castellanii	g185-395816-398240	hypothetical protein		269	N/A
A. castellanii	g904-22075-24548	H(+)-transporting ATPase family protein		270	N/A
A. castellanii	g641-620276-623240	Zinc finger, C3HC4 type (RING finger) domain containing protein		271	N/A
A. castellanii	g185-371587-375028	Agarase, putative		272	N/A
A. castellanii	gunplaced-225545-227583	hypothetical protein		273	N/A
A. castellanii	g626-155026-157500	Chaperonin GroL, putative		274	N/A
Noumeavirus	NMV_356	ATP-dependent DNA ligase	MEL_060	275	187
A. castellanii	g734-24891-27760	Tripeptidylpeptidase 1, putative		276	N/A
A. castellanii	g631-1049155-1053916	Protein binding		277	N/A
Noumeavirus	NMV_129	hypothetical protein	MEL_293	278	N/A
A. castellanii	g1297-265817-271878	Myosin-IA		279	N/A
A. castellanii	g430-68666-70450	Ras subfamily protein, putative		280	N/A
A. castellanii	g1229-124396-134855	Binding		281	N/A
A. castellanii	g218-42810-52640	GTPase-activator protein for Ras family GTPase		282	N/A
A. castellanii	g1185-1772532-1781891	Multidrug resistance associated protein, putative		283	N/A
A. castellanii	g1185-770357-772028	Multidrug resistance associated protein, putative		N/A	101
A. castellanii	g557-519644-520579	Vtype ATPase; C subunit		N/A	104
A. castellanii	g935-64777-65458	Ribosomal L40e family		N/A	108
Noumeavirus	NMV_401	hypothetical protein	MEL_348	N/A	110
	N/A	Transmembranedomaincontainingprotein	MEL_332	N/A	111
Noumeavirus	NMV_061	uracil-DNAglycosylase	MEL_355	N/A	118
Noumeavirus	NMV_301	hypothetical protein	MEL_102	N/A	121
A. castellanii	g97-87038-122304	Probable aggregation factor core protein MAFp3		N/A	122

A. castellanii	g97-85422-86868	hypothetical protein		N/A	123
	N/A	hypothetical protein	MEL_129b	N/A	124
Noumeavirus	NMV_231	Hypothetical protein	MEL_201	N/A	126
Noumeavirus	NMV_123	Zinc finger protein	MEL_286	N/A	129
Noumeavirus	NMV_122	hypothetical protein	MEL_287	N/A	136
	N/A	endonuclease	MEL_221b	N/A	137
	N/A	hypothetical protein	MEL_323b	N/A	148
A. castellanii	g284-245837-247148	hypothetical protein		N/A	151
Noumeavirus	NMV_188	hypothetical protein	MEL_238	N/A	153
Noumeavirus	NMV_360	hypothetical protein	MEL_049	N/A	154
A. castellanii	g40-197330-199538	Tubulin beta1 chain		N/A	156
A. castellanii	g422-400365-406076	Ubiquitin carboxylterminal hydrolase domain containing protein		N/A	157
Noumeavirus	NMV_152	eukaryotic translation initiation factor 5	MEL_273	N/A	160
Noumeavirus	NMV_260	hypothetical protein	MEL_176	N/A	162
	N/A	adenine-specific methyltransferase	MEL_016	N/A	165
A. castellanii	g1369-335797-338290	Calponin domain containing protein		N/A	171
Noumeavirus	NMV_418	hypothetical protein	MEL_171	N/A	174
Noumeavirus	NMV_067	helicase	MEL_359	N/A	183
Noumeavirus	NMV_253	proliferating cell nuclear antigen	MEL_190	N/A	184
A. castellanii	g1271-609961-613385	Pigment precursor permease subfamily protein		N/A	193
A. castellanii	gunplaced-1642808-1647994	Protein kinase domain containing protein		N/A	194
A. castellanii	g397-60057-61818	Hypothetical protein		N/A	196

* NMV_006 and MEL_040 are homologous non orthologous

131 **Table S2: Host proteins shared by the Noumeavirus and Melbournevirus virions and their ranks in**
132 **virions and non-infected *A. castellanii* cells (NI)**

Gene	Protein name	Real rank in Noumea virus	Real rank in Melbourne virus	Host relative rank in Noumea virus	Host relative rank in Melbourne virus	Rank in host (NI)	Other viruses
gunplaced-2126338-2127783	Actin 1	63	78	2	2	1	+
g422-496887-497476	fascin subfamily protein	124	119	13	11	2	+
g1525-75307-76608	Euk Porin protein	45	65	1	1	3	+
g1185-112487-113838	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic-like	229	172	47	37	4	0
g414-411909-412468	C2 domain containing protein	127	109	15	7	5	+
g696-284312-285231	LIM domain containing protein	118	145	10	24	9	+
g527-633449-637592	Cobalaminin dependent methionine synthase	247	195	50	52	10	0
g1390-216047-218303	High molecular weight heat shock protein	204	191	41	50	12	+
g1317-556358-557497	Hypothetical, protein binding	248	177	51	41	18	0
g218-68723-70868	Corornin3 putative	147	179	23	43	42	+
gunplaced-1973089-1973871	Copper/zinc superoxide dismutase	167	83	30	3	54	+
g254-213866-215830	Peptidase M16 inactive domain containing protein	212	188	43	47	72	+
g726-224146-225044	Wnt-10b	109	117	6	10	92	+
g1359-61676-62848	hypothetical protein	103	125	4	13	129	+
g757-590645-592959	C2 domain containing protein	123	127	12	14	138	+
g1390-714707-716810	Alpha tubulin, partial	230	132	48	17	140	0
g531-97521-99028	Actin related protein 2/3 complex, 41kDa subunit	179	178	33	42	152	+
g631-162107-163748	Tubulin beta chain	216	141	46	22	164	0
g376-450681-452588	peroxidase	114	106	9	5	174	+
gunplaced-1669823-1671094	Mn/Fe superoxide dismutase	214	166	45	32	188	0
cox1/2	cox1/2	192	192	38	51	220	+
g1406-111173-112358	hypothetical protein	158	175	28	39	224	0
g42-42086-43890	Rac-like protein	129	173	16	38	269	0
g446-42975-43598	hypothetical protein	132	114	17	8	286	+
g1369-497445-500306	Gelation factor, putative	209	169	42	35	290	0
g284-678893-681196	PH domain containing protein	190	161	37	30	308	+
g376-448106-449758	Peroxidase	187	149	35	26	309	+
g1369-729197-730236	Ras-like protein 1	135	133	18	18	331	0
g817-77076-78510	C8 sterol isomerase	102	107	3	6	333	+
g611-72236-73493	exported protein putative	111	115	7	9	338	0
g631-243047-244467	Rab GTPase	142	168	20	34	356	0
g185-111308-112608	autophagy-related protein 27	201	135	40	20	370	+
g397-418028-419548	Ras GTPase Rap1b	154	164	27	31	374	0
g1317-154778-156138	Rho family small GTPase Rac3	143	131	21	16	401	0
g1185-1867882-1870710	aspartic proteinase	113	147	8	25	475	+
g414-40155-41168	hypothetical protein	151	182	26	46	484	0
g527-461746-463738	Ras GTPase	180	155	34	27	487	+
g680-132237-134308	Protein kinase C4, partial	173	189	32	48	521	0
g454-198036-200147	EF-hand containing protein	160	120	29	12	524	+
g1007-410536-413510	Predicted protein	232	170	49	36	570	+
g935-165792-170359	Peroxidase	126	139	14	21	789	+
g1185-1806178-1808434	Protein phosphatase 1 catalytic subunit alpha, putative	194	134	39	19	837	+
g1007-257770-259328	hypotheticalprotein	150	180	25	44	865	+
g254-27747-28588	hypothetical protein	105	103	5	4	1080	+
g1317-272306-275600	Sec1 family protein, partial	260	190	52	49	1143	+
g1007-388106-392098	hypothetical protein	120	159	11	29	1391	0
g253-184876-188554	Hydrolase activity	136	181	19	45	1555	+

g1185-98157-99558	Histone cluster 1, H2bb, putative	213	158	44	28	1621	0
g1406-789283-793494	Glutamate carboxypeptidase	169	176	31	40	1783	0
g631-805994-806720	g1406-111173-112358	148	128	24	15	2601	+
g185-637578-643589	hypothetical protein	144	167	22	33	3020	+
gunplaced-452590-455256	Ef1a	188	142	36	23	NA	0

133

134

Table S3: Identification and ranking of proteins involved in transcription and transcript maturation in the virion proteomes of cytoplasmic viruses *versus* Noumeavirus and Melbournevirus (NA: the corresponding gene does not exist, noID: no identified peptides)

Gene name	Mimivirus				Pithovirus sibericum				Noumeavirus				Melbournevirus			
	Gene #	Theore-tical	Identified Peptides	Rank	Gene #	Theore-tical	Identified Peptides	Rank	Gene #	Theore-tical	Identified Peptides	Rank	Gene #	Theore-tical	Identified Peptides	Rank
Rpb1	R501	77	50	65	pv_366 -368	51	35	68	NMV_011	70	noID		MEL_047	72	noID	
Rpb2	L244	60	30	80	pv_242	61	40	78	NMV_436	55	noID		MEL_022	63	noID	
Rpb5	L235	12	7	54	pv_31	11	8	72	NMV_046	14	noID		MEL_342	9	noID	
polyA pol	R341	30	14	87	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
RNAse III	R343	28	noID		pv_75	20	15	63	NMV_212	18	10	56	MEL_219	16	8	51
mRNA capping enzyme	R382	72	41	83	pv_454	52	38	52	NMV_139	45	28	93	MEL_282	46	28	80

The table present the number of theoretically identifiable (http://web.expasy.org/peptide_mass) and experimentally identified peptides (no miscleavage allowed, minimum of 7 amino acids, m/z between 750 and 4000 Da) as well as the rank observed for the corresponding genes for Mimivirus, *Pithovirus sibericum*, Noumeavirus and Melbournevirus.

Table S4: Protein Comparison between Noumeavirus and Lausannevirus and between Noumeavirus and Melbournevirus

	Noumeavirus/Lausannevirus	Noumeavirus/Melbournevirus
Best Blast Reciprocal Hits	374	331
Minimal %ID	15.10	9.70
Median %ID	83.15	58.10
Average %ID	77.37	55.79
Maximal %ID	100.00	96.20

Supplementary References

1. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6,** 31 (2005).
2. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).
3. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27,** 1164–1165 (2011).
4. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinforma. Oxf. Engl.* **30,** 1312–1313 (2014).
5. Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* **13,** 2178–2189 (2003).

Résumé :

La découverte des virus géants il y a une dizaine d'années a véritablement bouleversé notre perception du monde viral. Historiquement, les virus sont des parasites qui sont plus petits que les bactéries, qui sont capables de passer à travers le filtre de Chamberland et qui ne peuvent pas être observés au microscope optique. Mimivirus, le premier virus géant, dont l'hôte est une amibe, a été initialement découvert en 1992, mais a été confondu avec une bactérie : ce n'est que dix ans plus tard que sa vraie nature a été révélée. Mimivirus possède une particule d'environ 750 nm de diamètre et un génome à ADN double-brin de 1.18 Mb : il code pour sa propre machinerie de transcription et de réparation de l'ADN, et même pour des protéines de la traduction. Mimivirus possède des caractéristiques exceptionnelles pour un virus, et a remis en question la définition même du mot "virus". Cette découverte, ainsi que celle d'autres *Mimiviridae*, a ouvert un débat sur l'origine et l'histoire évolutive de ces virus, et ravivé celui portant sur la nature des virus : peuvent-ils être considérés comme vivants ?

Cette thèse, dont le but était d'isoler de nouveaux virus géants à partir d'échantillons environnementaux, afin d'enrichir notre compréhension de leur mode d'évolution et de leur origine, se divise en deux parties. La première partie de ce manuscrit est consacrée à la caractérisation d'un nouveau *Marseilleviridae*, qui a été isolé à partir d'un échantillon prélevé en Nouvelle-Calédonie, et qui a été appelé Noumeavirus. Les *Marseilleviridae*, dont le membre fondateur a été découvert en 2009, sont des grands virus à ADN, qui possèdent des particules à symétrie icosaédrique d'environ 200 nm de diamètre, et des génomes à ADN double-brin de plus de 300 kb. Noumeavirus a été comparé avec Melbournevirus, le membre le plus éloigné des *Marseilleviridae*, qui a été isolé au laboratoire en 2012. Différentes approches de génomique, protéomique, ainsi que l'étude de leurs cycles infectieux (microscopie à fluorescence et électronique) nous ont permis de révéler que le cycle infectieux de ces virus n'était pas indépendant du noyau cellulaire comme nous le pensions mais recrutait transitoirement les fonctions nucléaires à l'usine virale.

La deuxième partie est consacrée à la caractérisation d'un virus qui appartient à la famille des *Pandoraviridae* : le premier Pandoravirus, qui a été découvert en 2012, possède une morphologie unique au sein des virus, ainsi qu'un génome colossal à ADN double-brin de 2.5 Mb. Ce nouveau *Pandoraviridae* a également été isolé à partir d'un échantillon qui a été prélevé en Nouvelle-Calédonie, et a été appelé *Pandoravirus neocaledonia*. Des études comparatives avec d'autres virus appartenant à la même famille ont été réalisées en combinant plusieurs approches (génomique, protéomique, transcriptomique et microscopie électronique à transmission), afin de mieux comprendre les caractéristiques de ces virus inédits. Enfin, la morphologie étonnante de ces virus nous a poussés à étudier la composition de leur enveloppe, qui est constituée d'un réseau de fibres formant des structures lamellaires, afin de déterminer leur nature. Se pourrait-il que les Pandoravirus, contrairement aux autres virus, détournent la machinerie cellulaire pour construire leurs particules ? Dans ce cas, quelle serait leur place dans l'histoire évolutive des virus ?

Mots-clés : Virus géants, *Pandoraviridae*, *Marseilleviridae*, Evolution, Microscopie Electronique à Transmission.

Abstract:

The discovery of giant viruses about a decade ago has truly shaken our perception of the viral world. Historically, viruses are defined as parasites smaller than bacteria, which are able to pass through a Chamberland filter, and cannot be observed under a light microscope. Mimivirus, the first giant virus, which infects amoebas, was originally discovered in 1992 but was mistaken for a bacterium: it is only ten years later that its true nature was revealed. Mimivirus has a particle of about 750 nm in diameter and a double-stranded DNA genome of about 1.18 Mb: it encodes its own transcription and DNA repair machineries, and even proteins involved in translation. Mimivirus has outstanding characteristics for a virus, and questioned the very definition of the word "virus". This discovery and that of other Mimiviridae have initiated a debate on the origin and evolutionary history of these viruses, and it revived the debate on their nature: are viruses alive?

This thesis, which purpose was to isolate novel giant viruses from environmental samples in order to enhance our understanding of their mode of evolution and origin, is divided into two parts. The first part of this manuscript focuses on the characterization of a new *Marseilleviridae*, which was isolated from a sample collected in New-Caledonia, and which was named Noumeavirus. The *Marseilleviridae*, which founding member was discovered in 2009, are large DNA viruses that have icosahedral particles of about 200 nm in diameter, and double-stranded DNA genomes of more than 300 kb. Noumeavirus was compared with Melbournevirus, the most distant member of *Marseilleviridae*, which was isolated in the laboratory in 2012. Various approaches, such as genomics, proteomics and the study of their infectious cycles (with fluorescence and transmission electron microscopy), allowed us to reveal that the infectious cycle of these viruses was not independent from the cell nucleus as we thought, but was transiently recruiting nuclear functions to the viral factory.

The second part focuses on the characterization of a virus that belongs to the *Pandoraviridae* family: the first Pandoravirus, which was discovered in 2012, has a unique morphology and a gigantic double-stranded DNA genome of about 2.5 Mb. This new *Pandoraviridae* was also isolated from a sample that was collected in New-Caledonia and was named *Pandoravirus neocaledonia*. Comparative studies with other viruses belonging to the same family were performed using several approaches (genomics, proteomics, transcriptomics and transmission electron microscopy) to better understand the characteristics of these original viruses. Finally, the astonishing morphology of these viruses led us to investigate the composition of their envelope, which is made of a mesh of fibers forming lamellar structures. Is it possible that Pandoraviruses, unlike other viruses, hijack the cellular machinery to build their particles? In this case, what would be their place in the evolutionary history of viruses?

Key-words: Giant viruses, *Pandoraviridae*, *Marseilleviridae*, Evolution, Transmission Electron Microscopy.

Résumé :

La découverte des virus géants il y a une dizaine d'années a véritablement bouleversé notre perception du monde viral. Historiquement, les virus sont des parasites qui sont plus petits que les bactéries, qui sont capables de passer à travers le filtre de Chamberland et qui ne peuvent pas être observés au microscope optique. Mimivirus, le premier virus géant, dont l'hôte est une amibe, a été initialement découvert en 1992, mais a été confondu avec une bactérie : ce n'est que dix ans plus tard que sa vraie nature a été révélée. Mimivirus possède une particule d'environ 750 nm de diamètre et un génome à ADN double-brin de 1.18 Mb : il code pour sa propre machinerie de transcription et de réparation de l'ADN, et même pour des protéines de la traduction. Mimivirus possède des caractéristiques exceptionnelles pour un virus, et a remis en question la définition même du mot "virus". Cette découverte, ainsi que celle d'autres *Mimiviridae*, a ouvert un débat sur l'origine et l'histoire évolutive de ces virus, et ravivé celui portant sur la nature des virus : peuvent-ils être considérés comme vivants ?

Cette thèse, dont le but était d'isoler de nouveaux virus géants à partir d'échantillons environnementaux, afin d'enrichir notre compréhension de leur mode d'évolution et de leur origine, se divise en deux parties. La première partie de ce manuscrit est consacrée à la caractérisation d'un nouveau *Marseilleviridae*, qui a été isolé à partir d'un échantillon prélevé en Nouvelle-Calédonie, et qui a été appelé Noumeavirus. Les *Marseilleviridae*, dont le membre fondateur a été découvert en 2009, sont des grands virus à ADN, qui possèdent des particules à symétrie icosaédrique d'environ 200 nm de diamètre, et des génomes à ADN double-brin de plus de 300 kb. Noumeavirus a été comparé avec Melbournevirus, le membre le plus éloigné des *Marseilleviridae*, qui a été isolé au laboratoire en 2012. Différentes approches de génomique, protéomique, ainsi que l'étude de leurs cycles infectieux (microscopie à fluorescence et électronique) nous ont permis de révéler que le cycle infectieux de ces virus n'était pas indépendant du noyau cellulaire comme nous le pensions mais recrutait transitoirement les fonctions nucléaires à l'usine virale.

La deuxième partie est consacrée à la caractérisation d'un virus qui appartient à la famille des *Pandoraviridae* : le premier Pandoravirus, qui a été découvert en 2012, possède une morphologie unique au sein des virus, ainsi qu'un génome colossal à ADN double-brin de 2.5 Mb. Ce nouveau *Pandoraviridae* a également été isolé à partir d'un échantillon qui a été prélevé en Nouvelle-Calédonie, et a été appelé *Pandoravirus neocaledonia*. Des études comparatives avec d'autres virus appartenant à la même famille ont été réalisées en combinant plusieurs approches (génomique, protéomique, transcriptomique et microscopie électronique à transmission), afin de mieux comprendre les caractéristiques de ces virus inédits. Enfin, la morphologie étonnante de ces virus nous a poussés à étudier la composition de leur enveloppe, qui est constituée d'un réseau de fibres formant des structures lamellaires, afin de déterminer leur nature. Se pourrait-il que les Pandoravirus, contrairement aux autres virus, détournent la machinerie cellulaire pour construire leurs particules ? Dans ce cas, quelle serait leur place dans l'histoire évolutive des virus ?

Mots-clés : Virus géants, *Pandoraviridae*, *Marseilleviridae*, Evolution, Microscopie Electronique à Transmission.

Abstract:

The discovery of giant viruses about a decade ago has truly shaken our perception of the viral world. Historically, viruses are defined as parasites smaller than bacteria, which are able to pass through a Chamberland filter, and cannot be observed under a light microscope. Mimivirus, the first giant virus, which infects amoebas, was originally discovered in 1992 but was mistaken for a bacterium: it is only ten years later that its true nature was revealed. Mimivirus has a particle of about 750 nm in diameter and a double-stranded DNA genome of about 1.18 Mb: it encodes its own transcription and DNA repair machineries, and even proteins involved in translation. Mimivirus has outstanding characteristics for a virus, and questioned the very definition of the word "virus". This discovery and that of other Mimiviridae have initiated a debate on the origin and evolutionary history of these viruses, and it revived the debate on their nature: are viruses alive?

This thesis, which purpose was to isolate novel giant viruses from environmental samples in order to enhance our understanding of their mode of evolution and origin, is divided into two parts. The first part of this manuscript focuses on the characterization of a new *Marseilleviridae*, which was isolated from a sample collected in New-Caledonia, and which was named Noumeavirus. The *Marseilleviridae*, which founding member was discovered in 2009, are large DNA viruses that have icosahedral particles of about 200 nm in diameter, and double-stranded DNA genomes of more than 300 kb. Noumeavirus was compared with Melbournevirus, the most distant member of *Marseilleviridae*, which was isolated in the laboratory in 2012. Various approaches, such as genomics, proteomics and the study of their infectious cycles (with fluorescence and transmission electron microscopy), allowed us to reveal that the infectious cycle of these viruses was not independent from the cell nucleus as we thought, but was transiently recruiting nuclear functions to the viral factory.

The second part focuses on the characterization of a virus that belongs to the *Pandoraviridae* family: the first Pandoravirus, which was discovered in 2012, has a unique morphology and a gigantic double-stranded DNA genome of about 2.5 Mb. This new *Pandoraviridae* was also isolated from a sample that was collected in New-Caledonia and was named *Pandoravirus neocaledonia*. Comparative studies with other viruses belonging to the same family were performed using several approaches (genomics, proteomics, transcriptomics and transmission electron microscopy) to better understand the characteristics of these original viruses. Finally, the astonishing morphology of these viruses led us to investigate the composition of their envelope, which is made of a mesh of fibers forming lamellar structures. Is it possible that Pandoraviruses, unlike other viruses, hijack the cellular machinery to build their particles? In this case, what would be their place in the evolutionary history of viruses?

Key-words: Giant viruses, *Pandoraviridae*, *Marseilleviridae*, Evolution, Transmission Electron Microscopy.