



**HAL**  
open science

# Approches génomiques de l'interaction *Ostreococcus tauri*-Prasinovirus

Julie Thomy

► **To cite this version:**

Julie Thomy. Approches génomiques de l'interaction *Ostreococcus tauri*-Prasinovirus. Sciences de la Terre. Sorbonne Université, 2022. Français. NNT : 2022SORUS245 . tel-04046461

**HAL Id: tel-04046461**

**<https://theses.hal.science/tel-04046461>**

Submitted on 26 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sorbonne Université

École doctorale Sciences de la Nature et de l'Homme : Evolution et Ecologie

ED227

*Observatoire océanologique de Banyuls-sur-Mer – UMR 7232 Biologie Intégrative des Organismes Marins – Equipe Génomique Evolutive et Environnementale du Phytoplancton*

## Approches génomiques de l'interaction

### *Ostreococcus tauri–Prasinovirus*

Par : Julie Thomy

Thèse de Doctorat de Biologie

Dirigée par Dr. Nigel Grimsley & Dr. Sheree Yau

Présentée et soutenue publiquement le : 25/03/2022

#### Devant un jury composé de :

Claire Geslin	MCF – Université de Bretagne Occidentale	Rapportrice
Guillaume Blanc	DR CNRS – Aix-Marseille Université	Rapporteur
Anne-Claire Baudoux	CR CNRS – Sorbonne Université	Examinatrice
Sandra Jeudy	CR CNRS – Aix-Marseille Université	Examinatrice
Guillaume Charrière	MCF – Université de Montpellier	Examineur
Marcelino Suzuki	Pr. – Sorbonne Université	Examineur
Yves Desdevises	Pr. – Sorbonne Université	Examineur
Nigel Grimsley	DR émérite CNRS – Sorbonne Université	Directeur de thèse
Sheree Yau	CR CNRS – Sorbonne Université	Membre invité



*A ma bonne étoile*



## REMERCIEMENTS

*...Si j'en suis à écrire ces dernières lignes c'est que la fin n'est maintenant plus très loin. Je n'aime pas les grands discours mais il m'était indispensable de remercier toutes les personnes qui m'ont entourée, m'ont soutenue et m'ont guidée tout au long de cette superbe aventure, qu'est la thèse.*

*Avant toute chose je souhaite remercier l'ensemble des membres du jury qui m'ont fait l'honneur de juger mon travail de thèse et d'avoir été présent le jour de ma soutenance : Claire Geslin, Guillaume Blanc, Sandra Jeudy, Guillaume Charriere, Anne-Claire Baudoux, Marcelino Suzuki, Yves Desdevises*

*Je tiens à remercier mes deux superviseurs Nigel Grimsley et Sheree Yau qui m'ont fait confiance, m'ont soutenue et m'ont guidée tout le long de ma thèse. Merci Nigel pour m'avoir ouvert les portes du monde de la recherche et de la virologie marine. Je te remercie pour ton implication, ton engagement et ton suivi durant ces trois ans et demi. Merci de m'avoir un peu « bousculé » avec l'anglais car cela a porté ses fruits et je t'en remercie. Un très grand merci à toi Sheree pour ton dévouement à ce co-encadrement en fin de thèse. Je suis fière d'avoir été ta première étudiante et d'avoir pu bénéficier de ton super encadrement. Sans ça, ma thèse n'aurait pas été la même aujourd'hui. Je te remercie d'avoir toujours su me rassurer et me remotiver quand il le fallait. Je te remercie également pour ta disponibilité, ton implication, ton engagement et tout ce que tu as pu m'apprendre durant cette thèse : c'était passionnant et tellement enrichissant !!!!*

*Je remercie toute l'équipe Genophy qui m'a accueillie, m'a aiguillée et soutenue tout au long de cette aventure : Gwenaél, Manuel, Hervé, Carlos, Gilles, Marc, Julie, Manon. Gilles, je ne te remercierai jamais assez pour toute ton implication et ton dévouement que tu as pu avoir en fin de thèse. Je te remercie pour nos discussions enrichissantes et captivantes que nous avons eues : que ce soit de la microscopie, au virus en passant par les vésicules, et sans oublier les éléments transposables. Tu as vraiment réveillé en moi l'âme de chercheuse passionnée que je suis. Un très grand merci Fred pour ton aide précieuse à la paillasse, mais également toute ton implication et ton soutien dans le travail expérimental. Tu as toujours su répondre présent quand j'en avais besoin, un grand merci. Je te remercie aussi pour ta bonne humeur constante et ton sens de l'humour qui m'ont toujours fait garder le sourire. Merci pour tous les partagent que l'on a pu avoir : ta passion dévorante pour le Japon, la bouffe, et bien d'autres... Tout simplement, merci pour l'ami que tu es devenu. Bien*

qu'elle ne soit plus aujourd'hui dans cette équipe, je remercie Lucie pour sa gentillesse, sa bonne humeur, et surtout ses SUPER bon gâteaux ☺ !!!!

Adrien, merci pour ta bonne humeur, ta gentillesse, ta folie, qui m'ont permis de garder le sourire durant ces deux premières années de thèse. Merci pour tous ces partages et les jeux délirants sur l'ordinateur que tu m'as fait faire après le travail.

Merci Marc pour ta bonne humeur, ta passion pour les chats et les chinchillas : j'ai enfin trouvé une personne aussi « gaga » que moi pour les chats et les petites bestioles. Je te remercie pour ton soutien mais également ta disponibilité et toutes nos conversations enrichissantes. J'en viens à remercier mes anciens collègues thésards qui m'ont été une réelle source d'inspiration et qui m'ont aidée, soutenue et toujours motivée dans cette aventure. Merci Laurie, pour ta bonne humeur, ta gentillesse, la Toulousaine que tu es. Je suis tellement heureuse aujourd'hui que tu aies pu rebondir de plus belle, tu es une fille super et tu le mérites !! J'espère te revoir très bientôt. Thanks to Felipe for your good mood, your kindness and your precious help with the codes. I wish you good luck for the success of your career in the USA!!! Merci Rémy pour la personne fabuleuse que tu es, rempli de savoirs, tu as été une réelle source d'inspiration pour moi. Nos échanges sur nos projets respectifs étaient tellement passionnants et enrichissants. C'était un véritable plaisir de travailler avec toi. Je te souhaite pleins de bonnes choses pour ton post-doc en Suisse. J'espère te revoir très prochainement pour la soutenance. Enfin, je n'oublie pas l'ensemble des stagiaires que nous avons accueillis qui ont permis d'apporter une touche de bonne humeur et de convivialité à l'équipe : Paul, Sophie, Victor, Matéo, Yness, Julien, Camille, Hippolyte.

Je remercie toutes les personnes des plateformes confondues qui m'ont permis de mener à bien mon projet de thèse. La plateforme BioPIC à Banyuls-sur-Mer : David Pecqueur, Christophe Salmeron pour toute la partie cytométrie et Marie-Line Escandé pour la microscopie électronique à transmission. Un grand merci Marie-Line pour m'avoir fait partager ta passion pour le théâtre et ton aide précieuse pour ma thèse en 180s. Je te remercie infiniment de m'avoir ouvert les portes du monde de la microscopie, un univers qui m'était encore peu familier, mais qui aujourd'hui me fascine tellement !!! Je te remercie pour tout ton soutien et ton implication dans cette fin de thèse, notamment dans la recherche de ces fameux virus qui nous ont bien menés par le bout du nez... Merci à la plateforme Bio2Mar pour leur soutien et pour la maintenance des appareils de biologie moléculaire. Je dois aussi remercier le pôle bio-informatique : Vladimir Darric et Hélène Mayer pour leur aide si précieuse dans l'écriture de scripts ou tout simplement leur soutien dans ce monde parallèle, qu'est la bio-informatique. Un univers très mystérieux, complexe m'est qui aujourd'hui me passionne et me donne envie de continuer... Mes travaux n'auraient pas pu se faire sans la

disponibilité des ressources et services bio-informatiques de la plateforme Genotoul à Toulouse, un grand merci à eux et pour leurs formations.

Je ne pouvais pas finir sans remercier notre secrétaire d'unité, Didier Peuzé pour toute son implication et son aide précieuse dans la paperasse administrative. Hormis cela, tu as toujours eu une oreille bienveillante et attentive à mon égard et je t'en remercie infiniment.

Les derniers remerciements sont le plus souvent attribués à nos proches et à notre famille. Alors je voulais tout simplement les remercier pour leur soutien permanent et pour avoir toujours cru en moi. J'ai une pensée particulière pour tous les anciens doctorants de l'équipe de Mamadou Daffé à l'IPBS à Toulouse qui sont aujourd'hui aux quatre coins du monde mais qui restent toujours dans mon cœur : Richard, Cyril, Aïcha, Hung, Nicolas et ma petite Laura. « Votre petite compote de cassis » ☺. Laura, je voulais te remercier car aujourd'hui si j'en suis là c'est en partie grâce à toi. Tu as été un modèle pour moi, tu es une fille pétillante qui n'a jamais rien lâché même dans les plus durs moments. Ta rigueur, ton audace, ton intelligence, et ta passion pour ton sujet de thèse ont été un véritable tremplin pour moi. J'espère tellement te revoir un jour, peut-être en nouvelle Calédonie qui c'est ?

Je voulais remercier les personnes qui me sont le plus proches : Audrey, Amélie, Flavien, ma petite Mélanie, Laure. Audrey et Amélie qui l'auraient cru...Ma petite Mélanie, merci pour la personne que tu es, merci pour nos très longues conversations au téléphone. J'ai découvert une personne extraordinaire, pleine de talents, si touchantes mais également à l'écoute de ses amis. Une véritable sœur de cœur. Je te souhaite bon courage pour la poursuite de ta thèse, ne lâche rien, tu vas y arriver, tu es une battante !!

J'en viens à remercier une personne si chère à mes yeux, c'est toi Julien. Ces quelques lignes ne sont pas suffisantes pour te dire à quel point je suis tellement reconnaissante pour tout ce que tu as fait pour moi durant cette longue aventure. Tu sais que si je n'ai rien lâché c'est en grande partie grâce à toi. Ta gentillesse, ta bonne humeur, ta bien vaillance ont été une véritable bouffée d'oxygène. Je te remercie de m'avoir ouvert les portes de ton univers, ta green peace attitude ont été une évidence pour moi. Un énorme merci pour le merveilleux voyage à NYC, je ne l'oublierai jamais, quelle aventure...Je remercie aussi ta famille pour leur bienveillance, leur écoute et leur gentillesse. Vous êtes une superbe famille avec un énorme cœur. Merci Joanna pour votre générosité, votre gentillesse, et pour votre si grand cœur avec vos proches !!!

Enfin, je remercie mes parents et ma famille pour m'avoir toujours soutenue et avoir toujours cru en moi... Une pensée particulière à ma petite sœur Eva.

*« La pire des choses dans la vie est de réaliser un jour que l'on a vécu au-dessous de ses moyens. Par manque d'audace ou de courage »*

*Catherine Rambert*

# TABLES DES MATIÈRES

<b>Liste des figures</b> .....	<b>i</b>
<b>Liste des tableaux</b> .....	<b>ii</b>
<b>Liste des abréviations</b> .....	<b>iv</b>
<b>Introduction</b> .....	<b>2</b>
1. Le phytoplancton eucaryote.....	2
1.1. Qu'est-ce que le phytoplancton marin ?.....	2
1.2. Impact sur les cycles biogéochimiques et les réseaux trophiques marins.....	2
1.3. Le phytoplancton au sein de l'arbre du vivant eucaryote.....	3
1.4. Les Archaeplastida : la branche de la lignée verte.....	4
1.5. Les Mamiellophyceae.....	6
1.6. L'ordre des Mamiellales.....	7
2. Les virus marins.....	13
2.1. Qu'est-ce qu'un virus ? .....	13
2.2. Contexte historique des virus marins .....	19
2.3. Diversité des virus eucaryotes du phytoplancton.....	19
2.4. Que sont les virus géants ? .....	21
2.5. Diversité des virus géants du phytoplancton eucaryote .....	27
2.6. Les virus des <i>Mamiellophyceae</i> : Le genre <i>Prasinovirus</i> .....	31
3. Dynamiques des interactions hôte-virus.....	40
3.1. Impact des virus marins sur les cycles biogéochimiques et le climat .....	40
3.2. Impact des virus marins sur la structure des communautés planctoniques .....	42
3.3. Coévolution dans le système hôte-virus .....	43
3.4. Mécanismes de défenses antivirales.....	46
4. Objectifs de la thèse .....	49
<b>Chapitre I – Caractérisation génomique des événements d'insertion et impacts de la transformation médiée par le polyéthylène glycol sur le génome chez l'algue marine <i>Ostreococcus tauri</i></b> .....	<b>53</b>
1. Introduction .....	56
2. Material and Methods.....	58
3. Results .....	63
4. Discussion.....	74
<b>Chapitre II – Approche intégrative de la résistance antivirale chez <i>Ostreococcus tauri</i> entre génome et phénotype</b> .....	<b>81</b>
1. Introduction .....	83
2. Matériels et méthodes.....	85
3. Résultats .....	88
4. Discussion.....	100
<b>Chapitre III – Caractérisation génomique des prasinovirus : description des bases génétiques associées à l'évolution de la gamme de l'hôte <i>Ostreococcus tauri</i></b> .....	<b>107</b>
1. Introduction .....	110
2. Material and Methods.....	111
3. Results and discussion.....	116
4. General discussion .....	134
5. Conclusion .....	136
<b>Chapitre IV – Caractérisation morphologique des virus d'<i>Ostreococcus tauri</i></b> .....	<b>139</b>
1. Introduction .....	141
2. Matériel et méthodes .....	143

3. Résultats et discussions .....	145
4. Discussion.....	154
5. Conclusion.....	158
<b>Discussion générale et perspectives.....</b>	<b>161</b>
<b>Références bibliographiques .....</b>	<b>175</b>
<b>Documents supplémentaires du chapitre II.....</b>	<b>209</b>
<b>Documents supplémentaires du chapitre III .....</b>	<b>212</b>
<b>Documents supplémentaires du chapitre IV.....</b>	<b>215</b>
<b>Résumé.....</b>	<b>216</b>

# LISTE DES FIGURES

<b>FIGURE 1.</b> NOUVELLE RECONSTRUCTION PHYLOGENETIQUE DE L'ARBRE DU VIVANT EUCARYOTE..	4
<b>FIGURE 2.</b> ARBRE PHYLOGENETIQUE DE LA LIGNEE VERTE (ARCHAEPLASTIDA).....	6
<b>FIGURE 3.</b> MORPHOLOGIE DES MAMIELLALES.....	8
<b>FIGURE 4.</b> PHYLOGENIE DES SEQUENCES ADN RIBOSOMIQUE 18S DES MAMIELLALES..	11
<b>FIGURE 5.</b> CARACTERISTIQUES DU GENOME D' <i>O. TAURI</i> RCC4221 ..	12
<b>FIGURE 6.</b> REPRESENTATION SCHEMATIQUE DE LA DIVERSITE SYMETRIQUE DES VIRUS. ....	15
<b>FIGURE 7.</b> REPRESENTATION DES SEPT GROUPES DE VIRUS SELON LA CLASSIFICATION DE BALTIMORE. ....	16
<b>FIGURE 8.</b> CYCLE DE REPLICATION LYTIQUE DES VIRUS A ADN « CLASSIQUE ». ....	17
<b>FIGURE 9.</b> RECONSTRUCTION PHYLOGENETIQUE DES 7 FAMILLES DE NCLDVS. ....	24
<b>FIGURE 10.</b> MORPHOLOGIE DES <i>PRASINOVIRUS</i> . ....	34
<b>FIGURE 11.</b> ILLUSTRATION SCHEMATIQUE DES RELATIONS ENTRE LES ORGANISMES PLANCTONIQUES DANS LES ECOSYSTEMES MARINS. ....	41
<b>FIGURE 12.</b> MODELES INTERACTIONS ET DYNAMIQUES DE COEVOLUTIONS HOTE–VIRUS. ....	45
<b>FIGURE 13.</b> SEQUENCE MAP AND READ COVERAGE OF THE POLK4 VECTOR.....	65
<b>FIGURE 14.</b> PROCEDURE FOR FINDING THE INSERTION SITES OF TRANSFORMING DNA IN THE HOST GENOME USING SHORT-READ NGS DATA. ....	67
<b>FIGURE 15.</b> IGV (INTEGRATED GENOME VIEWER) VISUALIZATION OF READ COVERAGE AND IDENTIFICATION OF INSERTION SITES. ....	69
<b>FIGURE 16.</b> SCHEMATIC ILLUSTRATION OF THE VECTOR INTEGRATION SITES AND ASSOCIATED STRUCTURAL VARIATIONS IN CONTIGS ASSEMBLED FROM TRANSFORMANTS COMPARED TO THE REFERENCE GENOME. ....	71
<b>FIGURE 17.</b> FIREFLY LUCIFERASE ASSAY OF TRANSGENIC CLONAL LINES .....	73
<b>FIGURE 18.</b> PLAN EXPERIMENTAL DU CRIBLAGE PHENOTYPIQUE RESISTANT-SENSIBLE DES LIGNEES CLONALES.....	89
<b>FIGURE 19.</b> DYNAMIQUE D'INFECTION DES LIGNEES CLONALES TRANSFORMEES NON INFECTEES CONTROLES (BLEUE) ET INFECTEES A OTV09-578 (ROUGE) DURANT 14 JOURS POST-INFECTION. ....	90
<b>FIGURE 20.</b> ANALYSE DE L'ELECTROPHORESE A CHAMPS PULSEE DES LIGNEES CLONALES TRANSFORMEES (TX) ET NON- TRANSFORMEES (NTX)..	94
<b>FIGURE 21.</b> DOT PLOT NUCLEOTIDIQUE DES LIGNEES CLONALES TRANSFORMEES <i>O. TAURI</i> RCC1115.....	96
<b>FIGURE 22.</b> CARTOGRAPHIE GENETIQUE DES ELEMENTS MOBILES A LTR (CLASSE I) ASSOCIES AUX GENOMES DES CINQ CLONES TRANSFORMES <i>O. TAURI</i> RCC1115 .....	97
<b>FIGURE 23.</b> LOCATIONS OF THE SEAWATER SAMPLING SITES.....	117
<b>FIGURE 24.</b> HOST–VIRUS CROSS-INFECTION PATTERN .....	118
<b>FIGURE 25.</b> PHYLOGENETIC ANALYSIS OF PRASINOVIRUSES.....	121
<b>FIGURE 26.</b> COMPARISON OF THE COMPLETE GENOMES OF NINETEEN <i>OSTREOCOCCUS TAURI</i> VIRUSES (OTVS) .....	123
<b>FIGURE 27.</b> ORTHOLOGOUS PROTEIN DISTRIBUTION OF <i>PRASINOVIRUSES</i> .....	125
<b>FIGURE 28.</b> GENOMIC MAP COMPARING THE HYPERVARIABLE REGION OF THE 18 NEWLY SEQUENCED OTVS (BOLD).....	131
<b>FIGURE 29.</b> GENOMIC AND EVOLUTIONARY ANALYSIS OF THE POTASSIUM ION (K <sup>+</sup> ) CHANNEL PROTEIN ..	133
<b>FIGURE 30.</b> MORPHOLOGIE DE HUIT OTVS. ....	146
<b>FIGURE 31.</b> MORPHOLOGIE D'OTV19-T-TYPE 2. ....	147
<b>FIGURE 32.</b> MORPHOLOGIE D'OTV09-565 .....	149
<b>FIGURE 33.</b> PROPORTION DU NOMBRE DE PARTICULES VIRALES A QUEUE (ORANGE) OU SANS QUEUE (GRIS).. .....	150
<b>FIGURE 34.</b> INFECTION DES CELLULES <i>O. TAURI</i> RCC4221 AVEC LE VIRUS OTV09-565 .....	152
<b>FIGURE 35.</b> PROCEDURE EXPERIMENTALE D'EVOLUTION DE LA GAMME D'HOTE .....	168

# LISTE DES TABLEAUX

<b>TABLEAU 1.</b> CARACTERISTIQUES MORPHOLOGIQUES ET GENOMIQUES DES VIRUS REPRESENTATIFS DE CULTURE ISOLEE ASSOCIEE AU PHYTOPLANCTON EUCARYOTE .....	30
<b>TABLEAU 2.</b> LISTE DES PRINCIPAUX <i>PRASINOVIRUS</i> ISOLES ET DE LEURS HOTES ASSOCIES. ....	32
<b>TABLEAU 3.</b> CARACTERISTIQUES GENOMIQUES DES VINGT-DEUX GENOMES DE <i>PRASINOVIRUS</i> . ....	38
<b>TABLE 4.</b> ILLUMINA SEQUENCING AND MAPPING STATISTICS OF READS FROM CLONAL LINES AGAINST THE <i>O. TAURI</i> RCC1115 REFERENCE GENOME. ....	64
<b>TABLE 5.</b> MAPPING SUMMARY OF SHORT-READS BACK TO VECTOR SEQUENCE (POLK4) AND ESTIMATED INSERT COPY NUMBER IN EACH TRANSFORMED (Tx) CLONAL LINE.....	66
<b>TABLEAU 6.</b> DESCRIPTION PHENOTYPIQUE DES LIGNEES CLONALES TRANSFORMEES (Tx) CANDIDATES FACE A L'INFECTION VIRALE OtV09-578.....	89
<b>TABLEAU 7.</b> TEST D'INFECTION CROISE DES LIGNEES TRANSFORMEES (Tx) ET NON-TRANSFORMEES (NTx). ....	93
<b>TABLEAU 8.</b> IDENTIFICATION DES ETs DANS LE GENOME DES LIGNEES <i>O. TAURI</i> RCC1115.....	99
<b>TABLE 9.</b> GENERAL CHARACTERISTICS OF NINETEEN NEWLY SEQUENCED <i>PRASINOVIRUS</i> GENOMES.....	119



# LISTE DES ABREVIATIONS

Les abréviations sont indiquées en français et/ou en *anglais* selon leur apparition dans le manuscrit de thèse.

## A

ADN ( <i>DNA</i> )	Acide désoxyribonucléique ( <i>Deoxyribonucleic acid</i> )
ADNdb	Acide désoxyribonucléique double-brin
ADNsb	Acide désoxyribonucléique simple-brin
aLRT	<i>Approximation likelihood ratio test</i>
ARN ( <i>RNA</i> )	Acide ribonucléique ( <i>Ribonucleic acid</i> )
ARNdb	Acide ribonucléique double-brin
ARNm	Acide ribonucléique messenger
ARNr	Acide ribonucléique ribosomique
ARNsb	Acide ribonucléique simple-brin
AU	Acetate d'uranyle

## B

<i>BER</i>	<i>Base Excision Repair</i>
BET	Bromure d'éthidium
<i>BLAST</i>	<i>Basic Local Search Tool</i>
<i>BOC</i>	<i>Big outlier chromosome</i>
BpV	<i>Bathycoccus virus</i>
<i>BWA</i>	<i>Burrows-Wheeler Aligner</i>

## C

<i>Cas</i>	<i>CRISPR- associated genes</i>
<i>CDS</i>	<i>Coding DNA Sequence</i>
CO <sub>2</sub>	Dioxyde de Carbone
<i>CRISPR</i>	<i>Clustered Regularly Interspaced Short Palindromic Repeats</i>
cryoEM	<i>Cryogenic Electron Microscopy</i>
CTAB	Bromure de cétyltriméthylammonium

## D

<i>DDR</i>	<i>DNA damage response</i>
DMS	Diméthylsulfure
DMSP	Diméthylsulfoniopropionate

## E

EDTA	Éthylènediaminetétraacétique
ET	Elément Transposable
<i>EVE</i>	<i>Endogenous Viral Element</i>

## G

G+C	Guanine/Cytosine
-----	------------------

## H

Hpi	Heure post-infection
HR	<i>Homologous Recombination</i>
HRAPs	<i>High rate algal ponds</i>
<b>I</b>	
ICTVs	<i>International committee on taxonomy of viruses</i>
IRES	<i>Internal ribosome entry sites</i>
iRNA	<i>interference RNA</i>
iTOL	<i>Interactive Tree Of Life</i>
ITS	<i>Internal transcribed spacer</i>
<b>J</b>	
JPI	Jours Post-infection
<b>K</b>	
Kb	Kilo paires de bases
<b>L</b>	
LGT	<i>Lateral Gene Transfer</i>
LIRR	<i>Large Inverted Repeat Region</i>
LTR	<i>Long Terminal Repeats</i>
<b>M</b>	
MA	<i>Matching alleles</i>
MAG	<i>Metagenome Assembled Genomes</i>
Mb	Méga paires de bases
MCP	Majeur capside protéine ( <i>Major capsid protein</i> )
MET	Microscopie électronique à transmission
MIMIVIRE	<i>MIMIVirus VTrophage Resistance Element</i>
ML	<i>Maximum-Likelihood</i>
MOD	Matière organique dissoute
MOI	<i>Multiplicity of infection</i>
MOP	Matière organique particulaire
MpV	<i>Micromonas Virus</i>
µm	Micro-mètre
<b>N</b>	
NCLDV	<i>Nucleocytoplasmic Large DNA Virus</i>
NGS	<i>Next-generation sequencing</i>
NHEJ	<i>Non-homologous end joining</i>
nm	Nanomètre (Nanometer)
NTx	Non-Transformé
<b>O</b>	
OG	<i>Orthogroup</i>
OIV	<i>Ostreococcus lucimarinus Virus</i>
OmV	<i>Ostreococcus mediterraneus Virus</i>
ONT	<i>Oxford Nanopore Technology</i>
OSD	<i>Ocean Sampling Day</i>

OtV	<i>Ostreococcus tauri Virus</i>
<b>P</b>	
pATPase	<i>A32-like packaging ATPase</i>
pb ( <i>bp</i> )	<i>Paire de bases (base paire)</i>
PCR	<i>Polymerase chain reaction</i>
PEG	<i>Polyéthylène glycol</i>
PFGE	<i>Pulsed-field gel electrophoresis</i>
piRNA	<i>piwi RNA</i>
pOLK	<i>plasmid Ostreococcus Luciferase Kanamycin</i>
<b>R</b>	
RCC	<i>Roscoff culture collection</i>
RLU	<i>Relative Luminescence Unit</i>
RNAP	<i>DNA-dependent RNA Polymerase</i>
RT	<i>Rétro-transcription</i>
RT-qPCR	<i>Quantitative reverse transcription PCR</i>
<b>S</b>	
SAG	<i>Single-cell Amplified Genome</i>
SARs	<i>Straménopila Alveolata Rhizaria</i>
SH	<i>Shimodaira-Hasegawa</i>
SOC	<i>Small Outlier Chromosome</i>
<b>T</b>	
T-DNA	<i>transfer-DNA</i>
TAC/TAP	<i>Tail Assembly Chaperon protein</i>
TAE	<i>Tris-Acetate-EDTA</i>
TALEN	<i>TAL Effector Endonuclease</i>
TFIIIS	<i>Transcription factor II</i>
TIR	<i>Terminal Inverted Repeats</i>
TRIM	<i>Terminal Repeat Retrotransposon in Miniature</i>
tRNA	<i>transfer-RNA</i>
Tx	<i>Transformé</i>
<b>U</b>	
UV	<i>Ultra-violet</i>
<b>V</b>	
VLTF3	<i>Viral late gene transcription factor 3</i>
<b>Z</b>	
ZFN	<i>ZINC-finger nuclease</i>



# INTRODUCTION

## 1. Le phytoplancton eucaryote

### 1.1. Qu'est-ce que le phytoplancton marin ?

Selon la définition du zoologiste Allemand Hensen (1887), le terme plancton désignerait l'ensemble des organismes vivants errant en suspensions dans les environnements aquatiques et incapables d'échapper aux courants d'eaux du fait de leur petite taille. Ces organismes très diversifiés regroupent des bactéries, des archées et des eucaryotes sous forme unicellulaire ou pluricellulaire. Du point de vue de leur morphologie, le plancton peut s'étendre à des sous-ensembles allant du megaplancton (20 à 200 cm) jusqu'au femtoplancton (incluant le virioplancton d'une taille inférieure à 0.2  $\mu\text{m}$ ) en passant par le picoplancton (0.2 cm à 2  $\mu\text{m}$ ). Ce travail de thèse concerne uniquement les organismes picoplanctoniques et virioplanctoniques.

Le nom phytoplancton (du termes grec *phuton* ou « plante ») par étymologie désigne le plancton végétal, c'est à-dire l'ensemble des organismes marins photosynthétiques vivants en suspension dans l'eau. Contrairement au zooplancton, qui lui est hétérotrophe, les organismes du phytoplancton sont eux autotrophes vis-à-vis du carbone, mais aussi mixotrophes pour certains. La croissance du phytoplancton dépend donc de la disponibilité de la lumière et des nutriments dans tous les milieux aquatiques qu'il colonise. Par conséquent, l'abondance du phytoplancton est inégale à la surface du globe. Dans les océans, il se développe dans la zone euphotique (i.e. surface de l'eau) de la colonne d'eau.

### 1.2. Impact sur les cycles biogéochimiques et les réseaux trophiques marins

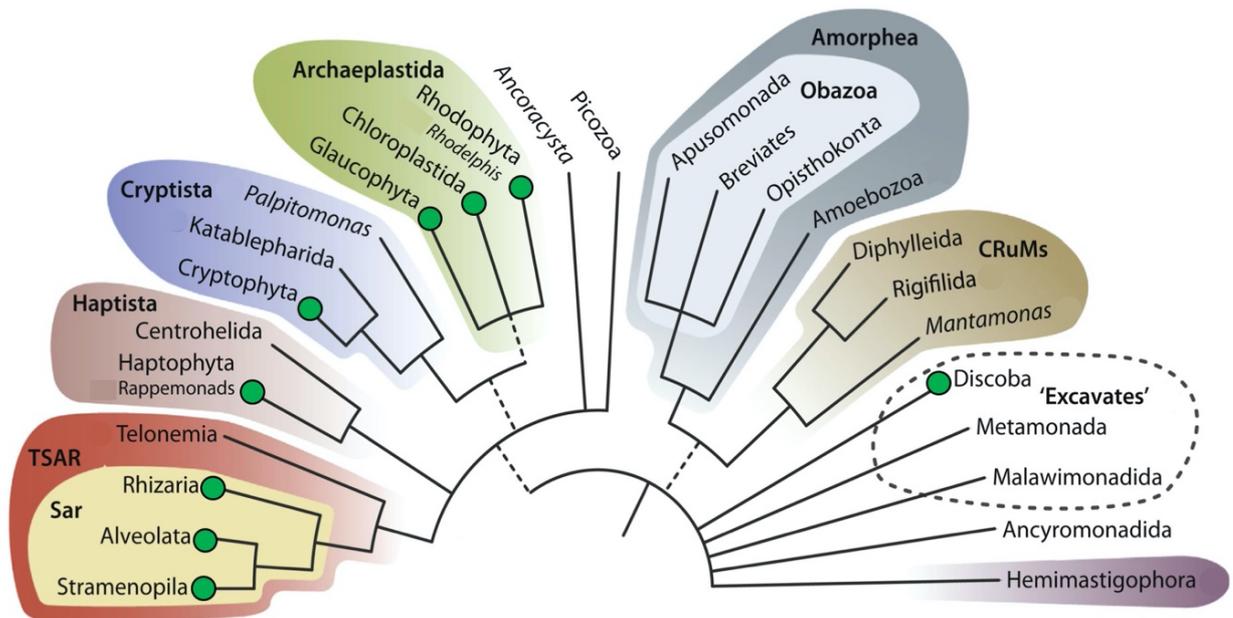
Bien que ne représentant qu'une très faible proportion de la biomasse à l'échelle planétaire (moins de 1%)<sup>1</sup>, le phytoplancton marin est responsable de plus de la moitié de la production primaire nette par an sur Terre<sup>2</sup>. A la base de la chaîne alimentaire en milieu marin, les producteurs primaires phototrophes (ou photoautotrophes) fixent 50% du dioxyde de carbone atmosphérique<sup>3</sup> vers la zone euphotique des océans en le convertissant en carbone organique et en dioxygène *via* le processus chimique de la photosynthèse oxygénique<sup>4</sup>.

Le carbone organique dissous libéré lors de la photosynthèse par le phytoplancton est ensuite réutilisé presque exclusivement par les bactéries et les archées hétérotrophes<sup>2</sup>. Les bactéries hétérotrophes seraient l'un des principaux contributeurs de la minéralisation et de la transformation de cette matière organique dissoute (MOD) en biomasse bactérienne dans la colonne d'eau, réutilisable par les maillons supérieurs de la chaîne trophique<sup>5,6</sup>. Les bactéries hétérotrophes participent donc activement aux flux de carbone et d'énergie comme un intermédiaire trophique entre les débris phytoplanctoniques et les réseaux trophiques supérieurs<sup>7</sup>. Ainsi, la production nette bactérienne équivaldrait à plus de 25% de la production primaire calculée, dont 45% serait directement canalisée par les bactéries hétérotrophes au sein de ce qui est appelée « la boucle microbienne »<sup>8</sup>. Nous verrons plus loin dans l'introduction, l'impact des virus sur les réseaux trophiques marins au sein de cette boucle (cf. partie 3.1 de l'introduction).

### 1.3. Le phytoplancton au sein de l'arbre du vivant eucaryote

Il est remarquable de noter que l'apparition de toute vie vertébrée et d'organismes multicellulaires complexes sur Terre incluant l'homme a été rendu possible grâce à l'apparition du premier organisme eucaryote photosynthétique il y a 1.5 milliards d'années<sup>9,10</sup>. Ce premier organisme photosynthétique serait à l'origine d'un ancêtre anaérobie des cyanobactéries qui aurait acquis la fonction de photosynthèse oxygénique avant l'oxygénation de la vie sur Terre<sup>11,12</sup>. Le premier organisme phytoplanctonique eucaryote aurait pris son origine dans le processus symbiotique entre une cellule eucaryote hétérotrophe (non photosynthétique) et une cyanobactérie photosynthétique, appelé « l'endosymbiose plastidique primaire »<sup>10,13,14</sup>.

Le supergroupe des Archaeplastida (autrement connu sous le nom de *Plantae*) émergea de ce premier événement d'endosymbiose, englobant le groupe des Glaucophyta (algues bleues), des Rhodophyta (algues rouges) et des Chloroplastida (ou Viridiplantae)<sup>15</sup>. La distribution extrêmement diversifiée des organismes phytoplanctoniques est aussi retrouvée dans les supergroupes des Cryptista (groupe Cryptophyta), Haptista (groupe Haptophyta) et les SARs (Stramenopila, Alveolata, et Rhizaria), provenant d'une endosymbiose secondaire entre une cellule eucaryote hétérotrophe et une algue rouge (**Figure 1**). Dans ce manuscrit, nous nous focaliserons uniquement sur le supergroupe des Archeplastida et plus précisément des organismes issus de la lignée verte Chloroplastida.



**Figure 1.** Nouvelle reconstruction phylogénétique de l’arbre du vivant eucaryote. Les supergroupes sont définis par les différentes zones colorées. Les ordres non résolus sont montrés par les multiples branchements. Les lignes en pointillées reflètent une incertitude quant à la monophylie de certains groupes. Les lignées qui regroupent des organismes phytoplanctoniques sont annotées par un cercle vert aux extrémités des feuilles de l’arbre. Figure adaptée de Burki *et al.* <sup>15</sup>.

#### 1.4. Les Archaeplastida : la branche de la lignée verte

La lignée verte désigne l’ensemble des organismes qui inclut les algues vertes et les plantes (embryophytes). Elle forme le groupe monophylétique le plus dominant des producteurs primaires sur Terre<sup>16</sup>, qui aurait émergé il y a plus d’un milliard d’année <sup>10,17,18</sup>. La caractéristique commune que partagent les algues vertes et les plantes terrestres est la présence d’un chloroplaste à double membrane contenant les principaux pigments tels que la chlorophylle *a* et la chlorophylle *b*, ainsi que des pigments accessoires tels que les carotènes et les xanthophylles<sup>19</sup>.

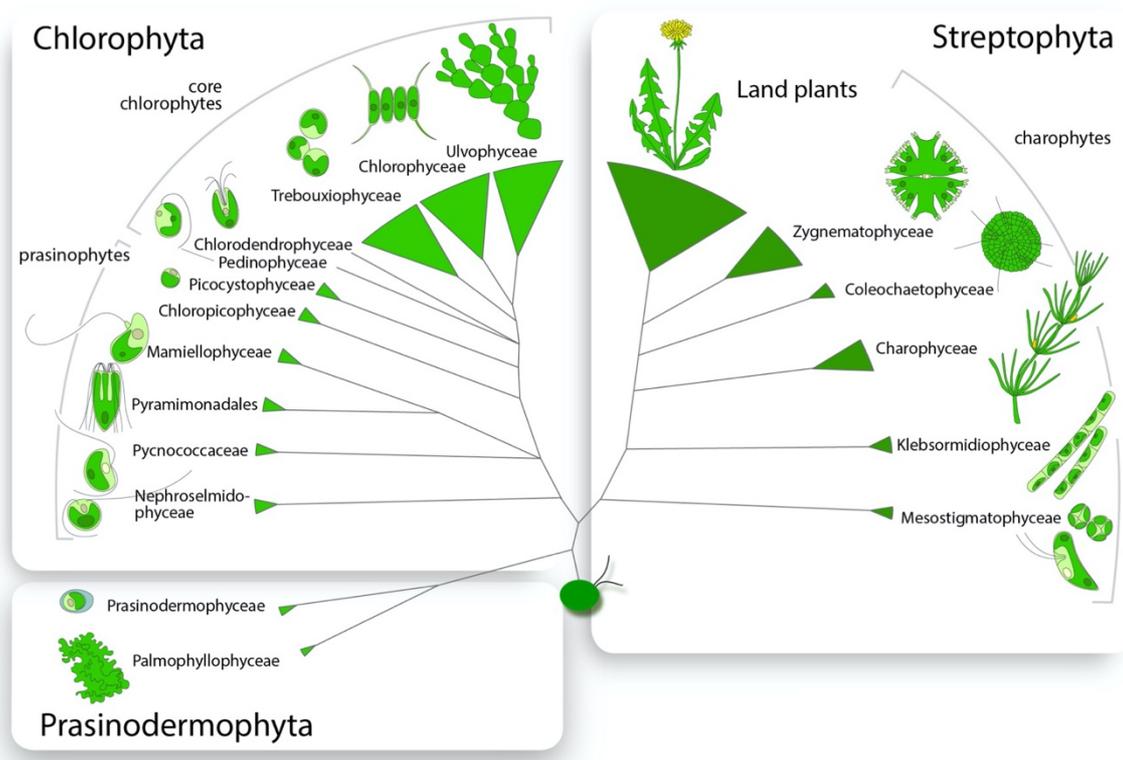
Avant les méthodes d’analyses moléculaires, la classification phylogénétique des algues vertes s’est longtemps basée sur les caractérisations morphologiques grâce à l’utilisation de microscopes. Ces observations ont mis en évidence une grande diversité comptant plus de 5000 espèces qui présente une importante variation architecturale, morphologique<sup>20</sup> et de tailles de corps (i.e. les thalles). Les thalles comportent des structures très complexes, allant de cellules

uniques (unicellulaires) nageuses à des cellules non mobiles, ainsi qu'aux colonies ou aux algues filamenteuses. Contrairement aux cellules filamenteuses ou aux multicellulaires très complexes, les algues unicellulaires partagent une architecture commune avec une morphologie cellulaire entre sphérique et allongée, avec ou sans flagelle, avec la présence d'écaille ou des parois recouvrant la cellule<sup>21</sup>.

Au cours des dernières décennies, les approches moléculaires ont permis d'élargir les connaissances des algues photosynthétiques eucaryotes et ont révélé une diversité qui était largement sous-estimée auparavant. Le développement de nouvelles méthodes de séquençage, tel que le metabarcoding ou métagénomique, a permis de mieux mesurer cette biodiversité, notamment pour les espèces phytoplanctoniques de très petites tailles avec peu de caractère morphologique différentiable<sup>22,23</sup>. La combinaison des approches microscopiques et moléculaires a alors favorisé une description plus fine des relations phylogénétiques des organismes issus de la lignée verte définissant le groupe des Archaeplastida.

L'ensemble des plantes et des algues de cette lignée verte (i.e. Archaeplastida) a donné naissance à deux grandes lignées que sont les Streptophyta et Chlorophyta<sup>19,21</sup>. Les Chlorophyta qui regroupent les espèces d'algues vertes planctoniques unicellulaire, contiennent les « chlorophytes » et les « prasinophytes ». Les prasinophytes (ou flagellés verts écailleux) forment un groupe paraphylétique d'algues essentiellement marines. Les chlorophytes se composent d'une grande diversité avec des espèces microscopiques unicellulaires et multicellulaires à des formes macroscopiques colonisant les environnements d'eau douce, terrestres et côtiers<sup>24</sup>. Les Streptophyta, eux, constituent l'ensemble des plantes terrestres (embryophytes) mais englobent aussi un groupe paraphylétique d'algue verte regroupant des espèces majoritairement d'eau douce et quelques espèces terrestres, les Charophytes<sup>19</sup>.

Plus récemment, basé sur des analyses phylogénomiques, une troisième grande lignée qui aurait divergé avant la séparation des Streptophyta et Chlorophyta a été proposée, les Prasinodermophyta. Cette nouvelle lignée comprend deux classes, les Prasinodermophyceae et les Palmophyllophyceae qui regroupent des espèces d'algues vertes retrouvées plus en profondeur dans les régions marines oligotrophes<sup>25</sup> (**Figure 2**).



**Figure 2.** Arbre phylogénétique de la lignée verte (Archaeplastida). D’après Leliaert *et al.*<sup>19</sup>, révisé le 2 janvier 2021.

## 1.5. Les Mamiellophyceae

Dans les Chlorophyta, les Mamiellophyceae forment la plus large classe monophylétique d’algue verte unicellulaire picoeucaryote ( $<3 \mu\text{m}$ )<sup>26,27</sup>. Du point de vue écologique, le picophytoplancton domine principalement les environnements oligotrophes pélagiques et fournit une contribution majeure dans les réseaux trophiques marins par la production primaire<sup>28</sup>. Parmi l’ensemble des picoeucaryotes, la classe des Mamiellophyceae est très largement distribuée géographiquement dans les régions marines côtières, détectées dans toutes les analyses métagénomiques décrites à ce jour<sup>29–32</sup>. Ces études ont mis en évidence une distribution massive de cette communauté picophycoplantonique caractérisée principalement dans des environnements tempérés ou Arctiques<sup>32–34</sup>.

Les Mamiellophyceae sont divisés en trois ordres : Monomastigales, Dolichomastigales et Mamiellales. Les ordres Monomastigales et Dolichomastigales contiennent respectivement les familles Monomastigaceae, Dolichomastigaceae et Crustomastigaceae ; les Mamiellales

englobent les familles des Bathycoocaceae et Mamiellaceae<sup>26</sup>. Les Monomastigales sont exclusivement d'eau douce, les Dolichomastigales regroupent une plus large diversité d'espèce distribuées dans les régions marines côtières, et les Mamiellales proviennent exclusivement de région marine<sup>29,35</sup>.

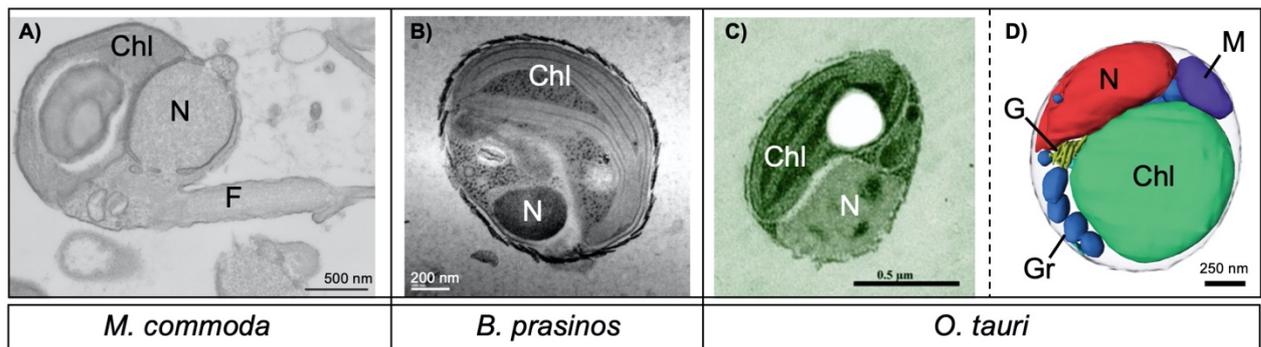
#### 1.4. L'ordre des Mamiellales

L'ordre des Mamiellales est représenté par les genres *Micromonas*, *Bathycoccus*, *Ostreococcus*, *Mantoniella* et *Mamiella* qui sont distribués exclusivement dans les environnements marins<sup>26</sup>. Les trois genres *Micromonas*, *Bathycoccus* et *Ostreococcus* sont les mieux décrits dans les données génomiques puisqu'ils apparaissent très fréquemment dans les échantillons d'écosystèmes marins. Du point de vue morphologique, ils sont décrits jusqu'à ce jour au sein de la lignée verte comme les plus petits eucaryotes photosynthétiques connus.

Les Mamiellales sont donc des algues unicellulaires de petite taille (diamètre cellulaire < 3 µm) possédant une morphologie commune, mais aussi quelques caractères particuliers comme la présence de flagelles (propre à la famille des Mamiellaceae) ou d'écailles non minérales (de type polysaccharidique) variables selon les espèces (aplaties, arrondies, ornementées ou réticulées) à la surface des cellules entourant la membrane plasmique<sup>26</sup>. Ces écailles sont absentes chez *Micromonas* et *Ostreococcus*, tandis que les flagelles sont absentes chez *Ostreococcus* et *Bathycoccus* (**Figure 3**). Ces picoalgues ont une architecture cellulaire très minimaliste et un rapport surface/volume élevé<sup>36</sup>. Elles sont pourvues d'un chloroplaste (parfois deux) qui détient les pigments photosynthétiques (chlorophylle *a* et *b*), ainsi qu'un mélange complexe de pigments accessoires incluant deux pigments caroténoïdes : la prasinoxanthine<sup>37</sup> et l'uriolide<sup>38</sup>, propres aux Mamiellales<sup>39</sup>. Elles possèdent aussi une mitochondrie, un appareil de Golgi, des grains d'amidon et un noyau détenant un génome de petite taille allant de 13 Mb (*Ostreococcus*) jusqu'à 23 Mb (*Micromonas*)<sup>26,40-43</sup>.

Des études se basant sur des analyses de metabarcoding et de métagénomique ont révélé que les trois genres *Micromonas*, *Bathycoccus* et *Ostreococcus* sont composées d'espèces cryptiques<sup>46-49</sup>. Sur le plan morphologique, ces espèces ne présentent aucune différence, mais, d'un point de vue génétique, elles mettent en avant des différences notables. En effet, les espèces au sein des genres de *Micromonas*, *Bathycoccus* et *Ostreococcus* se distinguent uniquement sur les séquences ribosomiques ITS (*Internal transcribed spacer*). Toutefois, certaines de ces espèces présentent une identité sur les séquences ribotypes de la petite sous-

unité 18S. La distribution géographique de ces espèces a permis de mettre en évidence des niches écologiques distinctes<sup>33,46,50,51</sup>.



**Figure 3.** Morphologie des Mamiellales. Observation par Microscopie Electronique à Transmission (MET) des cellules de (A) *Micromonas commoda* (RCC299) (B), *Bathycoccus prasinos* (RCC1105) et (C) *Ostreococcus tauri*. Visualisation de l’ultrastructure en trois dimensions de la cellule d’*O. tauri* par l’approche cryotomographie. Chl: chloroplaste ; N: noyau ; G: appareil de Golgi ; Gr: granules ; M: mitochondrie. Sources: (A) Van Baren *et al.* 2016<sup>44</sup>, (B) Moreau *et al.* 2012<sup>43</sup>, (C) Hervé Moreau à Banyuls-sur-Mer, (D) Henderson *et al.* 2007<sup>45</sup>.

### Le genre *Micromonas*

*Micromonas* est un genre ubiquiste avec des espèces isolées à partir d’un large éventail d’environnement allant des tropiques aux pôles mais reste prédominant dans les régions côtières pélagiques<sup>31,33,52</sup>. Récemment, des études phylogénétiques et écologiques ont permis de diviser le genre *Micromonas* en 9 clades majeurs<sup>31,53</sup>.

*M. bravo* est une espèce<sup>53</sup> (**Figure 4**) récemment divisée en deux sous-clades (B1 et B2), avec une divergence génétique estimée à 99.3% (région V4 de l’ARNr 18s)<sup>31</sup>. Le clade *M. bravo* B1 semble être spécifique aux stations côtières en Europe, tandis que *M. bravo* B2 apparaît plus largement répandu géographiquement dans les environnements tempérés sur les côtes et au large des océans.

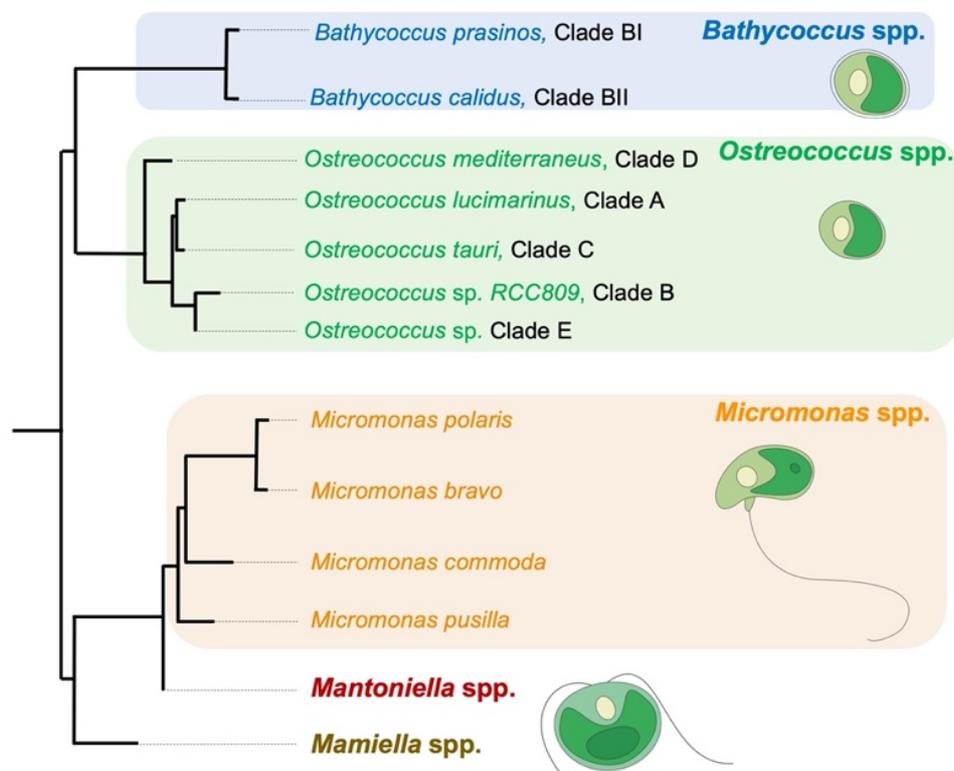
La seconde espèce la plus représentative de ce genre est *M. commoda* (**Figure 4**), qui a été divisée en sous-clades (A1 et A2) en se basant sur des analyses phylogénétiques, et qui montre des origines géographiquement distinctes<sup>31</sup>. *M. commoda* A1 domine les régions tropicales et subtropicales, tandis que *M. commoda* A2 est plus adaptée à des environnements d’eaux marines plus froides dans les régions côtières ainsi qu’au large des océans. Dans ce

travail de thèse, la souche *M. commoda* A1 RCC299 – nouvellement décrit comme RCC827 – a été utilisée pour l'isolement d'un nouveau prasinovirus (i.e. MpV20-T), à proximité des côtes chiliennes au sud-est de l'Océan Pacifique (cf. Chapitre 3).

*M. polaris* est une espèce nouvellement décrite<sup>53</sup> (**Figure 4**) dans les écosystèmes marins arctiques<sup>31</sup> (mais absente dans les environnements antarctiques<sup>54</sup>).

Enfin, *M. pusilla* – anciennement décrite comme *Chromulina pusilla* – est la première algue verte phytoplanctonique picoeucaryote à avoir été découverte<sup>55</sup>. En revanche, sa distribution est faiblement observée dans les écosystèmes marins des zones tempérées<sup>31</sup>.

Trois nouveaux clades génétiquement divergents (i.e. B3, B4, B5) ont été décrits récemment<sup>31</sup>. *Micromonas* sp. clade B3 est phylogénétiquement relié à *M. polaris*, bien qu'il soit prévalent dans des stations subarctiques ainsi que dans des régions plus tempérées dans la Mer du Nord. Le clade B4, proche génétiquement de *M. commoda*, a été plutôt retrouvé dans les eaux chaudes, ainsi que *Micromonas* sp. clade B5 qui montre une relation phylogénétique plus proche de *M. pusilla*.



**Figure 4.** Phylogénie des séquences ADN ribosomique 18S des Mamiellales. Adaptée de Benites *et al.* 2019<sup>56</sup> et Tragin et Vaultot. 2019<sup>31</sup>.

## *Le genre Bathycoccus*

*Bathycoccus* est représenté comme le deuxième genre le plus important en abondance dans les écosystèmes pélagiques marins au sein des Mamiellophyceae. Des études environnementales moléculaires indiquent la présence de deux espèces cryptiques avec des habitats environnementaux bien distincts qui ne coexistent que très rarement<sup>31,49,57</sup>. Ces espèces possèdent des séquences ARNr 18s identiques et se différencient uniquement sur les séquences ITS ainsi que des divergences plus importantes à l'échelle du génome<sup>49,58</sup>. L'écotype clade BI correspond à l'unique génome séquencé *B. prasinus* RCC1105 (**Figure 4**), qui a été isolé dans la Mer Méditerranée<sup>43</sup>. Par ailleurs, les données métagénomiques associés à ce clade ont également été détectées dans des écosystèmes mésotrophes côtiers plus frais étendu jusqu'aux pôles<sup>31</sup>. En revanche, le clade BII, redéfini très récemment comme l'espèce *B. calidus* (**Figure 4**)<sup>57,59</sup>, est représenté par les deux souches RCC715 et RCC716 échantillonnés dans l'Océan Indien, ainsi que par le génome amplifié unicellulaire (ou SAG : *Single-cell Amplified Genome*) TOSAG39-1, détecté dans les eaux chaudes de la mer d'Arabie<sup>49</sup>. Des analyses moléculaires quantitatives et métagénomiques indiquent aussi une distribution plus étendue du clade BII, qui serait adapté aux environnements marins plus salés incluant des régions subtropicales pauvres en éléments nutritifs<sup>31,60</sup>.

## *Le genre Ostreococcus : la plus petite cellule eucaryote connue*

### ***Découverte et caractères morphologiques***

*Ostreococcus* a été initialement isolée à la surface de l'étang de Thau (Mer Méditerranéenne) en France, et tire son nom *Ostreococcus tauri* de son lieu d'origine<sup>36</sup>. Des analyses de cytométrie en flux ont montré que ce picoeucaryote domine largement la population phytoplanctonique dans l'étang de Thau en terme de nombre de cellules et de biomasse annuelle estimée à environ 30% mais pouvant atteindre 86% en saisons estivales<sup>61,62</sup>. L'observation par microscopie électronique a révélé une ultrastructure cellulaire simple et de très petite taille (<1 µm) en forme de coccoïde, dépourvue de paroi cellulaire, sans écailles ni flagelles<sup>36,61</sup>. Remarquablement, *Ostreococcus* apparaît comme la plus petite cellule eucaryote décrite à ce jour. Une reconstruction tridimensionnelle par cryotomographie d'*O. tauri* a permis de détailler cette organisation architecturale cellulaire minimaliste. Sa cellule se compose d'un chloroplaste et d'un noyau (organites les plus volumineux), ainsi que d'une mitochondrie, d'un peroxyosome, d'un réticulum endoplasmique, d'un appareil de Golgi et des grains d'amidon situés dans un

compartiment cytoplasmique très restreint (i.e. environ 30% du volume cytoplasmique)<sup>45</sup> (**Figure 3**).

### ***Distribution géographique dans les régions marines***

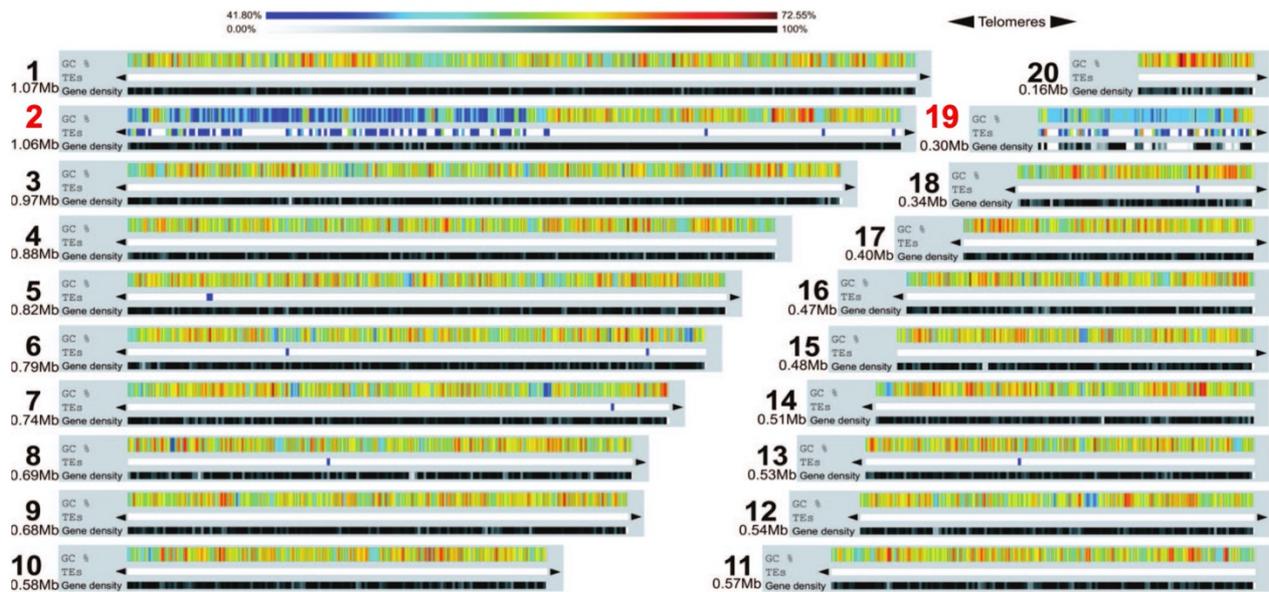
Depuis sa découverte, de nombreuses souches ont été isolées dans différents environnements, en surface ou dans la colonne d'eau jusqu'à 120 m de profondeur à l'échelle planétaire<sup>63</sup>. Le genre *Ostreococcus* est ubiquitaire mais domine principalement les régions oligotrophes tropicales et subtropicales côtières dans les stations méditerranéennes<sup>62</sup> et dans les estuaires marins<sup>64</sup>. Certaines souches sont toutefois plus adaptées aux environnements froids en régions côtières<sup>65</sup> ou aux ouverts des océans<sup>57,63,66-68</sup>. En revanche, *Ostreococcus* n'a pas été retrouvé dans les régions polaires<sup>31,33</sup>. Des analyses phylogénétiques moléculaires indiquent la présence de cinq clades (A, B, C, D et E) (reconnus aujourd'hui comme des espèces) (**Figure 4**). Contrairement aux autres membres des Mamiellales, la distribution de ces espèces ne correspond pas à leur distribution géographique mais plutôt à leur profondeur d'isolement et donc à la lumière à laquelle elles sont exposées dans les environnements naturels. Ainsi, les analyses en teneur de pigments et de la réponse aux niveaux de lumière ont permis de discriminer deux grands écotypes : les souches adaptées à des niveaux de forte intensité lumineuse (*Ostreococcus lucimarinus* (i.e. clade A), *Ostreococcus tauri* (i.e. clade C) et *Ostreococcus mediterraneus* (i.e. clade D)) et celles adaptées à de plus faibles teneurs en lumières (i.e. *Ostreococcus* clade B)<sup>46</sup>. Plus récemment, un nouveau clade (clade E) a été décrit dans les données métagénomiques, et apparaît largement distribué dans les océans. En revanche, aucune culture n'a été obtenue à ce jour<sup>31</sup>.

### ***Caractéristiques génomiques***

*Ostreococcus* reste à ce jour le genre le mieux caractérisé des Mamiellophyceae, avec un total de 16 génomes séquencés pour les quatre premières espèces décrites<sup>40,41,69-71</sup>. Pour les autres genres, aucun génome de population n'est disponible mais seulement 1 ou 2 génomes d'espèce.

La taille des génomes d'*Ostreococcus* est comparable à celle des levures *Saccharomyces cerevisiae* et *Schizosaccharomyces pombe*, mesurant environ 13 méga paires de base (Mb). Son contenu en guanine et cytosine (G+C) est élevé pour un organisme eucaryotes<sup>72</sup> (~ 60% comparé à 41% pour l'homme, 38% pour *Saccharomyces cerevisiae* et 36% pour *Arabidopsis*

*thaliana*). Une seconde remarquable caractéristique d'*O. tauri* est le degré de compaction de son génome, qui se répartit en 20 chromosomes comptabilisant environ 8000 gènes inter-espacés de très courtes régions (196 pb en moyenne)<sup>40</sup> (Figure 5). De plus, une des principales caractéristiques génomiques communes aux Mamiellales est la présence de deux chromosomes « aberrants » (dit *outlier*) qui ont des différences de structures, de compositions et de variations de tailles par rapport aux 18 autres chromosomes.



**Figure 5.** Caractéristiques du génome d'*O. tauri* RCC4221. Les 20 chromosomes sont représentés de haut en bas du premier au dernier chromosome et dans l'ordre croissant selon leur taille en Mega paires de base (Mb). Le numéro du chromosome et sa taille sont montrés à gauche. Le chromosome 2 et le chromosome 19 sont le BOC et SOC respectivement (indiquée en rouge gras). La barre colorée supérieure indique la composition en pourcentage de G+C, tandis que la barre inférieure représente le pourcentage en éléments transposables (TEs – *Transposable Elements*) répartis sur l'ensemble de chaque chromosome. La barre inférieure montre la densité des gènes (*gene density*). La présence de télomères à chaque extrémité des chromosomes est décrite par un triangle noir. Figure issue de Derelle *et al.* 2006<sup>40</sup>.

Chez *O. tauri*, ces deux chromosomes sont le chromosome 2 (*Big Outlier Chromosome* ou BOC) et le chromosome 19 (*Small Outlier Chromosome* ou SOC), qui se conservent chez *M. pusilla* tandis que l'ordre de ces chromosomes varie chez *Ostreococcus* sp. clade B (i.e. chromosome 18), *O. lucimarinus* (i.e. chromosome 18), *M. commoda* (i.e. chromosome 17) et chez *Bathycococcus prasinus* (i.e. chromosome 19)<sup>43,73</sup>.

Dans tous les génomes des Mamiellales, le BOC et le SOC sont marqués par une plus faible composition en G+C (55%), une densité des gènes et un taux d'expression inférieurs à ceux des autres chromosomes. De plus, dans le génome d'*O. tauri*, une répartition hétérogène des 417 éléments transposables (ETs) identifiés, a été observée. Environ 57% de ces éléments sont localisés sur le chromosome 2 (BOC) et 20% sur le chromosome 19 (SOC) : réunissant ainsi plus des trois quarts des ETs uniquement sur ces deux chromosomes<sup>40</sup> (**Figure 5**). Le faible taux en GC et la haute fréquence d'éléments transposables sur ces deux chromosomes semblent indiquer une absence de recombinaison pour les chromosomes *outliers*. Le BOC est proposé pour jouer le rôle de chromosome sexuel. Chez *O. tauri*, bien que la fréquence de méiose/mitose a été estimée à  $10^{-6}$ , le mécanisme de méiose reste encore obscur<sup>74</sup>. Le SOC quant à lui apparaît jouer un rôle dans la défense contre les pathogènes, notamment en raison du nombre élevé de protéines membranaires étrangères pour lesquelles ils codent<sup>40</sup>. Nous décrivons plus en détail dans la partie 3.4 de l'introduction le rôle du SOC dans le processus d'immunité antiviral contre les *Prasinovirus* chez *O. tauri*.

## 2. Les virus marins

### 2.1. Qu'est-ce qu'un virus ?

#### *Contexte historique et définitions*

Les plus anciennes études de virologie remontent au début des années 1892, où le biologiste russe Dimitri Ivanovski associe la maladie de la mosaïque du tabac à un agent infectieux de type non-bactérien capable de traverser le filtre de Chamberland. En complétant ses recherches, Martinus Beijerinck sera le premier chercheur à utiliser le terme « virus » pour décrire cet agent<sup>75</sup>. Toutefois, c'est dans les années 1930 que sera donnée une définition résiliente des virus.

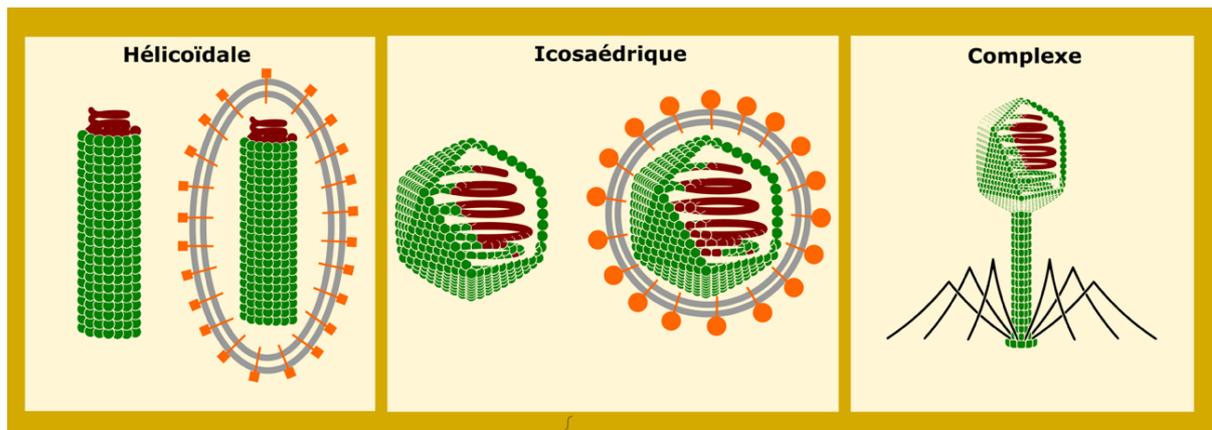
Un virus est un agent infectieux obligatoire qui nécessite une cellule vivante d'un organisme hôte pour survivre et se répliquer. Ces caractéristiques ont donc mené à une controverse qui est toujours d'actualité : les virus sont-ils vivants ? Si on tient compte qu'un virus n'est pas constitué de cellules (ni d'organelles), et qu'un virus, un virion et une particule virale sont considérés comme une et même chose, il est compréhensible que l'on puisse considérer un virus comme non-vivant. En 1961, Jacob et Wollman ont décrit que les virus pouvaient exister sous trois formes distinctes au cours de leur vie<sup>76</sup>. La première est la forme

reproductible du virus (virion) qui serait l'état « extracellulaire infectieux », comparable à une spore. La seconde est l'état inerte du virion ou phase intracellulaire qui serait la forme « vivante »<sup>77</sup>, associée à l'état « végétatif de réplication autonome » décrit par Jacob et Wollman. Sous cette forme, la particule virale devient métaboliquement active lors de la rencontre avec son hôte. En référence à cette proposition conceptuelle de ce qu'est un virus, la notion de « *virocell* » a été introduite par le professeur Patrick Forterre<sup>78-80</sup>. Ce nouveau concept rend compte de la redirection (partielle ou totale) du métabolisme cellulaire hôte avec la participation active de protéines codées par le virus vers sa propre réplication. Le dernier état proposé par les deux biologistes est l'état « proviral », où le virus serait intégré subtilement au génome de son hôte. Cette forme est souvent observée durant un cycle lysogénique que nous détaillerons plus tard.

### *Caractéristiques morphologiques*

Grâce aux avancées en microscopie électronique, la forme extracellulaire des virus a pu être déterminée. A ce jour, les virus montrent une grande diversité de tailles, de formes et de symétries mais ils partagent des éléments structurels communs. Les virions sont constitués d'une grande enveloppe protéique stable (i.e. la capsid) qui se caractérise par un assemblage d'une centaine de sous-unités identiques et répétées appelées « les capsomères ». Le rapport géométrique des capsomères conduit à deux types de symétries principales, qui déterminent la forme du virus : hélicoïdale et icosaédrique (**Figure 6**). L'hélicoïdale est une symétrie des capsomères en forme de spirale tandis que la symétrie icosaédrique est sphérique (i.e. 20 faces). Toutefois, certains virus, par exemple les bactériophages, se caractérisent par une symétrie complexe avec une tête icosaédrique et une queue de symétrie hélicoïdale (**Figure 6**).

Selon la classe de virus, cette capsid peut être enveloppée (i.e. virus enveloppé) ou non (i.e. virus « nu ») par une membrane lipoprotéique et décorée de glycoprotéines, en partie dérivée de la membrane cellulaire de son hôte (**Figure 6**). Enfin, le génome viral exploite toutes les formes naturellement possibles d'acide nucléique et peut-être constitué d'ARN ou d'ADN, simple ou double brin, circulaire ou linéaire. Le complexe capsid et acide nucléique constitue la « nucléocapsid ».



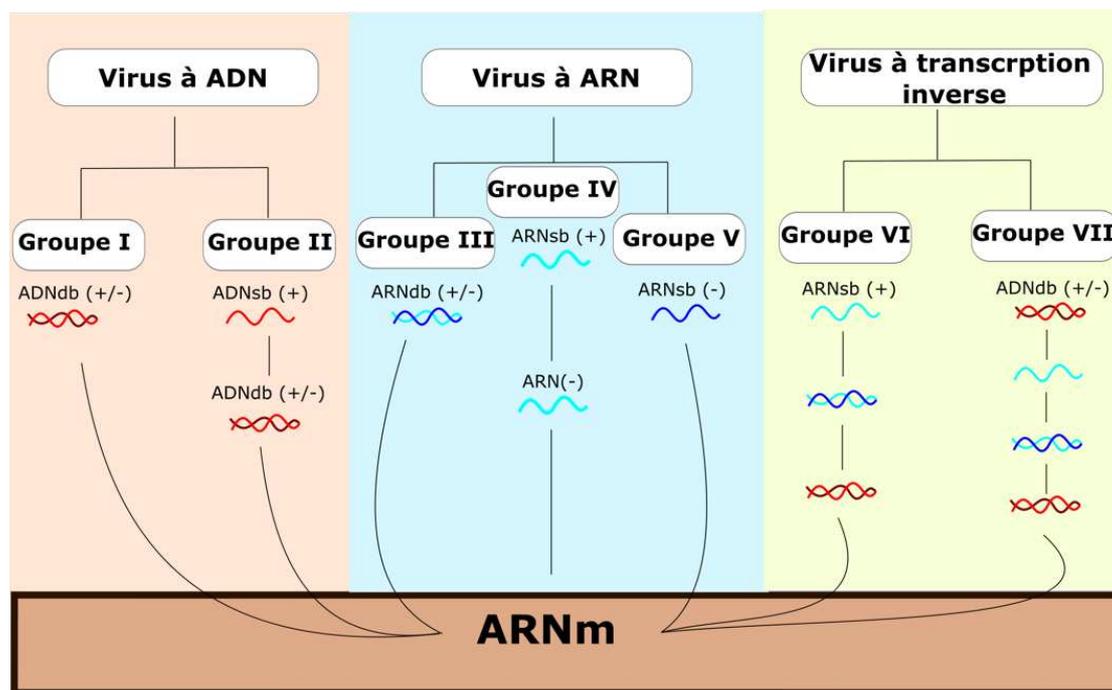
**Figure 6.** Représentation schématique de la diversité symétrique des virus. Les virus nus sont caractérisés par la présence d'une capsid (vert) qui protège les acides nucléiques (rouge). Tandis que les virus enveloppés présentent un manteau lipoprotéique (gris) avec ses glycoprotéines associées (orange), qui enveloppe la nucléocapside (capsid et acide nucléique). Les trois principales formes symétriques de capsides sont présentées comme suit : hélicoïdale, icosaédrique, et complexe.

### *Classifications taxonomiques*

Parallèlement à la classification fondée sur la caractérisation morphologique, ou selon l'histoire évolutive (i.e. la phylogénie) des virus, deux grands types de classification sont utilisés et reconnus :

La première, la classification de Baltimore a été nommée en hommage à son fondateur, le virologue David Baltimore<sup>81</sup>. Cette classification se base sur le type d'acide nucléique des virus et la manière dont ils synthétisent leur propre ARN messenger (ARNm). Sept groupes distincts sont reconnus (**Figure 7**) :

- Le groupe I avec les virus à ADN double-brin (ADNdb (+/-)).
- Le groupe II avec les virus à ADN simple-brin (ADNsb (+)).
- Le groupe III avec les virus à ARN double-brin (ARNdb (+/-)).
- Le groupe IV avec les virus à ARN simple brin-positif (ARNsb (+)).
- Le groupe V avec les virus à ARN simple-brin négatif.
- Le groupe VI avec les virus dont le génome d'ARN simple-brin positif nécessite un intermédiaire d'ADN pour se répliquer (ARNsb (+) - RT).
- Le groupe VII avec les virus ADN double-brin qui nécessitent un intermédiaire d'ARN dans son cycle de répllication.



**Figure 7.** Représentation des sept groupes de virus selon la classification de Baltimore. Adaptée de ©Microbiologynote.com.

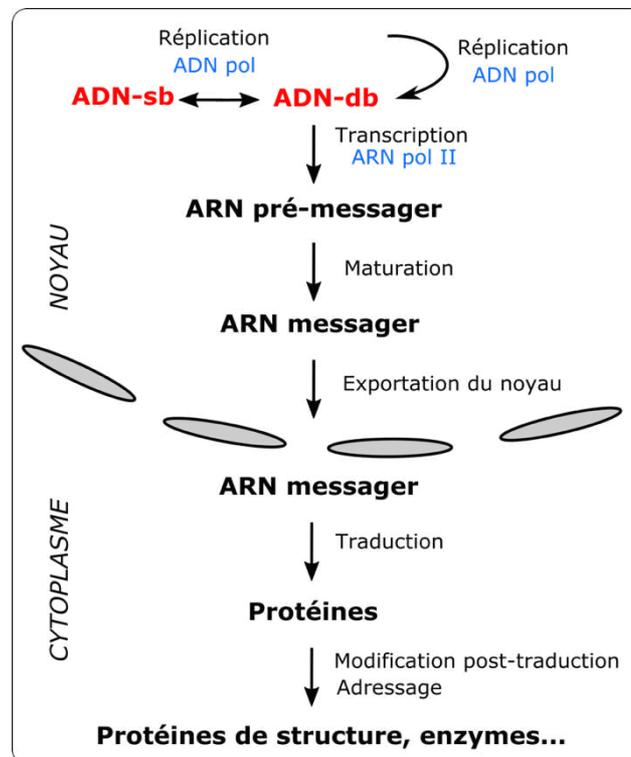
La seconde, la classification de l'*International Committee on Taxonomy of Viruses* (ICTV), se base sur des caractéristiques similaires à celle de la classification traditionnelle des organismes cellulaires (i.e. rangs taxonomiques). L'ICTV classe ainsi les virus aux rangs hiérarchiques de l'ordre, de la famille, du genre et de l'espèce et chaque taxon a un nom défini, unique, et réglementé. La taxonomie de l'ICTV tient compte de la gamme d'hôtes, du cycle de réplication, ainsi que de la structure et des propriétés des particules virales. Les virus tirent généralement leurs noms de leurs hôtes, mais aussi de l'origine d'isolement, des maladies qu'ils provoquent et de leurs caractères phénotypiques<sup>82</sup>. Le développement du séquençage à haut débit dans les échantillons environnementaux (i.e. les métagénomes) a modifié considérablement notre perception de la diversité des virus et a remis en question la manière de les reconnaître et de les classer<sup>83,84</sup>. Ceci est aussi valable pour des virus isolés qui présentent des caractères évolutifs communs mais qui n'infectent pas le même hôte. Ainsi, un nombre important de (méta)virus reste à ce jour non-classifié officiellement par l'ICTV.

Dans ce travail de thèse, nous aborderons uniquement les virus à ADN double-brin (i.e. groupe 1) qui infectent le phytoplancton eucaryote (i.e. la famille des *Phycodnaviridae*), spécifiques aux Mamiellophyceae (i.e. le genre *Prasinovirus*).

## Cycle de réplication

La plupart des virus à ADN ont un cycle dit « nucléocytoplasmique ». Ces virus se répliquent exclusivement dans le noyau, avec une haute dépendance au système réplcatif et transcriptionnel hôte. La traduction des gènes en protéines et l'assemblage des particules virales sont en revanche effectués dans le compartiment cytoplasmique de la cellule hôte (**Figure 8**). En revanche, certains virus d'eucaryotes à ADNdb codent dans leur génome tout le complexe enzymatique nécessaire pour répliquer et transcrire leurs propres gènes dans la cellule hôte. Leur cycle de réplication est alors uniquement cytoplasmique.

Pour le génome des ADNsb, une étape préalable est nécessaire pour convertir l'ADNsb en ADNdb, qui est réalisée par l'ADN-polymérase cellulaire dans le noyau. La réplication de l'ADNdb est assurée par l'ADN-polymérase, puis la transcription en ARNm par l'ARN polymérase (**Figure 8**). Nous verrons un peu plus loin que certains virus d'eucaryote à ADNdb, adoptent un schéma réplcatif encore plus complexe que les virus à ADN « classique » et présentent une faible dépendance à la cellule hôte pour la réplication et transcription de leur gènes.



**Figure 8.** Cycle de réplication lytique des virus à ADN « classique ». Adaptée de ©UCLouvain.

La réplication virale peut être assurée par deux voies principales : le cycle lytique et le cycle lysogénique. Ces deux processus reposent sur la reconnaissance, l'adsorption, et la pénétration du virus dans la cellule hôte.

Dans le cas d'une réplication lytique, le virus entre dans la cellule et détourne activement la machinerie cellulaire hôte pour se répliquer et synthétiser ses protéines structurales (**Figure 8**), transformant ainsi la cellule infectée en une « virocellule » (*virocell*). L'assemblage des particules virales est réalisé dans la cellule hôte. La libération des virions matures dans le milieu extracellulaire est gouvernée par deux processus distincts selon le type de virus. Chez les virus nus, la libération est décrite par l'éclatement et lyse de la cellule hôte, tandis qu'un processus par bourgeonnement est observé chez les virus enveloppés.

Dans le cas d'une réplication lysogénique, le génome viral est intégré sous forme de prophage (ou provirus dans le cas d'un hôte eucaryote) au génome de l'hôte. Les gènes du prophage ne sont pas transcrits, et le prophage se réplique passivement durant les phases de divisions cellulaires. Cette phase est apparentée à une phase de « dormance ». En revanche, cette dormance peut être interrompue par un processus « d'induction », qui conduit à l'excision du virus en dehors du génome et à la commutation vers un cycle lytique. Cette induction est causée spontanément dans certains cas<sup>85</sup>, ou par des stress environnementaux comme des changements de conditions abiotiques ou bien par une exposition de la cellule hôte à des agents mutationnels d'ADN<sup>86</sup>. Cette voie est principalement connue chez des bactériophages de bactérie<sup>87</sup> et d'archée<sup>88</sup>, qui sont qualifiés de virus « tempérés ».

L'infection par des virus à ADNdb lysogènes a aussi été décrite chez certaines espèces d'algues eucaryotes tels que le virus de l'algue brune filamenteuse *Ectocarpus siliculosus* (EsV-1) (*Phaeovirus*)<sup>89,90</sup>. Ce phénomène de lysogénie a aussi été reporté chez le virophage (i.e. virus à ADNdb qui infectent des virus géants à ADNdb de la famille des *Mimiviridae*) *mavirus*<sup>91</sup>. Dans ce cas, le génome du virophage est intégré au génome de l'algue marine hétérotrophe flagellée *Cafeteria roenbergensis* (récemment reclassifiée comme *Cafeteria burkhardae*) et l'induction du virus endogène survient pendant l'infection avec le virus géant *Cafeteria roenbergensis virus* (CroV). Les virophages peuvent aussi être intégrés au génome du virus géant sous forme de « provirophage »<sup>92</sup>. Hormis ces quelques cas, les virus d'algues eucaryotes connus ont un cycle exclusivement lytique.

## 2.2. Contexte historique des virus marins

Après la découverte du premier virus de la mosaïque du tabac, il aura fallu attendre presque un siècle pour observer le premier virus marin, réalisé par Spencer en 1955<sup>93</sup>. Ce n'est toutefois que 34 ans après que la virologie marine connaît un essor remarquable, en décrivant les virus marins comme les entités biologiques les plus abondantes dans les océans<sup>94</sup>. En effet, grâce à l'utilisation de la microscopie électronique, Bergh *et al.* estiment une concentration virale à plus de 10<sup>8</sup> particules par millilitre d'eau de mer. L'extension de la recherche des virus dans différents habitats marins a confirmé leur ubiquité dans l'ensemble des habitats aquatiques explorés jusqu'à aujourd'hui<sup>95,96</sup>. En plus d'être les plus abondants, les virus marins détiennent une extrême diversification de répertoire génétique, de processus de réplication, de formes et de structures morphologiques, avec une très large gamme d'hôtes.

## 2.3. Diversité des virus eucaryotes du phytoplancton

Si la plupart des virus marins connus sont des virus de bactéries (i.e. les bactériophages)<sup>97,98</sup>, les virus qui infectent le phytoplancton eucaryote sont tout aussi abondants dans les environnements aquatiques<sup>99,100</sup>. Les virus d'algues eucaryotes sont très diversifiés au niveau de leur génome (i.e. ADN ou ARN), et de leurs architectures génomiques (i.e. linéaire, circulaire, simple-brin, double-brin ou segmentés) ainsi que des hôtes qu'ils infectent<sup>101,102</sup>. Un concept écologique important concerne les virus d'algues : leur haute spécificité vis-à-vis de leurs hôtes et notamment des souches qu'ils infectent. Ainsi, la diversité des virus d'algues reflèterait celle de leurs hôtes. De nombreux taxons viraux issus de cultures isolées de microalgues, qui englobent à la fois la classification de Baltimore et la classification traditionnelle, ont été reportés.

Les *Marnaviridae* restent le groupe le plus prédominant des virus à ARN infectant les microalgues. Les *Marnaviridae* (ordre *Picornavirales*) est une large famille de virus à ARNs (+) composée de 7 genres qui infectent spécifiquement les protistes marins (incluant les microalgues). Le premier *Marnaviridae* isolé en culture est le virus HaRNA infectant l'algue rouge *Heterosigma akashiwo* (*Raphidophyceae*) formant des efflorescences toxiques<sup>103</sup>. D'autres virus du dinoflagellé *H. circularisquama* (HcRNAV)<sup>104</sup> et divers virus de diatomées qui infectent des espèces des genres *Chaetoceros*<sup>105-108</sup>, *Rhizosolenia*<sup>109</sup>, *Asterionellopsis* et *Thalassionema*<sup>110</sup> ont également été isolés au Japon. En parallèle de ces découvertes, de

nombreuses études métagénomiques ont permis de révéler la présence de ces virus dans les environnements côtiers marins et d'eau douce<sup>111-113</sup>, et très récemment dans les bassins d'algue à haut débit (*high rate algal ponds* – HRAP)<sup>114</sup>. Les virions de cette famille se caractérisent par une structure icosaédrique d'une taille située entre 22 à 35 nm de diamètre. Aucune enveloppe externe, ni de spicules n'ont été mises en évidence<sup>115</sup>. Concernant leur génome, ce sont des petits virus non-segmentés d'ARNsb (+) d'une longueur qui varie entre 8.6 à 9.6 kb. Les *Marnaviridae* codent pour 3 protéines non-structurales (hélicase à ARN, 3C-protéase et une ARN-polymérase à ARN dépendante) en extrémité 5' et 4 protéines structurales (VP1-VP4) en extrémité 3'. La particularité des génomes des *Marnaviridae* est la présence de sites internes d'entrée des ribosomes (*internal ribosome entry sites* – IRES) en extrémité 5' qui assurent l'initiation de la traduction et d'une queue poly(A(n)) en extrémité 3'<sup>115</sup>.

A ce jour, aucun virus avec un génome à ARNsb (-) n'a encore été isolé chez les algues eucaryotes marines.

Par ailleurs, le seul virus d'algue eucaryote avec un génome à ARNdb a été isolé de l'algue unicellulaire prasinophyte *Micromonas pusilla* (Mamiellophyceae)(reclassifiée comme *M. commoda* en 2017<sup>116-118</sup>. Les analyses de microscopie électronique révèlent des particules virales icosaédriques non-enveloppées d'environ 65 à 80 nm de diamètre. En raison de sa structure morphologique et génomique, *Micromonas pusilla reovirus* a été officiellement classifié au sein des *Reoviridae*, et constitue le seul taxon du genre *Mimoreovirus* par l'ICTV. Le génome linéaire et segmenté de *Mimoreovirus* a une longueur d'environ 25.5 kb et code pour 11 protéines.

Les virus à ADN incluent une incroyable diversité de virus qui se composent d'ADNsb et ADNdb. Actuellement tous les virus avec des génomes ADNsb circulaires ont été isolés uniquement à partir d'espèces de diatomées (*Bacillariophyceae*) et se regroupent au sein de la famille des *Bacilladnaviridae*<sup>105,107,108,110,119-121</sup>. Seul le virus à inclusion nucléaire « *Chaetoceros salsugineum* nuclear inclusion virus » (CsNIV) a été officiellement classé au sein des *Diatodnavirus*, par l'ICTV<sup>122</sup>. Les virions de cette famille de virus à ADNsb se caractérisent par une capsid de symétrie icosaédrique non-enveloppée d'environ 22-38 nm de diamètre, dépourvue de queue. Toutefois, des particules en forme de bâtonnet ont été localisées à proximité des particules icosaédriques au sein des noyaux<sup>107</sup>. Puisque ces réseaux en forme de tige se produisent au début du cycle d'infection, des auteurs suggèrent un rôle précurseur dans la maturation des virions hexagonaux<sup>121</sup>.

La plupart des virus à ADNdb infectant les algues eucaryotes font partie du groupe des « *Nucleocytoplasmic Large DNA Virus* » (NCLDV), officiellement reconnus dans l'embranchement des *Nucleocyotoviricota*<sup>123</sup>. Ces virus sont aussi connus sous le nom de virus « géants » et même comme girus pour la contraction de *giant viruses* en anglais<sup>124</sup>. Ce nom a été proposé en raison de leurs caractères morphologiques et génomiques. En effet, les NCLDVs regroupent l'ensemble des virus à ADNdb eucaryote dont le génome et la taille du virion dépassent 300 kb et 200 à 300 nm, respectivement. Les plus grands virus de ce groupe ont été associés aux virus d'amibes, les « Pandoravirus » (« *Pandoraviridae* ») qui ont des virions d'environ 1 µm de longueur et des tailles de génome supérieur à 2.5 Mb avec environ 2500 gènes<sup>125,126</sup>, et les Pithovirus (« *Pithoviridae* ») qui détiennent le record de taille de virion (environ 1.5 µm de longueur)<sup>127</sup>. En revanche, ces deux familles ne sont toujours pas officiellement classifiées par l'ICTV. Nous définirons les virus associés aux NCLDVs dans la section suivante (cf. partie 2.4 de l'introduction).

## 2.4. Que sont les virus géants ?

Cette section sera incluse dans la revue : « *Why so big? Genome evolution of giant viruses* », prochainement soumise au journal GBE « *Genome Biology and Evolution* ». Auteurs : Julie Thomy et Sheree Yau

### *Découvertes du premier virus géant*

La découverte des virus géants a remis en question l'utilisation de la filtration d'une porosité de 0.2 µm (i.e. filtre de Chamberland) pour étudier la diversité des virus à l'échelle mondiale. C'est en 1992 que la description morphologique du tout premier virus géant a été faite. Lors d'une épidémie de légionellose, l'eau d'une tour de climatisation d'un hôpital à Bradford est prélevée et observée par microscopie optique. Des chercheurs britanniques découvrent alors un microorganisme semblable à une bactérie Gram positive en forme de coque et le nomme Bradfordcoccus, en référence à son lieu d'isolement. Toutefois, les milieux bactériens standards ne permettent pas de le cultiver. Cela est rendue uniquement possible en co-culture avec des amibes du genre *Acanthamoeba*. L'amplification de l'ARNr 16S échoue également, laissant en suspens sa caractérisation pendant 11 années. C'est en 2003, que l'analyse de Bradfordcoccus est relancée par des chercheurs de l'Université d'Aix-Marseille II<sup>128</sup>. Cette fois-ci, l'observation par microscopie électronique de ce microorganisme – en co-culture avec l'amibe – a révélé une morphologie de type virale avec des particules matures de

400 nm de diamètre entourée d'une capsidie icosaédrique. Il est observé des fibrilles d'une longueur de 80 nm associées à la capsidie et une taille de génome comparable à celle d'une bactérie. Il est donc renommé en référence à son hôte : *Acanthamoeba polyphaga* Mimivirus (pour *microbe mimicking*). Mimivirus est donc le tout premier virus à être retenu par le filtre de Chamberland et devient le fondateur de la famille des *Mimiviridae*. Ainsi, Mimivirus devient le premier « géant » des grands virus à ADN double-brin eucaryote<sup>128,129</sup>. La découverte de ce virus a renversé la virologie moderne. En effet, depuis ce jour, les virus géants ont été définis comme de particules virales supérieures à 0.2 µm et facilement observable au microscope optique : l'isolement et la découverte des virus géants commencent alors.

### *Classification taxonomique*

C'est en 2001 que le groupe des virus géants à ADN nucléo-cytoplasmique (NCLDV) a été établi pour la première fois<sup>130</sup>. A ce jour, le nom NCLDV est utilisé comme nom informel. Pour des raisons pratiques, nous continuerons d'utiliser ce terme tout au long de ce manuscrit.

A l'origine, quatre familles composaient le groupe des virus à grand ADN infectant les eucaryotes : les *Poxviridae*, les *Asfarviridae*, les *Iridoviridae* (y compris les *Ascoviridae*) et les *Phycodnaviridae*<sup>130</sup>. La découverte de Mimivirus aura permis d'accélérer nettement la recherche de virus apparentés à la famille des *Mimiviridae*, et d'élargir considérablement la liste des NCLDVs, qui constituent maintenant un des groupes les plus diversifiés et abondants de la virosphère eucaryote<sup>131–133</sup>. Ils présentent un large éventail de morphologie, de taille de leurs virions, de caractéristiques génomiques, de gamme d'hôtes et de cycles de réplication. Malgré ces différences, des analyses comparatives génomiques et de reconstructions évolutives ont permis d'appuyer la monophylie de ce groupe<sup>130,134–136</sup> et proposent une divergence en deux supergroupes après la séparation des lignées archées et eucaryotes, qui se serait produite avant l'émergence du dernier ancêtre commun<sup>137</sup>. Toutefois, l'origine et l'évolution des NCLDVs restent un sujet encore très controversé<sup>138</sup>.

L'étude des relations évolutives entre les NCLDVs s'est le plus souvent appuyée sur un ensemble combiné ou individuel de gènes communs largement partagés entre les NCLDVs (définis comme les gènes « *cores* ») pour reconstruire des arbres phylogénétiques. En 2012, Yutin et Koonin ont défini 5 gènes strictement conservés parmi 45 génomes NCLDVs : *ADN polymérase B* (DNA polB), *Viral Late Gene Transcription Factor 3-like* (VLTf3-like), *D5-like helicase-primase* (primase), *major capsid protein* (MCP), et *A32-like packaging ATPase*

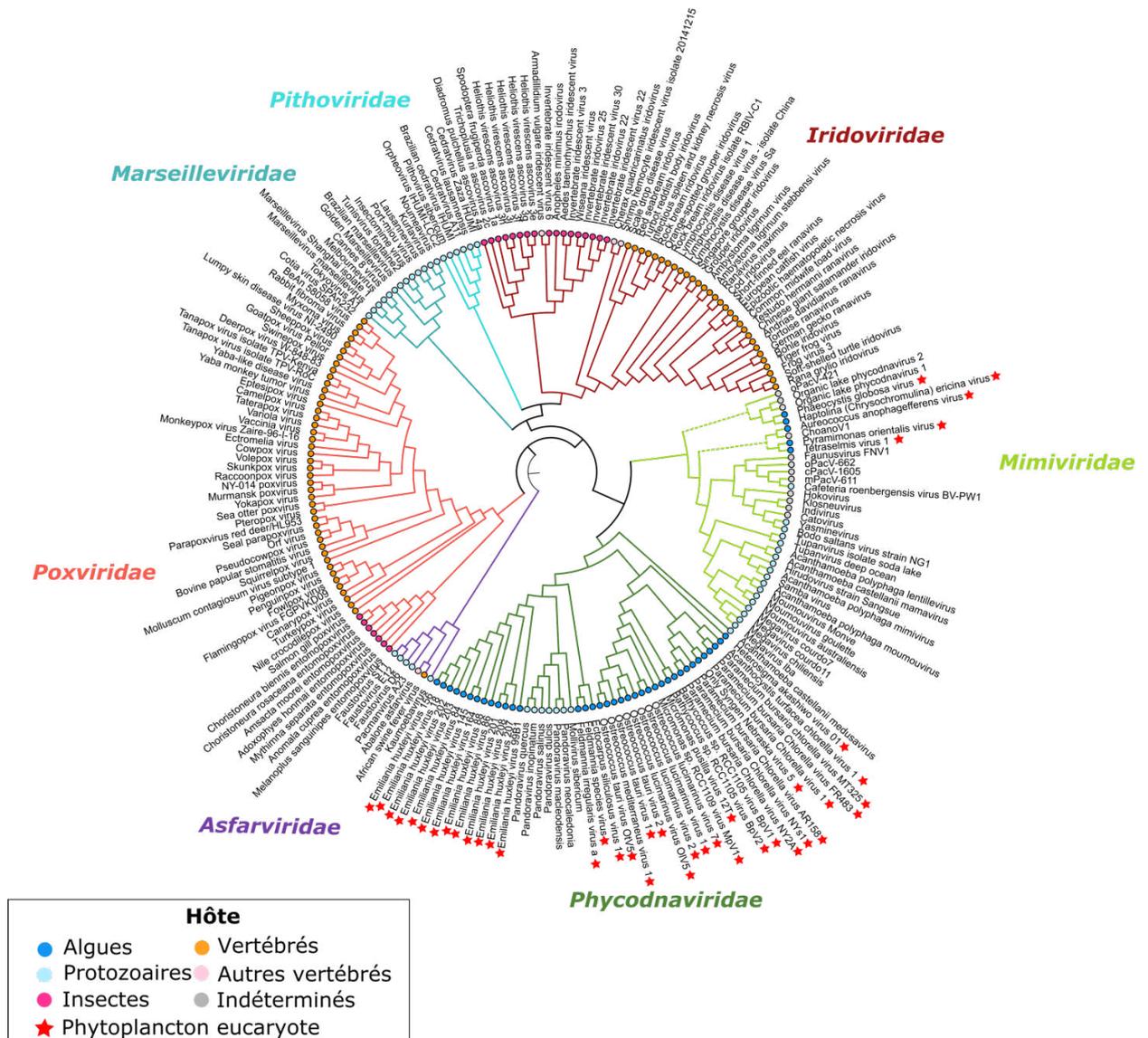
(pATPase)<sup>139</sup>. Toutefois, ces deux derniers gènes n'ont pas été retrouvés dans une récente analyse qui a recensé 73 génomes complets<sup>137</sup>. Ainsi, les NCLDV s partageraient 3 gènes strictement conservés (DNA polB, VLTF3-like et primase) et 5 gènes *cores* additionnels, présents au moins dans 92% des génomes : *Transcription Elongation Factor II-S* (TFIIS), MCP, pATPase, et les deux sous-unités de *DNA-dependent RNA Polymerase* (RNAP) (RNAP-a et -b)<sup>137</sup>.

L'ICTV a donc récemment classé les familles virales composantes des NCLDV s, dans un nouveau phylum nommé *Nucleocytoviricota*<sup>123</sup>. Les rangs taxonomiques reconnus dans ce phylum comprennent actuellement 2 classes, 5 ordres, 7 familles et 41 genres<sup>123,140,141</sup>. Ce supergroupe englobe les familles suivantes : les *Poxviridae*, *Ascoviridae*, *Iridoviridae*, *Asfarviridae*, *Marseilleviridae*, *Mimiviridae* et *Phycodnaviridae* ainsi que plusieurs lignées de virus non classés à ce jour. Les « Cedratvirus », « Pithovirus », « Orpheovirus » sont phylogénétiquement associés à la famille proposée des « Pithoviridae », et sont une branche sœur des « Mininucleoviridae ». Les « Mollivirus » et « Pandoravirus » se regroupent au sein des « Pandoraviridae », et les « Pacmanvirus » et « Faustovirus » au sein des *Asfarviridae*, seul les « Medusavirus » restent un groupe non attribué en raison de leur répertoire génétique extrêmement proche des gènes eucaryotes<sup>123</sup> (**Figure 9**).

Historiquement, l'isolement après infection de culture de cellules hôtes a longtemps été la principale méthode pour étudier la diversité des virus eucaryotes. Les contraintes d'isolement de nouvelles souches hôtes et leur maintien en culture par des méthodes classiques constituent encore à ce jour un biais à la description et caractérisation des virus. Les approches de métagénomique ou de génomique *single-cell* ont depuis facilité la découverte de nouveaux membres des NCLDV s à partir d'échantillons environnementaux sans culture<sup>99,145-149</sup>. Ces approches ont permis de mettre en évidence un immense réservoir de virus eucaryotes avec une diversité largement sous-estimé jusque-là.

La classification des métaviromes au sein des *Nucleocytoviricota* reste un challenge. Très récemment, une reclassification taxonomique basée sur l'utilisation de gènes ancestraux conservés a permis de proposer une nouvelle affiliation taxonomique plus juste des NCLDV s qui inclut les nouveaux virus issus d'échantillons environnementaux (MAG - *metagenome assembled genomes*) et les génomes de cultures isolées<sup>141</sup>. A travers cette analyse, les auteurs proposent une nouvelle répartition des *Nucleocytoviricota* en 6 ordres, 32 familles et 344 genres. La nomenclature des 6 familles reconnues par l'ICTV a été conservée (seuls les

*Ascoviridae* ont été inclus avec les *Iridoviridae*). Sur les 26 nouvelles familles, les auteurs en ont soumis 6 nouvelles nommées d'après les espèces hôtes qu'ils infectent : « *Prasinoviridae* » qui comprend les prasinovirus, « *Pandoraviridae* » qui comprend les pandoravirus et *Mollivirus sibericum*, « *Coccolithoviridae* » qui comprend les coccolithovirus, « *Pithoviridae* », qui comprend les pithovirus, cedratvirus, et orpheovirus, « *Mesomimiviridae* » qui comprend les virus d'algues précédemment définis comme « *Mimiviridae*-étendu », et les « *Mininucleoviridae* »<sup>141</sup>.



**Figure 9.** Reconstruction phylogénétique des 7 familles de NCLDVs basé sur la concaténation de 5 gènes *cores* : ADN polymérase B, D5-like hélicase primase, facteur de transcription II, VLTF3-like protéine, et l'ATPase. L'arbre est issu de Sun et Ku.<sup>142</sup> et édité sur iTOL (<http://itol.embl.de>)<sup>143,144</sup>. Les branches en pointillé des *Mimiviridae* représentent la sous-famille des *Mininucleoviridae*. L'arbre a été

enraciné de façon arbitraire avec les *Asfarviridae*. Les « Pandoraviridae » (y compris les « Mollivirus ») se regroupent au sein des *Phycodnaviridae*.

Bien que la métagénomique soit une approche qui a considérablement accéléré les connaissances sur la diversité des virus géants, il manque toutefois des connaissances sur leurs hôtes. Cette source d'information est indispensable puisque la réplication du virus dépend de son organisme hôte pour se répliquer<sup>150</sup>. L'accès aux informations génomiques et taxonomiques de l'hôte est une condition préalable notamment dans la compréhension des interactions hôtes–virus et du cycle de vie, mais aussi la mise en évidence de leur coévolution, notamment les interactions génétiques – transferts latéraux de gènes – entre les virus et leurs hôtes.

### *Diversité de la gamme d'hôtes*

A l'échelle de la famille, les NCLDV s issus de cultures isolées présentent une gamme étendue d'hôtes associés à divers embranchements eucaryotes. Les *Poxviridae*, *Iridoviridae/Ascoviridae* et *Asfarviridae* montrent une gamme d'hôte restreinte en infectant uniquement les Métazoaires (Opisthokonta) et plus spécifiquement les arthropodes et les vertébrés. Les *Marseilleviridae* et la plupart des virus non classifiés (« Pacmanvirus », « Faustovirus », « Cedratvirus », « Pithovirus », « Orpheovirus », « Mollivirus » et les « Pandoravirus ») ont une gamme d'hôte restreinte aux amibes. C'est aussi le cas pour les « Mininucleoviridae », qui regroupent trois virus spécifiques de crustacés<sup>151</sup>, contrairement aux *Mimiviridae* et aux *Phycodnaviridae*, qui sont associés à six et quatre supergroupes eucaryotes, respectivement. La gamme d'hôtes des *Mimiviridae* est plus complexe et s'étend des amibes jusqu'aux eucaryotes unicellulaires hétérotrophes aux microalgues. En revanche, les *Phycodnaviridae* sont plus restreints aux groupes des Archeplastida, Haptista ainsi que quelques SARs (**Figure 9**). Les virus associés aux *Phycodnaviridae* et *Mimiviridae* seront décrits plus en détail dans la partie 2.5 de l'introduction.

### *Cycle de réplication*

Les NCLDV s constituent un ensemble de virus bien plus complexes que les virus « classiques ». Contrairement à la plupart des virus à ADN, dont le cycle de réplication est exclusivement nucléaire, les NCLDV s peuvent accomplir leur cycle de vie complet dans le cytoplasme au sein des structures appelées « usines à virions » qui sont formées durant l'infection<sup>152,153</sup>.

Selon le degré d'autonomie du virus pour la transcription de ces propres gènes, le cycle de réplication peut se dérouler entièrement dans le cytoplasme ou bien être débuté dans le noyau et se terminer dans le cytoplasme de la cellule hôte eucaryote<sup>134</sup>. Cette relative indépendance des NCLDV's vis-à-vis de la cellule hôte, serait cohérente avec les gènes conservés entre les virus, qui coordonnent la plupart des processus essentiels à la reproduction virale.

Cette machinerie de réplication/transcription complète au sein des génomes et la dépendance exclusive au cytoplasme pour la réplication du virus ont été reportés chez les *Mimiviridae*<sup>154</sup>, les *Poxviridae*<sup>155</sup> et la plupart des *Marseilleviridae*<sup>156,157</sup>. Toutefois, les *Pandoraviridae*<sup>125,126</sup> (incluant le Mollivirus)<sup>158</sup> et les *Phycodnaviridae*<sup>159-161</sup> (à l'exception des *Coccolithovirus*)<sup>162</sup> présentent une machinerie de réplication/transcription partielle, notamment avec l'absence de la RNAP au sein de leur génome. Ainsi, les *Pandoraviridae* et les *Phycodnaviridae* sont caractérisés par une réplication nucléocytoplasmique, avec une dépendance vis-à-vis de l'ARN polymérase nucléaire de l'hôte.

Si la présence de la machinerie de transcription (notamment la RNAP) est un déterminant clé dans la compréhension du cycle réplcatif des virus géants, plusieurs exemples font exception à la règle. Ceci suggère une évolution divergente et une acquisition de fonctions additionnelles uniques à certains virus géants.

Chez les *Asfarividae*<sup>163</sup> et *Iridovidae/Ascoviridae*<sup>164</sup>, une phase nucléaire obligatoire a aussi été observée, bien qu'ils codent pour leur propre machinerie de transcription et de réplication.

*A contrario*, un cycle de réplication distinct a été mis en évidence par microscopie électronique à transmission (MET) et à fluorescence chez le Nouméavirus (*Marseilleviridae*)<sup>165</sup>. Le Nouméavirus est capable de recruter de façon transitoire les protéines nucléaires hôtes, nécessaires à la transcription des premiers transcrits viraux, vers l'usine virion située dans le cytoplasme. Ce processus conduit à une déformation du noyau de la cellule hôte et à une disparition du nucléole de manière transitoire et réversible<sup>165</sup>, contrairement à ce qui a été décrit pour un virus nucléocytoplasmique d'amibe « classique », comme les Pandoravirus<sup>125</sup> et le Mollivirus<sup>158</sup>.

Les Médusavirus sont caractérisés par une réplication vraisemblablement nucléaire (i.e. aucune RNAP et peu de gènes pour la machinerie de réplication), et présentent des

caractéristiques uniques à ce groupe<sup>166</sup>. La pénétration du génome viral au sein du noyau hôte pour initier la réplication ne conduit pas à la déformation de la membrane nucléaire ou une dégradation définitive du noyau de l'hôte, comme c'est le cas pour les autres virus d'amibe. De plus, la synthèse de l'ADN viral apparaît complète au sein du noyau hôte et aucune usine à virions cytoplasmiques n'a été observée par MET<sup>166</sup>. Ces données suggèrent un refaçonnement du noyau hôte par le Médusavirus au cours de son cycle de vie complet au sein de l'amibe<sup>166,167</sup>.

Ainsi, les NCLDV ont des degrés variables de dépendances aux fonctions nucléaires de l'hôte propres à chaque famille.

## 2.5. Diversité des virus géants du phytoplancton eucaryote

Sur la base d'analyse génomique, des études suggèrent la monophylie des *Mimiviridae* et *Phycodnaviridae*<sup>168-170</sup>. L'origine des virus géants reste encore très controversée entre les défenseurs de l'hypothèse de la « réduction du génome » selon laquelle les virus auraient dérivé d'un organisme cellulaire par une perte progressive de ses fonctions jusqu'à un véritable virus<sup>171</sup> et ceux qui soutiennent l'hypothèse de « l'expansion du génome » selon laquelle les virus géants auraient évolué à partir d'un ancêtre viral simple et se seraient complexifiés par accumulations et duplications de gènes<sup>172</sup>. Jonathan Filée propose une combinaison de ces deux hypothèses à travers un modèle évolutif dit « neutre » dans lequel les virus géants auraient évolué *via* un processus complexe « d'accordéon génomique », expliqué par des événements successifs de duplications et de pertes de gènes<sup>173,174</sup>. Ce schéma a été observé à plus petite échelle pour deux familles proches: les *Mimiviridae* et *Phycodnaviridae*<sup>174</sup>.

### *Les Mimiviridae*

Si de nombreuses études métagénomiques ont permis de révéler une immense richesse de diversité des NCLDV, elles ont également mis en évidence leur ubiquité et leur abondance dans les environnements aquatiques<sup>100</sup>. Après les virus de bactéries, les *Mimiviridae* constituent la plus grande fraction des NCLDV dans les océans<sup>133</sup> et dépassent de loin celle des bactéries et des archées<sup>175,176</sup>. De récentes études ont également montré que les membres de cette famille sont capables d'infecter une large gamme de communauté eucaryote de différentes tailles<sup>131</sup>.

A ce jour, les *Mimiviridae* représentent un vaste groupe complexe de virus avec de grandes différences en termes de contenu génétique, de morphologie, de taille de particule, et

de gamme d'hôtes. Malgré ces disparités, les *Mimiviridae* forment leur propre groupe, à part des *Phycodnaviridae*<sup>142</sup>. Jusqu'à très récemment, on pensait que l'ensemble des virus d'algues étaient regroupés dans une même famille, nommée d'après leurs hôtes, les *Phycodnaviridae*. Toutefois, l'isolement et la caractérisation d'un sous-ensemble de virus d'algues hétérotrophes et autotrophes a suggéré une affinité phylogénétique plus proches avec les *Mimiviridae* (**Figure 9**). Ces découvertes ont rendu confuse la classification nominative des virus géants. Ainsi, basés sur leurs relations phylogénétiques, Gallot-Lavallée et al. ont récemment proposé de diviser les *Mimiviridae* en deux sous-familles distinctes<sup>177</sup>.

Cette première sous-famille nommée les « Megavirinae » inclue l'ensemble des virus spécifiques des amibes, avec un large génome (~1250 kb), répartis en trois clades A, B et C. Les virus d'algues unicellulaires avec un plus petit génome (177-668 kb) se regroupent eux dans les « Mesomimivirinae ». Outre les caractères génomiques uniques aux Mesomimivirinae, les Megavirinae se distinguent des Mesomimivirinae avec une morphologie très similaire de Mimivirus – une capsidie icosaédrique avec un *vertex* unique en forme d'étoile de mer, enveloppée de fibrilles<sup>178-180</sup> – à l'exception des Tupanvirus qui présentent une grande queue cylindrique (550 nm de long et 450 nm de diamètre) à la base de la capsidie fibrillaire<sup>181</sup>.

En revanche, les Mesomimivirinae se caractérisent par une plus petite capsidie icosaédrique (entre 100 et 400 nm) et sont dépourvues de fibrilles, comparables aux *Phycodnaviridae* (**Tableau 1**). Les Mesomimivirinae regroupent les virus infectant des haptophytes efflorescentes telles que *Phaeocystis pouchetii* (PpV), *Phaeocystis globosa*, *Chrysochromulina ericina* (HeV), *Prymnesium parvum* (PpDVAV) et le straménopyle *Aureococcus anophagefferens* (AaV). Ce groupe inclut également l'haptophyte *Prymnesium kappa* (PkV), et les deux chlorophytes *Pyramimonas orientalis* (PoV) et *Tetraselmis* sp (TetV) (**Tableau 1**). Par ailleurs, des récentes études métagénomiques démontrent la possibilité d'un nouveau clade de virus de chrysophytes phylogénétiquement relié au *Mimiviridae* ou aux *Phycodnaviridae*<sup>133</sup>.

### *Les Phycodnaviridae*

Peu de temps après avoir caractérisé les virus comme les entités les plus abondantes dans les environnements marins en 1989<sup>94</sup>, de nombreuses recherches se sont tournées sur des approches moléculaires pour étudier la diversité des virus d'algues eucaryotes<sup>182</sup>. L'amplification par PCR de l'ADN polymérase (*polB*) des virus d'algues a mis en évidence une

grande diversité au sein des *Phycodnaviridae*<sup>183,184</sup>. Ces études ont également démontré leur ubiquité dans les océans<sup>185</sup> et milieux d'eau douce<sup>186–189</sup>. Ces observations ont été renforcées avec les premières études métagénomiques. Ainsi, on considère que les *Phycodnaviridae* constituent après les *Mimiviridae*, le groupe le plus abondant et le plus répandu des NCLDV dans les océans, prévalents principalement dans les régions côtières<sup>100,133,175</sup>.

La famille des *Phycodnaviridae* a fourni la première classification officielle des virus d'algues unicellulaires eucaryotes à ADNdb. A ce jour, les *Phycodnaviridae* constituent un large groupe polyphylétique<sup>190</sup> qui rassemble six genres – avec un total de 33 espèces reconnues par l'ICTV– nommés suivant leurs hôtes: *Chlorovirus*, *Prasinovirus*, *Prymnesiovirus*, *Phaeovirus*, *Coccolithovirus*, et *Raphidovirus* (**Tableau 1**). Les *Chlorovirus* sont le genre qui regroupe le plus d'espèces classifiées (i.e. 19 espèces), suivis des *Phaeovirus* avec 9 espèces, puis seule une espèce a été officiellement reconnue par l'ICTV chez les *Prymnesiovirus*, les *Coccolithovirus* et les *Raphidovirus*.

Bien que peu d'espèces soient formellement classées au sein des *Phycodnaviridae*, cela ne fait aucun doute que de nombreux autres virus infectant les microalgues pourront ultérieurement rejoindre cette famille. Alors que la classification taxonomique semble être claire pour certains virus non classifiés, d'autres présentent des histoires évolutives ou des schémas de classification plus complexes ou inconnus. C'est le cas pour les virus isolés des algues prymnesiophytes ou haptophytes (*Phaeocystis globosa* virus (PgV), *Chrysochromulina ericina* virus (CeV), *Prymnesium kappa* Virus (PkV), et *Aureococcus anophagefferens* virus (AaV), qui ne doivent pas être considérés comme *Prymnesiovirus* en raison de leur divergence phylogénétique. En effet, Gallot-Lavallée et al. démontrent que CeV, PgV et AaV partagent des caractères génétiques communs qui les classent dans le groupe des *Mimiviridae* au sein de la sous-famille des Mesomimivirinae<sup>177</sup>(**Tableau 1**).

Les *Phycodnaviridae* présentent une morphologie structurale similaire, cohérente avec une origine évolutive commune. Ils se caractérisent par une capsid hexagonale de symétrie icosaédrique d'une taille d'environ 120–200 nm de diamètre<sup>191</sup>. Ces analyses se basent sur l'observation des particules virales par MET au sein des cellules infectées, et n'ont pas fait l'objet d'études structurales beaucoup plus détaillées – comme c'est le cas pour les virus d'amibe<sup>192,193</sup>. Toutefois, il est décrit que la capsid des phycodnavirus est assemblée à partir de sous-structures capsidiques organisées de 20 trisymétrons et de 12 pentasymétrons<sup>194</sup>. Très récemment, une analyse de reconstruction structurale par cryo-microscopie a aussi mis en

**Tableau 1.** Caractéristiques morphologiques et génomiques des virus représentatifs de culture isolée associée au phytoplancton eucaryote. Ce tableau rend compte de la classification officielle de l'ICTV rapporté en 2020. <sup>1</sup> mesurée dans des cellules infectées.

	Genre	Nom du virus	Taille virus <sup>1</sup>	Taille génome	Espèce hôte	Références
Mimiviridae (« Mesomimivirinae »)	Non-classifié	<i>Chrysochromulina ericina Virus</i> (CeV)	160 nm	510 kb	<i>Haptolina ericina</i> (Haptophyta)	199
	Non-classifié	<i>Prymnesium kappa Virus</i> (PkV)	160–400 nm	0.507–1.4 Mb	<i>Prymnesium kappa</i> (Haptophyta)	200,201
	Non-classifié	<i>Aureococcus anophagefferens Virus</i> (AaV)	140 nm	~371 kb	<i>Aureococcus anophagefferens</i> (Ochrophyta)	169,202
	Non-classifié	<i>Phaeocystis globosa Virus</i> (PgV)	106–153 nm	177–466 kb	<i>Phaeocystis globosa</i> (Haptophyta)	168,203
	Non-classifié	<i>Pyramimonas orientalis Virus</i> (PoV)	220 × 180 nm	560 kb	<i>Pyramimonas orientalis</i> (chlorophyta)	199
	Non-classifié	<i>Tetraselmis Virus</i> (TetV)	226 nm	668 kb	<i>Tetraselmis sp.</i> (Chlorophyta)	204
Phycodnaviridae	<i>Raphidovirus</i>	<i>Heterosigma akashiwo Virus</i> (HaV)	202 nm	~294 kb	<i>Heterosigma akashiwo</i> (Ochrophyta)	205,206
	<i>Coccolithovirus</i>	<i>Emiliana Huxleyi Virus</i> (EhV)	120–200 nm	~400 kb	<i>Emiliana Huxleyi</i> (Haptophyta)	207
	<i>Chlorovirus</i>	<i>Paramecium Bursaria Chlorella Virus 1</i> (PBCV-1)	150–190 nm	~331 kb	<i>Chlorella sp.</i> (Chlorophyta)	208–210
	<i>Phaeovirus</i>	<i>Ectocarpus siliculosus Virus</i> (EsV)	130–150 nm	~335 kb	<i>Ectocarpus siliculosus</i> (Ochrophyta)	89,211,212
	<i>Prasinovirus</i>	<i>Micromonas pusilla Virus</i> (MpV)	130–135 nm	~184 kb	<i>Micromonas pusilla</i> (Chlorophyta)	213,214
	<i>Prymnesiovirus</i>	<i>Chrysochromulina brevifilum Virus</i> (CbV)	145–170 nm	/	<i>Chrysochromulina brevifilum</i> (Haptophyta)	215

évidence la spectaculaire organisation architecturale de la capsid du chlorovirus, *Paramecium bursaria chlorella virus 1* (PBCV-1), laquelle comptabilise 5040 copies de la protéine de capsid majeure, 60 copies de la protéine penton (capsomère pentamériques) et 1800 copies de protéines de capsid mineures de 13 types différents<sup>195</sup>. Chez tous les *Phycodnaviridae*, la capsid renferme une bicouche lipidique, qui elle-même séquestre et protège le matériel génétique (i.e. ADNdb d'une taille située entre 184 et 400 kb) et les protéines virales associées. Si la structure de la capsid apparaît vraisemblablement similaire entre les virus d'algues eucaryotes, il n'est pas étonnant de voir apparaître des caractéristiques spécifiques à leur

surface. En effet, au sein de cette famille, les *Coccolithovirus* sont pourvus d'une membrane lipidique externe vraisemblablement originaire de sa cellule hôte<sup>196</sup>. Chez les *Chlorovirus*, PBCV-1 présente une structure en forme de pointe (*spike*) située sur un *vertex* unique de la capsidie icosaédrique en symétrie rotationnelle *5-fold*<sup>197,198</sup>. Par analogie, les auteurs font références à une stratégie d'infection décrite chez les virus de bactéries (i.e. les bactériophages) (Cherrier et al., 2009). De telles différences reflètent la diversité des virus au sein d'une même famille. Ces structures propres à chaque virus seraient donc utilisées comme des stratégies particulières d'entrée et de sortie, exploitées par les *Phycodnaviridae*, reflétant ainsi la nature de leurs hôtes.

## 2.6. Les virus des *Mamiellophyceae* : Le genre *Prasinovirus*

### *Distribution géographique*

Des études moléculaires et métagénomiques, basées sur l'analyse de la diversité des NCLDV, ont démontré la prévalence du genre *Prasinovirus* dans les écosystèmes marins et d'eau douce ainsi que leur grande richesse phylotypique<sup>133,186,187,216,217</sup>. Au sein des NCLDV, les *Prasinovirus* constituent un des genres les plus abondants dans les océans, et sont probablement plus nombreux que les organismes eucaryotes dans les zones euphotiques et dans les environnements limités en nutriments<sup>100,133,218</sup>. Cette répartition ubiquitaire des *Prasinovirus* reflète celle de leurs hôtes dans la nature. D'ailleurs, plusieurs études ont permis de mettre en évidence une certaine synchronicité de l'abondance des *Prasinovirus* avec la dynamique temporelle de leurs hôtes, suggérant que les virus pourraient contrôler et réguler les populations des algues phytoplanctoniques eucaryotes pendant les périodes estivales<sup>216,219,220</sup>. Notamment, cette régulation se traduit par une mortalité des espèces hôtes, parfois observable à l'œil nu lors des terminaisons d'efflorescence d'algues<sup>65</sup>.

Les *Prasinovirus* décrits jusqu'à présent forment un groupe monophylétique constitué de virus infectant les trois genres représentatifs des Mamiellophyceae : *Micromonas*, *Bathycoccus* et *Ostreococcus*. A ce jour, seuls *Micromonas pusilla virus SP1* (MpV-SP1) et *Ostreococcus tauri virus 5* (OtV5) ont été officiellement classifiés au sein du genre *Prasinovirus* par l'ICTV. Toutefois, l'échantillonnage d'eau de mer a permis d'isoler à partir de ces trois genres hôtes un nombre accru de virus phylogénétiquement et morphologiquement relié à MpV-SP1 et OtV5 (**Tableau 2**). Il est donc clair que tous les virus présentant une histoire

évolutive commune, spécifiques des algues prasinophytes, devraient être associés au sein du même genre.

**Tableau 2.** Liste des principaux prasinovirus isolés et de leurs hôtes associés.

Genre hôte	Espèce hôte	Clade hôte	Génome viral séquencé	Isolat de prasinovirus	Site d'isolement	Références
<i>Micromonas</i>	<i>commoda</i>	A.ABC.12	MpV-12T	31	Océan Atlantique Nord, Mer du Nord	219-221
	<i>bravo</i>	B.E.3	-	7	Mer Méditerranée et Golf de Naples	220,222
	<i>pusilla</i>	C.D.5	MpV-SP1, MpV-PL1	35	Océan Pacifique Nord, Océan Atlantique Nord, Mer Méd., Mer du Nord	219,220,223-225
	<i>polaris</i>	Ea	-	4	Mers de Barents et du Groenland	118
	sp.	Inconnu	MpV1	16	Mer Méd. et Océan Atlantique Nord	214,220,224,226
<b>Non-classifié</b>	Inconnu	Inconnu	-	11	Océan Atlantique Nord	224
<i>Ostreococcus</i>	<i>lucimarinus</i>	A	OIV1, OIV2, OIV3, OIV4, OIV5, OIV6, OIV7	27	Mer Méd., Nord, Océan Pacifique Sud, Océan Atlantique Nord, Manche	217,227
	sp.	B	OtV2	1	Océan Atlantique Nord	228
	<i>tauri</i>	C	OtV5, OtV6, OtV1	54	Océan Atlantique, lagune méditerranéenne (France)	217,226,229-232
	<i>mediterraneus</i>	D	OmV1, OmV2	7	Lagune méditerranéenne (France)	71,226,227
	sp.	E	-	-	-	-
<i>Bathycoccus</i>	<i>prasinos</i>	BI	BpV1 et BpV2	9	Mer. Méd, Océan Atlantique Nord	214,224,226
	<i>calidus</i>	BII	BII-V1, BII-V2, BII-V3	3	Océan Atlantique Nord	59

Le premier *Prasinovirus* a été découvert par Mayer et Taylor en 1975 en utilisant comme hôte l'algue *Micromonas pusilla*, durant des observations de MET<sup>213</sup>. Puisque ce virus était capable d'infecter seulement cette espèce, les chercheurs le nommèrent *Micromonas pusilla Virus* (MpV). Cette découverte a permis à des collaborateurs d'utiliser cette souche comme hôte pour isoler de multiples virus spécifiques au genre *Micromonas* répartis dans la plupart des écosystèmes marins y compris arctiques<sup>118,219,220,222,226,233,234</sup> (**Tableau 2**).

Dans les régions côtières, les virus de *Micromonas* dominent la plupart des communautés de *Prasinovirus*<sup>235</sup>. Au vu de leur ubiquité dans les écosystèmes marins, il a été possible d'isoler au moins un virus représentatif de chaque clade – décrit comme des espèces – de *Micromonas* (i.e. *M. commoda*, *M. bravo*, *M. pusilla*, *M. polaris* et *Micromonas* sp. RCC11109) (**Tableau 2**).

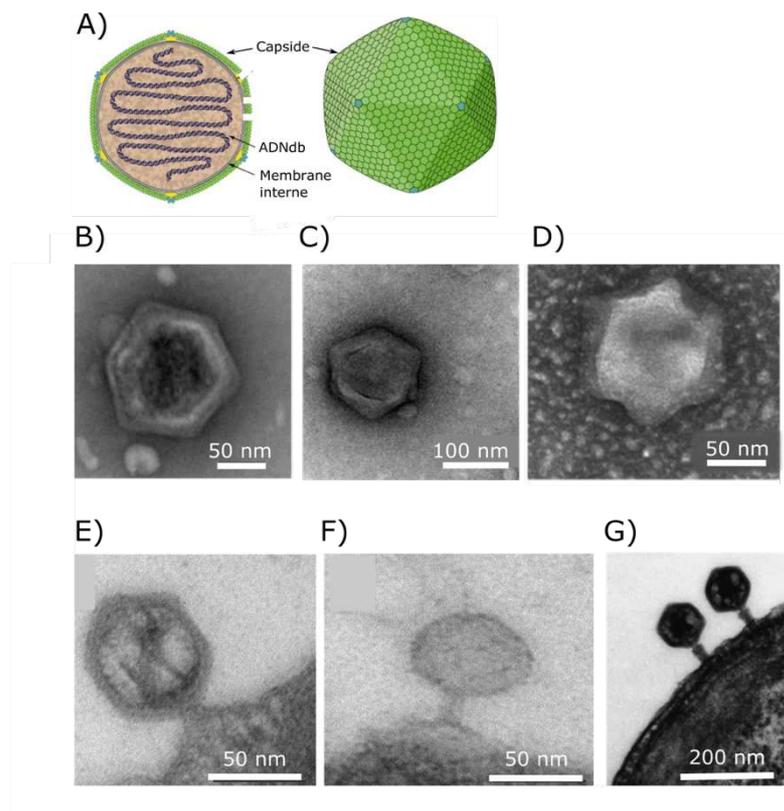
Si un déclin d'efflorescence d'*Ostreococcus* a été observée pour la première fois en 2003 sur la côte Est des États-Unis<sup>65</sup>, il aura fallu attendre 5 ans pour isoler et caractériser le premier virus infectant l'espèce *O. tauri* (i.e. OtV5) dans l'étang de Thau en France (Nord-Ouest de la Mer Méditerranée)<sup>229</sup>, où son hôte est originaire<sup>36</sup>. Comme pour les virus de *Micromonas*, les virus d'*Ostreococcus* sont ubiquitaires dans les régions côtières (**Tableau 2**), notamment les virus d'*O. tauri* (OtV) identifiés principalement dans les lagunes Méditerranéennes<sup>216</sup> où la concentration en nutriments est la plus forte<sup>235</sup>, et deux virus de *O. mediterraneus* (OmV) sur la côte en Mer Méditerranée<sup>71,214</sup>. À l'inverse, les virus d'*O. lucimarinus* (OlV) se répartissent principalement dans les océans géographiquement distincts, avec seulement un virus identifié en Mer Méditerranéennes<sup>214,217,227</sup>. Enfin, l'isolement du seul virus spécifique de l'écotype adapté à faible luminosité (*Ostreococcus* sp. clade B) vient de la côte Ouest de l'Océan Atlantique (i.e. dans la Manche). À ce jour, aucun virus du clade E n'a encore été identifié, en raison de l'absence de souche cultivée en laboratoire<sup>31</sup>.

Les virus du genre *Bathycoccus* (BV) restent encore non-classifiés officiellement par l'ICTV au sein des *Prasinovirus*. Le tout premier isolement et la caractérisation de deux virus spécifiques de l'espèce *Bathycoccus prasinos*, (i.e. clade BI : BpV1 et BpV2) a été faite dans la baie de Banyuls-sur-Mer (i.e. Mer Méditerranée) en France<sup>214</sup> (**Tableau 2**). Ces analyses phylogénétiques montrent une monophylie de BpV1 et BpV2 et les placent dans une position basale par rapport aux virus de *Micromonas* et *Ostreococcus*<sup>214</sup>. En raison des contraintes liées au maintien des cultures de *Bathycoccus*, contrairement à d'autres Mamiellophyceae, il aura fallu attendre plus de 10 ans pour isoler trois autres virus spécifiques de l'espèce *B. calidus* (i.e. clade BII : BII-V1, BII-V2 et BII-V3) dans l'Océan Atlantique Nord au large des Bermudes<sup>59</sup>. Par opposition aux virus de *B. prasinos* qui sont des virus adaptés aux systèmes mésotrophes et aux régions côtières dans des eaux à plus faible température, les virus de *B. calidus* sont prévalents dans des écosystèmes avec des salinités et des températures plus élevées<sup>59</sup> (**Tableau 2**). Bachy *et al.* ont aussi décrit la distribution et l'abondance des hôtes et de leur virus par des données métagénomiques<sup>59</sup>. Comme pour les virus de *Micromonas* et *Ostreococcus*, les virus

de *Bathycoccus* d'un même clade ont été retrouvés de manière concomitante avec leurs hôtes et démontrent clairement la grande spécificité intra-écotype et thermo-adapté des virus de *Bathycoccus*. L'ensemble de ces études rapportent une monophylie des virus de *Bathycoccus* au sein d'un même clade qui se ramifie vers les virus de *Micromonas* et *Ostreococcus*, formant un grand groupe de virus infectant les Mamiellophyceae<sup>59</sup>.

### Caractéristiques morphologiques

Il est assez frappant de constater que les Mamiellales, qui constituent avec *Ostreococcus* parmi les plus petits organismes eucaryotes connus, puissent être infectés par un virus « géant » seulement 10 fois plus petit. Avec une taille située entre 120 et 190 nm (dépendant de la méthode d'observation par MET utilisée – coupe ou coloration négative) (**Figure 10**), les *Prasinovirus* se caractérisent par une capsidie à symétrie icosaédrique et l'absence d'une enveloppe externe (**Figure 10**).



**Figure 10.** Morphologie des *Prasinovirus*. (A) Structure des prasinovirus à symétrie icosaédrique, image adaptée de ©ViralZone (<https://viralzone.expasy.org/>). Observation des virus (B) d'*Ostreococcus*, (C) de *Micromonas*, et (D) de *Bathycoccus* par la méthode de coloration négative, les images sont tirées de Derelle *et al.* 2014<sup>227</sup>, Weynberg *et al.* 2017<sup>236</sup> et Bachy *et al.* 2021<sup>59</sup>, respectivement. Adsorption des particules virales à la cellule hôte au cours des premières étapes

d'infection (E-G). (E-F) montrent une particule virale OtV5 reliée par une structure en goulot à la membrane plasmique de la cellule hôte *O. tauri* 30min après infection, images tirées de Derelle *et al.* 2008<sup>229</sup>. (E) Particule virale pleine avec une capsidie dense aux électrons. (F) Particule virale vide. (G) Les virus de *Micromonas* MpVN1 adsorbent à la surface de la cellule hôte par une structure protubérante en forme de queue, image tirée de Zingone *et al.* 2006<sup>222</sup>.

Si la majorité des *Prasinovirus* ne présentent pas de structure protubérante en forme de queue associée à la capsidie, elle a toutefois été reportée chez un virus de *Micromonas*. Au cours des premières étapes d'infection avec son hôte, Zingone *et al.* mettent en évidence par observation microscopique une structure tubulaire en forme de queue, rétractable pouvant mesurer jusqu'à 77 nm de long, qui relie la capsidie du virus MpVN1 à la cellule hôte<sup>222</sup> (**Figure 10**). Cette structure apparaît unique et n'a jamais été observée chez les autres membres des *Prasinovirus*.

### *Cycle de vie*

Les *Prasinovirus* ont un cycle de vie relativement rapide. Chez les virus d'*Ostreococcus* et *Micromonas*, la lyse cellulaire est obtenue 14 à 20 heures après inoculation (défini comme la période de latence)<sup>229,237</sup>. Cette période est réduite chez les virus de *Bathycoccus* BII-V1, et plus longue chez BII-V2 et BII-V3 (i.e. 16–18h)<sup>59</sup>.

La première étape du cycle de vie des *Prasinovirus* est décrite par la reconnaissance, l'adsorption et l'ancrage des particules virales aux récepteurs membranaires de la cellule hôte. Cette étape est un déterminant clé dans la spécificité des virus à reconnaître leurs hôtes. Les *Prasinovirus* sont capables d'infecter uniquement les espèces d'un même genre et aucune infection dépassant la barrière du genre n'a encore été décrite à ce jour. Par ailleurs, la spécificité des *Prasinovirus* s'étend principalement au niveau des souches avec des variations parfois importantes d'infection<sup>217,219,222,222,224,225,228,230,231</sup>. La spécificité et l'évolution de la gamme d'hôte chez les *Prasinovirus* est un mécanisme encore très mal connu aujourd'hui. C'est dans ce contexte que mon travail de thèse s'est articulé, cette étude est présentée dans le chapitre III.

Chez les *Prasinovirus*, contrairement aux virus d'amibes<sup>127,165,238,239</sup>, aux *Iridoviridae*<sup>240</sup>, *Asfarviridae*<sup>241,242</sup> et même aux *Coccolithovirus*<sup>196</sup>, qui utilisent les vacuoles de l'hôte par un processus de phagocytose ou d'endocytose pour leur internalisation dans le compartiment cellulaire. Cette étape résulte de l'attachement de la particule virale à la

membrane cellulaire puis la fusion des membranes virales et cellulaires aboutissant à l'injection du contenu viral dans le cytoplasme (**Figure 10E-G**). Les particules virales ne sont pas internalisées, mais le matériel génétique et les protéines associées sont expulsés dans le compartiment cytoplasmique hôte. Ce processus est souligné par l'observation de particules vides attachées à la membrane cellulaire (**Figure 10F**). D'autres virus eucaryotes icosaédriques semblent aussi utiliser un *vertex* unique dans les étapes précoces de leur cycle de vie telles que la reconnaissance, l'ancrage et la perforation de la membrane cellulaire hôte. Par ailleurs, l'asymétrie d'un virus au niveau d'un *vertex* peut aussi avoir d'autres fonctions biologiques notamment durant l'empaquetage et le bourgeonnement d'une particule virale<sup>243</sup>. Chez les *Prasinovirus*, ces processus liés à l'entrée et à la sortie des virions sont très peu décrits et encore mal compris à ce jour.

Chez les *Prasinovirus* (et tous les membres des *Phycodnaviridae* à l'exception des *Coccolithovirus*), aucun virus ne code pour une RNAP. Cette absence soutient l'idée que l'ADN viral, une fois éjecté dans le compartiment cytoplasmique de son hôte, est dirigé vers le noyau cellulaire et qu'une ARN polymérase hôte initie la transcription des gènes codants viraux, éventuellement soutenue par des facteurs de transcription fournis par le virus<sup>214,244</sup>. Par ailleurs, des études de MET de cellules infectées suggèrent que la membrane nucléaire de l'hôte reste intacte, au moins durant les premières étapes de réplication du virus. Aucune particule virale n'est visible dans le noyau, ni dans le chloroplaste ou dans la mitochondrie<sup>229</sup>. Une fois la transcription des gènes viraux effectuée, l'assemblage des premières particules virales est observée dans le cytoplasme de l'hôte, assez tôt au cours de l'infection (i.e. approximativement 6 heures après l'infection)<sup>225,229</sup>.

Comme pour l'entrée des virus, la sortie des *Prasinovirus* est distincte de celle décrite chez d'autres membres des NCLDV. Chez les virus enveloppés tels que les *Poxviridae*<sup>245</sup>, certains *Iridoviridae*<sup>246-248</sup> et même les *Coccolithovirus*<sup>196</sup>, un processus par bourgeonnement à partir du plasmalemme hôte est utilisé par le virus dans la libération des progénitures dans le milieu extracellulaire. Chez les *Prasinovirus*, cette dernière étape est caractérisée par une rupture physique de la membrane plasmique hôte conduisant à l'éclatement de la cellule et à la libération des virions matures dans le milieu extracellulaire.

Bien que les différentes étapes d'un cycle lytique apparaissent comparables chez tous les membres des *Prasinovirus*, en revanche le nombre moyen de particules virales nouvellement synthétisées, libérées par une cellule infectée – définit comme le « *burst size* » – est très

variables entre les virus et s'étend de 25 (i.e. les OtV) à 700 (i.e. les virus de *Micromonas* clade C (MicVC))<sup>59,220,229,249</sup>.

Chez les virus lytiques, le *burst size* et la période de latence sont deux paramètres intimement reliés qui contrôlent la dynamique entre l'hôte et son virus, et peuvent varier selon certaines conditions environnementales abiotiques et même selon le contenu en gène et la taille des génomes des deux partenaires<sup>250,251</sup>. La taille de la cellule hôte et donc le volume disponible à l'assemblage de particules virales néo-synthétisées est également une des limites du *burst size* entre les virus de différents genres (i.e. *Ostreococcus* la plus petite algue des Mamiellales, présente le plus petit *burst size* à l'infection virale)

### *Caractéristiques génomiques*

A ce jour, 22 génomes de *Prasinovirus* ont été séquencés et sont disponibles sur les bases de données (**Tableau 2**). Les génomes des *Prasinovirus*, allant de 184 à 217 kb (**Tableau 3**), sont les plus petits des *Phycodnaviridae* comparés par exemple à ceux des *Chlorovirus* ou des *Coccolithovirus* qui contiennent jusqu'à 370 kb et 410 kb respectivement<sup>252,253</sup>. La réduction de taille des génomes peut être expliquée en partie par le nombre plus restreint des régions codantes (*coding DNA sequence* – CDS) (i.e. entre 203 et 269 CDSs) (**Tableau 3**) mais aussi à une tendance générale à une réduction des régions CDS et des régions intergéniques<sup>214</sup>. Les génomes des *Prasinovirus* ont une composition en GC relativement plus faible que celle du génome de leurs hôtes, de 35–45%<sup>59,229</sup> comparé à 48–64%<sup>42,43</sup>, respectivement. Toutefois, des variations au sein des genres sont observées. Chez les virus de *Bathycoccus*, la composition en GC est relativement plus basse (~36%) que les virus de *Micromonas* (~41%) ou même d'*Ostreococcus* (~42%) (**Tableau 3**). De plus, le nombre d'ARN de transfert (tRNA) varie selon les virus, entre quatre et neuf (**Tableau 3**).

L'ordre des gènes le long des génomes linéaires des *Prasinovirus* apparaît conservé entre les virus (i.e. la synténie) à l'exception des extrémités dans lesquelles des longues régions inversées répétées terminales (*terminal inverted repeats* – TIR) sont présentes. Toutefois, les gènes partagés par tous les membres des *Prasinovirus* apparaissent plus enrichis dans la région centrale des génomes<sup>59,214,227</sup>.

Si seulement 3 gènes sont strictement partagés parmi les NCLDVs<sup>137</sup>, on retrouve en revanche 83 gènes *cores* parmi 18 génomes de *Prasinovirus*<sup>59</sup>. Bien que la plupart des gènes

communs présente des fonctions inconnues, environ un quart d'entre eux a été assigné à des fonctions biologiques requises par le virus pour mener à bien son cycle répliatif.

**Tableau 3.** Caractéristiques génomiques des vingt-deux génomes de *Prasinovirus*.

Genre	Souche	Taille des génomes (kb)	#CDSs	#tRNAs	%GC	#capsides
<i>Bathycoccus</i>	<b>BpV1</b>	198	203	4	37	7
	<b>BpV2</b>	187	210	4	37	7
	<b>BII-V1</b>	174	220	/	35	7
	<b>BII-V2</b>	208	235	/	36	7
	<b>BII-V3</b>	212	230	/	36	7
<i>Micromonas</i>	<b>MpV1</b>	184	244	6	41	8
	<b>MpV-SP1</b>	173	242	6	41	8
	<b>MpV-PL1</b>	197	261	5	43	8
	<b>MpV12-T</b>	206	253	7	40	7
<i>Ostreococcus</i>	<b>OIV1</b>	194	250	5	41	8
	<b>OIV2</b>	196	269	5	41	8
	<b>OIV3</b>	191	264	5	41	8
	<b>OIV4</b>	217	256	5	40	8
	<b>OIV5</b>	186	254	5	42	8
	<b>OIV6</b>	185	252	5	42	8
	<b>OIV7</b>	182	243	5	41	8
	<b>OmV1</b>	193	252	5	45	8
	<b>OmV2</b>	192	221	7	43	7
	<b>OtV2</b>	184	237	5	42	8
	<b>OtV1</b>	192	230	4	45	8
	<b>OtV6</b>	190	250	9	42	8
	<b>OtV5</b>	187	247	5	45	8

Cela inclut la réplication, la recombinaison/réparation de l'ADN, la transcription des gènes, la modification d'ARN, le métabolisme nucléotidique, la synthèse/modification et dégradation lipidique et des acide-aminés, ainsi que la signalisation, la méthylation de l'ADN, le métabolisme des sucres, la morphogénèse pour l'assemblage de la capsid, et d'autres fonctions diverses non classifiées<sup>59,214,227-230</sup>. En revanche, comme pour l'ensemble des *Phycodnaviridae* (excepté les *Coccolithovirus*), aucune RNAP n'est trouvée chez les *Prasinovirus*, suggérant une totale dépendance à la machinerie de transcription de la cellule hôte pour transcrire les gènes viraux.

En tant que gènes *cores*, l'ADN polB (*polB*) et la MCP sont les marqueurs moléculaires les plus communément utilisés pour étudier la diversité des *Prasinovirus* dans les environnements naturels et dans les cultures isolées<sup>217,218,220,222,224–226</sup>. Les *Prasinovirus* et les membres des NCLDV s codent généralement une seule copie de *polB*, tandis que huit gènes distincts de capsides dont la MCP, sont répartis le long du génome des *Prasinovirus* (**Tableau 3**). Les virus de *Bathycoccus*, de MpV12-T et de OmV2 présentent une copie en moins, suggérant une perte de ce gène au cours de leur histoire évolutive. A noter, un nombre élevé de protéines de capsides (4–5 copies) a aussi été détecté chez les *Chlorovirus*<sup>139</sup>. Chez les *Prasinovirus*, des analyses phylogénétiques réalisées sur l'ensemble des copies de capsides, suggèrent que la 6<sup>e</sup> copie – déterminée comme le 6<sup>e</sup> gène de capside le long des génomes – partage une origine évolutive commune avec les *Phycodnaviridae* et même les NCLDV s<sup>218,230</sup>, et serait donc vraisemblablement la capside majeure. Le nombre étonnamment élevé de gènes de capside doit impacter sur la morphologie structurale de ces virus. Il a été proposé que les protéines minoritaires de capside (i.e. capsides mineures) participent au processus d'encapsidation des *Prasinovirus*, peut-être durant les premières étapes d'infection dans l'adsorption et la fusion des membranes de la cellule hôte<sup>236</sup>. Ce rôle est sous-jacent aux observations de MET qui mettent en évidence la fusion des membranes entre virion et cellule avant injection du contenu de la capside virale dans le cytoplasme : un phénomène commun au *Prasinovirus*.

Par ailleurs, les génomes des *Prasinovirus* contiennent un certain nombre de gènes uniques. Ces gènes sont notamment connus dans des processus de biosynthèse des acides aminés (e.g. *acetolactate synthase*, *thiamine diphosphate-dependent enzyme*), du métabolisme du carbone (i.e. *6-phosphofructokinase*), de la synthèse et/ou au transfert de carbohydrate sur des protéines, des lipides et même sur l'ADN (nombreuses glycosyltransférases de familles distinctes)<sup>228,230,236</sup>. Encore plus surprenant, certains gènes codés par les *Prasinovirus* sont décrits pour jouer un rôle dans la régulation métabolique, le transfert d'énergie et l'activation de certaines protéines, chez la cellule hôte (e.g. la protéine de la famille PhoH), mais aussi dans la régulation du photosystème hôte (e.g. *FtsH metalloendopeptidase*)<sup>228,230,236</sup>. Par ailleurs, un cas unique de *Prasinovirus* a été reporté pour le virus OtV6, dans lequel un transporteur d'ammonium, dérivé de son hôte *Ostreococcus tauri*, a été retrouvé dans le génome du virus<sup>232</sup>. Ces analyses phylogénétiques indiquent une acquisition de ces gènes spécifiques de l'hôte, mais aussi des bactéries marines vers le virus à travers des transferts latéraux géniques (*lateral gene*

*transfer* – LGT), qui se serait produite plus ou moins tôt au cours du processus évolutif et de spéciation des *Prasinovirus*.

### **3. Dynamiques des interactions hôte–virus**

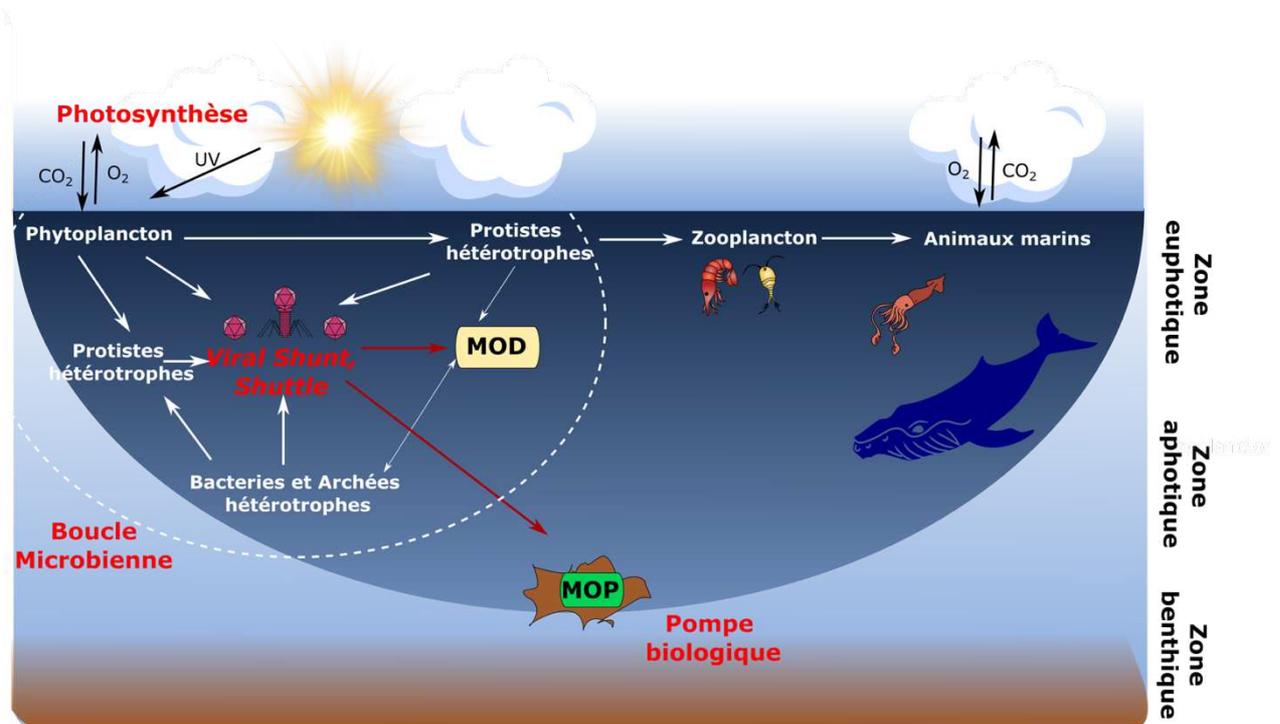
La répartition ubiquitaire des virus marins dans les environnements aquatiques a suscité beaucoup d'interrogations concernant l'influence des virus sur les écosystèmes marins à différents niveaux. Notamment quels sont les impacts des virus lytiques 1) sur la production primaire et le cycle du carbone dans les océans, 2) sur la structure des communautés et l'évolution des hôtes, et enfin 3) comment les virus et leurs hôtes peuvent coexister stablement dans les mêmes environnements ? Toutes ces questions sont abordées dans cette dernière section d'introduction.

#### **3.1. Impact des virus marins sur les cycles biogéochimiques et le climat**

Avec un taux d'infection estimé à  $10^{23}$  par seconde dans les océans, il est clair que les virus marins constituent une des principales sources de mortalité des organismes hôtes<sup>96</sup>. On estime que 20 à 40% des procaryotes marins et un quart des populations phytoplanctoniques sont tués chaque jour par des virus<sup>96,254</sup>. En raison du caractère unicellulaire du phytoplancton, l'infection par un virus lytique conduit inévitablement à la mort cellulaire de l'hôte, impactant immédiatement l'abondance de sa population hôte. De par leur activité lytique, les virus sont des agents importants dans la régulation des cycles biogéochimiques<sup>95,255–258</sup>.

En court-circuitant le flux de carbone organique ainsi que les nutriments des réseaux trophiques inférieurs, les virus marins lysent les cellules hôtes et libèrent ainsi la MOD dans la zone euphotique par le processus appelé « *shunt viral* ». Cette MOD est ensuite réutilisée comme source d'énergie par les bactéries et réintégrée dans le réseau trophique, maintenant ainsi la croissance des communautés virales et les populations bactériennes. Contrairement au *shunt viral* qui détourne le carbone de la chaîne alimentaire, la dynamique des populations phytoplanctoniques est contrebalancée par les prédateurs des protistes hétérotrophes et des brouteurs microbiens au sein de la chaîne alimentaire de « pâturage ». Le broutage microbien représente la redirection du carbone la plus importante de la chaîne alimentaire marine vers les trophiques supérieurs<sup>4</sup>. Si la plus grande majorité de la matière organique est recyclée dans la boucle microbienne sous forme de MOD, une fraction de moins de 20% s'écoule sous forme de matière organique particulaire (MOP) et descend dans les profondeurs vers la zone benthique

par la pompe biologique, servant comme un immense réservoir de séquestration du carbone dans les océans<sup>259</sup>. Le concept initial du *shunt viral* proposait que la lyse virale contribue à la libération de la MOD, en permettant d'alimenter la boucle microbienne. Aujourd'hui, il est aussi proposé que la lyse virale augmenterait la MOP et la pompe biologique dans les océans : ce concept a été défini comme le « *viral shuttle* »<sup>133,260</sup> (Figure 11).



**Figure 11.** Illustration schématique des relations entre les organismes planctoniques dans les écosystèmes marins. Les échanges gazeux entre l'océan et l'atmosphère sont représentés par les flèches en noir, tandis que les flèches pleines représentent les transferts de matière organique dissoute (MOD) ainsi que les échanges trophiques entre les différents planctoniques. Le phytoplancton et d'autres organismes du plancton (bactéries, archées) sont consommés par les protistes hétérotrophes, lui-même consommé par le zooplancton, et ce dernier consommé par des petits carnivores mangeurs de plancton. La population phytoplanctonique, les protistes hétérotrophes ainsi que les bactéries et les archées hétérotrophes sont régulés par le virioplancton (virus marin) au travers du *shunt viral*, et sont reminéralisés par les bactéries dans la boucle microbienne. Ce mécanisme détourne le flux du carbone en le recyclant en MOD, réutilisable par les réseaux trophiques inférieurs. La mort cellulaire programmée ou induite du plancton par des parasites, conduit à la libération de la matière organique particulaire (MOP) et à sa séquestration au fond des océans dans la zone benthique. Cette séquestration du carbone est également favorisée par la lyse virale au travers du processus nommé *viral shuttle*. La séquestration du carbone est réduite grâce à la respiration et au relargage de CO<sub>2</sub> dans l'atmosphère par les bactéries, le plancton hétérotrophe, le zooplancton, les carnivores, et quelques taxons photosynthétiques nocturnes. Illustration adaptée de Breitbart *et al.* 2018<sup>261</sup>.

Les virus marins jouent aussi un rôle notable dans la régulation du climat. Un exemple concerne la production de molécules sulfurées, le diméthylsulfoniopropionate (DMSP) et le diméthylsulfure (DMS), dans les océans. Le coccolithophore *Emiliana huxleyi* est un des principaux producteurs de ces composés<sup>262,263</sup>. La production de DMSP a été aussi rapportée chez *Micromonas pusilla*<sup>264</sup>. La dégradation du DMSP par les bactéries produit du DMS, un gaz volatil qui joue un rôle majeur dans la régulation du climat en favorisant la formation de nuage et de pluies acides<sup>265</sup>. En tuant et régulant les populations phytoplanctoniques productrices de molécules sulfurées, les virus marins agissent directement sur la libération massive de DMSP et de DMS dans l'atmosphère, et sont donc des régulateurs importants du climat et des cycles biogéochimiques à l'échelle planétaire<sup>266</sup>.

### **3.2. Impact des virus marins sur la structure des communautés planctoniques**

L'infection virale est uniquement possible si la distance entre le virus et son hôte est suffisamment proche. Ainsi, une forte dilution des deux partenaires dans l'environnement permettrait un équilibre entre le taux de croissance de l'hôte et de la lyse virale. Cette relation a donné naissance au modèle « *killing the winner* » où le « gagnant » serait la première cible des virus et ferait donc référence à la population la plus active et probablement la plus abondante<sup>267</sup>. En d'autres termes, les virus maintiendraient une compétition intraspécifique en régulant négativement les souches les plus abondantes et en permettant aux autres espèces planctoniques de se développer<sup>268</sup>. La nature spécifique de l'hôte restreinte à la souche de l'infection virale, restreinte à la souche, fait des virus marins des agents puissants dans le contrôle des communautés microbiennes. Dans le milieu naturel, la dissolution d'efflorescence d'algues ou de cyanobactéries est une des conséquences visibles de la propagation rapide des virus dans les populations hôtes. Ce phénomène a principalement été observé et étudié chez *E. huxleyi*<sup>269,270</sup>. La disparition des efflorescences d'algues provoquée par les virus a aussi été démontrée chez d'autres espèces telles que *Phaeocystis globosa*<sup>203</sup>, *Heterosigma akahiwo*<sup>205</sup>, le prymnesiophyte *Chrysochromulina*<sup>215</sup> et *Ostreococcus*<sup>65</sup>. La disparition d'efflorescence est l'un des aspects les plus évidents de l'influence des virus marins sur le maintien d'une grande diversité des communautés hôtes. Ainsi, face au concept de sélection naturelle, les espèces les plus abondantes, (i.e. les plus compétitives pour une ressource) sont aussi exposées à une plus forte pression de contre-sélection par les virus. Cette dynamique maintiendrait alors les espèces les plus rares.

Par opposition à l'extinction, les virus marins peuvent réguler à l'échelle spatio-temporelle, la composition et l'abondance des populations hôtes, photoautotrophes, non-efflorescentes à un niveau minimal. Les virus marins agissent donc comme des acteurs « préventifs » contre le développement massif de communautés hôtes trop homogènes<sup>271</sup> favorisant ainsi une coexistence stable des espèces en compétition.

Le paradoxe du plancton est une question fondamentale qui a été soulevé en 1961 par Hutchinson<sup>272</sup> : comment est-il possible pour un certain nombre d'espèces de coexister dans un environnement relativement limité en nutriment tout en étant en concurrence pour les mêmes ressources, alors que des modèles théoriques prédisent que seules les espèces compétitives dominantes peuvent persister ? Aujourd'hui il est clair que l'implication des virus marins sur la structure des communautés planctoniques par des infections lytiques fournit une première explication à ce paradoxe. La coexistence entre l'hôte et son virus a principalement été étudiée dans le système cyanobactérie–cyanophages<sup>273,274</sup>. La présence simultanée et répétée d'algues phytoplanctoniques et de leurs virus ont également été rapportée<sup>203,205,215,234,275</sup>.

Si les virus marins sont des acteurs clés dans le contrôle de la structure des communautés microbiennes, en revanche la façon dont il régule la diversité microbienne dans la nature reste ambiguë.

### 3.3. Coévolution dans le système hôte–virus

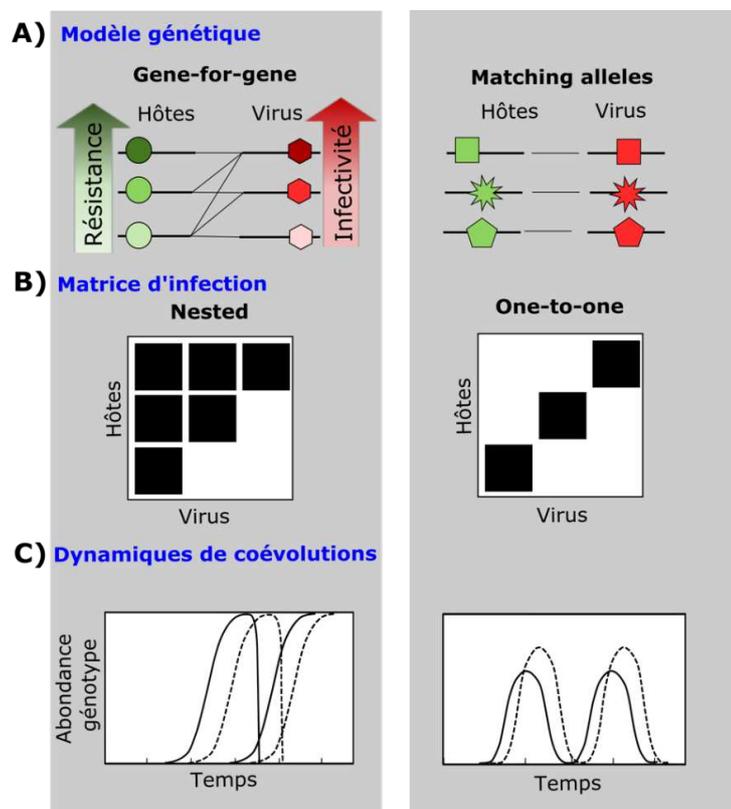
A l'échelle macroévolutive, la coexistence stable de l'hôte et du virus en concurrence conduit généralement à un profil de coévolution phylogénétique (cophylogénétique) entre les deux partenaires, selon lequel l'évolution des virus reflète celle de leurs hôtes et *vice versa*<sup>276</sup>. La congruence topologique entre l'hôte et son virus a notamment été soulignée dans le système marin chez les Mamiellales et les *Prasinovirus*, suggérant l'existence d'une dynamique évolutive commune à plus long terme<sup>224</sup>. Toutefois, une plus faible congruence topologique entre l'hôte et les virus chez *Ostreococcus* et *Bathycoccus* a été observée, suggérant un changement d'hôte de *Bathycoccus* à *Micromonas* assez tôt dans l'histoire évolutive des *Prasinovirus*<sup>224</sup>. La dynamique coévolutive à l'échelle macroévolutive conduit généralement à des occurrences concomitantes de la spéciation de l'hôte et son virus appelé « cospéciation »<sup>276</sup>. La cospéciation significative entre les *Prasinovirus* et les Mamiellales peut notamment expliquer la gamme de spécificité virale restreinte à l'espèce hôte (spécificité inter-clade chez les virus de *Micromonas*) avec des variations d'infection entre souches hôtes<sup>220,231</sup>. Cette

perspective étroite de la gamme d'hôte pourrait avoir des implications sur la dynamique coévolutive à plus court-terme (i.e. à l'échelle microévolutive) ainsi que des conséquences sur une coévolution dite « antagoniste ».

La coexistence dans un système hôte–parasite est possible si les espèces en compétition coévoluent avec des forces opposées pour maintenir cet équilibre à plus long terme. A l'échelle microévolutive, la coévolution implique généralement des modifications génétiques alléliques entre les espèces en compétition au cours des générations successives<sup>276</sup>. Cette dynamique de forces opposées entre les espèces en compétition constitue donc un processus majeur dans la diversification génétique chez les deux partenaires. Dans le système hôte–virus, les deux espèces subissent une coévolution antagoniste, selon laquelle les hôtes seraient forcés à développer des mécanismes rapides de résistance contre l'infectivité du virus, et le virus à acquérir des mécanismes de contre-défense en retour<sup>277</sup>. Ainsi, la coévolution antagoniste conduirait à de fortes pressions de sélection et une diversité cooccurrence au sein des populations des hôtes et des virus. Fondamentalement, ce concept repose sur la compatibilité génétique et moléculaire de reconnaissance entre le virus et son hôte. Deux principaux modèles qui décrivent les bases génétiques des interactions hôtes–virus ont été rapportés<sup>278–280</sup> (**Figure 12**). Ces deux modèles impliquent des gènes liés à la résistance de l'hôte et à l'infectivité du virus, mais avec des contraintes distinctes :

-Le premier modèle nommé « *matching alleles* » (MA), a été fondé sur la notion de système de reconnaissance spécifique du soi et du non-soi dans le système hôte–virus chez les animaux<sup>281</sup>. Ce modèle est expliqué par une correspondance génétique compatible unique entre le génotype viral et le génotype hôte : un virus, ne peut donc infecter qu'un seul hôte que s'il a adapté ses contre-défenses au type de système immunitaire de ce même hôte. Si on transpose ce modèle à une organisation matricielle d'infection hôte–virus, le modèle MA serait associé à une organisation de type « *one-to-one* »<sup>282</sup>. Selon ce modèle, l'hôte et le virus évoluent simultanément selon « une dynamique de sélection fluctuante », où les virus s'adaptent pour infecter le génotype hôte le plus commun, entraînant son remplacement par des génotypes hôte/virus plus rares. Aucun changement directionnel dans la gamme de résistance n'est observé dans ce cas (**Figure 12**)

-Le second modèle nommé « *gene-for-gene* » (GFG), a été initialement décrit par Flor durant les années 1950 dans le système plante–pathogène<sup>283,284</sup>. Ce modèle prédit que chaque pair allélique de résistance chez l’hôte présente une paire d’allèles de virulence correspondantes chez le virus, mais avec des coûts de fitness associés<sup>285</sup>. En d’autres termes, l’hôte évolue de manière à augmenter son spectre de résistance, en réponse à l’évolution des virus de sa gamme d’hôte à l’infection. L’accumulation d’allèles liée à la virulence permet donc au virus d’infecter une plus large gamme de génotypes hôtes sensibles (i.e. virus généralistes). A l’inverse, un nombre d’allèles de virulence restreints limite l’infection du virus à un seul génotype hôte résistant (i.e. virus spécialistes) (**Figure 12**).



**Figure 12.** Modèles interactions et dynamiques de coévolutions hôte–virus qui en résultent : *gene-for-gene* (GFG) et *Matching allele* (MA). (A) Les gènes impliqués dans les interactions hôtes–virus sont représentés par des symboles : le rond et l’hexagone représentent les gènes spécifiques associés à l’hôte et au virus respectivement. Une infection entre deux génotypes compatibles est possible lorsqu’une ligne relie les deux espèces en compétition. Plus la résistance de l’hôte augmente plus l’infectivité du virus est grande en retour. (B) Matrice d’infection croisée entre les génotypes de l’hôte et du virus. Un carré noir indique une infection réussie. (C) Dynamiques coévolutive entre plusieurs génotypes d’hôtes et de virus qui coexistent à un moment donné. Les deux modèles représentent les fréquences alléliques de l’hôte (trait plein) et du virus (trait discontinu).

Ce type d'interaction a notamment été décrit dans le système phage–bactérie, associant la théorie GFG à une organisation matricielle d'infection de type « *nested* »<sup>282,286,287</sup>. Le modèle GFG est associé à une dynamique coévolutive dite de « courses aux armements », définie par l'évolution successive et la sélection directionnelle de nouveaux génotypes de virulence chez le virus et de génotypes de résistance chez l'hôte<sup>288</sup>. Récemment, des analyses en laboratoire ont permis de démontrer la dynamique évolutive rapide de résistance de l'hôte et de l'infectivité du virus, semblable à une dynamique de course aux armements, dans le système *Chlorella variabilis*–PBCV-1<sup>289</sup>.

La coexistence stable entre plusieurs espèces en compétition peut aussi être renversée par une dynamique axée sur du « compromis », qui soutient une coexistence entre une espèce hôte dominante généralement résistante et une minorité d'espèces sensibles à l'infection à croissance plus rapide, maintenant la petite population virale au sein du système<sup>290</sup>. Ce phénomène est aussi décrit comme une évolution (ou acquisition) de la résistance des espèces hôtes en réponse à l'infection virale. Bien que l'acquisition de la résistance soit un processus biologique encore mal compris aujourd'hui, de nombreuses études ont reporté une telle dynamique chez des algues taxonomiquement diverses (*Chlorophyceae*, *Dinophyceae*, *Haptophyta* et *Mamiellophyceae*)<sup>222,275,291,292</sup> ainsi que chez la cyanobactérie marine *Synechococcus* sp.<sup>273</sup>, souvent suivies par une coexistence stable entre l'algue et le virus sur de nombreuses générations successives<sup>71,291,293</sup>.

### 3.4. Mécanismes de défenses antivirales

#### *Inhibition à l'attachement et à l'entrée*

L'incapacité d'adsorption du virus en surface de sa cellule hôte constitue une des premières barrières physiques à l'infection, un déterminant clé de la gamme de spécificité d'hôte. La résistance physique de la surface cellulaire hôte peut être causée par plusieurs raisons : 1) l'altération, 2) la modification, 3) la réduction et/ou 4) la perte totale des sites récepteurs<sup>279,294,295</sup>. L'ensemble de ces changements résultent de mutations génétiques principalement responsables de la biosynthèse de structures de la paroi cellulaire hôte<sup>296</sup>. Dans le système cyanobactérie–cyanophage, la plupart des mutations conférant la résistance à l'infection se sont avérées être situées dans une unique région spécifique hypervariable (i.e. îlot génomique), soumise à une très forte pression de sélection par le virus<sup>297,298</sup>. Le caractère hautement spécifique des virus est soutenu par cette variabilité phénotypique des souches qui

s'exprime au travers d'une grande diversité génétique, physiologique et différents mécanismes moléculaires.

#### *Mécanisme intracellulaire « classique »*

Même si la première ligne de défense peut être franchie par le virus, des mécanismes d'immunité intracellulaire peuvent bloquer la réplication du génome viral. Deux principaux modèles sont décrits : 1) Le système de restriction-modification (R-M *system*), est l'un des dispositifs biologiques les plus connus chez les bactéries et les archées<sup>299</sup>. Ce dispositif antiviral est fondé sur le recrutement d'endonucléase hôte permettant la reconnaissance de site spécifique, le clivage et la destruction d'ADNdb étranger non-méthylé. 2) Plus récemment, un autre système immunitaire microbien, le système CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) -Cas (CRISPR-*associated genes*), a été identifié chez les archées (~90%) et les bactéries (40%), les protégeant des attaques virales et plasmidiques<sup>300,301</sup>. Ce mécanisme de défense plus complexe repose sur l'intégration d'une région virale ou plasmidique dans le génome hôte, qui confère alors à la cellule hôte une résistance hautement spécifique à ce virus. Le système antiviral CRISPR-Cas a été rapporté dans les écosystèmes marins<sup>302</sup> et semble largement distribué chez les cyanobactéries à l'exception des genres *Prochlorococcus* and *Synechococcus*<sup>303</sup>. Par ailleurs, bien que le système CRISPR-Cas a seulement été reporté chez les bactéries et les archées (à l'exception d'un bactériophage<sup>304</sup>), un mécanisme similaire a été décrit chez le mimivirus<sup>305</sup>. Ce système de défense unique a été nommé le MIMIVIRE (*mimivirus virophage resistance element*), puisqu'il confère spécifiquement une immunité antivirale au Mimivirus, uniquement la lignée A, contre l'infection aux virophages de type Zamilon.

L'incorporation d'un prophage dans le génome de l'hôte (i.e. la lysogénie) est un autre mécanisme de défense antiviral bien connu chez les bactéries. Il permettrait notamment d'augmenter le *fitness* de l'hôte infecté en l'immunisant contre de nouvelles infections par des virus homologues, et lui conférer de nouvelles fonctions adaptatives comme c'est le cas chez les souches toxigènes de *Vibrio cholerae*<sup>306</sup>. Bien que la lysogénie ait seulement été décrite chez les *Phaeovirus*<sup>90</sup>, l'insertion de fragments d'ADN viral dans les génomes des algues eucaryotes, définis comme les « *endogenous viral elements* » (EVEs), semble être un trait commun aux NCLDVs<sup>307</sup>. En revanche, le rôle de ces séquences dans les mécanismes de défenses contre les virus reste à ce jour un sujet ouvert au débat.

## *Mécanisme immunitaire antivirale dans l'interaction Ostreococcus–Prasinovirus*

Dans le système *Ostreococcus–Prasinovirus*, l'acquisition de la résistance apparaît de manière rapide en culture (3–5 jours après-infection), et se maintient stablement sur plusieurs générations successives, parfois plusieurs années<sup>293,308,309</sup>. Chez *Ostreococcus*, les souches devenues résistantes au virus ont une grande plasticité de taille du chromosome 19 (SOC)<sup>310</sup>. En ce sens, une corrélation positive entre la longueur de ce chromosome et la résistance a été démontrée<sup>70</sup>. Ces importantes variations de tailles ont été corrélées à des réarrangements structuraux massifs de portions de chromosome de type délétion, duplication, et translocation<sup>310</sup>.

Des analyses transcriptomiques de souches *O. tauri* RCC4221 résistantes et sensibles au virus OtV5 ont permis de révéler des changements transcriptionnels importants localisés principalement sur le chromosome SOC, qui soutient près d'un tiers des 170 gènes différenciellement transcrits<sup>310</sup>. Ce dernier concentre essentiellement des protéines de surface membranaire, de production de glycoconjugués important pour les interactions hôtes–virus, ainsi que des protéines associées au transports et aux métabolismes des glucides (i.e. glycosyltransférases et des méthyltransférases). De façon remarquable, la plupart de ces gènes sont massivement surexprimés dans les souches résistantes au virus par rapport à celles sensibles, et sont localisés uniquement dans une grande région répétée inversée (*large inverted repeat region* – LIRR) sur le chromosome 19. Cette région a la particularité de présenter une structure transcriptionnelle bipartite : la partie 5' de la LIRR est surexprimée chez les souches résistantes et réprimée chez les souches sensibles, tandis que chez ces dernières l'expression est renforcée coté partie 3'. Il est clair que l'expression de ces gènes par les cellules résistantes peut-être une première cause de la résistance au travers de camouflage ou de modifications de récepteurs à la surface d'*O. tauri*, afin d'inhiber l'adsorption du virus. En ce sens, les souches résistantes ont montré une très forte réduction d'adsorption aux virus OtV5 ainsi qu'à un ensemble de virus génétiquement apparentés, suggérant un mécanisme immunitaire antiviral spécifique et qui évolue rapidement<sup>308</sup>.

La remarquable richesse d'éléments transposables (ETs) localisés dans le chromosome SOC (i.e. 20% des 400 recensés) soulève également beaucoup d'interrogation quant à leur rôle dans la résistance au virus. Yau et al. ont démontré une expression différentielle des ETs dans les souches résistantes par rapport à celles sensibles<sup>310</sup>. Par ailleurs, une analyse temporelle transcriptionnelle de souches *O. tauri* RCC4221 sensibles infectées et non-infectées à OtV5 a

été réalisée<sup>309</sup>. Cette analyse a mis en évidence une très forte expression d'une transcriptase inverse sur le chromosome 8, durant la phase de division cellulaire. La forte transcription des gènes qui composent cet élément, (i.e. réplique/intégrase) a été observée uniquement dans les cellules infectées. Les auteurs proposent que le stress causé par l'infection virale soit une des principales raisons de l'activation du rétroélément à *long terminal repeats* (LTRs), conduisant à la transposition du gène lui-même et à l'activation du rétrotransposon miniature, non-autonome (*terminal repeat retrotransposon in miniature* – TRIM) du chromosome 19. Ainsi, une telle activation est suspectée de conduire aux importants réarrangements chromosomiques observés et par conséquent de promouvoir l'expression ou la répression de certains gènes du chromosome 19. A ce jour, l'ensemble des travaux réalisés avec le modèle *Ostreococcus–Prasinovirus* supporte fortement l'implication du chromosome SOC dans les mécanismes de défenses immunitaires à l'infection virale.

#### **4. Objectifs de la thèse**

Mon projet de recherche est partie prenante du projet ANR – ALGALVIRUS « Adaptations génomiques des algues Marines aux Virus » qui visait à caractériser les mécanismes moléculaires et génétiques associés à l'interaction hôte–virus, dans le cadre de la résistance de l'algue *Ostreococcus tauri* aux *Prasinovirus*. Cette thèse est sous-divisée en trois parties principales, organisées en 4 chapitres qui incluent 2 articles dont un a été accepté et publié dans le journal MDPI *Cells*.

La résistance innée ou l'acquisition de la résistance aux virus a été mise en évidence chez de nombreuses algues taxonomiquement distinctes. Chez *Ostreococcus tauri*, l'immunité antivirale est vraisemblablement associée à d'importants réarrangements structuraux d'un chromosome atypique nommé SOC. Durant l'infection virale des cellules sensibles au virus et chez les lignées résistantes, l'expression différentielle de gènes localement situés sur le SOC sont majoritairement des protéines de surface membranaire, de production de glycoconjugués important pour les interactions hôtes–virus, ainsi que des protéines associées au transports et aux métabolismes des glucides (i.e. glycosyltransférases et des méthyltransférases). Par ailleurs, ces travaux ont convergé vers une hypothèse impliquant les éléments transposables comme *driver* de l'expression des gènes et l'évolution des génomes de la picoalgue dans le cadre de l'immunité antivirale. C'est donc dans ce contexte que mon projet de thèse s'est articulé.

Le premier objectif de ma thèse a été de gagner en connaissances sur **les bases génétiques de la résistance aux *Prasinovirus* chez l'algue marine *Ostreococcus tauri***. Pour répondre à cette question, nous avons choisi d'utiliser une approche génétique classique de mutagenèse insertionnelle aléatoire *via* la transformation médiée par le polyéthylène glycol (PEG). Dans le premier chapitre nous décrivons pour la première fois la caractérisation détaillée des événements individuels d'intégration génomique d'ADN étranger au cœur de la transformation par une approche génomique de séquençage de nouvelle génération à haut débit (*next-generation sequencing* – NGS). Dans le prolongement de cette étude, nous avons tenté dans le chapitre II de corrélérer le caractère phénotypique des lignées clonales transformées aux données génomiques dans le cadre de la résistance aux *Prasinovirus*.

Compte tenu de l'abondance des *Prasinovirus* dans les régions marines côtières contrastant avec la faible quantité de données génomiques disponibles pour les virus spécifiques d'*O. tauri*, la seconde thématique de cette thèse a rendu compte de la **diversité des génomes de *Prasinovirus* à l'échelle planétaire**. Cette étude a donc eu pour second objectif de **gagner en connaissances sur les bases génétiques qui sous-tendent l'évolution de la gamme d'hôte chez les *Prasinovirus***. Pour cela, une analyse comparative de 18 génomes de virus d'*O. tauri* (OtVs), dont 5 virus nouvellement isolés, a été conduite. Afin de répondre aux deux problématiques de ce troisième chapitre, nous avons séquencé des génomes OtVs qui présentaient des origines géographiques distinctes ainsi que des virus qui présentaient un spectre d'infection très variables entre les souches d'*O. tauri* testées.

Dans la continuité de ce travail, le dernier objectif de la thèse a été de **caractériser la morphologie des *Prasinovirus* spécifiques de l'espèce *O. tauri***, dans le but d'essayer de relier la diversité génomique décrite dans le chapitre III aux caractères morphologiques observés. Cette étude a été réalisée *via* la microscopie électronique à transmission (MET). La caractérisation des structures morphologiques des virus est nécessaire pour mieux comprendre les bases moléculaires qui sous-tendent les interactions hôtes–virus dans le cadre de la résistance et de la gamme d'hôte au sein du système *O. tauri*–*Prasinovirus*.





# CHAPITRE I – CARACTERISATION GENOMIQUE DES EVENEMENTS D'INSERTION ET IMPACTS DE LA TRANSFORMATION MEDIEE PAR POLYETHYLENE GLYCOL SUR LE GENOME CHEZ L'ALGUE MARINE OSTREOCOCCUS TAURI

Ce premier chapitre présente l'article « *Combining Nanopore and Illumina sequencing permits detailed analysis of insertion mutations and structural variations produced by PEG-mediated transformation in *Ostreococcus tauri** », accepté en Mars 2021 dans le journal *Cells* MDPI.

## **Objectifs du chapitre :**

Le premier objectif de ma thèse avait pour but d'utiliser des nouvelles méthodes de transformation génétique permettant d'approcher chez l'algue marine *O. tauri* les mécanismes moléculaires de l'immunité antivirale aux *Prasinovirus*. L'article présenté dans ce chapitre montre pour la première fois une caractérisation détaillée par une approche génomique de séquençage à haut débit (NGS) des événements individuels d'intégration génomique d'ADN étranger après transformation médiée par polyéthylène glycol (PEG) dans le génome d'*O. tauri*.

## **Contributions de l'auteur :**

Au travers d'une collaboration internationale (le financement du programme de recherche et d'innovation Horizon 2020 de l'Union européenne sous la convention de subvention n° 824110-EASI-Genomics et la subvention de projet de l'Agence nationale de la recherche française (ANR ALGALVIRUS), le laboratoire CNAG-CRG à Barcelone (Marta Gut, Tyler Alioto et Fernando Cruz) a pris en charge le séquençage des deux banques de données (Illumina et Nanopore) ainsi que l'assemblage des génomes des lignées clonales *O. tauri* RCC1115. Par ailleurs, j'ai conduit tous les travaux expérimentaux en amont et les analyses génomiques des sept génomes assemblés d'*O. tauri* et j'ai rédigé l'article. Frédéric Sanchez a participé aux expériences de transformation et d'expression. Nigel Grimsley et Sheree Yau ont édité l'article.

**Combining Nanopore and Illumina sequencing permits detailed analysis of insertion mutations and structural variations produced by PEG-mediated transformation in in *Ostreococcus tauri***

Julie Thomy <sup>1</sup>, Frederic Sanchez <sup>1</sup>, Marta Gut <sup>2,3</sup>, Fernando Cruz <sup>2</sup>, Tyler Alioto <sup>2,3</sup>, Gwenael Piganeau <sup>1</sup>, Nigel Grimsley <sup>1,\*</sup>, Sheree Yau <sup>1,\*</sup>

**Author affiliations**

<sup>1</sup> Sorbonne Université, CNRS, UMR 7232 Biologie Intégrative des Organismes Marins (BIOM), Observatoire Océanologique, F-66650 Banyuls-sur-Mer, France ; [thomy@obs-banyuls.fr](mailto:thomy@obs-banyuls.fr) (J.T) ; [sanchezf@obs-banyuls.fr](mailto:sanchezf@obs-banyuls.fr) (F.S) ; [gwenael.piganeau@obs-banyuls.fr](mailto:gwenael.piganeau@obs-banyuls.fr) (G.P)

<sup>2</sup> CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, 08028 Barcelona, Spain; [marta.gut@cnag.crg.eu](mailto:marta.gut@cnag.crg.eu) (M.G); [fernando.cruz@cnag.crg.eu](mailto:fernando.cruz@cnag.crg.eu) (F.C); [tyler.alioto@cnag.crg.eu](mailto:tyler.alioto@cnag.crg.eu) (T.A)

<sup>3</sup> Universitat Pompeu Fabra (UPF), Barcelona, Spain

**Corresponding authors** \*: [nigel.grimsley@obs-banyuls.fr](mailto:nigel.grimsley@obs-banyuls.fr) (N.G), [sheree.yau@obs-banyuls.fr](mailto:sheree.yau@obs-banyuls.fr)

**Supplementary Materials:** The following documents are available online at <https://www.mdpi.com/2073-4409/10/3/664/s1>.

## Résumé

Ce premier chapitre décrit pour la première fois chez l'algue *O. tauri* une caractérisation détaillée des événements individuels d'intégration génomique d'ADN étranger d'origine plasmidique après transformation médiée par le PEG. Par une analyse génomique, nous mettons en évidence que l'intégration de l'ADN exogène se produit de façon aléatoire à un seul locus dans le génome de l'algue et principalement en une copie unique. Parallèlement, nous décrivons finement les mécanismes d'intégration du vecteur dans le génome au niveau des loci d'insertion. À partir de ces observations, nous suggérons qu'un mécanisme de type NHEJ (*non-homologous end joining*) soit utilisé lors d'événements d'insertion aléatoires, comme décrit chez les plantes et d'autres modèles d'algues d'eau douce. La transformation médiée par le PEG est un outil prometteur dans la poursuite des études génomiques fonctionnelles, notamment dans le cadre de l'immunité antivirale chez *O. tauri*.

## Abstract:

*Ostreococcus tauri* is a simple unicellular green alga representing an ecologically important group of phytoplankton in oceans worldwide. Modern molecular techniques must be developed in order to understand the mechanisms that permit adaptation of microalgae to their environment. We present for the first time in *O. tauri* a detailed characterization of individual genomic integration events of foreign DNA of plasmid origin after PEG-mediated transformation. Vector integration occurred randomly at a single locus in the genome and mainly as a single copy. Thus, we confirmed the utility of this technique for insertional mutagenesis. While the mechanism of double-stranded DNA repair in the *O. tauri* model remains to be elucidated, we clearly demonstrate by genome resequencing that the integration of the vector leads to frequent structural variations (deletions/insertions and duplications) and some chromosomal rearrangements in the genome at the insertion loci. Furthermore, we often observed variations in the vector sequence itself. From these observations, we speculate that a non-homologous end joining-like mechanism is employed during random insertion events, as described in plants and other freshwater algal models. PEG-mediated transformation is therefore a promising molecular biology tool, not only for functional genomic studies, but also for biotechnological research in this ecologically important marine alga.

## 1. Introduction

In most living organisms a high proportion of genes have unknown functions and the determination of these functions is a laborious task. To facilitate this functional annotation, different genetic tools are available, such as radiation or chemical mutagenesis<sup>311</sup>, genetic engineering<sup>312</sup>, RNA interference<sup>313,314</sup> and gene expression systems<sup>315</sup>. Beyond the advantages of molecular biology providing new methods, these genetic tools allow us to increase the precision of analyses, the throughput and the diversity of organisms that it is possible to study. Random insertional mutagenesis, which entails the integration of exogenous DNA (typically transposable elements, plasmids or viruses) into the genome, is one of the most powerful genetic tools to disturb gene function. This has emerged as the method of choice to produce random genome-wide mutant collections in cultivable organisms with short life-cycles that are amenable to genetic transformation<sup>316</sup>. Different methods have been developed for nuclear transformation in many model organisms to enable functional genetic characterisation, but there is now growing recognition of the need to study a larger number of “non-model” or “emerging model” systems<sup>317</sup>.

In unicellular algae, for many years, the development and application of molecular genetic approaches have been largely confined to the freshwater model alga, *Chlamydomonas reinhardtii*<sup>318</sup>. Microalgae span an immense breadth of taxonomic and genetic diversity, and although they offer opportunities for biotechnology, they are poorly exploited among microorganisms for genetic engineering<sup>319</sup>. The need to develop marine model organisms has been recognised by the Environmental Model Systems (EMS) Project, a multi-team collaboration aiming to advance DNA delivery protocols for marine protists. Prior to the EMS project, only a handful of marine protists were genetically transformable, and this figure has increased to ~20<sup>317</sup>. The continued development of genetic manipulation of marine algae opens the door to realizing high-throughput functional studies and development of strains to synthesize high-value products, such as biofuels, food additives, pigments and pharmaceuticals.

The marine green alga *Ostreococcus tauri* (class Mamiellophyceae) is the smallest free-living eukaryote known<sup>36</sup> and a representative of an ecologically important group of marine phytoplankton<sup>30</sup>. Its ease of growth in the laboratory, simple cell organisation (one chloroplast, one mitochondrion, a single Golgi body, no cell wall), and its small haploid compact genome (8,000 genes, 13 Mbp)<sup>40</sup>, makes it an attractive model marine alga. *O. tauri* was shown to stably

integrate an overexpression plasmid<sup>320</sup> and was amenable to targeted insertion by homologous recombination by electroporation-mediated transformation<sup>321</sup>. Recently, a simplified method to transform *O. tauri* using polyethylene glycol (PEG) has been described<sup>322</sup>. This approach was first presented in the tobacco plant<sup>323</sup>, also in yeast and bacterial models<sup>324–326</sup>, as well as in protoplasts<sup>327–329</sup>. In algae, the first optimization of PEG transformation was reported in the unicellular red alga *Cyanidioschyzon merolae*<sup>330,331</sup>. Overall, PEG-mediated transformation has shown many advantages, including being simple and fast to perform, with a high transformation efficiency. However, genomic characterisation of how exogenous DNA is integrated using this method, such as determination of the insertion site and copy number has yet to be explored. Currently, the genomic characterisation of transformation in algae has been largely limited to *Chlamydomonas*. Molecular approaches that allowed the copy number of the vector to be determined and the genomic sequences flanking the foreign DNA to be described have been based on the hybridization of radioactive DNA fragments<sup>332</sup>, the plasmid rescue technique<sup>333</sup>, thermal asymmetric interlaced Polymerase Chain Reaction (PCR)<sup>334,335</sup>, restriction enzyme digestion combined with PCR<sup>336,337</sup>, the adaptator linked PCR<sup>338</sup>, hairpin-PCR<sup>339</sup> or *Chlamydomonas* MmeI-based insertion site Sequencing (ChlaMmeSeq)<sup>340</sup>.

Nonetheless, most of these molecular tools are labour-intensive and provide relatively low-resolution information on the DNA insertion events. Neither can they detect the presence of small insertion or deletions (indels), duplications and chromosomal rearrangements, often observed at the junction site of transfer DNA (T-DNA) integration in *Agrobacterium*-mediated plant transformation<sup>341,342</sup>. The development of high throughput Next-Generation Sequencing (NGS) technologies has enabled significant progress by quickly providing complete high-resolution genomes at relatively low cost<sup>343</sup>. NGS has proven feasible as a complementary or alternative method for the characterization of transgene insertion events, having been applied in model plants and animals<sup>344,345</sup> where the presence of DNA structural variations at the insertion site have also been observed<sup>346,347</sup>.

Here, we report the use of Illumina re-sequencing combined with Oxford Nanopore Technology (ONT) sequencing to produce *de novo* genome assemblies of five independent clonal transformant lines of *Ostreococcus tauri* strain RCC1115 in order to characterize insertional events. These data demonstrate that PEG-mediated transformation is a robust molecular tool leading to the integration of a single or few copies of transforming DNA that is stably maintained in the algal genome. Moreover, we provide a reliable demonstration of the

existence of frequent structural variations at the insertion locus, which suggests that *O. tauri* uses a non-homologous repair mechanism for the introduction of foreign DNA into its genome.

## 2. Material and Methods

### *Algal and viral culture conditions*

All progeny clonal lines (around 1,200 transformed and 200 untransformed clonal lines) derived from *Ostreococcus tauri* RCC1115 (Roscoff Culture Collection) strain were maintained in liquid L1 medium (NCMA, Bigelow Laboratory for Ocean Sciences, USA) made with autoclaved offshore seawater (MOLA station: 42° 27' 11" N, 3°8'42" E), diluted with Milli-Q water to give a final salinity of 30 g·L<sup>-1</sup>, and filter sterilized through 0.22 µm filters. Cultures were kept under a photoperiod cycle of 12 h: 12 h light: dark (50 µmol photon m<sup>-2</sup> s<sup>-1</sup> white light) at 15°C, unless otherwise specified.

Cell densities were counted using a Beckman-Coulter Cytoflex flow cytometer (excitation wavelength laser 488 nm) by chlorophyll autofluorescence for algae (detection filter >620 nm) and by SYBR Green I fluorescence for bacteria (detection bandwidth 525–540 nm channel). Cells were fixed using glutaraldehyde (0.25% final concentration) and pluronic acid (0.10% final concentration) for 15 min in the dark, then SYBR Green I (Ozyme ref LON50512) was added for 15 min to stain bacteria.

A selection of twenty viruses specific to *Ostreococcus tauri* with a range of host strain specificities as described in Clerissi *et al.*<sup>231</sup> were grown in liquid culture under the same conditions as previously described. Viral lysates obtained three days post-inoculation were filtered through 0.22-µm pore-size filters, then stored at 4°C for use in virus specificity tests that were performed using the plaque assay described in Clerissi *et al.*<sup>231</sup>.

### *Transformation and screening for host susceptibility or resistance to prasinoviruses*

Insertional mutagenesis was performed by transformation using PEG, as described in Sanchez *et al.*<sup>322</sup>. For the transformation experiments in this study, 50 mL of culture at a density of 50 × 10<sup>6</sup> cells·mL<sup>-1</sup> was used, in order to obtain at least 2,000 transformed clones. The culture was centrifuged for 10 min at 6000× g at 20°C, and the pellet was resuspended and gently mixed with 500 µL PEG 4000 (30% PEG final concentration), and 2 µg carrier tRNA. Then, 10 µg of ScaI-digested pOLK4 DNA<sup>322</sup> was added followed by incubation of the mixture

for 2 min at 20°C. The cells were then diluted into 40 mL of fresh L1 medium, and transferred to a growth chamber for 6 hours to maximize the frequency of transformation. As a control, a cell suspension without PEG treatment and without the addition of DNA was prepared in parallel, described here as the untransformed clonal lines. Finally, both transformed and untransformed control cell suspensions (1 mL volume final) were mixed with a solution of 2.1% (w/v) low-melting-point agarose (1 mL) in sterile Milli-Q (Merck) water (maintained at 60°C in a water bath) and 8 mL of L1 medium at 20°C with 2 mg·mL<sup>-1</sup> G418 (Sigma-Aldrich, A1720) or without antibiotic, respectively, and poured into Petri dishes (diameter 55 mm). The Petri dishes were transferred into a culture chamber at 20°C with 100% humidity and incubated until colonies appeared, usually for at least 3 weeks. Individual *O. tauri* colonies were picked from the semi-solid medium with a sterile pipette tip and transferred into 200 µL liquid L1 medium in 96 well-plates. Over 1,200 transformed clones were screened for by antibiotic (G418) selection and 200 untransformed clones were collected. Secondly, a phenotypic screen for susceptibility or resistance to the virus strain OtV09-578, to which the parental RCC1115 is susceptible, was carried out on all clonal lines (untransformed and transformed). The phenotypic screening was performed in liquid L1 medium in 96-well plates. Control uninfected cultures were mock-inoculated with 50 µL of L1 medium, while a concentrated viral suspension of 50 µL was added to the test cell cultures to a final volume of 200 µL. The dynamics of infection were followed under the growth conditions described above by visual inspection of the green coloration of the culture over 3 weeks. Cultures were considered virus-susceptible if visible loss of green color occurred (indicative of viral lysis of cells) and as potentially virus-resistant candidates if there was no visible color change compared to the uninfected controls.

#### *Illumina and Oxford Nanopore Technologies (ONT) sequencing*

Seven clonal lines including two untransformed controls and five transformed lines (**Table 4**) were selected for genome sequencing. Genomic DNA for each clonal line was extracted with the CTAB protocol, as previously described<sup>348</sup>. DNA quality was evaluated by absorbance at both 260/230 nm and 260/280 nm ratios in a Nanodrop and confirmed by electrophoresis and visualization in a 0.8% agarose gel. DNA was quantified in 1 µL by fluorimetry (Quantus™) using QuantiFluor® dsDNA system kit (Promega, E2670).

The short-insert paired-end libraries (2 × 101 bp) for whole genome sequencing (Illumina Hiseq 3000/4000 system) were prepared with KAPA HyperPrep kit (Roche) with some modifications. As a function of the material availability, 0.07–1.0 µg of genomic DNA

was sheared on a Covaris™ LE220-Plus (Covaris). The fragmented DNA was size-selected with AMPure XP beads (Agencourt, Beckman Coulter), end-repaired, adenylated and Illumina platform compatible adaptors with unique dual indexes and unique molecular identifiers (Integrated DNA Technologies) were ligated. Depending on the library concentration, the adaptor-modified end library was enriched by 4–15 PCR cycles, except for libraries with starting material of  $\geq 1.0$   $\mu\text{g}$  that were further processed without PCR amplification. The libraries were quality controlled on an Agilent 2100 Bioanalyzer with the DNA 7500 assay (Agilent) and quantified by Kapa Library Quantification Kit for Illumina platforms (Roche).

ONT sequencing was performed on the seven transformed lines (**Table 4**). The 1D Genomic libraries for the long-read Nanopore sequencing were prepared using Native Barcoding Expansion 1-12 and 13-24 kits (ONT) and Ligation sequencing kit SQK-LSK109 (ONT) following the manufacturer's recommendations. 2.5  $\mu\text{g}$  of genomic DNA was used in the DNA repair (NEBNext FFPE DNA Repair Mix) and end-repair reactions (NEBNext Ultra II End repair / dA-tailing module reagents) (New England Biolabs, NEB), followed by native barcode ligation with Blunt/TA Ligase Master Mix (NEB) and Native Barcoding Expansion 1-12 and 13-24 kits. After ligation of the barcodes the samples were combined into an equimolar pool of 12 or 13 samples and adapter ligation reaction was carried out using Adapter Mix II (ONT) together with the NEBNext quick ligation reaction buffer 5X (NEB) and Quick T4 DNA Ligase (NEB). The library was purified with 0.4X AMPure XP Beads (Beckman Coulter), washed with Short Fragment Buffer (ONT) and eluted with Elution Buffer (ONT). Nine sequencing runs were performed using R9.4.1 flow cells (ONT) on the GridION instrument (ONT). The MinKNOW GUI (v3.5.4) interface QC (ONT) was run in order to assess the flow cell quality and this was followed by the flow cell priming. After flow cell priming, the pre-sequencing mix comprising Sequencing Buffer and Loading beads (ONT) were combined with the final library and loaded in the flow cell. The sequencing data were collected over 48 hours and the quality parameters of the sequencing runs were further monitored by the MinKNOW platform. The raw data were base called with Guppy 3.2.6.

### *Long-read assembly*

The sequencing data from each of the five transformed clonal lines (**Table 4**) were used to produce a long-read assembly (using ONT reads only) and a hybrid assembly (using both Illumina and ONT reads). The long-read assemblies were obtained with Flye version 2.6<sup>349</sup> with a minimum overlap between reads of 1 Kb (-m 1000), 2 polishing iterations (-i 2) and genome

size 14 mega bases (-g 14m). In addition, we obtained a hybrid genome assembly with MaSuRCA v3.4.1<sup>350,351</sup> using raw Illumina Paired-End libraries (2x101 bp, 517-837x) and the raw ONT reads (N50 7.4-13.4 Kb and 42-96x) to construct mega-reads and assembled them with Flye version 2.5.

#### *Analysis of insertion sites using Illumina short-read sequencing*

The quality of raw reads was checked using FastQC (v0.11.7). For each clonal line, the reads were mapped with the Burrows-Wheeler Aligner (BWA v0.7.17) using maximum exact matches (BWA-MEM) and a minimum seed length of 20 while keeping the other default parameters<sup>352</sup>. Reads were mapped to the chromosomal sequences the *Ostreococcus tauri* RCC1115 (v1.0) reference genome (BioProject number on NCBI: PRJNA337288), including the chloroplast and mitochondrial genomes, (see **Table S1** for the GenBank contig identifiers and accession numbers) as well as the 6Kbp pOLK4 vector sequence. Throughout this study, the chromosome numbers we reported for RCC1115 were the equivalent chromosome numbers to the genome of *O. tauri* strain RCC4221<sup>69</sup>, as whole genome comparisons have revealed a low level of polymorphism ( $\Phi=0.01$ ) between these two strains<sup>70</sup> (**Table S1**). Mapping results were visualized in the Integrative Genomic Viewer (IGV v2.4.16) implemented with the reference genome annotation available from the JGI website ([https://phycocosm.jgi.doe.gov/Ostta1115\\_2/Ostta1115\\_2.home.html](https://phycocosm.jgi.doe.gov/Ostta1115_2/Ostta1115_2.home.html)). To localize the insertion site, first all unmapped reads and mates were removed with Samtools (options: -view -F 12) software (v1.10), then reads that spanned the junction between the chromosome and vector sequence were kept by manual methods *i.e.* reads that spanned the junction were manually parsed from the alignment file (in sam format) by text string searches for the vector identifier.

#### *Analysis of insertion sites in ONT long-read assemblies*

Analyses were carried out on assemblies from both assembly methods (Flye and Masurca), but only the results from the hybrid assembly are presented in view of its better assembly quality at the integration site that was confirmed by PCR resequencing (see Confirmation of structural variations flanking the insertion junctions using PCR below). Assembly quality was also assessed by remapping the Illumina reads against the *de novo* ONT contigs using BWA-MEM (v0.7.17) (options: -k 20). The visualization and the analysis of assemblies was performed using Geneious software (v11.0.3+7). The contig containing the

vector's sequence was identified by a BLASTn (v2.10.1+) search using the pOLK4 sequence as a query (accepting the top high scoring pair with e-value <1e-5 and at least 95% identity). The integration of the vector into the chromosome was visualised by dotplot using Mummer (v4.0.0).

#### *Confirmation of structural variations flanking the insertion using PCR*

PCRs were performed on transformants in which small indels flanking the insertion sites had been detected in the *de novo* hybrid assemblies (T3, T6 and T16). Two sets of primers for each sample were designed to amplify the 5'-end and the 3'-end junctions of each insertion site (**Figure S7**). PCRs were conducted as follows: 2  $\mu\text{L}$  of extracted DNA was added to a 48  $\mu\text{L}$  reaction containing 1X PCR buffer, 1 mM  $\text{MgCl}_2$ , 0.2 mM deoxyribonucleoside triphosphate (dNTP), 0.5  $\mu\text{M}$  of each primer, 0.1  $\text{mg}\cdot\text{mL}^{-1}$  BSA and 1.25 U of *Taq* DNA polymerase (GOTaq® G2 Flexi M7805, Promega), which was amplified in a Mastercycler nexus system (Eppendorf) under thermocycling conditions comprising initial denaturation at 94°C (3 min), 40 rounds of denaturation 94°C (30 s), annealing 56°C (30 s), extension 72°C (30 s) and final extension at 72°C (4 min). PCR products were run on a 0.8% agarose electrophoresis gel in 0.5% TAE (Tris-Acetate-EDTA) buffer and visualized under ultraviolet light after ethidium bromide staining. PCR products were purified using a Wizard® SV gel and PCR clean-up system kit (Promega, A9281) and then sequenced by Sanger sequencing technology. Sequences were cleaned and analysed with Geneious® software (version 2.1).

#### *Bioluminescence assay*

The cellular concentration for each culture was adjusted to  $20 \times 10^6$  cells·mL<sup>-1</sup> in a final volume of 200  $\mu\text{L}$  in each well of a 96-well white plate (Greiner Bio-One™ LUMITRAC, SAS, Les Ulis, France), then D-luciferin (Pierce, Waltham, MA, USA, ref 88293) was added to a final concentration of 100  $\mu\text{M}$ . The plate was placed at room temperature in the dark for 10 min before measuring luminescence with a Victor Nivo 3F Microplate Reader Spectrofluorimeter (PerkinElmer Reader. Ref: 396532, Dominique Dutscher) by digital photon counting. A bar plot was constructed using the ggplot2 data visualization package implemented within the RStudio environment.

### *Data availability*

The parental *O. tauri* strain RCC1115 used for transformation and the seven clonal lines whose complete genomes have been sequenced in this study have been deposited in the Roscoff Culture Collection (<http://roscoff-culture-collection.org/>) with the following strain numbers: T3 (RCC7079), T6 (RCC7080), T12 (RCC7081), T14 (RCC7082), T16 (RCC7083), NT1 (RCC7084) and NT10 (RCC7085). The Illumina-ONT hybrid assemblies of the *O. tauri* lines analysed in this study are available from the European Nucleotide Archive (ENA) under project accession number PRJEB43294. The sequence of the pOLK4 vector is available under GenBank accession number MW598458.

## **3. Results**

### *Illumina paired-end sequencing data and mapping to wild-type and transformed lines*

High throughput sequencing yielded a total of  $90 \times 10^6$  and  $98 \times 10^6$  paired-end reads for untransformed clonal lines NT1 and NT10, respectively, and  $122 \times 10^6$ ,  $75 \times 10^6$ ,  $86 \times 10^6$ ,  $79 \times 10^6$  and  $100 \times 10^6$  for the transformed lines T3, T6, T12, T14 and T16, respectively. By mapping paired-end reads to the reference *O. tauri* RCC1115 v1.0 nuclear and organellar genomes (14.8 Mbp), we obtained between 500–800× average coverage of the nuclear genome (**Table 4**). The coverage for each chromosome was similar to the average coverage of the nuclear genome, whereas the coverage of the mitochondrial and chloroplast genomes was 4-fold more than the nuclear genome (**Table S1**), indicating there were four organelle genome copies per cell.

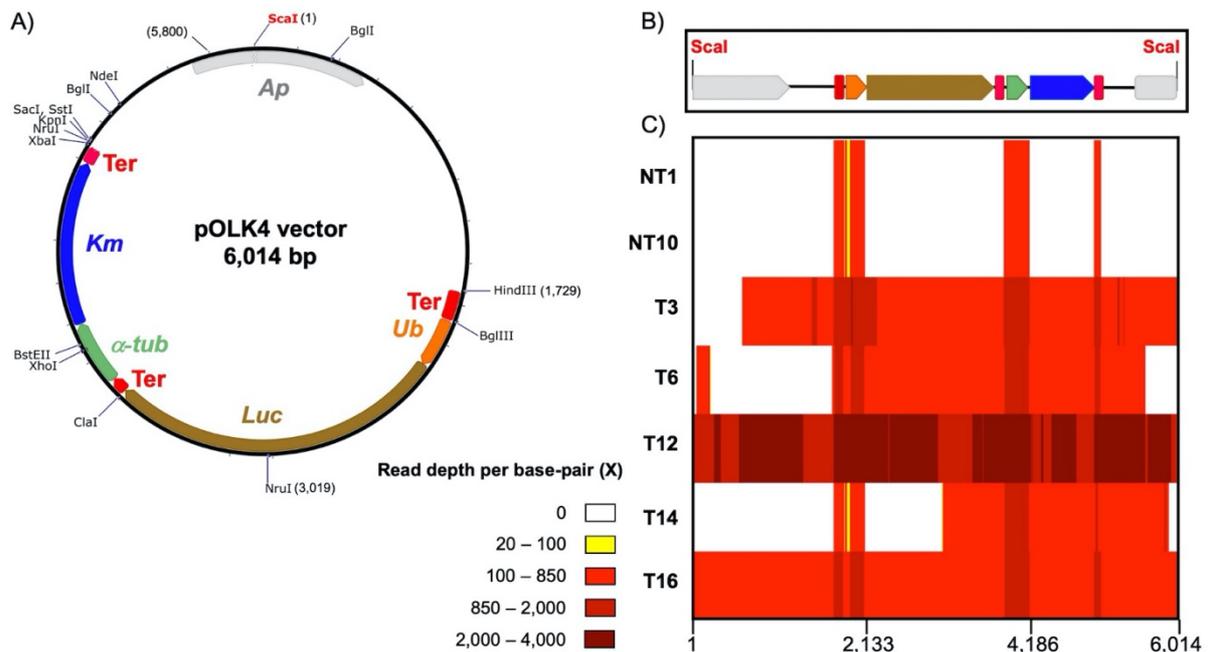
**Table 4.** Illumina sequencing and mapping statistics of reads from clonal lines against the *O. tauri* RCC1115 reference genome. Untransformed (NTx), Transformed (Tx) clones. Total raw reads correspond to mapped and unmapped read pairs. The total length of the reference genome is 14,762,682 bp.

Clones	Total raw reads	Total lengths of the mapped reads (bp)	Average genome coverage* (X)	Read pairs mapped	Proper pairs mapped	Mapped reads** (%)	Average insert size
NT1	90,144,988	9,104,643,788	617	62,018,612	61,681,022	68.8	249.7
NT10	98,800,514	9,978,851,914	676	75,700,644	75,183,524	76.6	275.3
T3	122,387,308	12,361,118,108	837	112,843,908	111,677,142	92.2	267.5
T6	75,540,256	7,629,565,856	517	70,533,906	69,873,506	93.4	265.7
T12	86,367,696	8,723,137,296	591	66,928,772	66,494,262	77.5	250.1
T14	79,501,816	8,029,683,416	544	66,065,344	65,680,576	83.1	260.7
T16	100,660,452	10,166,705,652	688	85,058,636	84,481,042	84.5	277.6

\* Total lengths of the mapped reads / Total reference genome length. \*\* (Reads pairs mapped / Raw total reads) x 100.

*The integrity of the transgene in each line is variable*

To determine the integrity of the transformed DNA, we measured the coverage of the Illumina read pairs that mapped to the whole pOLK4 vector (**Figure 13C**). Note that the pOLK4 construct used for this study contained sequences from *O. tauri* that were expected to recruit genomic reads (**Figure 13**): a ubiquitin promoter sequence found on RCC1115 chromosome 12 (position 173,766–175,000 bp), as well as  $\alpha$ -tubulin promoter and terminator sequences (position 119,519–120,967 bp) from the same chromosome<sup>322</sup>. Reads from both NT1 and NT10 untransformed lines mapped only to the regions of the vector encoding promoter and terminator sequences, confirming the control lines did not contain the vector sequence. Reads from the transformed lines T3, T12 and T16 mapped to more than 90% of the vector sequence length, while reads from T6 and T14 covered less than 75%. This showed the integrity of the transformed vector was variable; in some lines it was complete, whereas in others, portions of the vector sequence were lost. Specifically, deletions occurred mainly at the linearized vector extremities, with the exception of line T6 where a fragment of the vector's 5'-end containing a part of the ampicillin resistance gene was retained (**Figure 13C**). Consistent with the selection of transformants on G418 (Km), no deletions occurred in the G418 antibiotic resistance selectable marker gene.



**Figure 13.** Sequence map and read coverage of the pOLK4 vector. **(A)** Map of the circular pOLK4 vector. Gray arrow: Ap, ampicillin resistance gene; blue arrow: Km, kanamycin (G418) antibiotic resistance gene used as a selectable marker for algal transformation; brown arrow: Luc, luciferase used as a luminescence reporter gene; orange arrow: ubiquitin promoter; green arrow:  $\alpha$ -tubulin promoter; and red block: Ter,  $\alpha$ -tubulin terminator. Positions of restriction sites are shown. **(B)** Map of the linear pOLK4 vector that the algal cells were transformed with highlighting the Sca-I restriction site (red) used to linearize vector DNA. Colors of the gene maps as in **(A)**. **(C)** Heatmap of short-reads from untransformed (NTx) and transformed (Tx) lines mapped against the linearized vector, showing no read coverage (white) to high read coverage (dark red).

*Most of the transformants (4 out of 5) carried one copy of the transforming DNA*

To quantify the number of insertions of the vector sequence in each transformed line, we used an NGS strategy to compare the coverage of transformed DNA to the coverage of the nuclear genome. Effective coverage of the pOLK4 vector was 649 $\times$ , 306 $\times$ , 1,947 $\times$ , 208 $\times$  and 531 $\times$  for T3, T6, T12, T14 and T16 lines, respectively (**Table S2**), which was in the same order of magnitude as the average coverage of the nuclear genome (**Table 4**). However, given the heterogeneous integrity of the pOLK4 vector, to determine the copy number of the transforming DNA as accurately as possible, we calculated the ratio of read coverage of the G418 selectable marker to the coverage of a single-copy housekeeping gene, DNA directed DNA polymerase II, family B (OtaPolB) (**Table 5**). We estimated that four transformants (T3, T6, T14, T16)

harbored one copy of the transforming DNA in their genomes, while T12 carried four copies (**Table 5**).

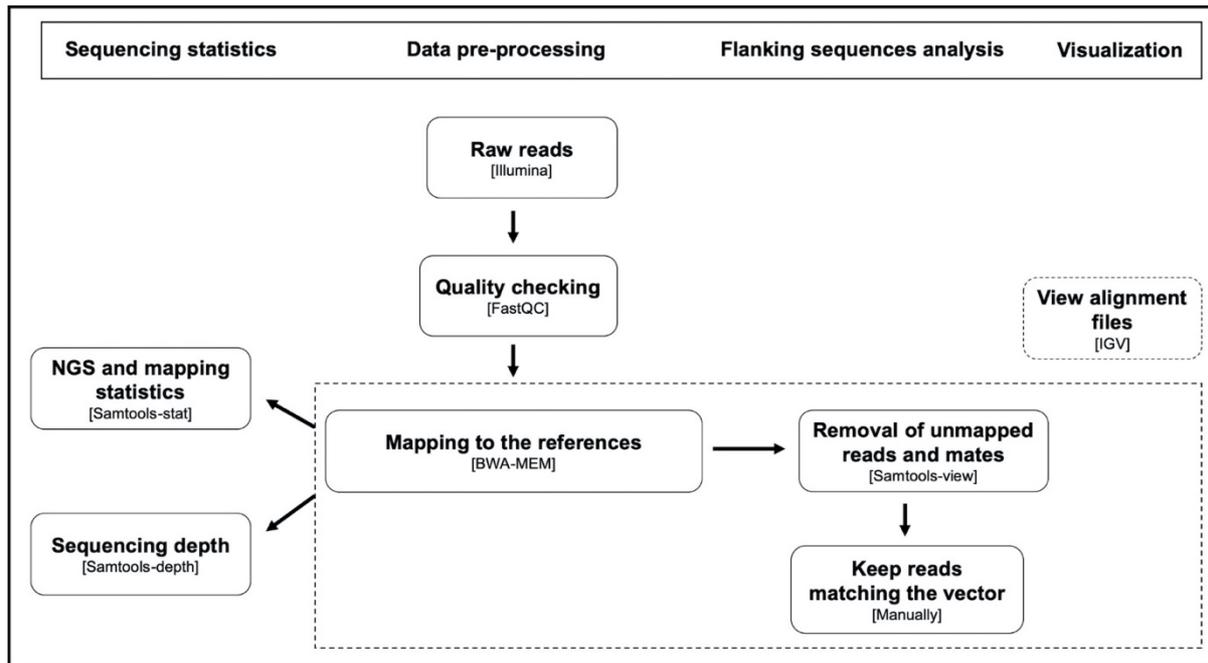
**Table 5.** Mapping summary of short-reads back to vector sequence (pOLK4) and estimated insert copy number in each transformed (Tx) clonal line. The vector selectable marker encodes resistance to the antibiotic G418 (809 bp). The housekeeping gene is the DNA directed DNA polymerase II gene (BE221DRAFT\_187986) (3,527 bp), present in a single copy in the reference genome.

Clone	Sequence feature	Total lengths of the mapped reads (bp)	Average coverage* (X)	Average copy number **
T3	Vector G418 gene	680,437	841	1.01
	Housekeeping gene	2,944,554	835	
T6	Vector G418 gene	441,673	546	1.06
	Housekeeping gene	1,813,960	514	
T12	Vector G418 gene	1,833,049	2,266	4.68
	Housekeeping gene	1,707,102	484	
T14	Vector G418 gene	389,557	482	0.99
	Housekeeping gene	1,718,919	487	
T16	Vector G418 gene	512,979	634	1.04
	Housekeeping gene	2,143,927	608	

\* Total lengths of the mapped reads / total gene sequence length. \*\*Average coverage of vector G418 gene / Average coverage of housekeeping gene.

*Vector insertions occurred at a single genomic location in each transformant*

The insertion events of the pOLK4 vector were characterized according to the bioinformatic workflow presented in **Figure 14**, based on the strategies used in genetically modified crops<sup>344</sup> and also as described by Yang *et al.*, in module 1<sup>353</sup>. Firstly, the quality of the raw read data was checked using the FastQC program (v0.11.5), confirming that the quality was sufficiently high to carry out analyses on the raw reads. Secondly, we mapped the short-reads from the transformed lines against the reference genome and the vector sequence with maximal exact matches. Next, we manually retrieved only reads of a pair mapping at one end to the reference genome and at the other end mapping to the vector sequence, which we called “junction pairs”, to locate the vector insertion site. Finally, we visualized the coverage around the predicted insertion site to indicate any local insertions and/or deletions of the genome.



**Figure 14.** Procedure for finding the insertion sites of transforming DNA in the host genome using short-read NGS data.

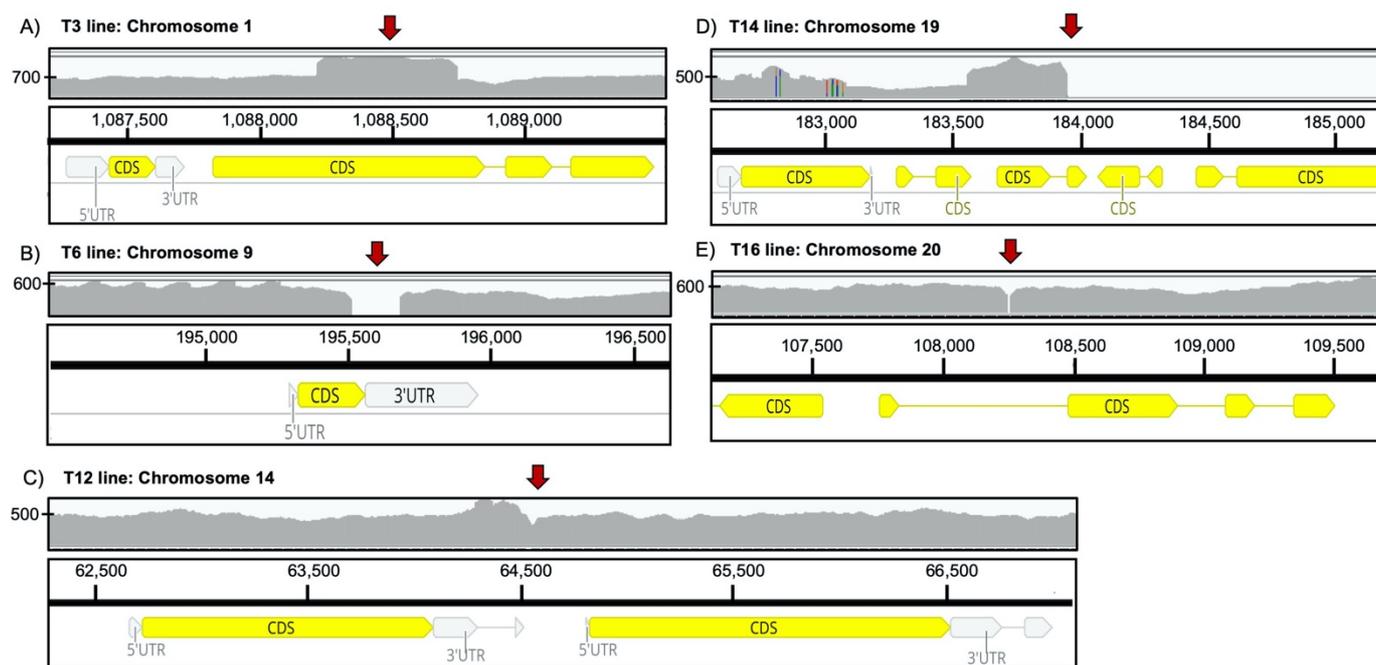
For lines T3, T6 and T16, the junction pairs on both ends of transforming DNA were detected, enabling the characterization of the insertion site. In line T3, the vector insertion was predicted to be on chromosome 1 in an exonic region of a gene coding for a methionine aminopeptidase 2 (MetAP2), where the insertion was flanked by a small duplication of the genomic sequence (**Figure 15A & 15F**). The vector insertion site of the T6 lineage was located on chromosome 9 inducing a deletion of about 200 bp from the 3'-extremity and part of the 3'-UTR of a coding sequence with no known function (**Figure 15B & 15F**). In line T16, the vector insertion mapped to a single locus on chromosome 20, which also produced a deletion of 10 bp in the intron of ribosomal protein L19 (**Figure 15E & 15F**).

In lines T14 and T12, although we could locate the insert, we could not fully determine the structure of the insertion site from this strategy using short-read mapping. For the T14 line, only junction reads spanning the left-hand side of the insertion were found, which were located on chromosome 19 within a partial gene encoding a type I polyketide synthase (PKS) (**Figure 15D & 15F**). The absence of junction reads to the right-hand side of the insertion site was likely due to deletion of the equivalent region of chromosome 19 in T14, evidenced by the lack of coverage downstream of the insertion (**Figure 15D**). This chromosome in *O. tauri* is termed the Small Outlier Chromosome (SOC) due to its lower GC content and higher proportions of

repeats and species-specific genes than the other chromosomes, but nonetheless is conserved in all the sequenced genomes of Mamiellales<sup>354</sup>. Furthermore, altered transcription and hypervariability in size and genomic content of the SOC are linked to resistance to viruses in *Ostreococcus* species<sup>48,70,71,310</sup>. In keeping with the SOC's higher propensity for structural variation, it is likely that DNA integration has promoted a deletion of more than 100 kb of the SOC, as the deletion was adjacent to the insertion locus and short-read mapping in the other six lines do not show similarly large structural variants (**Figure S1**). We observed a loss of 31 predicted unique genes encoded by the wild-type (**Table S3**), however ~35% of the deleted region comprised repeat sequences located elsewhere on chromosome 19 (**Figure S1**), including nine presumably redundant genes. Finally, the T12 transformant was estimated to bear four vector copies (**Table 5**), and it accordingly recruited the highest number of junction pairs, all of which were located on chromosome 14 at a single locus spanning the left-hand end of the insert. This insertion fell in an intergenic region between a small nuclear ribonucleoprotein gene (Prp4p) involved in RNA splicing and Plus-3 domain-containing protein gene (Rtf1) (**Figure 15C & 15F**). Rtf1 forms part of the Paf1 complex (Paf1C) that consist of five proteins (Paf1, Rtf1, Cdc73, Ctr9, and Leo1) in eukaryotic systems where it functions with RNA polymerase II in transcriptional processes and is also associated with RNA polymerase I<sup>355</sup>. Overall, as all junction pairs were located in the same region on chromosome 14 at approximately four times the average read coverage; this suggested multiple insertions occurred at the same locus.

#### *Vector integration is associated with structural variation at the insertion site*

In order to gain insight into the structural changes at the vector integration site, we re-sequenced all of the transformed clonal lines using ONT and performed de novo assemblies with the long reads, as well as hybrid assemblies with both long and short-reads. The alignment of the assembled contigs back to the chromosome of the reference strain where the insertion was predicted from the short-read analysis confirmed the exact location of the integration site. Likewise, aligning the contigs containing the insert against the pOLK4 vector (**Figure S2**) confirmed the variable completeness of the vector sequence that we observed from the Illumina short-read data (**Figure 13C**). The hybrid assembly enabled the description of structural variations, such as insertions, duplications, deletions, and inversions, in both the chromosome and the vector (**Figure 16**), which could not be resolved from the short-read mapping strategy.



F)

	Clonal line	Chromosome	SequenceIDs	Location	CDS putative function	Genbank locus tag
A)	T3	1	BE221ch_0100	1,088,227-1,088,742	Methionine aminopeptidase 2 ( <i>MetAP2</i> )	BE221DRAFT_62303
B)	T6	9	BE221unitig_03	195,500-195,700	Unknown function	BE221DRAFT_189169
C)	T12	14	BE221unitig_17	64,500	Putative small nuclear ribonucleoprotein ( <i>Prp4p</i> )	BE221DRAFT_11509
					Plus-3 domain-containing protein ( <i>Rtf1</i> )	BE221DRAFT_192465
D)	T14	19	BE221ch_1900	183,900	Polyketide synthase partial	BE221DRAFT_64653
E)	T16	20	BE221unitig_33	108,000-108-500	Ribosomal protein L19	BE221DRAFT_47391

**Figure 15.** IGV (Integrated Genome Viewer) visualization of read coverage and identification of insertion sites. Genomic maps of the vector insertion site in (A) T3, (B) T6, (C) T12, (D) T14 and (E) T16 transformed clonal lines. The coverage plot is shown in the top panel (Y-axis the coverage (X) scale), the locations of insertion sites are shown with red arrows and the gene annotation of the reference genome is shown in the bottom panel. (F) Exact location of the sites of vector integration and putative functions of genes surrounding the insertion site.

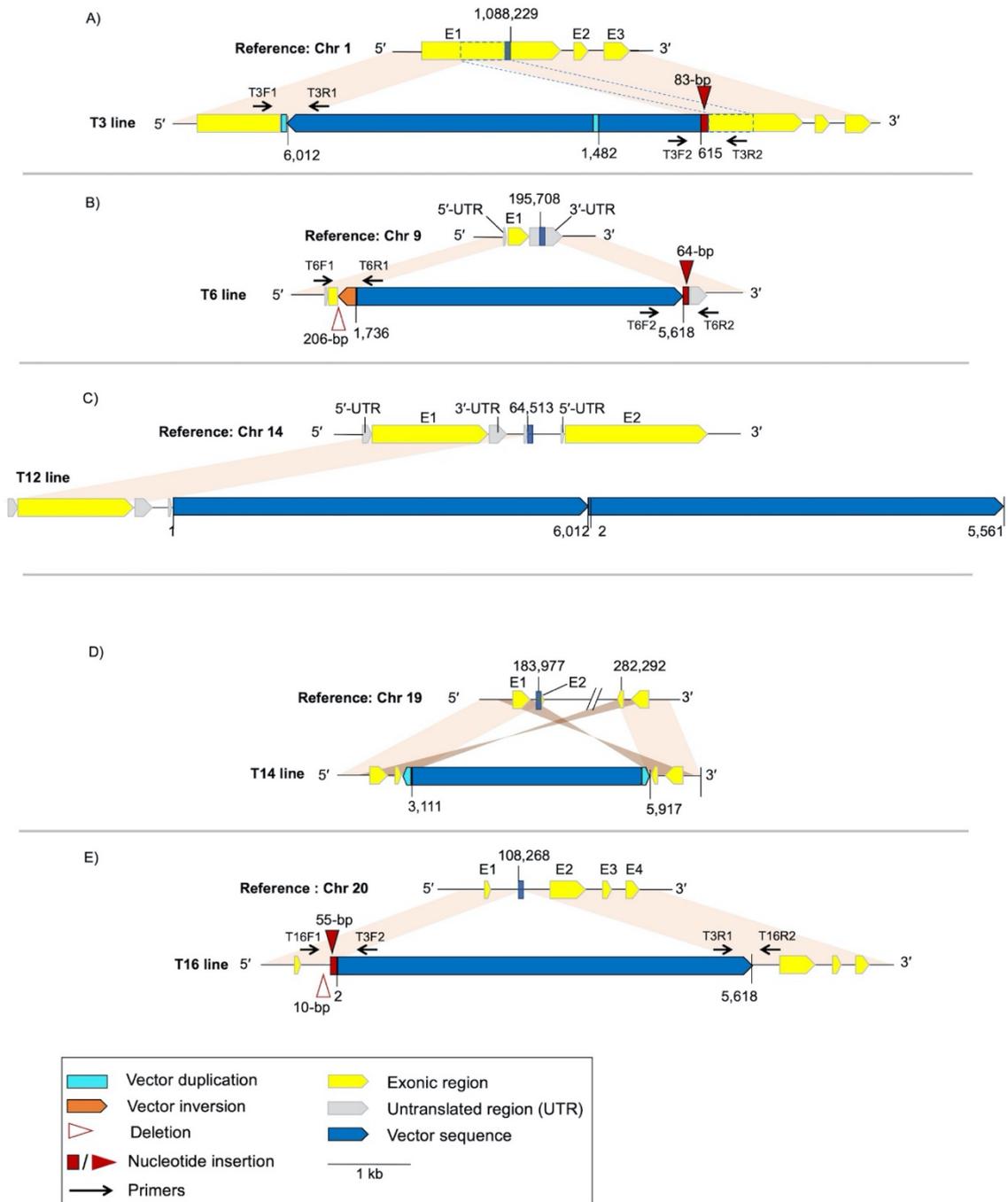
In the T3 line, the insertion of the vector was accompanied by an 83-bp insertion of unknown origin at the vector's 5'-end and a 62-bp duplication of the non-coding region of the pOLK4 vector sequence (position 1,482–1,543 bp) at the 3'-end. The inserted sequence was flanked by a 518-bp direct duplication of part of the *MetAP2* gene at the integration (**Figure 16A**).

In the T6 line, several variations were observed (**Figure 16B**): a 206-bp deletion of 3'-end of the gene at the integration site; a 64-bp insertion of sequence of unknown origin at the vector's 3'-end; and rearrangement in the vector sequence itself, with a 171-bp inversion corresponding to part of the ampicillin resistance gene (position 40–210 bp) at the vector's 5'-end.

In the T16 line we observed a 10-bp deletion of genomic sequence and a 55-bp insertion of viral origin (a partial gene encoding an integrase Genbank protein ID: XP\_022839977.1) at the vector's 5'-end (**Figure 16C**).

In the T14 line the integration of the vector was more complex than in the other lines, involving a 90-bp inverted duplication of the part of the ampicillin resistance gene from the vector's 3'-end (position: 5,828–5,917 bp), and the insertion of the vector occurring between two 414-bp inverted repeat sequences of chromosome 19 origin (**Figure 16D**). In the reference chromosome 19, these 414-bp inverted repeat sequences are separated by ~98-kb and it is primarily this intervening sequence that has been deleted in T14. We observed the vector integration occurred adjacent to the telomeric sequence at the end of the contig, indicating that the integration locus is located at the chromosome end in T14 and the local structure of the vector insertion was resolved in the assembly. As the region homologous to the vector integration site is not at the chromosome end in the reference (**Figure S2**), this further indicates a larger rearrangement event occurred within chromosome 19 of T14 that placed this region at the chromosome extremity.

In the T12 line, although four copies of the vector have been estimated from the Illumina short-read coverage, the hybrid assembly showed tandem insertion of the vector in only two copies (**Figure 16E**). However, as the vector sequences were located at the contig's 3'-end (**Figure S2C**) and alignments of ONT long-reads against partially polished assemblies showed compressed coverage, the tandem vector repeats likely led to an assembly breakpoint and thus the insertion structure was not completely resolved. From examination of the long-read alignments it was clear that the vector has been inserted multiple times in tandem with at least one vector copy inserted in the inverse orientation.



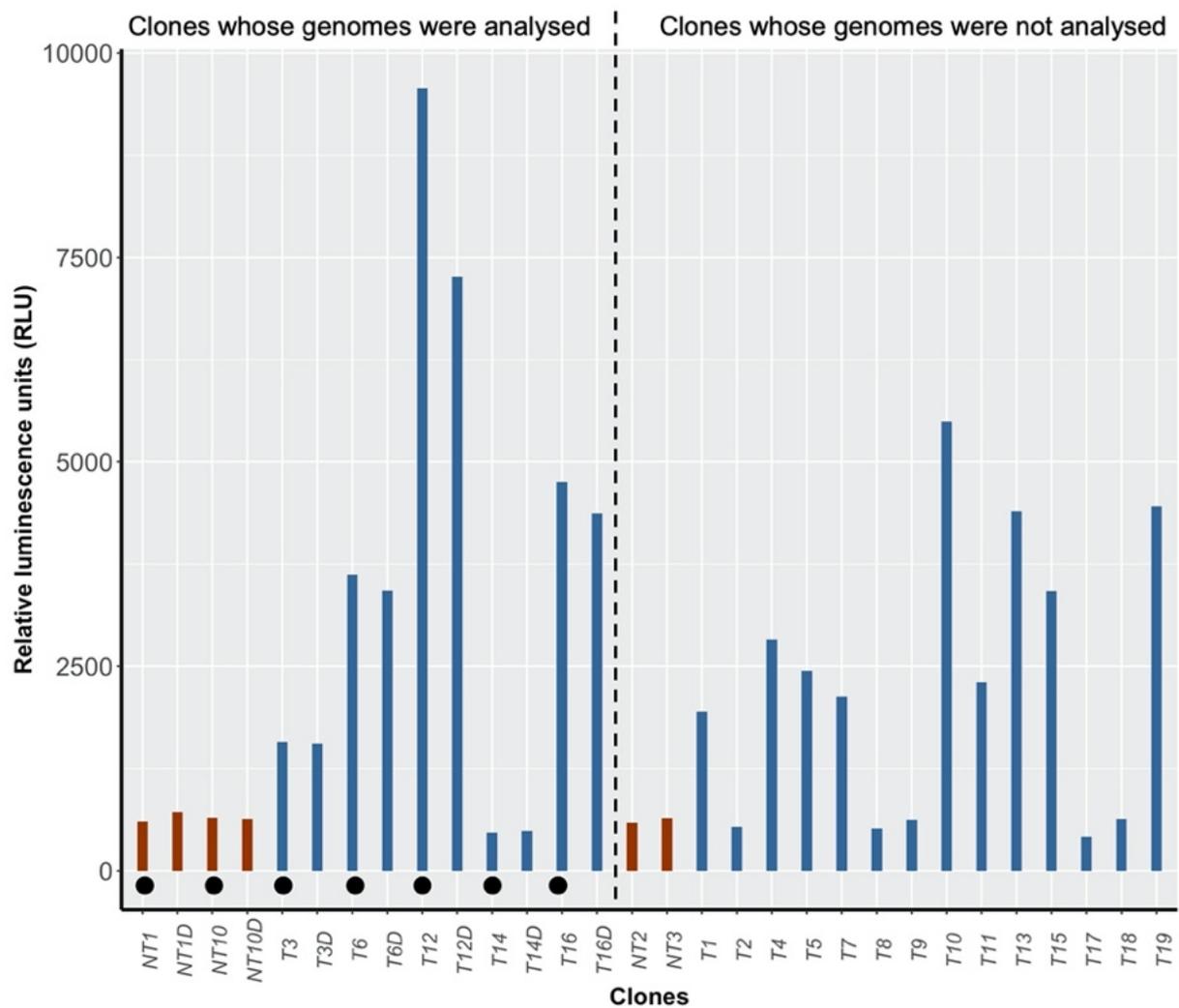
**Figure 16.** Schematic illustration of the vector integration sites and associated structural variations in contigs assembled from transformants compared to the reference genome. Homologous sequences between the reference and the clonal lines are connected by colored blocks (light orange; same orientation; dark orange; inverse orientation) (A) T3 line. (B) T6 line. (C) T12 line, the entire vector sequence occurs twice in tandem at the end of the contig. (D) T14 line. Note that the map of the reference chromosome 19 sequence shows two distant regions separated by a break. (E) T16 line. Primers (black arrows) targeting the flanking regions have been designed to confirm the insertion-deletion events (red) in the T3, T6 and T16 lines.

To verify the presence of the small indels flanking the vector integration sites, we designed primers to the junctions between the vector and chromosomal sequences (**Figure 16**) and performed PCR. The amplification products (**Figure S5**) were subjected to Sanger sequencing that confirmed the sequence of these small indels for T3, T6 and T16 lines (**Figure S6**) corresponded exactly to the sequence in the hybrid assemblies.

#### *Inserted DNA is transcribed in transformed lines*

Thousands of clonal lines, both transformed and untransformed, were generated in this study from which we selected the 7 clonal lines whose genomes were sequenced. A selection of these clones was maintained in liquid-culture. In order to enlarge our sample size and to have a broader view of the level of expression of the inserted DNA, we included an additional 14 transformed and 2 non-transformed clonal lines to this analysis. The expression of the inserted DNA was measured from the vector's luciferase reporter across 23 independent clonal lines (**Figure 17**). Control lines exhibited a low basal level luminescence, establishing background levels in the absence of luciferase expression. In order to assess the stability of integration of the vector over time, we thawed a subset of seven lines that had been cryopreserved six months earlier (NT1D, NT10D, T3D, T6D, T12D, T14D, T16D) and assayed their luciferase expression in parallel with the 23 lines, which had been maintained by repeated subculture. The expression of luciferase was similar between cryopreserved lines and lines that had not been frozen (**Figure 17**), indicating vector integration and reporter expression were stable over this 6-month timeframe. We observed a high variation in luminescence in transformed lines demonstrating highly variable protein expression (**Figure 17**). Six out of 19 transformed lines (~32%) showed background luminescence levels in agreement with previous work that found several transformants lacked reporter expression<sup>322</sup>. We observed the absence of luciferase expression for T14 line, which is consistent with the deletion of the reporter gene in this line that we determined in the NGS analysis. This result suggests that the lack of luciferase expression previously observed was similarly due to deletion of the reporter gene. We established that luciferase activity was not related solely to the number of copies of the vector, as evidenced by the range of luminescence in the T3, T6 and T16 lines, which each contained a single copy of the vector. In these lines, the variation in luminescence was likely due to variation in gene regulation specific to the genomic context of the integrated vector. Nonetheless, the T12 line showed the highest luciferase activity and also had the highest

number of vector copies, as determined by the genome resequencing, which provides some evidence that vector copy number also contributes substantially to reporter expression levels.



**Figure 17.** Firefly luciferase assay of transgenic clonal lines. Emission of light measured expressed in relative luminescence units (RLU) from nineteen transformed clonal lines (blue bars) and four untransformed clonal lines as a control (red bars). The stability of the vector within the host genome was analyzed in the seven lines studied in this study by measuring the fluorescence emitted from samples (NT1D, NT10D, T3D, T6D, T12D, T14D, T16D) thawed 6 months after transformation. Clonal lines studied in this work are distinguished by filled black circles. Raw data are available in supplementary data (**Table S4**). Each sample was measured once.

## 4. Discussion

*What could be the mechanism underlying the random insertion of DNA by PEG-mediated transformation?*

External mutagenic agents, such as ultraviolet (UV) radiation or ionization radiation (IR), which are naturally present in the environment<sup>356</sup>, can severely damage eukaryotic cells. The DNA lesions are detected by DNA damage response (DDR) pathways, which include direct repair, base excision repair (BER), nucleotide excision repair, cross-link repair and double-stranded DNA break (DSB) repair<sup>357,358</sup>. DSBs are one of the most severe types of DNA damage experienced in the lifetime of a cell. To survive and maintain homeostasis, two main pathways of DSB repair are known. First, homologous recombination (HR) is often associated with error-free repair, since it uses a homologous chromosome as a template. In contrast, the second pathway, non-homologous end joining (NHEJ), also known as illegitimate recombination, may align a few complementary bases of "microhomology" that often leads to errors such as deletions, duplications, translocation or to the addition of sequences from elsewhere at the break site<sup>359-361</sup>. The choice between which of these pathways is used is highly regulated and depends on the type of cell, the stage of the cell cycle<sup>361,362</sup> and the type of genotoxic stress.

In *Ostreococcus*, DDR mechanisms have not been directly characterised. One study achieved targeted insertion of foreign DNA by transforming constructs containing sequence homologous to the target locus by electroporation, indicating that HR repair is functional in *O. tauri*<sup>321</sup>. In our study, DNA insertion occurred in regions of the genome that had no homology with the vector, indicating that insertion does not primarily occur by HR, but rather by a process akin to NHEJ. This is consistent with observations of transformation methods that similarly do not use homologous regions, such as *Agrobacterium*-mediated T-DNA insertion in plants<sup>363</sup>. In order to gain insight into the DDR pathways that may be utilized in *O. tauri*, we searched for genes known to be required in DDR in related and model organisms. Forward genetic experiments combined with a screen for sensitivity to mutagen treatments in *Chlamydomonas* have implicated the following genes in DDR: the *REX1* gene<sup>364</sup>, cytosolic thioredoxin h1 (*Trxh1*) (required for BER<sup>365</sup> and also observed in yeast<sup>366</sup> and mammalian cells<sup>367</sup>, DNA polymerase zeta, DNA polymerase theta, SAE2/COM1 endonuclease, RaD17 and ERCC1<sup>339</sup>. The RaD17 homolog in *Arabidopsis* was also associated with non-homologous DSB repair<sup>368</sup>, as was DNA polymerase theta in *Arabidopsis*<sup>369</sup>. Other proteins known to play

a role in NHEJ repair include the Ku70/Ku80 heterodimer and DNA ligase IV in plants<sup>370,371</sup> and in yeast<sup>372,373</sup>. We confirmed by homology searches on the comparative genomics platform Pico-PLAZA<sup>374</sup> that *O. tauri* encodes four of the above-mentioned plant and algal gene families that could mediate NHEJ: the RaD17 complex (Genbank accession: OUS48190.1), which shares 20% amino acid identity with *Arabidopsis thaliana*; DNA polymerase theta (OUS43494.1); DNA ligase IV (OUS46583.1); and the Ku70/Ku80 heterodimer (OUS47173.1).

PEG is thought to enable transformation by modification of membrane fluidity and the aggregation of phospholipid vesicles<sup>375</sup>, as well as enhancing transformation by altering transcriptional and metabolic response in *Saccharomyces cerevisiae*<sup>376</sup>. Although high M.W. PEG at high concentrations was found to reduce growth of *O. tauri* cells<sup>322</sup>, to our knowledge, it is not mutagenic. In animals or plants, it has been shown to be completely safe to use, even for medical applications in humans<sup>377</sup>, whereas in plants it has been used for transformation and regeneration of healthy individuals<sup>378</sup>. Although the exact mechanism by which foreign DNA is incorporated into the genome during PEG treatment in *O. tauri* remains to be elucidated, the combination of short and long read sequencing in this study has shed light on genomic events involved. Our study suggests that DNA integration mediated by PEG occurred randomly, as none of our transformed lines shared the same integration site. This observation of random insertion is consistent with previous work<sup>322</sup> where vector integration occurred in different chromosomes in 11 clonal transformants. In addition, vector insertions were accompanied frequently by structural variations at the insertion site. We found that most of the insertion events of non-vector DNA observed may have been produced *de novo* synthesis from the cell's recombination/repair systems, since sequences large enough to be tested by BLAST (from 55 bp to 83 bp) bore no similarities to those in common public databases. Taken together, we propose that integration of foreign DNA in *O. tauri* occurs primarily *via* a mechanism similar to NHEJ to repair spontaneous damage arising during cell growth.

*Does the integration of the inserted DNA have an impact on the deregulation of gene expression?*

Since the genome of *O. tauri* is very compact, with only about 8,000 genes and short intergenic regions<sup>40</sup>, and these cells are haploid with few gene duplications, it is likely that a large proportion of insertion events will have been lethal and cannot be characterized using this approach. However, we easily produced thousands of transformants, demonstrating many

insertions are clearly viable and transformation may be used to introduce DNA in different genomic locations. The integration of foreign DNA using PEG-mediated transformation seemed to occur randomly in the host genome, mainly as a single or low copy. We showed the vector may reside in coding regions, such as in the T3 line where integration was in the coding region of the *MetAP2* gene, perhaps leading to silencing, or altered function in the interrupted gene or altered expression in adjacent or disrupted genes. Therefore, this method of PEG-mediated transformation holds promise for generating collections of insertional mutants. In particular non-essential genes, such as those that may be induced by attack of certain pathogens or by exposure to specific environments, may be disrupted with little or no apparent effect on cell growth in culture.

In the context of this work, we performed PEG-mediated random insertion as a possible approach to answer our initial biological question, which was to elucidate the molecular basis of resistance to prasinoviruses, known to exist from previous work in *O. tauri*<sup>310</sup>, by interrupting the genes potentially involved in this antiviral immunity system. For this reason, the choice of the host strain was based on a previous phenotypic screen for viral infection in thirteen wild-type strains of *O. tauri* against forty prasinoviruses<sup>231</sup>. We chose *O. tauri* RCC1115 since its genome was completely sequenced<sup>70</sup> and it showed a range of susceptibility to different strains of prasinoviruses. The choice of candidate transformants in this study was based on phenotypic screening for susceptibility to a panel of viruses, which initially indicated these lines had altered patterns of viral susceptibility (**Figure S3**). However, we found that the phenotype of resistance or susceptibility to viruses in clonal lines of RCC1115 was too unstable for use in the longer culture periods required for re-growth of strains (**Figure S3**). In addition, more detailed observations carried out on host cell dynamics following viral infection (viral strain OtV09-578) revealed rapid fluctuations in growth in liquid culture (**Figure S4**). Importantly, we also showed that growth of transformants that were not virus infected was not substantially affected (**Figure S4**), indicating that the vector insertion at these loci did not impact cell growth.

An experimental evolution experiment in the related alga *O. mediterraneus* showed that independent clonal culture lines started from single cells switched between resistant and susceptible phenotypes and this was linked to changes in size of the SOC<sup>71</sup>. Similarly, rapid changes in SOC size have been observed upon acquisition of resistance to prasinoviruses in *O. tauri* strain RCC4221<sup>308</sup> and these resistant lines also contained a minority population of virus-susceptible cells<sup>310</sup>. These works are consistent with our observation of instability in the

resistance phenotype in *O. tauri* RCC1115. Furthermore, it suggests that if antiviral immunity depends on the rapid generation of hypervariability in the SOC in the genus *Ostreococcus*, clones of RCC1115 would similarly display changes in SOC size. Indeed, an analysis of the karyotypes of seven transformed and three untransformed clonal lines of RCC1115 revealed genomic plasticity between all the lines independent of the transformation (data not shown); we could exclude the hypothesis that variations in chromosome size were caused by the PEG treatment, since the control (NT) lines were also affected. Moreover, the NGS data provide a clear description of rearrangements occurring in chromosome 19 leading to a large deletion of about 100 kb in line T14. Although we cannot go further in the characterisation of the phenotypic effects correlated to each insertion event in the transformed lines, we could exclude a direct link between the vector insertion and the phenotypic instability of viral resistance/susceptibility during viral infection, which was better explained by a distinct mechanism of genomic plasticity.

#### *PEG-mediated transformation as a robust and powerful biomolecular tool*

To obtain a high transformation efficiency, optimal conditions were chosen according to Sanchez *et al.*<sup>322</sup>. We similarly achieved high transformation efficiency in our experiment ( $1.84 \times 10^5$  transformed cells/ $\mu\text{g}$  of DNA), showing that PEG-mediated transformation is robust and reproducible. Even with high transformation efficiency, we found that DNA insertion by PEG in *O. tauri* usually led to the stable integration of a single copy per clone in the cells' genome. Single or low copy numbers of inserts were also reported using an *Agrobacterium*-mediated transformation method in plants<sup>345,379</sup> and in *Clamylomonas reinhardtii* algae<sup>380</sup>. Having single vector insertion events will facilitate future work aiming to characterise insertional mutants of a specific gene of interest or the effects of introducing an exogenous gene. Specifically, we showed that stable expression of the luciferase reporter introduced by PEG-mediated transformation was due to integration of the vector DNA and not transient expression, which lays the foundation for introduction of other exogenous genes of interest. Furthermore, we also showed lack of reporter gene expression was most likely due to spontaneous deletion of the reporter gene during DNA integration rather than gene silencing in *O. tauri*, which is in keeping with the paucity of heterochromatin in this organism<sup>381</sup>. Our genomic analyses also brought to light the existence of frequent indels around the vector insertion site, as well as in the vector, which should be taken into consideration, especially when aiming to precisely insert a sequence for expression of a foreign gene. Based on all of this new

knowledge about PEG-mediated transformation, we envisage using this approach as a robust and powerful biomolecular tool to go further in the study of various functional mechanisms, which has particular promise in biotechnological applications. Engineering of oleaginous algae to enhance their biomass and the bioactivity of lipid production was reported in *Nannochloropsis* sp.<sup>382</sup>, *Chlamydomonas reinhardtii*<sup>383</sup>, and also in *Chlorella pyrenoidosa* for the production of biofuel and bioplastic<sup>384</sup>. Recently, a study showed that *Ostreococcus tauri* produced and secreted high-lipid content droplets under conditions close to standard laboratory conditions<sup>385</sup>, highlighting its potential as a marine algal model. Another study has shown some early indications that *O. tauri* extracts were protective against cytotoxic effects of polyaromatic hydrocarbons in human cell lines<sup>386</sup>, further indicating a potential avenue of development in bioactive compounds' production. We propose that PEG-mediated transformation could be used as a molecular genetic approach to determine the genetic basis for a phenotype or also as a biotechnological tool. *Ostreococcus tauri* is a representative species of the order Mamiellales, including the genera *Ostreococcus*, *Bathycoccus* and *Micromonas*, that are distributed worldwide in our oceans<sup>387,388</sup>. The population densities of these species in different oceanic regions are considered to be indicators of climate change and nutrient availability<sup>388-390</sup>, and will be important for monitoring and modelling the evolution of planetary ecosystem health and biodiversity.





## CHAPITRE II – APPROCHE INTEGRATIVE DE LA RESISTANCE ANTIVIRALE CHEZ OSTREOCOCCUS TAURI ENTRE GENOME ET PHENOTYPE

### **Objectifs du chapitre :**

Ce chapitre vient dans le prolongement des travaux du chapitre I dans le but d'approcher les bases moléculaires de la résistance aux *Prasinovirus* dans le système d'immunité antivirale chez la souche *O. tauri* RCC1115. Dans ce cas, nous avons conduit des analyses expérimentales de criblage phénotypique à l'infection virale, en essayant de relier le génotype au phénotype immunitaire.

### **Contributions de l'auteurs :**

Dans ce chapitre, j'ai conduit l'ensemble des travaux expérimentaux ainsi que les analyses génomiques.

**Données supplémentaires :** Les documents suivants sont retrouvés dans la partie des données supplémentaires à la page 209

## **Résumé :**

Ce travail a révélé l'importance des choix méthodologiques dans la sélection des lignées clonales à l'infection. En effet, la méthode employée dans la sélection des lignées clonales à l'infection virale en culture liquide en plaque de 96 puits ne permettait pas de discriminer le « vrai » phénotype immunitaire observé. L'ensemble des cultures étaient sensibles au virus OtV09-578, bien que certaines montraient une infection plus faible avec une augmentation très rapide de la résistance. Nous avons mis en évidence que cette fluctuation rapide pouvait être corrélée à la forte plasticité observée au sein du génome. D'autre part, nos résultats montrent indirectement que les éléments transposables (ETs) pourraient avoir un rôle crucial dans les changements rapides phénotypiques face à l'infection virale et suggèrent qu'ils pourraient potentiellement aider dans l'adaptation aux changements environnementaux. Par un mécanisme encore inconnu, les ETs pourraient affecter l'expression des gènes cellulaires impliqués dans l'infection ou dans l'immunité virale. Des études futures devront être envisagées pour montrer si une telle diversification phénotypique peut conférer finalement un avantage adaptatif et si les ETs sont moteurs de cette adaptation.

## 1. Introduction

A ce jour, il est bien établi que les virus marins sont des contributeurs importants sur le plan écologique dans les océans à l'échelle planétaire. En effet, ces entités biologiques sont responsables de la majeure partie de la mortalité des populations planctoniques et bactériennes dans les environnements aquatiques. Au travers des infections lytiques, les virus contribuent au transfert du carbone et de nutriments sous forme particulaire et dissoute au travers de la boucle microbienne vers les réseaux trophiques inférieurs, influençant ainsi les cycles biogéochimiques et le climat<sup>96,391</sup>. Par ailleurs, la prédation virale favorise la diversité des hôtes en réduisant les populations les plus abondantes et en permettant la croissance des espèces les plus rares<sup>268</sup>. Cette coexistence stable a suscité la question de comment l'équilibre entre les populations hôtes et virales peut être maintenu<sup>272</sup>. Cette coexistence a pu être mise en évidence chez plusieurs espèces d'algues, notamment chez l'algue verte d'eau douce *Chlorella variabilis*<sup>210</sup> chez l'haptophyte *Emiliana huxleyi*<sup>291</sup> mais également chez les Mamiellales<sup>222,293</sup>.

La dynamique de course aux armements peut être une des réponses à ce phénomène<sup>288</sup>. Elle propose que l'évolution simultanée des deux espèces en interaction devrait accélérer leur évolution génétique et moléculaire grâce aux évolutions constantes de l'espèce antagoniste. Ainsi, les hôtes développeraient des mécanismes de résistances pour échapper aux infections virales et les virus évolueraient sans cesse en retour<sup>392</sup>. La coévolution rapide entre l'hôte et son virus a notamment été décrite dans le système modèle *Emiliana huxleyi* et son virus (EhV)<sup>393</sup>. En effet, les auteurs ont observé que la phase diploïde du coccolithophore, sensible à l'infection virale, conduit à la transition de la phase haploïde résistante durant l'infection. La résistance réversible des cellules du dinoflagellé toxique *Heterocapsa circularisquama* à l'infection après co-culture avec son virus à ARN simple brin est un autre exemple qui illustre cette théorie évolutive. Les auteurs proposent que le caractère réversible des lignées cellulaires au virus pourrait être corrélé à un mécanisme interne de suppression, empêchant la réplication du génome viral dans la cellule hôte. Chez la microalgue *Ostreococcus tauri*, le taux de transition d'un phénotype de sensibilité vers un phénotype de résistance des lignées cellulaires à son virus, OtV5 a été estimé entre 1/1000 et 1/10000<sup>308,310</sup>. Récemment, Yau et al. ont proposé d'expliquer ce changement de phase chez *O. mediterraneus* et son virus OmV2 à travers une modélisation mathématique d'interaction hôte-virus<sup>71</sup>. Ainsi, ce modèle soutient l'idée que la coexistence stable serait renversée par une dynamique axée sur du « compromis », selon laquelle l'espèce hôte résistante deviendrait dominante, et la minorité d'espèces sensibles à l'infection, mais à croissance plus rapide, maintiendrait la population virale sur plusieurs générations successives.

Alors qu'il n'est pas évident de les observer au sein des populations naturelles, l'évolution de la résistance et son maintien dans le système sont des phénomènes courants en culture. Jusqu'à présent les stratégies de défenses antivirales ont été établies principalement chez le modèle bactérie-phage<sup>297,394-396</sup>. En se basant sur ces connaissances, la défense antivirale de l'hôte pourrait être assurée au cours des différentes étapes de la réplication du virus. La résistance des cellules infectées par un *Prasinovirus* est un phénomène qui apparaît commun aux Mamiellales<sup>222,237,293,308</sup>. Yau et al. ont proposé que l'altération de l'attachement viral à la surface de la cellule serait une première ligne de défense des algues unicellulaires<sup>308</sup>.

Le génome des *Mamiellophyceae* présentent deux chromosomes atypiques, nommés *outliers* dont le plus petit des deux, le SOC (*small outlier chromosome*) serait impliqué dans l'acquisition de la résistance au virus lors de l'infection<sup>308</sup>. Ce chromosome présente des caractéristiques communes au *Mamiellales* qui sont : -une plus faible teneur en GC (50%) que celui du génome (59%), -une très grande proportion d'éléments transposables (20%) , -un taux plus élevé de mutations<sup>70</sup>. Le SOC présente une plasticité remarquable en taille et en composition dans les communautés naturelles<sup>48</sup>. Il a été proposé que cette hypervariabilité serait favorisée par la très forte pression de sélection induite sur les communautés phytoplanctoniques par les populations denses en *Prasinovirus* dans les environnements naturels<sup>308</sup>. Les mécanismes moléculaires qui sous-tendent les traits phénotypiques de résistance ont été étudiés et caractérisés chez *O. tauri* et son virus OtV5 par des approches génomiques et transcriptomiques. L'acquisition de la résistance à OtV5 induite chez *O. tauri*, apparaît fortement corrélée à un taux élevé de grands réarrangements spontanés concentrés sur le SOC, conduisant à une hypervariabilité de la taille de ce chromosome<sup>310</sup>. En ce sens, une corrélation linéaire a été établie entre la taille du chromosome SOC et la résistance des souches résistantes ; les souches les plus résistantes présentent un chromosome SOC plus long<sup>70</sup>. L'analyse transcriptomique de souches d'*O. tauri* résistantes a mis en évidence la surexpression de gènes la plupart impliqués dans le transport, la synthèse et la modification de sucre, majoritairement de type glycosyltransférases. Ainsi, le système immunitaire antiviral chez *Ostreococcus* pourrait être un mécanisme intracellulaire complexe, qui résulte de réarrangements massifs du chromosome immunitaire SOC, impliquant des modifications transcriptionnelles de gènes à l'origine de changements adaptatifs.

C'est dans ce contexte que nous avons cherché à étudier les bases fonctionnelles de la résistance au *Prasinovirus* chez l'algue marine *Ostreococcus tauri*. Pour identifier des gènes candidats impliqués dans ce système immunitaire antivirale, une approche moléculaire de mutagenèse d'insertion aléatoire chez la souche *O. tauri* RCC1115 a été médiée par le

polyéthylène glycol (PEG). L'utilisation d'un vecteur (pOLK) conférant la résistance à un antibiotique (G418 analogue à la kanamycine) a été privilégiée pour sélectionner les clones transformés. Dans le chapitre précédent, la caractérisation par approche NGS des événements d'intégration a été présentée pour cinq lignées clonales. Dans ce second chapitre, nous présentons la stratégie expérimentale utilisée pour la sélection des lignées clonales transformées, qui s'est basée sur le caractère phénotypique des cultures à l'infection virale. Dans un second temps nous avons essayé de corrélérer les données phénotypiques aux données génomiques afin d'avoir une approche globale dans les mécanismes de l'immunité antivirale associés à l'acquisition de la résistance dans le modèle d'interaction *O. tauri*–*Prasinovirus*.

## 2. Matériels et méthodes

### *Conditions de culture des algues et des virus*

L'ensemble des lignées clonales (i.e. plus de 1200 lignées clonales transformées et 200 lignées clonales non-transformées) originaires de la lignée parentale *O. tauri* RCC1115, sont les mêmes cultures décrites précédemment et leurs conditions de cultures sont reportées dans le matériel et méthodes du chapitre I. La production des virus et leurs conditions de culture sont décrites dans le chapitre III. Pour la dynamique d'infection, les cellules *O. tauri* RCC1115 en phase exponentielle de croissance ont été infectées au virus OtV09-578, décrit précédemment<sup>231</sup>, à une MOI de 5.

### *Suivi de croissance des algues par cytométrie en flux*

Afin de mesurer la densité de la population cellulaire, les cellules ont été fixées à l'aide de glutaraldéhyde (concentration finale de 0.25%) et d'acide pluronique (concentration finale de 0.10%) pendant 15 min dans l'obscurité, puis le SYBR Green I (Ozyme réf. LON50512) a été ajouté pendant 15 min pour colorer les bactéries. Les cellules ont été comptées à l'aide d'un cytomètre en flux Beckman-Coulter Cytotflex (i.e. longueur d'onde d'excitation laser 488 nm), par autofluorescence de la chlorophylle pour les algues (i.e. filtre de détection > 620 nm) et par fluorescence du SYBR Green I pour les bactéries (i.e. canal de détection 525–540 nm).

### *Criblage phénotypique des lignées clonales*

Une première sélection des lignées clonales transformées a été réalisée en présence d'un antibiotique (G418). Les 200 clones non-transformés ont été obtenus par croissance de la souche parentale *O. tauri* RCC1115 sur agarose *low-melting* à 1.5% (concentration finale de 0,15%) dans une boîte de Pétri (55 mm). Les colonies ont été repiquées et placées dans une culture liquide L1 dans des plaques 96-puits. Un criblage phénotypique (i.e. sensibilité ou résistance) à la souche virale OtV09-578 (i.e. la plus généraliste) a été effectué sur toutes les lignées clonales (i.e. non-transformées et transformées). Le criblage phénotypique a été réalisé en milieu liquide L1 dans des plaques de 96 puits. Des cultures témoins non-infectées ont été inoculées avec 50  $\mu$ L de milieu L1, tandis qu'une suspension virale concentrée de 50  $\mu$ L a été ajoutée aux cultures cellulaires jusqu'à un volume final de 200  $\mu$ L. Les cultures ont été suivies par inspection visuelle sous des conditions de croissance décrites ci-dessus.

### *Test de spécificité aux virus*

Un total de 34 lignées clonales sélectionnées (i.e. transformées et non-transformées) ainsi que la lignée parentale *O. tauri* ont été testées pour leur sensibilité à 20 virus spécifiques d'*O. tauri*. Les souches d'algues en croissance exponentielle ont été incorporées dans le milieu L1 semi-solide par l'ajout d'agarose fondu à 1.5% (concentration finale de 0.15%) et versées dans des boîtes de Pétri (9 cm de diamètre). Trois microlitres de chaque lysat viral ont été déposés en *duplicata* sur chaque souche d'algue et la sensibilité au virus a été évaluée par l'apparition de plaques visualisées après 7 jours.

### *Électrophorèse sur gel à champs pulsé (Pulsed Filed Gel Electrophoresis – PFGE)*

Le caryotype de toutes les lignées clonales de cette étude a été réalisé en utilisant le protocole d'électrophorèse sur gel en champ pulsé (PFGE) tel que décrit précédemment<sup>229</sup> avec les modifications suivantes. Les cultures de cellules ( $8.7 \times 10^7$  cellules) en phase exponentielle de croissance ont été centrifugées ( $8000 \times g$  pendant 20 min), remises en suspension dans 150  $\mu$ L de tampon TE (10 mM Tris-HCl, 125 mM EDTA, pH 8), puis intégrées dans des *plugs* mélangés à un volume égal d'agarose fondu à bas point de fusion à 1 % dans du tampon TE pré-refroidi à 45°C. Ensuite, les cellules dans les *plugs* ont été lysées en utilisant un tampon 10 mM Tris HCl pH 8, 0.5 M EDTA pH 8, 1% lauryl sarcosinate avec 1 mg/mL de protéinase K

(concentration finale), à 37°C pendant 24 h sous agitation constante, puis lavées 3 fois (12 h chacun) dans 0.5 M EDTA pH 8 avant le transfert des *plugs* à 4°C pour stockage. Pour chaque lignée clonale, les *plugs* ont été chargés dans les puits du gel d'agarose. La PFGE a été réalisé avec le système CHEF-DR III (Bio-Rad) et le gel d'agarose a été incubé dans un tampon TBE 0.5 X (final 44.5 mM Tris, 44.5 mM acide borique, 1 mM EDTA à pH 8) à 14°C. L'électrophorèse a été réalisée en 2 étapes : la première d'une durée de 15 heures, 6 V cm<sup>-1</sup> avec un angle d'impulsion de 120° avec un temps de commutation de 60 s, et la seconde étape d'une durée de 9 heures et un temps de commutation de 90 s. Le gel a été finalement incubé pendant 1 heure dans un bain de Bromure d'Ethidium (BET) à une concentration de 0.6 µg/ml puis visualisé à la lumière ultra-violet.

### *Caractérisation génomique des éléments transposables (ETs) à LTR et comparaison des séquences nucléotidiques*

L'identification des séquences associés à des ETs à proximité des sites d'intégration d'ADN a été réalisée manuellement à partir du logiciel Geneious® (v2019.2.1). La prédiction des gènes des séquences d'ETs a été réalisée à l'aide de Prokka (v1.14.5) en spécifiant « virus » comme royaume taxonomique<sup>397</sup>. La fonction pour chaque gène a ensuite été prédite par BlastP à partir de la banque de données du NCBI « nr ». Enfin, une annotation fonctionnelle de toutes les séquences de protéines a été recherchée par une recherche BLASTp par rapport à la base de données NCBI « nr » en acceptant une valeur de *e-value* <1e-5, ne gardant que le meilleur résultat. L'analyse comparative des génomes a été réalisée en utilisant le package Rstudio gggenes v0.4.1. La matrice de distance nucléotidique a été calculée sur la base des séquences nucléotidiques à partir du logiciel Geneious® (v2019.2.1) avec les fonctions prédéfinies. Les récentes intégrations d'ETs dans les contigs assemblés des lignées clonales ont été visualisées par dotplot en utilisant Mummer (v4.0.0).

### *Données séquencées disponibles*

Le génome de référence de la lignée parentale *O. tauri* RCC1115 utilisée a été séquencé en 2017<sup>70</sup> avec le numéro BioProject du NCBI : PRJNA337288. Les cultures des sept lignées clonales ont été déposées à la Roscoff Culture Collection (<https://roscoff-culture-collection.org/>) avec le numéro des souches suivantes : T3 (RCC7079), T6 (RCC7080), T12 (RCC7081), T14 (RCC7082), T16 (RCC7083), NT1 (RCC7084), NT10 (RCC7085). Les

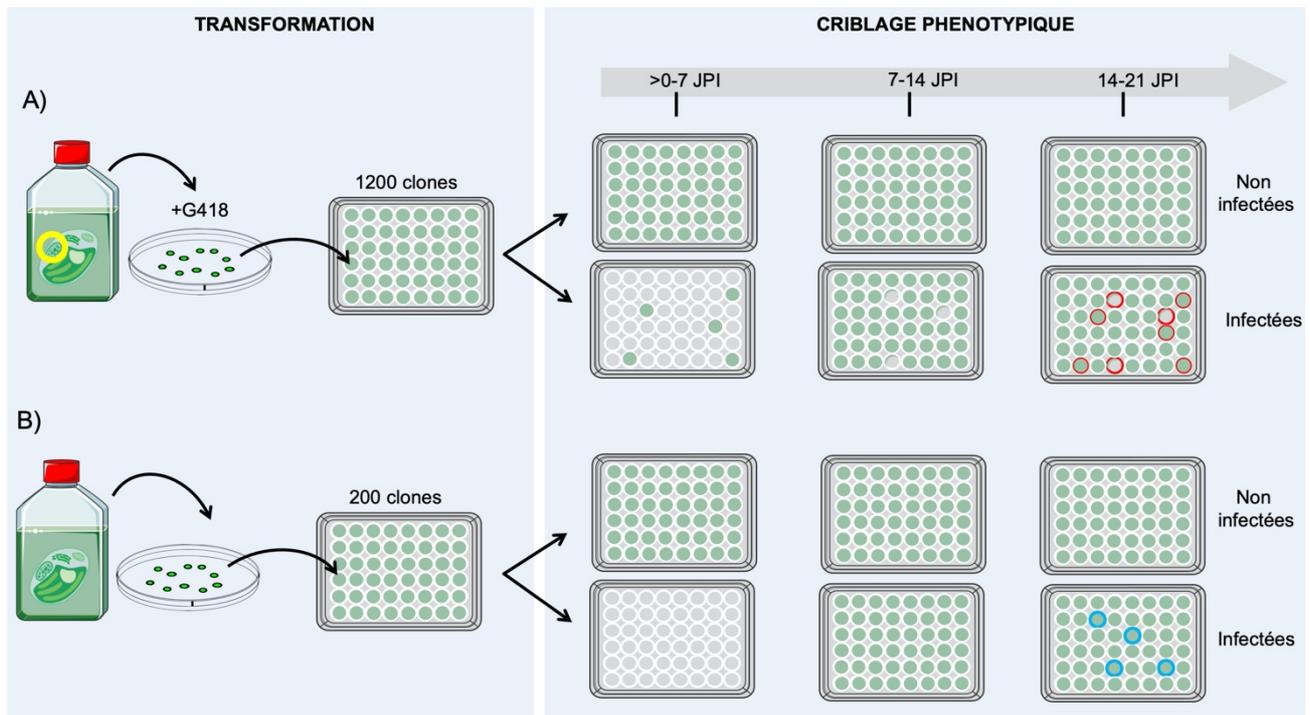
assemblages hybrides Illumina-ONT et les lectures brutes Illumina et ONT des lignées d'*O. tauri* analysées dans cette étude sont disponibles auprès de l'European Nucleotide Archive (ENA) sous le numéro d'accès au projet PRJEB43294 avec le numéro suivant : AP0985 (T3), AP0986 (T6), AP0993 (T12), AP0994 (T14), AP0982 (T16), AP0965 (NT1), AP0966 (NT10).

### 3. Résultats

#### *Observation préliminaire du caractère phénotypique des lignées clonales sélectionnées*

Afin d'étudier les mécanismes de résistance au virus chez l'algue *O. tauri*, nous avons sélectionné des lignées clonales transformées montrant des caractères phénotypiques différents vis-à-vis de la lignée parentale *O. tauri* RCC1115. Suite à la transformation, plus de 1200 clones ont été sélectionnés sur agarose par antibiotique. Ils ont alors été testés pour leur capacité d'infection au virus OtV09-578. Le plan expérimental est présenté dans la **Figure 18**.

La dynamique d'infection a été suivie par visualisation directe de la coloration verte de la culture sous 3 semaines. Les cultures ont été considérées comme sensibles au virus en cas de perte de la coloration verte dans les sept premiers jours d'infection, et comme résistantes au virus, si aucun changement de couleur n'apparaissait comparé aux cultures témoins non infectées (T1 à T9 ; **Tableau 6**). D'autre part, les cultures sensibles aux virus ne montrant pas d'acquisition de la résistance dans les 3 semaines ont été considérées aussi comme des candidats d'intérêts dans cette analyse (T10 à T19 ; **Tableau 6**). Avec ces critères, nous avons réussi à sélectionner 19 lignées clonales transformées (T), alors que 15 lignées contrôles (non-transformées – NT) ont été sélectionnées indépendamment et aléatoirement, sans changement phénotypique après comparaison avec la lignée parentale *O. tauri* RCC1115. Le génome des lignées transformées T3, T6, T12, T14 et T16 et les deux cultures non-transformées NT1 et NT10 ont été séquencées, et la caractérisation des événements d'intégration de l'ADN a été présentée dans le chapitre I.



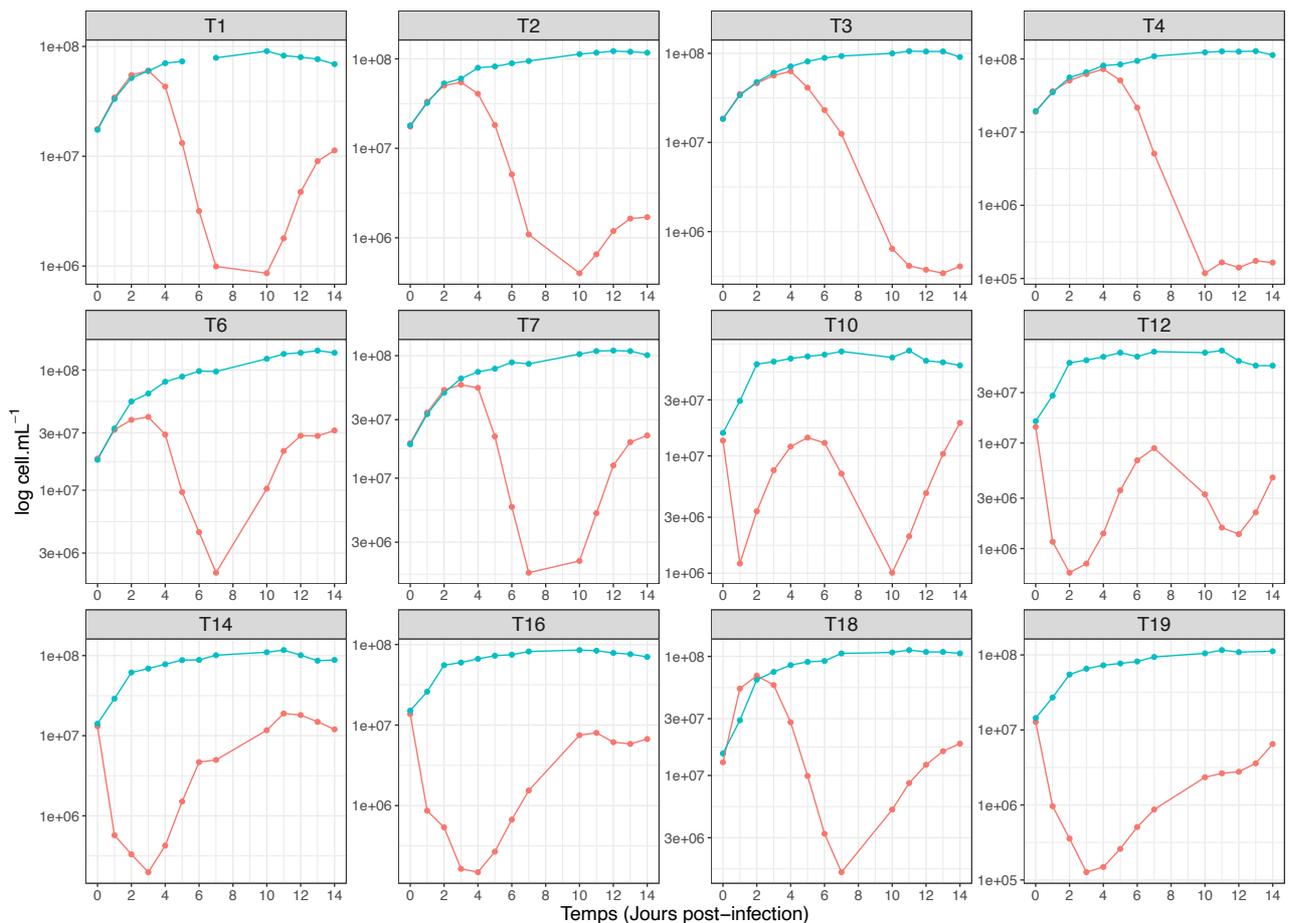
**Figure 18.** Plan expérimental du criblage phénotypique résistant-sensible des lignées clonales transformées (A) et non-transformées (B) d'*O. tauri* à OtV09-578, au cours du temps. Les cultures transformées candidates ont été sélectionnées à la fin des 21 jours post-infection (JPI) selon les deux critères : une résistance des cellules infectées ou une incapacité des cellules à recroître (cerclée en rouge) après infection par le virus, tandis que les cultures non-transformées ont été choisies aléatoirement (cerclée en bleu). Le phénotype des cultures non-transformées et transformées sélectionnées candidates dans cette étude sont présentées dans le **Tableau 6**.

**Tableau 6.** Description phénotypique des lignées clonales transformées (Tx) candidates face à l'infection virale OtV09-578. Les cultures résistantes ont été marquées par un moins (-) tandis que celles qui montraient une sensibilité au virus mais sans évolution de la résistance visible après infection ont été notées par un plus (+).

Lignées	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19
Phénotype	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+

## Dynamique d'infection de 12 lignées clonales transformées

Étant donné la difficulté de suivre la dynamique d'infection seulement par critère visuel, nous avons cherché à étudier plus finement la dynamique d'infection de 12 lignées transformées sur les 19 sélectionnées par suivi cinétique. Contrairement à ce qui était attendu, le nombre de cellules viables diminuent fortement dans les jours qui suivent l'ajout du virus OtV09-578 à temps zéro ( $t_0$ ). L'ensemble des cultures sélectionnées sont donc sensibles au virus (**Figure 19**) et ne reflètent pas le caractère phénotypique décrit précédemment (**Tableau 6**). La dynamique des cellules hôtes après infection virale révèlent des profils de lyse qui diffèrent entre les cultures. Globalement, le délai qui s'écoule entre l'exposition au virus et la diminution du nombre de cellules viables en culture est de 4 jours pour la majorité des cultures. En revanche, cette période se montre plus courte pour les lignées T14, T16 et T19, où le nombre de cellules viables infectées diminue drastiquement 1 jours post-infection.



**Figure 19.** Dynamique d'infection des lignées clonales transformées non infectées contrôles (bleue) et infectées à OtV09-578 (rouge) durant 14 jours post-infection. La concentration cellulaire (log de cellule/mL) est présentée sur l'axe des ordonnées et la durée (jour) du suivi cellulaire sur l'axe des abscisses.

Dans cette expérience de cinétique d'infection, une stabilisation puis dans la plupart des cas une augmentation de la coloration verte des cultures est aussi observée. Ce retour à la viabilité montre que les cellules en culture sont devenues résistantes au virus. L'acquisition de la résistance des cellules infectées apparaît dans l'ensemble des cultures, avec une dynamique de résistance des cellules des lignées T3 et T4 apparemment plus lente. Ce phénomène survient plus rapidement (i.e. entre 3 et 4 jours après infection) chez les lignées T14, T16 et T19 qui montrent une phase de lyse plus courte que pour les autres lignées (i.e. entre 9 et 13 jours après infection). La dynamique d'infection des lignées T10 et T12 révèle une fluctuation rapide de la croissance des cellules infectées. Dans ces deux cultures, une première lyse des cellules est observée très rapidement entre 24 et 48 heures après infection, suivie par une acquisition de la résistance. Une seconde chute des cellules viables en culture est observée 7 jours après infection et conduit à une nouvelle acquisition de la résistance de ces mêmes cellules. Une oscillation de la population phytoplanctonique est présente dans les cultures T10 et T12. Cette dynamique au cours du temps met en évidence une évolution réciproque entre l'hôte et son virus, maintenant un état d'équilibre entre résistance et infectivité au sein de la culture.

#### *Spectre d'infection des lignées clonales à un large panel de virus spécifique d'*O. tauri**

Étant donné que l'ensemble des lignées clonales T présentaient finalement un caractère sensible à OtV09-578 – mais non distinguable par l'approche visuelle –, nous avons cherché à mettre en évidence des changements phénotypiques à l'infection sur un plus large panel de virus d'*O. tauri* (OtVs). Une grande partie des virus utilisés (14 sur les 20 OtVs totaux) a été décrite dans des travaux précédents<sup>231</sup>, tandis que les autres l'ont été dans ce travail de thèse (Cf. Chapitre III). Nous avons testé la spécificité de ces OtVs par plage de lyse sur les 19 lignées T et les 15 lignées NT sélectionnées précédemment.

La lignée parentale *O. tauri* RCC1115 utilisée dans cette étude présente un spectre de spécificité similaire à celui réalisé antérieurement par Clerissi *et al.*<sup>231</sup>. 45% des virus présentent le même patron de spécificité entre la culture parentale et l'ensemble des cultures clonales (**Tableau 7**). Contrairement aux résultats présentés précédemment (**Tableau 6**), l'ensemble des lignées T ne montrent aucun changement phénotypique par l'approche de plage de lyse par rapport à la lignée parentale *O. tauri* RCC1115, et confirme la sensibilité des cellules à l'infection de OtV09-578 comme observé par suivi cinétique (**Figure 19**). En revanche, des changements phénotypiques sont observés dans certaines cultures. Les lignées T montrent un

*switch* du phénotype de résistance (R) vers un phénotype de sensibilité (S) pour environ 90% (17/19) à OtV09-570 et pour environ 80% (15/19) à OtV06-12. 53%, 47% et 37% des lignées T sont aussi passées de résistantes à sensibles pour les virus OtV09-584, Sylt2-5 et OtV19-R, respectivement. Il est important de noter que sur les 74 modifications phénotypiques observées dans les cultures T, 62 proviennent d'un passage d'un caractère résistant vers un caractère sensible. Ces observations suggèrent que l'intégration d'ADN lors de la transformation conduit à des perturbations intrinsèques au sein du génome se traduisant par des modifications phénotypiques du spectre d'infection par le virus.

Face à cette hypothèse, nous avons testé le patron d'infection des 15 lignées clonales NTs sélectionnées précédemment à ce même panel d'OtVs (**Tableau 7**). Des variations phénotypiques sont présentes chez les NTs, mais globalement en plus faible proportion que chez les lignées T. Contrastant avec le reste, 93% des lignées NTs passent de résistant à sensible à OtV09-584, un taux bien supérieur à celui des lignées T. Par ailleurs, un basculement de résistant vers sensible apparaît pour l'infection par OtV09-570 (8/15) et en moindre proportion pour les virus OtV06-12 (3 sur 15), OtV09-556 (3 sur 15) et OtV06-1 (dans ce cas, seulement NT8). Il est important de noter que le *switch* de R vers S est largement prédominant, alors que le *switch de S vers R* n'apparaît que pour les infections par OtV09-582 et OtV09-585 et dans le cas des lignées NT seulement pour la lignée NT8, suggérant un mécanisme spécifique commun de résistance à ces deux virus.

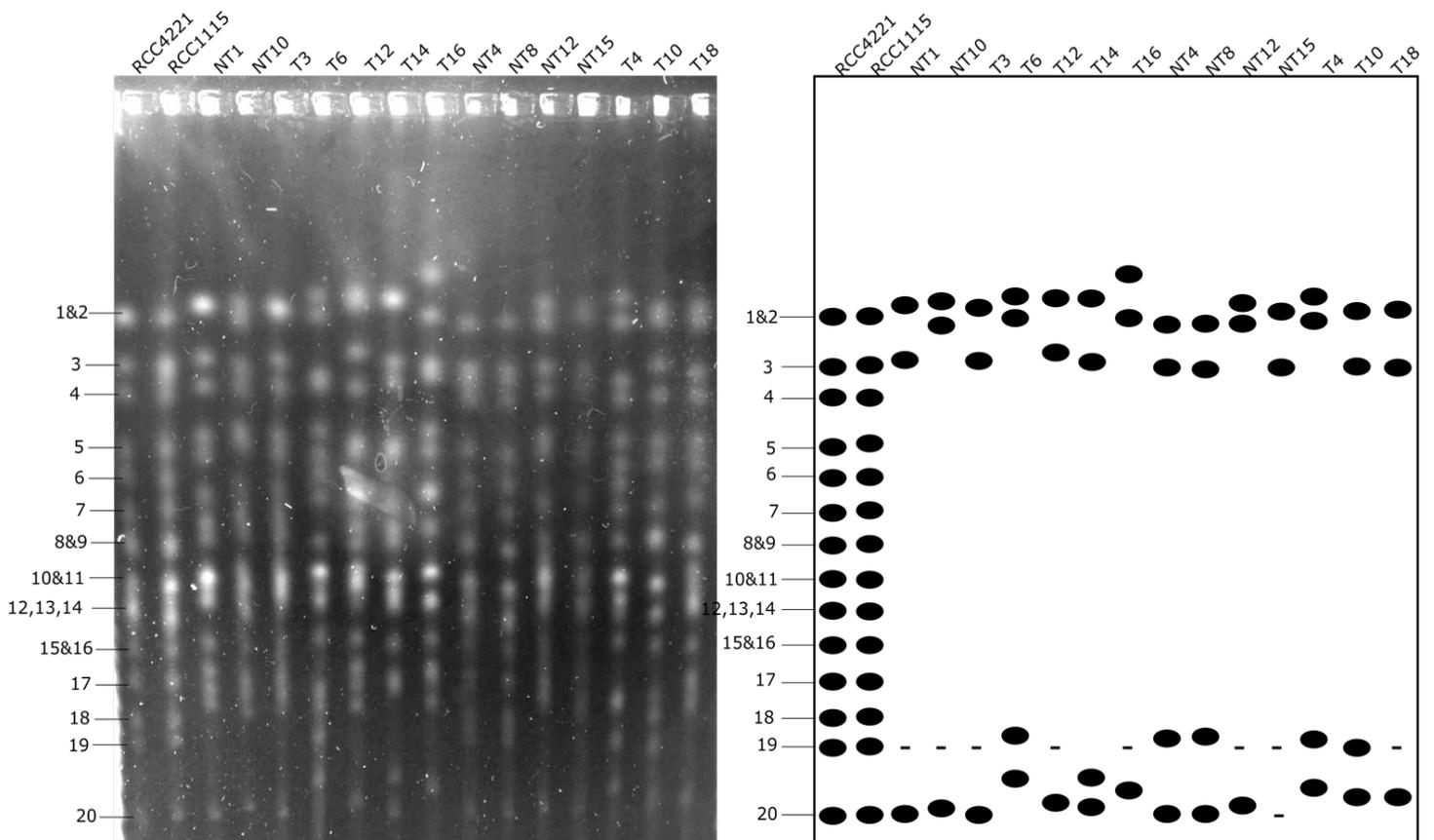
**Tableau 7.** Test d'infection croisé des lignées transformées (Tx) et non-transformées (NTx) à 20 *Ostreococcus tauri* Virus (OtVs) (\*\*virus issus de Clerissi *et al*, 2012<sup>231</sup>). Le fond en jaune indique un changement de phénotype des lignées clonales par rapport à la lignée parentale *O. tauri* RCC1115. Le signe négatif (-) indique une absence de lyse, le signe positif (+) une forte lyse, tandis que le signe positif et négatif (+/-) indique une faible lyse. Les lignées clonales séquencées dans le chapitre I sont indiquées par un \*. Ces observations ont été suivies jusqu'à 7 jours post-infection.

Spectre d'infection virale

	Syl2-1	Syl2-5	OtV06-1**	OtV5	OtV09-573**	OtV06-4**	OtV09-557**	OtV09-559**	OtV19-P	OtV19-O	OtV19-T	OtV19-R	OtV09-556**	OtV09-584**	OtV09-570**	OtV06-12**	OtV09-582**	OtV09-585**	OtV09-565**	OtV09-578**
RCC1115	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
T1	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	+/-	+	+	+	+	+
T2	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	+/-	+	+	+	+	+
T3*	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	-	+	+	+	+/-	+/-
T4	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	+/-	+	+	+	+	+
T5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
T6*	-	-	-	-	-	-	-	-	-	-	-	+/-	-	+/-	+/-	+	+	+	+	+
T7	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	+/-	+/-	+	+	+	+
T8	-	-	-	-	-	-	-	-	-	-	-	-	+/-	+/-	+/-	+/-	+	+	+	+
T9	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	+/-	+/-	+	+	+	+
T10	-	+	-	-	-	-	-	-	-	-	+/-	+	-	-	+	+	-	-	+	+
T11	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	+	+	+/-
T12*	-	+	-	-	-	-	-	-	-	-	-	+/-	-	-	+	+	-	-	+	+
T13	-	+	-	-	-	-	-	-	-	-	-	-	-	+/-	+	-	+	+	+	+
T14*	-	+	-	-	-	-	-	-	-	-	-	+	-	-	+	+	-	-	+	+
T15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	+	+	+
T16*	-	+	-	-	-	-	-	-	-	-	-	+	-	-	+	+	-	-	+	+
T17	-	+	-	-	-	-	-	-	-	-	-	-	-	+/-	+	-	+	+	+	+
T18	-	+	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	+	+
T19	-	+	-	-	-	-	-	-	+/-	-	+/-	+	-	-	+	+	-	-	+	+
NT1*	-	-	-	-	-	-	-	-	-	-	-	-	+/-	+/-	+/-	+	+	+	+	+
NT2	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	-	-	+	+	+	+
NT3	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	-	-	+	+	+	+
NT4	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	-	-	+	+	+	+
NT5	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	-	-	+	+	+	+
NT6	-	-	-	-	-	-	-	-	-	-	-	-	+/-	+/-	+/-	-	+	+	+	+
NT7	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	+/-	-	+	+	+	+
NT8	-	-	+/-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	+	+
NT9	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	+/-	-	+	+	+	+
NT10*	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	-	+/-	+	+	+	+
NT11	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	-	-	+	+	+	+
NT12	-	-	-	-	-	-	-	-	-	-	-	-	+/-	+/-	+/-	-	+	+	+	+
NT13	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	+/-	-	+	+	+	+
NT14	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	-	-	+	+	+	+
NT15	-	-	-	-	-	-	-	-	-	-	-	-	-	+/-	+/-	-	+	+	+	+

*Analyse du caryotype des lignées clonales d'O. tauri RCC1115*

Étant donné l'existence de tels changements phénotypiques induits rapidement dans l'ensemble des cultures suite à l'infection par une large gamme de virus, nous avons cherché à mettre en évidence si ces variations reflèteraient de possibles changements au sein des génomes. Pour cela, les caryotypes de 8 lignées clonales T et 6 lignées clonales NT ont été comparés à celui de la lignée parentale RCC1115 par électrophorèse sur gel à champ pulsé (*Pulsed-Field Gel Electrophoresis – PFGE*). Cette lignée parentale RCC1115 montre un caryotype très similaire à la souche *O. tauri* RCC4221 (**Figure 20**). En revanche, 12 lignées sur 14 (~86%) présentent un caryotype très différent de la lignée parentale avec de grandes variations en taille sur la majorité des chromosomes.



**Figure 20.** Analyse de l'électrophorèse à champs pulsée des lignées clonales transformées (Tx) et non-transformées (NTx). Le gel original est montré à gauche tandis qu'une représentation schématique est présentée à droite de la figure ; seuls les chromosomes avec d'importants réarrangements ont été illustrés. Le numéro des chromosomes associé à la souche *O. tauri* RCC4221 sont indiqués à gauche du gel. Le signe moins (-) indique une ambiguïté quant à la localisation du chromosome.

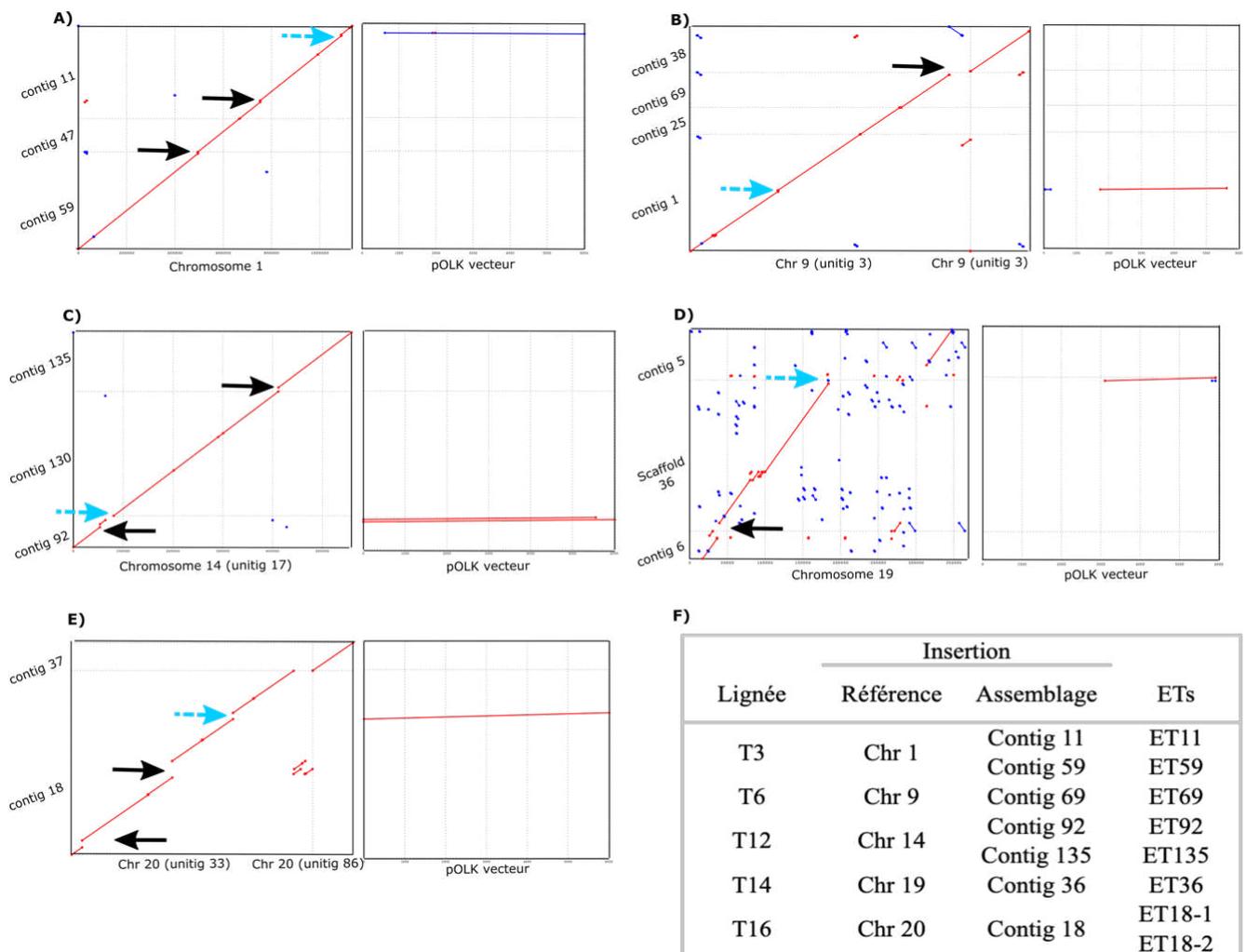
Nous pouvons écarter l'hypothèse d'une éventuelle contamination au sein de la culture parentale puisque les lignées NT4 et NT8 présentent un caryotype semblable à *O. tauri* RCC1115 et RCC4221. Ces changements sont principalement observés dans le chromosome 19 ou SOC et une grande proportion des souches (9 sur 14 lignées clonales) montre une localisation ambiguë de ce chromosome. Ces événements sont possiblement dû à des remaniements structuraux de types translocations, insertions ou des délétions, qui ont été déjà observés dans les cultures *O. tauri* RCC4221 résistantes au virus<sup>310</sup>. La lignée T14 présente une taille du SOC plus réduite que celle de la lignée parentale *O. tauri* RCC1115. Ce phénomène a été associé précédemment dans le chapitre I, à une délétion de plus de 100 kb, induite par l'événement d'intégration du vecteur pOLK4 (**Figure S1**). Conformément à la forte prédisposition du SOC aux variations structurelles, il est probable que l'insertion de l'ADN ait favorisé cette délétion, étant donné que la délétion est adjacente à l'intégration. Des changements en taille du chromosome 20 sont aussi mis en évidence, particulièrement pour la lignée T16 qui présente une augmentation de sa taille, probablement aussi corrélée à l'intégration de l'ADN exogène qui aurait induit des remaniements au sein de ce chromosome<sup>398</sup>. Par ailleurs, comme pour les lignées transformées, d'importants réarrangements structuraux du chromosome 19 ainsi que dans d'autres chromosomes (chr1, chr2, chr3, chr20) ont été aussi mis en évidence dans les lignées contrôles (NT). Ainsi, au moins deux hypothèses peuvent être proposées : 1) la souche *O. tauri* RCC1115 serait une population de cellules hétérogènes au départ, c'est-à-dire non clonales, avec un type dominant donnant un caryotype unique dans le PFGE ; 2) la plasticité génomique intrinsèque de la souche *O. tauri* RCC1115 donnerait constamment lieu à des variations entre les lignées clonales, mais à des échelles de temps plus courtes. Avec la première hypothèse, le clonage des lignées NTs permettrait de révéler la diversité de la population originale, comme le résultat de l'accumulation de cellules hétérogènes pendant des années.

#### *Analyse génomique des lignées clonales O. tauri RCC1115*

Afin de mieux comprendre ce phénomène, nous avons cherché à corréler les importants réarrangements structuraux observés chez les lignées *O. tauri* RCC1115 avec les données génomiques. Dans le chapitre I, les événements d'intégrations aléatoires d'ADN exogène dans les lignées T3, T6, T12, T14 et T16 ont été caractérisés par des approches NGS.

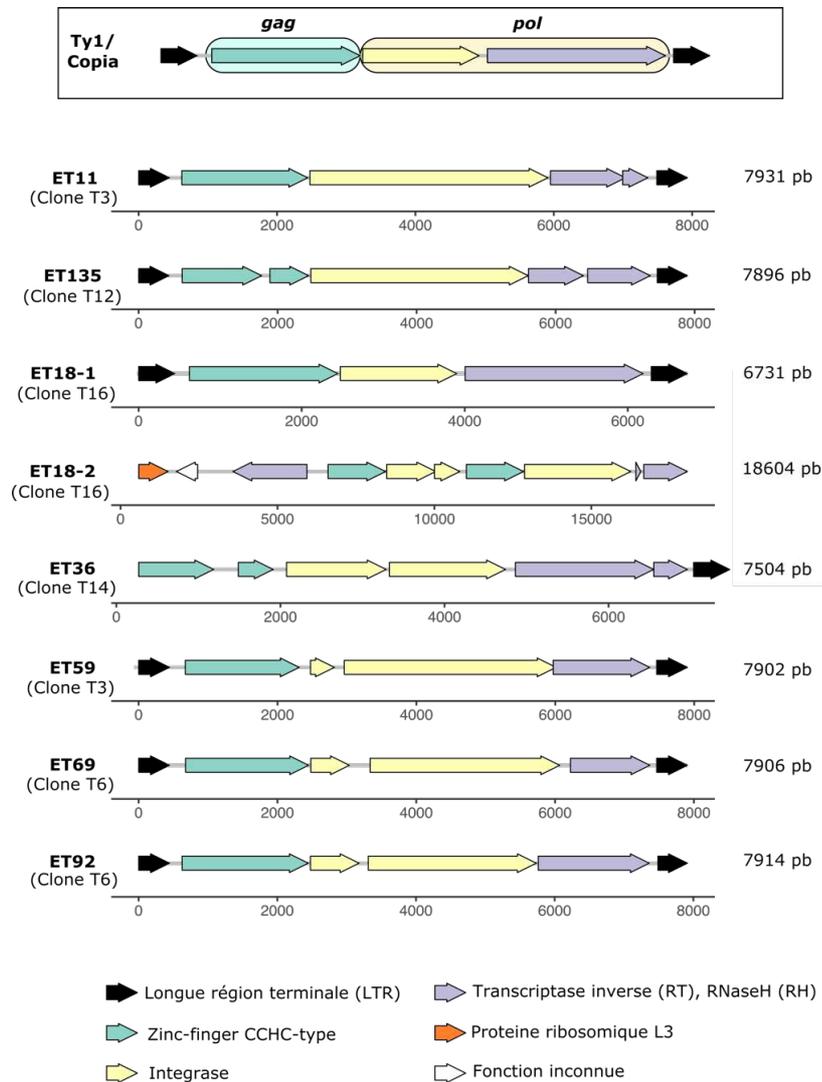
L'analyse nucléotidique en tracé de points (*dot plot*) a révélé la présence de grandes mutations structurales de type insertion-délétion (indel) ou translocation, adjacentes aux intégrations du vecteur pOLK qui se sont produites dans le génome des lignées T (**Figure 21**).

Ces événements d'insertion ont une taille d'environ 8,000 pb, un taux plus faible en GC (55%) que le génome de référence (60%). D'autre part, ces régions sont bornées par deux longues régions non-inversées répétées (~445 pb) et codent pour des domaines fonctionnels enzymatiques : un motif *Zinc-finger* de type CCHC, une intégrase et une transcriptase inverse (RT) avec une ribonucléase H (RH) (**Figure 22**).



**Figure 21.** Dot plot nucléotidique des lignées clonales transformées *O. tauri* RCC1115. Alignement des contigs issu de l'assemblage hybride MaSuRCA-Flye dans lesquels les événements d'insertions du vecteur pOLK se sont produits (axe-y) contre le génome de référence et contre la séquence pOLK (axe-x) pour chaque transformant (A) T3, (B) T6, (C) T12, (D) T14, (E) T16. La flèche pointillée bleue indique le locus d'insertion pOLK tandis que les flèches noires en trait plein montre les événements d'insertion d'éléments transposables (ETs) à LTR dans les lignées transformées. Les lignes rouges

correspondent aux alignements dans le même sens que le chromosome de référence et les lignes bleues aux alignements anti-sens. Le nom et la localisation des insertions des ETs dans chaque lignée sont référencés (F).



**Figure 22.** Cartographie génétique des éléments mobiles à LTR (classe I) associés aux génomes de cinq clones transformés *O. tauri* RCC1115. Un modèle structurel de l'élément à LTR Ty1/copia est présenté ci-dessus. Les éléments mobiles à LTR sont délimités par de *Large Terminal Repeat* (LTR) en orientation directe. Ces éléments codent deux protéines principales *gag* (*group specific antigens*) et *pol*. Au sein de ces protéines plusieurs domaines sont identifiés. Le premier domaine contient un motif *Zinc-finger* de type CCHC de la région *gag*, tandis que les domaines de la région *pol* sont représentés par une intégrase (INT), une transcriptase inverse – Reverse transcriptase (RT), et la ribonucléase H (RH). ET36 ne présente pas de LTR en 5', vraisemblablement causé par un problème d'assemblage. L'élément ET18-2 montre une structure plus complexe. L'alignement pairwise des huit éléments transposables sont présentés dans les données supplémentaires. Le locus de chaque ET au sein des génomes des clones transformés est indiqué dans la **Figure 21**.

Ces caractéristiques génétiques indiquent la présence d'éléments transposables (ETs) de classe I à LTR (i.e. rétrotransposons), intégrés récemment dans les génomes des lignées transformées. Chez les rétrotransposons à LTR, le motif *Zinc-finger* de type CCHC est retrouvé au sein de la protéine *gag*, tandis que l'intégrase, la RT et la RH sont des domaines associés à la protéine *pol*<sup>399</sup>. La structure des éléments et la position de l'intégrase donnent une information supplémentaire quant à la nature des ETs en suggérant qu'il s'agit d'éléments de type Ty1/Copia (**Figure 22**). L'alignement nucléotidique des séquences entières a montré une étroite relation des huit ETs, bien qu'ils ne soient pas identiques. Ils présentent un pourcentage d'identité situé entre 67 et 98 % (**Tableau S1**). Malgré leur faible divergence, il a été possible de distinguer chaque ET comme un élément unique et nommé individuellement selon leur site d'intégration dans le génome du clone transformé.

Chez T3, deux ETs (i.e. ET59 et ET11) se sont transposés dans le chromosome 1 en positions 494,811 et 752,808 pb, respectivement, tous deux en amont de région promotrice des gènes codant pour une glutamine aminotransferase (*ostta01g03620*) et une porphobilinogen synthase (*ostta01g02390*) (**Tableau S1**).

Une insertion de l'élément ET69 a été localisée au sein d'un gène codant pour une 4-hydroxy-3-méthylbut-2-en-1-yl diphosphate synthase (position 578,994 pb) dans le chromosome 9 de la lignée T6 (**Tableau S1**). Chez les plantes, cette enzyme est essentielle au développement des chloroplastes et nécessaire pour la résistance aux agents pathogènes médiés par l'acide salicylique<sup>400</sup>.

Pour la lignée T12, deux ETs ont été localisés dans le chromosome 9 à proximité de site d'intégration du vecteur : ET92 se situe en amont d'une région promotrice d'un gène codant pour un canal ionique mécano-sensible (*ostta14g00230*) et ET135 a été localisé en position 608,305 pb également en région intergénique en amont d'un gène codant pour une enzyme qui participe à la synthèse de peptide de la paroi cellulaire chez les plantes (UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase)(*ostta14g02240*) et d'un gène dont la fonction est inconnue (*ostta14g02235*).

Enfin chez T16, nous avons identifié une insertion de l'élément ET18-1 en amont du gène codant pour un facteur de transcription eucaryote *eIF4* (*ostta01g01600*) en position 7,063 pb et en amont d'un gène codant une protéine avec un domaine à répétition ankyrine (*ostta20g00010*) dans le chromosome 20.

Dans ce travail, nous mettons donc en avant que des insertions récentes sont retrouvées à proximité des sites d'insertion des vecteurs ADN et qu'il s'agit d'évènements de rétrotransposition d'éléments à LTR. Les rétrotransposons à LTR sont pratiquement retrouvés dans tous les grands groupes eucaryotes<sup>401</sup>, et semblent être la classe prédominante d'ETs chez les diatomées<sup>402</sup>. Leur transcriptase inverse est retrouvée en grande abondance dans le plancton marin par métagénomique et transcriptomique<sup>403</sup>. Les éléments transposables sont directement impliqués dans la grande plasticité des génomes et dans leurs adaptations environnementales. Leur activation est observée très fréquemment en réponse à divers stress biotiques ou abiotiques<sup>404,405</sup>. Ici, en comparaison à la lignée de référence, l'ensemble des lignées T a clairement été sujette à l'intégration d'éléments mobiles. Nous pouvons donc postuler que cette intégration à la suite de la transformation soit en relation directe avec des évènements de rétrotransposition.

**Tableau 8.** Identification des ETs dans le génome des lignées *O. tauri* RCC1115. Le chromosome où l'ET tire son origine est coloré en gris. L'ET présent au même locus que dans les autres clones transformés (T) et non-transformés (NT) est indiqué d'une croix (X), et absent par un moins (-). Ref : génome de référence *O. tauri* RCC1115.

<i>Insertion</i>	<i>ET</i>	<i>T3</i>	<i>T6</i>	<i>T12</i>	<i>T14</i>	<i>T16</i>	<i>NT1</i>	<i>NT10</i>	<i>Ref</i>
<i>Chr 1</i>	ET11	X	-	-	-	X	-	X	-
	ET59	X	X	-	-	-	-	-	-
<i>Chr 9</i>	ET69	X	X	X	-	-	-	X	-
<i>Chr 14</i>	ET92	-	X	X	X	-	X	-	-
	ET135	-	-	X	X	-	X	-	-
<i>Chr 19</i>	ET36	-	-	X	X	X	-	-	-
<i>Chr 20</i>	ET18-1	-	-	-	-	X	-	X	-
	ET18-2	-	-	-	-	X	-	-	-

Pour tester cette hypothèse, nous avons réalisé une analyse comparative des chromosomes entre les lignées clonales transformées et non-transformées avec le génome de référence. Sur un total des 8 éléments transposables polymorphiques, 5 (62.5%) étaient présents dans au moins deux génomes, suggérant une variation allélique commune. Seuls ET59, ET36 et ET18-2 n'ont pas été localisés sur le même locus chez les lignées NTs, comme chez la souche de référence (**Tableau 8**). Nous avons localisé ET11 chez T16 et NT10 au même locus

chromosomique que chez la lignée T3. L'élément ET69 était présent aussi bien chez T6 que T3, T12 et NT10 sur le chromosome 9. D'autre part, nous avons retrouvé les deux éléments ET92 et ET135 associés au chromosome 14 chez la lignée T12, également chez T14, NT1, tandis que l'ET135 était absent chez la lignée T6 et seul ET92 a été localisé au même locus. Enfin, la transposition d'ET18-1 a eu lieu uniquement chez la lignée NT10 sur le chromosome 20.

Les transpositions d'ETs observées dans ce travail semblent être des événements d'origine récente vraisemblablement qui se sont produites en amont du séquençage du génome de référence *O. tauri* RCC1115 par Blanc-Matthieu *et al.*<sup>70</sup> et par cette étude (inférieur à 10 ans). Au vu de cette analyse, nous pouvons affirmer que les transpositions apparaissent indépendamment de la transformation puisque les mêmes événements d'ETs ont été retrouvés à la fois dans les génomes des lignées transformées et des non-transformées. Par ailleurs, nous avons pu observer que certains de ces événements étaient partagés entre certaines lignées T et NT, suggérant que la transposition d'ETs chez *O. tauri* est un phénomène qui ne se produit pas totalement indépendamment au sein des génomes T et NT.

#### **4. Discussion**

*L'instabilité du phénotype de la souche RCC1115 au cours de l'infection virale, une conséquence d'importants réarrangements structuraux au sein de son génome*

Le premier objectif de ce travail était de gagner en connaissance sur les bases génétiques et fonctionnelles des mécanismes immunitaires associées à la résistance virale. Afin de déterminer la fonction de gènes candidats impliqués dans cette réponse, une stratégie moléculaire par transformation a été utilisée visant à interrompre aléatoirement l'expression de ces gènes sur la base de la méthode présentée dans le chapitre précédent (cf. Chapitre I). Puisque le génome d'*O. tauri* est haploïde avec seulement environ 8000 gènes, et de courtes régions intergéniques<sup>40</sup>, nous avons cherché au cours d'une transformation à « saturer » le génome en ADN transgène afin de cibler un maximum de gènes. Nous avons choisi la souche *O. tauri* RCC1115 dans ce contexte, puisque son génome était complètement séquencé<sup>70</sup> et qu'elle montrait une résistance à un plus large panel de souches de prasinovirus que la souche *O. tauri* RCC4221<sup>231</sup>.

Une sélection de plus de 1200 clones transformés issus de RCC1115 a alors été obtenue. Nous avons ensuite isolé des lignées clonales transformées ayant une réponse phénotypique à l'infection virale différente de celle de la lignée parentale RCC1115. Le choix de ces lignées candidates transformées, s'est porté sur le dépistage phénotypique pour la résistance à un virus dit « généraliste » (i.e. OtV09-578), pour lequel la souche *O. tauri* RCC1115 était sensible. Au vu du grand nombre de clones, nous avons réalisé une première sélection des candidats résistants à l'infection en culture liquide en plaque de 96 puits par suivi visuel pendant environ 3 semaines. Cependant, le phénotype de résistance ou de sensibilité aux virus dans les lignées clonales *O. tauri* RCC1115 s'est avéré très instable. En effet, la cinétique d'infection des cellules transformées révèle des fluctuations rapides de la croissance et de la lyse des cultures en milieu liquide. Par ailleurs, un bon nombre de cultures transformées et non-transformées présente un *switch* phénotypique, d'une résistance vers une sensibilité à l'infection sur un large panel de prasinovirus. Ces changements peuvent être expliqués par les importants réarrangements structuraux au sein des génomes. Chez l'algue *O. mediterraneus* une expérience d'évolution expérimentale a montré qu'une proportion des lignées de cultures clonales indépendantes, issues de cellules uniques, étaient capables de passer avec une fréquence de 12.5% à 22% d'un phénotype R à un phénotype S et de repasser à R sur des périodes très courtes. Le passage de R à S serait lié à l'existence d'une nette variation de la taille du chromosome SOC dans les lignées devenues sensibles par rapport aux lignées résistantes parentales. En revanche aucun changement systématique du SOC n'a été visible lié au passage de S vers R<sup>71</sup>. Dans la souche *O. tauri* RCC4221, des changements rapides de la taille du SOC ont été observés lors de l'acquisition de résistance à OtV5, et ces lignées résistantes contenaient également une population plus minoritaire de cellules sensibles<sup>310</sup>. D'autre part, des analyses antérieures réalisées sur le système *O. tauri*-OtV ont montré que les souches résistantes à une plus large gamme de virus ont tendance à avoir des tailles du SOC plus grandes<sup>70</sup> et une expression différentielle des gènes codés par le SOC<sup>310</sup>.

Dans l'ensemble de ces travaux, les auteurs proposent que les changements de phénotype chez l'algue *Ostreococcus* seraient associés à un processus de mutations affectant l'immunité virale du SOC que ce soit de R à S comme chez *O. mediterraneus* ou de S à R chez *O. tauri*.

### *L'activation des éléments rétrotransposables, une possible cause des remaniements chromosomiques*

Chez *O. tauri* et *O. mediterraneus*, plusieurs gènes peuvent être à l'origine de tels réarrangements massifs. De nombreuses copies d'intégrase et de transcriptase inverse ont été détectées sur le SOC et sont surexprimées pour le phénotype résistant<sup>310</sup>(Marcellin-Gros *et al*, en cours de soumission). Ces protéines sont associées à une famille d'éléments transposables, les rétrotransposons, trouvés en multiples copies dans le génome de nombreux organismes. Ces éléments génétiques mobiles sont considérés comme ayant un impact important sur l'évolution des génomes et contribuent ainsi fortement aux réarrangements chromosomiques, en augmentant ou réduisant la taille physique d'un génome<sup>406</sup>. Chez *O. tauri*, il a été proposé que les stress biotiques pourraient conduire à leur activation dans le génome, et entraîneraient ces réarrangements et variation en taille du chromosome 19<sup>310</sup>. Ce type d'activation serait aussi lié aux réarrangements chromosomiques observés chez les lignées ayant acquis une résistance à l'infection, puisque les cultures sensibles au virus ne montraient aucun changement du caryotype<sup>310</sup>. En revanche, les changements structurels et génétiques ont toujours été observés sur le SOC lors d'une infection virale chez la souche domestiquée *O. tauri* RCC4221 et chez *O. mediterraneus*<sup>71,310</sup>. Ce travail sur la souche *O. tauri* RCC1115 met en avant l'existence d'une dynamique qui s'étend sur l'ensemble du génome au sein d'une population clonale naturelle. La réactivation des rétrotransposons y a aussi été observée en même temps que des réarrangements chromosomiques. Cette réactivation pourrait promouvoir des variabilités structurelles génomiques à plus long terme et à plus grandes échelles.

### *L'intégration des ETs comme facteur de façonnage du devenir évolutif et adaptatif de l'hôte.*

Les ETs sont *à priori* des parasites obligatoires qui dépendent de leur hôte pour survivre et se répliquer. Des stratégies adaptatives de coévolution entre les deux partenaires sont nécessaires et reposent sur des forces opposées qui répriment l'activité des ETs tout en maintenant une dynamique de leur réactivation opportuniste, maintenant de la sorte une dynamique favorable à la co-habitation. Les ETs auraient donc développé des mécanismes propres dans le but de cibler des loci spécifiques où leurs insertions ne seraient pas nuisibles pour l'hôte mais favorables à leur propagation<sup>407</sup>.

De nombreux travaux tendent aussi à prouver que les ETs fournissent un réservoir de matériel génétique *in situ* pour la régulation et l'expression des gènes eucaryotes, par exemple en modulant le niveau de méthylation des gènes en amont. En effet, ils sont porteurs de nombreux éléments *cis*-régulateurs (i.e. promoteurs, activateurs, facteur de transcriptions, répresseurs) qui peuvent être dispersés dans le génome et affecter les réseaux transcriptionnels de l'hôte<sup>408-411</sup>. Chez le palmier à huile, la floraison serait contrôlée par le niveau d'expression d'un ET situé à proximité d'un gène important dans ce processus du développement floral de la plante<sup>412</sup>. Plus récemment, l'intégration d'un ET associée à une région régulatrice du gène *CrFLC* conduirait à une floraison plus précoce chez la plante *Capsella rubella*. Les auteurs proposent alors que la variation allélique causée par l'insertion de l'ETs en amont du gène *CrFLC* serait responsable du déterminisme phénotypique pour la différence de temps de floraison<sup>413</sup>. Les ETs auraient donc un rôle fondamental dans le façonnage du devenir du caractère phénotypique chez les organismes eucaryotes, en particuliers dans les organismes unicellulaires indépendamment de leur sexualité<sup>414</sup>. Étant donné qu'une importante variabilité du phénotype apparaît rapidement face à l'infection virale chez *O. tauri* RCC1115, il est tentant de spéculer que les ETs polymorphiques en soient à l'origine chez les lignées clonales qu'elles soient non-transformées ou transformées. En revanche, les mécanismes par lesquels les éléments mobiles influencent ces phénotypes restent à découvrir. Dans ce travail, aucune preuve ne permet de démontrer que les insertions des ETs induisent directement des changements de l'expression des gènes en altérant leur transcription ou en modulant la stabilité de leur ARNm. Des analyses expérimentales sont donc requises entre les différentes lignées clonales afin d'évaluer l'implication fonctionnelle de ces variations, plus particulièrement sur les gènes adjacents des sites d'insertion.

*Quelles seraient les causes potentielles de ces réactivations ?*

La dynamique des ETs est régulée à des niveaux différents par l'hôte et la régulation transcriptionnelle est la première étape pour limiter leur activation. Chez les plantes et les cellules animales, cette régulation répressive est souvent obtenue par l'établissement et le maintien d'une hétérochromatine répressive et/ou la méthylation de l'ADN, notamment par la médiation des petits ARN non-codants: les ARN interférents (*iRNA*)<sup>415-417</sup> et les ARN piwi (*piRNA*)<sup>418,419</sup>. Des perturbations épigénétiques peuvent donc être le résultat de la réactivation des ETs au sein d'un génome<sup>420,421</sup>. L'activation transcriptionnelle d'un grand nombre des rétrotransposons végétaux a été associée à divers facteurs de stress biotiques et

abiotiques<sup>404,422,423</sup>. Chez le tabac, des auteurs ont suggéré que l'expression de rétrotransposons *Tnt1A* était étroitement liée aux premières étapes des voies d'activation des gènes de défense<sup>424</sup>, ainsi que chez le maïs avec la transposition de l'élément *Bs1*<sup>425</sup>.

Chez *O. tauri*, une analyse transcriptomique des souches infectées ou non par OtV5 a été réalisée<sup>309</sup>. Elle met en avant une très forte induction de l'expression d'une transcriptase inverse associée à un rétrotransposon sur le chromosome 8 chez les cultures infectées, tandis que ce gène n'est pas transcrit dans les souches contrôles. L'expression de cette enzyme pourrait être activement régulée par l'infection virale. En parallèle, avec ces données de transcriptomiques, un mécanisme alternatif au *silencing* du chromosome 19 chez les cellules sensibles est suggéré par l'absence de corrélation entre la diminution de l'activité de la méthyltransférase et une augmentation de l'activité des déméthylases<sup>310</sup>. Au niveau épigénétique, le mécanisme moléculaire responsable de la méthylation *de novo* et la maintenance des CpG chez les Mamiellales est encore mal compris. Il a été proposé que l'ADN-méthyltransférase, DNMT5, soit impliquée dans la maintenance de méthylation des îlots CpG, comme DNMT1 et DNMT3. Toutefois, DNMT1 et DNMT5 apparaissent absentes chez les Mamiellales et semblent donc présenter un schéma de méthylation différent de celui des eucaryotes<sup>426</sup>. Il reste donc ici beaucoup à faire pour mieux comprendre les transactions possibles entre épigénétique, rétrotransposition et variation phénotypique chez *O. tauri*.

Bien que les réarrangements structuraux détectés dans cette étude semblent varier aussi bien dans les lignées transformées (T) que non-transformées (NT), une adaptation phénotypique à l'infection semble plus importante chez les lignées T. La transformation pourrait donc aussi impliquer – en relation avec l'intégration du vecteur ADN – des perturbations épigénétiques, conduisant à une sur-activation des éléments transposables dans les lignées T. En revanche, nous ne pouvons pas écarter l'hypothèse qu'un facteur abiotique conduise à l'activation des ETs chez les lignées NTs dans les cultures.





# CHAPITRE III – CARACTERISATION GENOMIQUE DES PRASINOVIRUS : DESCRIPTION DES BASES GENETIQUES ASSOCIEES A L'EVOLUTION DE LA GAMME DE L'HÔTE OSTREOCOCCUS TAURI

Ce chapitre présente l'article « *Insight into the genetic diversity of viruses infecting the green picoalga *Ostreococcus tauri* and the molecular bases host range evolution* », qui est actuellement en préparation pour une soumission dans le journal *Cell Host & Microbe*.

## **Objectifs du chapitre :**

Ce travail tient compte du dernier objectif de la thèse qui visait à gagner en connaissances sur la diversité des virus infectant l'espèce *Ostreococcus tauri* et ainsi de mieux comprendre les bases moléculaires de l'évolution de la gamme d'hôte.

## **Contributions de l'auteur :**

Ce travail a été réalisé en collaboration avec plusieurs auteurs. J'ai pris en charge toute la partie expérimentale et conduit l'ensemble des analyses bio-informatiques de ce projet. Le séquençage des génomes a été réalisé par la plateforme Bioenvironnement à l'UPVD de Perpignan. Vladimir Darric et Hélène Mayer de la plateforme bio-informatique à Banyuls-sur-mer, m'ont soutenue dans la conceptualisation de scripts bio-informatiques. Je tiens à remercier Morgan Gaïa pour son expertise dans l'analyse phylogénétique. Je tiens également à remercier la plateforme Genotoul à Toulouse pour l'accès aux outils bio-informatiques et à la maintenance informatique. Pour finir, j'ai rédigé l'article. Sheree Yau et Nigel Grimsley ont édité ce travail.

**Insight into the genetic diversity of viruses infecting the green picoalga  
*Ostreococcus tauri* and the molecular bases host range evolution**

Julie Thomy<sup>1</sup>, Vladimir Darric<sup>1</sup>, Helene Mayer<sup>1</sup>, Frederic Sanchez<sup>1</sup>, Nigel Grimsley<sup>1</sup>, Sheree Yau<sup>1</sup>.

**Author affiliations**

<sup>1</sup>Integrative Biology of Marine Organisms (BIOM), Sorbonne University, CNRS, Oceanological Observatory of Banyuls, Banyuls-sur-Mer, France.

**Corresponding authors:** [sheree.yau@obs-banyuls.fr](mailto:sheree.yau@obs-banyuls.fr), [nigel.grimsley@obs-banyuls.fr](mailto:nigel.grimsley@obs-banyuls.fr)

**Supplementary figures** in appendix chapter III; page 212.

**All supplementary tables** are found on an online server for purposes of space:

<https://www.php.obs-banyuls.fr/publications/data/3/>

## Résumé

Au travers d'une analyse génomique comparative de 19 génomes entiers de virus d'*Ostreococcus tauri* (OtVs) d'origine géographiquement distincte, ce chapitre met en lumière la diversité des *Prasinovirus* à l'échelle mondiale. Sur la base de leur diversité phylogénétique, trois sous-groupes ont été distingués. La plupart des OtVs de l'Océan Pacifique présentent une relation phylogénétique étroite avec les virus *Micromonas*, suggérant que le genre *Micromonas* peut être un hôte alternatif dans leur environnement naturel ou un changement récent d'hôte de *Micromonas* à *O. tauri*. Par ailleurs, les génomes d'OtVs exposent un génome globalement synténique mais interrompu par une région hypervariable. Cette région est d'autant plus variable chez les OtVs qui montrent un plus large spectre d'infection. Cette région plastique encode principalement des gènes impliqués dans la reconnaissance et l'adhésion à la cellule hôte. Au travers de l'ensemble de ces analyses, nous suggérons que cette région pourrait être un déterminant clé de la gamme d'hôtes d'OtV, évoluant dans une dynamique de course aux armements avec son hôte.

## Abstract

The *Prasinovirus* genus comprises double-stranded DNA viruses that are members of the *Phycodnaviridae* within the “giant virus” *Nucleocytoviricota* phylum. Prasinoviruses are thus far known to specifically infect cosmopolitan unicellular green algae in the order Mamiellales (class Mamiellophyceae). They are easily isolated from coastal marine waters, where their host species are the most abundant. Even though several prasinoviruses have been isolated and characterized to date, they are the most abundant of the giant viruses in global ocean metagenomic surveys indicating much remains to be discovered about their diversity and how they interact with their hosts. In this context, we performed a comparative genomic analysis of nineteen whole *Ostreococcus tauri* virus (OtV) genomes that were isolated from three different marine sites and that show distinct infection patterns, as well as a *Micromonas commoda* RCC827 virus, newly isolated from the Pacific Ocean. Three subgroups of OtV were distinguished on the basis of their genetic diversity. Most OtVs from the Pacific Ocean showed a close phylogenetic relationship with *Micromonas* viruses, suggesting that the genus *Micromonas* may be an alternative host in their natural environment or a recent host switch from *Micromonas* to *O. tauri* in this subgroup. In addition, the OtV genomes showed a genome that was globally syntenic but interrupted by a hypervariable region. The genes encoding the hypervariable region are involved in recognition and adhesion to the host cell. We suggest that this hypervariable region is a key determiner of OtV host range, evolving in an arms-race dynamic with its host.

## 1. Introduction

Viruses are described as the most common and diverse biological entities on earth and in the oceans, even though they represent only a tiny fraction of the biomass due to their small size<sup>95,96</sup>. The importance of algal viruses in marine ecosystems became clear after a series of observations of virus-like particles associated with important bloom-forming algae<sup>207,427–429</sup>. Through host cell lysis, viruses have been widely regarded as major contributors to the microbial loop in marine food webs by facilitating recycling of organic matter and nutrients *via* the viral shunt<sup>256,258,430</sup>. They also drive the evolution of their hosts through selection and genetic exchange, as well as modify host population dynamics and microbial community structures<sup>431,432</sup>. Furthermore, viral predation promotes genetic diversity by reducing the most abundant populations and allowing the rarest species to grow<sup>268</sup>.

One answer to how the equilibrium between host and virus populations can be maintained<sup>272</sup> is by an arms-race dynamic<sup>288</sup>. This race proposes that constant antagonism between two interaction species drives reciprocal selection for survival of both partners that eternally change their pathogen virulence or host resistance, accelerating their simultaneous genetic evolution (coevolution). Theoretically, an arms-race dynamic results in a “nested” infection pattern in which hosts span a continuum of resistance to viral strains and viruses equally span a continuum in their ability to infect host strains<sup>282</sup>. This is mediated by a gene-for-gene (GFG) process of sequential adaptations whereby host mutations confer resistance to ancestral viruses and new viral mutations allow host range expansion while maintaining their ability to infect ancestral hosts<sup>280,282,286,287</sup>. Viral genetic modifications have been mainly observed to occur in genes related to host attachment for virus entry<sup>433</sup>, which appears to be a key determinant of viral host range<sup>434–437</sup>. For instance in phage–bacteria systems, the conserved tail fiber genes involved in attachment to host cells often evolve in response to acquired resistance by the host<sup>392</sup>. However, the genetic basis of host range in eukaryotic marine viruses remains poorly investigated.

Viruses infecting the cosmopolitan green picoalga *Ostreococcus tauri* (OtVs) were first reported by Derelle *et al*<sup>229</sup>. They were found to co-occur with their hosts, which inhabit in a wide range of marine environments, but especially predominate in Mediterranean lagoons<sup>36,61</sup>. OtVs belong to the *Prasinovirus* genus in the *Phycodnaviridae* family and appear to be the smallest members of the “giant” viruses of the *Nucleocytoviricota* phylum in terms of virion (~200 nm capsid diameter) and genome size (~200 kb encoding ~250 genes).

The infection pattern of *O. tauri* strains and OtVs isolated from the Mediterranean Sea showed a nested structure suggesting that arms-race dynamics is occurring in the environment<sup>231</sup>. Thus far in prasinoviruses, the DNA polymerase B (*polB*) gene has been used as a genetic marker to explore the relationship between genetic diversity evolution and host range. Bellec *et al.* demonstrated the absence of correlation between geographical and phylogenetic distance based on the partial viral *polB* gene<sup>217</sup>. Clerissi *et al.* showed that although OtVs strains with similar *polB* genotypes tended to infect the same host strains, the *polB* gene alone underestimated OtV host range diversity<sup>231</sup>. Hence, complete genomes are required to investigate the determiners of host–virus interactions.

Given the lack of knowledge on the global diversity prasinoviruses, we first performed a comparative phylogenomic analysis of eighteen newly sequenced whole OtV genomes that have distinct geographical origins and infection patterns, a *Micromonas commoda* (strain RCC827) virus isolated from the Pacific Ocean, and the complete published *Prasinovirus* genomes. In addition, we investigated the genetic bases of viral host range expansion in OtVs at the whole genome-level. Our data suggest that broad host ranges in OtVs can occur under a wide variety of ecological conditions, and we propose they result from mutations, genetic exchange and recombination events concentrated in a unique hypervariable region encoding for host cell recognition and attachment.

## 2. Material and Methods

### *Culture of host algal and viruses*

All algal cultures and viruses (**Figure 24**) were grown in liquid L1 medium (NCMA, Bigelow Laboratory for Ocean Sciences, USA) made with autoclaved offshore seawater (MOLA station: 42°27'11"N, 3°8'42" E), diluted with Milli-Q water to give a final salinity of 30 g·L<sup>-1</sup>, and filter sterilized through 0.22 µm filters. The cultures were kept at 15°C using a photoperiod cycle of 12 h: 12 h light: dark (50 µmol photon m<sup>-2</sup> s<sup>-1</sup> white light).

### *Isolation of new viruses and host-virus specificity tests*

Viruses from the Mediterranean Sea had been previously isolated between 2006 and 2009<sup>231</sup>. One virus from the North Sea and six viruses from the Pacific Ocean were isolated in this study. These viruses were isolated from half a liter of seawater collected from each marine

station site (**Table S3**) that was filtered through a 0.45- $\mu\text{m}$  PES membrane filter (SARSTEDT reference; 83.1826), concentrated ten times by tangential flow filtration (100 kDa PES Vivaflow200, Satorius, Germany) and then clonal viruses were procured by plating for individual plaques in exponentially growing *O. tauri* RCC4221 cultures, as described previously<sup>229</sup>. We considered that the virus strains were clonal after three successive infections of the host strain by the viruses from one lysis plaque. Fresh viral lysate was stored at 4°C. For host–virus specificity tests, a panel of Mamiellophyceae strains from the Roscoff Culture Collection ([www.roscoff-culture-collection.org](http://www.roscoff-culture-collection.org)) (*Ostreococcus tauri* RCC4221, RCC1123, RCC1117, RCC1116, RCC1110, RCC1115, RCC1108, RCC1114, RCC1561; *O. mediterraneus* RCC789; *M. commoda* RCC827; and *B. prasinus* RCC4222) were screened for their susceptibility to the viruses analyzed in this study. Algal strains in exponential growth were incorporated into semi-solid L1 medium by addition of 1.5% molten agarose held at 58°C in a water bath (0.15% final concentration) and poured into Petri dishes (9 cm diameter). Three microliters of each viral lysate were deposited in duplicate on the surface of the semi-solid medium for each algal strain and the susceptibility to viral lysis was evaluated by the appearance of plaques visualized after 7 days.

#### *Partial DNA polymerase B gene amplification and sequencing*

We determined that the newly isolated viruses were prasinoviruses based on PCR screening of the partial *polB* gene, which is a viral marker gene, using the primer sets in Clerissi *et al.* 2014<sup>218</sup>. PCR reactions were done as follows: 2  $\mu\text{L}$  of fresh viral lysate was added to a 48  $\mu\text{L}$  of reaction mixture containing 1X PCR buffer, 1 mM  $\text{MgCl}_2$ , 0.2 mM deoxyribonucleoside triphosphate (dNTP) mix, 0.5  $\mu\text{M}$  of each primer, 0.1  $\text{mg}\cdot\text{m}^{-1}$  BSA and 1.25 U of Taq DNA polymerase (GOTaq® G2 Flexi M7805, Promega). PCR was performed in a Mastercycler Nexus system (Eppendorf) using an initial denaturation at 95°C (3 min), 38 rounds of denaturation at 95°C (30 s), annealing at 49°C (60 s), extension at 72°C (3 min) and final extension at 72°C (4 min). PCR products were loaded into a 0.8% agarose electrophoresis gel in 0.5% TAE (Tris-acetate-EDTA) buffer and visualized under ultraviolet light after ethidium bromide staining. Before Sanger sequencing, PCR products were purified using a Wizard® SV gel and PCR clean-up system kit (Promega, ref A9281). Sequences were cleaned by manually trimming low quality base pairs from the 3' and 5' ends using Geneious® software (version 2.1).

### *Concentration and purification of viral particles*

A total of two liters of viral lysate from each strain was produced and filtered through a 0.2- $\mu\text{m}$  PES membrane filter (SARSTEDT, ref 83.1826) to remove cellular debris and bacteria. Viral particles filtered were concentrated 40-fold by tangential flow filtration using a 100 kDa MWCO PES membrane Vivaflow200 unit (Satorius, Germany, ref VF20P4). We concentrated the filtered lysate a second time with a 50,000 MWCO PES membrane Vivaspin20 ultracentrifugation unit (Satorius, Germany, ref VS2031). Viral particles were further purified in an iodixanol (Optiprep™, Sigma-Aldrich, ref D1556) density gradient by ultracentrifugation rotor (SW41 Ti rotor, Beckman Coulter) at 175,117 g for 5 h at 4°C. For each virus, a single visible band in the 30% fraction, corresponding to the purified concentrated viral particles, was collected and pelleted by ultracentrifugation at 245,418.9 g for 2.5 h at 4°C (SW55 Ti rotor, Beckman Coulter). The pellets were resuspended in a sterile SM buffer (100 mM NaCl, 8mM MgSO<sub>4</sub>-7H<sub>2</sub>O, 50mM Tris pH 7.5) and stored at 4°C.

### *Extraction and sequencing of viral DNA*

DNA was extracted from the purified viral-fraction with the CTAB protocol, as previously described<sup>348</sup> with the following modifications. After CTAB buffer incubation (65°C for 30 min), an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added and the samples were centrifuged (15,000 g for 10 min at 4°C). The aqueous phase was recovered, mixed with an equal volume of chloroform:isoamyl alcohol (24:1) and centrifuged (15,000 g for 10 min at 4°C). 2  $\mu\text{L}$  of RNase A (1 M) (DNase free) was added to the aqueous phase for 30 min at 37°C, and then the enzymes were inactivated by heating for 10 min at 60°C. One tenth volume of ammonium acetate (3 M, pH 5.3) was added in order to favor DNA precipitation. Nucleic acids were precipitated using 2.5 volumes of absolute ethanol, gently homogenized and transferred at -80°C overnight. The precipitated DNA was centrifuged at 15,000 g for 30 min at 15°C and the pellet washed in 70% ethanol. Finally, the samples were centrifuged (15,000 g for 10 min at 15°C) air-dried and DNA dissolved in 50  $\mu\text{L}$  of DNase/RNase free water and stored at 4°C. DNA in 1  $\mu\text{L}$  was quantified by fluorimetry (Quantus™) using QuantiFluor® dsDNA system kit (Promega, ref E2670).

Viral DNA (1  $\mu\text{g}$ ) was submitted to Illumina NEXTFLEX library preparation (average insert size 241 bp) and sequenced using Illumina NextSeq 550 technology in paired-end mode (2x150 bp) giving a yield of 120 million reads total. Reads were cleaned with Trimgalore

software (v0.6.5) (with options --paired --fastqc --illumina --quality 26 --stringency 3 --length 35 --three\_prime\_clip\_R1 10 --three\_prime\_clip\_R2 10 --clip\_R1 10 --clip\_R2 10) and the read quality evaluated with FastQC (v0.11.7) with default settings.

### *De novo genome assembly*

We found that optimal *de novo* assembly of viral genomes required read data down-sampling, as using the entire dataset with estimated coverage of 6,000x induced assembly fragmentation. Down-sampling was performed using the software seqtk version 1.3 (<https://github.com/lh3/seqtk>). A total of eight independent sets of randomly down-sampled reads (seed numbers 100) were obtained given a coverage of 5x, 10x, 50x, 75x, 100x, 500x, 1,000x and 2,500x for each virus genomic library. *De novo* assembly was performed using Spades (v3.15.1)<sup>438</sup> for all sets of independently down-sampled reads, as well as for the total set of cleaned reads. In order to ensure optimal assembly, and since some samples were revealed to have two independent viruses, we further assessed two assembly options: ‘--isolate’ and ‘--meta’. The 18 assemblies for each viral strain were compared and the choice of the ‘best’ assembly was based on optimizing for 3 quality criteria given by QUAST (v5.0.0)<sup>439</sup>. Specifically, we sought to maximize the size of the longest contig and to reduce ambiguous nucleotide positions (lowest number of N’s) and the total number of contigs. To facilitate the comparative genomic analysis, genomes comprising several contigs were manually scaffolded using Geneious software (v11.0.3+7), to ensure that the order of genes was conserved relative to that of viral genomes that were assembled as a single contig. The reads were mapped to the final genome assembly with the Burrows-Wheeler Aligner (BWA v0.7.17) using maximum exact matches (BWA-MEM) and a minimum seed length of 20 while keeping the other default parameters in order to further assess a correct assembly<sup>440</sup>. Specifically, we verified re-mapped read depth was even along the length of the genome.

### *Gene prediction and functional annotation*

Gene prediction of OtV genomes was performed using Prokka (v1.14.5) specifying ‘virus’ as the taxonomic kingdom<sup>397</sup>. First, protein-coding genes (CDS) were annotated using the databases implemented in Prokka with the following parameters: e-value, 1e-5; and genetic code, standard (--gcode 1). Secondly, tRNA were predicted using tRNAscan-SE with default parameters (v2.0.2)<sup>441</sup>. Thirdly, additional functional annotation of all protein sequences was

sought by BLASTp search against the NCBI nr database accepting an e-value of  $<1e-5$  keeping only the best hit. Finally, taxonomic affiliation was associated for each protein accession using Entrez Direct (EDirect) (epost -db protein | esummary -db protein | xtract -pattern DocumentSummary -element Caption,TaxId)<sup>442</sup>.

### *Comparative genomic analysis*

The complete genomes of the previously available prasinoviruses were downloaded from NCBI (**Table S1**). A total of 37 genomes were used to define groups of orthologous genes (orthogroups), where we excluded some already published prasinovirus genomes that were incomplete or too fragmented. All translated CDS were included in the orthogroup analysis using Orthofinder with default settings<sup>443</sup>. Each orthogroup was annotated *via* the EggNOG-mapper toolkit (v2.0.4-*rf1*) against EggNOG protein database (-m diamond) using setting parameters<sup>444</sup>. The Euclidean distance between the viruses based on the presence/absence pattern of orthogroups was calculated using pvelust R package (v2.2-0) based on the “ward.D2” hierarchical clustering linkage method and approximately unbiased (AU) *p*-values are computed using 1,000 bootstrap replicates (default settings). Orthogroup distribution in each genome was represented using the pheatmap R package (v1.0.12). Pairwise whole-genome alignments of OtVs were carried out using Mauve software version 2.3.1 with default parameters<sup>445</sup>. Average Nucleotide Identity (ANI) between all orthologous genes shared between all pairs of OtV genomes was computed using FastANI (v.1.32)<sup>446</sup>. Genomic maps of the hypervariable region between specialist and generalist viruses were generated using Easifig v2.2.2<sup>447</sup>.

### *Phylogenetic analysis*

Phylogenetic reconstruction based on both the partial and full-length DNA polB protein and 19 core proteins shared between *Prasinovirus* and *Chlorovirus* genomes was performed. Protein sequences were aligned using MAFFT (v.7.313) with the L-INS-i algorithm<sup>448</sup>. We trimmed each alignment removing positions containing more than 80% gaps with Goalign (v0.3.2)<sup>449</sup>. Phylogenetic trees of single-proteins and concatenated proteins (comprising 5483 amino acid positions) were built using Maximum-Likelihood (ML) method with IQ-TREE version 2.0.6<sup>450</sup>. For mixture model analysis, the LG+F+R4, LG+R3 and LG best models were chosen (option -m MFP) according to the Bayesian information criterion (BIC) for

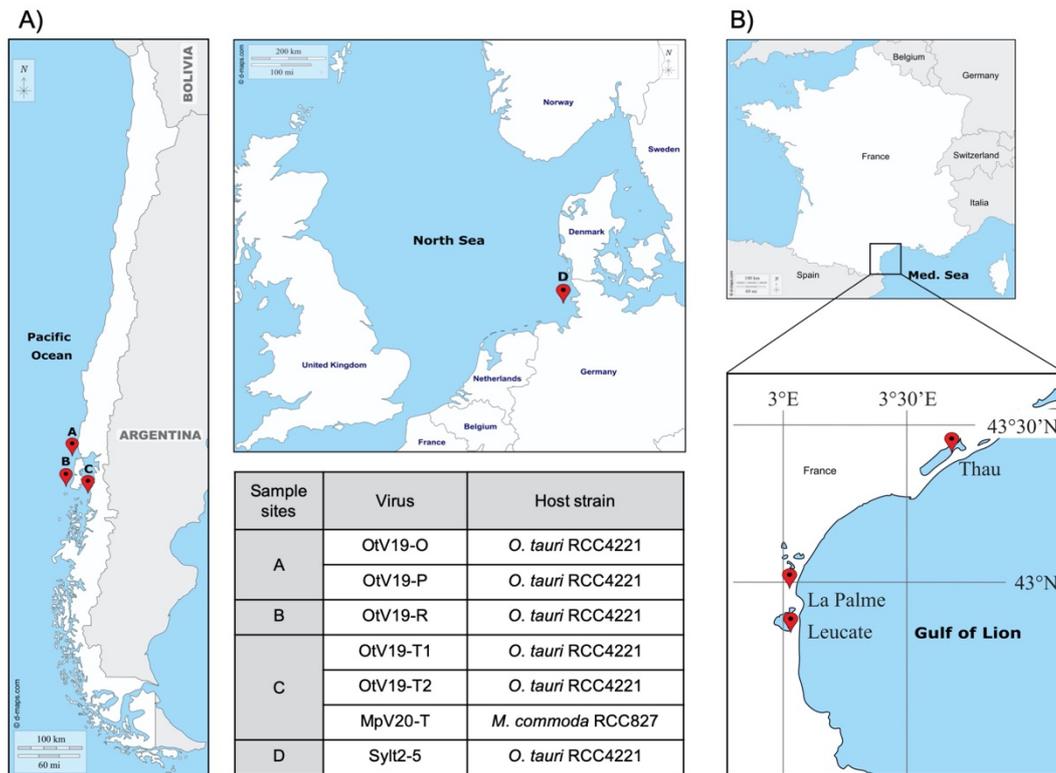
concatenated proteins (as well as the potassium ion channel protein), full-length and partial DNA polB protein, respectively<sup>451</sup>. The branch support values were computed from 1,000 replicates for the Shimodaira-Hasegawa (SH)-like approximation likelihood ratio test (aLRT)<sup>452</sup> and 1,000 ultrafast bootstrap approximation (UFBoot)<sup>453</sup>. The trees were visualized with Interactive Tree Of Life (iTOL) v6<sup>143</sup>(last version unpublished).

### 3. Results and discussion

#### *Isolation of viruses from distant geographical origins from their host*

The main goals of this work were to gain insight into the genome-wide diversity of OtVs and to identify genes that may determine host specificity. Hence, we first sought to isolate new prasinoviruses from worldwide sampling sites in order to expand the diversity of viruses from our initial set of 12 previously isolated genetically distinct OtVs<sup>231</sup> from the Mediterranean Sea (**Figure 23**). Surface seawater samples collected from two globally distant locations (coastal south Pacific Ocean and North Sea) (**Figure 23**) were tested for their ability to lyse different clonal microalgal strains (**Table S3**). Interestingly, we were able to obtain lysis plaques on the strain *O. tauri* RCC4221 (cloned from RCC745), which itself had been isolated from a Mediterranean Sea lagoon in France<sup>61</sup>, and therefore from a very distant geographical origin. Additionally, we observed cell lysis using seawater from the Pacific Ocean on the warm thermotype strain *M. commoda* RCC827 (originally RCC299), which had been isolated from a coastal equatorial Pacific site<sup>389</sup>.

In total, we isolated six viruses infecting *Ostreococcus tauri* and one infecting *Micromonas commoda* from surface seawater samples. To determine if newly isolated viruses belong to the genus *Prasinovirus*, like all previously isolated OtVs, amplification of the viral *polB* gene with degenerate primers and sequencing was performed, showing all new viruses were indeed members of *Prasinovirus*.

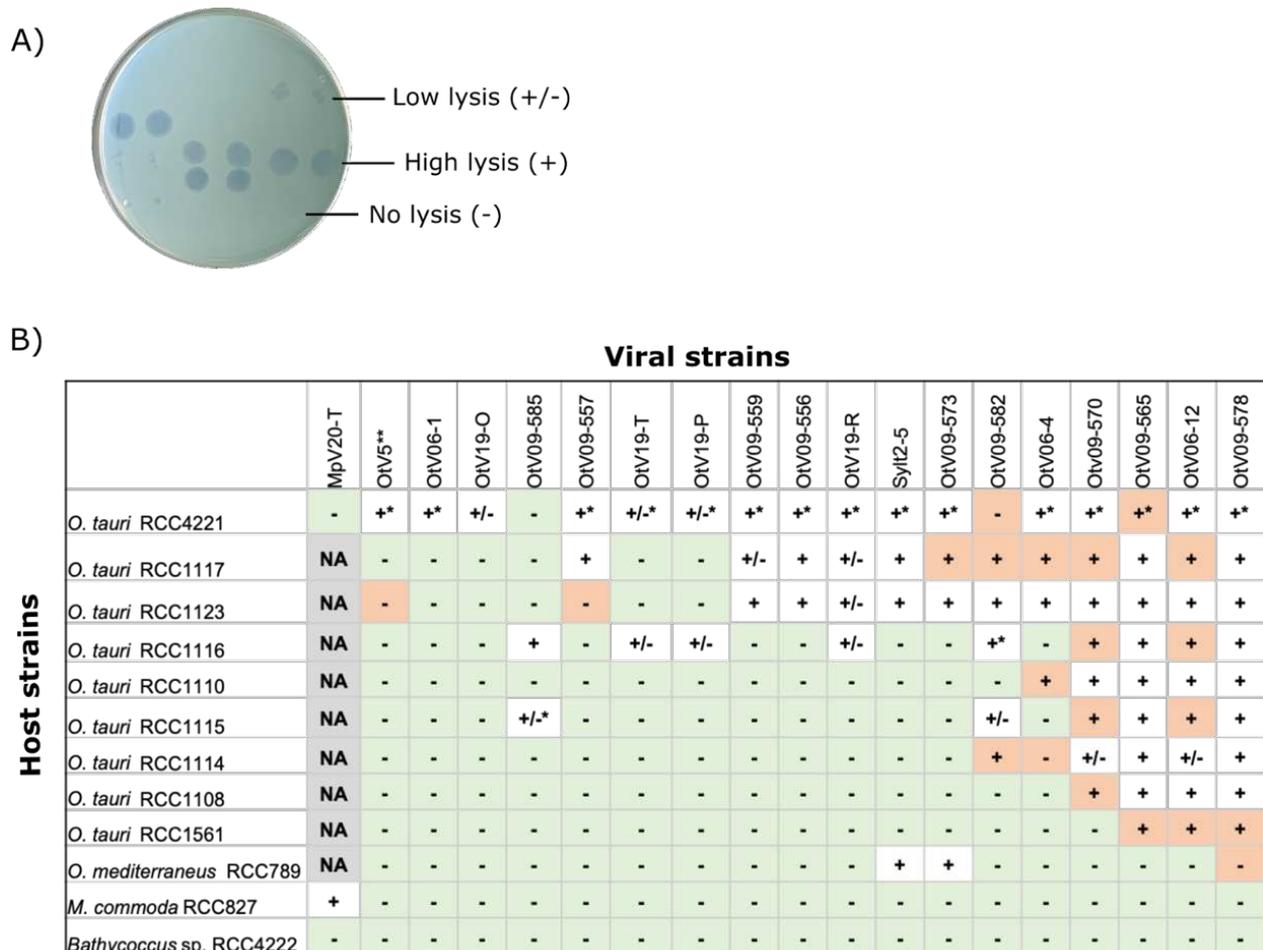


**Figure 23.** Locations of the seawater sampling sites. A) Sampling sites off the Chilean coast in the Pacific Ocean and the German coast in the North Sea from where the six *Ostreococcus tauri* viruses and the *Micromonas commoda* virus were isolated in this study are plotted in red on the map. B) Sampling sites in the Mediterranean Sea and lagoons in southern France from where 12 OtVs had been isolated in a previous study<sup>231</sup>. The exact geographical coordinates for each site and each virus are found in supplemental data (**Table S1**).

#### *The O. tauri–OtV infection pattern displayed a nested structure*

Cross-infection tests between OtVs and a large panel of *Ostreococcus* strains, as well as representative strains of the related genera *Micromonas* and *Bathycoccus*, were performed (**Figure 24**) to determine viral host ranges. The infection pattern showed a “nested” or hierarchical structure, as described in a prior report, where viral host range existed on a continuum from specialist viruses infecting a restricted range of host strains to generalist viruses infecting a wider range of hosts<sup>231</sup>. All OtVs were strictly *Ostreococcus* specific, since no lysis was observed for *Micromonas* and *Bathycoccus* strains. Interestingly, Sylt2-5 and OtV09-573 (from the North Sea and Mediterranean, respectively) showed cross-species specificity as they were able to infect *Ostreococcus mediterraneus*.

Three viruses isolated from the Pacific Ocean (OtV19-O, OtV19-P, OtV19-T) only infected one to two host strains like the most specialist viruses (OtV5, OtV06-1, OtV09-557, OtV09-585) previously isolated from the Mediterranean Sea. Altogether, this indicates that the host ranges of OtVs was not strongly related to geographic origin, but rather all viruses fit along a continuum of host specificity suggestive of arms-race dynamics in natural populations.



**Figure 24.** Host–virus cross-infection pattern. A) Example of plaque assay test showing low and high lysis and no lysis results. B) Cross-infectivity tests of OtVs against a panel of *Ostreococcus tauri* strains and representative species of the most closely related genera within the Mamiellophyceae class. Minus (-) corresponds to no lysis observed, plus (+) indicates lysis was observed three days post infection and plus and minus signs (+/-) indicate lower lysis, where lysis plaques were smaller or less turbid. Note that if a phenotypic change was observed compared to the data of Clerissi *et al.* 2012<sup>231</sup>, we marked the box in orange. NA: Cross infectivity was not tested. \*Strain used to produce the virus. \*\*Strains sequenced in Derelle *et al.* 2008<sup>229</sup>.

## Characteristics of viral genomes

Most of the genomes were assembled to form a single contig with a size ranging from 180 to 200 kb, and inverted repeat sequences were observed at each end of most of the genomes, suggesting full assemblies (**Table 9**).

**Table 9.** General characteristics of nineteen newly sequenced prasinovirus genomes. \*number of CDSs/number total predicted genes x 100.

Virus	Genome size (bp)	Contig (no.)	CDS (no.)	Predicted gene (no.)	Mean gene size (bp/gene)	Coding density (genes/kb)	Coding proportion (%)*	GC content (%)	tRNA (no.)	Capsid (no.)
MpV20-T	209,762	1	265	271	774	1.292	97.8	40.58	6	8
OtV06-1	188,487	1	246	247	763	1.305	99.5	44.27	4	8
OtV06-4	190,369	1	248	251	758	1.302	98.8	44.66	5	8
OtV06-12	191,735	1	231	234	819	1.205	98.7	43.69	5	9
OtV09-556	190,575	1	251	254	750	1.317	98.8	44.73	5	8
OtV09-557	189,897	1	247	250	760	1.300	98.8	44.69	5	8
OtV09-559	189,416	1	249	252	752	1.314	98.8	44.63	5	8
OtV09-565	191,124	1	232	235	813	1.214	98.7	43.79	5	8
OtV09-570	190,769	1	250	253	754	1.310	98.8	44.90	5	8
OtV09-573	190,656	1	245	248	769	1.285	98.8	45.04	5	8
OtV09-578	183,592	1	235	238	771	1.280	98.7	43.65	5	8
OtV09-582	191,259	1	248	248	771	1.296	100	44.71	4	8
OtV09-585	193,432	1	251	254	762	1.298	98.8	44.75	5	8
OtV19-O	189,391	1	268	275	766	1.415	97.5	41.86	8	8
OtV19-P	191,803	1	272	279	687	1.418	97.5	41.91	8	8
OtV19-R	193,132	11	248	251	769	1.284	98.8	44.64	5	7
OtV19-T1	189,558	1	267	274	692	1.408	97.5	41.89	8	8
OtV19-T2	200,246	2	251	253	791	1.253	99.2	41.18	5	9
SYLT2-5	189,609	1	247	250	758	1.303	98.8	44.68	5	8

The MpV20-T genome was assembled into a single contig with a larger size genome, and a GC content lower than the OtVs but similar to the other *Micromonas* viruses described so far<sup>454</sup>. Despite stringent purification of viral strains from each seawater sample by picking of individual virus plaques through three rounds of plating, one of the viral lysates, OtV19-T, contained two separate viral genomes, which we designated OtV19-T1 and OtV19-T2 (**Table 9**). This came to light when we found two different *polB* genes (a single-copy gene in *Prasinovirus*) on two separate contigs. Overall, the OtVs have a similar genome size, an average of 44% GC content, significantly lower than their corresponding hosts (~59%)<sup>40</sup>, a similar

number of ~250 predicted coding DNA sequences (CDSs) (98.7% of gene coding), an average gene length of 761 bp, a coding density of 1.31 genes per kbp, and an average of 8 capsid proteins (**Table 9**). Four viruses showed genomic characteristics divergent from other OtVs previously described. With the exception of OtV19-R, four viruses isolated from the Pacific Ocean (OtV19-O, OtV19-P, OtV19-T1 and OtV19-T2) displayed clearly lower GC content (~41%), as well as generally more tRNA genes and higher coding density (**Table 9**).

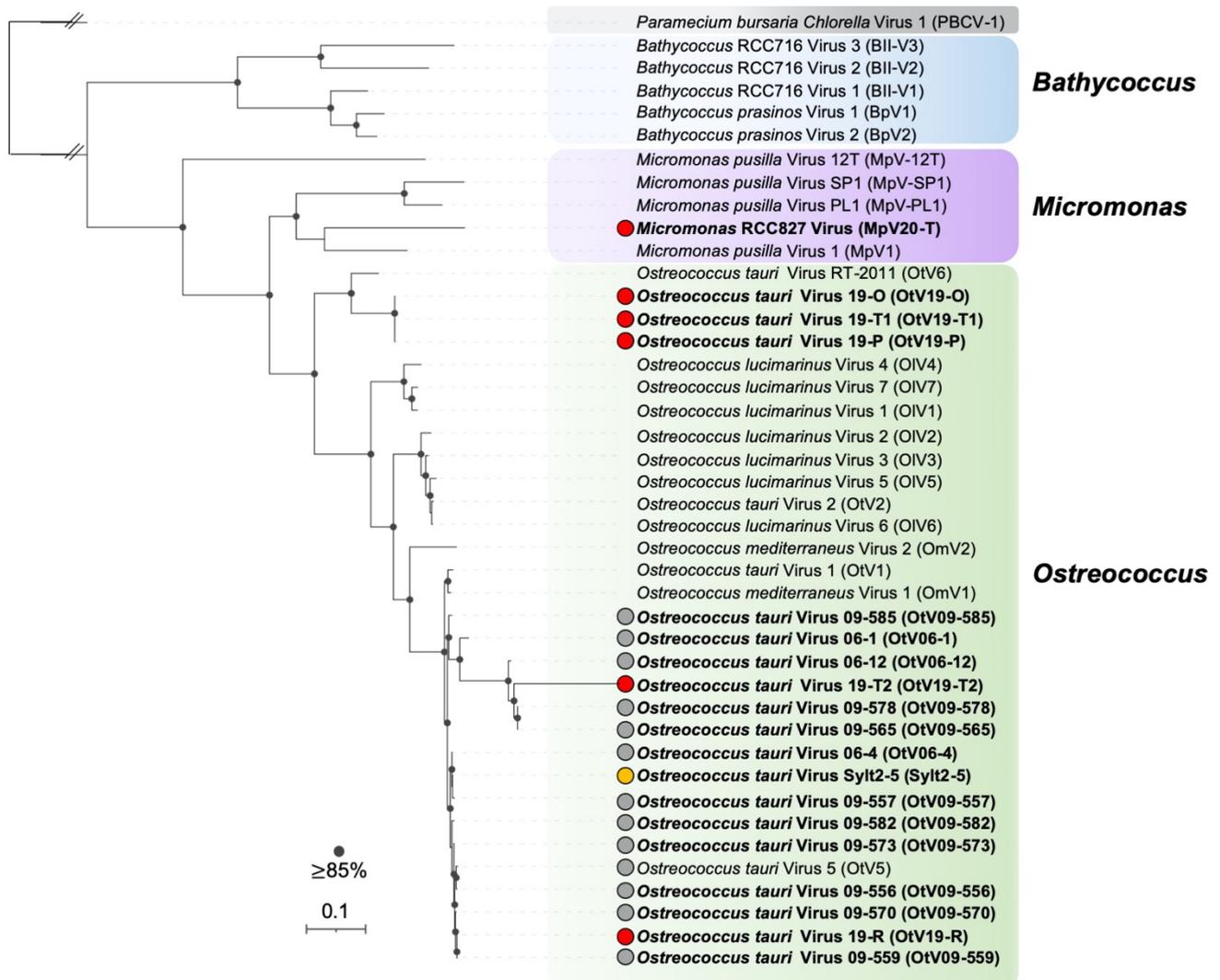
### *Phylogenomic analysis of prasinoviruses*

We performed a phylogenetic analysis based on concatenation of 19 conserved proteins that are a subset of core genes previously determined to be shared between *Prasinovirus* and *Chlorovirus* (present in at least 80% of the genomes)(**Table S2**)<sup>59,227,232</sup>. We chose 19 of 22 previously defined core genes, since 3 of them (i.e. proliferating cell nuclear antigen homolog 1 – PCNA1, Patatin-like phospholipase and ribonucleoside-diphosphate reductase large subunit) were duplicated in some viral genomes, resulting in incongruences in relationships. We constructed a robust phylogenetic tree including 23 reference genomes (**Figure 25**).

The concatenated phylogeny grouped viruses infecting algae from the same genus, except for MpV-12T, which formed an independent group from the other viruses infecting *Micromonas*. Overall, the evolutionary relationship observed agreed with previous work<sup>43,59,227,232</sup>. Likewise, the DNA polymerase B protein phylogenetic tree displayed the same topology as the concatenated alignment (**Figure S1**). In the clade of *Micromonas*-infecting viruses, phylogenetic reconstruction grouped MpV20-T with the MpV1. The *Ostreococcus* viruses formed a single large clade divided into four distinct subclades. The three Pacific Ocean viruses OtV19-O, OtV19-T1 and OtV19-P clustered together with a virus from the Mediterranean Sea OtV6, as the earliest diverging branch of *Ostreococcus* viruses from *Micromonas* viruses. The *O. lucimarinus* viruses (OIVs) segregated into two subclades as previously observed<sup>227</sup>.

In contrast, all other OtVs clustered together with *Ostreococcus mediterraneus* viruses, strongly supporting them belonging to a single subclade. Viruses in this large *O. tauri*/*O. mediterraneus* virus subclade were mainly isolated from the Mediterranean Sea; although Sylt2-5, as well as OtV19-T2 and OtV19-R, which were isolated in this study from the North

Sea and the Pacific Ocean, respectively, were also part of this subclade. Intriguingly, we revealed that within this large *O. tauri*/*O. mediterraneus* virus subclade there was an internal fast-diverging branch, encompassing OtV06-12, OtV09-565, OtV09-578 and OtV19-T2. Three of these OtVs (OtV06-12, OtV09-565 and OtV09-578) also were the viruses with the broadest host ranges, infecting all nine *O. tauri* strains tested (**Figure 24**).

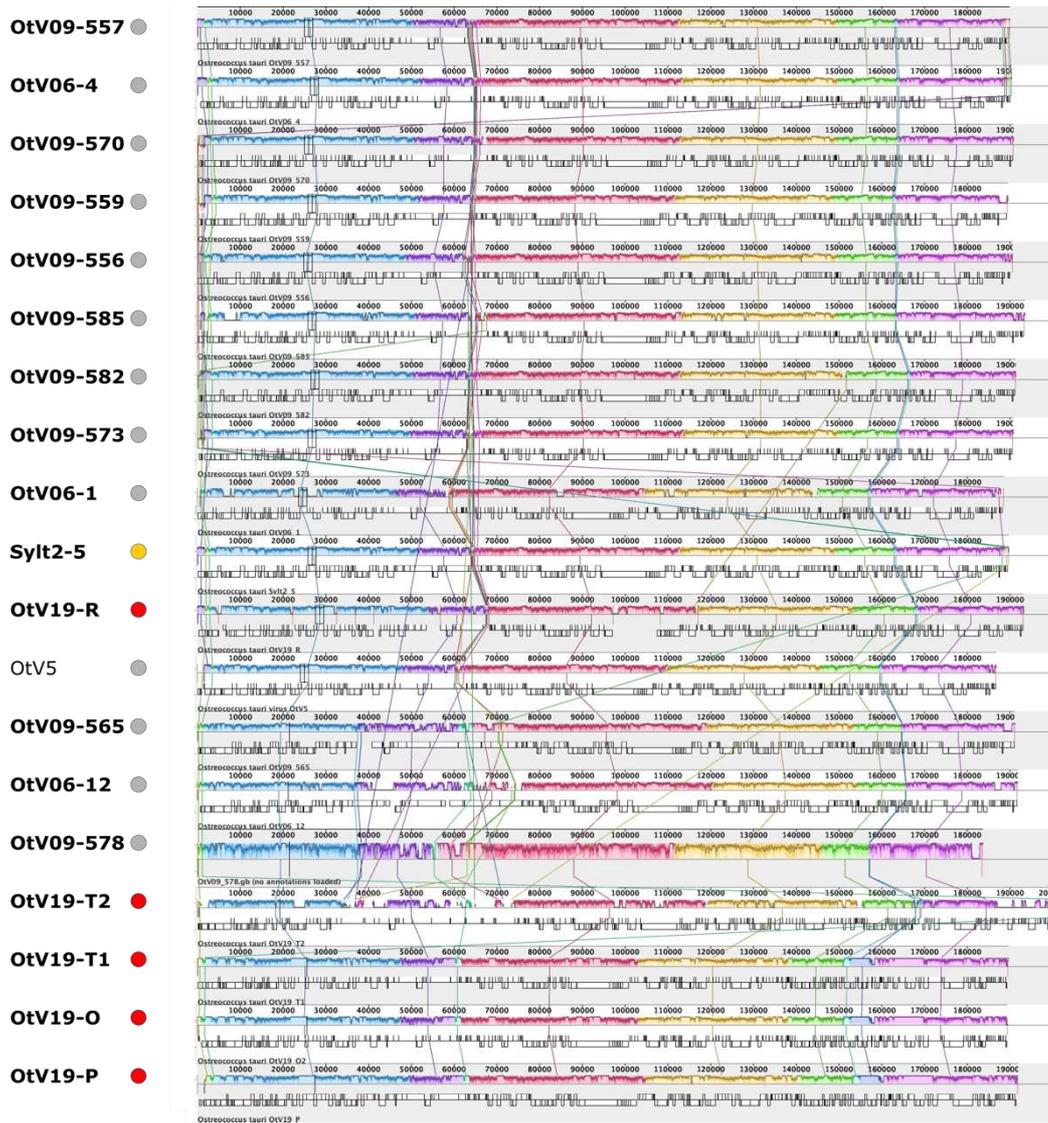


**Figure 25.** Phylogenetic analysis of prasinoviruses. Maximum-likelihood (ML) phylogenetic reconstruction of nineteen core proteins shared among the viruses (5,482 positions) infecting the genera *Bathycoccus* (blue block), *Micromonas* (purple block) and *Ostreococcus* (green block) and *Chlorovirus* (PBCV-1), which was used as an outgroup with the connecting branch truncated for display. Nodes with bootstrap support over 85% are indicated by black dots. The scale bar represents the average number of substitutions per site. Virus names in bold were sequenced in this study and their isolation site is shown by colored dots: red, Pacific Ocean; yellow, North Sea; and gray, Mediterranean Sea and lagoons.

### *Genome-wide alignment of OtVs shows high synteny*

Whole-genome alignments of OtVs showed high collinearity and gene synteny between all genomes (**Figure 26**). The divergence of OtVs was very low with a pairwise average nucleotide identity (ANI) values ranging from 71% to ~99% (**Table S4**). A structural organization of gene functions was observed, represented as Locally Collinear Blocks (LCBs), of which major LCBs were conserved between all viruses. The blue block clustered a broad range of genes associated with carbohydrate, nucleotide and amino acid biosynthesis and metabolism, while the purple block included genes with unknown functions and three capsid proteins. The central blocks (red and yellow) contained the bulk of the core genes, which were associated with DNA transcription, replication and post transcriptional modification. Finally, the green block located towards the 3' end of the genomes encoded genes involved in amino acid and polysaccharide synthesis and metabolism. The last block at the 3' end (pink) was less distinct in its functional complement, but encoded core genes including two capsid genes and DNA replication genes, including DNA polymerase B and DNA topoisomerase II.

Minor LCBs, not shared by all viruses, were observed at the genome extremities, as well as two others minor LCBs located between the major conserved blocks just described. The most complex rearrangement was observed between the purple and red blocks (close to position 60,000 bp in the OtV5 genome), identified as a hypervariable region among viruses. This hypervariable region corresponded to an additional cyan block present in OtV09-565, OtV06-12, OtV09-578, OtV19-T1, OtV19-T2, OtV19-P and OtV19-O. The genes encoded in this region comprised chaperones of endosialidase domain-containing proteins (peptidase S74 family; PF13884) (OG0000405, OG0000211, OG0000205, OG0000001; **Table S5**). Another visible variable feature from the whole-genome alignments was the expansion of the purple block in OtV09-565, OtV06-12, OtV09-578, OtV19-T2, which starts in the position 40,000 bp in these viruses and included an upstream dark purple block of ~2 kb in length. These were the same viruses found to be in a faster diverging branch in the phylogenetic analysis (**Figure 25**).



**Figure 26.** Comparison of the complete genomes of nineteen *Ostreococcus tauri* viruses (OtVs). Colored blocks correspond to conserved regions that are locally collinear and homologous between viral genomes. Such segments are referred to as Locally Collinear Blocks (LCBs). More complex rearrangements are shown by the connecting lines between the LCBs. OtV isolation site is shown by colored dots next to the virus name: red, Pacific Ocean; yellow, North Sea; and Gray, Mediterranean Sea and lagoons. OtV5 was isolated and sequenced in Derelle *et al.*, 2008.

The expanded part of the purple block encoded one gene (OtV09-565\_00057, OtV09\_12\_00055, OtV09\_578\_00057 and OtV19\_T2\_00053) associated with the trimeric autotransporter adhesin (TAA) family, which is known to play a role in the cell recognition and adhesion, and shared 40% amino-acid identity with *Acanthocystis turfacea* *Chlorella* virus Canal-1 (Genbank accession: AGE50088.1), as well as other chloroviruses and prasinoviruses.

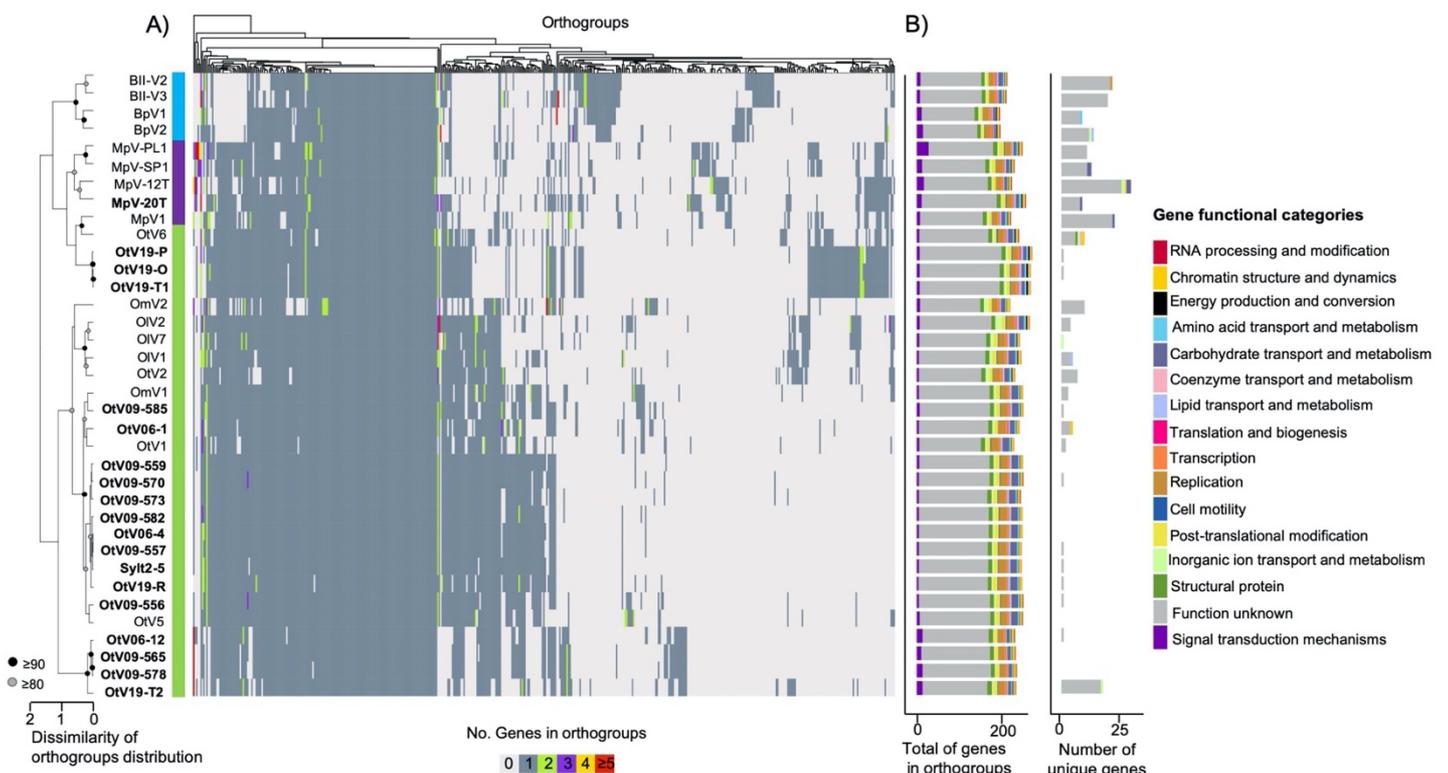
The second variable region (blue-purple block) was observed inserted in three viruses isolated from the Pacific Ocean (OtV19-O, OtV19-P and OtV19-T1) towards the 3' part of the genome (corresponding to position 160,000 bp in OtV5). Interestingly, this region was only in these three OtVs, which all belonged to the same phylogenetic subclade (**Figure 25**). This region encoded thirteen genes, of which only three had a known function. One gene encoded for dCMP deaminase (OG0000248; **Table S5**), which is involved in the nucleotide transport and metabolism. The other two genes were associated with DNA replication, recombination and repair. They were the DNA polymerase X (OG0000244; **Table S5**) and the SNase (staphylococcal nuclease homolog) (OG0000231; **Table S5**), which shares ~36% amino-acid identity with *Mimiviridae* sp. ChoanoV1 (Genbank accession: QDY52382.1).

#### *Gene family analysis in the whole prasinovirus genomes*

Gene family analysis was used to explore the evolutionary relationship of the OtVs with other published prasinovirus genomes. Of the 483 orthogroups (OGs), 87 core proteins were shared among the 36 prasinoviruses of which 73 were present in single copy (singletons). Among the core prasinovirus proteins, functional categories could be assigned to 31 of them. The most representative functions were DNA replication, recombination and repair (14%), three were related to transcription and three others were involved in the post-translational modification. Note that the capsid protein represents one of the core genes present in a high copy number that was distributed throughout the prasinovirus genome. An all-versus-all ortholog presence and absence pattern enabled us to perform a hierarchical clustering of the complete set of *Prasinovirus* genes and thus to describe their evolutionary relationship at a whole genome-scale (**Figure 27**). The topology of the ortholog clustering was fairly similar to the multi-protein tree (**Figure 25**). The main difference is that *Ostreococcus* viruses do not form a single large cluster. Instead, taking their whole genome content into account appeared to separate OtV groups according to other factors shaping the non-core genes or pan-genome, which potentially include specific environmental or host-interacting genes.

As the multigene phylogeny (**Figure 25**) and genome-wide alignment (**Figure 26**) revealed broad clustering of OtVs based on both geographic origin and host range, we sought to test if there were any strict geography–gene content and host specificity–gene content correlations. We tested three specific questions: 1) Are there any genes shared only among the

viruses isolated from the Pacific Ocean? 2) Are there any genes shared only among viruses with a broad host specificity pattern? 3) Conversely, are there any genes shared only among viruses with narrow host specificity pattern? Based on these questions, we separated the OtVs into three groups. The first group included OtV19-O, OtV19-P, OtV19-R, OtV19-T1, and OtV19-T2, which were isolated from the Pacific Ocean; the second group encompassed OtV06-12, OtV09-565, OtV09-570 and OtV09-578, which were the most generalist viruses described in this study infecting at least 8/9 *O. tauri* strains tested (**Figure 24**); and the third group was composed of OtV5, OtV06-1 and OtV19-O, which infected only *O. tauri* RCC4221. No gene families could be associated only with geographical origin or infection pattern, even though broad trends were observed, suggesting evolutionary and ecological constraints both shape the genomic diversity and phenotypes of OtVs.



**Figure 27.** Orthologous protein distribution of prasinoviruses. A) Heatmap showing the distribution and frequency of all orthogroups in *Bathycoccus* (vertical blue stripe), *Micromonas* (purple stripe) and *Ostreococcus* (green stripe) viruses. Virus names in bold were sequenced in this study. The heatmap was ordered vertically by hierarchical clustering of viral genomes based on the occurrence pattern of all orthogroups (left dendrogram) and horizontally by clustering of the orthogroups based on their distribution in the viral genomes (top dendrogram). Nodes marked with a gray dot show genome clusters with over 80% bootstrap support and black dots show over 90% support. B) Bar charts showing

(left) the total number of genes assigned to an orthogroup and (right) the number of unique genes (not shared with any other viruses analyzed) in each viral genome. Bar charts colored according to the distribution of functional categories based on the EggNOG database.

Therefore, we propose OtV groups according to both their genome-wide diversity (**Figure 27**) and their evolutionary relationships described previously (**Figure 25**). OtV-type 1 includes the three Pacific Ocean viruses (OtV19-O, OtV19-P and OtV19-T1) and OtV6, which form a large cluster within the *Micromonas*-infecting viruses clade. OtV-type 2a comprises all OtVs (except OtV2, which is more closely related to OIVs group). Finally, we define a subgroup OtV-type 2b that includes the fast-diverging strains OtV06-12, OtV09-565, OtV09-578, and OtV19-T2 (**Figure 27**).

The OtV-type 1 viruses shared 18 genes (specific to these viruses and present in 3 out of 4 genomes) of which 2 of them have a predicted function. These proteins include an alternative oxidase (AOX) protein (OG0000373; **Table S5**). This enzyme forms part of the electron transport chain, not only in mitochondria in eukaryotic cells but also in prokaryotic organisms, and the expression of this gene is thought to be influenced by different stresses such as pathogen infection and ROS in the host cell<sup>455</sup>. An inverse correlation between the expression of AOX protein and the ROS production has been demonstrated in plant cells<sup>456</sup>, suggesting a way for the cell to resist these stresses in maintaining the cellular metabolic and energy balance<sup>457</sup>. The AOX protein has been reported in the freshwater alga *Chlamydomonas reinhardtii*<sup>458,459</sup> and many other algae<sup>460</sup>. We observed that these viruses have acquired the AOX protein by a recent horizontal gene transfer (HGT) event from their host (**Figure S2B**). The second gene shared encoded a chaperone of an endosialidase domain-containing protein (peptidase S74 family; PF13884) (OG0000405; **Table S5**). Remarkably, endosialidases are components of the tail proteins commonly found in tailed bacteriophages in the order *Caudovirales*. This protein has been studied for its specific degradation of  $\alpha$ 2,8-linked polysialic acid of bacterial polysaccharides and its role in adhesion molecule N-CAM to the bacterial host cell<sup>461,462</sup>. The tail and fiber proteins are components involved in virus-host interactions, and are key determinants of host specificity. Since these endosialidases appear to be specific to OtV-type 1 viruses, which could thereby have evolved host recognition factors independently of other OtV groups, perhaps allowing them to recognize divergent hosts in natural environments or recognize their hosts by an alternative mechanism.

According to their gene composition, OtV-type 1 viruses seem to share a common evolutionary history and thus common biological functions with *Micromonas* viruses. OtV-type 1 viruses and all *Micromonas* viruses share 6 genes (present in 8 out of 9 genomes) of which only 1 of them had a predicted function. This protein included a mannitol dehydrogenase domain near its C-terminal (OG0000281; **Table S5**), which catalyzes the reaction of D-mannitol 1-phosphate to D-fructose 6-phosphate in bacteria<sup>463</sup>. Interestingly, a copper/zinc superoxide dismutase protein (Cu-Zn SOD) (OG0000303; **Table S5**) is present in the OtV-type 1 viruses and *Micromonas* viruses, except for MpV12-T and MpV20-T. This metalloprotein plays an important role in protecting the organisms from reactive oxygen species (ROS) by catalyzing the conversion of superoxide radicals to hydrogen peroxide and molecular oxygen<sup>464</sup>. The Cu-Zn SOD protein was reported to be present in several NCLDV families, such as poxviruses<sup>465,466</sup>, baculoviruses<sup>467</sup>, some *Mimiviridae* members<sup>468</sup> and chlorovirus<sup>469</sup>. This protein shows high similarity to many *Chlorovirus* proteins (~60% of amino-acid identity) whose molecular function has been described in PBCV-1<sup>469</sup>. The authors demonstrated that virus-encoded Cu-Zn SOD enzymes were active and decreased ROS in the cell as a strategy to manipulate its host and promote its replication during the earlier stages of infection. The OtV-type 1 and *Micromonas* viruses may therefore have retained this gene from the ancestral group that gave rise to both *Chlorovirus* and *Prasinovirus*. Furthermore, a mitochondrial 18 kDa protein (MTP18) (OG0000304; **Table S5**) was found in the OtV-type 1 strains and both MpV-PL1 and MpV1. BLASTP searches showed that the best hits originated from the green algae. Therefore, a phylogenetic tree was constructed and confirms a recent HGT event occurred from the alga host to its virus (**Figure S2A**). In mammalian cells, MTP18 is a housekeeping protein localized in mitochondria, and the decline in the level of this protein leads to a signaling cascade causing programmed cell death in the cell<sup>470</sup>. Taken together, we speculate that the OtV-type 1 viruses acquired these genes to hijack the molecular mechanism of their host to express and replicate their genome. Taken together, we speculate that the OtV-type 1 viruses acquired these genes as specific virulence factors. The fact that some genes are shared by OtV-type 1 and *Micromonas*-infecting viruses suggests a possible switch of host from *Micromonas* to *Ostreococcus* occurred in this lineage.

Although the OtV-type 2a subgroup appeared to be closely related phylogenetically, no common group-specific genes were detected. These viruses share many genes with some *Ostreococcus* viruses and even more widely with *Prasinovirus*. These data suggest that OtV-type 2a may be more 'general' viruses that gave rise to more specific viruses (i.e. OtV-type 2b

or even OtV-type 1) through the acquisition of specific genes, particularly for cell recognition and adhesion function, during their evolutionary history.

The OtV-type 2b viruses shared 10 genes (specific to these viruses and present in 3 out of 4 genomes) of which 2 of them have a predicted function. It is particularly interesting to note that one of these two proteins were a coiled stalk of TAA (OG0000365; **Table S5**). As described above, TAA is known to play a role in host cell attachment, as outer membrane proteins in many pathogenic Gram-negative bacteria of animals (e.g. YadA of *Yersinia enterocolitica*<sup>471</sup> or BadA of *Bartonella henselae*<sup>472,473</sup>). TAAs are trimeric fibers subjected to significant rearrangement during their evolution, resulting in considerable diversity in terms of sequence, and their lengths can vary by over almost two orders of magnitude<sup>473</sup>. More interestingly, in our work, these proteins have been found in the most variable region of the OtV genomes (**Figure 26**). The unique genome organization, gene composition (**Figure 27**), and host specificity patterns (**Figure 24**) are therefore consistent with OtV-type 2b viruses having independently evolved an expanded host range through HGT and rearrangements of genes involved in host cell recognition and adhesion. Finally, the second shared gene was *N*-acetylglucosaminyltransferase III, which belongs to glycosyltransferase family 17 (OG0000362; **Table S5**) and is a transmembrane protein involved in transferring *N*-acetylglucosamine (GlcNAc) from UDP-GlcNAc to a core mannose with a  $\beta$ 1,4 linkage of *N*-glycosyl oligosaccharides<sup>474</sup>. In animal cells, the addition of this residue drives structural changes in glycoprotein oligosaccharide chains, which has functional consequences for receptor signaling, cell adhesion, and tumor progression<sup>475-477</sup>.

*Ostreococcus tauri* viruses exhibit a virulence region that might be a key determinant of viral host range.

Previously in this work, the genome-wide alignments of OtVs distinguished a hypervariable region between the purple and red blocks near position 60,000 bp in OtV5 that was expanded in the OtV-type 1 and OtV-type 2b viruses. To go further, we sought to describe this region in more detail by conducting a genomic comparison limited to this hypervariable block in all OtVs, shown in (**Figure 28**). As most of the genes in this region had no match by similarity search with BLASTX (e-value of 1e-5) (**Table S6**), we extended the annotation by BLASTP search (e-value of 1e-5) against Refseq, Interproscan, and Conserved Domains

Databases (CDD), as well as by protein structural predictions using Phyre V2.0 (**Table S6**). The genes in this region were color-coded into broad functional categories based on conserved domains and their structural domain homology (**Figure 28**).

At the 5' and 3' ends of the hypervariable region, the genes were syntenic between all OtVs with a high level of sequence similarity. The genes that occurred in all OtV genomes on the 5' side of the hypervariable region included an ABC1 atypical kinase-like domain, NAD(P)-binding Rossmann-like domain, the DEAD-like helicase protein and a capsid protein. The conserved genes on the 3' side of the hypervariable region were a second capsid, NUDIX domain and a prim-pol4 protein.

The two capsids protein sequences at each end allowed the delineation of a more complex central region, which showed less sequence identity between the different groups. The gene order was only maintained in OtV-type 1 and OtV-type 2a groups. The OtV-type 2b showed a more complex and variable architecture between strains. Note that a third copy of the capsid gene is found in this central region in all OtVs genomes, with the exception of both OtV06-12 and OtV19-T2 viruses that have a duplication of this gene.

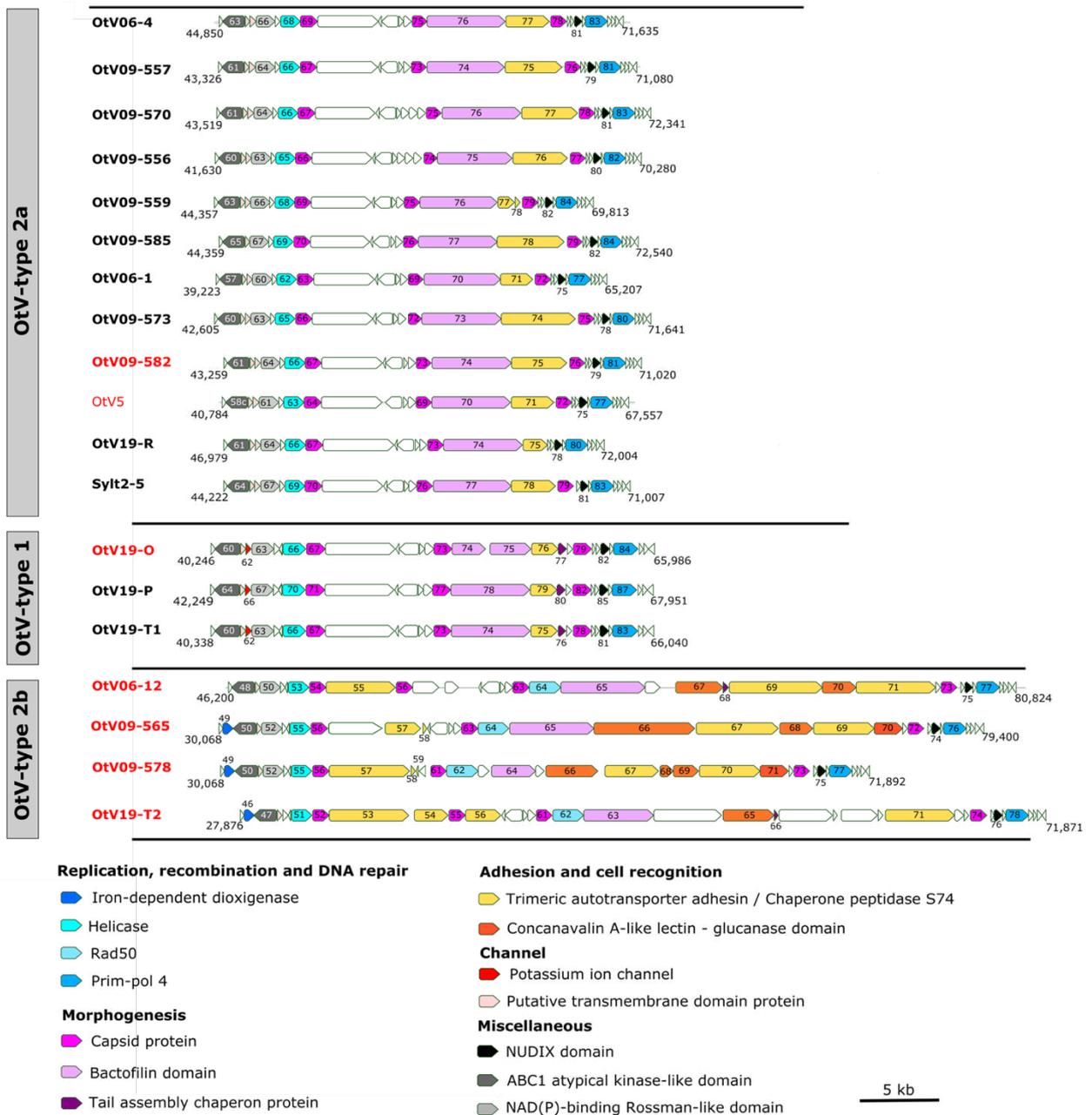
In OtV-type 1 and OtV-type 2a, the set of genes at the 5' end of the central region encoded unknown functions. In contrast, the downstream end carries genes with conserved protein domains, known as TAA/chaperone peptidase S74, which are involved in adsorption (Yellow; **Figure 28**) and host cell recognition, as well as a small bactofilin domain (Pink; **Figure 28**). Remarkably, variation in length of the TAA-containing domain protein was observed mainly in OtV-type 2a. In addition, a high levels of sequence similarity occurred only between OtVs of the same group, whereas synteny and the sequence identity were poorly conserved among OtVs of different groups (**Figure 28** et **Figure S3**). Although OtV-type 1 showed conservation of the size of putative TAA/chaperone peptidase S74 gene among different viruses, we were surprised to identify a gene coding for a putative phage Tail Assembly Chaperone protein (TAC) (OtV19\_O\_00077) (dark purple; **Figure 28**). A homolog of the TAC gene was also found in OtV06-12 (OtV06\_12\_00068) and OtV19-T2 (OtV19\_T2\_00066) genomes, but with less conservation of sequence identity. TAC is present in *Siphoviridae* phage genomes, where it is required for the tail assembly of all long-tailed phages<sup>478</sup> by acting as a chaperone for tape measure proteins<sup>479</sup>. In tailed phage genomes, TACs are located between the genes encoding the tail tube and tape measure proteins and possess

overlapping ORFs that are related by a translational frameshift. In our study, this gene is located downstream of the putative TAA/chaperone peptidase S74 domain-containing protein.

In OtV-type 2b, the central region appeared to be highly complex with lower synteny than the first two groups. Even in otherwise closely related OtVs, the sequence and structure of this locus had diverged substantially. TAA/chaperone peptidase S74 homologs in OtV-type 2b have duplicated and diverged, producing what appear to be chimeric proteins which showed a mixed protein domain. Chimeric proteins can be created by crossovers either between small homologies within the structural part of the gene or in conserved motifs. These data suggest that OtVs might employ different classes of adhesin-like proteins that likely function in the interaction with their hosts to determine host range, and it is variability of this locus that mediates the acquisition of diversity in host-range determinants.

OtVs-type 2b genomes uniquely had several concanavalin A-like lectin/glucanase domain-containing proteins (Orange; **Figure 28**), located between the last two capsids of the central region. Although these genes displayed highly conserved sequence domains, they were highly variable in size. Proteins containing lectin domains recognize sugar structures and mediate a variety of biological processes, such as cell–cell and host–pathogen interactions, serum glycoprotein turnover and innate immune responses<sup>480</sup>. Proteins that interact with carbohydrates occur widely in nature and are found in most organisms, ranging from viruses and bacteria, to plants and animals<sup>481</sup>. Several proteins with C-type lectin-like domains have been reported in animal viruses, especially in *Poxviridae*<sup>482</sup>, *Herspesviridae*<sup>483</sup> and *Asfarviridae*<sup>484</sup>, where they are involved in the hemadsorption process, cell to cell spreading and many other functions.

A DNA double-strand break repair ATPase Rad50 protein (Cyan; **Figure 28**) was found in all OtV-type 2b genomes adjacent to a capsid protein and upstream of the variable central region. In eukaryotes, the Rad50 protein is involved in the DNA recombination and repair mechanism<sup>485,486</sup>. Hence, we speculate that the OtV-type 2b group generate hypervariation in this region through recombination events mediated by this Rad50 protein, suggesting they have an infection strategy similar to that described in the tailed phages<sup>487</sup>.



**Figure 28.** Genomic map comparing the hypervariable region of the 18 newly sequenced OtVs (Bold). The OtV5 genome was used as a reference genome. Genomes were sorted according to their OtV-type based on their phylogenetic relationship and gene family composition. The search for gene function and/or conserved domain and structural domains was performed in the virus names in red that served as a reference for each type (**Table S6**). Homologous genes were colored identically according to the functional categories. White open reading frames have no predicted function. Each open reading frame was numbered according to its associated locus tag. BLASTn comparisons between the OtV06-12, OtV09-578, OtV09-565, OtV19-T12, OtV5, OtV09-582, and OtV19-O genomic regions were shown in **Figure S3**.

### *The OtVs encode a putative potassium ion channel protein*

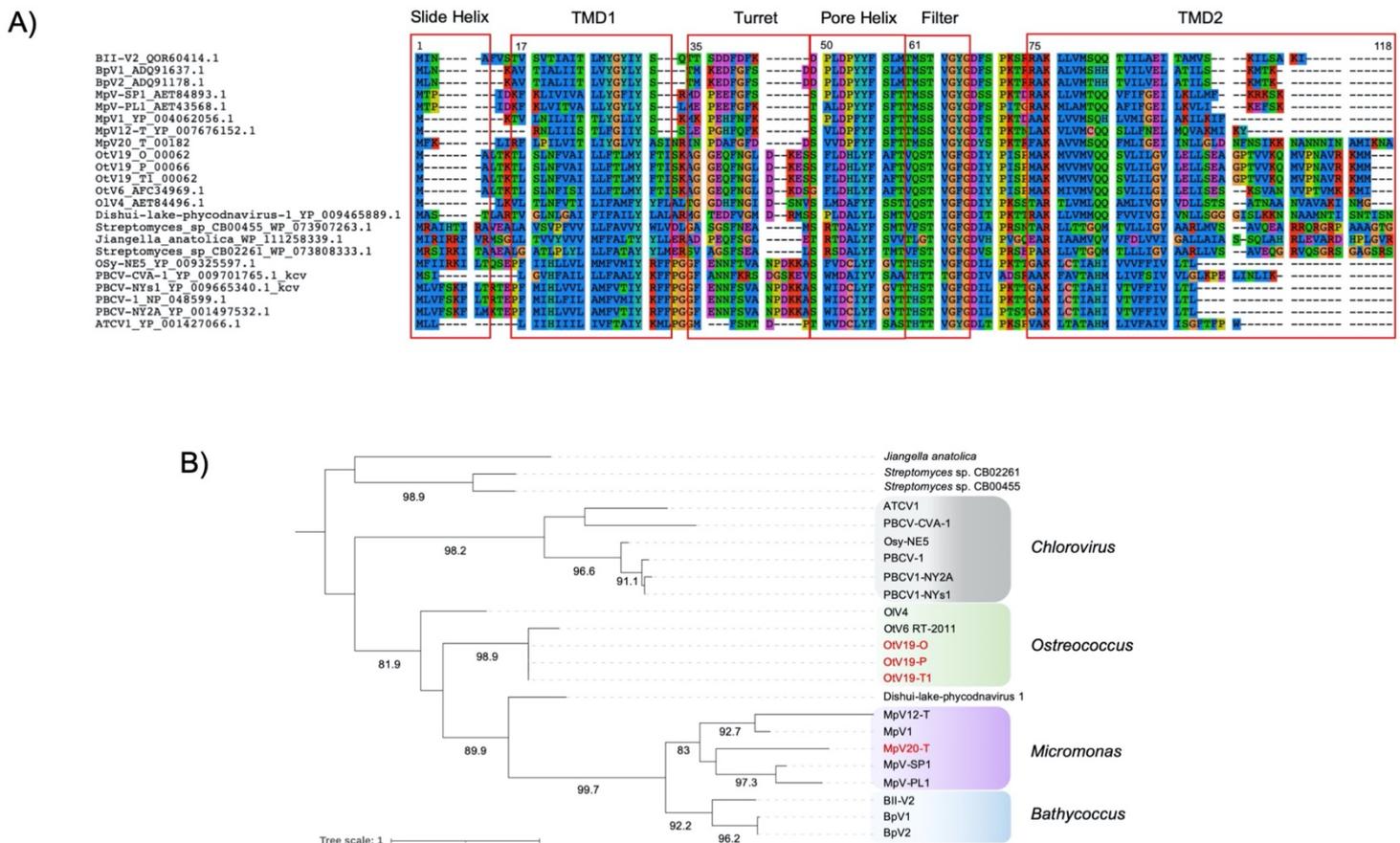
The steps used by prasinoviruses to initiate infection is still an unexplored question. In the closest relatives of prasinoviruses, the chloroviruses, considerable experimental data support that canonical K<sup>+</sup> channels, localized in the internal membrane, play a key role during in the early phase of viral infection. By fusing the virus membrane with the host membrane<sup>488</sup> viral K<sup>+</sup> channels mediate host cell membrane depolarization<sup>489</sup> enabling viral DNA ejection into the cell<sup>490</sup>.

In this work, we predicted from sequence similarity search that only OtV-type 1 encoded a K<sup>+</sup> ion channel protein (OtV19\_O2\_00062) at the 5' end of the hypervariable module (Red; **Figure 28**). The K<sup>+</sup> channel has been associated to a gene family (OG0000247; **Table S5**), which is shared between all *Micromonas* viruses, including the MpV20-T virus isolated in this study, as well as all *Bathycoccus* viruses (except BII-V3). Surprisingly, no hits to the K<sup>+</sup> channel were found for the other OtVs in this study, and only the two *Ostreococcus* viruses OtV6 and OIV4 are known to encode this gene<sup>491</sup>. However, in this work, several clues support the idea that all OtVs encode the K<sup>+</sup> channel gene, albeit with no significant similarity detected by BLAST. Firstly, the conservation of gene order among OtV genomes showed a small ORF in the same relative position as the K<sup>+</sup> channel gene in OtV-type 1 strains. Secondly, the small sequence size of this ORF corresponds with that of known K<sup>+</sup> channels. Finally, these small ORFs had predicted transmembrane domains (using a prediction program, TMHMM2.0 algorithm<sup>492</sup>) consistent with them being K<sup>+</sup> channels (Pink; **Figure 28**). A detailed analysis of the protein in PBCV-1 (ksv protein) showed that K<sup>+</sup> channel consisted of a short slide helix, an outer transmembrane helix (TMD1), a turret, a pore helix, a filter and an inner transmembrane helix (TMD2)<sup>493</sup>, which together form a functional K<sup>+</sup> channel.

To better describe the K<sup>+</sup> channel diversity in the OtV19-O, OtV19-P, OtV19-T1 and MpV20-T, an amino-acid alignment was carried out including the ksv protein and the three best hits of the BLASTP search (e-value 10<sup>-5</sup>) in the Refseq database NCBI (**Figure 29A**). The most conserved regions for all K<sup>+</sup> channels were the pore helix and the filter, while the turret, the TMD1 and TMD2 transmembrane appears to be more variable but fairly conserved between each genus (**Figure 29A**). Moreover, unlike kcv, the K<sup>+</sup> channel length in the OtVs-type 1

showed a larger amino-acid sequence (106 aa) and shared around 25% amino-acid identity with the chloroviruses.

Phylogenetic reconstruction of the K<sup>+</sup> channel protein showed that although *Chlorovirus* and *Prasinovirus* are genetically distant, they nevertheless share an ancient common ancestor. However, only homologous proteins in bacteria gave the highest BLAST scores (**Figure 29B**); these results support the work described in the chlorovirus model suggesting a bacterial origin<sup>491,494,495</sup>.



**Figure 29.** Genomic and evolutionary analysis of the potassium ion (K<sup>+</sup>) channel protein. A) Multiple amino-acid alignment of the K<sup>+</sup> channel protein described in *Prasinovirus* and *Chlorovirus*, as well as the OtV19-O, OtV19-T1, OtV19-P and MpV20-T by MAFFT v7.313 with L-INS-i algorithm, and 20% of sites in the alignment with gaps were removed using goalign v0.3.2. Different functional protein units were shown by the red box and described from N-terminal to C-terminal; slide helix, outer transmembrane (TMD1), turret, pore helix, filter and inner transmembrane (TMD2). B) Maximum likelihood phylogenetic tree of amino acid sequences of the K<sup>+</sup> channel gene from *Prasinovirus* and

*Chlorovirus*. Viruses newly isolated in this study have been annotated in red. The tree was rooted with the bacteria branch and bootstrap support > 80% were shown. The scale shows substitutions per site.

#### 4. General discussion

The genera *Bathycoccus*, *Micromonas* and *Ostreococcus* belonging to class Mamiellophyceae are widespread and dominate the eukaryotic picophytoplankton fraction in coastal water<sup>30</sup>. Their abundance supports a dense population of prasinoviruses present everywhere in marine waters where their hosts have been isolated<sup>216</sup>, where they are thought to play a significant role in regulating phytoplankton numbers through cell lysis. Despite their abundance and ecological impact, little is known about the diversity of these viruses and, in particular, in algal strains that are challenging to grow in culture. In this work, we isolated four viruses from *O. tauri*, for the first time a virus of *Micromonas commoda* RCC827 from the Pacific Ocean near the Chilean coast, and also a single virus isolated infecting an *O. tauri* strain from the North Sea. We sequenced these viral genomes along with 12 representative viruses from a set of 40 OtVs that were mainly species-specific (85%) and highlight viruses able to infect a wide range of hosts (described as generalist viruses) unlike others that can infect few strains (specialists)<sup>231</sup>.

The issue of the genomic basis for determining the infection pattern of viruses and their relationship with the geographical origin seems to be much more complex than we had thought. In this work several approaches have been used to address this question. Based on the specificity pattern we have identified three main groups of viruses. On the one hand, OtVs with a broad host range infecting all *O. tauri* strains tested, on the other hand viruses able to infect only one *O. tauri* strain (RCC4221), and finally two viruses which are able to infect strains from another *Ostreococcus* species (Clade D – *O. mediterraneus*) as well as *O. tauri*. While the most specialist viruses comprised some viruses of Pacific Ocean origin, as well as OtV5 isolated from a Mediterranean marine station, we inferred that the infection spectrum was not closely related to geographical origin. Furthermore, phylogenetic analysis illustrated that the OtV-type 1 forms a monophyletic group in a basal position to other OtVs-types, closely related to the *Micromonas* virus clade. Their taxonomic affiliation revealed that most of the genes encoded by the OtV-type 1 were more similar to OtV6 and *Micromonas* virus homologs than other *Ostreococcus* viruses. Our data further support the possibility that this subclade might represent an intermediate lineage of *Ostreococcus* viruses, as described for OtV6<sup>232</sup>. It is even tempting to speculate that the OtVs-type 1 are also able to infect other hosts than *Ostreococcus*, which

could be more closely related to the genus *Micromonas* in natural environmental communities and would therefore be the first wider host-range virus described to date among the prasinoviruses. For this purpose, the isolation of more strains from the same regions where the viruses were isolated would extend our view of the correlation between geographical origin and the pattern of infection as well as the original host strain of these Pacific viruses.

Through phenotyping the viruses on a broad selection of *O. tauri* strains, we were able to show that OtV09-570 could be classes as a generalist virus, infecting 8 of 9 strains. However, comparative genomic and phylogenetic analyses showed OtV09-570 was affiliated with the largest group of OtVs (OtV-type 2a), and not with the most generalist viruses (OtVs-type 2b) infecting 9 of 9 strains. These findings suggest that multiple gene determinants of host-range are possibly acquired over the course of evolutionary history such that that there might be distinct evolutionary pathways to becoming a generalist that do not depend on the same functional biological process.

*The arms-race between host and virus shown by antiviral immune system plasticity and viral modular shuffling in the hypervariable region*

Coevolution between hosts and viruses can be subject to an arms-race dynamic. This scenario is generally seen as a continuum of reciprocal coadaptations, whereby host and virus evolve with adaptation and counteradaptation through escape-recognition cycles<sup>496,497</sup>, thereby driving molecular evolution<sup>498</sup>. Host escape from viral infection results from genetic modification (spontaneous mutations) of the structure of surface receptor regions where recognition/binding of viral proteins occurs<sup>294</sup>. These defense mechanisms are commonly observed in taxonomically divergent host organisms (fungi, bacteria, mammalian cells) in response to parasite infection<sup>297,499–502</sup>. In *Prochlorococcus*, resistance to cyanophages has been associated to mutations in a single hypervariable genomic island, affecting the viral attachment to the cell surface<sup>297,298</sup>.

In *Ostreococcus*, the receptors used by the virus in the process of infection are not yet known. However, transcriptomic analyses in *O. tauri* revealed differentially expressed genes mainly related to carbohydrate metabolism and modification were clustered on chromosome 19 (known as Small Outlier Chromosome – SOC) in resistant clonal lines<sup>310</sup>. Some of these genes include CMP-N-acetylneuraminic acid hydrolase (ostta19g00610), and triose-phosphate transporter (ostta19g00600), which play a role in the biosynthesis of cell surface sialic acid N-

glycolylneuraminic acid<sup>503</sup>. Likewise, the glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase (ostta19g00120) is involved in the protein glycosylation pathway and in protein modification, mainly in the synthesis of extended mucin type O-linked glycans. Hence, *Ostreococcus* might evade viral infection by masking or changing the structure of the receptor of carbohydrate origin, altering the virus adsorption<sup>308,310</sup>. Under this scenario, viruses must evolve in response to the host defense, which favors selection of genetic variants adapted to the new receptor structure, enabling the interaction to be restored.

In contrast to virus resistance in algae, the genetic basis of host-range evolution in algal viruses is less known. In tailed phages, host recognition is mainly achieved *via* adhesin proteins of their long tail fibers. Hence, adhesin are subjected to frequent modular shuffling, modifying host receptor specificity<sup>504-506</sup>. In the T4 phage superfamily, genomic analysis has demonstrated that the tail fiber module – responsible for recognition and attachment to the host cells – was the most plastic structural region in the viral genomes<sup>487</sup>. Paterson *et al.*, have demonstrated that conserved tail fiber genes often have evolved in response to acquired resistance by the host cell receptors, and evolved more rapidly when both partners co-evolved<sup>392</sup>. Here, we revealed for the first time the existence of a modular architecture – a conserved core genome and a hypervariable region – in OtV genomes. The evidence of a diversification of gene sequences, such as the trimeric autotransporter adhesin and the acquisition of additional specific genes such as concanavalin A-like lectin/glucanase domain containing protein in OtV-type 2b, suggests these regions are involved in the interaction with the host and might be the result of arms-race co-evolution.

While we could not establish the ratio of nonsynonymous substitution ( $d_N$ )/ synonymous substitution ( $d_S$ ) for these genes, due to too high divergence in protein and nucleotide sequences, we nevertheless suggest that these genes evolved very fast as they contained completely new structures in each strain, indicating they are subject to strong selection pressure in nature. This hypervariable locus could be a recombination hotspot mediating diversification in the recognition of host cells and thus host-range evolution in OtVs.

## 5. Conclusion

While most of the viruses studied here have been isolated from a single *O. tauri* RCC4221 strain, the OtV genomes showed diversification split into two distinct modules: core and hypervariable regions. The highly conserved genes required for the successful completion of the virus life cycle have been located in the core region of the genome. By contrast, the

hypervariable region appeared to concentrate all the genes involved in the host-virus interaction that most likely function in recognition and adhesion to the host cell during viral infection. The distinction in host range between the three OtV types supports the idea that the loss or non-acquisition of genes encoding proteins with adhesin or lectin domains could explain the narrower host range of OtV-type 2a and OtV-type 1. If this is the case, these proteins should be potential targets for host defense. Hence, the evolution of genes in the hypervariable region has probably been shaped by opposing forces in an arms-race dynamic: The need for the virus to maintain the function of these recognition genes for entry into the host cells, the need to escape recognition by host defenses and the acquisition or loss of genes for host range evolution, and the need for the host to genetically modify the surface receptor structure as a means of antiviral defense.



# CHAPITRE IV – CARACTERISATION MORPHOLOGIQUE DES VIRUS D’OSTREOCOCCUS TAURI

## **Objectifs du chapitre :**

Les résultats obtenus dans le chapitre précédent nous ont poussé vers une caractérisation morphologique des virus d’*Ostreococcus tauri*. Nous avons essayé de corrélérer les données génomiques et les caractères morphologiques pour tenter d’approcher à l’échelle microscopique des indices sur la spécificité de la gamme d’hôte chez les *Prasinovirus*.

## **Contributions de l’auteur :**

Ce travail a été réalisé en étroite collaboration avec Marie-Line Escande de la plateforme de Microscopie Électronique (BioPIC) à Banyuls-sur-Mer. Marie-Line Escande a fourni l’expertise et conduit le développement des protocoles de microscopie. Elle a également réalisé toute la partie expérimentale de préparation des échantillons en coupe pour la MET. J’ai effectué le maintien et la préparation des cultures d’algues et de virus ainsi que la préparation des échantillons en coloration négative. J’ai également réalisé toutes les observations au MET ainsi que les clichés.

**Les documents supplémentaires** de ce chapitre sont référencés à la page 215

## Résumé

Ce chapitre met en lumière la caractérisation morphologique des cinq virus d'*Ostreococcus tauri* (OtVs) nouvellement isolés dans ce travail de thèse ainsi que les trois OtVs au large spectre d'hôte isolés précédemment. Par l'approche de microscopie électronique à transmission (MET), ce travail a permis de mettre en évidence des structures uniques encore non décrites chez les virus d'eucaryote. Premièrement, deux sous-populations à morphotype distinct provenant d'un échantillon environnemental marin de l'Océan Pacifique (OtV19-T) ont été mises en évidence. L'une d'elle se caractérise par une forme semblable à un « citron », dont l'asymétrie semble interrompue par un *vertex* unique au centre d'un enfoncement de la capsid. D'autre part, le virus OtV09-565 à large spectre présente une structure protubérante flexible, semblable à une queue située sur un des *vertex* de la capsid. Nous présentons d'abord les verrous empêchant son observation, puis finalement des hypothèses concernant la présence de queue chez OtV09-565 et d'autre part ce qui pourrait favoriser son observation. Cette structure apparaîtrait dans les premières étapes d'infection, et *via* son déploiement, pourrait favoriser la reconnaissance et l'adsorption du virus à la cellule hôte. Des travaux complémentaires devront être apportés pour tester nos hypothèses.

## 1. Introduction

A ce jour, les virus sont décrits comme les entités biologiques les plus abondantes et ubiquitaires sur Terre en raison de leur capacité à infecter toute vie cellulaire et de leur omniprésence dans l'ensemble des écosystèmes explorés<sup>95,96</sup>. Au-delà de leur prédominance quantitative, les virus ont une immense diversification tant dans leur répertoire génétique que dans la morphologie des particules avec des formes et tailles extrêmement variables<sup>507</sup>. La découverte des virus géants a aussi remis en question la définition même d'un virus<sup>508</sup>. Récemment, ces virus géants ont été classifiés dans l'embranchement *Nucleocytoviricota*, (aussi connu sous le nom plus informel, NCLDV). Les sept familles officielles qui le constituent sont des virus à ADN double-brin qui infectent les eucaryotes<sup>123</sup>. Les caractères génomiques et morphologiques que partagent les NCLDVs suggèrent que ces virus ont évolué à partir d'un ancêtre commun<sup>130,134</sup>, bien que cela reste contesté à ce jour<sup>138</sup>. Ce type de virus est décrit comme « géant » puisque la taille des virions peut atteindre 1.5 µm et la taille du génome correspondant 2.5 Mb, donc comparable à de nombreuses bactéries, archées et même aux picoeucaryotes<sup>508,509</sup>. Bien que les NCLDV partagent un certain nombre de caractères en commun, ils présentent toutefois une énorme diversité de formes (i.e. des formes non-icosaédriques : virions ovoïdes ou formes imbriquées des *Pandoravirus*, *Poxviridae*, ou même les *Ascoviridae* en forme de bacilliforme/allantoïde<sup>510</sup>), de leurs cycles de réplication, ou même de gammes d'hôtes puisqu'ils infectent toutes les espèces de l'arbre eucaryote<sup>140</sup>.

La microscopie électronique est un outil fréquemment utilisé pour fournir des informations structurelles de base, particulièrement la forme et les caractéristiques de surface. De nombreuses analyses ont permis de mettre en évidence des caractéristiques morphologiques communes à de nombreux NCLDVs. Ils se composent de trois éléments principaux tels qu'une capsidie protéique icosaédrique, une membrane lipidique interne et un noyau génomique renfermant le matériel génétique. Deux virus associés aux *Asfarviridae* font exception à cette règle. Les Faustovirus échangent la membrane interne lipidique par un manteau de bicouche protéique pour encapsider et protéger son génome<sup>511</sup>, tandis que l'*African swine fever virus* (ASFV) se caractérise par une structure multicouche complexe qui renferme un nucléoïde contenant l'ADNdb, entouré de deux capsides protéiques icosaédriques distinctes (internes et externes) et de deux membranes lipoprotéiques (internes et externes)<sup>512-514</sup>.

Les *Phycodnaviridae* sont les prédateurs d'une large gamme d'algues marines et d'eau douce et font parties des plus petits virus géants puisque la taille de leur particule est estimée

entre 120 et 220 nm de diamètre et la taille de leur génome entre 180 et 400 kb. Des travaux de reconstruction en trois dimensions par cryogénique microscopie électronique (*Cryogenic Electron Microscopy* – cryoEM) à transmission ont permis de montrer que les *Phycodnaviridae* possèdent certaines caractéristiques structurales communes<sup>198,515</sup>. Ces virus se composent d'une structure stratifiée avec une capsid glycoprotéique externe icosaédrique assemblée à partir de centaines de capsomères trimériques et pentamériques pseudo-hexagonaux, étroitement empaquetés, et d'une membrane interne lipidique, renfermant et protégeant le matériel génétique. Leur enveloppe interne joue un rôle majeur dans le processus d'infection puisqu'elle fusionne avec la membrane cellulaire hôte pour permettre l'expulsion de l'ADN viral dans le compartiment cytoplasmique de son hôte<sup>516,517</sup>.

Au sein des *Phycodnaviridae*, les chlorovirus constituent à ce jour un des groupes les mieux caractérisés aux niveaux morphologique, génomique et répliatif<sup>253</sup>. Les chlorovirus infectent l'algue verte eucaryote unicellulaire *Chlorella* par une première étape de fixation rapide à la paroi cellulaire de l'algue, qui se produit *via* un *vertex* unique de la capsid virale. Ce point d'attachement est suivi par une seconde étape de perforation de la paroi cellulaire. La dernière étape consiste en une fusion de la membrane virale avec la membrane de l'hôte pour permettre le transfert du matériel génétique et de certaines protéines associées. Ce phénomène est caractérisé par la présence de capsid vide à la surface de la cellule hôte<sup>208,518</sup>. Des analyses de cryoEM et reconstruction tridimensionnelle ont permis de révéler une structure unique en forme de pointe (*spike*) située sur ce même *vertex* de la capsid icosaédrique à symétrie *five-fold*<sup>197,198</sup>. En raison de sa forme, de sa taille et de son emplacement, les auteurs proposent que cette structure constituerait un dispositif de reconnaissance / contact puis de perforation de la paroi cellulaire hôte, comme c'est le cas pour les bactériophages à queue<sup>519</sup>. En revanche, l'hypothèse du transfert de l'ADN à travers ce dispositif est rejetée, puisque la *spike* apparaît trop étroite chez PBCV-1.

Au sein des *Phycodnaviridae*, les *prasinovirus* se caractérisent par une capsid de symétrie icosaédrique d'un diamètre d'environ 100 à 140 nm qui recouvre une membrane lipidique interne protégeant le génome du virus. Toutefois aucune structure complexe n'a été observée sur un des sommets de la capsid<sup>222,227,230</sup>. Quelques exemples de structures comme des projections à la surface de la cellule hôte ont été illustrées chez le virus infectant *Micromonas*<sup>222,520</sup>, ainsi que chez le virus OtV-1<sup>230</sup>. Des analyses de microscopie électronique à transmission (MET) ont permis de mettre en évidence l'attachement du virus OtV5 sur un *vertex* unique à la membrane externe de l'algue *O. tauri* *via* la formation d'un « pont » qui semble fusionner la capsid protéique du virus avec la membrane plasmique de la cellule

hôte<sup>229</sup>. Il est donc possible que d'autres virus eucaryotes icosaédriques utilisent un des *vertex* pour accomplir à bien leur cycle de vie. La queue des bactériophages est l'un des meilleurs exemples de structures non-icosaédriques avec un *vertex* unique, brisant ainsi la symétrie globale de la capsid du virus<sup>521</sup>. Chez les bactériophages, les structures de la queue sont des déterminants majeurs de la spécificité de l'hôte et du processus d'infection propres aux virus concernés<sup>522,523</sup>. Ces structures complexes et très diverses fonctionnent comme des détecteurs qui reconnaissent spécifiquement des récepteurs à la surface de la cellule hôte.

La caractérisation morphologique de la particule virale est donc essentielle pour mieux comprendre les interactions *Ostreococcus*-OtVs mais aussi ce qui détermine la spécificité de ces virus. Nous avons donc choisi de visualiser par MET 1) cinq OtVs nouvellement isolés d'origine géographique distincte et 2) trois OtVs qui présentent une gamme de spécificité plus large d'hôtes. Ces observations ont été comparées à celle de notre OtV modèle, OtV5. Nous nous focalisons ici sur deux virus d'*O. tauri* avec des caractéristiques structurales distinctes de celles déjà décrites pour des virus d'algues eucaryotes. Celui nommé OtV19-T, provenant d'un échantillon environnemental originaire de la côte Est de l'Océan Pacifique, présente deux morphotypes distincts dont l'un deux se caractérise par une symétrie non-icosaédrique avec un *vertex* plus arrondi. L'autre nommé OtV09-565 se caractérise par la présence sporadique d'une structure protubérante flexible, semblable à une queue située sur un des *vertex* de la capsid à symétrie icosaédrique. Cette structure s'est avérée difficilement observable dans les préparations de virions isolés ce qui nous a conduit à faire varier les conditions de préparations avant visualisation et parallèlement d'observer aussi par MET le suivi de l'infection de la cellule hôte. L'ensemble de ces observations sera mis en perspective avec les données génomiques rapportées dans le chapitre III et discutées d'après la littérature.

## 2. Matériel et méthodes

### *Conditions des cultures d'algue et des virus*

Les cultures d'algues *O. tauri* RCC4221 et de virus (OtV5, OtV09-578, OtV09-565, OtV06-12, OtV19-P, OtV19-T, OtV19-R, OtV19-O, Sylt2-5) ont été cultivées dans les mêmes conditions décrites dans le chapitre III. Une densité cellulaire en phase exponentielle de croissance, à environ  $15 \times 10^6$  cellules/mL, a été infectée à OtV09-578 à une MOI 10 pour suivre la dynamique d'infection au cours du temps. La densité des cellules a été mesurée à

l'aide d'un cytomètre en flux BD FACS Canto II (BD Biosciences, Franklin Lakes, NJ, USA). L'auto-fluorescence cellulaire de la chlorophylle a été utilisée pour estimer le nombre de cellules d'algues viables.

### *Microscopie électronique*

L'observation des virus en coloration négative a été réalisée sur des grilles recouvertes d'un film de Formvar et d'un film de carbone (FCF200-Cu – EMS), rendues hydrophiles par effluage (20 secondes de décharge lumineuse à l'air,-ve, GloQube® Plus Glow Discharge System). La grille a été retournée sur une goutte de lysat viral frais filtré à 0.45 µm pendant 30 minutes pour permettre aux virus d'adhérer sur le film support de la grille. Après élimination de l'excédent d'échantillon par adsorption sur un papier filtre, la grille a été déposée sur une goutte de solution d'acétate d'Uranyle aqueux à 2% pendant 45 secondes. Après avoir absorbé l'excédent de colorant avec du papier filtre, la grille a été séchée à l'air libre. L'observation a été réalisée au MET Hitachi 7500.

Pour l'observation en coupe en MET, les cellules *O. tauri* infectées et non-infectées à OtV09-565 ont été préparées selon Chretiennot-Dinet *et al.* (1995)<sup>61</sup> et Derelle *et al.* (2008)<sup>229</sup>. Une concentration de  $15 \times 10^6$  cellules.mL<sup>-1</sup> contenues dans des aliquotes de 50 mL d'échantillons préfixés au glutaraldéhyde à 1% (Electron Microscopy Sciences, #16320), ont été recueillies par centrifugation à 2500 g pendant 30 min. Les cellules agglomérées ont ensuite été mélangées rapidement avec 40 µL d'agarose (Low melting point agar - Sigma, #A9414) à 37°C à l'aide d'un microrécepteur (SMI,Emerville,CA,USA). Une fois l'agarose solidifié, le morceau d'agar obtenu contenant les cellules, a été fixé pendant 2 h à 4°C dans du glutaraldéhyde à 2.5% avec un volume de tampon cacodylate 0.4M et deux volumes de milieu de culture (milieu L1), et lavé 3 fois 15 min dans un volume de tampon cacodylate 0.4M et un volume de milieu de culture. La post-fixation a été réalisée dans du tétroxyde d'osmium (OsO4) à 1% (Electron Microscopy Sciences, #19150) dans un tampon cacodylate 0.2M pendant 1 heure. Après deux lavages dans ce même tampon, les petits morceaux d'agar, ont été déshydratés progressivement dans l'éthanol (70%, 95%, 100%) avant imprégnation et inclusion en Epon 812 (Electron Microscopy Sciences, #14120). Des coupes ultrafines (80 nm) ont été réalisées à l'ultra-microtome (Leica Ultracut R) et observées au microscope électronique à transmission (Hitachi 7500) après coloration à l'acétate d'uranyle et au citrate de plomb. Les images obtenues ont été traitées avec le logiciel ImageJ.

### 3. Résultats et discussions

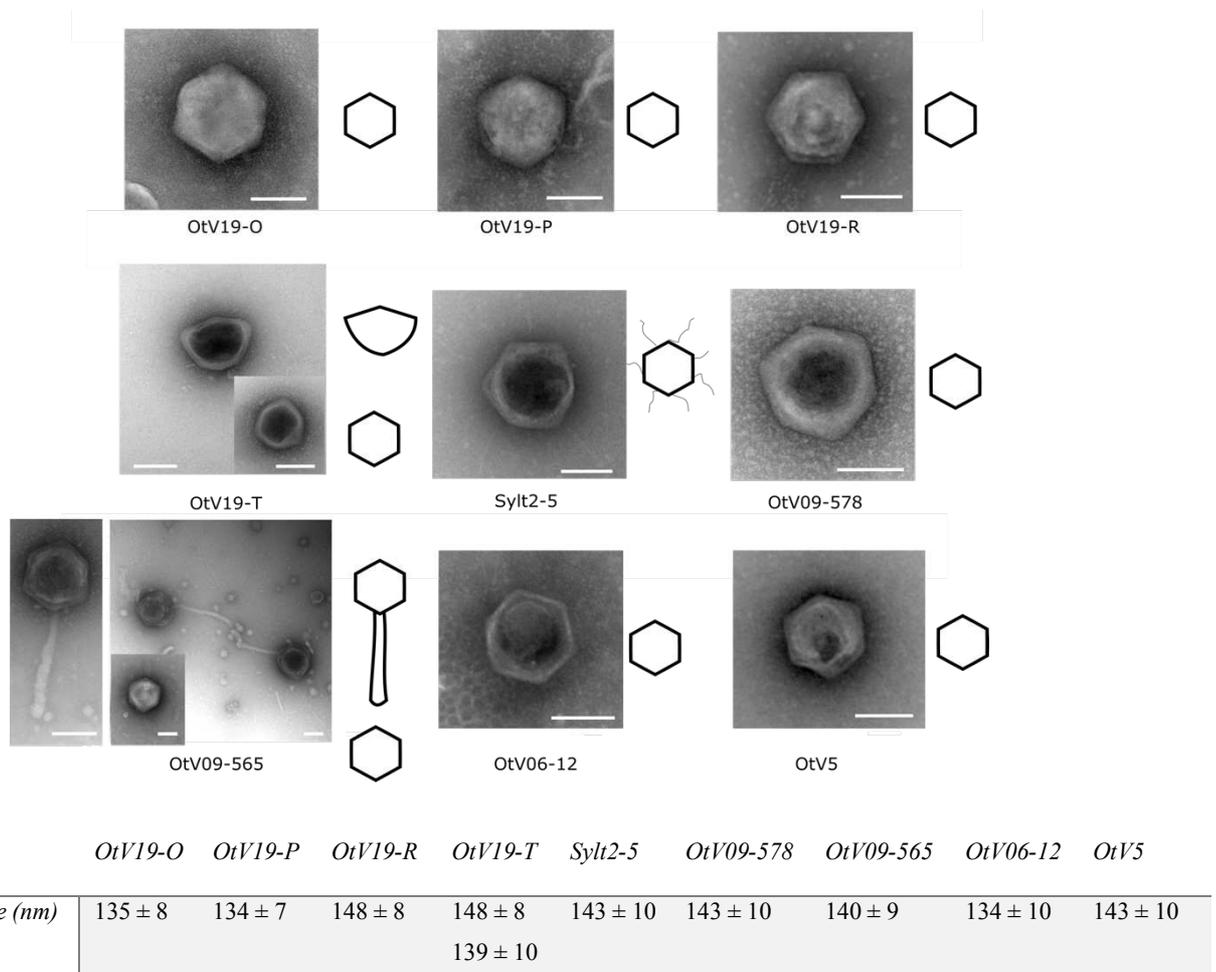
#### *Observation préliminaire de la morphologie de neuf OtVs*

Dans un premier temps, nous avons cherché à mettre en évidence si le caractère morphologique pouvait expliquer le patron de spécificité et l'origine géographique des OtVs. Pour tester cette hypothèse, les virus de l'Océan Pacifique (i.e. OtV19-O, OtV19-P, OtV19-T, OtV19-R) et de la Mer du Nord (i.e. Sylt2-5) ont été comparés aux virus de la Mer Méditerranée qui présentaient un plus large spectre d'hôte qu'OtV5 (i.e. OtV09-565, OtV09-578 et OtV06-12). Pour tous les virus étudiés, les particules virales libres ont été observées par MET en coloration négative.

Globalement, la structure des nucléocapsides apparaît hexagonale et se caractérise par une symétrie vraisemblablement icosaédrique (**Figure 30**). Les observations microscopiques mettent en évidence une large capsid non-enveloppée avec une taille située entre 134 et 148 nm. Nous devons noter que l'usage de l'acétate d'uranyle (AU) conduit aléatoirement à une coloration positive et négative des particules virales dans certains cas. Ainsi, des particules apparaissent grisâtres et plus en relief, tandis que d'autres laissent visible la membrane interne protégeant l'ADNdb dense aux électrons.

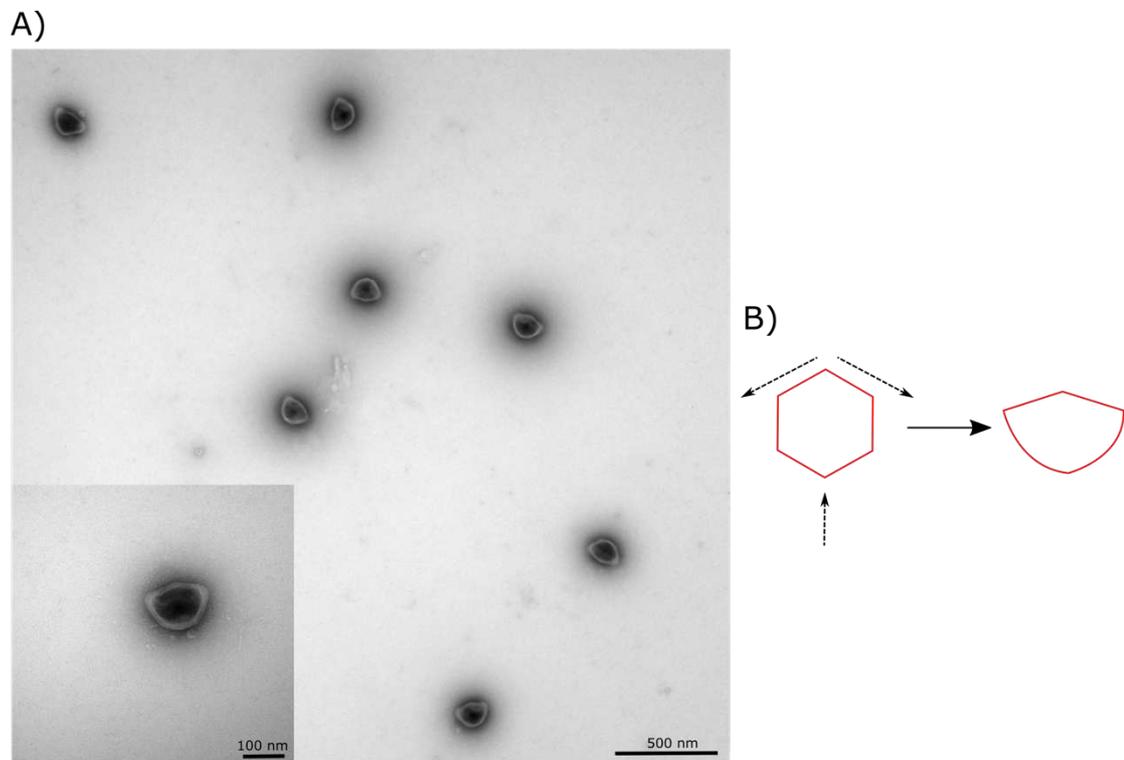
Bien qu'aucune différence morphologique structurelle ne soit observée entre les virus de la Mer Méditerranée et OtV19-O, OtV19-P, OtV19-R et Sylt2-5, ce dernier laisse intimement apparaître de longues fines fibrilles, possiblement ancrées à la capsid (**Figure 30**). Par ailleurs, nous avons pu mettre en évidence deux morphologies distinctes retrouvées dans le même lysat viral d'OtV19-T (**Figure 30**). Les deux types de virus présentent des capsides non-enveloppées avec des structures distinctes. Le premier type possède une capsid de symétrie icosaédrique similaire aux autres OtVs, avec un diamètre d'environ  $139 \pm 10$  nm (axe le plus court), tandis que le second type se caractérise par une forme plus atypique. Il présente une capsid icosaédrique interrompue au niveau d'un seul *vertex* plus arrondie avec un diamètre plus allongé d'environ  $148 \pm 8$  nm (axe le plus court). Un étirement latéral des deux côtés de la capsid et un « enfoncement » d'un *vertex* unique peut être une première raison de la géométrie irrégulière observée chez OtV19-T type-2 (**Figure 31**).

De façon très surprenante, les analyses de MET révèlent une structure remarquable du virion OtV09-565. Ce virus présente une capsidie similaire à celle des *Prasinovirus* en taille ( $140 \pm 9$  nm) (axe le plus court) et en structure avec une symétrie icosaédrique (**Figure 30**). Cependant, OtV09-565 expose une longue queue non-structurée qui apparaît flexible, émergeant d'un *vertex* unique de la capsidie. Ainsi, ce virus se définit par une symétrie complexe et bilatérale, comme c'est le cas pour les bactériophages à queue.



**Figure 30.** Morphologie de huit OtVs. La morphologie des virions des OtVs a été analysée au moyen d'études de la microscopie électronique par coloration négative. Les virus isolés de l'Océan Pacifique (OtV19-O, OtV19-P, OtV19-T, OtV19-R) ainsi que le virus de la Mer du Nord (Sylt2-5) ont été comparés aux virus de Méditerranée qui présentent un spectre d'hôte plus large (OtV09-565, OtV09-578 et OtV06-12) qu'OtV5. Un dessin représentatif de la morphologie virale est illustré à droite de chaque virus. Les particules virales libres observées sont issues d'un lysat fraîchement obtenu après 3 jours post-infection (JPI) (à l'exception du virus OtV09-565 où des queues localisées à un vertex unique ont été observées à 4 JPI), et filtrées à 0.45  $\mu$ m. La barre d'échelle correspond à 100 nm. Un nombre de 15 particules virales a été comptabilisé pour estimer le diamètre (Moyenne  $\pm$  écart-type standard) des

capsides (axe le plus court), tandis que les dimensions de la queue d'OtV09-565 ont été estimées sur une base de 60 particules virales. A noter que l'utilisation de l'acétate d'uranyle (AU) provoque des colorations négatives (capsides grisâtres) et positives (capside avec un matériel génétique denses aux électrons) de façon aléatoire et indépendante. Ce phénomène est associé à l'absorption de l'AU uniquement par les larges particules virales à ADNdb<sup>524</sup>.



**Figure 31.** Morphologie d'OtV19-T type-2. A) Observation par MET des virions en suspensions d'un filtrat à 0.45  $\mu\text{m}$ , issus d'une culture *O. tauri* RCC4221 infectée à 3 JPI. B) Représentation schématique des possibles forces exercées (flèches en pointillé) sur la capsid virale à symétrie icosaédrique hexagonale régulière conduisant vers (flèche trait plein) une symétrie irrégulière arrondie de la particule OtV19-T type-2.

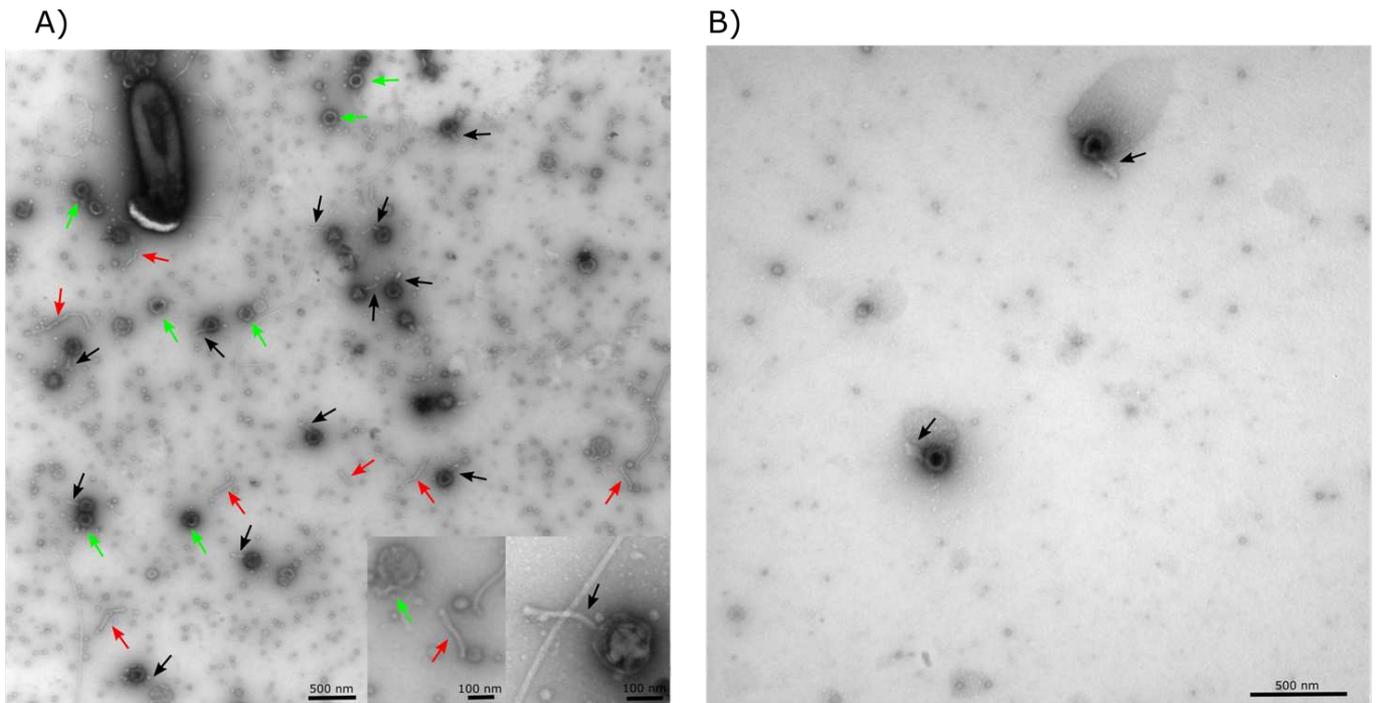
La longueur et l'épaisseur de la queue ont été mesurées : respectivement  $154 \pm 46$  nm et  $21 \pm 5$  nm. Bien que moins dense aux électrons, cette queue d'OtV09-565 rappelle celle des bactériophages à queue (*Siphoviridae*)<sup>523</sup>. Toutefois, les structures plus complexes de reconnaissance et d'adhésion (i.e. fibre de queue – *tail fiber* ou *spike* central) situées à l'extrémité de la queue du phage, n'ont pas été observées dans notre cas. D'autre part, la queue du virion ne présente aucun disque empilé visible comme c'est également le cas pour les *Siphoviridae*<sup>525</sup>. En revanche, son extrémité montre un élargissement de l'épaisseur d'une taille d'environ 30 nm, qui est observable pour certaines particules. Cette structure pourrait

correspondre à une plaque basale de type *Myoviridae* ou *Siphoviridae*, qui est connue pour jouer un rôle dans la reconnaissance et l'ancrage à différents types de récepteurs glycosidiques situés à la surface de la cellule hôte<sup>526,527</sup>.

### *Optimisation de l'observation morphologique d'OtV09-565*

Précédemment, nous avons mis en évidence que la queue d'OtV09-565 semble posséder des caractéristiques biologiques et physiques particulières. En effet, cet appendice se montre flexible et peu structuré. D'autre part, de nombreuses particules abîmées ou sans queue ont été observées, à proximité vraisemblablement de queues libres dans le filtrat (**Figure 32A**). Toutes ces indications laissent à penser que la queue d'OtV09-565 est une structure extrêmement labile. De ce fait, nous avons envisagé que les conditions de préparations utilisées précédemment pourraient ne pas être assez favorables à leur visualisation. Ainsi, nous avons cherché à déterminer si cette structure est plus ou moins visible selon des conditions de préparations expérimentales ou si sa présence et son état varient dans certaines conditions biologiques.

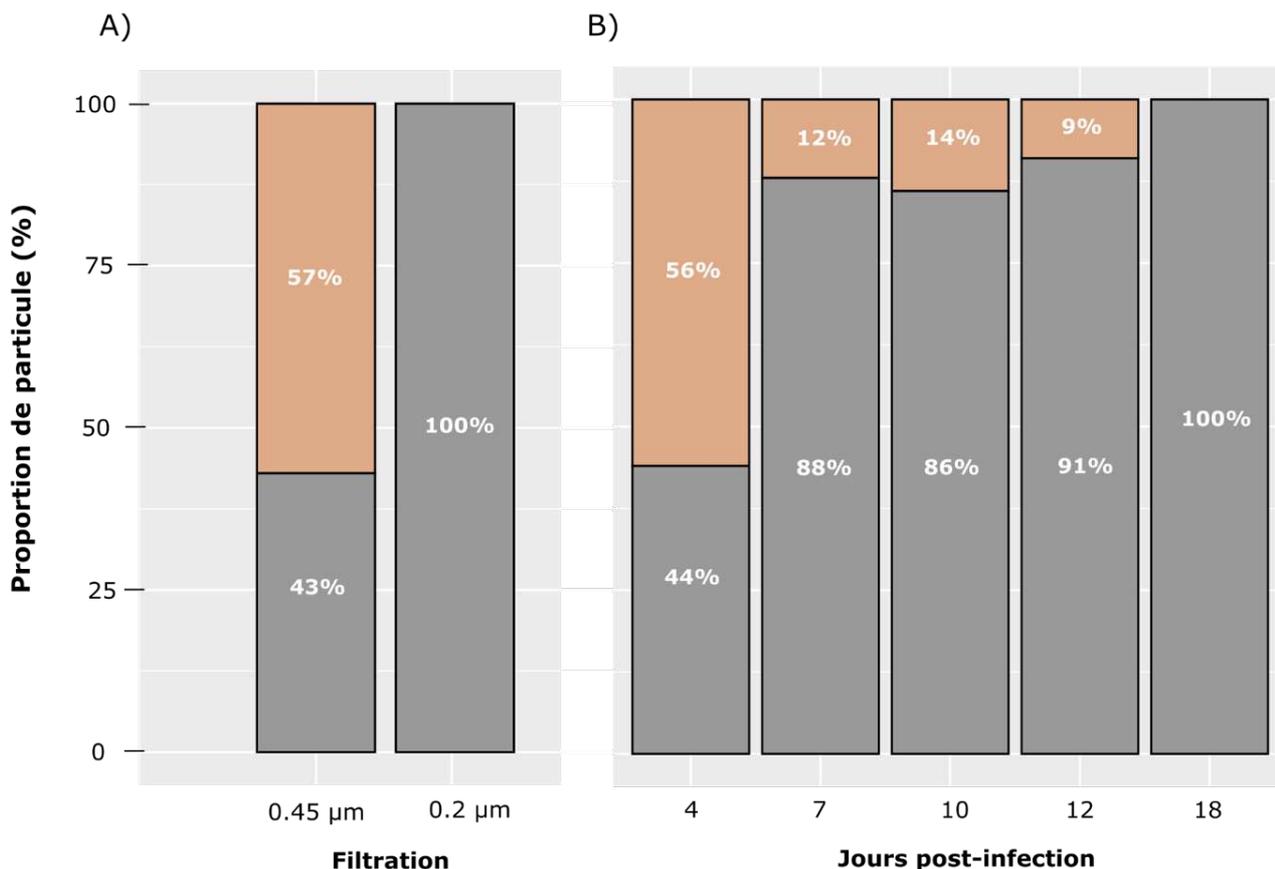
Dans un premier temps, nous avons adopté une stratégie visant à améliorer les conditions de préparation de la culture. Une homogénéisation plus douce de la culture infectée a été réalisée puis visualisée à 4 JPI par MET (**Figure 32B**). L'observation du filtrat à 0.45  $\mu\text{m}$  montre l'absence de queues libres en suspension dans le milieu et met en évidence une caractéristique morphologique des particules d'OtV09-565, semblable à celle décrite précédemment, avec une proportion d'environ 57% de capsides associées à une queue. Nous suggérons donc qu'une trop vive homogénéisation de la culture infectée fournirait une première casse mécanique, aboutissant à la présence de queues libres en suspension. Par ailleurs, la filtration à 0.2  $\mu\text{m}$  de cette même culture a mis en évidence une proportion de 100% de capsides sans queue (**Figure S1**), ce qui suggère que la filtration à 0.2  $\mu\text{m}$  ait conduit à une rétention des particules à queue sur la membrane. Étant donné le caractère très labile de cet appendice, une seconde hypothèse peut aussi être envisagée : la porosité à 0.2  $\mu\text{m}$  de filtration favoriserait la dissociation entre ces structures et les capsides, laissant traverser uniquement ces dernières.



**Figure 32.** Morphologie d’OtV09-565 produit à partir A) d’une culture concentrée par la méthode de flux tangentiel (TFF), fortement homogénéisée et filtrée à 0.45  $\mu\text{m}$  et d’une autre B) culture homogénéisée à plus faible intensité et filtrée à 0.45  $\mu\text{m}$ . Les flèches noires montrent les particules virales associées à une queue sur un *vertex* unique à la capsid. Les flèches rouges indiquent une potentielle queue libre dans le filtrat et les flèches vertes des particules virales dépourvues de queues.

Il est probable que la variation de la présence des queues soit liée avec une dégradation, et/ou une rupture des queues au cours du temps. Pour tester cette hypothèse, nous avons suivi l’infection d’une culture *O. tauri* RCC4221 par le virus OtV09-565. Des filtrats à 0.45  $\mu\text{m}$  de la culture ont été observés par microscopie électronique en coloration négative à différents jours post-infection. La proportion de particules virales avec et sans queue a été reportée dans la **Figure 33**.

A 4 JPI la proportion de particules à queue apparaît comparable à celle obtenue dans l’expérience précédente de culture filtrée à 0.45  $\mu\text{m}$  (**Figure 33A**). Ainsi, les particules à queue apparaissent légèrement majoritaires (56%) par rapport à celles sans queue (44%). Cette proportion diminue drastiquement à 7 JPI, puisque seulement 6 particules sur 50 (12%) présentent une queue localisée sur un *vertex* unique de la capsid. A 10 et 12 JPI cette proportion est maintenue au sein de la culture, tandis qu’elle est ramenée à zéro environ trois semaines après, avec une population de particule sans queue de 100%.



**Figure 33.** Proportion du nombre de particules virales à queue (orange) ou sans queue (gris). Les particules virales libres issues A) d'un lysat fraîchement obtenu à 4 jours post-infection filtrée à 0.45 µm ou 0.2 µm, ou B) d'une infection lytique au cours du temps, ont été visualisées et comptabilisées en coloration négative par microscopie électronique. Pour l'ensemble des observations un total de 50 particules virales indépendantes a été comptabilisé. L'observation en microscopie du lysat OtV09-565 filtré à 0.45 µm et 0.2 µm est présentée dans la **Figure S1**.

Dans cette expérience, les deux cultures infectées indépendamment (**Figure 33**) ont permis de montrer une tendance de particules à queue majoritaires à 4 JPI. Nous avons cherché à reproduire l'expérience une troisième fois et à comptabiliser les particules à 1, 2, 3 et 4 JPI, afin de démontrer si ces structures sont prédominantes dans les premiers jours d'infection. Malheureusement, bien que les conditions de préparations aient été inchangées, aucune queue associée à la capsid n'a pu être observée à nouveau. Par ailleurs, aucune queue libre n'a été retrouvée dans le filtrat extracellulaire. L'addition de glutaraldéhyde dans la culture juste avant la filtration a également été testée afin de favoriser un pontage protéique entre la capsid et cet appendice : aucune structure de ce type n'a pu être visualisée. La morphologie de ce virus

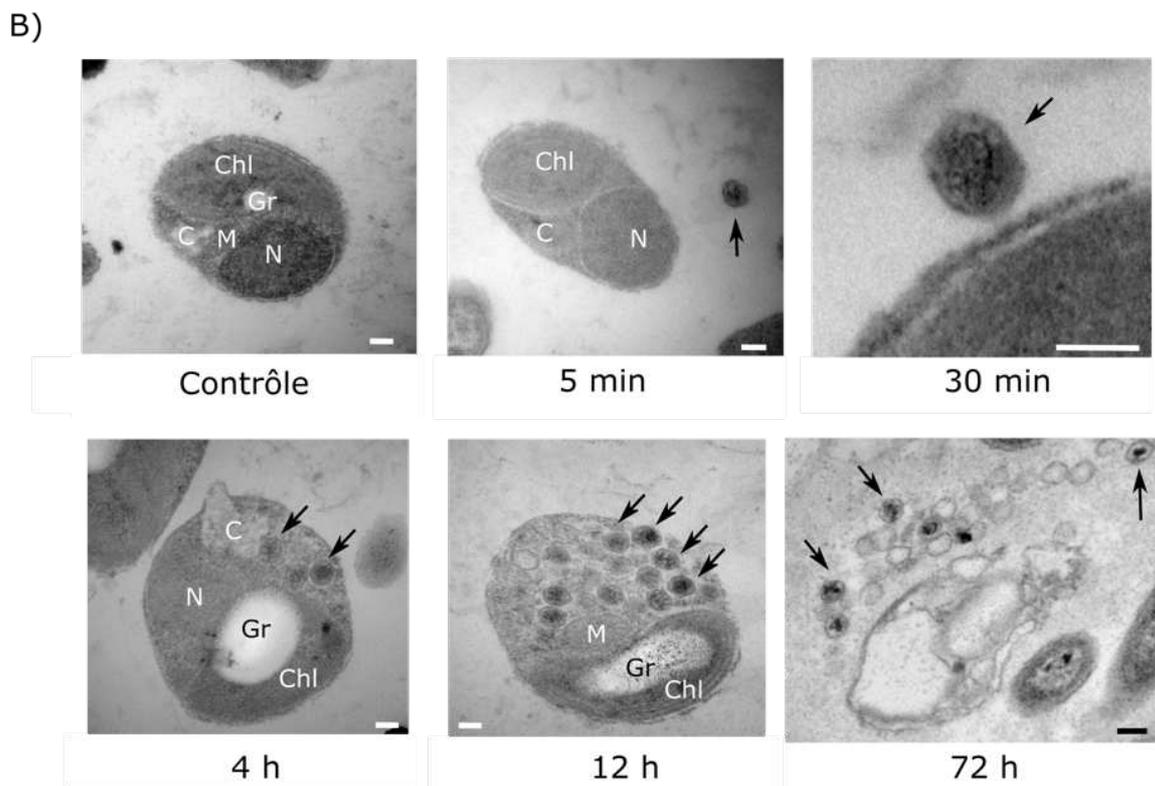
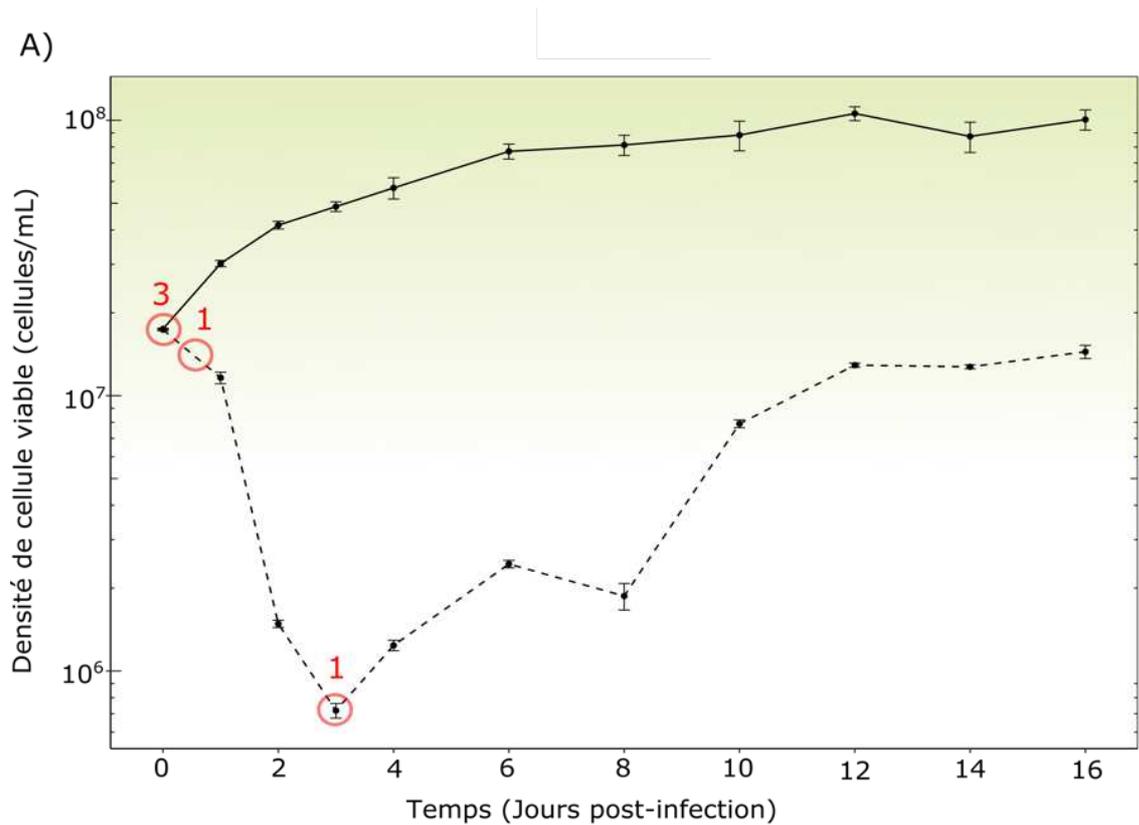
pourrait donc être associée à un caractère phénotypique très plastique que nos observations préliminaires révèlent en mettant en lumière les difficultés d'observer de telles particules à queue. Ainsi, afin de pouvoir mieux comprendre le rôle des queues dans le processus d'infection du virus, une meilleure maîtrise expérimentale de la formation de celle-ci est donc indispensable.

#### *Dynamique d'infection d'O. tauri* RCC4221 et son virus OtV09-565

Dans la continuité de ces travaux, nous avons réalisé un suivi cinétique de la dynamique d'infection d'une co-culture d'*O. tauri* RCC4221 et son virus OtV09-565. En parallèle, l'observation des cellules infectées a été réalisée à des temps précoces après infection (5 min, 30 min 4 heures, 12 heures et 72 heures) par MET.

La dynamique d'infection d'*O. tauri* présente une fluctuation rapide de la croissance des cellules infectées (**Figure 34A**). Après inoculation avec OtV09-565, le nombre de cellules viables diminue drastiquement de  $1.74 \times 10^7$  cellules.mL<sup>-1</sup> au premier jour d'infection à  $7.18 \times 10^5$  cellules.mL<sup>-1</sup> à 3 JPI. Au même temps, la densité des cellules de la culture non-infectée (contrôle) augmente de  $1.75 \times 10^7$  cellules.mL<sup>-1</sup> à  $3.02 \times 10^7$  cellules.mL<sup>-1</sup> à 1 JPI et de  $4.16 \times 10^7$  cellules.mL<sup>-1</sup> à 2 JPI, cohérent avec environ une division cellulaire par jour sous un cycle de lumière/obscurité de 12h :12h<sup>528</sup>. En revanche, une augmentation des cellules viables de la culture infectée ( $1.24 \times 10^6$  cellules.mL<sup>-1</sup>) est observée à partir du 4<sup>ème</sup> JPI, un jour de plus que la co-culture d'*O. tauri* et OtV5<sup>308</sup>. Ces cellules sont devenues résistantes au virus. Une telle dynamique des cellules infectées (i.e. lyse puis résistance) est observée une seconde fois mais plus atténuée entre 6 et 9 JPI. A partir, de 12 JPI les cellules viables qui ont évolué vers la résistance atteignent un plateau avec une concentration maximale d'une moyenne de  $1.29 \times 10^7$  cellules.mL<sup>-1</sup>.

A noter, que les cellules infectées n'atteignent pas les mêmes valeurs de fluorescence que les cellules non infectées contrôles à la fin de l'expérience à 16 JPI, contrairement à la dynamique d'infection d'*O. tauri* et OtV5 qui a été observée par Yau *et al.*<sup>308</sup>. D'ailleurs, la recoloration verte de la culture infectée après 16 JPI disparaît et aucune recoloration de la co-culture à OtV09-565 n'a jamais été observée, comme c'est le cas avec OtV5<sup>229,308</sup>. Parallèlement, nous avons observé des cellules infectées à des stades précoces de l'infection. Pour cela, cinq temps (5 minutes, 30 minutes, 4 heures, 12 heures et 72 heures) ont été choisis (**Figure 34B**).



**Figure 34.** Infection des cellules *O. tauri* RCC4221 avec le virus OtV09-565. A) Dynamique d'infection (le nombre de jours post-infection – JPI). Chaque point représente une moyenne de trois cultures indépendantes d'*O. tauri*. La culture non-infectée contrôle est indiquée en trait plein et la culture infectée avec le virus en trait discontinu. Les cultures échantillonnées à différents temps après infection pour les

observations par microscopie électronique, sont cerclées en rouge et le nombre d'échantillons prélevés est indiqué en rouge. La moyenne et l'erreur standard pour chaque temps sont montrés par les points noirs et les barres verticales, respectivement. Les données brutes sont disponibles dans les données supplémentaires de ce chapitre. B) Observation par microscopie électronique à transmission des cellules infectées à 5 min, 30 min, 4 heures, 12 heures et 72 heures après infection et non infectées (contrôle à 30 min). Les barres représentent 100 nm. Les flèches noires indiquent les particules virales pleines avec le matériel génétique dense aux électrons. Chl – Chloroplaste, N – Noyau, M – Mitochondrie, C – Cytoplasme, Gr – Grain d'amidon. Les observations à 30 min montrent une particule virale remplie, attachée à la membrane plasmique de la cellule hôte.

La culture contrôle montre des cellules de petite taille (~1 µm), surchargée d'organelles (i.e. mitochondrie, chloroplaste, appareil de Golgi, et noyau) laissant peu de place pour le compartiment cytoplasmique, qui représente seulement 30% du volume total de la cellule d'*O. tauri*. A une MOI de 10, nous observons à 5 min après l'infection, une particule virale proche de la cellule *O. tauri* prête à être infectée. Comparé à la taille de la cellule hôte, la particule virale est relativement large (~110 nm) avec une forme icosaédrique. Les capsides virales remplies présentent une région centrale dense aux électrons. En revanche aucune structure protubérante associée à un des *vertex* de la capsid n'est rendu visible à ce stade de l'infection.

A 30 min post-infection, les premières étapes d'adsorption sont observées. On observe souvent une particule virale accrochée à la membrane plasmique de la cellule hôte (**Figure 34B**). Cette dernière semble être reliée au virus par une forme de « pont » d'une longueur d'environ 20 nm et 30 nm d'épaisseur et qui apparaît moins dense aux électrons. Au vu de la morphologie de cette structure, nous ne pouvons pas affirmer que celle-ci soit une queue comme observée dans les colorations négatives décrites précédemment. Toutefois, l'adhésion entre la particule virale et la cellule hôte se produit vraisemblablement sur ce même *vertex* unique de la capsid. Une telle structure morphologique a aussi été décrite entre *O. tauri* et OtV5<sup>229</sup>. Le cycle de vie du virus apparaît relativement rapide puisque des particules virales complètes avec un matériel dense aux électrons sont apparues à 4 heures post-infection (hpi) dans le cytoplasme de la cellule. En revanche, les organelles apparaissent intactes et aucune particule n'a été observée dans le noyau ou le chloroplaste. Les étapes de réplication du virus et de son génome n'ont pas ou peu été étudiés chez les virus d'*O. tauri*. On suppose qu'OtV5 possède une phase nucléaire pendant son cycle d'infection car les *Prasinovirus* sont dépourvus d'ARN polymérase ADN-dépendante pour transcrire leurs propres gènes<sup>229</sup>.

A 12 hpi, les nouvelles particules virales néoformées s'accumulent dans le cytoplasme, et occupent un volume maximal de la cellule hôte. L'espace permettant la formation de nouvelles particules virales est ainsi fortement restreint. Il n'est donc pas étonnant que la moyenne de production virale (i.e. *burst size*) chez *O. tauri* soit seulement de 25 virions par cellule<sup>229</sup>. Les observations de MET à 12 hpi montrent un maximum de dix particules dans la cellule infectée dans cette section. La morphologie des particules virales semble très similaire avec une large capsidie icosaédrique, et ne présentent également aucune queue visible. Des particules incomplètes sont aussi mises en évidence ; elles sont constituées d'une capsidie icosaédrique avec une région centrale non-dense aux électrons.

Enfin, à 72 hpi, le nombre de cellule viable est le plus bas (**Figure 34A**). La plupart des cellules sont lysées, ce qui est « compatible » avec la perte visible de la coloration verte de la culture infectée. Ce phénomène est associé à l'éclatement des cellules hôtes pour permettre la libération des particules virales dans le milieu extracellulaire. Les particules virales libres ont une structure morphologique semblable en taille et en forme, bien que certaines apparaissent vides. En revanche, même à ce stade du cycle d'infection, aucune queue visible n'a été observée. Par cette méthode d'observation, il n'a donc pas été possible de retrouver une structure similaire à celle visualisée précédemment par coloration négative, bien que ce même filtrat ait été utilisé comme inoculum pour l'expérience de MET.

## 4. Discussion

### *Importance de structure asymétrique chez les virus icosaédriques*

Chez la plupart des virus, la symétrie icosaédrique de la capsidie est définie par une organisation structurale de sous-unités protéiques répétées appelées capsomères trimériques et pentamériques, qui s'assemblent pour former une grande enveloppe protéique stable<sup>529-531</sup>. Ce haut degré de symétrie des capsides virales est fondé sur le principe que Crick et Watson ont appelé « l'économie génétique ». Cette théorie exige que la structure icosaédrique de la capsidie soit construite uniquement à partir de quelques copies de sous-unité protéiques identiques afin de minimiser la partie du génome pour coder la capsidie finale. Sous pression de sélection naturelle, les virus auraient donc évolué pour faciliter autant que possible l'encapsidation du génome, une étape cruciale dans le cycle de vie du virus<sup>532,533</sup>. Toutefois, après cette étape, certains virus icosaédriques subissent des processus de maturation supplémentaire qui

produisent une asymétrie de la capsid. Ce phénomène est facilité par des composants additionnels qui brisent sa géométrie globale. Les bactériophages avec une queue hélicoïdale<sup>523</sup>, le *vertex* unique des herpesvirus<sup>534</sup> ainsi que celui de symétrie *five-fold*, en forme d'étoile de Mimivirus<sup>535-538</sup>, et la *spike* du chlorovirus PBCV-1 sont des exemples frappants qui illustrent l'asymétrie des capsides des virus icosaédrique à ADN. Un point de rupture de symétrie peut jouer un rôle fondamental dans le cycle de vie des virus. Chez PBCV-1, la *spike* est positionnée pour jouer un rôle précis dans l'adsorption/la liaison aux récepteurs de la surface à la cellule hôte. Chez le Mimivirus, l'asymétrie concerne davantage le transfert d'ADN viral entre capsid et cellule : l'éjection de l'ADN dans la cellule lors de l'entrée du virus au début de l'infection et son empaquetage dans la capsid naissante pour former une nouvelle particule virale prête à infecter, comme c'est le cas chez les bactériophages à queue<sup>539,540</sup> et les herpesvirus<sup>534</sup>. Par ailleurs, l'asymétrie a aussi été observée chez des virus à ADNsb ou même à ARN, impliquée dans d'autres processus clés du cycle de vie<sup>243</sup>.

#### *Le OtV19-T présente une asymétrie de sa capsid icosaédrique sur un vertex unique*

Globalement, les OtVs étudiés dans ce travail présentent une morphologie similaire à celle des virus géants eucaryotes icosaédriques décrits à ce jour. Ils présentent une capsid de symétrie icosaédrique avec une taille située entre 134 et 148 nm, et une membrane lipidique interne entourant le matériel génétique du virus. Toutefois, un d'entre eux, OtV19-T type-2 se caractérise par une asymétrie irrégulière et inhabituelle de sa capsid icosaédrique : la présence d'un *vertex* « enfoncé », conduisant à un étirement latéral de la capsid (**Figure 31**). Par rapport à une capsid icosaédrique virale, il est très surprenant de voir apparaître ce type d'asymétrie. D'ailleurs, à ce jour aucun autre NCLDV ne présente une telle forme. A la vue de cette anomalie, on pourrait penser que la capsid d'OtV19-T type-2 présente des modifications supplémentaires en plus du simple enfoncement d'un seul *vertex*, suggérant une adaptation possible de ce virus pour piloter différemment l'une ou l'autre des étapes critiques impliquant la capsid icosaédrique.

#### *Le virus OtV09-565 pourrait utiliser un processus biologique actif dans le déploiement d'une structure protubérante en forme de queue*

Nos travaux ont révélé une incroyable structure morphologique asymétrique de la particule de OtV09-565. En effet, ce virus se caractérise par la présence d'une capsid de forme

hexagonale avec une taille d'environ  $140 \pm 9$  nm, associée à une longue queue, non-cylindrique et flexible, située à la base d'un *vertex* unique. La morphologie de la capsid e d'OtV09-565 ressemble à celle décrite chez les autres *Prasinovirus*, et à notre connaissance aucune propriété d'asymétrie n'avait encore été observée<sup>59,71,213,229</sup>. En revanche, quelques travaux ont reporté la présence transitoire de structure en forme de queue chez certains autres virus d'algue verte eucaryote<sup>210,541,542</sup>. Une telle structure a été mise en évidence chez le virus *Micromonas* MpVN1 durant l'étape d'adsorption à la paroi de la cellule hôte<sup>222</sup>.

Les virus d'eucaryotes, de bactéries et même d'archées montrent diverses caractéristiques morphologiques à leur surface. Par exemple, le Tupanvirus se caractérise par la présence d'une couche dense de longues fibres, comme c'est le cas pour de nombreux virus d'amibes. En revanche, il possède également une grande structure cylindrique unique attachée à la base de la capsid e semblable à une queue<sup>181</sup>. L'ordre des *Caudovirales* rassemble l'ensemble des virus à queue qui infectent aussi bien les bactéries que les archées<sup>521</sup>. Les virus d'archées présentent une incroyable diversité qui se caractérisent par des morphotypes uniques jamais observés parmi les virus de bactérie et eucaryotes<sup>543</sup>. Parmi eux, les *Fusellovirus*<sup>544,545</sup>, *Salterprovirus*<sup>546,547</sup>, le *Bicaudavirus*<sup>548,549</sup>, et les *Thaspiviridae*<sup>550</sup> présentent en plus, un ou deux appendices en forme de queue, attachés à un ou deux pôles des particules. Des analyses très récentes ont permis de mettre en évidence le premier virus à queue infectant les archées hyperthermophiles méthanogènes de grandes profondeurs<sup>551</sup>. Si la morphologie et la structure de ces queues apparaissent distinctes, elles présentent vraisemblablement une fonction biologique commune parmi les virus. Notamment ces structures ont été le plus souvent observées durant les premières étapes d'infection : la reconnaissance, l'adsorption, la perforation de la paroi cellulaire et même le transfert de protéines associées et du génome dans le cytoplasme de la cellule hôte.

A ce stade, il est parfois possible d'observer des changements de morphotype du virion lors de la rencontre avec la cellule hôte sensible. Le déploiement d'un tube chez des virus normalement dépourvu de queue a été mise en évidence chez les *Tectiviridae* durant la libération de l'ADN dans la cellule hôte ou lors d'un traitement chloroformique<sup>552</sup>. Un autre exemple remarquable de modification de morphogénèse virale a été décrit chez le virus d'archée membre des *Bicaudavirus*, mais dans ce cas, ce phénomène se produit de façon totalement indépendante de son hôte<sup>548,549</sup>. Bien que la raison de son développement soit encore obscure, on estime que le virus a acquis des stratégies adaptatives pour survivre dans des conditions extrêmes à faible densité cellulaire<sup>549</sup>. Ainsi, cette structure protubérante permettrait

d'augmenter favorablement la probabilité de contact et d'adsorption entre le virus et la cellule hôte. Il est donc tentant de spéculer qu'OtV09-565 pourrait utiliser un processus biologique actif qui implique la protrusion d'appendice en forme de queue à partir de la particule, et qui serait uniquement induit dans des conditions spécifiques pour la reconnaissance et l'adhésion à la cellule d'*O. tauri*.

*Le virus OtV09-565 présente une stratégie évolutive adaptative semblable aux Siphoviridae dans les mécanismes de spécificité hôtes*

Au vu de la morphologie du virus OtV09-565 on pourrait penser que ces virus infectent les algues bleues ou cyanobactérie (i.e. les cyanophages). En revanche, la structure des particules de ce virus présente une taille plus grande et une forme légèrement différente de celle observée chez les cyanophages<sup>553</sup>. De plus, cet appendice se montre moins structuré que les bactériophages à queue et semble avoir un caractère très labile. Ainsi, nous pouvons nous interroger quant à la nature et à la fonction de cette structure dans le cycle de vie du virus. Toutefois, la comparaison entre les *Siphoviridae* et le virus OtV09-565 nous a paru pertinente dans cette analyse pour confronter les données génomiques aux observations de MET.

Les *Siphoviridae* sont une famille de virus de bactéries mais aussi d'archée, qui se caractérisent par une longue queue flexible non-contractile très structurée, attachée à une capsidie icosaédrique<sup>525</sup>. Dans cette discussion, nous nous focaliserons uniquement sur les *Siphoviridae* de bactérie (i.e. les bactériophages). La queue du bactériophage est une structure créée pour reconnaître les cellules de l'hôte, pénétrer la barrière de l'enveloppe cellulaire et assurer une libération efficace du génome dans le cytoplasme<sup>554</sup>. A ce jour, il est clairement reconnu que la structure de la queue du virus est un déterminant clé dans la spécificité de l'hôte<sup>523</sup>. Ce déterminisme est notamment favorisé par le développement de stratégie par le virus pour adapter leur *tail* à la diversité des récepteurs polysaccharidiques produits par leur hôte dans le but d'élargir leur spectre d'hôte<sup>555,556</sup>. Des changements mutationnels de gènes structuraux associés au module de morphogénèse chez les phages à queue ont été observés, conduisant ainsi à des régions centrales hypervariables au sein des génomes. Le caractère génomique est donc corrélé à la morphologie et au patron de spécificité chez les bactériophages à queue.

Dans les analyses génomiques (cf. Chapitre III), un sous-groupe d'OtVs (i.e type 2b) se caractérisait par un module hypervariable de gènes vraisemblablement impliqués dans la reconnaissance et l'adhésion à la cellule hôte. Ce module semble être comparable au module

de morphogénèse/reconnaissance des phages à queue, puisqu'une chaperonne de l'assemblage de la queue (*Tail Assembly Chaperon protein* – TAP ou TAC) – homologue aux phages marins – a pu être également identifiée chez OtV06-12, et OtV19-T2. Le plus remarquable est la présence de ce même module encodé dans le génome d'OtV09-565, bien que la chaperonne n'ait pas été détectée dans ce cas. Toutefois, les analyses de MET ont mis en évidence une structure protubérante en forme de queue, associée à la capsidie de ce virus, sur un *vertex* unique. Avec l'ensemble des analyses génomiques et microscopiques, il est donc tentant de spéculer que les OtVs du type 2b aurait acquis par transfert horizontal ce module hypervariable au cours de l'évolution pour moduler leur spectre d'hôte et s'adapter à un nouvel environnement comme c'est le cas pour les bactériophages. Il est encore très spéculatif de dire que ce module soit associé directement à la morphogénèse de la queue, mais la coïncidence est frappante. Cette région apparaît très variable et seuls les domaines en N-terminal et C-terminal des protéines sont conservés entre les virus de ce même groupe. Par ailleurs, nous ne pouvons pas assurer que cette région soit implicitement liée à la morphologie d'OtV09-565 puisqu'une telle structure n'a pas été observée pour les autres virus étudiés dans ce sous-groupe.

## 5. Conclusion

Les virus de morphologie asymétrique sont répandus à travers les règnes de la vie. Dans ce travail nous mettons en évidence que deux virus d'*O. tauri* présentent une asymétrie *via* un *vertex* de leur capsidie. Le premier, OtV19-T, présente un *vertex* en retrait aboutissant à une structure non-icosaédrique. Des études de la particule à plus haute résolution et au cours de l'infection avec l'hôte sont nécessaires pour comprendre l'intérêt d'une telle adaptation au niveau du cycle de vie du virus. Pour le deuxième virus, malgré la difficulté de son observation, cette protubérance en forme de queue attachée sur un *vertex* d'OtV09-565 laisse penser que ce virus utiliserait un processus biologique actif dans le déploiement de cette structure comme un dispositif de reconnaissance et d'attachement à la cellule hôte. Nous suggérons également que cette structure puisse être exprimée par le virus uniquement dans des conditions expérimentales spécifiques. Toutefois, des analyses complémentaires par des approches de MET devront être menées à bien pour tester ces hypothèses.





# DISCUSSION GENERALE ET PERSPECTIVES

## *Chapitre I et II – Bases génétiques sous-jacentes de la résistance antivirale chez *Ostreococcus tauri**

Le premier objectif de ma thèse a été de gagner en connaissances sur les bases génétiques de la résistance aux *Prasinovirus* chez l'algue marine *Ostreococcus tauri*. Pour répondre à cette question, nous avons choisi d'utiliser une approche génétique classique de mutagenèse insertionnelle aléatoire par intégration d'ADN exogène d'origine plasmidique dans le génome nucléaire de l'algue. Le développement de la transformation médiée par le polyéthylène glycol (PEG) a constitué une avancée considérable notamment dans la réduction du délai, la facilité d'utilisation mais également l'augmentation considérable de l'efficacité de transformation en comparaison à l'électroporation<sup>322</sup>.

La mutagenèse insertionnelle aléatoire constitue un premier pas dans la caractérisation fonctionnelle des gènes. Si cet outil génétique reste l'une des méthodes les plus puissantes et facilement applicables sur des organismes taxonomiquement divers, qui se reproduisent rapidement et sont génétiquement faciles à manipuler comme c'est le cas pour notre modèle d'étude, il pose toutefois un certain nombre de problèmes<sup>557</sup>. En effet, la mutagenèse insertionnelle aléatoire conduit généralement à une perte incomplète de la fonction du gène, voire même à une mutation silencieuse selon le site d'insertion. La nature non-dirigée de cette méthode requiert donc la génération d'un très grand nombre d'individus pour obtenir un mutant présentant les caractéristiques phénotypiques désirées. La large collection de mutants ainsi générée nécessite un criblage phénotypique massif accompagné d'une caractérisation génomique des évènements d'intégrations.

Dans le premier chapitre de thèse, la génération d'une très grande collection de mutants (plus de 1200 lignées clonales transformées) a été réalisée. Le séquençage à haut débit de nouvelle génération (NGS) de cinq génomes complets de lignées clonales mutantes nous a démontré que la transformation médiée par le PEG conduit à un très faible nombre d'évènements insertionnels et totalement aléatoires dans le génome de l'algue, ce qui est un avantage considérable dans le cadre de notre étude. Parmi ces génomes, l'intégration d'ADN exogène a conduit à de fréquentes variations structurelles ainsi que quelques réarrangements chromosomiques à proximité des sites d'insertions, suggérant l'utilisation d'un mécanisme de

recombinaison non-homologue de type NHEJ (*non-homologous end-joining*) par l'algue. Puisque la majorité des loci insertionnels (i.e. 80% soit 4 génomes sur 5) ont été détectés dans des régions non-codantes (i.e. introns, UTR ou intergénique), il est probable que l'insertion n'ait pas entraîné l'inactivation de gènes, mais qu'elle ait plutôt modifié son expression. A cette occasion, il serait intéressant d'évaluer leurs niveaux d'expression à partir d'une quantification relative par PCR à transcription inverse (*Quantitative reverse transcription PCR – RT-qPCR*). Toutefois – à l'exception de la lignée T14 qui présentait une importante délétion sur son chromosome SOC – nous ne sommes pas certains que les mutants séquencés soient des candidats intéressants dans le cadre de notre étude qui visait à décrire les bases génétiques de la résistance antivirale chez l'algue *O. tauri*.

Ainsi, toujours en utilisant le PEG comme facteur d'incorporation d'ADN exogène dans le noyau hôte, il serait plus pertinent de conduire une expérience de mutagenèse dirigée pour l'extinction de gènes d'intérêts. Le ciblage de gènes par recombinaison homologue (RH) a été décrit une seule fois chez l'algue *O. tauri*<sup>321</sup>. Il est fortement possible que NHEJ soit le processus biologique prédominant du mécanisme de réparation de l'ADN, au détriment du mécanisme RH. Ceci pourrait être une première explication aux obstacles rencontrés en utilisant un segment d'ADN exogène homologue comme modèle du gène ciblé. La très faible efficacité par la RH pourrait constituer un autre biais à son utilisation. En effet, si le gène cible n'a aucun moyen de sélection phénotypique, le criblage de nombreux clones doit obligatoirement être réalisé par PCR.

D'autre part, la suppression d'expression de gènes par la technologie d'interférence à ARN (*RNAi*) n'est pas envisagée chez *O. tauri* en raison de l'absence des enzymes clés de type Dicer et Argonaute-Piwi (*AGO-Piwi*) dans son génome<sup>558</sup>. Dans ce cas, des technologies moléculaires alternatives pour l'édition des génomes pourraient être utilisées. En effet, plusieurs approches classiques d'ingénierie des génomes existent, nécessitant des nucléases modifiées telles que *TAL effector endonuclease* (TALEN) *ZINC-finger nuclease* (ZFN), *homing meganucleases*, et *Clustered Regularly Interspaced Short Palindromic Repeat* (CRISPR)<sup>559–563</sup>. Parmi ces nucléases, la technologie CRISPR de type II, qui utilise une seule protéine effectrice, la Cas9 (*CRISPR associated 9*) et un guide ARN (sgRNA)<sup>564</sup>, a été utilisée avec succès depuis quelques années pour l'extinction ciblée de gènes dans divers organismes modèles<sup>565–568</sup>. Grâce à son efficacité, sa précision, et sa simplicité amplement plus élevées par rapport aux autres nucléases, cette technologie a aussi été appliquée pour modifier efficacement le génome de microalgues taxonomiquement diverses, dont des diatomées<sup>569,570</sup>, *Chlamydomonas reinhardtii*, *Phaeodactylum tricornutum* ou *Nannochloropsis oceanica*<sup>571–573</sup>.

Il est donc fortement probable que cette technologie puisse être également adaptée au modèle *O. tauri*.

Dans cette étude, nous avons mis en évidence que la souche *O. tauri* RCC1115 semble contenir un plus grand nombre d'éléments transposables (ETs) de classe I (rétrotransposons à LTR) constitutivement actifs et mobiles et donc un génome plus dynamique en comparaison à la souche domestiquée *O. tauri* RCC4221. Ainsi, les prochaines expériences de mutagenèses devront être conduites à partir d'une souche clonale afin de réduire au maximum la variabilité génomique naturelle observée dans ce travail. Par ailleurs, il serait plus judicieux d'utiliser la souche domestiquée *O. tauri* RCC4221 comme modèle d'étude – pour la suppression de gènes d'intérêts dans les mécanismes immunitaires de résistance antivirale – en raison de sa plus grande stabilité génomique et phénotypique, dans le cas où les lignées clonales ne sont pas soumises à l'infection virale<sup>310</sup>.

*Les ETs actifs participent à la dynamique du génome d'O. tauri : une possible conséquence de la variation phénotypique rapide face à l'infection virale*

Dans le chapitre II, l'identification et la dynamique des ETs dans les sept lignées clonales *O. tauri* RCC1115 ont été mises en évidence – uniquement dans les chromosomes où les insertions d'ADN plasmidique s'étaient produites – par la comparaison de leur génome respectif à celui de leur souche parentale d'origine. Cette exploration nous a conduit à conclure sur le rôle potentiel des ETs dans la dynamique des génomes des souches naturelles et de manière plus générale dans la plasticité du génome de l'espèce *O. tauri*. Tenant compte de l'importance des ETs dans la variabilité génétique, ces éléments sont aussi de conséquents régulateurs de l'expression des gènes de leurs hôtes<sup>574</sup>, pouvant conduire à une grande diversité phénotypique entre espèces. En référence à ce propos, une analyse génomique des populations a été réalisée chez deux espèces florales dans le but de mieux comprendre les mécanismes qui sous-tendent leur adaptabilité aux changements environnementaux<sup>413</sup>. Cette étude a mis en évidence le rôle phare des ETs dans la diversification phénotypique rapide reliée à la diversité, la distribution et l'expression de ces éléments dans le génome de ces deux espèces hôtes.

En raison d'un faible nombre d'études visant à l'identification et la classification des ETs dans le génome des microalgues<sup>402,575,576</sup>, il serait intéressant de réaliser une analyse complète du paysage des ETs chez la souche *O. tauri* RCC1115. D'autre part, une analyse comparative de la dynamique des ETs entre les sept lignées clonales indépendantes obtenues à

partir de la souche cryptique *O. tauri* RCC1115, nous permettrait peut-être de mieux comprendre les rapides changements phénotypiques face à l'infection virales (i.e. réversion sensible-résistant) : une conséquence plausible de la variation génétique chromosomique observée dans ce travail.

Chez *O. tauri*, une annotation exhaustive des ETs a été réalisée<sup>40</sup>. Toutefois, aucune analyse génomique comparative de ces éléments entre différentes espèces et souches n'a encore été effectuée. Considérant cela, il serait pertinent de conduire une étude à plus grande échelle visant à identifier et caractériser les ETs chez différentes souches de l'espèce *O. tauri*. Cette analyse permettrait de comparer la diversité, la proportion et la distribution des ETs dans les génomes d'une même espèce, présentant une grande variation phénotypique entre les souches face à l'infection virale. Cette étude permettrait de mieux comprendre les raisons d'un spectre de sensibilité plus large chez l'algue *O. tauri* RCC4221 et plus restreint chez la souche *O. tauri* RCC1115 face à un large panel de virus (OtVs) testés. En ce sens, une analyse de génomique des populations a été conduite chez la picoalgue *O. tauri* pour 13 souches différentes<sup>70</sup>. Cette étude met en avant un important polymorphisme génétique associé à de grandes différences phénotypiques entre les souches avec une corrélation positive entre la taille du chromosome SOC et la résistance aux *Prasinovirus*. La capacité dynamique des ETs à créer un large panel de mutations lors de leur transposition<sup>577</sup> pourrait donc être une des principales raisons expliquant une telle plasticité génomique chez *O. tauri*. Ainsi, notre analyse permet de souligner l'implication des ETs dans le façonnement de la diversité des génomes d'*O. tauri* et plus largement dans l'évolution adaptative des microalgues face aux pressions environnementales.

Par ailleurs, l'étude de la diversité des ETs chez des espèces cryptiques d'*Ostreococcus* pourrait renseigner sur des éventuels événements de transferts latéraux d'ETs entre des espèces taxonomiquement proches coexistant dans l'environnement naturel. En conséquence, ceci permettrait aux espèces d'acquérir de nouvelles fonctions biologiques et donc de s'adapter<sup>577-579</sup>. De plus, le rôle des virus comme vecteurs de tels transferts est une piste à ne pas écarter, étant donné l'existence d'une étroite association entre les deux partenaires dans les régions marines tropicales. Vérifier cette hypothèse implique une identification de ETs potentiels au sein de différentes souches virales taxonomiquement et géographiquement proches, puis une analyse génomique comparative des ETs entre l'hôte et son virus associé. De tels événements ont été mis en évidence récemment chez les insectes et leurs virus associés à ADN double-brin, les Baculovirus<sup>580</sup> : un processus biologique vraisemblablement commun à plusieurs lignées eucaryotes<sup>581</sup>.

*Quels sont les mécanismes qui sous-tendent le contrôle de la dynamique des éléments transposables dans le génome d'O. tauri ?*

Pour aller plus loin dans la compréhension des mécanismes de résistance associés aux ETs chez *O. tauri*, il serait intéressant de mesurer le niveau d'expression des ETs dans le génome d'une culture en condition de stress (i.e. infection virale) vis-à-vis d'une condition contrôle (culture non infectée). Quelques études ont fourni l'évidence claire de la réactivation de rétroéléments dans le génome de microalgues en conditions de stress<sup>402,582-584</sup>. Chez la souche *O. tauri* RCC4221, des travaux préliminaires ont montré une surexpression d'une transcriptase inverse associée à un rétrotransposon sur le chromosome 8 chez les cultures infectées, tandis que ce gène n'était pas transcrit dans les souches contrôles<sup>309</sup>. L'expression différentielle des ETs chez l'algue *O. tauri* apparaît comme un processus biologique actif, vraisemblablement induit durant un stress biotique telle qu'est l'infection virale<sup>310</sup>. Cette analyse pourrait alors donner des indications sur les mécanismes de régulation de l'expression des ETs chez *O. tauri*, et également permettre de mieux comprendre les processus biologiques qui maintiennent un faible taux d'insertion des ETs dans le génome de cette espèce. Chez la plupart des organismes eucaryotes, la méthylation de l'ADN et d'histone ainsi que RNAi et piRNA, sont les principaux mécanismes épigénétiques contrôlant la répression des transposons au sein des génomes<sup>585</sup>. En revanche, en raison de l'absence d'un certain nombre de composants nécessaires au *silencing*, il se pourrait qu' *O. tauri* requiert un mécanisme alternatif<sup>426</sup> encore méconnu à ce jour. Dans ce contexte, le projet ANR ALGALVIRUS, dont je fais partie, a pour principal objectif de gagner en connaissances sur la méthylation de l'ADN et la structure de la chromatine dans les interactions *Ostreococcus-Prasinovirus*. Ce travail est coordonné par le Laboratoire Génome et Développement des Plantes (LGDP) à Perpignan par Thierry Lagrange et Dominique Pontier.

*Chapitre III – Caractérisation génomique des prasinovirus : description des bases génétiques associées à la sélectivité de reconnaissance des hôtes de l'espèce Ostreococcus tauri*

Le second objectif de ma thèse a été de décrire plus largement la diversité des *Prasinovirus* spécifiques de l'espèce *O. tauri* à l'échelle de l'océan mondial. Pour cela, une analyse génomique des populations de 18 génomes complets a été réalisée. Ces virus ont été isolés entre 2006 et 2019 dans trois régions côtières marines géographiquement distinctes : la

côte Est de l'Océan Pacifique, les lagunes à l'Ouest de la Mer Méditerranée dans le Golfe du Lion, ainsi que la Mer du Nord. Dans cette étude, nous avons pu mettre en évidence que la biogéographie des isolats viraux ne présentait pas un patron clair en rapport avec la distance génétique entre différents sous-groupes d'OtVs. Toutefois, le groupe désigné comme OtV-type 1, qui englobe la plupart des virus de l'Océan Pacifique, a montré une relation phylogénétique et taxonomique plus étroite avec OtV6 et avec les virus de *Micromonas* qu'avec les autres virus OtVs majoritairement issus de la Mer Méditerranée. Tenant compte de la cospéciation et du schéma cophylogénétique observé entre les *Prasinovirus* et leurs hôtes associés<sup>224</sup>, il est probable qu'un changement d'hôte se soit produit récemment chez les OtVs-type 1, élargissant donc la sélectivité de reconnaissance de leurs hôtes en s'affranchissant de la barrière de l'espèce et du genre. Il est donc plausible que l'hôte « naturel » de ces virus soit affilié au genre *Micromonas* ou alors à un clade plus éloigné du genre *Ostreococcus* encore non décrit à ce jour. Il serait pertinent d'élargir le test croisé de spécificité des OtVs sur une plus large gamme d'hôtes, notamment en utilisant différentes souches de *Micromonas*. Par ailleurs, les prochaines campagnes d'isolement de nouveaux virus devront porter davantage attention sur l'isolement des virus et de leurs hôtes associés issus du même échantillon environnemental.

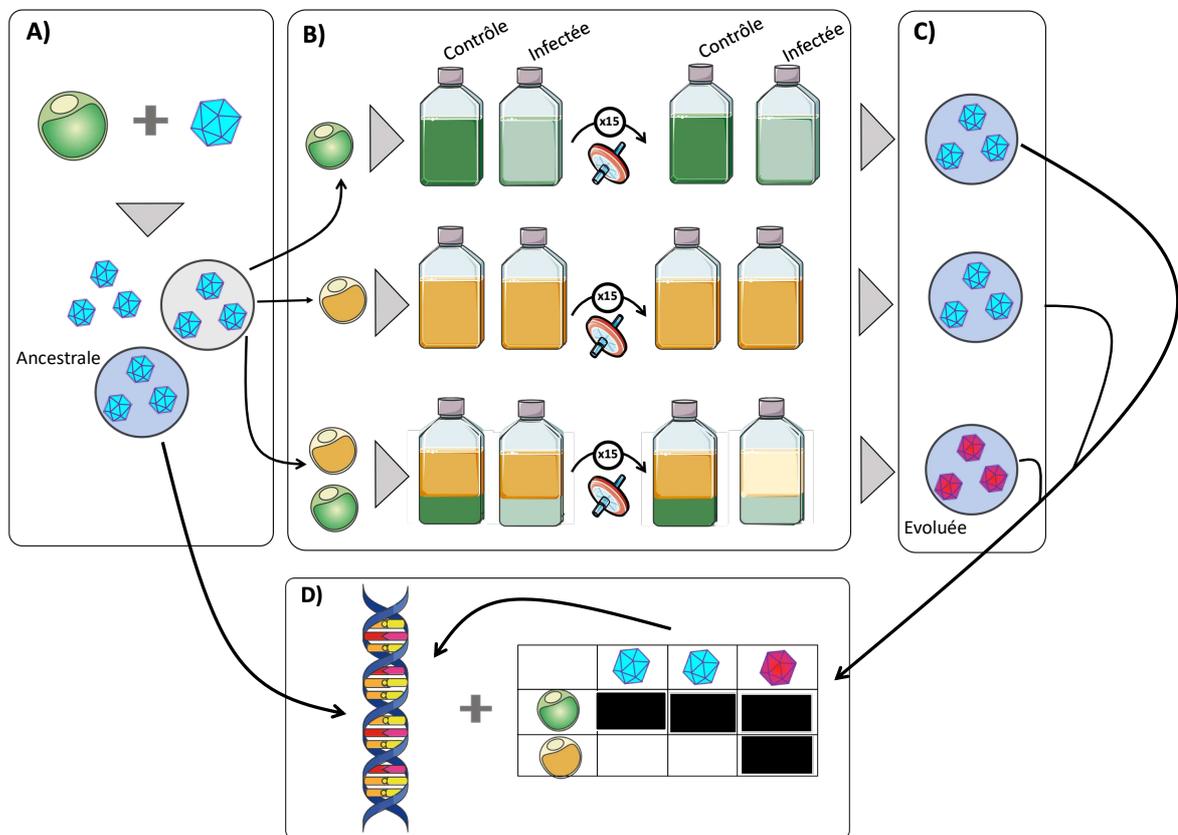
L'absence de relation entre la biogéographie et la diversité génétique des OtVs décrit dans ce travail pourrait être expliquée par le trop faible jeu de données, qui par conséquent, ne serait pas suffisamment représentatif de la répartition et de la diversité « réelle » des virus dans les communautés naturelles. Dans le prolongement de notre étude, il serait donc pertinent de conduire une analyse visant à évaluer la distribution biogéographique des isolats OtVs séquencés dans les océans, en utilisant les données métagénomiques actuellement disponibles qui couvrent de nombreuses régions géographiques marines (e.g. les projets « *Ocean Sampling Day* » (OSD) et *Tara Oceans*). De plus, considérant que la structure des communautés virales reflète la distribution des hôtes, nous pourrions essayer de corrélérer la distribution des OtVs à celle des picoeucaryotes afin d'évaluer les possibles associations communautaires entre le virus et son hôte.

Le second objectif de ce chapitre a été de gagner en connaissances sur les bases génétiques de l'évolution du patron d'infection des *Prasinovirus*. Pour ce faire, une collection d'OtVs présentant une variabilité dans la gamme d'hôtes a été sélectionnée en constituant 2 groupes (spécialistes vs. généralistes) pour le séquençage et l'analyse de leurs génomes complets. Cette analyse a mise en évidence une région centrale hyperplastique dans les génomes

des OtVs. Suite à l'annotation et la prédiction de domaines fonctionnels des gènes encodés dans cette région, nous avons découvert qu'une majorité d'entre eux sont associés à des fonctions impliquées dans la reconnaissance et l'adhésion de la cellule hôte (i.e. adhésine /chaperonne peptidase S74, et lectine). A notre grande surprise, cette région est d'autant plus étendue avec un nombre de gènes additionnels que le spectre d'infection est large (les virus généralistes, comme OtV-type 2b). Chez les *Caudovirales*, des études ont démontré la corrélation positive entre l'hyperplasticité de la séquence de l'adhésine, modulant la spécificité de reconnaissance de la queue du virus aux récepteurs des cellules hôtes, et l'expansion de la gamme d'hôte des virus<sup>487,504,505</sup>. Notre étude met en évidence pour la première fois qu'un tel schéma coévolutif avec leurs hôtes pourrait se produire chez les virus phytoplanctoniques eucaryotes, engagés dans une dynamique de course aux armements. Dans ce contexte, la grande diversité de séquences observée dans la région plastique entre les génomes viraux – d'autant plus importante entre des virus de groupe différents – pourrait être expliquée par la forte pression de sélection et l'accumulation de mutations successives rapides : une possible conséquence de l'expansion du spectre d'hôte chez les virus généralistes.

Pour aller plus loin dans cette analyse et valider notre hypothèse, la construction de protéines de fusion exprimant les gènes d'intérêts uniquement codés dans le génome de virus généralistes pourrait être envisagée. Cette étude validerait ou non la fonction biologique des gènes suspectés dans l'élargissement du spectre d'hôtes. Dans cette expérience, il serait intéressant de réaliser une comparaison de l'infection d'une souche hautement résistante (i.e. *O. tauri* RCC1561) non-hôte à un virus spécialiste (i.e. OtV5) comme condition contrôle vis-à-vis d'une culture infectée (i.e. *O. tauri* RCC1561–OtV5) en présence de la protéine purifiée exprimée d'intérêt. Si notre hypothèse est valable, on s'attend à observer un changement de phénotype de l'algue et donc un élargissement de la gamme d'hôte du virus.

Une méthode alternative à l'utilisation d'outils génétiques moléculaires serait de conduire une approche expérimentale pour suivre l'évolution de la spécificité des virus. Cette méthode a été utilisée dans le système phage-bactérie afin de mieux comprendre les processus sous-jacents, les conséquences évolutives et les coûts d'adaptation de l'expansion de la gamme d'hôtes<sup>286,586,587</sup>. Considérant que la compétition intraspécifique pour des ressources est un agent moteur de la sélection divergente génétique, celle-ci favoriserait l'émergence de nouveaux phénotypes pour l'utilisation de ces mêmes ressources<sup>588</sup>. Ce propos a été démontré lors d'une expérience de coévolution dans le système *Pseudomonas*-Ø6, dans lequel des populations virales étaient plus susceptibles d'évoluer vers une gamme d'hôtes élargie lorsque la concurrence pour les hôtes étaient fortes<sup>589</sup>.



**Figure 35.** Procédure expérimentale d'évolution de la gamme d'hôte. A) Génération d'une population clonale virale ancestrale à partir d'une population hôte clonale. B) Adaptations à des hôtes spécifiques. Un échantillon de la population virale clonale ancestrale sera incubé dans trois cultures d'algues indépendamment : la souche hôte contrôle (vert), une souche non-hôte (jaune) et un mélange en proportion 1:1 (v:v) des deux souches (vert et jaune). A la suite d'une lyse (~J+3) de la culture infectée hôte, le lysat viral filtré à 0.2µm sera transféré pour infecter à nouveau la population hôte « naïve ». Ceci pourra être réalisé sur 15 transferts successifs ou plus. C) Populations évoluées. D) Caractérisation du phénotype et du génotype de la population ancestrale et population évoluée. Un test

d'infection croisée par plage de lyse sera réalisé. Une case noire indique une infection réussie tandis qu'une case blanche indique une absence d'infection (droite). Le séquençage des génomes complets de la population évoluée sera comparé à celui de la population ancestrale (gauche). Figure adaptée de Enav et al. 2018.

La **Figure 35** présente le plan expérimental de coévolution qui pourrait être conduit dans de futures études. L'évolution d'une population virale au spectre restreint serait réalisée par transfert en série successif dans des cultures de microalgues fraîches contenant un mélange 1:1 de l'hôte standard et d'un nouvel hôte (non-hôte). Afin de pallier aux éventuels effets liés au temps ou aux conditions expérimentales sur le nombre de mutations présentes dans la population virale, une condition contrôle devra être menée dans laquelle les virus ont évolué avec une population « non-évolutive » d'un seul génotype d'hôte. A chaque transfert, les profils d'infectivité des populations virales individuelles obtenues au cours de l'évolution seront évalués par test croisé sur plage de lyse. Pour finir, le séquençage du génome complet des populations virales individuelles – qui comprend la population ancestrale et la population évoluée qui présenterait une expansion de son spectre d'hôte capable d'infecter la nouvel hôte – sera réalisé afin d'établir une relation éventuelle entre phénotype et génotype.

#### Chapitre IV – *Caractérisation morphologique des virus d'Ostreococcus tauri*

Ce travail a été réalisé dans le prolongement du chapitre III dans le but de mieux comprendre les interactions *Ostreococcus-Prasinovirus* et les éventuelles raisons pouvant expliquer la spécificité de ces virus à l'échelle microscopique. Par ailleurs, nous avons cherché à gagner en connaissances sur le paysage des morphotypes d'OtVs nouvellement isolés dans ce travail de thèse avec une origine géographiquement distincte. Si la plupart des virus ont une morphologie comparable aux virus d'eucaryotes phytoplanctoniques déjà décrit dans la littérature (i.e. une symétrie icosaédrique), cette exploration a conduit à la découverte de deux virus qui avaient des caractéristiques structurales très surprenantes encore jamais décrites chez les NCLDVs. Le virus OtV19-T, provenant d'un échantillon environnemental originaire de la côte Est de l'Océan Pacifique, a montré une population morphologique et génomique hétérogène de deux virions. L'une d'entre elle présente une capsid non-enveloppée non-icosaédrique, avec une interruption de la symétrie au niveau d'un seul *vertex*, vraisemblablement plus arrondie. Ce défaut serait lié à l'allongement latéral des deux côtés de la capsid et à « l'enfoncement » de ce même *vertex*. A notre grande surprise, les deux sous-

populations de morphotype viral ont présenté également une diversité génétique distincte (présenté dans le chapitre III): OtV19-T1 rejoint phylogénétiquement aux OtV-type1 alors qu'OtV19-T2 est proche d'OtV-type2b. Néanmoins, la double identification morphologique et génomique n'a pas été possible à ce stade. A cette occasion, il serait judicieux de réitérer un clonage de cet échantillon par plage de lyse ou par dilution successive, séparant ainsi les deux sous-populations, dans le cas où les deux virus ne forment pas un consortium mutualiste symbiotique. Une amplification par PCR d'un gène discriminant les deux génomes pourrait ensuite être conduite afin de différencier la morphologie du génome et de confirmer l'identité de ces deux virus par séquençage par la méthode de Sanger.

Le second virus à la morphologie atypique que nous avons décrit dans cette étude est OtV09-565. Les observations de ce virus nous ont permis de révéler une remarquable structure protubérante en forme de queue située le long de sa capsidie icosaédrique. En raison des propriétés hautement labiles et des difficultés d'observation de cette queue, nous avons réalisé des observations microscopiques du lysat viral dans différentes conditions de préparations, à différents temps après l'infection, et durant les étapes clés du cycle de vie du virus avec la cellule infectée. Nous avons fait l'hypothèse que le virus OtV09-565 utiliserait un processus biologique actif dans le déploiement d'une telle structure au cours de son cycle de reproduction, certainement impliquée dans la reconnaissance et l'adhésion de la cellule hôte. Afin de vérifier ce propos, nous devons à l'avenir réitérer l'expérience en repartant d'un échantillon frais, et réaliser une confirmation d'identité du virus par amplification partielle de la *PolB* par PCR et séquençage par la méthode de Sanger. La production et la concentration du lysat viral devront être suivies d'une purification sur un gradient de sucrose par ultracentrifugation. L'observation par MET sera alors réalisée en coloration négative à l'acétate d'uranyle à partir de la bande de particules virales concentrées et hautement purifiées.

Dans la littérature, nombre d'études ont souligné l'influence majeure des facteurs environnementaux sur le processus d'infection virale. Une perte ou une diminution de l'infectivité virale a par exemple été mise en évidence après l'augmentation de température de la culture infectée<sup>118,203,225,590,591</sup>. A ce propos, une augmentation de la température induit une résistance chez le coccolithophore *E. huxleyi* à son virus (EhV)<sup>592</sup>. Il serait donc intéressant, dans le cadre de notre étude, de faire varier les conditions de température de la culture et de réaliser un suivi de l'infection vis-à-vis d'une condition optimale contrôle. On pourrait envisager qu'un changement physiologique et métabolique de l'hôte conduise à des

changements morphologiques chez le virus, tel que le déploiement de la queue par exemple. En ce sens, le développement d'une longue queue chez un virus d'une archée hyperthermophile membre des *Bicaudavirus* a été mis en évidence durant des conditions de température plus élevées que celles de l'habitat naturel. Toutefois, ce processus biologique apparaît indépendamment de son hôte<sup>548,549</sup>.

Bien que la morphologie commune aux *Phycodnaviridae* a été mise en évidence dans un grand nombre d'études, l'approche par MET n'est pas suffisamment résolutive pour évaluer les caractéristiques structurelles de la capsidie icosaédrique nucléaire des virus. La reconstruction d'images en trois-dimension combinée à la cryomicroscopie électronique (*cryogenic Electron Microscopy* – cryoEM) a été réalisée chez un certain nombre de NCLDV<sub>s</sub><sup>193,194,511,515,537,593–595</sup>, dont le chlorovirus PBCV-1<sup>194,195,197,198</sup>. Néanmoins, aucune approche de ce type n'a été conduite jusqu'à présent chez les *Prasinovirus*. Nous avons démontré dans cette étude que la morphologie des *Prasinovirus* n'est pas un caractère commun et que des formes et des structures particulières peuvent subsister dans les populations naturelles. Des études de reconstructions 3D en cryoEM pourraient donc être envisagées chez les OtVs. L'objectif de ce travail permettrait de mieux comprendre la disposition des sous-unités protéiques formant la capsidie et d'identifier la « vrai » capsidie majeure, puisque plusieurs gènes de capsidie non-identiques (i.e. entre 7 et 9 copies) existent dans les génomes des *Prasinovirus*. De plus, cette approche permettrait de mieux évaluer les éventuelles raisons de symétries non-icosaédrique décrites ici. La microscopie à force atomique pourrait être utilisée en complément de la cryoEM afin d'obtenir plus de détails. Cette méthode de reconstruction (i.e. symétrie *five-fold*) a permis de révéler le *vertex* unique et spécifique en forme d'étoile de mer chez le Mimivirus, et la présence d'une *spike* chez le chlorovirus PBCV-1. Considérant cela, il est plausible que cette approche puisse mettre en évidence des caractères structuraux atypiques chez les deux virus OtV19-T et OtV09-565 et même chez l'unique virus de la Mer du Nord Sylt2-5, puisque de fines fibrilles décorant la capsidie du virion en question ont été observées dans notre étude.

La reconstruction structurale de l'enveloppe protéique externe des *Prasinovirus* à l'échelle atomique est donc à prendre en grande considération pour de futures études. Elle permettrait de faire un saut considérable dans nos connaissances sur l'interaction hôte-virus et plus particulièrement sur les processus d'infection qui sous-tendent la spécificité de la gamme d'hôte chez les virus.

## *Conclusion générale*

Ce travail de thèse a eu pour objectif de décrire pour la première fois les interactions *Ostreococcus tauri*–*Prasinovirus* régissant la dynamique de coévolution dans ce système hôte-virus à l'échelle génomique. L'identification et la dynamique de rétroéléments actifs dans le génome de la souche *O. tauri* RCC115 semblent clairement indiquer leurs rôles clés dans les mécanismes intrinsèques de l'immunité antivirale chez cette picoalgue. Afin d'apporter des réponses à nos résultats préliminaires, le paysage global des éléments transposables dans le génome de différentes souches d'*O. tauri* devra être décrit et détaillé à l'avenir. Par ailleurs, une analyse comparative du niveau d'expression et la dynamique des ETs devra être conduite dans une culture contrôle vis-à-vis d'une culture infectée. Parallèlement, nous avons pu apporter des réponses encore jamais décrites à ce jour sur les bases génétiques de la sélectivité de reconnaissance des hôtes chez les *Prasinovirus*, ce qui ouvre vers de nouvelles études sur les mécanismes d'infection dans le système *Ostreococcus tauri*–*Prasinovirus*.

L'ensemble de ces résultats très prometteurs confirment l'intérêt d'étudier l'interaction hôte-virus en combinant différentes échelles (génomique et microscopique) : cette combinaison nous a permis de démasquer des mécanismes biologiques au cœur de ce système *Ostreococcus tauri*–*Prasinovirus*. Finalement, ces travaux de thèse offrent une nouvelle porte d'entrée à l'étude des interactions hôte-virus dans les communautés naturelles pour aider à mieux comprendre le phénomène de coévolution à plus grande échelle.





## REFERENCES BIBLIOGRAPHIQUES

1. Falkowski, P. G., Barber, R. T. & Smetacek, V. Biogeochemical controls and feedbacks on ocean primary production. *Science* **281**, 200–206 (1998).
2. Field, C. B. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
3. Behrenfeld, M. J. *et al.* Biospheric primary production during an ENSO transition. *Science* **291**, 2594–2597 (2001).
4. Worden, A. Z. *et al.* Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science* **347**, (2015).
5. Azam, F. *et al.* The ecological role of water-column microbes in the Sea. *Mar. Ecol. Prog. Ser.* **10**, 257–263 (1983).
6. Azam, F. Microbial control of oceanic Ccarbon flux: The plot thickens. *Science* **280**, 694–696 (1998).
7. Fuhrman, J. A. & Azam, F. Bacterioplankton secondary production estimates for coastal waters of british Columbia, Antarctica, and California. *Appl Environ Microbiol* **39**, 1085–1095 (1980).
8. Larsson, U. & Hagström, A. Phytoplankton exudate release as an energy source for the growth of pelagic bacteria. *Mar. Biol.* **52**, 199–206 (1979).
9. Hedges, S. B., Blair, J. E., Venturi, M. L. & Shoe, J. L. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evolutionary Biology* **4**, 2 (2004).
10. Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G. & Bhattacharya, D. A Molecular timeline for the origin of photosynthetic eukaryotes. *Molecular Biology and Evolution* **21**, 809–818 (2004).
11. Buick, R. When did oxygenic photosynthesis evolve? *Philos Trans R Soc Lond B Biol Sci* **363**, 2731–2743 (2008).
12. Soo, R. M., Hemp, J., Parks, D. H., Fischer, W. W. & Hugenholtz, P. On the origins of oxygenic photosynthesis and aerobic respiration in *Cyanobacteria*. *Science* **355**, 1436–1440 (2017).
13. Archibald, J. M. The puzzle of plastid evolution. *Current Biology* **19**, R81–R88 (2009).
14. Keeling, P. J. The endosymbiotic origin, diversification and fate of plastids. *Philos Trans R Soc Lond B Biol Sci* **365**, 729–748 (2010).
15. Burki, F., Roger, A. J., Brown, M. W. & Simpson, A. G. B. The new tree of eukaryotes. *Trends in Ecology & Evolution* **35**, 43–55 (2020).
16. Simpson, A. G. B. & Eglit, Y. Protist diversification. in *Encyclopedia of Evolutionary Biology* 344–360 (Elsevier, 2016).
17. Morris, J. L. *et al.* The timescale of early land plant evolution. *PNAS* **115**, E2274–E2283 (2018).
18. Butterfield, N., Knoll, A. & Swett, K. A bangiophyte red alga from the Proterozoic of arctic Canada. *Science* **250**, 104–107 (1990).
19. Leliaert, F. *et al.* Phylogeny and molecular evolution of the green algae. *Critical Reviews in Plant Sciences* **31**, 1–46 (2012).

20. Not, F. *et al.* Diversity and ecology of eukaryotic marine phytoplankton. in *Advances in Botanical Research* vol. 64 1–53 (Elsevier, 2012).
21. Lewis, L. A. & McCourt, R. M. Green algae and the origin of land plants. *Am J Bot* **91**, 1535–1556 (2004).
22. Toulza, E., Blanc-Mathieu, R., Gourbière, S. & Piganeau, G. Chapter ten - environmental and evolutionary genomics of microbial algae: Power and challenges of metagenomics. in *Advances in Botanical Research* (ed. Piganeau, G.) vol. 64 383–427 (Academic Press, 2012).
23. Bik, H. M. *et al.* Sequencing our way towards understanding global eukaryotic biodiversity. *Trends Ecol Evol* **27**, 233–243 (2012).
24. Leliaert, F., Verbruggen, H. & Zechman, F. W. Into the deep: New discoveries at the base of the green plant phylogeny. *BioEssays* **33**, 683–692 (2011).
25. Li, L. *et al.* The genome of *Prasinoderma coloniale* unveils the existence of a third phylum within green plants. *Nat Ecol Evol* **4**, 1220–1231 (2020).
26. Marin, B. & Melkonian, M. Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-encoded rRNA operons. *Protist* **161**, 304–336 (2010).
27. Vaultot, D., Eikrem, W., Viprey, M. & Moreau, H. The diversity of small eukaryotic phytoplankton ( $\leq 3 \mu\text{m}$ ) in marine ecosystems. *FEMS Microbiol Rev* **32**, 795–820 (2008).
28. Massana, R. Eukaryotic picoplankton in surface oceans (2011).
29. Monier, A., Worden, A. Z. & Richards, T. A. Phylogenetic diversity and biogeography of the Mamiellophyceae lineage of eukaryotic phytoplankton across the oceans. *Environmental Microbiology Reports* **8**, 461–469 (2016).
30. Tragin, M. & Vaultot, D. Green microalgae in marine coastal waters: The Ocean Sampling Day (OSD) dataset. *Scientific Reports* **8**, 14020 (2018).
31. Tragin, M. & Vaultot, D. Novel diversity within marine Mamiellophyceae (Chlorophyta) unveiled by metabarcoding. *Scientific Reports* **9**, 5190 (2019).
32. Belevich, T. A., Milyutina, I. A., Abyzova, G. A. & Troitsky, A. V. The pico-sized Mamiellophyceae and a novel *Bathycoccus* clade from the summer plankton of Russian Arctic Seas and adjacent waters. *FEMS Microbiology Ecology* **97**, (2021).
33. Lovejoy, C. *et al.* Distribution, phylogeny, and growth of cold-adapted Picoprasinophytes in Arctic Seas I. *Journal of Phycology* **43**, 78–89 (2007).
34. Lovejoy, C. & Potvin, M. Microbial eukaryotic distribution in a dynamic Beaufort Sea and the Arctic Ocean. *Journal of Plankton Research* **33**, 431–444 (2011).
35. Viprey, M., Guillou, L., Ferréol, M. & Vaultot, D. Wide genetic diversity of picoplanktonic green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR approach. *Environ Microbiol* **10**, 1804–1822 (2008).
36. Courties, C. *et al.* Smallest eukaryotic organism. *Nature* **370**, 255–255 (1994).
37. Foss, P., Guillard, R. R. L. & Liaen-Jensen, S. Prasinoxanthin—a chemosystematic marker for algae. *Phytochemistry* **23**, 1629–1633 (1984).
38. Foss, P., Guillard, R. R. L. & Synnøve, L.-J. Carotenoids from eucaryotic ultraplankton clones (prasinophyceae). *Phytochemistry* **25**, 119–124 (1985).
39. Latasa, M., Scharek, R., Gall, F. L. & Guillou, L. Pigment suites and taxonomic groups in *Prasinophyceae*. *Journal of Phycology* **40**, 1149–1155 (2004).

40. Derelle, E. *et al.* Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc Natl Acad Sci U S A* **103**, 11647–11652 (2006).
41. Palenik, B. *et al.* The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *PNAS* **104**, 7705–7710 (2007).
42. Worden, A. Z. *et al.* Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* **324**, 268–272 (2009).
43. Moreau, H. *et al.* Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biology* **13**, R74 (2012).
44. Van Baren, M. J. *et al.* Evidence-based green algal genomics reveals marine diversity and ancestral characteristics of land plants. *BMC Genomics* **17**, 1–22 (2016).
45. Henderson, G. P., Gan, L. & Jensen, G. J. 3-D Ultrastructure of *O. tauri*: Electron cryotomography of an entire eukaryotic cell. *PLOS ONE* **2**, e749 (2007).
46. Rodríguez, F. *et al.* Ecotype diversity in the marine picoeukaryote *Ostreococcus* (Chlorophyta, Prasinophyceae): Ecotype diversity in the picoeukaryote *Ostreococcus*. *Environmental Microbiology* **7**, 853–859 (2005).
47. Šlapeta, J., López-García, P. & Moreira, D. Global dispersal and ancient cryptic species in the smallest marine eukaryotes. *Molecular Biology and Evolution* **23**, 23–29 (2006).
48. Subirana, L. *et al.* Morphology, genome plasticity, and phylogeny in the genus *Ostreococcus* reveal a cryptic species, *O. mediterraneus* sp. nov. (Mamiellales, Mamiellophyceae). *Protist* **164**, 643–659 (2013).
49. Vannier, T. *et al.* Survey of the green picoalga *Bathycoccus* genomes in the global ocean. *Sci Rep* **6**, 37900 (2016).
50. Foulon, E. *et al.* Ecological niche partitioning in the picoplanktonic green alga *Micromonas pusilla*: evidence from environmental surveys using phylogenetic probes. *Environmental Microbiology* **10**, 2433–2443 (2008).
51. Jancek, S., Gourbière, S., Moreau, H. & Piganeau, G. Clues about the genetic basis of adaptation emerge from comparing the proteomes of two *Ostreococcus* ecotypes (Chlorophyta, Prasinophyceae). *Mol Biol Evol* **25**, 2293–2300 (2008).
52. Not, F. *et al.* A Single species, *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in the Western English Channel. *Appl. Environ. Microbiol.* **70**, 4064–4072 (2004).
53. Simon, N. *et al.* Revision of the genus *Micromonas* Manton et Parke (Chlorophyta, Mamiellophyceae), of the type species *M. pusilla* (Butcher) Manton & Parke and of the Species *M. commoda* van Baren, Bachy and Worden and description of two new Species based on the genetic and phenotypic characterization of cultured isolates. *Protist* **168**, 612–635 (2017).
54. Tragin, M., Lopes dos Santos, A., Christen, R. & Vaultot, D. Diversity and ecology of green microalgae in marine systems: an overview based on 18S rRNA gene sequences. *pip* **3**, 141–154 (2016).
55. Butcher, R. W. Contributions to our knowledge of the smaller marine algae. *J. Mar. Biol. Ass.* **31**, 175–191 (1952).

56. Benites, L. F. *et al.* Single cell ecogenomics reveals mating types of individual cells and ssDNA viral infections in the smallest photosynthetic eukaryotes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20190089 (2019).
57. Simmons, M. P. *et al.* Abundance and biogeography of picoprasinophyte ecotypes and other phytoplankton in the Eastern North Pacific Ocean. *Applied and Environmental Microbiology* **82**, 1693–1705 (2016).
58. Vaultot, D. *et al.* Metagenomes of the picoalga *Bathycoccus* from the Chile coastal upwelling. *PLOS ONE* **7**, e39648 (2012).
59. Bachy, C. *et al.* Viruses infecting a warm water picoeukaryote shed light on spatial co-occurrence dynamics of marine viruses and their hosts. *The ISME Journal* 1–19 (2021) doi:10.1038/s41396-021-00989-9.
60. Limardo, A. J. *et al.* Quantitative biogeography of picoprasinophytes establishes ecotype distributions and significant contributions to marine phytoplankton. *Environmental Microbiology* **19**, 3219–3234 (2017).
61. Chrétiennot-Dinet, M.-J. *et al.* A new marine picoeucaryote: *Ostreococcus tauri* gen. et sp. nov. (Chlorophyta, Prasinophyceae). *Phycologia* **34**, 285–292 (1995).
62. Vaquer, A., Troussellier, M., Courties, C. & Bibent, B. Standing stock and dynamics of picophytoplankton in the Thau Lagoon (northwest Mediterranean coast). *Limnology and Oceanography* **41**, 1821–1828 (1996).
63. Guillou, L. *et al.* Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. *Protist* **155**, 193–214 (2004).
64. Countway, P. D. & Caron, D. A. Abundance and distribution of *Ostreococcus* sp. in the San Pedro Channel, California, as revealed by quantitative PCR. *Applied and Environmental Microbiology* **72**, 2496–2506 (2006).
65. O’Kelly, C. J., Sieracki, M. E., Thier, E. C. & Hobson, I. C. A Transient bloom of *Ostreococcus* (chlorophyta, Prasinophyceae) in West Neck Bay, Long Island, New York. *Journal of Phycology* **39**, 850–854 (2003).
66. Demir-Hilton, E. *et al.* Global distribution patterns of distinct clades of the photosynthetic picoeukaryote *Ostreococcus*. *The ISME Journal* **5**, 1095–1107 (2011).
67. Treusch, A. H. *et al.* Phytoplankton distribution patterns in the northwestern Sargasso Sea revealed by small subunit rRNA genes from plastids. *ISME J* **6**, 481–492 (2012).
68. Romari, K. & Vaultot, D. Composition and temporal variability of picoeukaryote communities at a coastal site of the English Channel from 18S rDNA sequences. *Limnology and Oceanography* **49**, 784–798 (2004).
69. Blanc-Mathieu, R. *et al.* An improved genome of the model marine alga *Ostreococcus tauri* unfolds by assessing Illumina *de novo* assemblies. *BMC Genomics* **15**, 1103 (2014).
70. Blanc-Mathieu, R. *et al.* Population genomics of picophytoplankton unveils novel chromosome hypervariability. *Science Advances* **3**, e1700239 (2017).
71. Yau, S. *et al.* Virus-host coexistence in phytoplankton through the genomic lens. *Science Advances* **6**, eaay2587 (2020).
72. Li, X.-Q. & Du, D. Variation, evolution, and correlation analysis of C+G content and genome or chromosome size in different kingdoms and phyla. *PLOS ONE* **9**, e88339 (2014).

73. Verhelst, B., Van de Peer, Y. & Rouzé, P. The complex intron landscape and massive intron invasion in a picoeukaryote provides insights into intron evolution. *Genome Biol Evol* **5**, 2393–2401 (2013).
74. Grimsley, N., Péquin, B., Bachy, C., Moreau, H. & Piganeau, G. Cryptic sex in the smallest eukaryotic marine green alga. *Molecular Biology and Evolution* **27**, 47–54 (2010).
75. Lecoq, H. Découverte du premier virus, le virus de la mosaïque du tabac : 1892 ou 1898 ? *Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie* **324**, 929–933 (2001).
76. Jacob, F. & Wollman, E. L. Viruses and Genes. *Sci Am* **204**, 92–110 (1961).
77. Bândeă, C. I. A new theory on the origin and the nature of viruses. *J Theor Biol* **105**, 591–602 (1983).
78. Forterre, P. Manipulation of cellular syntheses and the nature of viruses: The virocell concept. *Comptes Rendus Chimie* **14**, 392–399 (2011).
79. Patrick, F. Virocell Concept, The. in *eLS* (John Wiley & Sons, Ltd, 2012).
80. Forterre, P. The virocell concept and environmental microbiology. *ISME J* **7**, 233–236 (2013).
81. Baltimore, D. Expression of animal virus genomes. *Bacteriological Reviews* **35**, 235–241 (1971).
82. Part III. The ICTV. in *Virus Taxonomy* (eds. King, A. M. Q., Adams, M. J., Carstens, E. B. & Lefkowitz, E. J.) 1261–1291 (Elsevier, 2012).
83. Simmonds, P. 2015. Methods for virus classification and the challenge of incorporating metagenomic sequence data. *Journal of General Virology* **96**, 1193–1206 (2015).
84. Simmonds, P. *et al.* Virus taxonomy in the age of metagenomics. *Nat Rev Microbiol* **15**, 161–168 (2017).
85. Czyz, A., Los, M., Wrobel, B. & Wegrzyn, G. Inhibition of spontaneous induction of lambdoid prophages in *Escherichia coli* cultures: simple procedures with possible biotechnological applications. *BMC Biotechnol* **1**, 1 (2001).
86. Casjens, S. R. & Hendrix, R. W. Bacteriophage lambda: Early pioneer and still relevant. *Virology* **479–480**, 310–330 (2015).
87. Howard-Varona, C., Hargreaves, K. R., Abedon, S. T. & Sullivan, M. B. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J* **11**, 1511–1520 (2017).
88. Krupovič, M. & Bamford, D. H. Archaeal proviruses TKV4 and MVV extend the PRD1-adenovirus lineage to the phylum Euryarchaeota. *Virology* **375**, 292–300 (2008).
89. Müller, D. G., Kapp, M. & Knippers, R. Viruses in marine brown algae. in *Advances in Virus Research* (eds. Maramorosch, K., Murphy, F. A. & Shatkin, A. J.) vol. 50 49–67 (Academic Press, 1998).
90. Meints, R. H., Ivey, R. G., Lee, A. M. & Choi, T.-J. Identification of two virus integration sites in the brown alga *Feldmannia* chromosome. *J Virol* **82**, 1407–1413 (2008).
91. Fischer, M. G. & Hackl, T. Host genome integration and giant virus-induced reactivation of the virophage mavirus. *Nature* **540**, 288–291 (2016).
92. Desnues, C. *et al.* Provirophages and transpovirons as the diverse mobilome of giant viruses. *PNAS* **109**, 18078–18083 (2012).

93. Spencer, R. A Marine Bacteriophage. *Nature* **175**, 690–691 (1955).
94. Bergh, Ø., BØrsheim, K. Y., Bratbak, G. & Haldal, M. High abundance of viruses found in aquatic environments. *Nature* **340**, 467–468 (1989).
95. Suttle, C. A. Viruses in the sea. *Nature* **437**, 356–361 (2005).
96. Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nat Rev Microbiol* **5**, 801–812 (2007).
97. Gregory, A. C. *et al.* Marine DNA viral macro- and microdiversity from pole to pole. *Cell* **0**, (2019).
98. Roux, S. *et al.* Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**, 689–693 (2016).
99. Schulz, F. *et al.* Giant virus diversity and host interactions through global metagenomics. *Nature* (2020).
100. Hingamp, P. *et al.* Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *ISME J* **7**, 1678–1695 (2013).
101. Coy, S. R., Gann, E. R., Pound, H. L., Short, S. M. & Wilhelm, S. W. Viruses of Eukaryotic Algae: Diversity, Methods for Detection, and Future Directions. *Viruses* **10**, (2018).
102. Short, S. M., Staniewski, M. A., Chaban, Y. V., Long, A. M. & Wang, D. Diversity of Viruses Infecting Eukaryotic Algae. *Curr Issues Mol Biol* **39**, 29–62 (2020).
103. Tai, V. *et al.* Characterization of HaRNAV, a single-stranded RNA virus causing lysis of *Heterosigma akashiwo* (Rhaphidophyceae). *Journal of Phycology* **39**, 343–352 (2003).
104. Tomaru, Y. *et al.* Isolation and characterization of two distinct types of HcRNAV, a single-stranded RNA virus infecting the bivalve-killing microalga *Heterocapsa circularisquama*. (2004)
105. Shirai, Y. *et al.* Isolation and characterization of a single-stranded RNA virus infecting the marine planktonic diatom *Chaetoceros tenuissimus* Meunier. *Appl Environ Microbiol* **74**, 4022–4027 (2008).
106. Tomaru, Y., Takao, Y., Suzuki, H., Nagumo, T. & Nagasaki, K. Isolation and characterization of a single-stranded RNA virus infecting the bloom-forming diatom *Chaetoceros socialis*. *Appl Environ Microbiol* **75**, 2375–2381 (2009).
107. Kimura, K. & Tomaru, Y. Discovery of two novel viruses expands the diversity of single-stranded DNA and single-stranded RNA viruses infecting a cosmopolitan marine diatom. *Applied and Environmental Microbiology* **81**, 1120–1131 (2015).
108. Kimura, K. & Tomaru, Y. Isolation and Characterization of a single-stranded DNA virus infecting the marine diatom *Chaetoceros* sp. Strain SS628-11 isolated from western Japan. *PLOS ONE* **8**, e82013 (2013).
109. Nagasaki, K. *et al.* Isolation and Characterization of a novel single-stranded RNA virus infecting the bloom-forming diatom *Rhizosolenia setigera*. *Appl Environ Microbiol* **70**, 704–711 (2004).
110. Tomaru, Y. *et al.* First evidence for the existence of pennate diatom viruses. *ISME J* **6**, 1445–1448 (2012).
111. Culley, A. I., Lang, A. S. & Suttle, C. A. The complete genomes of three viruses assembled from shotgun libraries of marine RNA virus communities. *Virology* **4**, 1–9 (2007).

112. Greninger, A. L. & DeRisi, J. L. Draft genome sequences of marine RNA viruses SF-1, SF-2, and SF-3 recovered from San Francisco wastewater. *Genome Announcements* **3**, e00653-15 (2015).
113. Vlok, M., Lang, A. S. & Suttle, C. A. Marine RNA virus quasispecies are distributed throughout the oceans. *mSphere* **4**, e00157-19 (2019).
114. Chase, E. E. *et al.* A High rate algal pond hosting a dynamic community of RNA viruses. *Viruses* **13**, 2163 (2021).
115. Lang, A. S. *et al.* ICTV virus taxonomy profile: *Marnaviridae* 2021. *Journal of General Virology* **102**, 001633 (2021).
116. Brussaard, C. P. D., Noordeloos, A. A. M., Sandaa, R.-A., Heldal, M. & Bratbak, G. Discovery of a dsRNA virus infecting the marine photosynthetic protist *Micromonas pusilla*. *Virology* **319**, 280–291 (2004).
117. Attoui, H. *et al.* *Micromonas pusilla reovirus*: a new member of the family *Reoviridae* assigned to a novel proposed genus (Mimoreovirus). *Journal of General Virology* **87**, 1375–1383 (2006).
118. Maat, D. S. *et al.* Characterization and temperature dependence of Arctic *Micromonas polaris* viruses. *Viruses* **9**, 134 (2017).
119. Tomaru, Y., Shirai, Y., Suzuki, H., Nagumo, T. & Nagasaki, K. Isolation and characterization of a new single-stranded DNA virus infecting the cosmopolitan marine diatom *Chaetoceros debilis*. *Aquatic Microbial Ecology* **50**, 103–112 (2008).
120. Tomaru, Y., Shirai, Y., Toyoda, K. & Nagasaki, K. Isolation and characterisation of a single-stranded DNA virus infecting the marine planktonic diatom *Chaetoceros tenuissimus*. *Aquatic Microbial Ecology* **64**, 175–184 (2011).
121. Eissler, Y., Wang, K., Chen, F., Eric Wommack, K. & Wayne Coats, D. Ultrastructural characterization of the lytic cycle of an intranuclear virus infecting the diatom *Chaetoceros Cf. Wighamii*(bacillariophyceae) from Chesapeake Bay, Usa1. *Journal of Phycology* **45**, 787–797 (2009).
122. Nagasaki, K. *et al.* Previously unknown virus infects marine diatom. *Applied and Environmental Microbiology* **71**, 3528–3535 (2005).
123. Koonin, E. V. *et al.* Global organization and proposed megataxonomy of the virus World. *Microbiol Mol Biol Rev* **84**, e00061-19 (2020).
124. Van Etten, J. L., Lane, L. C. & Dunigan, D. D. DNA viruses: the really big ones (giruses). *Annu Rev Microbiol* **64**, 83–99 (2010).
125. Philippe, N. *et al.* *Pandoraviruses*: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* **341**, 281–286 (2013).
126. Legendre, M. *et al.* Diversity and evolution of the emerging *Pandoraviridae* family. *Nat Commun* **9**, 1–12 (2018).
127. Legendre, M. *et al.* Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *PNAS* **111**, 4274–4279 (2014).
128. La Scola, B. L. *et al.* A Giant Virus in Amoebae. *Science* **299**, 2033–2033 (2003).
129. Raoult, D. *et al.* The 1.2-megabase genome sequence of Mimivirus. *Science* **306**, 1344–1350 (2004).
130. Iyer, L. M., Aravind, L. & Koonin, E. V. Common origin of four diverse families of large eukaryotic DNA viruses. *Journal of Virology* **75**, 11720–11734 (2001).

131. Carradec, Q. *et al.* A global ocean atlas of eukaryotic genes. *Nat Commun* **9**, 373 (2018).
132. Schulz, F. *et al.* Hidden diversity of soil giant viruses. *Nat Commun* **9**, 4881 (2018).
133. Endo, H. *et al.* Biogeography of marine giant viruses reveals their interplay with eukaryotes and ecological functions. *Nature Ecology & Evolution* 1–11 (2020).
134. Iyer, L. M., Balaji, S., Koonin, E. V. & Aravind, L. Evolutionary genomics of nucleocytoplasmic large DNA viruses. *Virus Research* **117**, 156–184 (2006).
135. Boyer, M., Madoui, M.-A., Gimenez, G., Scola, B. L. & Raoult, D. Phylogenetic and Phyletic Studies of Informational Genes in Genomes Highlight Existence of a 4th Domain of Life Including Giant Viruses. *PLOS ONE* **5**, e15530 (2010).
136. Yutin, N., Colson, P., Raoult, D. & Koonin, E. V. Mimiviridae: clusters of orthologous genes, reconstruction of gene repertoire evolution and proposed expansion of the giant virus family. *Virol J* **10**, 1–13 (2013).
137. Guglielmini, J., Woo, A. C., Krupovic, M., Forterre, P. & Gaia, M. Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. *PNAS* **116**, 19585–19592 (2019).
138. Claverie, J.-M. Fundamental Difficulties Prevent the reconstruction of the deep phylogeny of viruses. *Viruses* **12**, 1130 (2020).
139. Yutin, N. & Koonin, E. V. Hidden evolutionary complexity of Nucleo-Cytoplasmic Large DNA viruses of eukaryotes. *Virol J* **9**, 1–18 (2012).
140. Koonin, E. V. & Yutin, N. Evolution of the large nucleocytoplasmic DNA viruses of eukaryotes and convergent origins of viral gigantism. in *Advances in Virus Research* vol. 103 167–202 (Elsevier, 2019).
141. Aylward, F. O., Moniruzzaman, M., Ha, A. D. & Koonin, E. V. A phylogenomic framework for charting the diversity and evolution of giant viruses. *PLOS Biology* **19**, e3001430 (2021).
142. Sun, T.-W. & Ku, C. Unraveling gene content variation across eukaryotic giant viruses based on network analyses and host associations. *Virus Evolution* **7**, (2021).
143. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).
144. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**, W256–W259 (2019).
145. Larsen, J. B., Larsen, A., Bratbak, G. & Sandaa, R.-A. Phylogenetic Analysis of Members of the Phycodnaviridae Virus Family, Using Amplified Fragments of the Major Capsid Protein Gene. *Applied and Environmental Microbiology* **74**, 3048–3057 (2008).
146. Monier, A. *et al.* Marine mimivirus relatives are probably large algal viruses. *Virol J* **5**, 1–8 (2008).
147. Monier, A. *et al.* Horizontal gene transfer of an entire metabolic pathway between a eukaryotic alga and its DNA virus. *Genome Res* **19**, 1441–1449 (2009).
148. Andreani, J., Verneau, J., Raoult, D., Levasseur, A. & La Scola, B. Deciphering viral presences: two novel partial giant viruses detected in marine metagenome and in a mine drainage metagenome. *Virol J* **15**, 1–5 (2018).
149. Needham, D. M. *et al.* A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *PNAS* **116**, 20574–20583 (2019).
150. Mihara, T. *et al.* Linking Virus genomes with host taxonomy. *Viruses* **8**, 66 (2016).

151. Subramaniam, K. *et al.* A New family of DNA viruses causing disease in crustaceans from diverse aquatic biomes. *mBio* **11**, (2020).
152. Suzan-Monti, M., Scola, B. L., Barrassi, L., Espinosa, L. & Raoult, D. Ultrastructural characterization of the giant volcano-like virus factory of *Acanthamoeba polyphaga* *Mimivirus*. *PLOS ONE* **2**, e328 (2007).
153. Katsafanas, G. C. & Moss, B. Colocalization of transcription and translation within cytoplasmic Poxvirus factories coordinates viral expression and subjugates host functions. *Cell Host & Microbe* **2**, 221–228 (2007).
154. Abergel, C., Legendre, M. & Claverie, J.-M. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiol Rev* **39**, 779–796 (2015).
155. Condit, R. C., Moussatche, N. & Traktman, P. In a nutshell: structure and assembly of the vaccinia virion. in *Advances in Virus Research* vol. 66 31–124 (Academic Press, 2006).
156. Boyer, M. *et al.* Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *PNAS* **106**, 21848–21853 (2009).
157. Blanca, L., Christo-Foroux, E., Rigou, S. & Legendre, M. Comparative analysis of the circular and highly asymmetrical *Marseilleviridae* genomes. *Viruses* **12**, 1270 (2020).
158. Legendre, M. *et al.* In-depth study of *Mollivirus sibericum*, a new 30,000-y-old giant virus infecting *Acanthamoeba*. *PNAS* **112**, E5327–E5335 (2015).
159. Van Etten, J. L., Burbank, D. E., Joshi, J. & Meints, R. H. DNA synthesis in a *Chlorella*-like alga following infection with the virus PBCV-1. *Virology* **134**, 443–449 (1984).
160. Van Etten, J. L. & Meints, R. H. Giant viruses infecting algae. *Annu. Rev. Microbiol.* **53**, 447–494 (1999).
161. Kang, M., Dunigan, D. D. & VAN Etten, J. L. Chlorovirus: a genus of *Phycodnaviridae* that infects certain *chlorella*-like green algae. *Mol Plant Pathol* **6**, 213–224 (2005).
162. Wilson, W. H. *et al.* Complete genome sequence and lytic phase transcription profile of a *Coccolithovirus*. *Science* **309**, 1090–1092 (2005).
163. García-Beato, R., Salas, M. L., Viñuela, E. & Salas, J. Role of the host cell nucleus in the replication of African swine fever virus DNA. *Virology* **188**, 637–649 (1992).
164. Goorha, R. Frog virus 3 DNA replication occurs in two stages. *J Virol* **43**, 519–528 (1982).
165. Fabre, E. *et al.* *Noumeavirus* replication relies on a transient remote control of the host nucleus. *Nat Commun* **8**, 15087 (2017).
166. Yoshikawa, G. *et al.* *Medusavirus*, a novel large DNA virus discovered from hot spring water. *J Virol* **93**, (2019).
167. Zhang, R., Endo, H., Takemura, M. & Ogata, H. RNA sequencing of Medusavirus suggests remodeling of the host nuclear environment at an early infection stage. *Microbiology Spectrum* **9**, e00064-21 (2021).
168. Santini, S. *et al.* 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* **110**, 10800–10805 (2013).
169. Moniruzzaman, M. *et al.* Genome of brown tide virus (AaV), the little giant of the Megaviridae, elucidates NCLDV genome expansion and host–virus coevolution. *Virology* **466–467**, 60–70 (2014).

170. Koonin, E. V. & Yutin, N. Multiple evolutionary origins of giant viruses. *F1000Res* **7**, F1000 Faculty Rev-1840 (2018).
171. Claverie, J.-M. Viruses take center stage in cellular evolution. *Genome Biol* **7**, 110 (2006).
172. Yutin, N., Wolf, Y. I. & Koonin, E. V. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. *Virology* **0**, 38–52 (2014).
173. Filée, J. Route of NCLDV evolution: the genomic accordion. *Current Opinion in Virology* **3**, 595–599 (2013).
174. Filée, J. Genomic comparison of closely related Giant Viruses supports an accordion-like model of evolution. *Frontiers in Microbiology* **6**, 593 (2015).
175. Monier, A., Claverie, J.-M. & Ogata, H. Taxonomic distribution of large DNA viruses in the sea. *Genome Biol* **9**, 1–15 (2008).
176. Mihara, T. *et al.* Taxon Richness of ‘Megaviridae’ Exceeds those of Bacteria and Archaea in the Ocean. *Microbes Environ* **33**, 162–171 (2018).
177. Gallot-Lavallée, L., Blanc, G. & Claverie, J.-M. Comparative genomics of *Chrysochromulina ericina virus* and other microalga-infecting large DNA viruses highlights their intricate evolutionary relationship with the established *Mimiviridae* family. *Journal of Virology* **91**, e00230-17 (2017).
178. Arslan, D., Legendre, M., Seltzer, V., Abergel, C. & Claverie, J.-M. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proc Natl Acad Sci U S A* **108**, 17486–17491 (2011).
179. Yoosuf, N. *et al.* Related Giant Viruses in Distant Locations and Different Habitats: *Acanthamoeba polyphaga moumouvirus* Represents a Third Lineage of the *Mimiviridae* that is close to the Megavirus Lineage. *Genome Biology and Evolution* **4**, 1324–1330 (2012).
180. Boratto, P. V. M. *et al.* A mysterious 80 nm amoeba virus with a near-complete “ORFan genome” challenges the classification of DNA viruses. *bioRxiv* 2020.01.28.923185 (2020).
181. Abrahão, J. *et al.* Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nat Commun* **9**, 749 (2018).
182. Chen, F. & Suttle, C. A. Amplification of DNA polymerase gene fragments from viruses infecting microalgae. *Appl. Environ. Microbiol.* **61**, 1274–1278 (1995).
183. Chen, F. & Suttle, C. A. Evolutionary relationships among large double-stranded DNA viruses that infect microalgae and other organisms as inferred from DNA polymerase genes. *Virology* **219**, 170–178 (1996).
184. Park, Y., Lee, K., Lee, Y. S., Kim, S. W. & Choi, T.-J. Detection of diverse marine algal viruses in the South Sea regions of Korea by PCR amplification of the DNA polymerase and major capsid protein genes. *Virus Research* **159**, 43–50 (2011).
185. Short, S. M. & Suttle, C. A. Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. *Applied and Environmental Microbiology* **68**, 1290–1296 (2002).
186. Short, S. M. & Short, C. M. Diversity of algal viruses in various North American freshwater environments. *Aquatic Microbial Ecology* **51**, 13–21 (2008).

187. Clasen, J. L. & Suttle, C. A. Identification of freshwater *Phycodnaviridae* and their potential phytoplankton hosts, using DNA pol sequence fragments and a genetic-distance analysis. *Applied and Environmental Microbiology* **75**, 991–997 (2009).
188. Short, S., Rusanova, O. & Staniewski, M. Novel phycodnavirus genes amplified from Canadian freshwater environments. *Aquat. Microb. Ecol.* **63**, 61–67 (2011).
189. Zhong, X. & Jacquet, S. Contrasting diversity of phycodnavirus signature genes in two large and deep western European lakes. *Environ. Microbiol.* **16**, 759–773 (2014).
190. Moniruzzaman, M., Martinez-Gutierrez, C. A., Weinheimer, A. R. & Aylward, F. O. Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. *Nat Commun* **11**, 1710 (2020).
191. Wilson, W. H., Van Etten, J. L. & Allen, M. J. The *Phycodnaviridae*: The story of how tiny giants rule the world. *Curr Top Microbiol Immunol* **328**, 1–42 (2009).
192. Klose, T. *et al.* The three-dimensional structure of Mimivirus. *Intervirology* **53**, 268–273 (2010).
193. Okamoto, K. *et al.* Cryo-EM structure of a *Marseilleviridae* virus particle reveals a large internal microassembly. *Virology* **516**, 239–245 (2018).
194. Yan, X. *et al.* Structure and assembly of large lipid-containing dsDNA viruses. *Nat Struct Mol Biol* **7**, 101–103 (2000).
195. Fang, Q. *et al.* Near-atomic structure of a giant virus. *Nat Commun* **10**, 1–11 (2019).
196. Mackinder, L. C. M. *et al.* A unicellular algal virus, *Emiliana huxleyi* virus 86, exploits an animal-like infection strategy. *Journal of General Virology* **90**, 2306–2316 (2009).
197. Cherrier, M. V. *et al.* An icosahedral algal virus has a complex unique vertex decorated by a spike. *Proc Natl Acad Sci U S A* **106**, 11085–11089 (2009).
198. Zhang, X. *et al.* Three-dimensional structure and function of the *Paramecium bursaria* chlorella virus capsid. *PNAS* **108**, 14837–14842 (2011).
199. Sandaa, R.-A., Haldal, M., Castberg, T., Thyraug, R. & Bratbak, G. Isolation and characterization of two viruses with large genome size infecting *Chrysochromulina ericina* (Prymnesiophyceae) and *Pyramimonas orientalis* (Prasinophyceae). *Virology* **290**, 272–280 (2001).
200. Johannessen, T. V. *et al.* Characterisation of three novel giant viruses reveals huge diversity among viruses infecting Prymnesiales (Haptophyta). *Virology* **476**, 180–188 (2015).
201. Blanc-Mathieu, R. *et al.* A persistent giant algal virus, with a unique morphology, encodes an unprecedented number of genes involved in energy metabolism. *Journal of Virology* (2021) doi:10.1128/JVI.02446-20.
202. Rowe, J. M. *et al.* Isolation of a non-phage-like lytic virus infecting *Aureococcus Anophagefferens*. *Journal of Phycology* **44**, 71–76 (2008).
203. Baudoux, A.-C. & Brussaard, C. P. D. Characterization of different viruses infecting the marine harmful algal bloom species *Phaeocystis globosa*. *Virology* **341**, 80–90 (2005).
204. Schvarcz, C. R. & Steward, G. F. A giant virus infecting green algae encodes key fermentation genes. *Virology* **518**, 423–433 (2018).
205. Nagasaki, K. & Yamaguchi, M. Isolation of a virus infectious to the harmful bloom causing microalga *Heterosigma akashiwo* (Raphidophyceae). *Aquat. Microb. Ecol.* **13**, 135–140 (1997).

206. Nagasaki, K., Shirai, Y., Tomaru, Y., Nishida, K. & Pietrokovski, S. Algal viruses with distinct intraspecies host specificities include identical intein elements. *Appl Environ Microbiol* **71**, 3599–3607 (2005).
207. Wilson, W. H. *et al.* Isolation of viruses responsible for the demise of an *Emiliania huxleyi* bloom in the English Channel. *Journal of the Marine Biological Association of the United Kingdom* **82**, 369–377 (2002).
208. Meints, R. H., Lee, K., Burbank, D. E. & Van Etten, J. L. Infection of a *chlorella*-like alga with the virus, PBCV-1: Ultrastructural studies. *Virology* **138**, 341–346 (1984).
209. Van Etten, J. L. *et al.* Isolation and characterization of a virus from the intracellular green alga symbiotic with *Hydra viridis*. *Virology* **113**, 704–711 (1981).
210. Van Etten, J. L., Lane, L. C. & Meints, R. H. Viruses and virus-like particles of eukaryotic algae. *Microbiol Rev* **55**, 586–620 (1991).
211. Müller, D. Marine virioplankton produced by infected *Ectocarpus siliculosus* (Phaeophyceae). *Mar. Ecol. Prog. Ser.* **76**, 101–102 (1991).
212. Delaroque, N. *et al.* The complete DNA sequence of the *Ectocarpus siliculosus virus EsV-1* genome. *Virology* **287**, 112–132 (2001).
213. Mayer, J. A. & Taylor, F. J. R. A virus which lyses the marine nanoflagellate *Micromonas pusilla*. *Nature* **281**, 299–301 (1979).
214. Moreau, H. *et al.* Marine prasinovirus genomes show low evolutionary divergence and acquisition of protein metabolism genes by horizontal gene transfer. *Journal of Virology* **84**, 12555–12563 (2010).
215. Suttle, C. & Chan, A. Viruses infecting the marine *Prymnesiophyte* *Chrysochromulina* spp.: isolation, preliminary characterization and natural abundance. *Mar. Ecol. Prog. Ser.* **118**, 275–282 (1995).
216. Bellec, L., Grimsley, N., Derelle, E., Moreau, H. & Desdevises, Y. Abundance, spatial distribution and genetic diversity of *Ostreococcus tauri* viruses in two different environments. *Environmental Microbiology Reports* **2**, 313–321 (2010).
217. Bellec, L., Grimsley, N. & Desdevises, Y. Isolation of prasinoviruses of the green unicellular algae *Ostreococcus* spp. on a worldwide geographical scale. *Appl. Environ. Microbiol.* **76**, 96–101 (2010).
218. Clerissi, C. *et al.* Unveiling of the diversity of prasinoviruses (*Phycodnaviridae*) in marine samples by using high-throughput sequencing analyses of PCR-amplified DNA polymerase and major capsid protein genes. *Appl Environ Microbiol* **80**, 3150–3160 (2014).
219. Sahlsten, E. Seasonal abundance in Skagerrak-Kattegat coastal waters and host specificity of viruses infecting the marine photosynthetic flagellate *Micromonas pusilla*. (1998)
220. Baudoux, A.-C. *et al.* Interplay between the genetic clades of *Micromonas* and their viruses in the Western English Channel. *Environ Microbiol Rep* **7**, 765–773 (2015).
221. Martin, W. F., Garg, S. & Zimorski, V. Endosymbiotic theories for eukaryote origin. *Philos Trans R Soc Lond B Biol Sci* **370**, 20140330 (2015).
222. Zingone, A. *et al.* Diversity in morphology, infectivity, molecular characteristics and induced host resistance between two viruses infecting *Micromonas pusilla*. (2006)

223. Cottrell, M. & Suttle, C. Wide-spread occurrence and clonal variation in viruses which cause lysis of a cosmopolitan, eukaryotic marine phytoplankter *Micromonas pusilla*. *Mar. Ecol. Prog. Ser.* **78**, 1–9 (1991).
224. Bellec, L. *et al.* Cophylogenetic interactions between marine viruses and eukaryotic picophytoplankton. *BMC Evol. Biol.* **14**, 59 (2014).
225. Martínez, J. M. *et al.* New lipid envelope-containing dsDNA virus isolates infecting *Micromonas pusilla* reveal a separate phylogenetic group. *Aquatic Microbial Ecology* **74**, 17–28 (2015).
226. Bellec, L., Grimsley, N., Moreau, H. & Desdevises, Y. Phylogenetic analysis of new Prasinoviruses (*Phycodnaviridae*) that infect the green unicellular algae *Ostreococcus*, *Bathycoccus* and *Micromonas*. *Environmental Microbiology Reports* **1**, 114–123 (2009).
227. Derelle, E. *et al.* Diversity of viruses infecting the green microalga *Ostreococcus lucimarinus*. *J Virol* **89**, 5812–5821 (2015).
228. Weynberg, K. D., Allen, M. J., Gilg, I. C., Scanlan, D. J. & Wilson, W. H. Genome sequence of *Ostreococcus tauri* virus OtV-2 throws light on the role of picoeukaryote niche separation in the ocean. *J Virol* **85**, 4520–4529 (2011).
229. Derelle, E. *et al.* Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus tauri*. *PLoS One* **3**, (2008).
230. Weynberg, K. D., Allen, M. J., Ashelford, K., Scanlan, D. J. & Wilson, W. H. From small hosts come big viruses: the complete genome of a second *Ostreococcus tauri* virus, OtV-1. *Environmental Microbiology* **11**, 2821–2839 (2009).
231. Clerissi, C., Desdevises, Y. & Grimsley, N. Prasinoviruses of the marine green alga *Ostreococcus tauri* are mainly species specific. *J Virol* **86**, 4611–4619 (2012).
232. Monier, A. *et al.* Host-derived viral transporter protein for nitrogen uptake in infected marine phytoplankton. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7489–E7498 (2017).
233. Cottrell, M. T. & Suttle, C. A. Genetic diversity of algal viruses which lyse the photosynthetic picoflagellate *Micromonas pusilla* (Prasinophyceae). *Appl Environ Microbiol* **61**, 3088–3091 (1995).
234. Zingone, A., Sarno, D. & Forlani, G. Seasonal dynamics in the abundance of *Micromonas pusilla* (Prasinophyceae) and its viruses in the Gulf of Naples (Mediterranean Sea). *Journal of Plankton Research* **21**, 2143–2159 (1999).
235. Clerissi, C. *et al.* Prasinovirus distribution in the Northwest Mediterranean Sea is affected by the environment and particularly by phosphate availability. *Virology* **466–467**, 146–157 (2014).
236. Weynberg, K. *et al.* Marine Prasinoviruses and their tiny plankton hosts: A review. *Viruses* **9**, 43 (2017).
237. Waters, R. E. & Chan, A. T. *Micromonas pusilla* Virus: the virus growth cycle and associated physiological events within the host cells; host range mutation. *Journal of General Virology*, **63**, 199–206 (1982).
238. Abrahão, J. S. *et al.* *Acanthamoeba polyphaga mimivirus* and other giant viruses: an open field to outstanding discoveries. *Virol J* **11**, 1–12 (2014).
239. Arantes, T. S. *et al.* The large Marseillevirus explores different entry pathways by forming giant infectious vesicles. *Journal of Virology* **90**, 5246–5255 (2016).

240. Wang, S. *et al.* Entry of a novel marine DNA virus, singapore grouper Iridovirus, into host cells occurs via clathrin-mediated endocytosis and macropinocytosis in a pH-dependent manner. *Journal of Virology* **88**, 13047–13063 (2014).
241. Alcamí, A., Carrascosa, A. L. & Viñuela, E. The entry of *African swine fever virus* into Vero cells. *Virology* **171**, 68–75 (1989).
242. Galindo, I. & Alonso, C. *African Swine Fever Virus: A Review*. *Viruses* **9**, 103 (2017).
243. Goetschius, D. J., Parrish, C. R. & Hafenstein, S. Asymmetry in icosahedral viruses. *Curr Opin Virol* **36**, 67–73 (2019).
244. Van Etten, J. L., Graves, M. V., Müller, D. G., Boland, W. & Delaroque, N. Phycodnaviridae– large DNA algal viruses. *Arch. Virol.* **147**, 1479–1516 (2002).
245. Smith, G. L. & Law, M. The exit of *Vaccinia virus* from infected cells. *Virus Research* **106**, 189–197 (2004).
246. Braunwald, J., Nonnenmacher, H. & Tripiet-Darcy, F. 1985. Ultrastructural and biochemical study of *Frog Virus 3* uptake by BHK-21 Cells. *Journal of General Virology* **66**, 283–293 (1985).
247. Chinchar, V. G. & Waltzek, T. B. Ranaviruses: Not Just for Frogs. *PLOS Pathogens* **10**, e1003850 (2014).
248. Liu, Y. *et al.* Visualization of assembly intermediates and budding vacuoles of Singapore grouper Iridovirus in grouper embryonic cells. *Sci Rep* **6**, 18696 (2016).
249. Zimmerman, A. E. *et al.* Closely related viruses of the marine picoeukaryotic alga *Ostreococcus lucimarinus* exhibit different ecological strategies. *Environ Microbiol* **21**, 2148–2170 (2019).
250. Edwards, K. F. & Steward, G. F. Host traits drive viral life histories across phytoplankton viruses. **16** (2018).
251. Record, N. R., Talmy, D. & Våge, S. Quantifying tradeoffs for marine viruses. *Frontiers in Marine Science* **3**, 251 (2016).
252. Schroeder, D. C., Oke, J., Malin, G. & Wilson, W. H. *Coccolithovirus (Phycodnaviridae)*: Characterisation of a new large dsDNA algal virus that infects *Emiliana huxleyi*. *Arch. Virol.* **147**, 1685–1698 (2002).
253. Van Etten, J. L., Agarkova, I. V. & Dunigan, D. D. Chloroviruses. *Viruses* **12**, 20 (2020).
254. Suttle, C. A. The significance of viruses to mortality in aquatic microbial communities. *Microb Ecol* **28**, 237–243 (1994).
255. Wilhelm, S. W. & Suttle, C. A. Viruses and nutrient cycles in the Sea: Viruses play critical roles in the structure and function of aquatic food webs. *BioScience* **49**, 781–788 (1999).
256. Wommack, K. E. & Colwell, R. R. Virioplankton: Viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* **64**, 69–114 (2000).
257. Weinbauer, M. G. Ecology of prokaryotic viruses. *FEMS Microbiol Rev* **28**, 127–181 (2004).
258. Middelboe, M. & Brussaard, C. P. D. Marine viruses: Key players in marine ecosystems. *Viruses* **9**, 302 (2017).
259. Omand, M. M., Govindarajan, R., He, J. & Mahadevan, A. Sinking flux of particulate organic matter in the oceans: Sensitivity to particle characteristics. *Sci Rep* **10**, 5582 (2020).
260. Poulton, A. J. Shunt or shuttle. *Nat. Geosci.* **14**, 181–183 (2021).

261. Breitbart, M., Bonnain, C., Malki, K. & Sawaya, N. A. Phage puppet masters of the marine microbial realm. *Nat Microbiol* **3**, 754–766 (2018).
262. Malin, G., Turner, S., Liss, P., Holligan, P. & Harbour, D. Dimethylsulphide and dimethylsulphonioacetate in the Northeast Atlantic during the summer coccolithophore bloom. *Deep Sea Research Part I: Oceanographic Research Papers* **40**, 1487–1508 (1993).
263. Matrai, P. A. & Keller, M. D. Dimethylsulfide in a large-scale coccolithophore bloom in the Gulf of Maine. *Continental Shelf Research* **13**, 831–843 (1993).
264. Hill et al., R. W. Virus-mediated total release of dimethylsulfoniopropionate from marine phytoplankton: a potential climate process. *7* (1998).
265. Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature* **326**, 655–661 (1987).
266. Bratbak, G. et al. Viral activity in relation to *Emiliania huxleyi* blooms: a mechanism of DMSP release? *Mar. Ecol. Prog. Ser.* **128**, 133–142 (1995).
267. Winter, C., Bouvier, T., Weinbauer, M. G. & Thingstad, T. F. Trade-Offs between competition and defense specialists among unicellular planktonic organisms: the “Killing the Winner” hypothesis revisited. *Microbiol Mol Biol Rev* **74**, 42–57 (2010).
268. Thingstad, T. F. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography* **45**, 1320–1328 (2000).
269. Bratbak, G., Egge, J. & Heldal, M. Viral mortality of the marine alga *Emiliania huxleyi* (Haptophyceae) and termination of algal blooms. *Marine Ecology Progress Series* **93**, 39–48 (1993).
270. Brussaard, C., Kempers, R., Kop, A., Riegman, R. & Heldal, M. Virus-like particles in a summer bloom of *Emiliania huxleyi* in the North Sea. *Aquat. Microb. Ecol.* **10**, 105–113 (1996).
271. Brussaard, C. P. D. Viral Control of Phytoplankton Populations—a Review 1. *Journal of Eukaryotic Microbiology* **51**, 125–138 (2004).
272. Hutchinson, G. E. The Paradox of the plankton. (1961)
273. Waterbury, J. B. & Valois, F. W. Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Appl Environ Microbiol* **59**, 3393–3399 (1993).
274. Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**, 1047–1051 (2003).
275. Tarutani, K., Nagasaki, K. & Yamaguchi, M. Viral impacts on total abundance and clonal composition of the harmful bloom-forming phytoplankton *Heterosigma akashiwo*. *Appl Environ Microbiol* **66**, 4916–4920 (2000).
276. Page, R. D. M. Tangled Trees: Phylogeny, Cospeciation, and Coevolution. (University of Chicago Press, 2003).
277. Woolhouse, M. E. J., Webster, J. P., Domingo, E., Charlesworth, B. & Levin, B. R. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat Genet* **32**, 569–577 (2002).

278. Gandon, S., Buckling, A., Decaestecker, E. & Day, T. Host–parasite coevolution and patterns of adaptation across time and space. *Journal of Evolutionary Biology* **21**, 1861–1866 (2008).
279. Martiny, J. B. H., Riemann, L., Marston, M. F. & Middelboe, M. Antagonistic coevolution of marine planktonic viruses and their hosts. *Ann Rev Mar Sci* **6**, 393–414 (2014).
280. Agrawal, A. F. & Lively, C. M. Modelling infection as a two-step process combining gene-for-gene and matching-allele genetics. *Proceedings of the Royal Society of London. Series B: Biological Sciences* (2003).
281. Grosberg, R. K. & Hart, M. W. Mate selection and the evolution of highly polymorphic self/nonself recognition genes. *Science* **289**, 2111–2114 (2000).
282. Weitz, J. S. *et al.* Phage–bacteria infection networks. *Trends in Microbiology* **21**, 82–91 (2013).
283. Frank, S. A. Models of plant–pathogen coevolution. *Trends in Genetics* **8**, 213–219 (1992).
284. Flor, H. H. Current Status of the Gene-For-Gene Concept. *Annual Review of Phytopathology* **9**, 275–296 (1971).
285. Thompson, J. N. & Burdon, J. J. Gene-for-gene coevolution between plants and parasites. *Nature* **360**, 121–125 (1992).
286. Poullain, V., Gandon, S., Brockhurst, M. A., Buckling, A. & Hochberg, M. E. The evolution of specificity in evolving and coevolving antagonistic interactions between a bacteria and its phage. *Evolution* **62**, 1–11 (2008).
287. Flores, C. O., Meyer, J. R., Valverde, S., Farr, L. & Weitz, J. S. Statistical structure of host–phage interactions. *PNAS* **108**, E288–E297 (2011).
288. Van Valen, L. Body size and numbers of plants and animals. *Evolution* **27**, 27–35 (1973).
289. Frickel, J., Theodosiou, L. & Becks, L. Rapid evolution of hosts begets species diversity at the cost of intraspecific diversity. *Proc Natl Acad Sci U S A* **114**, 11193–11198 (2017).
290. Frickel, J., Sieber, M. & Becks, L. Eco-evolutionary dynamics in a coevolving host–virus system. *Ecology Letters* **19**, 450–459 (2016).
291. Thyrhaug, R., Larsen, A., Thingstad, T. & Bratbak, G. Stable coexistence in marine algal host-virus systems. *Mar. Ecol. Prog. Ser.* **254**, 27–35 (2003).
292. Tomaru, Y., Mizumoto, H. & Nagasaki, K. Virus resistance in the toxic bloom-forming dinoflagellate *Heterocapsa circularisquama* to single-stranded RNA virus infection. *Environmental Microbiology* **11**, 2915–2923 (2009).
293. Thomas, R. *et al.* Acquisition and maintenance of resistance to viruses in eukaryotic phytoplankton populations. *Environmental Microbiology* **13**, 1412–1420 (2011).
294. Luria, S. E. & Delbrück, M. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**, 491–511 (1943).
295. Brenner, S. & Hinshelwood, C. N. The adsorption of bacteriophages by sensitive and resistant cells of *Escherichia coli* strain B. *Proceedings of the Royal Society of London. Series B - Biological Sciences* **144**, 93–99 (1955).
296. Bohannan, B. J. M., Kerr, B., Jessup, C. M., Hughes, J. B. & Sandvik, G. Trade-offs and coexistence in microbial microcosms. *Antonie Van Leeuwenhoek* **81**, 107–115 (2002).
297. Avrani, S., Wurtzel, O., Sharon, I., Sorek, R. & Lindell, D. Genomic island variability facilitates *Prochlorococcus*–virus coexistence. *Nature* **474**, 604–608 (2011).

298. Coleman, M. L. *et al.* Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**, 1768–1770 (2006).
299. Williams, R. J. Restriction endonuclease. *Mol Biotechnol* **23**, 225–243 (2003).
300. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–170 (2010).
301. Sorek, R., Kunin, V. & Hugenholz, P. CRISPR — a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6**, 181–186 (2008).
302. Wietz, M., Millán-Aguiñaga, N. & Jensen, P. R. CRISPR-Cas systems in the marine actinomycete *Salinispora*: linkages with phage defense, microdiversity and biogeography. *BMC Genomics* **15**, 1–10 (2014).
303. Cai, F., Axen, S. D. & Kerfeld, C. A. Evidence for the widespread distribution of CRISPR-Cas system in the Phylum Cyanobacteria. *RNA Biology* **10**, 687–693 (2013).
304. Seed, K. D., Lazinski, D. W., Calderwood, S. B. & Camilli, A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**, 489–491 (2013).
305. Levasseur, A. *et al.* MIMIVIRE is a defence system in mimivirus that confers resistance to virophage. *Nature* **531**, 249–252 (2016).
306. Waldor, M. K. & Mekalanos, J. J. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**, 1910–1914 (1996).
307. Wang, L. *et al.* Endogenous viral elements in algal genomes. *Acta Oceanol. Sin.* **33**, 102–107 (2014).
308. Yau, S. *et al.* Rapidity of genomic adaptations to *Prasinovirus* infection in a marine microalga. *Viruses* **10**, (2018).
309. Derelle, E., Yau, S., Moreau, H. & Grimsley, N. H. *Prasinovirus* attack of *Ostreococcus* is furtive by day but savage by night. *J Virol* **92**, (2018).
310. Yau, S. *et al.* A Viral immunity chromosome in the marine picoeukaryote, *Ostreococcus tauri*. *PLOS Pathogens* **12**, e1005965 (2016).
311. Bhattacharya, A. *et al.* Coupling chemical mutagenesis to next generation sequencing for the identification of drug resistance mutations in *Leishmania*. *Nature Communications* **10**, 5627 (2019).
312. Mosey, M., Douchi, D., Knoshaug, E. P. & Laurens, L. M. L. Methodological review of genetic engineering approaches for non-model algae. *Algal Research* **54**, 102221 (2021).
313. Cullen, B. R. RNA interference: antiviral defense and genetic tool. *Nat Immunol* **3**, 597–599 (2002).
314. Alsford, S. *et al.* High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Res* **21**, 915–924 (2011).
315. Specht, E., Miyake-Stoner, S. & Mayfield, S. Micro-algae come of age as a platform for recombinant protein production. *Biotechnol Lett* **32**, 1373–1383 (2010).
316. Moresco, E. M. Y., Li, X. & Beutler, B. Going forward with genetics: Recent technological advances and forward genetics in mice. *The American Journal of Pathology* **182**, 1462–1473 (2013).
317. Faktorová, D. *et al.* Genetic tool development in marine protists: emerging model organisms for experimental cell biology. *Nature Methods* **17**, 481–494 (2020).

318. Pazour, G. J. & Witman, G. B. Forward and Reverse Genetic Analysis of Microtubule Motors in *Chlamydomonas*. *Methods* **22**, 285–298 (2000).
319. Kumar, G. *et al.* Bioengineering of microalgae: Recent advances, perspectives, and regulatory Challenges for Industrial Application. *Front. Bioeng. Biotechnol.* **8**, (2020).
320. Van Ooijen, G., Knox, K., Kis, K., Bouget, F.-Y. & Millar, A. J. Genomic transformation of the picoeukaryote *Ostreococcus tauri*. *J Vis Exp* e4074 (2012).
321. Lozano, J.-C. *et al.* Efficient gene targeting and removal of foreign DNA by homologous recombination in the picoeukaryote *Ostreococcus*. *The Plant Journal* **78**, 1073–1083 (2014).
322. Sanchez, F. *et al.* Simplified transformation of *Ostreococcus tauri* using polyethylene glycol. *Genes* **10**, 399 (2019).
323. Paszkowski, J. *et al.* Direct gene transfer to plants. *EMBO J* **3**, 2717–2722 (1984).
324. Ito, H., Fukuda, Y., Murata, K. & Kimura, A. Transformation of intact yeast cells treated with alkali cations. *J Bacteriol* **153**, 163–168 (1983).
325. Bibb, M. J., Ward, J. M. & Hopwood, D. A. Transformation of plasmid DNA into *Streptomyces* at high frequency. *Nature* **274**, 398–400 (1978).
326. Klebe, R. J., Harriss, J. V., Sharp, Z. D. & Douglas, M. G. A general method for polyethylene-glycol-induced genetic transformation of bacteria and yeast. *Gene* **25**, 333–341 (1983).
327. Abel, S. & Theologis, A. Transient transformation of *Arabidopsis* leaf protoplasts: a versatile experimental system to study gene expression. *The Plant Journal* **5**, 421–427 (1994).
328. Rasmussen, J. O. & Rasmussen, O. S. PEG mediated DNA uptake and transient GUS expression in carrot, rapeseed and soybean protoplasts. *Plant Science* **89**, 199–207 (1993).
329. Liu, Y.-C. & Vidali, L. Efficient Polyethylene Glycol (PEG) Mediated Transformation of the Moss *Physcomitrella patens*. *J Vis Exp* (2011).
330. Ohnuma, M., Yokoyama, T., Inouye, T., Sekine, Y. & Tanaka, K. Polyethylene glycol (PEG)-mediated transient gene expression in a red alga, *Cyanidioschyzon merolae* 10D. *Plant Cell Physiol* **49**, 117–120 (2008).
331. Ohnuma, M. *et al.* Optimization of polyethylene glycol (PEG)-mediated DNA introduction conditions for transient gene expression in the unicellular red alga *Cyanidioschyzon merolae*. *J. Gen. Appl. Microbiol.* **60**, 156–159 (2014).
332. Kathir, P. *et al.* Molecular Map of the *Chlamydomonas reinhardtii* Nuclear Genome. *Eukaryot Cell* **2**, 362–379 (2003).
333. Tam, L. W. & Lefebvre, P. A. Cloning of flagellar genes in *Chlamydomonas reinhardtii* by DNA insertional mutagenesis. *Genetics* **135**, 375–384 (1993).
334. Dent, R. M., Haglund, C. M., Chin, B. L., Kobayashi, M. C. & Niyogi, K. K. Functional genomics of eukaryotic photosynthesis using insertional mutagenesis of *Chlamydomonas reinhardtii*. *Plant Physiology* **137**, 545–556 (2005).
335. González-Ballester, D., Montaigu, A. de, Higuera, J. J., Galván, A. & Fernández, E. Functional genomics of the regulation of the nitrate assimilation pathway in *Chlamydomonas*. *Plant Physiology* **137**, 522–533 (2005).
336. Meslet-Cladière, L. & Vallon, O. A new method to identify flanking sequence tags in *Chlamydomonas* using 3'-RACE. *Plant Methods* **8**, 21 (2012).

337. González-Ballester, D., de Montaigu, A., Galván, A. & Fernández, E. Restriction enzyme site-directed amplification PCR: A tool to identify regions flanking a marker DNA. *Analytical Biochemistry* **340**, 330–335 (2005).
338. Pollock, S. V. *et al.* A robust protocol for efficient generation, and genomic characterization of insertional mutants of *Chlamydomonas reinhardtii*. *Plant Methods* **13**, 1–9 (2017).
339. Plecenikova, A., Slaninova, M. & Riha, K. Characterization of DNA repair deficient strains of *Chlamydomonas reinhardtii* generated by insertional mutagenesis. *PLoS One* **9**, (2014).
340. Zhang, R. *et al.* High-throughput genotyping of green algal mutants reveals random distribution of mutagenic insertion sites and endonucleolytic cleavage of transforming DNA. *The Plant Cell* **26**, 1398–1409 (2014).
341. Schouten, H. J. *et al.* Re-sequencing transgenic plants revealed rearrangements at T-DNA inserts, and integration of a short T-DNA fragment, but no increase of small mutations elsewhere. *Plant Cell Rep* **36**, 493–504 (2017).
342. Jupe, F. *et al.* The complex architecture and epigenomic impact of plant T-DNA insertions. *PLoS Genet* **15**, (2019).
343. Liu, L. *et al.* Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology* vol. 2012 e251364 (2012).
344. Guttikonda, S. K. *et al.* Molecular characterization of transgenic events using next generation sequencing approach. *PLOS ONE* **11**, e0149515 (2016).
345. Qin, Y. *et al.* Flanking sequence and copy-number analysis of transformation events by integrating next-generation sequencing technology with Southern Blot hybridization. *Plant Breed. Biotech.* **5**, 269–281 (2017).
346. Abel, H. J. & Duncavage, E. J. Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. *Cancer Genet* **206**, 432–440 (2013).
347. Goodwin, L. O. *et al.* Large-scale discovery of mouse transgenic integration sites reveals frequent structural variation and insertional mutagenesis. *Genome Res.* **29**, 494–505 (2019).
348. Winnepeninckx, B., Backeljau, T. & De Wachter, R. Extraction of high molecular weight DNA from molluscs. *Trends Genet* **9**, 407 (1993).
349. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* **37**, 540–546 (2019).
350. Zimin, A. V. *et al.* The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669–2677 (2013).
351. Zimin, A. V. *et al.* Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res.* **27**, 787–792 (2017).
352. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
353. Yang, L. *et al.* Characterization of GM events by insert knowledge adapted re-sequencing approaches. *Scientific Reports* **3**, 2839 (2013).
354. Piganeau, G., Grimsley, N. & Moreau, H. Genome diversity in the smallest marine photosynthetic eukaryotes. *Research in Microbiology* **162**, 570–577 (2011).

355. Jaehning, J. A. The Paf1 complex: platform or player in RNA polymerase II transcription? *Biochim Biophys Acta* **1799**, 379–388 (2010).
356. Kovács, E. & Keresztes, Á. Effect of gamma and UV-B/C radiation on plant cells. *Micron* **33**, 199–210 (2002).
357. Zhou, B.-B. S. & Elledge, S. J. The DNA damage response: putting checkpoints in perspective. *Nature* **408**, 433–439 (2000).
358. Sancar, A., Lindsey-Boltz, L. A., Unsal-Kaçmaz, K. & Linn, S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu Rev Biochem* **73**, 39–85 (2004).
359. Onozawa, M. *et al.* Repair of DNA double-strand breaks by templated nucleotide sequence insertions derived from distant regions of the genome. *PNAS* **111**, 7729–7734 (2014).
360. Pâques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* **63**, 349–404 (1999).
361. Shrivastav, M., De Haro, L. P. & Nickoloff, J. A. Regulation of DNA double-strand break repair pathway choice. *Cell Research* **18**, 134–147 (2008).
362. Nagaria, P., Robert, C. & Rassool, F. V. DNA double-strand break response in stem cells: mechanisms to maintain genomic integrity. *Biochim Biophys Acta* **1830**, 2345–2353 (2013).
363. Mayerhofer, R. *et al.* T-DNA integration: a mode of illegitimate recombination in plants. *EMBO J* **10**, 697–704 (1991).
364. Cenkci, B., Petersen, J. L. & Small, G. D. REX1, a Novel Gene Required for DNA Repair. *Journal of Biological Chemistry* **278**, 22574–22577 (2003).
365. Sarkar, N., Lemaire, S., Wu-Scharf, D., Issakidis-Bourguet, E. & Cerutti, H. Functional specialization of *Chlamydomonas reinhardtii* cytosolic thioredoxin h1 in the response to alkylation-induced DNA damage. *Eukaryot Cell* **4**, 262–273 (2005).
366. Boiteux, S. & Guillet, M. Abasic sites in DNA: repair and biological consequences in *Saccharomyces cerevisiae*. *DNA Repair* **3**, 1–12 (2004).
367. Kelley, M. R., Kow, Y. W. & Wilson, D. M. Disparity between DNA base excision repair in yeast and mammals: translational implications. *Cancer Res* **63**, 549–554 (2003).
368. Heitzberg, F. *et al.* The Rad17 homologue of Arabidopsis is involved in the regulation of DNA damage repair and homologous recombination. *The Plant Journal* **38**, 954–968 (2004).
369. Van Kregten, M. *et al.* T-DNA integration in plants results from polymerase- $\theta$ -mediated DNA repair. *Nature Plants* **2**, 1–6 (2016).
370. Gallego, M. E., Bleuyard, J.-Y., Daoudal-Cotterell, S., Jallut, N. & White, C. I. Ku80 plays a role in non-homologous recombination but is not required for T-DNA integration in Arabidopsis. *Plant J* **35**, 557–565 (2003).
371. Friesner, J. & Britt, A. B. Ku80- and DNA ligase IV-deficient plants are sensitive to ionizing radiation and defective in T-DNA integration. *Plant J* **34**, 427–440 (2003).
372. Van Attikum, H., Bundock, P. & Hooykaas, P. J. Non-homologous end-joining proteins are required for Agrobacterium T-DNA integration. *EMBO J* **20**, 6550–6558 (2001).
373. Daley, J. M., Palmbo, P. L., Wu, D. & Wilson, T. E. Nonhomologous end joining in yeast. *Annu. Rev. Genet.* **39**, 431–451 (2005).

374. Van Bel, M. *et al.* PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Research* **46**, D1190–D1196 (2018).
375. Yamazaki, M., Ohnishi, S. & Ito, T. Osmoelastic coupling in biological structures: decrease in membrane fluidity and osmophobic association of phospholipid vesicles in response to osmotic stress. *Biochemistry* **28**, 3710–3715 (1989).
376. Kawai, S. *et al.* Transcriptional and metabolic response in yeast *Saccharomyces cerevisiae* cells during polyethylene glycol-dependent transformation. *Journal of Basic Microbiology* **49**, 73–81 (2009).
377. D'souza, A. A. & Shegokar, R. Polyethylene glycol (PEG): a versatile polymer for pharmaceutical applications. *Expert Opin Drug Deliv* **13**, 1257–1275 (2016).
378. Hansen, null & Wright, null. Recent advances in the transformation of plants. *Trends Plant Sci* **4**, 226–231 (1999).
379. Does, M. P., Dekker, B. M., de Groot, M. J. & Offringa, R. A quick method to estimate the T-DNA copy number in transgenic plants at an early stage after transformation, using inverse PCR. *Plant Molecular Biology* **17**, 151–153 (1991).
380. Kumar, S. V., Misquitta, R. W., Reddy, V. S., Rao, B. J. & Rajam, M. V. Genetic transformation of the green alga—*Chlamydomonas reinhardtii* by *Agrobacterium tumefaciens*. *Plant Science* **166**, 731–738 (2004).
381. Gan, L., Ladinsky, M. S. & Jensen, G. J. Chromatin in a marine picoeukaryote is a disordered assemblage of nucleosomes. *Chromosoma* **122**, 377–386 (2013).
382. Radakovits, R. *et al.* Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. *Nature Communications* **3**, 686 (2012).
383. Scranton, M. A., Ostrand, J. T., Fields, F. J. & Mayfield, S. P. *Chlamydomonas* as a model for biofuels and bio-products production. *Plant J* **82**, 523–531 (2015).
384. Das, S. K., Sathish, A. & Stanley, J. Production of biofuel and bioplastic from *Chlorella pyrenoidosa*. *Materials Today: Proceedings* **5**, 16774–16781 (2018).
385. Smallwood, C. R. *et al.* *Ostreococcus tauri* is a high-lipid content green algae that extrudes clustered lipid droplets. *bioRxiv* 249052 (2018).
386. Goff, M. L. *et al.* Protective action of *Ostreococcus tauri* and *Phaeodactylum tricornutum* extracts towards Benzo[a]Pyrene-Induced Cytotoxicity in Endothelial Cells. *Mar Drugs* **18**, (2019).
387. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* **348**, (2015).
388. Leconte, J. *et al.* Genome resolved biogeography of Mamiellales. *Genes* **11**, 66 (2020).
389. Demory, D. *et al.* Picoeukaryotes of the *Micromonas* genus: sentinels of a warming ocean. *The ISME Journal* **13**, 132–146 (2019).
390. Schaum, C.-E., Rost, B. & Collins, S. Environmental stability affects phenotypic evolution in a globally distributed marine picoplankton. *The ISME Journal* **10**, 75–84 (2016).
391. Brussaard, C. P. D. *et al.* Global-scale processes with a nanoscale drive: the role of marine viruses. *ISME J* **2**, 575–578 (2008).
392. Paterson, S. *et al.* Antagonistic coevolution accelerates molecular evolution. *Nature* **464**, 275–278 (2010).

393. Frada, M., Probert, I., Allen, M. J., Wilson, W. H. & Vargas, C. de. The “Cheshire Cat” escape strategy of the coccolithophore *Emiliania huxleyi* in response to viral infection. *PNAS* **105**, 15944–15949 (2008).
394. Bohannan, B. j. m. & Lenski, R. e. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. *Ecology Letters* **3**, 362–377 (2000).
395. Golais, F., Holly, J. & Vítkovská, J. Coevolution of bacteria and their viruses. *Folia Microbiol* **58**, 177–186 (2013).
396. Bernheim, A. & Sorek, R. The pan-immune system of bacteria: antiviral defence as a community resource. *Nat Rev Microbiol* **18**, 113–119 (2020).
397. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
398. Thomy, J. *et al.* Combining nanopore and illumina sequencing permits detailed analysis of insertion mutations and structural variations produced by PEG-mediated transformation in *Ostreococcus tauri*. *Cells* **10**, (2021).
399. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* **8**, 973–982 (2007).
400. Gil, M. J., Coego, A., Mauch-Mani, B., Jordá, L. & Vera, P. The *Arabidopsis* *csb3* mutant reveals a regulatory link between salicylic acid-mediated disease resistance and the methyl-erythritol 4-phosphate pathway. *Plant J* **44**, 155–166 (2005).
401. Malik, H. S. & Eickbush, T. H. Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. *Genome Res.* **11**, 1187–1197 (2001).
402. Maumus, F. *et al.* Potential impact of stress activated retrotransposons on genome evolution in a marine diatom. *BMC Genomics* **10**, 624 (2009).
403. Lescot, M. *et al.* Reverse transcriptase genes are highly abundant and transcriptionally active in marine plankton assemblages. *ISME J* **10**, 1134–1146 (2016).
404. Grandbastien, M.-A. Activation of plant retrotransposons under stress conditions. *Trends in Plant Science* **3**, 181–187 (1998).
405. Grandbastien, M.-A. LTR retrotransposons, handy hitchhikers of plant regulation and stress response. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1849**, 403–416 (2015).
406. Kumar, A. & Bennetzen, J. L. Plant Retrotransposons. *Annu. Rev. Genet.* **33**, 479–532 (1999).
407. Sultana, T., Zamborlini, A., Cristofari, G. & Lesage, P. Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat Rev Genet* **18**, 292–308 (2017).
408. Bejerano, G. *et al.* A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **441**, 87–90 (2006).
409. Bourque, G. *et al.* Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* **18**, 1752–1762 (2008).
410. Jacques, P.-É., Jeyakani, J. & Bourque, G. The Majority of primate-specific regulatory sequences are derived from transposable elements. *PLOS Genetics* **9**, e1003504 (2013).
411. Sundaram, V. *et al.* Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* **24**, 1963–1976 (2014).

412. Ong-Abdullah, M. *et al.* Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature* **525**, 533–537 (2015).
413. Niu, X.-M. *et al.* Transposable elements drive rapid phenotypic variation in *Capsella rubella*. *PNAS* **116**, 6908–6913 (2019).
414. Ferreira de Carvalho, J., de Jager, V., van Gurp, T. P., Wagemaker, N. C. A. M. & Verhoeven, K. J. F. Recent and dynamic transposable elements contribute to genomic divergence under asexuality. *BMC Genomics* **17**, 1–12 (2016).
415. Tabara, H. *et al.* The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**, 123–132 (1999).
416. Sijen, T. & Plasterk, R. H. A. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* **426**, 310–314 (2003).
417. Shpiz, S. *et al.* Characterization of *Drosophila* telomeric retroelement TAHRE: transcription, transpositions, and RNAi-based regulation of expression. *Mol Biol Evol* **24**, 2535–2545 (2007).
418. Kavi, H. H., Fernandez, H. R., Xie, W. & Birchler, J. A. RNA silencing in *Drosophila*. *FEBS Letters* **579**, 5940–5949 (2005).
419. Brennecke, J. *et al.* Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. *Cell* **128**, 1089–1103 (2007).
420. Slotkin, R. K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**, 272–285 (2007).
421. Lisch, D. Epigenetic regulation of transposable elements in plants. *Annu Rev Plant Biol* **60**, 43–66 (2009).
422. Ito, H. *et al.* An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* **472**, 115–119 (2011).
423. Makarevitch, I. *et al.* Transposable Elements Contribute to Activation of Maize Genes in Response to Abiotic Stress. *PLOS Genetics* **11**, e1004915 (2015).
424. Grandbastien, M. A. *et al.* The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses. *Genetica* **100**, 241–252 (1997).
425. Johns, M. A., Mottinger, J. & Freeling, M. A low copy number, copia-like transposon in maize. *EMBO J* **4**, 1093–1101 (1985).
426. Huff, J. T. & Zilberman, D. Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell* **156**, 1286–1297 (2014).
427. Castberg, T. *et al.* Microbial population dynamics and diversity during a bloom of the marine coccolithophorid *Emiliana huxleyi* (Haptophyta). *Marine Ecology Progress Series* **221**, 39–46 (2001).
428. Evans, C., Archer, S., Jacquet, S. & Wilson, W. Direct estimates of the contribution of viral lysis and microzooplankton grazing to the decline of a *Micromonas* spp. population. *Aquat. Microb. Ecol.* **30**, 207–219 (2003).
429. Gastrich, M. D. *et al.* Viruses as potential regulators of regional brown tide blooms caused by the alga, *Aureococcus anophagefferens*. *Estuaries* **27**, 112–119 (2004).
430. Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541–548 (1999).
431. Rodriguez-Valera, F. *et al.* Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* **7**, 828–836 (2009).

432. Friman, V.-P. & Buckling, A. Effects of predation on real-time host–parasite coevolutionary dynamics. *Ecology Letters* **16**, 39–46 (2013).
433. Baranowski, E., Ruiz-Jarabo, C. M. & Domingo, E. Evolution of cell recognition by viruses. *Science* (2001).
434. Hashemolhosseini, S., Holmes, Z., Mutschler, B. & Henning, U. Alterations of receptor specificities of coliphages of the T2 family. *Journal of Molecular Biology* **240**, 105–110 (1994).
435. Hashemolhosseini, S., Montag, D., Krämer, L. & Henning, U. Determinants of receptor specificity of coliphages of the T4 family. A chaperone alters the host range. *J Mol Biol* **241**, 524–533 (1994).
436. Blanc, G. *et al.* Molecular evolution of rickettsia surface antigens: evidence of positive selection. *Molecular Biology and Evolution* **22**, 2073–2083 (2005).
437. Rakhuba, D. V., Kolomiets, E. I., Dey, E. S. & Novik, G. I. Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Pol J Microbiol* **59**, 145–155 (2010).
438. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455–477 (2012).
439. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
440. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
441. Chan, P. P. & Lowe, T. M. tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol Biol* **1962**, 1–14 (2019).
442. Kans, J. *Entrez Direct: E-utilities on the unix command line. Entrez programming utilities help [Internet]* (National Center for Biotechnology Information (US), 2021).
443. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**, 238 (2019).
444. Huerta-Cepas, J. *et al.* Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Molecular Biology and Evolution* **34**, 2115–2122 (2017).
445. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: Multiple alignment of conserved genomic sequence With rearrangements. *Genome Res* **14**, 1394–1403 (2004).
446. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* **9**, 5114 (2018).
447. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009–1010 (2011).
448. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059–3066 (2002).
449. Lemoine, F. & Gascuel, O. Gotree/Goalign: toolkit and Go API to facilitate the development of phylogenetic workflows. *NAR Genomics and Bioinformatics* **3**, (2021).
450. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* **37**, 1530–1534 (2020).

451. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**, 587–589 (2017).
452. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**, 307–321 (2010).
453. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**, 518–522 (2018).
454. Finke, J. F., Winget, D. M., Chan, A. M. & Suttle, C. A. Variation in the genetic repertoire of viruses infecting *Micromonas pusilla* reflects horizontal gene transfer and links to their environmental distribution. *Viruses* **9**, (2017).
455. Vanlerberghe, G. C. & McIntosh, L. Alternative Oxidase: From gene to function. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* **48**, 703–734 (1997).
456. Maxwell, D. P., Wang, Y. & McIntosh, L. The alternative oxidase lowers mitochondrial reactive oxygen production in plant cells. *PNAS* **96**, 8271–8276 (1999).
457. Vishwakarma, A., Bashyam, L., Senthilkumaran, B., Scheibe, R. & Padmasree, K. Physiological role of AOX1a in photosynthesis and maintenance of cellular redox homeostasis under high light in *Arabidopsis thaliana*. *Plant Physiol Biochem* **81**, 44–53 (2014).
458. Weger, H. G., Guy, R. D. & Turpin, D. H. Cytochrome and alternative pathway respiration in green algae. *Plant Physiol* **93**, 356–360 (1990).
459. Weger, H. G. & Dasgupta, R. Regulation of alternative pathway respiration in *Chlamydomonas Reinhardtii* (chlorophyceae). *Journal of Phycology* **29**, 300–308 (1993).
460. Liu, Y. *et al.* Distinct roles of alternative oxidase pathway during the greening process of etiolated algae. *Sci. China Life Sci.* **64**, 816–827 (2021).
461. Pelkonen, S., Pelkonen, J. & Finne, J. Common cleavage pattern of polysialic acid by bacteriophage endosialidases of different properties and origins. *J Virol* **63**, 4409–4416 (1989).
462. Jakobsson, E., Schwarzer, D., Jokilampi, A. & Finne, J. Endosialidases: Versatile tools for the study of polysialic acid. *Top Curr Chem* **367**, 29–73 (2015).
463. Wolfe, J. B. & Kaplan, N. O. D-mannitol 1-phosphate dehydrogenase from *Escherichia coli*. *Journal of Biological Chemistry* **218**, 849–869 (1956).
464. Fridovich, I. Superoxide radical and superoxide dismutases. *Annu Rev Biochem* **64**, 97–112 (1995).
465. Cao, J. X., Teoh, M. L. T., Moon, M., McFadden, G. & Evans, D. H. Leporipoxvirus Cu-Zn superoxide dismutase homologs inhibit cellular superoxide dismutase, but are not essential for virus replication or virulence. *Virology* **296**, 125–135 (2002).
466. Teoh, M. L. T., Walasek, P. J. & Evans, D. H. Leporipoxvirus Cu,Zn-superoxide dismutase (SOD) homologs are catalytically inert decoy proteins that bind copper chaperone for SOD. *J Biol Chem* **278**, 33175–33184 (2003).
467. Tomalski, M. D., Eldridge, R. & Miller, L. K. A baculovirus homolog of a Cu/Zn superoxide dismutase gene. *Virology* **184**, 149–161 (1991).

468. Lartigue, A. *et al.* The Megavirus chilensis Cu,Zn-superoxide dismutase: the first viral structure of a typical cellular copper chaperone-independent hyperstable dimeric enzyme. *Journal of Virology* (2014).
469. Kang, M. *et al.* Chlorovirus PBCV-1 encodes an active copper-zinc superoxide dismutase. *Journal of Virology* (2014).
470. Tondera, D. *et al.* Knockdown of MTP18, a novel phosphatidylinositol 3-Kinase-dependent protein, affects mitochondrial morphology and induces apoptosis. *Journal of Biological Chemistry* **279**, 31544–31555 (2004).
471. Bölin, I., Norlander, L. & Wolf-Watz, H. Temperature-inducible outer membrane protein of *Yersinia pseudotuberculosis* and *Yersinia enterocolitica* is associated with the virulence plasmid. *Infection and Immunity* (1982) doi:10.1128/iai.37.2.506-512.1982.
472. Riess, T. *et al.* Bartonella Adhesin a mediates a proangiogenic host cell response. *Journal of Experimental Medicine* **200**, 1267–1278 (2004).
473. Bassler, J., Hernandez Alvarez, B., Hartmann, M. D. & Lupas, A. N. A domain dictionary of trimeric autotransporter adhesins. *International Journal of Medical Microbiology* **305**, 265–275 (2015).
474. Narasimhan, S. Control of glycoprotein synthesis. UDP-GlcNAc:glycopeptide beta 4-N-acetylglucosaminyltransferase III, an enzyme in hen oviduct which adds GlcNAc in beta 1-4 linkage to the beta-linked mannose of the trimannosyl core of N-glycosyl oligosaccharides. *J Biol Chem* **257**, 10235–10242 (1982).
475. Miyoshi, E. *et al.* Transfection of N-acetylglucosaminyltransferase III gene suppresses expression of hepatitis B virus in a human hepatoma Cell Line, HB611 (\*). *Journal of Biological Chemistry* **270**, 28311–28315 (1995).
476. Bhattacharyya, R., Bhaumik, M., Raju, T. S. & Stanley, P. Truncated, inactive N-acetylglucosaminyltransferase III (GlcNAc-TIII) induces neurological and other traits absent in mice that lack GlcNAc-TIII. *Journal of Biological Chemistry* **277**, 26300–26309 (2002).
477. Ihara, H. *et al.* A catalytically inactive  $\beta$ 1,4-N-acetylglucosaminyltransferase III (GnT-III) behaves as a dominant negative GnT-III inhibitor. *European Journal of Biochemistry* **269**, 193–201 (2002).
478. Xu, J., Hendrix, R. W. & Duda, R. L. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Molecular Cell* **16**, 11–21 (2004).
479. Pell, L. G. *et al.* A Conserved spiral structure for highly diverged phage tail assembly chaperones. *Journal of Molecular Biology* **425**, 2436–2449 (2013).
480. Vijayan, M. & Chandra, N. Lectins. *Current Opinion in Structural Biology* **9**, 707–714 (1999).
481. Lis, H. & Sharon, N. Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* **98**, 637–674 (1998).
482. Blasco, R., Sisler, J. R. & Moss, B. Dissociation of progeny vaccinia virus from the cell membrane is regulated by a viral envelope glycoprotein: effect of a point mutation in the lectin homology domain of the A34R gene. *Journal of Virology* (1993).
483. Voigt, S., Sandford, G. R., Ding, L. & Burns, W. H. Identification and characterization of a spliced C-type lectin-like gene encoded by rat Cytomegalovirus. *Journal of Virology* (2001).

484. Galindo, I., Almazán, F., Bustos, M. J., Viñuela, E. & Carrascosa, A. L. *African Swine Fever Virus* EP153R open reading frame encodes a glycoprotein involved in the hemadsorption of infected cells. *Virology* **266**, 340–351 (2000).
485. Game, J. C. DNA double-strand breaks and the RAD50-RAD57 genes in *Saccharomyces*. *Semin Cancer Biol* **4**, 73–83 (1993).
486. Petrini, J. H., Bressan, D. A. & Yao, M. S. The RAD52 epistasis group in mammalian double strand break repair. *Semin Immunol* **9**, 181–188 (1997).
487. Comeau, A. M., Bertrand, C., Letarov, A., Tétart, F. & Krisch, H. M. Modular architecture of the T4 phage superfamily: A conserved core genome and a plastic periphery. *Virology* **362**, 384–396 (2007).
488. Mehmel, M. *et al.* Possible function for virus encoded K<sup>+</sup> channel Kcv in the replication of chlorella virus PBCV-1. *FEBS Lett* **552**, 7–11 (2003).
489. Frohns, F. *et al.* Potassium ion channels of chlorella ciruses cause rapid depolarization of host cells during infection. *J Virol* **80**, 2437–2444 (2006).
490. Neupärtl, M. *et al.* Chlorella viruses evoke a rapid release of K<sup>+</sup> from host cells during the early phase of infection. *Virology* **372**, 340–348 (2008).
491. Siotto, F. *et al.* Viruses infecting marine picoplankton encode functional potassium ion channels. *Virology* **466–467**, 103–111 (2014).
492. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567–580 (2001).
493. Murry, C. R. *et al.* Genetic diversity of potassium ion channel proteins encoded by chloroviruses that infect *Chlorella heliozoae*. *Viruses* **12**, 678 (2020).
494. Hamacher, K. *et al.* Phycodnavirus potassium ion channel proteins question the virus molecular piracy hypothesis. *PLOS ONE* **7**, e38826 (2012).
495. Thiel, G., Moroni, A., Blanc, G. & Van Etten, J. L. Potassium ion channels: Could they have evolved from viruses? *Plant Physiology* **162**, 1215–1224 (2013).
496. Janzen, D. H. When is it coevolution? *Evolution* **34**, 611–612 (1980).
497. Stenseth, N. Chr. & Smith, J. M. Coevolution in ecosystems: Red queen evolution or stasis? *Evolution* **38**, 870–880 (1984).
498. Van Valen, L. Molecular evolution as predicted by natural selection. *J Mol Evol* **3**, 89–101 (1974).
499. Prehm, P., Jann, B., Jann, K., Schmidt, G. & Stirn, S. On a bacteriophage T3 and T4 receptor region within the cell wall lipopolysaccharide of escherichia coli B. *Journal of Molecular Biology* **101**, 277–281 (1976).
500. Marin, M., Tailor, C. S., Nouri, A., Kozak, S. L. & Kabat, D. Polymorphisms of the cell surface receptor control mouse susceptibilities to xenotropic and polytropic leukemia viruses. *J Virol* **73**, 9362–9368 (1999).
501. Cuomo, C. A. *et al.* The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* (2007).
502. Middelboe, M., Holmfeldt, K., Riemann, L., Nybroe, O. & Haaber, J. Bacteriophages drive strain diversification in a marine *Flavobacterium*: implications for phage resistance and physiological properties. *Environmental Microbiology* **11**, 1971–1982 (2009).

503. Malykh, Y. N., Shaw, L. & Schauer, R. The role of CMP-N-acetylneuraminic acid hydroxylase in determining the level of N-glycolylneuraminic acid in porcine tissues. *Glycoconj J* **15**, 885–893 (1998).
504. Tétart, F., Repoila, F., Monod, C. & Krisch, H. M. Bacteriophage T4 host range is expanded by duplications of a small domain of the tail fiber adhesin. *Journal of Molecular Biology* **258**, 726–731 (1996).
505. Tétart, F., Desplats, C. & Krisch, H. M. Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. *Journal of Molecular Biology* **282**, 543–556 (1998).
506. Dunne, M. *et al.* Salmonella phage S16 tail fiber adhesin features a rare polyglycine rich domain for host recognition. *Structure* **26**, 1573-1582.e4 (2018).
507. Koonin, E. V. & Dolja, V. V. A virocentric perspective on the evolution of life. *Curr Opin Virol* **3**, 546–557 (2013).
508. Raoult, D. & Forterre, P. Redefining viruses: lessons from Mimivirus. *Nat Rev Microbiol* **6**, 315–319 (2008).
509. Koonin, E. V. & Yutin, N. Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. *Intervirology* **53**, 284–292 (2010).
510. Sun, T.-W. *et al.* Host range and coding potential of eukaryotic giant viruses. *Viruses* **12**, 1337 (2020).
511. Klose, T. *et al.* Structure of faustovirus, a large dsDNA virus. *PNAS* **113**, 6206–6211 (2016).
512. Wang, N. *et al.* Architecture of *African swine fever virus* and implications for viral assembly. *Science* **366**, 640–644 (2019).
513. Liu, S. *et al.* Cryo-EM structure of the *African swine fever virus*. *Cell Host & Microbe* **26**, 836-843.e3 (2019).
514. Andrés, G., Charro, D., Matamoros, T., Dillard, R. S. & Abrescia, N. G. A. The cryo-EM structure of *African swine fever virus* unravels a unique architecture comprising two icosahedral protein capsids and two lipoprotein membranes. *Journal of Biological Chemistry* **295**, 1–12 (2020).
515. Yan, X., Chipman, P. R., Castberg, T., Bratbak, G. & Baker, T. S. The Marine algal virus PpV01 has an icosahedral capsid with T=219 quasisymmetry. *Journal of Virology* **79**, 9236–9243 (2005).
516. Huiskonen, J. T. & Butcher, S. J. Membrane-containing viruses with icosahedrally symmetric capsids. *Current Opinion in Structural Biology* **17**, 229–236 (2007).
517. Peralta, B. *et al.* Mechanism of membranous tunnelling nanotube formation in viral genome delivery. *PLOS Biology* **11**, e1001667 (2013).
518. Meints, R. H., Burbank, D. E., Van Etten, J. L. & Lampion, D. T. A. Properties of the chlorella receptor for the virus PBCV-11. *Virology* **164**, 15–21 (1988).
519. Johnson, J. E. & Chiu, W. DNA packaging and delivery machines in tailed bacteriophages. *Current Opinion in Structural Biology* **17**, 237–243 (2007).
520. Mayer, J. A. Isolation and ultrastructural study of a lytic virus in the small marine phytoflagellate, *Micromonas pusilla* (prasinophyceae). 175 (1978).
521. Order - Caudovirales. in *Virus Taxonomy* (eds. King, A. M. Q., Adams, M. J., Carstens, E. B. & Lefkowitz, E. J.) 39–45 (2012).

522. Casey, A., Jordan, K., Neve, H., Coffey, A. & McAuliffe, O. A tail of two phages: genomic and functional analysis of *Listeria monocytogenes* phages vB\_LmoS\_188 and vB\_LmoS\_293 reveal the receptor-binding proteins involved in host specificity. *Frontiers in Microbiology* **6**, 1107 (2015).
523. Nobrega, F. L. *et al.* Targeting mechanisms of tailed bacteriophages. *Nat Rev Microbiol* **16**, 760–773 (2018).
524. Ackermann, H.-W. & Haldal, M. Basic electron microscopy of aquatic viruses. in *Manual of Aquatic Viral Ecology* (eds. Wilhelm, S., Weinbauer, M. & Suttle, C.) 182–192 (2010).
525. Family - *Siphoviridae*. in *Virus Taxonomy* (eds. King, A. M. Q., Adams, M. J., Carstens, E. B. & Lefkowitz, E. J.) 86–98 (2012).
526. McCabe, O. *et al.* The targeted recognition of *Lactococcus lactis* phages to their polysaccharide receptors. *Molecular Microbiology* **96**, 875–886 (2015).
527. Hu, B., Margolin, W., Molineux, I. J. & Liu, J. Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *PNAS* **112**, E4919–E4928 (2015).
528. Farinas, B. *et al.* Natural synchronisation for the study of cell Division in the green unicellular alga *Ostreococcus tauri*. *Plant Mol Biol* **60**, 277–292 (2006).
529. Crick, F. H. C. & Watson, J. D. Structure of small viruses. *Nature* **177**, 473–475 (1956).
530. Caspar, D. L. D. & Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* **27**, 1–24 (1962).
531. Xian, Y. & Xiao, C. Current capsid assembly models of icosahedral *Nucleocytoviricota* viruses. *Adv Virus Res* **108**, 275–313 (2020).
532. Marinescu, D. C., Yongchang Ji & Marinescu, G. M. Genetic economy, symmetry, and computational complexity for atomic virus structure determination. in *Proceedings 16th International Parallel and Distributed Processing Symposium* 8 pp (2002).
533. Dechant, P.-P. & Twarock, R. Models of viral capsid symmetry as a driver of discovery in virology and nanotechnology. *The Biochemist* **43**, 20–24 (2021).
534. Newcomb, W. W. *et al.* The UL6 gene product forms the portal for entry of DNA into the *Herpes simplex virus* Capsid. *Journal of Virology* **75**, 10923–10932 (2001).
535. Zauberman, N. *et al.* Distinct DNA exit and packaging portals in the virus *Acanthamoeba polyphaga mimivirus*. *PLOS Biology* **6**, e114 (2008).
536. Kuznetsov, Y. G. *et al.* Atomic force microscopy investigation of the giant mimivirus. *Virology* **404**, 127–137 (2010).
537. Xiao, C. *et al.* Structural studies of the giant Mimivirus. *PLOS Biology* **7**, e1000092 (2009).
538. Klose, T. *et al.* The three-dimensional structure of Mimivirus. *Intervirology* **53**, 268–273 (2010).
539. Black, L. W. DNA packaging in dsDNA bacteriophages. *Annu Rev Microbiol* **43**, 267–292 (1989).
540. Hendrix, R. W. Bacteriophage DNA packaging: RNA gears in a DNA transport machine. *Cell* **94**, 147–150 (1998).
541. Swale, E. M. F. & Belcher, J. H. A light and electron microscope study of the colourless flagellate *Aulacomonas Skuja*. *Archiv. Mikrobiol.* **92**, 91–103 (1973).

542. Allen Dodds, J. & Cole, A. Microscopy and biology of *Uronema gigas*, a filamentous eucaryotic green alga, and its associated tailed virus-like particle. *Virology* **100**, 156–165 (1980).
543. Baquero, D. P. *et al.* Chapter Four - Structure and assembly of archaeal viruses. in *Advances in Virus Research* (eds. Kielian, M., Mettenleiter, T. C. & Roossinck, M. J.) vol. 108 127–164 (2020).
544. Contursi, P., Fusco, S., Cannio, R. & She, Q. Molecular biology of fuselloviruses and their satellites. *Extremophiles* **18**, 473–489 (2014).
545. Stedman, K. M., DeYoung, M., Saha, M., Sherman, M. B. & Morais, M. C. Structural insights into the architecture of the hyperthermophilic Fusellovirus SSV1. *Virology* **474**, 105–109 (2015).
546. Bath, C. & Dyall-Smith, M. L. His1, an Archaeal Virus of the Fuselloviridae Family That Infects *Haloarcula hispanica*. *J Virol* **72**, 9392–9395 (1998).
547. Bath, C., Cukalac, T., Porter, K. & Dyall-Smith, M. L. His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, Salterprovirus. *Virology* **350**, 228–239 (2006).
548. Häring, M. *et al.* Independent virus development outside a host. *Nature* **436**, 1101–1102 (2005).
549. Prangishvili, D. *et al.* Structural and genomic properties of the hyperthermophilic Archaeal virus ATV with an extracellular stage of the reproductive cycle. *Journal of Molecular Biology* **359**, 1203–1216 (2006).
550. Kim, J.-G. *et al.* Spindle-shaped viruses infect marine ammonia-oxidizing thaumarchaea. *Proc Natl Acad Sci U S A* **116**, 15645–15650 (2019).
551. Thiroux, S. *et al.* The first head-tailed virus, MFTV1, infecting hyperthermophilic methanogenic deep-sea archaea. *Environmental Microbiology* **23**, 3614–3626 (2021).
552. Family - *Tectiviridae*. in *Virus Taxonomy* (eds. King, A. M. Q., Adams, M. J., Carstens, E. B. & Lefkowitz, E. J.) 317–321 (2012).
553. Padan, E. & Shilo, M. Cyanophages-viruses attacking blue-green algae. *Bacteriological Reviews* **37**, 343–370 (1973).
554. Fokine, A. & Rossmann, M. G. Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage* **4**, e28281 (2014).
555. Stummeyer, K. *et al.* Evolution of bacteriophages infecting encapsulated bacteria: lessons from *Escherichia coli* K1-specific phages. *Molecular Microbiology* **60**, 1123–1135 (2006).
556. Schwarzer, D. *et al.* A Multivalent adsorption apparatus explains the broad host range of phage phi92: a comprehensive genomic and structural analysis. *J Virol* **86**, 10384–10398 (2012).
557. Hardy, S., Legagneux, V., Audic, Y. & Paillard, L. Reverse genetics in eukaryotes. *Biology of the Cell* **102**, 561 (2010).
558. Cerutti, H., Ma, X., Msanne, J. & Repas, T. RNA-mediated silencing in algae: Biological Roles and Tools for Analysis of Gene Function. *Eukaryotic Cell* (2011).
559. Stoddard, B. L. Homing endonuclease structure and function. *Q Rev Biophys* **38**, 49–95 (2005).

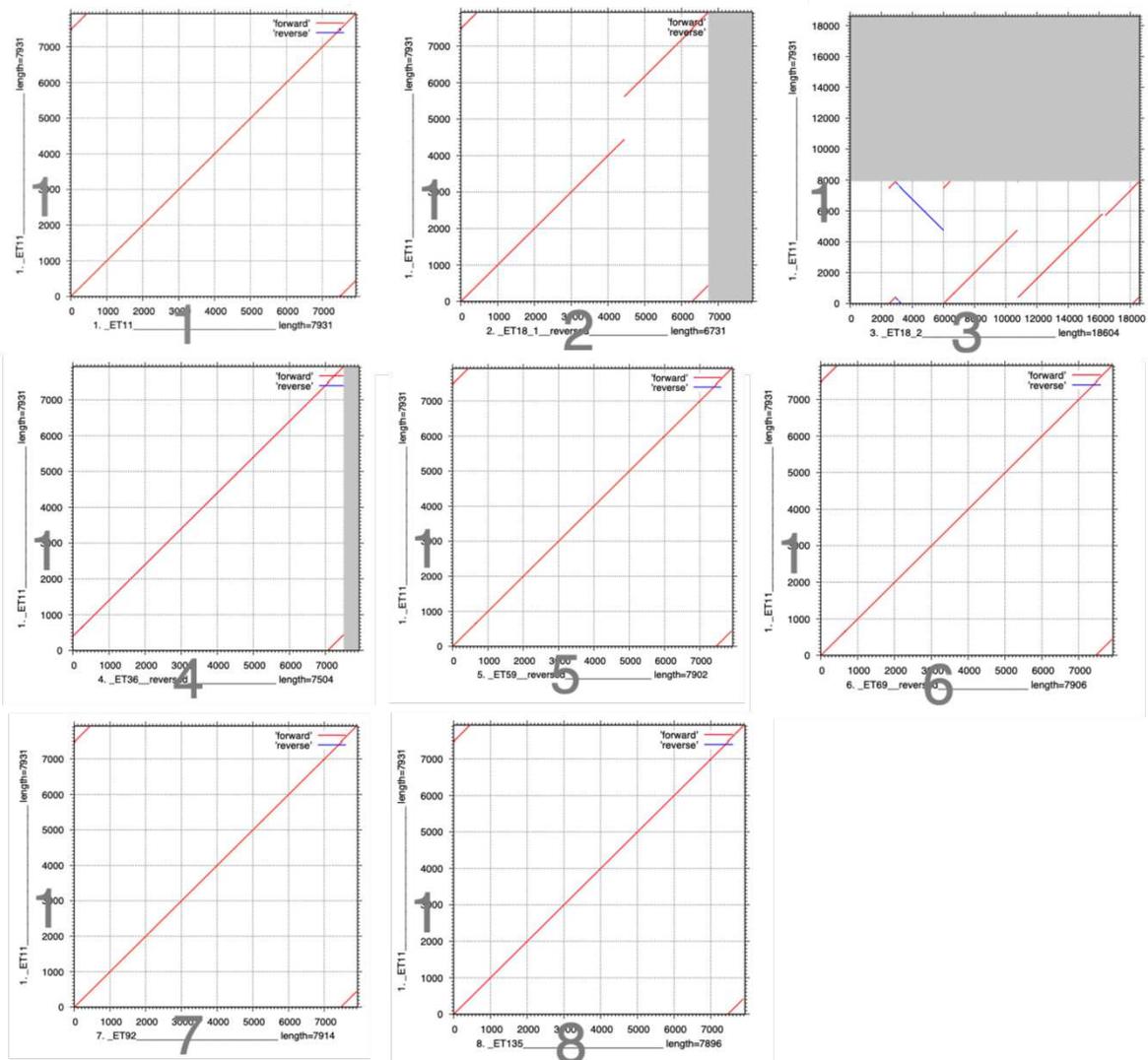
560. Miller, J. C. *et al.* An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* **25**, 778–785 (2007).
561. Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757–761 (2010).
562. Wood, A. J. *et al.* Targeted genome editing across species using ZFNs and TALENs. *Science* **333**, 307 (2011).
563. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* (2013).
564. Mir, A., Edraki, A., Lee, J. & Sontheimer, E. J. Type II-C CRISPR-Cas9 biology, mechanism, and application. *ACS Chem. Biol.* **13**, 357–365 (2018).
565. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* (2013).
566. Karkute, S. G., Singh, A. K., Gupta, O. P., Singh, P. M. & Singh, B. CRISPR/Cas9 mediated genome engineering for improvement of horticultural crops. *Front Plant Sci* **8**, 1635 (2017).
567. Hess, G. T., Tycko, J., Yao, D. & Bassik, M. C. Methods and applications of CRISPR-mediated base editing in eukaryotic genomes. *Mol Cell* **68**, 26–43 (2017).
568. Schuster, M. & Kahmann, R. CRISPR-Cas9 genome editing approaches in filamentous fungi and oomycetes. *Fungal Genetics and Biology* **130**, 43–53 (2019).
569. Jeon, S. *et al.* Current status and perspectives of genome editing technology for microalgae. *Biotechnology for Biofuels* **10**, 267 (2017).
570. Patel, V. K. *et al.* CRISPR–Cas9 system for genome engineering of photosynthetic microalgae. *Mol Biotechnol* **61**, 541–561 (2019).
571. Jiang, W., Brueggeman, A. J., Horken, K. M., Plucinak, T. M. & Weeks, D. P. Successful transient expression of Cas9 and single guide RNA genes in *Chlamydomonas reinhardtii*. *Eukaryot Cell* **13**, 1465–1469 (2014).
572. Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M. & Winge, P. A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci Rep* **6**, 24951 (2016).
573. Wang, Q. *et al.* Genome editing of model oleaginous microalgae *Nannochloropsis* spp. by CRISPR/Cas9. *The Plant Journal* **88**, 1071–1081 (2016).
574. Elbarbary, R. A., Lucas, B. A. & Maquat, L. E. Retrotransposons as regulators of gene expression. *Science* (2016).
575. Bowler, C. *et al.* The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239–244 (2008).
576. Berthelie, J. *et al.* A transposable element annotation pipeline and expression analysis reveal potentially active elements in the microalga *Tisochrysis lutea*. *BMC Genomics* **19**, 1–14 (2018).
577. Casacuberta, E. & González, J. The impact of transposable elements in environmental adaptation. *Mol Ecol* **22**, 1503–1517 (2013).
578. Pinsky, W., Haring, E., Hagemann, S. & Miller, W. J. The evolutionary life history of P transposons: from horizontal invaders to domesticated neogenes. *Chromosoma* **110**, 148–158 (2001).
579. Pace, J. K., Gilbert, C., Clark, M. S. & Feschotte, C. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proc Natl Acad Sci U S A* **105**, 17023–17028 (2008).

580. Loiseau, V. *et al.* Wide spectrum and high frequency of genomic structural variation, including transposable elements, in large double-stranded DNA viruses. *Virus Evolution* **6**, (2020).
581. Gilbert, C. & Cordaux, R. Viruses as vectors of horizontal transfer of genetic material in eukaryotes. *Current Opinion in Virology* **25**, 16–22 (2017).
582. Oliver, M. J., Schofield, O. & Bidle, K. Density dependent expression of a diatom retrotransposon. *Marine Genomics* **3**, 145–150 (2010).
583. Egue, F. *et al.* Expression of the retrotransposons Surcouf and Blackbeard in the marine diatom *Phaeodactylum tricornutum* under thermal stress. *Phycologia* **54**, 617–627 (2015).
584. Chen, J. E., Cui, G., Wang, X., Liew, Y. J. & Aranda, M. Recent expansion of heat-activated retrotransposons in the coral symbiont *Symbiodinium microadriaticum*. *ISME J* **12**, 639–643 (2018).
585. Mita, P. & Boeke, J. D. How retrotransposons shape genome regulation. *Current Opinion in Genetics & Development* **37**, 90–100 (2016).
586. Hall, A. R., Scanlan, P. D. & Buckling, A. Bacteria-Phage Coevolution and the Emergence of Generalist Pathogens. *The American Naturalist* **177**, 44–53 (2011).
587. Koskella, B. & Meaden, S. Understanding bacteriophage specificity in natural microbial communities. *Viruses* **5**, 806–823 (2013).
588. Schluter, D. *The Ecology of Adaptive Radiation*. (2000).
589. Bono, L. M., Gensel, C. L., Pfennig, D. W. & Burch, C. L. Competition and the origins of novelty: experimental evolution of niche-width expansion in a virus. *Biology Letters* **9**, 20120616 (2013).
590. Tomaru, Y., Tanabe, H., Yamanaka, S. & Nagasaki, K. Effects of temperature and light on stability of microalgal viruses, HaV, HcV and HcRNAV. (2005).
591. Demory, D. *et al.* Temperature is a key factor in *Micromonas*–virus interactions. *The ISME Journal* **11**, 601–612 (2017).
592. Kendrick, B. J. *et al.* Temperature-induced viral resistance in *Emiliana huxleyi* (Prymnesiophyceae). *PLOS ONE* **9**, e112134 (2014).
593. Yan, X. *et al.* The capsid proteins of a large, icosahedral dsDNA Virus. *Journal of Molecular Biology* **385**, 1287–1299 (2009).
594. Andreani, J. *et al.* Pacmanvirus, a new giant icosahedral virus at the crossroads between *Asfarviridae* and Faustoviruses. *Journal of Virology* (2017).
595. Xiao, C. *et al.* Cryo-EM reconstruction of the *Cafeteria roenbergensis* virus capsid suggests novel assembly pathway for giant viruses. *Sci Rep* **7**, 5484 (2017).





## DOCUMENTS SUPPLEMENTAIRES DU CHAPITRE II



**Figure S1.** Analyse génomique comparative des séquences nucléotidiques des huit éléments transposables, transposés à proximité des intégrations de l'ADN transgène chez *O. tauri* RCC1115. L'élément E11 (axe y) a été comparé (axe x) à 1) lui-même, 2) ET18-1, 3) ET18-2, 4) ET36, 5) ET59, 6) ET69, 7) ET92, 8) ET135.

**Tableau S1.** Matrice de distance (%) des séquences nucléotides des éléments transposables (ET).

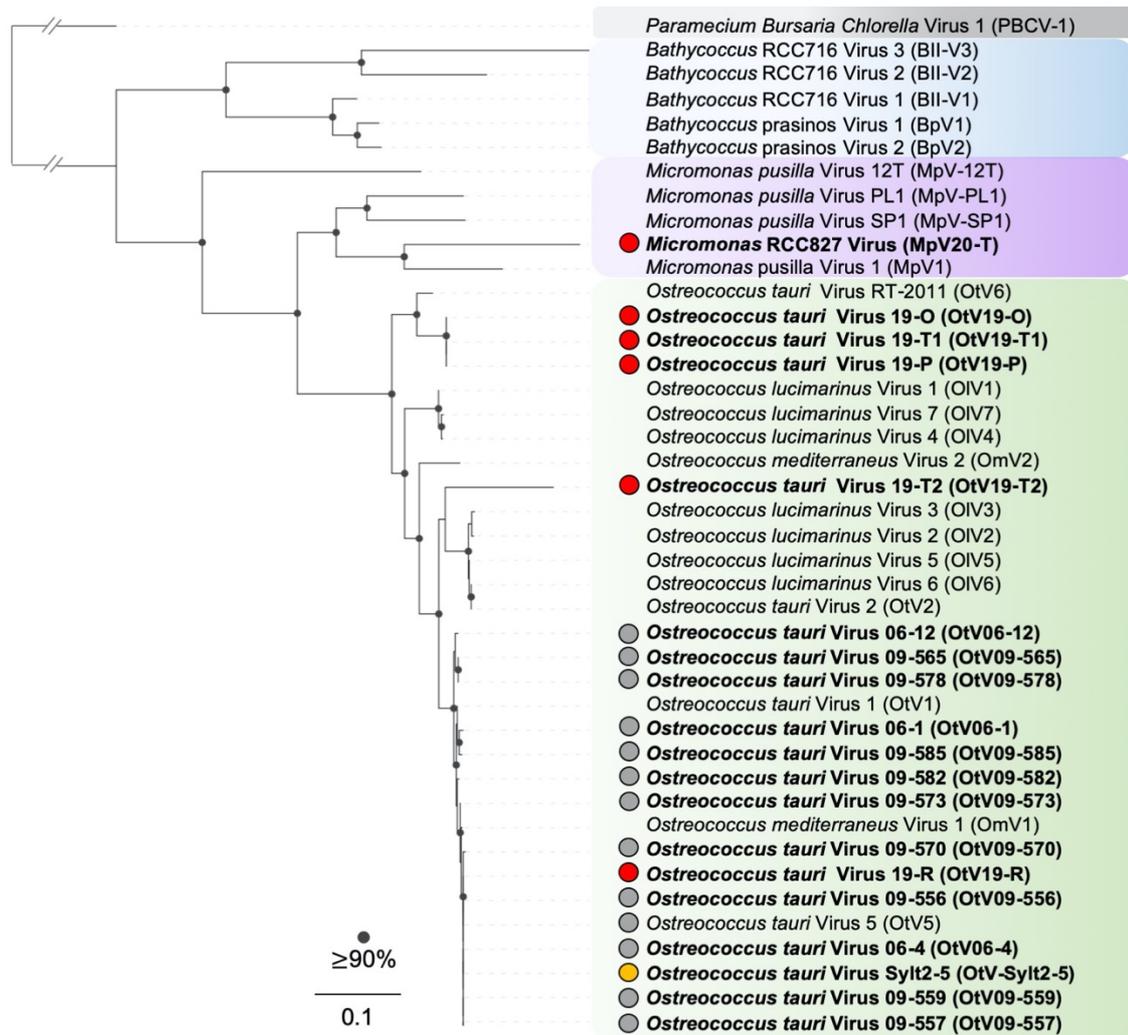
	<b>ET11</b>	<b>ET18-1</b>	<b>ET18-2</b>	<b>ET36</b>	<b>ET59</b>	<b>ET69</b>	<b>ET92</b>	<b>ET135</b>
<b>ET11</b>		74.4	90.7	97.8	98.0	97.9	96.7	97.8
<b>ET18-1</b>	74.4		66.9	72.9	74.5	74.4	73.2	74.3
<b>ET18-2</b>	90.7	66.9		94.1	91.1	90.9	89.8	90.8
<b>ET36</b>	97.8	72.9	94.1		98.3	98.1	96.8	98.1
<b>ET59</b>	98.0	74.5	91.1	98.3		98.3	97.1	98.4
<b>ET69</b>	97.9	74.4	90.9	98.1	98.3		96.8	98.2
<b>ET92</b>	96.7	73.2	89.8	96.8	97.1	96.8		96.9
<b>ET135</b>	97.8	74.3	90.8	98.1	98.4	98.2	96.9	

**Tableau S2.** Site d'insertions et la localisation des ETs dans les génomes de chaque lignée *O. tauri* RCC1115 transformées (T).

Lignée	Contig	Start	Stop	Chr	Localisation	Locus tag RCC4221	Fonction	Taille (bp)	GC (%)	Classe	Ordre	Sous-famille	ET
T3	11	85,565	93,454	1	Intergénique	osta01g02390	Porphobilinogen synthase	7,931	55.2	I	LTR	Ty1/Copia	ET11
	59	494,97	502,871	1	Intergénique	osta01g03620	Class II glutamine amidotransferase domain	7,902	55.3	I	LTR	Ty1/Copia	ET59
						osta01g03630	Nucleoporin FG repeat						
T6	69	107,549	115,454	9	Intragénique	osta09g00960	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	7,906	55.2	I	LTR	Ty1/Copia	ET69
T12	92	53,417	61,33	14	Intergénique	osta14g00230	Mechanosensitive ion channel	6,914	55.5	I	LTR	Ty1/Copia	ET92
	135	146,677	154,572	14	Intergénique	osta14g02235	Inconnue	7,896	55.4	I	LTR	Ty1/Copia	ET135
						osta14g02240	UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase						
T14	Scaffold 36	136,082	143,585	19	Intergénique	osta19g00440	Inconnue	7,504	55.4	I	LTR	Ty1/Copia	ET36
T16	18	6,937	13,667	20	Intergénique	osta01g01600	putative eukaryotic initiation factor eIF4	6,731	55.8	I	LTR	Ty1/Copia	ET18-1
						osta20g00010	Ankyrin repeat						
	18	72,218	90,821	20	Intergénique	osta20g00320	CDP-diacylglycerol-inositol 3-phosphatidyltransferase	18,6	56.8	I	LTR	Ty1/Copia	ET18-2
						osta20g00310	Inconnue						

# DOCUMENTS SUPPLEMENTAIRES DU CHAPITRE

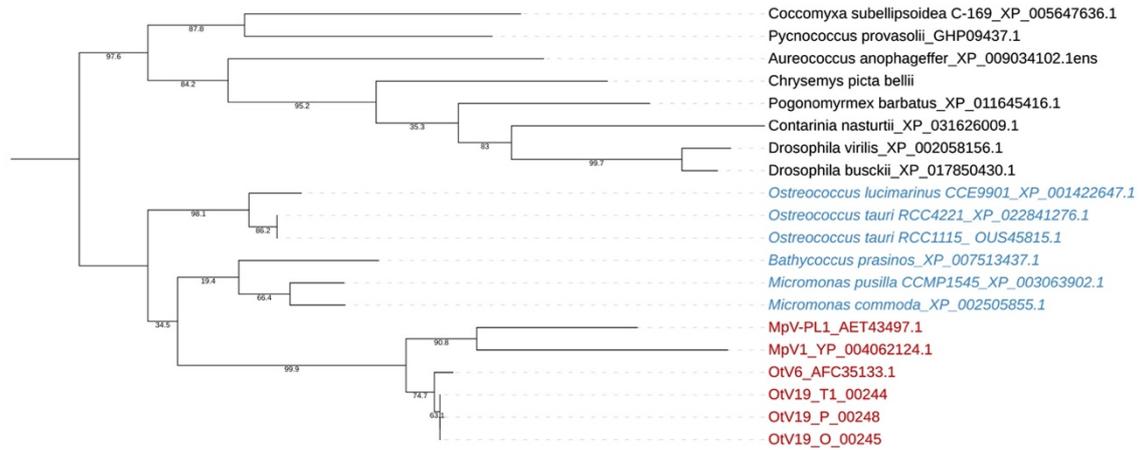
## III



**Figure S1.** Phylogenetic analysis of prasinoviruses. Maximum-likelihood (ML) phylogenetic reconstruction of full-length DNA polymerase B protein of prasinoviruses infecting the genera *Bathycoccus* (blue block), *Micromonas* (purple block) and *Ostreococcus* (green block). The Chlorovirus (PBCV-1) was used as an outgroup with the branch connecting truncated for display. The LG+F+R3 model was used based on the Bayesian Information Criterion (BIC). The support nodes were calculated by the SH-like aLRT (1,000 replicates) and UFBoot (1,000 replicates). Bootstrap supports higher to 90% are indicated as black dots and the scale bar represents the average number of substitutions per site. Red dots: OtVs isolated from the Pacific Ocean, Yellow dot: OtVs isolated from the Baltic Sea, Gray dots: OtVs isolated from the Mediterranean Sea (lagoon) in Clerissi *et al*, 2012.

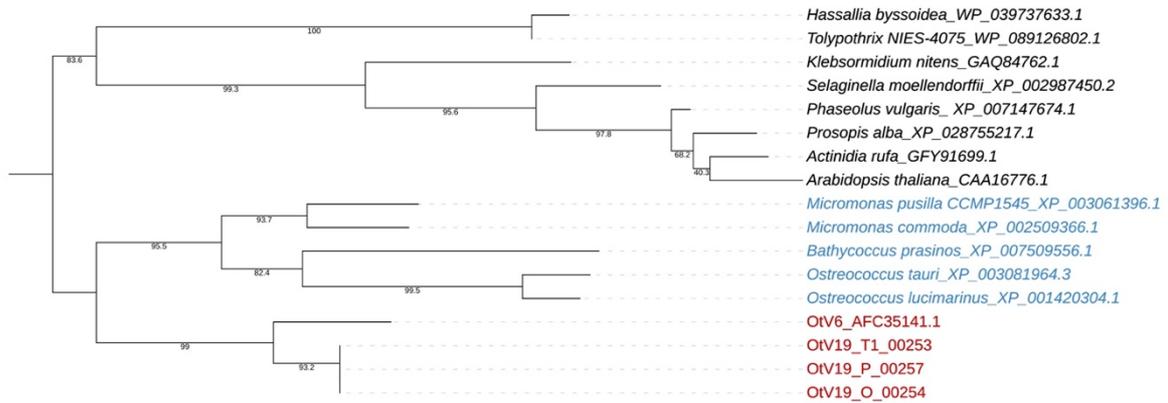
A)

Tree scale: 0.1

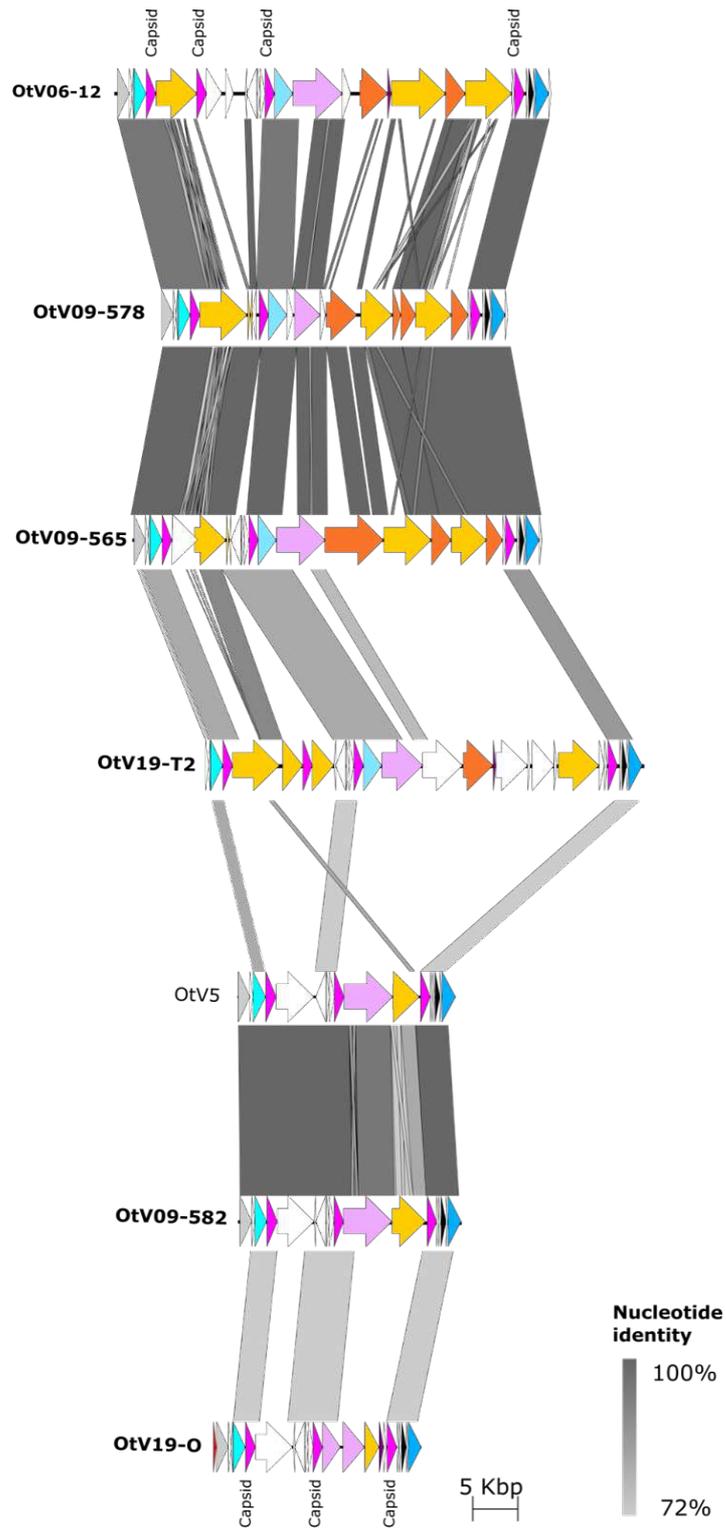


B)

Tree scale: 0.1



**Figure S2.** Maximum likelihood phylogenetic trees of A) mitochondrial 18KDa (MTP18) protein and B) alternative oxidase (AOX) protein. Prasinoviruses are labelled in red and their corresponding Mamiellophyceae algae host strains in blue.



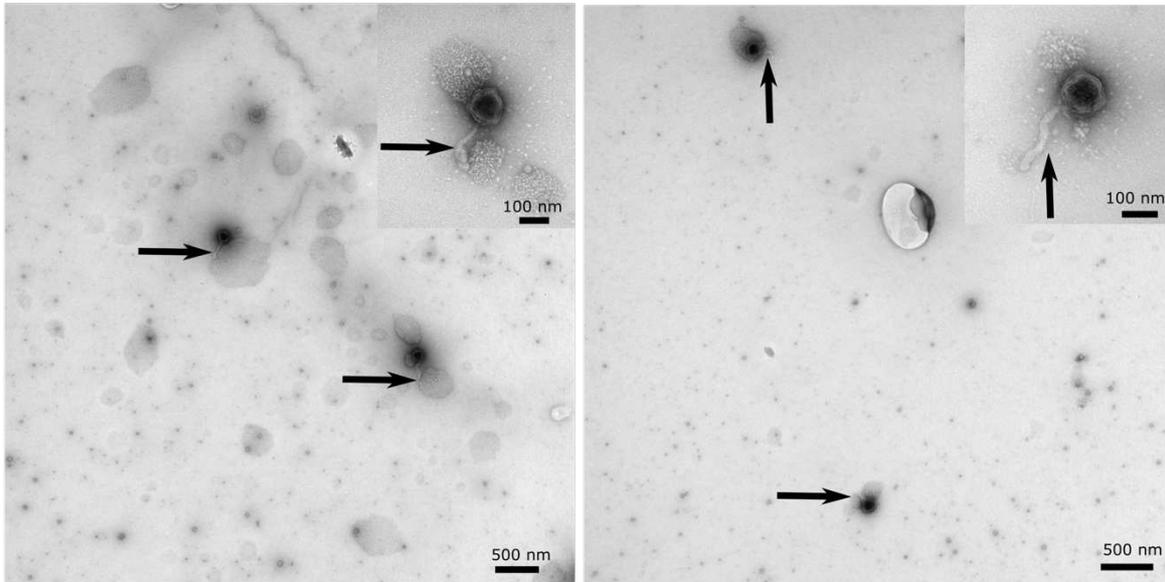
**Figure S3.** BLASTN comparison of the hypervariable region between representative OtVs. Shaded gray blocks join the regions between viruses with shared nucleotide identity (72–100%). BLASTN results are shown in the **Table S7**. Arrows represent CDSs and are colored according to the functional categories as shown in

**Figure 28.**

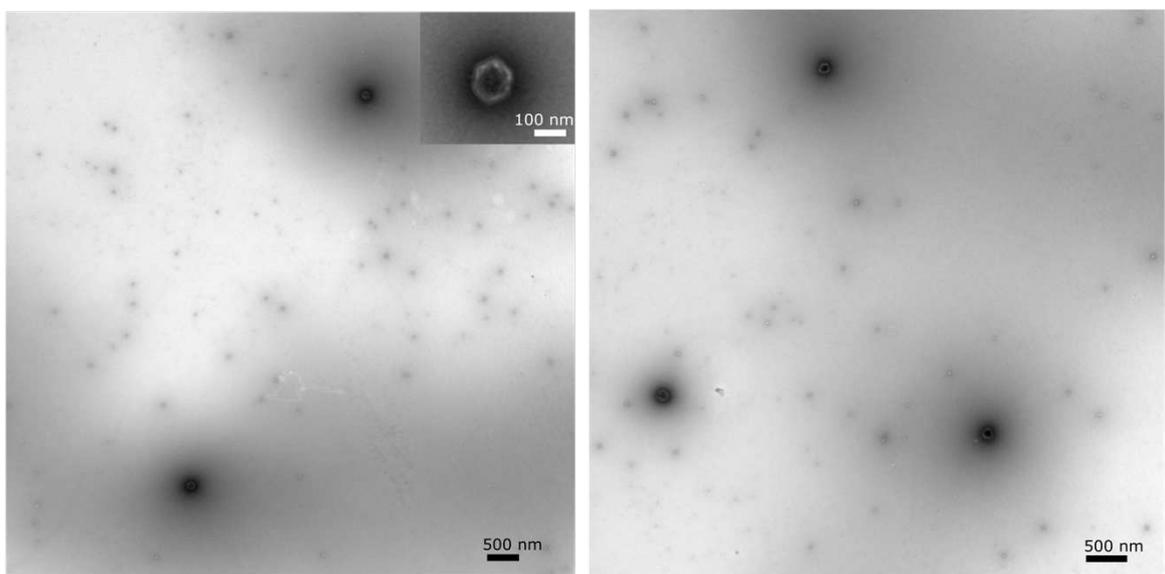
# DOCUMENTS SUPPLEMENTAIRES DU CHAPITRE

## IV

A)



B)



**Figure S1.** Observation des particules virales d'OtV09-565 issu d'un lysat à 4 jours-post infection, A) filtré à 0.45  $\mu\text{m}$  et B) filtré à 0.2  $\mu\text{m}$ . Les flèches noires indiquent la présence d'une queue présente sur un *vertex* de la capsid, tandis qu'aucune queue n'a été observée dans le même lysat viral, filtré à 0.2  $\mu\text{m}$ . Les photographies

## RESUME

Dans les écosystèmes marins, la population phytoplanctonique est sujette à l'infection par des virus géants affiliés au phylum *Nucleocytoviricota*. Nous nous sommes concentrés sur l'un des systèmes hôte-virus les plus abondants et les plus répandus dans les régions marines côtières : *Ostreococcus tauri*-*Prasinovirus*. En culture, l'émergence d'une résistance chez les microalgues sensibles aux virus a été observée et semble être le produit de variations structurelles importantes dans un chromosome spécialisé dans « l'immunité », appelé SOC (*small outlier chromosome*). Dans ce contexte, nous avons cherché à acquérir des connaissances sur les bases génétiques de la résistance aux *Prasinovirus* par une approche génomique. Ce travail a fourni une description préliminaire de l'activité des rétroéléments dans la souche RCC1115 d'*O. tauri*, attestant de leur importance dans l'évolution des génomes des microalgues. Le second objectif de cette thèse était de décrire la diversité génomique et les mécanismes sous-jacents à l'évolution de la gamme d'hôtes des virus infectant *O. tauri* (OtVs). Nous suggérons que l'évolution de la gamme d'hôtes des prasinovirus est principalement médiée par l'échange rapide et la recombinaison de gènes impliqués dans la reconnaissance de l'hôte qui sont principalement concentrés dans un locus génomique hautement variable, façonnant ainsi leur diversification et leur évolution. Globalement, ce travail de thèse nous a permis de mieux décrire les corrélats génomiques de la coévolution hôte-virus dans ce système enfermé dans une dynamique de course aux armements.

## ABSTRACT

In marine ecosystems, phytoplankton population is subject to infection by giant viruses affiliated with the phylum *Nucleocytoviricota*. We focused on one of the most abundant and widespread host-virus systems in coastal marine regions: *Ostreococcus tauri*-*Prasinovirus*. In culture, the emergence of resistance in virus-sensitive microalgae has been observed and appears to be the product of important structural variations in a chromosome specialised in 'immunity', called the SOC (small outlier chromosome). In this context, we sought to gain knowledge on the genetic basis of resistance to prasinoviruses through a genomic approach. This work provided a preliminary description of the activity of retroelements in *O. tauri* strain RCC1115, attesting to their importance in the and evolution of microalgal genomes. The second objective of this thesis was to describe the genomic diversity and mechanisms underlying the evolution of host range of *O. tauri*-infecting viruses (OtVs). We suggest that host range evolution in prasinoviruses is primarily mediated through rapid gene exchange and recombination in genes involved host recognition that are mainly concentrated in one genomic locus highly variable, thus shaping their diversification and evolution. Overall, this thesis work has allowed us to better describe the genomic correlates of host-virus coevolution in this system locked in an arms-race dynamic.