

Master Bioinformatique et Master IMHE

HAU802I : BILL

## Étude de l'évolution *in vitro* du Cyprinid herpesvirus 3 (CyHV-3)

Autrices : Marine DALLO & Laurene ROBERT

### Résumé :

Provoquant des symptômes à la fois physiques et comportementaux, le *Cyprinid herpesvirus 3* cause des dégâts socio-économiques majeurs dans les pays d'Asie et d'Afrique. Le CyHV-3 est inscrit dans la liste des pathogènes à surveiller au World Organisation for Animal Health (OIE) et dans l'Union Européenne (EU). Co-évoluant avec la carpe depuis de nombreuses années, le CyHV-3 est soudainement passé à l'état pathogène en 2016 et à ainsi retenu l'attention des chercheurs du projet BILL. Ils se sont concentrés sur l'étude de la souche italienne (KHV-1) du CyHV-3 qui est un virus à ADN double brin de 295 kpb et possédant 156 ORF. L'ADN a été extrait à partir d'échantillons de différents passages de cultures cellulaires, puis séquencé dans le but d'étudier les variations structurales de type INDELs suite à un stress thermique à 15°C et 28°C. Ce rapport se concentre sur l'étude des variations structurales suite au stress chaud à 28°C.

Pour répondre à cette problématique la mise en place d'un protocole expérimentale a vu le jour. Ce protocole concerne les étapes allant de l'extraction de l'ADN du virus à l'analyse bio-informatique des résultats du séquençage. Par la suite, une analyse des résultats par l'application de différents filtres a eu lieu, ce qui a permis de retenir au total 23 mutations. De ces mutations en a découlé la recherche des gènes, de leurs ORFs associés, ainsi que de leurs fonctions.

Cette étude a permis, bien qu'aucune conclusion a pu être clairement défini, de soulever une hypothèse importante sur l'impact d'un stress thermique de 28°C, sur le virus. En effet, cette hypothèse concerne la reproduction du virus en conditions de stress chaud, qui aurait tendance à décliner.

En parallèle plusieurs méthodes permettraient de confirmer cette hypothèse et ainsi d'établir une fonction avérée aux ORFs.

[Version du 15 mars 2023 | Version finale]

---

# Abréviations

<b>BAM</b>	Binary Alignment Map
<b>CyHV-3</b>	Cyprinid Herpesvirus 3
<b>IGV</b>	Integrative Genomics Viewer
<b>INDEL</b>	INsertion-DELetion
<b>KHV</b>	Koi Herpesvirus
<b>NGS</b>	New Generation Sequencing
<b>ORF</b>	Open Reading Frame cadre de lecture ouvert
<b>SAM</b>	Sequence Alignment Map
<b>SNV</b>	Single Nucleotide Variant
<b>SNP</b>	Variation d'un seul nucléotide Single Nucleotid Polymorphism Polymorphisme d'un seul Nucléotide
<b>VAF</b>	Variant Allelique Frequency Fréquence Allélique du Variant
<b>VCF</b>	Variant Calling Format Format d'Appel de Variant

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Contexte : Présentation de l'étude et objectifs . . . . .	1
1.2	Présentation du modèle : <i>Cyprinid Herpesvirus 3 (CyHV-3)</i> . . .	1
<b>2</b>	<b>Matériels et méthodes</b>	<b>2</b>
2.1	Cultures virales . . . . .	2
2.2	Extraction et contrôle qualité . .	2
2.3	Barecoding et Séquençage . . . .	3
2.4	Pipeline bio-informatique . . . .	3
<b>3</b>	<b>Résultats</b>	<b>3</b>
3.1	Extraction et séquençage . . . . .	4
3.1.1	Résultats de l'extraction de l'ADN du CyHV-3 . .	4
3.1.2	Résultats du séquençage des échantillons . . . . .	5
3.2	Sélection des échantillons . . . . .	5
3.3	Pré-traitement de la lignée souche P01 . . . . .	5
3.4	Filtres d'analyse . . . . .	6
3.5	Résultats obtenus . . . . .	7
<b>4</b>	<b>Discussion</b>	<b>8</b>
4.1	Les ORFs . . . . .	8
4.2	La qualité des échantillons . . . .	9
<b>5</b>	<b>Conclusion</b>	<b>10</b>
<b>6</b>	<b>Bibliographie</b>	
<b>A</b>	<b>Annexes :</b>	<b>A</b>
A.1	Annexe I : pipeline d'analyse bio-informatique . . . . .	A
A.2	Annexe II : image des profondeurs des échantillons . . . . .	B
A.3	Annexe III : localisation des ORFs	C
A.4	Annexe IV : mutations retenues après tri et comparaison . . . . .	D
A.5	Annexe V : Tableau récapitulatif des fonctions des ORFs . . . .	E
A.6	Annexe VI : Résultat des flagstats	F

# 1 Introduction

## 1.1 Contexte : Présentation de l'étude et objectifs

La mutualisation des Masters bio-informatique et IMHE (Interaction Hôte Micro-organismes Environnement), ainsi que la participation des chercheurs, vise à étudier le comportement génétique du virus *Cyprinid herpesvirus 3* (CyHV-3) chez les carpes communes.

Durant sur ces 5 dernières années, ces équipes de recherches ont étudiés ce virus avec des techniques différentes de celles qui seront présentées dans cet article. En effet, les chercheurs ont travaillé sur des shorts reads grâce au kit Nextera XT DNA Library Prep kit de la société Illumina [1] et ont utilisé le kit Nucleospin de Macherey Nagelun afin d'extraire l'ADN du CyHV-3 [1]. Cette année, un des nombreux objectifs de ces chercheurs, est de mettre en évidence l'impact éventuel d'un choc thermique à la fois chaud (28°C) et froid (15°C) sur les variations structurales de l'ADN de CyHV-3.

Bien que la présence du CyHV-3 et de l'aquaculture de carpes soit minime en France, ces recherches scientifiques ont un intérêt majeur pour le reste du monde, notamment pour les pays d'Asie et d'Afrique pour lesquels l'élevage représente un enjeu socio-économique très important [1].

La problématique choisie est de rechercher les variations structurales après un stress chaud à 28°C et d'étudier l'impact potentiel de ces variations sur le virus.

Afin de répondre à la problématique, un protocole expérimental a été mis en place. Il s'agit de l'adaptation d'un protocole d'extraction d'ADN humain à l'ADN du virus, ainsi qu'à la mise en place, pour l'analyse bio-informatique, d'un pipeline expérimental permettant de répondre au mieux aux attentes biologiques.

## 1.2 Présentation du modèle : *Cyprinid Herpesvirus 3* (CyHV-3)

Mesurant entre 170 et 200 nm de diamètre, le *Cyprinid herpesvirus 3* est composé d'un ADN double brin qui porte un génome de plus de 295 000 bases. Contenu dans le tégment constitué

de nombreuses protéine, l'ADN du CyHV-3 est situé dans une nucléocapside (Figure 1 [2])

Son cycle de reproduction durant une dizaine

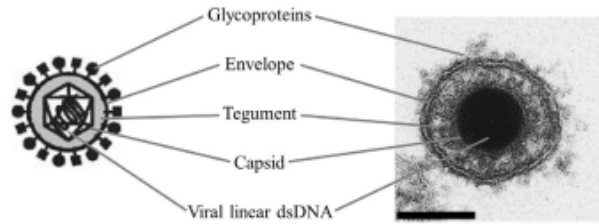


FIGURE 1 – Image et représentation du virus. *Schéma et photo au microscope du virus CyHV-3 et sa composition.*[3]

de jours, le virus présent dans les excréments se diffuse dans l'eau et entre en contact avec les cellules de la carpe (Figure 2). En commençant par se fixer à un récepteur transmembranaire, l'enveloppe du virus fusionne avec la membrane de la cellule du poisson et le tégment, en se dissociant, conduit à l'introduction de la nucléocapside dans la cellule, qui migre jusqu'à la membrane nucléaire de la cellule. L'ADN du virus est ensuite libéré dans le noyau de la cellule et y sera répliqué et transcrit grâce à la machinerie cellulaire de la carpe. La capsid est ensuite assemblée dans le noyau et est libérée dans le cytoplasme emportant avec lui la membrane nucléaire qui constitue son enveloppement primaire. Le tégment, les glycoprotéines et l'enveloppe secondaires, produit par l'appareil de Golgi, sont assemblées dans le cytoplasme. Une fois les différents constituants du virus réunis, il est libéré dans le milieu extérieur par exocytose, provoquant ainsi, la lyse de la cellule du poisson ([3]). Responsable de la maladie koi herpesvirus disease chez les carpes communes, le *Cyprinid herpesvirus 3* est un virus actif entre 15 et 28°C, qui provoque des symptômes à la fois physiques et comportementaux. En effet, une infection due au CyHV-3 cause l'apparition de taches et de cloques sur la peau et les branchies, ainsi qu'un enfoncement des yeux. De plus, il peut provoquer désorientation, de l'hyperactivité et l'isolation de la carpe contaminée de ses congénères. [4]

Après infection, le CyHV-3 a la capacité de persister dans l'hôte sous forme latente, jusqu'à que la température lui soit favorable (entre 15 et 28°C). De même, tout en étant actif, il est

capable d'être porté asymptomatiquement par l'hôte.

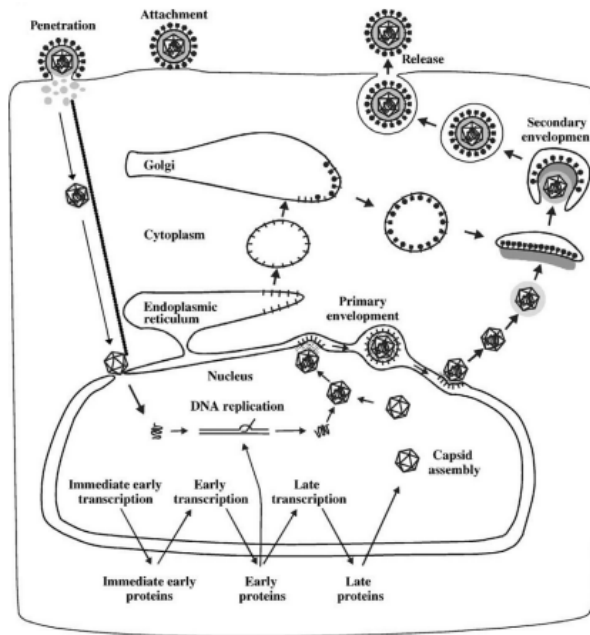


FIGURE 2 – Schéma du mécanisme de réplication du virus.

*Schéma du mécanisme d'invasion et de reproduction du virus CyHV-3 dans une cellule cible.[3]*

P40 et P50. Pour les lignées C1 et C5, le stress qui a été appliqué est une température froide, de 15°C, après le passage P20 pour une durée de 24 heures. Pour les lignées C6 et C10, le stress appliqué a été une température chaude de 28°C, également après le passage P20 et également pour une durée de 24 heures. Le schéma 3 représente les étapes de la culture cellulaire effectuée pour le CyHV-3.

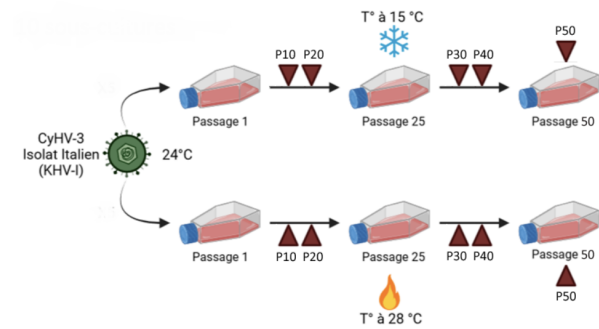


FIGURE 3 – Schéma des sous cultures du CyHV-3.

*Ce schéma représente les sous cultures du virus CyHV-3 réalisées pour cette étude. Jupyter Notebook de Mme FISTON LAVIER and al visible sur le dépôt github du projet.*

## 2 Matériels et méthodes

Dans le but d'étudier l'impact d'un stress de thermique sur le CyHV-3, une mise en culture de celui-ci a été faite. Une fois cette sous culture établie, une extraction de l'ADN du virus aux différents passages et pour chaque lignée a été réalisée. Suite à cela, dans un but d'identification des échantillons, un barcoding et un séquençage NGS de ces divers échantillons ont été fait. Visant à aligner ces séquences sur un génome de référence afin d'analyser les variations structurales recherchées, un pipeline expérimental a été mis en place.

### 2.1 Cultures virales

Contenant le virus, des cellules de carpes cultivées *in vitro* ont été prélevées. L'échantillon nommé P01 est la culture de départ de laquelle ont découlé les différentes lignées. De cette culture a été fait 4 sous cultures de lignées différentes, C1, C5, C6 et C10, chacune ayant subi plusieurs passages P10, P20, P30,

### 2.2 Extraction et contrôle qualité

L'extraction de l'ADN du CyHV-3 a été réalisée avec le kit MagAttract de la marque Qia-gen qui comprend plusieurs étapes clefs.

Tout d'abord, la protéinase K de ce kit permet la lyse des membranes cellulaires de la carpe et des protéines présentes dans la suspension cellulaire. Ensuite, de part sa charge négative, l'ADN se lie à des billes magnétiques facilitant ainsi les différents lavages permettant d'éliminer les contaminants. Enfin, l'étape d'élution est un lavage qui, grâce à la neutralité du tampon AE, permet de ré-équilibrer les charges et de décrocher l'ADN de la surface des billes. (cf. figure4)

Le protocole d'extraction est disponible sur le dépôt github : <https://github.com/marine30190/projetBILL.git>. Une fois l'extraction de l'ADN effectuée, sa qualité a été dosée par le nanodrop et sa quantité par le qubit.

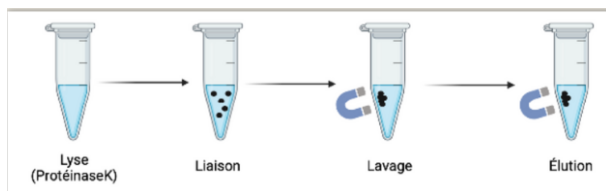


FIGURE 4 – Schéma simplifié des étapes d'extraction de l'ADN du virus CyHV-3.

Ce schéma simplifié représente les étapes d'extraction de l'ADN du virus CyHV-3 réalisées grâce au kit MagAtract de la société Qiagen.

## 2.3 Barecoding et Séquençage

Afin de différencier les échantillons après le séquençage, un barcode, grâce au kit "Rapid Barcoding Sequencing (SQK-RBK004)" de la société Nanopore (cf. dépôt github : <https://github.com/marine30190/projetBILL.git>), a été attribué à chaque échantillon qui ont ensuite été "poolés" c'est à dire la mise en commun dans une même cupule pour la suite du protocole. Dans le but d'analyser les longs reads, une technique de séquençage haut débit sans PCR a été utilisée. Il s'agit du nanopore d'oxford et plus précisément du MinION. [5] Cette technologie repose sur l'utilisation de pores protéiques nanométriques, ou "nanopore" qui, intégrés dans une membrane de polymère électriquement résistante, servent de biocapteurs. En passant par ces biocapteurs, chaque base émet un signal électrique qui, grâce au "base calling" inclus dans le MinION est automatiquement convertie en base (ATGC).

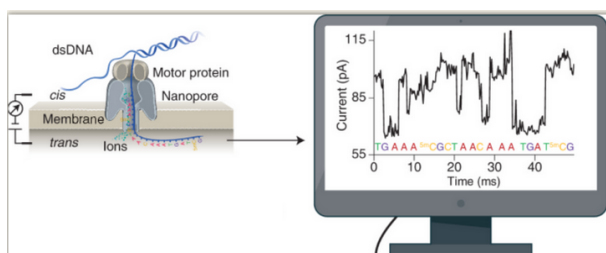


FIGURE 5 – Schéma simplifié du fonctionnement d'un pore de la technologie Nanopore et du basecalling.

Ce schéma simplifié du fonctionnement d'un pore Nanopore et de l'étape de basecalling.[5]

## 2.4 Pipeline bio-informatique

La partie bio-informatique peut se résumer en trois grandes étapes :

- **Analyse primaire** : "L'alignement des reads sur le génome de référence" ;

- **Analyse secondaire** : "L'appel des variants" ;
- **Analyse tertiaire** : "L'annotation des variants".

Ces trois étapes constituent ce que l'on appelle un "pipeline" d'analyse, qui est une suite d'outils permettant l'analyse des grandes quantités de données issues d'un séquençage haut débit (NGS). Ce pipeline est géré par un "snakemake" qui permet l'exécution des outils à la suite et de façon autonome.(cf. figure 11 en annexe I page A).

Premièrement les outils "PycoQC" et "Seqkit" vont permettre d'analyser la qualité du séquençage. PycoQC se base sur les fichiers "fastQ" et créer des graphiques montrant la qualité du run entier. Seqkit y ajoutera des informations statistiques telles que le nombre de bases ou la longueur des reads de chaque échantillon.

Deuxièmement, l'outil "Minimap2" a été utilisé pour réaliser le mapping, qui permet après alignement sur le P01 (souche italienne), d'aligner les reads de l'échantillon viral sur la souche CyHV-3 japonaise, utilisée comme génome de référence du virus. Des fichiers "sam" sont ainsi générés et peuvent être traités par l'outil "Samtools" qui possède différentes options telles que "Samtools sort" qui trie les séquences sur la position auxquelles elles sont associées ; "Samtools flagstat" qui compte le nombre d'alignements ; "Bam coverage" et "Plot coverage" qui permettent respectivement de connaître la couverture et la profondeur du séquençage et enfin "Samtools index" qui prépare les reads pour l'appel des variants. Cet outil génère des fichiers "bai" qui complètent les fichiers "bam" nécessaires pour la suite des étapes d'analyses.

Pour finir, l'outil "Sniffles" a été utilisé pour effectuer l'appel des variants. Cet outil de bioinformatique détecte les variations de nucléotides entre le génome de référence et les échantillons séquencés. Sniffles est spécialisé dans les longs reads et permet par la suite aux opérateurs, de faire un lien entre les génotypes. En effet, grâce à l'outil "IGV", les fichiers "vcf" ont pu être analysés, permettant ainsi d'émettre quelques hypothèses pour répondre à notre problématique. Le pipeline détaillé et commenté est disponible sur le lien github dans le jupyter notebook. (cf. <https://github.com/marine30190/projetBILL.git>).

Chacun des outils de ce pipeline utilise un type de fichier en input et output différent. La figure 6 schématise le déroulé des différents types de fichier en fonction de l'évolution des étapes du pipeline d'analyse.

## 3 Résultats

La première étape de l'analyse des résultats est l'extraction d'ADN du *Cyprinid-HerpesVirus 3*. A

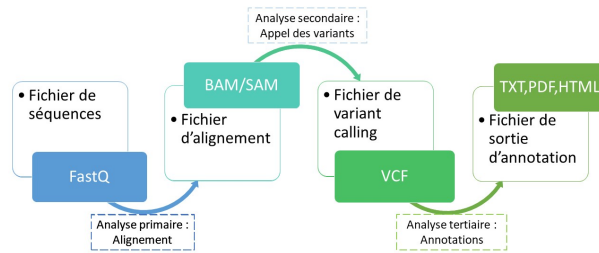


FIGURE 6 – Schéma simplifié des fichiers d'un pipeline.

*Ce schéma simplifié représente les différents fichiers en input et output au cours du pipeline.*

l'issue de cette étape la vérification de la quantité et la qualité de l'ADN extrait est possible. Ensuite une autre étape dont les résultats sont observables est le séquençage de l'ADN. Grâce aux premiers résultats il est possible de discriminer les échantillons qui ont été correctement séquencés et ceux dont le séquençage va émettre des doutes concernant la fiabilité des mutations.

Grâce au pipeline expérimental et suite à l'étape du séquençage des échantillons, un listing brut (non traité et non filtré) de mutations a été généré au format VCF. En supp DATA, sur github, se situe un exemple d'un fichier VCF pour un échantillon (cf. <https://github.com/marine30190/projetBILL.git>). Pour rappel, les échantillons sont issus de quatre lignées différentes (C1, C5, C6 et C10) et des différents passages (P10, P20, P30, P40 et P50).

Afin d'analyser ces mutations et de sélectionner uniquement les mutations fiables, différents filtres ont été appliqués, ainsi qu'un pré-traitement de la lignée souche appelé P01. Dans cette partie seront présentés, les résultats de l'extraction de l'ADN du CyHV3, du séquençage, du pré-traitement de la lignée souche P01 et par la suite, l'explication des filtres utilisés ainsi que les résultats obtenus.

A l'issue de ces analyses le choix des échantillons sélectionnés s'est porté sur les échantillons P10 et P30 de la lignée C1, lignée de stress thermique froid, qui nous servira de contrôle sur les mutations relevées. Les échantillons permettant de répondre à la problématique sont ceux des passages de la lignée C10, qui est une lignée concernée par le stress thermique à 28°C. De plus, le P50C6 de la lignée de stress de thermique chaud, sera également utilisé comme contrôle lors de l'analyse des mutations. Dans cette partie, l'analyse se portera uniquement sur ces échantillons sélectionnés et non sur toutes les sous-cultures.

### 3.1 Extraction et séquençage

L'extraction et le séquençage sont des étapes importantes dans l'étude des variations structurales. En effet, les résultats qui découlent de ces étapes sont des facteurs pouvant éventuellement être un risque de biais. En effet les biais d'extraction et de séquençage sont de plusieurs natures, par exemple, le facteur humain, autrement dit l'opérateur. C'est à dire que lors de l'expérimentation, l'erreur de manipulation est la première cause d'erreur, cela peut être dû à une mauvaise technique ou une à une contamination.

#### 3.1.1 Résultats de l'extraction de l'ADN du CyHV-3

Après extractions des échantillons, une analyse au laboratoire a été effectuée afin de déterminer la pureté du matériel génétique du CyHV-3. L'outil Qubit, a permis de doser la quantité d'ADN, tandis que le dosage au Nanodrop a permis de déterminer la qualité de ce même ADN. Comme le montre la figure 7, l'on peut voir la répartition des absorbances des échantillons.

Le ratio 260/280 est un indicateur de contamination par des protéines, alors que le rapport d'absorbance 260/230 est un indicateur de contamination par des agents chaotropiques ou par des composés organiques. Afin que l'ADN soit caractérisé de "pur" et pour continuer la suite des analyses, ces deux rapports d'absorbances doivent être proche d'une DO de 2. Bien que les absorbances 260/230 ne respectent pas les normes, elle ne sera pas prise en compte car le ratio 260/280, qui représente la donnée la plus importante, est valide.

Les autres résultats primaires importants sont les quantités d'ADN des échantillons. Cette information est indispensable au lancement du séquençage. Le séquençage demande une quantité minimum, qui est de 400ng, pour assurer un séquençage correct et fiable.

Comme le montre le tableau 1, deux groupes se distinguent de part leurs quantités d'ADN. En effet, les échantillons de P10C1 à P30C10, sont faiblement concentrés en ADN (inférieur à 100ng). En outre, les échantillons de P40C10 à P50C10 détiennent des concentrations élevées en ADN (supérieure à 400ng). Ces concentrations vont être utilisées pour valider l'extraction d'ADN, mais aussi pour préparer la répartition des échantillons sur les flowcells du séquenceur.



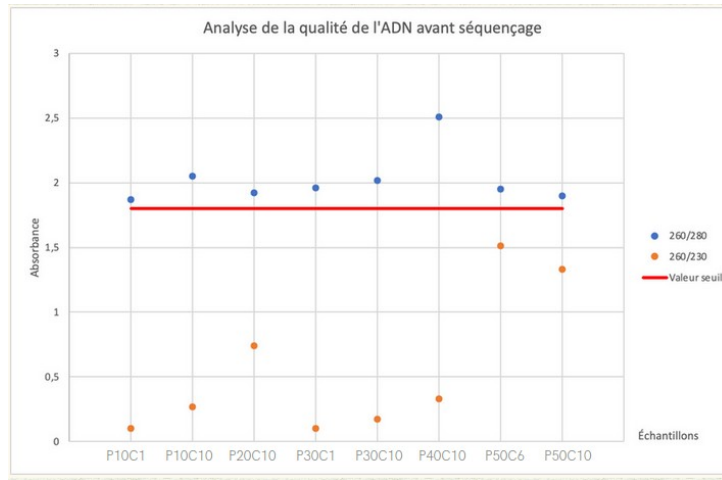


FIGURE 7 – Graphique des absorbances du Nanodrop.

Ce graphique représente l'absorbance 260/280 et 260/230 pour chaque échantillon sélectionné.

Echantillons	Qbit (ng/μl)	Nanodrop (ng/μl)	Quantité totale (ng)
P10C1	3,49	3,9	36
P10C10	5,59	6,3	62
P20C10	5,78	18,2	66
P30C1	6,37	13,7	73
P30C10	6,08	12,7	70
P40C10	35,1	17	404
P50C6	48,6	60,3	559
P50C10	38,8	66,9	446

TABLE 1 – Tableau récapitulatif des résultats de dosage.

Ce tableau représente la quantité d'ADN pour chaque échantillon sélectionné.

### 3.1.2 Résultats du séquençage des échantillons

Après le barecoding des échantillons, ils ont été séquencés. Le choix du regroupement des échantillons sur les flowcells s'est fait en fonction de leurs concentrations. Une fois le séquençage effectué les premiers résultats bruts de chaque séquençage en fonction de la flowcell sont lisibles grâce un fichier HTML comme montré en annexe sur le dépôt github : <https://github.com/marine30190/projetBILL.git>. Dans ce document plusieurs renseignements sont présents, mais seuls les éléments concernant le nombre de reads par échantillon sont utilisés pour cette étude.

Grâce à la figure 8, qui est un exemple d'une des flowcells et qui représente quelques échantillons, il est possible de relever le nombre de reads pour chaque échantillon qui ont passé le filtre de qualité du séquençage qui était Q8. Cet indice de qualité permet d'accepter 10% d'erreur lors du séquen-

çage. En annexe sur un dépôt github <https://github.com/marine30190/projetBILL.git> sont répertoriés tous les nombres de reads par échantillon. Ces résultats sont une aide pour déterminer quels échantillons ont été séquencés correctement et ceux qui ont rencontré un problème lors du séquençage.

### 3.2 Sélection des échantillons

Pour rappel, deux types de lignée étaient disponibles, la lignée ayant subi un stress froid au nombre de deux sous-cultures (C1 et C5) et la lignée ayant subi un stress chaud également au nombre (C6 et C10). Afin de répondre à la problématique, un choix concernant les échantillons a été fait grâce à l'outil IGV. Cet outil permet de visualiser les profondeurs de chaque échantillon et de les comparer. Comme le montre la figure 12 en annexe II page B, les échantillons ne sont pas tous uniformes et la profondeur de certains échantillons nous laissent penser que certaines zones du génome ne sont pas couvertes.

### 3.3 Pré-traitement de la lignée souche P01

Pour obtenir des résultats les plus fiables possible une étape de pré-traitement du P01, qui est la souche de départ prélevée sur la carpe commune, a été effectuée. Cette étape consiste à aligner le P01 à la séquence de référence, avec laquelle l'ensemble des séquences ont été alignés. Cette étape a eu pour but de vérifier la présence de variations structurales dans l'échantillon P01. Grâce à ce pré-traitement, les mutations appartenant à P01 pourront être éliminées, car elles ne sont pas dues au stress thermique ou à un biais de passage.

La fonction BLAST de NCBI a été utilisée pour

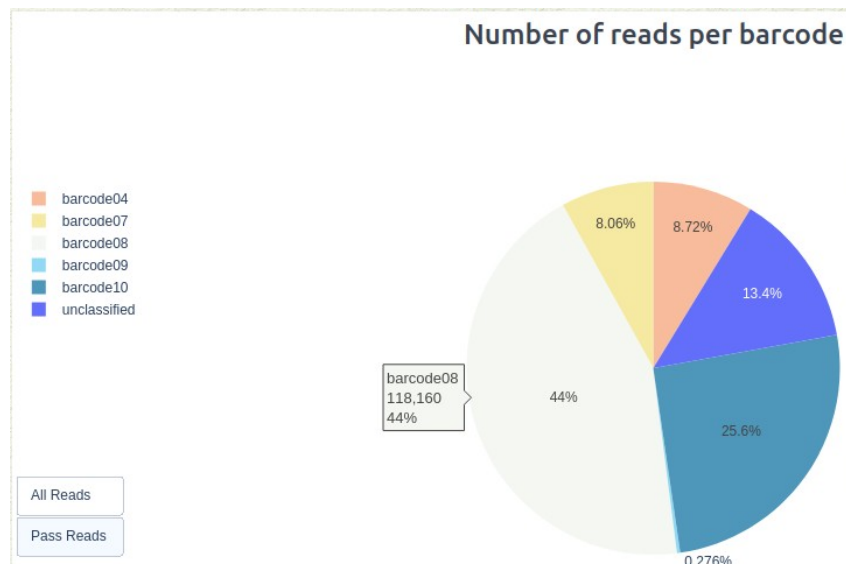


FIGURE 8 – Graphique représentant le nombre de reads par échantillon.

*Ce graphique représente le nombre de reads par échantillon ainsi que les reads non identifiés.*

aligner les séquences. Cet alignement a été contrôlé via MAFFR, qui est un outil permettant les alignements multiples. Ce résultat a démontré que la séquence de référence et le P01 étaient identiques (voir figure 9).

Comme montre la figure 9, les séquences du P01 et de la séquence de référence sont alignées grâce à la droite centrale sur laquelle il est possible d'observer la similarité. Un autre constat a été fait, la présence de séquences répétées en début et en fin du génome est observable. La figure 13, en annexe III page C, confirme bien la présence des zones répétées dans le génome.

Ce résultat permet de valider l'ensemble des mutations retrouvées et de n'en exclure aucune. Les mutations retenues sont hypothétiquement dues à un stress de température ou à un biais de passage. De même, cela permet d'aligner sur un génome de référence avec des annotations connues et non pas sur la souche de départ, le P01.

### 3.4 Filtres d'analyse

Le premier filtre utilisé pour la sélection des échantillons a été de visionner les BEDgraphs, qui permettent d'observer les profondeurs de chacune des séquences. Grâce à cette visualisation (12 de l'annexe II page B), 8 échantillons ont été sélectionnés. Il s'agit des échantillons P10C1 et le P30C1, qui serviront de contrôle pour le stress froid, le P10C10, P20C10, P30C10, P40C10 et P50C10, pour étudier les impacts du stress à 28°C et le P50C6 pour le contrôle entre deux stress chauds au dernier passage.

Après l'analyse du P01, l'analyse des séquences issues des échantillons de quatre lignées différentes

(C1, C5, C6 et C10) et des différents passages (P10, P20, P30, P40 et P50), ont été effectuées. Par ailleurs, avant de se concentrer sur l'impact du stress de température sur des potentielles variations structurelles, un tri des mutations obtenues a été réalisé. Ce tri a pu se faire grâce aux fichiers VCF générés par Sniffle.

Le fichier VCF se compose d'un entête qui contient différentes informations sur le run analysé, mais également plusieurs colonnes qui présentent des informations diverses sur les mutations. Dans l'entête est référencée le nom du run, le génome de référence analysé, la date à laquelle le run a été lancé ou encore les filtres déjà appliqués lors du séquençage. En annexe sur le dépôt <https://github.com/marine30190/projetBILL.git> se trouve un exemple d'un fichier VCF de l'un de ces échantillons préalablement sélectionné.

Les colonnes standards sont au nombre de huit minimums et correspondent au chromosome, à la position des mutations, aux identifiants, à la référence de base, à la qualité de la mutation, aux filtres, ainsi qu'aux informations sur la mutation (insertion ou délétion). Dans les fichiers générés pour les échantillons de cette étude d'autres colonnes supplémentaires ont pu être observées. Elles sont en lien avec le type d'analyse, les renseignements du séquençage en long read de Nanopore, comme par exemple, la fréquence allélique des mutations, le nombre de read possédant la mutation et d'autres colonnes. Au total le fichier VCF utilisé est composé de 24 colonnes.

Pour l'analyse des mutations, des filtres ont été appliqués sur certains paramètres afin d'éviter certaines erreurs comme celles liées au séquençage. En



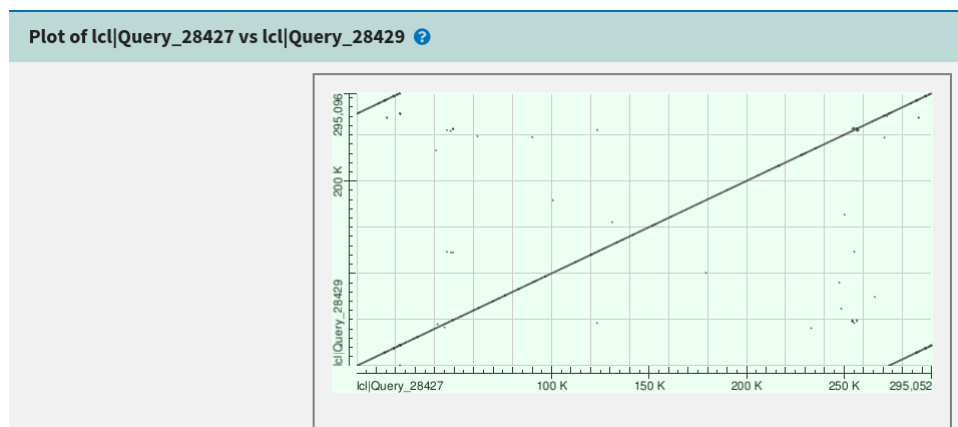


FIGURE 9 – Comparaison de P01 et du génome de référence.

*Ce graphique représente l'alignement entre P01 et le génome de référence.*

effet, les colonnes suivantes ont été utilisées pour trier les mutations :

- La première colonne représente "FILTER", seules les mutations dont la mention "PASS" est affichée ont été sélectionnées, garantissant ainsi, le passage du filtre de qualité.
- La deuxième colonne représente les fréquences alléliques dont le seuil choisit a été à partir de 10% pour corriger le biais lors du séquençage du Q8 qui était de 10% d'erreur acceptée.
- La troisième colonne est "PRECISE", seules les mutations "PRECISE" ont été retenues. Cette colonne représente une estimation de la fréquence allélique et des positions des mutations dans le cas où la mention "IMPRECISE" est écrite.

### 3.5 Résultats obtenus

Lorsque les mutations ont passé les différents filtres choisit une liste de 355 mutations a été obtenue pour nos 8 échantillons sélectionnés. Cette liste, des 355 mutations, est consultable en annexe sur le dépôt github <https://github.com/marine30190/projetBILL.git>. La première analyse effectuée sur ces mutations a été de relever les mutations communes aux échantillons, visible en annexe IV page D au document14. Grâce à ce tri, les mutations présentent avant et après stress chaud et froid ont pu être répertoriées et, à cette liste, a été inclus les contrôles permettant d'exclure des mutations qui ne seraient pas uniquement dues à un stress chaud. Après une deuxième analyse, il a été possible de retenir 23 mutations liées à ce stress, qui sont présentes uniquement chez les échantillons P30C10, P40C10, P50C10 et P50C6. Sur ces 23 mutations, 7 sont des insertions et 16 sont des délétions. (cf. tableau2)

Pour pouvoir affirmer qu'il y ait une présence de

délétion supérieur au nombre d'insertion liée au stress chaud ainsi qu'aux INDELs non liées à un stress quel qu'il soit soit, l'application d'un test statistique a été effectué. En prouvant que les résultats sont significatifs, le test du Khi2 valide l'hypothèse précédente. Le taux de délétion liée au stress thermique à 28°C est de 69.6% et le taux d'insertion est de 30.4%.

Suite à cette conclusion, le choix a été de relever tous les gènes sur lesquels étaient présents les mutations. Pour cela l'utilisation de l'outil IGV et des annotations fournies par le NCBI a été choisi. Les gènes sur lesquels des mutations ont été relevées sont : KHVJ020, KHVJ033 , KHVJ036, KHVJ039, KHVJ078, KHVJ088, KHVJ093, KHVJ107, KHVJ111, KHVJ169, KHVJ168 et KHVJ179. Les autres mutations se situent sur des introns avant les gènes KHVJ023, KHVJ060, KHVJ062, KHVJ109 et KHVJ131. Le choix de garder ces mutations a été fait car selon la position de l'insertion ou de la délétion avant ou après le gène, il est possible qu'elle ait un effet sur le gène tel que, le décalage du cadre de lecture ou bien encore la modification de la protéine par ajout ou suppression d'acides aminés.

A partir de ces gènes, la recherche d'ORF et de leur fonction associée a été effectuée. Le tableau 3 montre le lien entre gènes et ORFs. Cette recherche a été effectuée grâce au site NCBI [https://www.ncbi.nlm.nih.gov/nuccore/NC\\_009127.1?report=graph](https://www.ncbi.nlm.nih.gov/nuccore/NC_009127.1?report=graph).

L'objectif à présent est de pouvoir associer les fonctions potentielles aux ORFs et ainsi émettre des hypothèses sur l'impact et les effets des variations structurales découvertes sur le virus.

Résultats des nombres d'INDELs des mutations relevées :						Test d'indépendance (Khi-deux)	
NON lié au stress	45	insertions	27	délétions	18	Alpha	0,05
lié au stress	23	insertions	7	délétions	16	df	1
total mutations	68	total insertions	34	total délétions	34	Valeur P	0,02
Pvalue = 0,02						Statistique du test	5,32
Résultats significatifs avec le test du Khi2 (res=0.02 et significatif quand p<0.05)						Valeur critique	3,84

TABLE 2 – Tableau récapitulatif du nombre d'INDELs.

Ce tableau récapitule les insertions et délétions obtenues dans les mutations retenues après application des différents filtres avec un test statistique.

Gènes	ORF
KHVJ088	ORF 75, ORF 77 et ORF 78
KHVJ169	proche de l'ORF 01-2 (position : 27300) → proche ORF 1-2 (position : 773100)
KHVJ179 LTR	ORF 7-2 et ORF 8-2
KHVJ020	ORF11 et ORF 12
intron avant le gène 023	proche ORF 16 (position 31288)
KHVJ033	ORF 25
KHVJ036	ORF 27
KHVJ039	ORF 30
intron avant KHVJ060	ORF 49
intron après le KHVJ062	ORF55
KHVJ078	ORF 68 et ORF 69
KHVJ093	ORF 84 et ORF 85
KHVJ107	ORF 97
après le codon stop de KHVJ109	ORF 99
KHVJ111	ORF 101
intron après KHVJ131	ORF 122
KHVJ168	ORF 156

TABLE 3 – Tableau des gènes et leur ORFs associés.

Ce tableau récapitule gènes présentant des mutations sélectionnées ainsi que les ORFs associés à ces gènes.

## 4 Discussion

Bien que les ORFs n'aient pas de fonctions clairement définies, l'étude et la comparaison de ces ORF avec des virus et gènes homologues ont pu permettre d'établir des fonctions hypothétiques. Bien qu'ils ne soient qu'hypothétiques, les ORFs qui présentent des variations structurales peuvent potentiellement avoir un impact direct sur le virus. Malgré cela l'étude de ces mutations et de leurs impacts reste possible. C'est donc le but de cet article et plus précisément de cette discussion. Un autre point important à aborder, est la qualité des échantillons pour la recevabilité des hypothèses et reste une part importante pour toutes études et travaux de recherche.

### 4.1 Les ORFs

Comme le montre le tableau 4 de l'annexe V de la page E, tous les ORFs n'ont pas de fonction clairement définie. Pour cette étude, seuls les ORFs avec des fonctions établies vont être discutées. Le tableau 4 de l'annexe V de la page E permet également de mettre en avant les fonctions principalement liées à une activité de la membrane lipidique du virus. Les ORFs et les gènes retenus codent donc majoritairement pour des protéines membranaires. Au vu de ce choix, une partie du but de cette discussion est d'émettre des hypothèses sur l'impact des variations structurales au niveau de la membrane lipidique du virus, qui est aussi appelée enveloppe virale.

L'enveloppe joue un rôle important pour le virus. En effet, grâce à ses glycoprotéines de surface, le virus adhère à la paroi de la cellule cible, favorisant ainsi la pénétration du CyHV-3 dans la cellule hôte. Par ailleurs, l'enveloppe offre une protection à la capsid en faisant obstacle aux anticorps de l'organisme hôte. En effet, comportant des constituants de la membrane de la cellule hôte, le virus enveloppé est moins reconnaissable par le système immunitaire de l'hôte [6]. Bien que cette membrane soit un élément essentiel au virus, elle reste néanmoins fragile et sensible aux variations environnementales. C'est pour cela que l'étude des mutations sur les gènes impliqués dans la structure de la membrane est intéressante.

De plus, une mutation au niveau de l'enveloppe du virus pourrait influencer sur la virulence de celui-ci. En effet, une mutation dans les gènes codent pour des glycoprotéines transmembranaires, pourrait entraîner une perte d'adhésion du virus à la cellule hôte. De plus, une mutation au niveau de la capsid pourrait induire une perte de la capacité de fusion de la capsid avec la membrane nucléaire de la cellule cible. Cela provoquerait ainsi une diminution de la quantité d'ADN viral injecté dans le noyau, une baisse du taux de réplication et in fine une baisse du taux d'infection des carpes par le CyHV-3.

Cette hypothèse pourrait être appuyée par la diminution, après le stress chaud, du pourcentage

de reads mappés sur le génome de référence.

Lors de la lecture de l'article de Boutier and al, [3], les fonctions hypothétiques des ORFs ainsi que d'autres renseignements ont pu être relevés. Le constat établi est le suivant : sur l'ensemble des ORFs d'intérêts relevés, seul l'ORF78 est conservé au sein de la famille des *Alloherpesviridae*. Les autres ORFs sont supplémentaires au CyHV-3, c'est à dire qui lui sont propres, avec des caractéristiques reconnaissables. [3]

D'après les recherches effectuées sur les fonctions des ORFs sélectionnées, la figure 10 représente la localisation de ces ORFs sur le CyHV-3.

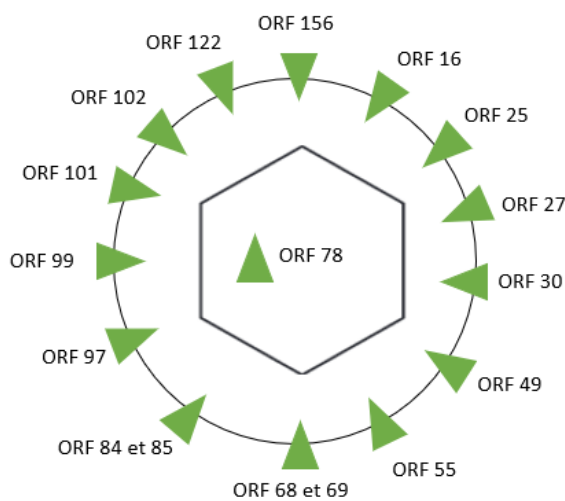


FIGURE 10 – Schéma des emplacements des ORFs sur le virus CyHV-3.

*Ce schéma représente l'emplacement des ORFs retenus sur le virus Cyprinid Herpesvirus 3.*

Dans l'article de BOUTIER and al [3], une partie était consacré à relever le nombre de transcriptions à deux températures différentes. Le constat étant qu'à 30°C le nombre de transcription a diminué, passant à une moyenne de 1 contrairement à la température de 22°C qui était en moyenne au nombre de 4 transcriptions. Cela permet de conforter l'hypothèse de cette discussion, qui est la diminution de la réplication du virus après un stress thermique à 28°C.

## 4.2 La qualité des échantillons

La qualité des échantillons est essentielle pour la fiabilité des résultats, c'est la raison pour laquelle certains échantillons sont à discuter dans tous les travaux de recherche. C'est également le cas dans

cette étude pour 3 échantillons en particulier qui remettent certains résultats en cause.

Concernant la variation structurale relevée pour l'ORF 78, s'agissant d'une mutation uniquement présente sur l'échantillon P50C6, plusieurs questions se posent. En effet pour cet échantillon les mutations relevées sont au nombre de 3, or pour les autres échantillons il y en a, au minimum, dix fois plus. La question se pose concernant la qualité et la remise en question de cet échantillon. Dans le cas où il serait comparé à son homologue de la lignée C10, lignée principalement étudiée pour l'impact du stress thermique chaud, le nombre de mutations est considérablement supérieur. En toute logique, il faudrait s'attendre à un nombre de mutations proches pour les deux passages 50 de la lignée C6 et C10. Un autre problème, qui sera discuté par la suite, de cet échantillon est le mapping des reads sur le génome de référence bien que la quantité soit importante le pourcentage reste faible. Pour ce point l'échantillon homologue P50C10 le constat est le même. Cela ne serait peut être pas due à la qualité de l'échantillon. La remise en cause de cet échantillon est établie et le re-séquençage pour confirmer et valider la mutation serait une option préférable.

Un autre axe de discussion est le résultat des flagstats concernant les échantillons P40C10, P50C10 et P50C6. Comme le montre le document 15 de l'annexe VI page F, à partir de l'échantillon P40 le mapping des reads séquencés ne se mappent plus autant que les autres. Ils sont inférieurs à 54% d'alignement sur le génome de référence contrairement aux autres échantillons qui sont supérieurs à 99% d'alignement. Bien que que ces pourcentages surprennent, la quantité pour le P50C6 et le P50C10 restent correcte avec 113910 et 56074 reads mappés. En revanche, pour le P40C10 le nombre de reads mappés est de 1836 sachant qu'en moyenne pour les autres échantillons, ce nombre est de 40000.

Une des hypothèses qui pourrait également être émise est la suivante, si les échantillons ne sont pas rejetés, est qu'au fil du temps et après un stress thermique à 28°C, la présence du virus diminue. Ce qui pourrait expliquer que lors du séquençage, les cellules du poisson soient plus abondamment séquencées que le virus. Cela entraînerait une diminution du taux de mapping des reads du virus. Toutefois, l'hypothèse d'une erreur de manipulation ou de séquençage, qu'il s'agisse d'une erreur humaine ou matérielle, n'est pas à écarter.

En résumé, les trois hypothèses principales, de l'effet des variations structurales sur le virus, qui découlent de ces analyses sont :

- Une diminution de l'adhérence du virus à la carpe, due à des mutations sur les gènes co-

- dant pour les glycoprotéines membranaires.
- Une perte de capacité de fusion entre la capside et la membrane de la carpe, due à une mutation sur le gène codant pour la capside.
- Une diminution de la présence du virus dans les cellules de carpe, due aux mutations retrouvées après un stress thermique à 28°C.

## 5 Conclusion

Pour rappel, la problématique était d'observer l'impact d'un stress de température à 28°C sur le *Cyprinid HerpesVirus3*. Lors de la discussion, des hypothèses sur l'impact des mutations sur le virus ont été émises, mais aucune conclusion claire n'a pu être établie sans tests complémentaires. Aucune affirmation ne peut être avancée sans la vérification et la confirmation d'un effet positif ou négatif de ces mutations sur le virus.

Dans l'optique d'étendre l'étude, plusieurs pistes sont à étudier.

La première étant de récupérer les reads non mappés des 3 échantillons douteux et d'effectuer un blast sur NCBI par exemple, afin de confirmer que ce qui a été amplifié en plus de l'ADN du virus, soit bien de l'ADN de carpe. Cela permettrait ne de pas remettre en cause la technique d'extraction et de séquençage et de pouvoir émettre d'autres hypothèses.

La deuxième étant d'explorer et de confirmer les fonctions de chaque gène/ORF afin de pouvoir appuyer et confirmer toutes les hypothèses citées dans le paragraphe précédent.

La troisième piste, est d'étudier les protéines membranaires qui constituent une protection supplémentaire au virus augmentant ainsi sa virulence. En effet, dans le but de réduire, voire, d'éliminer le taux d'infection des carpes par le CyHV-3, l'étude des protéines transmembranaires du virus pourrait permettre de comprendre le rôle de chacune d'entre elles, à la fois dans le déplacement, l'adhésion et la pénétration du virus dans la cellule hôte et ainsi réduire la prolifération du CyHV-3.

Plusieurs méthodes peuvent être appliquées afin de confirmer la fonction des ORFs, autrement dit la fonction des protéines membranaires, ainsi que l'impact des mutations sur la virulence du virus. Un exemple de possibilité et de muter volontairement des sites d'intérêt du le CyHV-3, puis grâce à une culture *in vitro* de cellules de carpe, cultiver le virus dans le but de comparer l'activité enzymatique [7] avant et après ajout de la mutation dans le virus. En effet, la culture cellulaire permet d'étudier l'impact des mutations sur le fonctionnement des cellules. Cette expérience est utile pour étudier l'effet des mutations qui affectent

positivement ou négativement la virulence du virus.

De nombreuses méthodes ont vu le jour grâce aux avancées technologiques, voici quelques exemples de ce qui peut être envisagé afin de confirmer et valider les hypothèses citées précédemment.

- L'étude des liaisons entre les protéines membranaires du virus et les molécules de surface des cellules de carpes, représente une piste de recherche qui permettrait de déterminer avec quelles molécules et comment, les protéines de surfaces du CyHV-3 interagissent-elles ? Il pourrait s'agir de ligands, de cofacteurs mais aussi de partenaires d'interaction. Cette information peut être obtenue en utilisant des techniques telles que la spectrométrie de masse, la calorimétrie isotherme titration (ITC) ou la résonance plasmonique de surface (SPR).[8]
- L'analyse de la structure tridimensionnelle de protéines potentiellement impliquées dans la virulence du virus peut être déterminée grâce aux techniques de cristallographie aux rayons X, ou de résonance magnétique nucléaire (RMN). Ainsi, la connaissance de la structure de la protéine peut aider à comprendre sa fonction au sein du virus.[9]
- La modélisation et la visualisation 3D des protéines peut être une méthode. Cela consiste à créer une représentation en 3D d'une protéine à partir de sa séquence d'acides aminés, permettant ainsi de simuler des mutations sur ces mêmes protéines et d'en étudier les effets.[10]

## 6 Bibliographie

### Références

- [1] Sandro KLAFAK, Anna-Sophie FISTON-LAVIER, Sven M. BERGMANN, Saliha HAMMOUMI, Lars SCHRÖDER, Walter FUCHS, Angela LUSIASTUTI, Pei-Yu LEE, Sarahi Vega HEREDIA, Anne-Sophie GOSSELIN-GRENET et Jean-Christophe AVARRE. “Cyprinid herpesvirus 3 Evolves In Vitro through an Assemblage of Haplotypes that Alternatively Become Dominant or Under-Represented”. In : *Viruses* 11.8 (15 août 2019), p. 754. ISSN : 1999-4915. DOI : 10.3390/v11080754. URL : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6723609/> (visité le 15/03/2023).
- [2] *Les virus herpétiques - I-3. Mécanisme de la réplication virale*. URL : [http://unt-ori2.crihan.fr/unspf/2010\\_Marseille\\_Regli\\_Virus\\_Herpetiques/co/I-3\\_Mecanisme%5C%20de%5C%20la%5C%20replication%5C%20virale.html](http://unt-ori2.crihan.fr/unspf/2010_Marseille_Regli_Virus_Herpetiques/co/I-3_Mecanisme%5C%20de%5C%20la%5C%20replication%5C%20virale.html) (visité le 15/03/2023).
- [3] Maxime BOUTIER, Maygane RONSMANS, Krzysztof RAKUS, Joanna JAZOWIECKA-RAKUS, Catherine VANCOSK, Léa MORVAN, Ma. Michelle D. PEÑARANDA, David M. STONE, Keith WAY, Steven J. van BEURDEN, Andrew J. DAVISON et Alain VANDERPLASSCHEN. *Chapter Three - Cyprinid Herpesvirus 3 : An Archetype of Fish Alloherpesviruses*. Sous la dir. de Margaret KIELIAN, Karl MARAMOROSCH et Thomas C. METTENLEITER. T. 93. Academic Press, 1<sup>er</sup> jan. 2015, p. 161-256. DOI : 10.1016/bs.aivir.2015.03.001. URL : <https://www.sciencedirect.com/science/article/pii/S0065352715000743> (visité le 06/03/2023).
- [4] *Herpès Labial : Causes, Traitement Naturel*. Doctonat. Section : Maladies & symptômes. 25 mars 2021. URL : <https://doctonat.com/herpes-labial-traitement-naturel/> (visité le 15/03/2023).
- [5] *Nanopore sequencing technology, bioinformatics and applications Nature Biotechnology*. URL : <https://www.nature.com/articles/s41587-021-01108-x> (visité le 15/03/2023).
- [6] *Enveloppe (virale)*. Sous le microscope. URL : <https://souslemicroscope.com/enveloppe-virale/> (visité le 15/03/2023).
- [7] Sindhu RAVEENDRAN, Binod PARAMESWARAN, Sabeela Beevi UMMALYMA, Amith ABRAHAM, Anil Kuruvilla MATHEW, Aravind MADHAVAN, Sharrel REBELLO et Ashok PANDEY. “Applications of Microbial Enzymes in Food Industry”. In : *Food Technology and Biotechnology* 56.1 (mars 2018), p. 16-30. ISSN : 1330-9862. DOI : 10.17113/ftb.56.01.18.5491. URL : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5956270/> (visité le 15/03/2023).
- [8] Jeremy M BERG, John L TYMOCZKO et Lubert STRYER. “Biochemistry. 5th editio”. In : *New York : NCBI* (2002).
- [9] Laurent CHAMPANEY. “Une nouvelle approche modulaire pour l’analyse d’assemblages de structures tridimensionnelles”. Thèse de doct. École normale supérieure de Cachan-ENS Cachan, 1996.
- [10] Donald VOET, Judith G VOET et Charlotte W PRATT. *Fundamentals of biochemistry : life at the molecular level*. John Wiley & Sons, 2016.

---

## Abstract :

Causing both physical and behavioral symptoms, Cyprinid herpesvirus 3 (CyHV-3) is causing major socio-economic damage in Asian and African countries. The CyHV-3 is listed as a pathogen of concern by the World Organisation for Animal Health (OIE) and the European Union (EU). Co-evolving with carp for many years, CyHV-3 suddenly became a pathogen in 2016 and caught the attention of BILL researchers. They focused on studying the Italian strain (KHV-1) of CyHV-3 which is a 295 kpb double-stranded DNA virus with 156 ORFs. DNA was extracted from samples of different cell culture passages and sequenced to study INDELs-like structural variations following heat stress at 15°C and 28°C. This report focuses on the study of structural variations following heating stress at 28°C. To answer this problem, an experimental protocol was developed. This protocol concerns the steps from the extraction of the virus DNA to the bioinformatics analysis of the sequencing results. Subsequently, an analysis of the results by the application of different filters took place, which allowed to retain a total of 23 mutations. These mutations resulted in the search for genes, their associated ORFs, and their functions. This study allowed, although no conclusion could be clearly defined, to raise an important hypothesis on the impact of a hot temperature stress on the virus. Indeed, this hypothesis concerns the reproduction of the virus in hot stress conditions, which would tend to decline. Concurrently, several methods would allow to confirm this hypothesis and thus to establish a proven function to the ORFs.

---

## Remerciements :

Nous remercions tout d'abord l'équipe pédagogique, qui regroupe les enseignants, les chercheurs ainsi que les ingénieurs qui les épaulent, sans qui ce projet n'aurait pas pu être réalisé. Nous les remercions également pour leurs investissements et leurs réponses à nos nombreuses questions.

Nous tenons également à remercier le binôme, Laurie et Thomas, avec qui nous avons partagé la présentation orale et avec qui nous avons fait nos premières analyses.

De plus, nous remercions Odran qui nous a grandement aidé avec l'utilisation de overleaf pour la résolution d'un problème de codage concernant la bibliographie.

Enfin, un grand merci à tous les participants de ce projet, nos autres camarades, pour les échanges autour de cette thématique.

Cette mutualisation a été, pour nous, une expérience enrichissante et innovatrice.

---



## A Annexes :

Toutes ces annexes sont également disponibles sur le dépôt github du projet : <https://github.com/marine30190/projetBILL.git>

### A.1 Annexe I : pipeline d'analyse bio-informatique

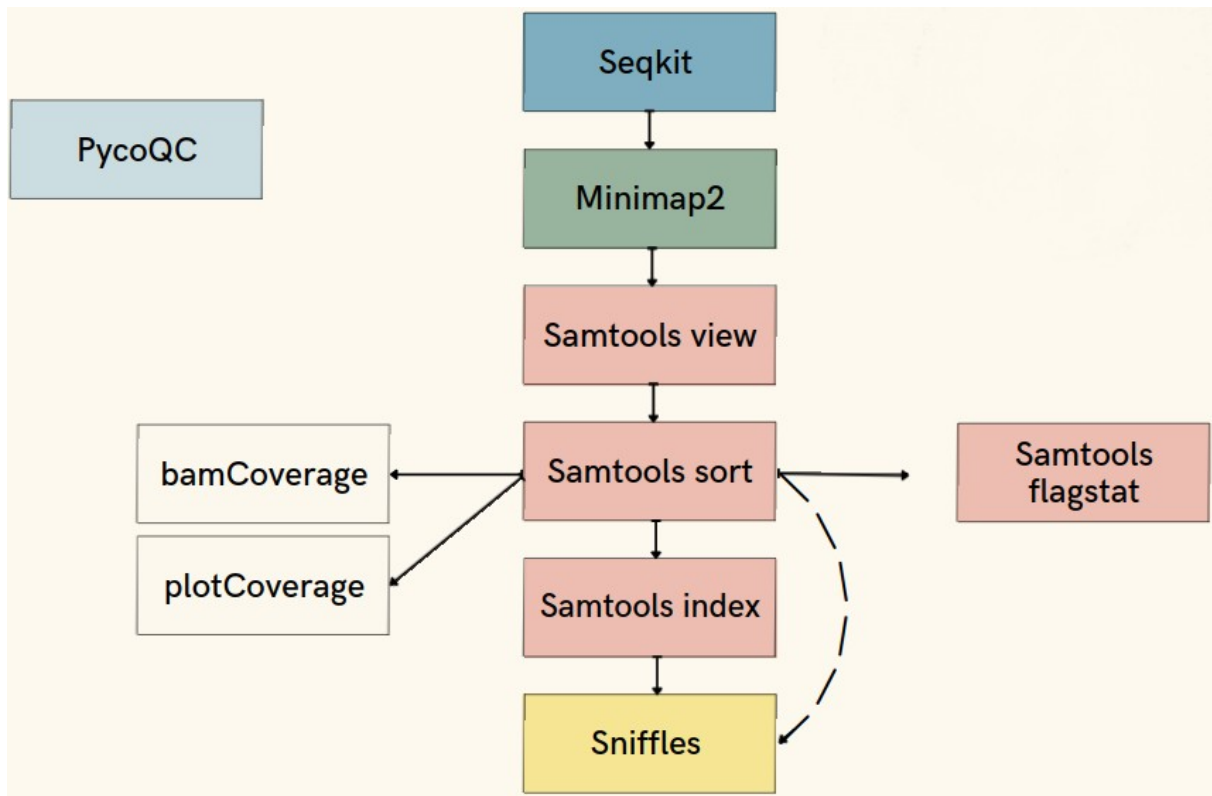


FIGURE 11 – Schéma du pipeline d'analyse bio-informatique.

*Ce schéma représente le pipeline d'analyse bio-informatique utilisé lors de l'analyse des séquences de nos échantillons.*

## A.2 Annexe II : image des profondeurs des échantillons

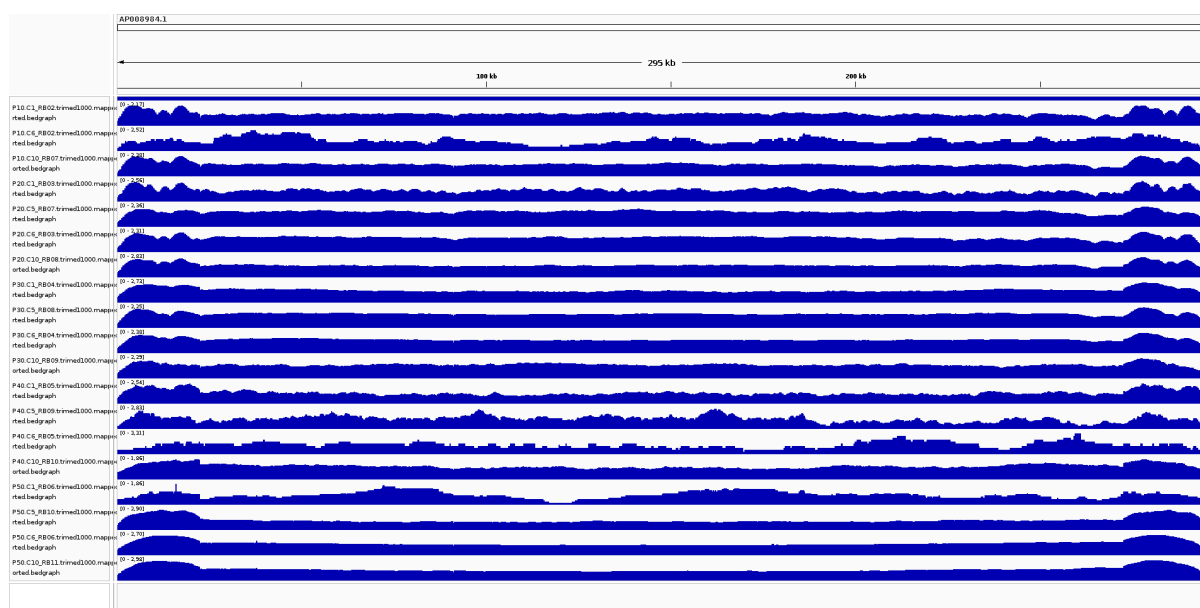


FIGURE 12 – Cette image représentant les profondeurs des échantillons séquencés.  
*Cette image représente toutes les profondeurs de chaque échantillon séquencé comprenant avant et après les stress froid et chaud.*

### A.3 Annexe III : localisation des ORFs

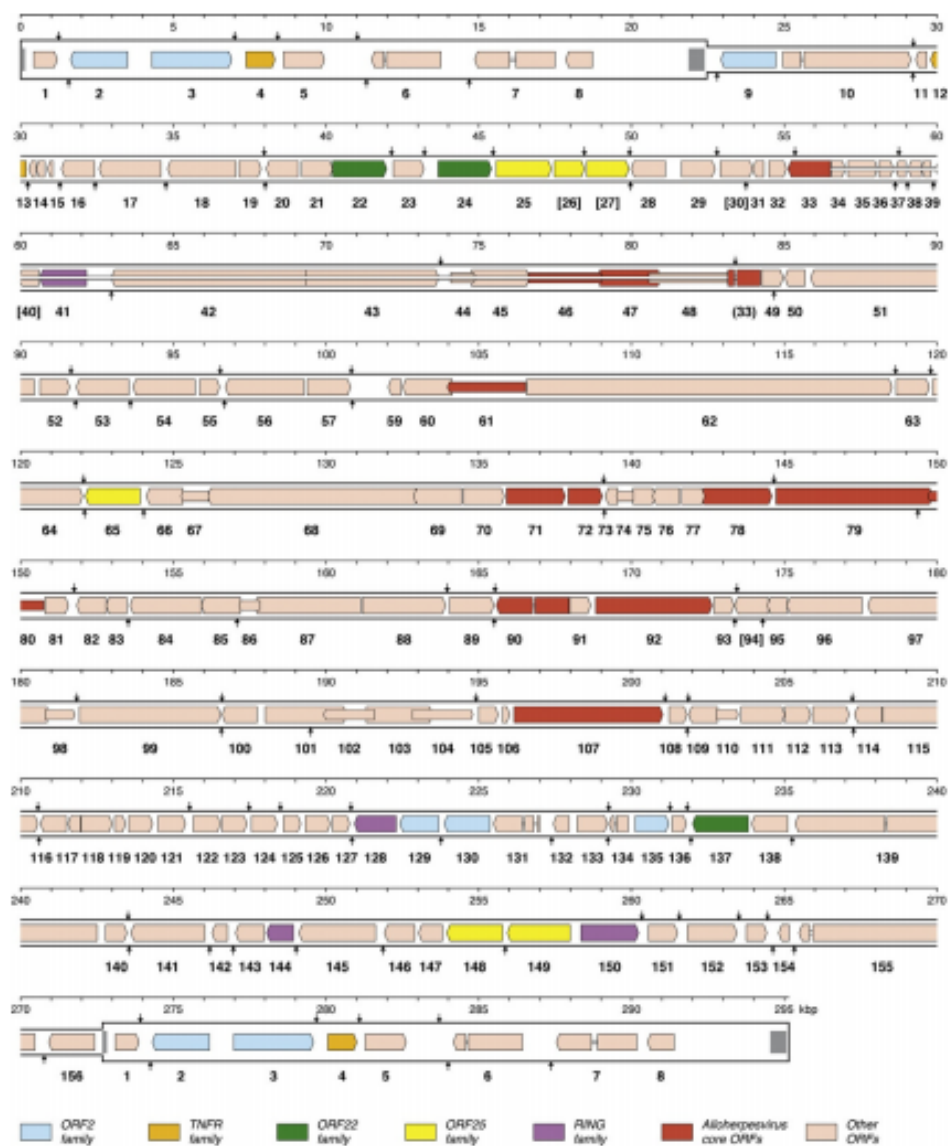


FIGURE 13 – Schéma de localisation des ORFs sur le KHVJ.

Ce schéma représente la localisation sur le génome pour la souche du virus KHVJ. article : BOUTIER

## A.4 Annexe IV : mutations retenues après tri et comparaison

position	P10C1	P30C1	P10C10	P20C10	P30C10	P40C10	P50C10	P50C6	Type de mutation
16050		X		X			X		INSERTION
32535	X	X	X		X	X	X		INSERTION
37855	X	X	X	X			X	X	INSERTION
60341/60342		X	X		X	X	X		INSERTION
70111/70135		X	X		X	X	X		INSERTION
90757		X	X		X	X	X		INSERTION
91777		X	X	X	X	X	X		INSERTION
131128					X	X	X		INSERTION
143580/143594					X	X	X		INSERTION
157828		X	X	X	X	X	X		INSERTION
167879	X		X		X	X	X		INSERTION
203393	X	X			X	X	X		INSERTION
270725				X	X	X	X		INSERTION
272754						X	X		INSERTION
287209							X		DELETION
288688							X	X	INSERTION
114253								X	INSERTION
22309	X				X	X			DELETION
22579/22553/22556/22454	X				X	X			INSERTION
26438/26436					X	X			INSERTION
26440	X								INSERTION
31249					X	X			DELETION
47119					X	X			DELETION
49297/49275					X	X			DELETION
53429/53428					X	X			DELETION
78789/78785	X	X	X		X	X			INSERTION
83023/83027/83028	X				X	X			DELETION
90631					X	X			DELETION
95871					X	X			DELETION
124201	X				X	X			DELETION
127321					X	X			DELETION
155133/154627					X	X			DELETION
165003/164991/164990	X				X	X			DELETION
178680					X	X			DELETION
186412					X	X			DELETION
188510/188513					X	X			DELETION
192410/192408	X				X	X			DELETION
201007	X	X	X	X	X	X			INSERTION
203107	X		X		X	X			INSERTION
216128/216120					X				DELETION
216143					X	X			DELETION
216143	X				X	X			DELETION
271717/271718/271720	X				X	X			DELETION
272143					X	X			DELETION
272123/272125		X			X				DELETION
272871/272880					X	X			INSERTION
272852			X						INSERTION
273975	X				X	X			DELETION
133353	X		X		X				INSERTION
177510	X	X	X		X				INSERTION
11334		X							INSERTION
14372		X							DELETION
49314				X					INSERTION
216398		X							DELETION
219137	X	X	X						INSERTION
254189/254183	X	X	X						INSERTION
284172		X							INSERTION
218452/218433	X		X						INSERTION
240052	X		X						INSERTION
240005	X		X						INSERTION
250497/250499	X		X						DELETION
91102	X								DELETION
120100	X								DELETION
127313	X								DELETION
215027	X								DELETION
210279	X								DELETION
231483	X								DELETION
250507	X								DELETION

Mutations non sélectionnées car pas liées au stress de température chaude

FIGURE 14 – Fichier des mutations retenues.

Ce fichier relève toutes les mutations communes des échantillons. Elles sont triées et classées en deux codes couleurs : blanc = mutations liées au stress thermique à 28°C et gris = mutations non liées au stress thermique à 28°C. Le type de mutation, insertion ou délétion, est également ajouté en dernière colonne.

## A.5 Annexe V : Tableau récapitulatif des fonctions des ORFs

ORF	Fonction estimée
ORF 75, ORF 77 et ORF 78	ORF 77 : Protéine thymidylate kinase utile au métabolisme des nucléotides. ORF 78 : protéase de maturation de la capside
proche de l'ORF 01-2 (position : 27300 et 773100)	protéine hypothétique
ORF 7-2 et ORF 8-2	protéine hypothétique
ORF11 et ORF 12	ORF 12 : Protéine de codage TMD
proche ORF 16 (position 31288)	ORF 16. Protéine transmembranaire Récepteur couplé aux protéines G
ORF 25	Famille ORF25 : glycoprotéine membranaire
ORF 27	Famille ORF25 : glycoprotéine membranaire
ORF 30	- Troncage de la protéine membranaire - "frameshift" (induit codon stop) de 23pb par délétion.
ORF 49	Protéine transmembranaire type 3 M,
ORF55	Thymidine kinase, Changement de cadre une fois en J
ORF 68 et ORF 69	ORF 68 : Similaire à la myosine et aux protéines apparentées et peut-être à l'ORF22 de l'IcHV-1.
ORF 84 et ORF 85 ORF 97	protéine hypothétique protéine hypothétique - similaire à la glycoprotéine Ictalurid (poisson )
ORF 99	- Protéine membranaire Allo46; Related to IcHV-1 ORF46; - Code pour des VTP essentiel à la croissance virale in vitro.
ORF 101	protéine hypothétique
ORF 122	protéine hypothétique
ORF 156	protéine similaire à glycoprotéine GP2 des chevaux

TABLE 4 – Tableau récapitulatif des fonctions des ORFs.

*Ce tableau associe les ORFs à leurs fonctions.*

## A.6 Annexe VI : Résultat des flagstats

Echantillon : P01	Echantillon : P10C1
86411 + 0 in total (QC-passed reads + QC-failed reads)	52996 + 0 in total (QC-passed reads + QC-failed reads)
9414 + 0 secondary	5670 + 0 secondary
222 + 0 supplementary	58 + 0 supplementary
0 + 0 duplicates	0 + 0 duplicates
86162 + 0 mapped (99.71% : N/A)	52763 + 0 mapped (99.56% : N/A)
0 + 0 paired in sequencing	0 + 0 paired in sequencing
0 + 0 read1	0 + 0 read1
0 + 0 read2	0 + 0 read2
0 + 0 properly paired (N/A : N/A)	0 + 0 properly paired (N/A : N/A)
0 + 0 with itself and mate mapped	0 + 0 with itself and mate mapped
0 + 0 singletons (N/A : N/A)	0 + 0 singletons (N/A : N/A)
0 + 0 with mate mapped to a different chr	0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)	0 + 0 with mate mapped to a different chr (mapQ>=5)
Echantillon : P30C1	Echantillon : P10C10
25717 + 0 in total (QC-passed reads + QC-failed reads)	47190 + 0 in total (QC-passed reads + QC-failed reads)
3007 + 0 secondary	5195 + 0 secondary
108 + 0 supplementary	71 + 0 supplementary
0 + 0 duplicates	0 + 0 duplicates
25638 + 0 mapped (99.69% : N/A)	47167 + 0 mapped (99.95% : N/A)
0 + 0 paired in sequencing	0 + 0 paired in sequencing
0 + 0 read1	0 + 0 read1
0 + 0 read2	0 + 0 read2
0 + 0 properly paired (N/A : N/A)	0 + 0 properly paired (N/A : N/A)
0 + 0 with itself and mate mapped	0 + 0 with itself and mate mapped
0 + 0 singletons (N/A : N/A)	0 + 0 singletons (N/A : N/A)
0 + 0 with mate mapped to a different chr	0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)	0 + 0 with mate mapped to a different chr (mapQ>=5)
Echantillon : P20C10	Echantillon : P30C10
40306 + 0 in total (QC-passed reads + QC-failed reads)	4997 + 0 in total (QC-passed reads + QC-failed reads)
4570 + 0 secondary	507 + 0 secondary
111 + 0 supplementary	28 + 0 supplementary
0 + 0 duplicates	0 + 0 duplicates
40284 + 0 mapped (99.95% : N/A)	4982 + 0 mapped (99.70% : N/A)
0 + 0 paired in sequencing	0 + 0 paired in sequencing
0 + 0 read1	0 + 0 read1
0 + 0 read2	0 + 0 read2
0 + 0 properly paired (N/A : N/A)	0 + 0 properly paired (N/A : N/A)
0 + 0 with itself and mate mapped	0 + 0 with itself and mate mapped
0 + 0 singletons (N/A : N/A)	0 + 0 singletons (N/A : N/A)
0 + 0 with mate mapped to a different chr	0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)	0 + 0 with mate mapped to a different chr (mapQ>=5)
Echantillon : P40C10	Echantillon : P50C10
3372 + 0 in total (QC-passed reads + QC-failed reads)	176401 + 0 in total (QC-passed reads + QC-failed reads)
213 + 0 secondary	7629 + 0 secondary
22 + 0 supplementary	195 + 0 supplementary
0 + 0 duplicates	0 + 0 duplicates
1836 + 0 mapped (54.45% : N/A)	56074 + 0 mapped (31.79% : N/A)
0 + 0 paired in sequencing	0 + 0 paired in sequencing
0 + 0 read1	0 + 0 read1
0 + 0 read2	0 + 0 read2
0 + 0 properly paired (N/A : N/A)	0 + 0 properly paired (N/A : N/A)
0 + 0 with itself and mate mapped	0 + 0 with itself and mate mapped
0 + 0 singletons (N/A : N/A)	0 + 0 singletons (N/A : N/A)
0 + 0 with mate mapped to a different chr	0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)	0 + 0 with mate mapped to a different chr (mapQ>=5)
Echantillon : P50C6	
459979 + 0 in total (QC-passed reads + QC-failed reads)	
16487 + 0 secondary	
557 + 0 supplementary	
0 + 0 duplicates	
113910 + 0 mapped (24.76% : N/A)	
0 + 0 paired in sequencing	
0 + 0 read1	
0 + 0 read2	
0 + 0 properly paired (N/A : N/A)	
0 + 0 with itself and mate mapped	
0 + 0 singletons (N/A : N/A)	
0 + 0 with mate mapped to a different chr	
0 + 0 with mate mapped to a different chr (mapQ>=5)	

FIGURE 15 – Fichier des résultats des flagstats.

*Ce fichier relève tous les résultats des statistiques générés par le flagstat pour les échantillons retenus et le P01(lignée de départ).*