



Université de Bourgogne
UFR des Sciences de Santé
Circonscription Pharmacie



N° de thèse :

MÉMOIRE

DU DIPLÔME D'ÉTUDES SPÉCIALISÉES DE BIOLOGIE MEDICALE

Tenant lieu de Thèse pour l'obtention du Diplôme d'État de Docteur en Pharmacie

soutenu publiquement le 27/11/2020

par

BAILLY Eloïse

29/10/1994 à Cosne sur Loire (58)

Evaluation du séquençage haut débit pour l'étude de *Cryptosporidium* spp.

JURY : **Me. le Professeur Catherine Vergely (Président)**

M. le Professeur Frédéric Dalle (Directeur)

M. le Professeur Patrick Callier

M. le Docteur Stéphane Valot

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à la disposition de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur.

Ceci implique une obligation de citation et de référencement dans la rédaction de vos travaux.

D'autre part, toutes contrefaçons, plagiats, reproductions illicites encourent une poursuite pénale.

De juridiction constante, en s'appropriant tout ou partie d'une œuvre pour l'intégrer dans son propre document, l'étudiant se rend coupable d'un délit de contrefaçon (au sens de l'article L.335.1 et suivants du code de la propriété intellectuelle). Ce délit est dès lors constitutif d'une fraude pouvant donner lieu à des poursuites pénales conformément à la loi du 23 décembre 1901 dite de répression des fraudes dans les examens et concours publics.



Université de Bourgogne
UFR des Sciences de Santé
Circonscription Pharmacie



N° de thèse :

MÉMOIRE

DU DIPLÔME D'ÉTUDES SPÉCIALISÉES DE BIOLOGIE MÉDICALE

Tenant lieu de Thèse pour l'obtention du Diplôme d'État de Docteur en Pharmacie

soutenu publiquement le 27/11/2020

par

BAILLY Eloïse

29/10/1994 à Cosne sur Loire (58)

Evaluation du séquençage haut débit pour l'étude de *Cryptosporidium* spp.

JURY :

- Me. le Professeur Catherine Vergely (Président)**
- M. le Professeur Frédéric Dalle (Directeur)**
- M. le Professeur Patrick Callier**
- M. le Docteur Stéphane Valot**

Vice – Doyen : Mme Christine MARIE

Professeurs

CHAMBIN Odile
GROS Claude
HEYDEL Jean-Marie
LESNIEWSKA Eric
MARIE Christine
OFFER Anne-Claire
TESSIER Anne
VERGELY-VANDRIESSE Catherine

Pharmacotechnie
Chimie organique
Biochimie, biologie moléculaire
Biophysique
Physiologie
Pharmacognosie
Physiologie
Physiopathologie, génétique

PU-PH

BOULIN Mathieu
KOHLI Evelyne
GIRODON François

Pharmacie clinique
Immunologie, Virologie
Hématologie

Professeurs Emérites

ROCHETTE Luc
BELON Jean-Paul
LACAILLE-DUBOIS Marie-Aleth

Physiologie
Pharmacologie
Pharmacognosie

Maîtres de Conférences

ANDRES Cyrille
ASSIFAUI Ali
BASSET Christelle
BERARD Véronique
BOUYER Florence
BOUYER Frédéric
CACHIA Claire
COLLIN Bertrand
DESBOIS Nicolas
FAURE Philippe
GUELDRY Serge
GUERRIAUD Matthieu
LEMAITRE Jean-Paul
NEIERS Fabrice
ROCHELET Murielle
SEGUY Nathalie
VIENNEY Fabienne
WENDREMAIRE Maëva

Pharmacotechnie
Pharmacotechnie
Immunologie, hématologie
Pharmacotechnie
Pharmacologie
Chimie physique, Chimie générale
Biomathématiques
Pharmaco-imagerie, radiopharmacie
Chimie organique
Biochimie générale et clinique
Biologie cellulaire
Droit pharmaceutique
Bactériologie
Biochimie, biologie moléculaire, enzymologie
Chimie analytique
Mycologie médicale, botanique
Biophysique
Toxicologie

MCU-PH

FAGNONI Philippe
SAUTOUR Marc
SCHMITT Antonin

Pharmacie clinique
Biodiversité végétale et fongique
Pharmacologie, Pharmacie clinique

PAST

BERNARD Dominique-Alain
CADOT Rachel
CRANSAC Amélie

ATER

BARBIER Elodie
BRUGUIERE Antoine

Chimie analytique
Pharmacognosie



Université de Bourgogne
UFR des Sciences de Santé
Circonscription Pharmacie



L'UFR des Sciences de Santé - Circonscription Pharmacie de Dijon déclare que les opinions émises dans les thèses qui lui sont présentées doivent être considérées comme propres à leurs auteurs, et qu'elle entend ne leur donner ni approbation, ni improbation.

SERMENT

En présence des Maîtres de la Faculté, des Conseillers de l'Ordre des Pharmaciens et de mes condisciples, je jure :

D'honorer ceux qui m'ont instruit dans les préceptes de mon art et de leur témoigner ma reconnaissance en restant fidèle à leur enseignement.

D'exercer, dans l'intérêt de la santé publique, ma profession avec conscience et de respecter non seulement la législation en vigueur, mais aussi les règles de l'honneur, de la probité et du désintéressement.

De ne jamais oublier ma responsabilité et mes devoirs envers le malade et sa dignité humaine.

Que les hommes m'accordent leur estime si je suis fidèle à mes promesses.

Que je sois couvert d'opprobre et méprisé de mes confrères si j'y manque.

REMERCIEMENTS

A Madame le Professeur Catherine VERGELY Présidente de ce jury

Professeur des Universités, Faculté de Pharmacie, UFR des Sciences de Santé, Dijon

Recevez ici l'expression de ma gratitude pour avoir accepté la présidence ce jury et pour l'intérêt que vous avez porté à ce travail.

A Monsieur le Professeur Frédéric DALLE Directeur de ce travail

Professeur des Universités, Faculté de Médecine, UFR des Sciences de Santé, Dijon
Praticien Hospitalier, Laboratoire de Parasitologie-Mycologie Médicale, C.H.U. de Dijon

Je vous remercie pour votre patience, votre attention et votre professionnalisme. Merci pour la confiance que vous m'avez porté au cours de mon internat et tout au long de ce projet. Je vous suis pleinement reconnaissante pour tout cela.

A Monsieur le Professeur Patrick CALLIER

Professeur des Universités, Faculté de Médecine, UFR des Sciences de Santé, Dijon
Praticien Hospitalier, Laboratoire de Cytogénétique, C.H.U. de Dijon

Je vous remercie vivement d'avoir accepté de vous joindre à ce jury et de porter votre attention sur ce travail.

A Monsieur le Docteur Stéphane VALOT

Praticien Hospitalier, Laboratoire de Parasitologie-Mycologie Médicale, C.H.U. de Dijon.

Je tiens à te remercier pour ta disponibilité, ta gentillesse et ton écoute. C'est un réel plaisir de travailler avec toi au laboratoire et également d'avoir pu le faire au cours de cette thèse. Je ne doute pas que je continuerai d'apprendre à tes côtés au cours de cette dernière année d'internat au laboratoire de parasitologie.

A Anne, Véronique et Nadège

Un immense merci pour votre participation à ce projet. Vous avez été d'une grande patience et finalement, on aura pas mal appris ensemble sur le NGS, et ses bonnes et mauvaises surprises. Merci pour votre gentillesse et votre bonne humeur à toute épreuve. On finira par manger ce cookie de la victoire ensemble.

A l'équipe des biologistes du laboratoire de parasitologie – mycologie

Frédéric, Stéphane, Louise, Marc et Christelle. Merci pour votre accueil et d'avoir toujours à cœur de partager vos connaissances. Je vous en suis très reconnaissante.

Aux techniciens du laboratoire de parasitologie – mycologie

Raphaëlle, Isabelle, Corentin, Françoise, Véronique, et Anne. Merci pour votre patience et votre gentillesse. Vous avez rendu mes stages au labo encore meilleurs.

A ma famille

A mes parents. Merci pour votre soutien sans faille, votre patience et votre amour. Vous êtes ma plus grande source d'inspiration et ma plus grande fierté.

A mes grands-parents. Merci pour votre présence et votre attention. Une pensée pour Pépé qui, j'en suis sûre, aurait été fier de voir sa petite fille passer sa thèse.

A Ludo, Estelle, Théophile et Mathilde. Merci pour tous les bons moments passés en famille, et vivement les prochains.

A Fred. Merci pour ta présence et ton amour en toutes circonstances. Tout simplement merci d'être toi.

A mes amis.

A Claire. Merci d'être là pour moi depuis toutes ces années. Quinze ans ça passe vite ! Et ce n'est que le début.

A Caroline. Merci d'avoir été là pendant nos années pharma. Nos soirées révision et nos soirées tout court... De très bons souvenirs et d'autres à venir j'en suis sûre.

A Florian. Merci pour ton écoute et ta gentillesse. Merci pour les soirées tacos, qui ont été un peu perturbées par cette année 2020 un peu particulière, mais ce n'est que partie remise.

A mes co-internes

Merci d'avoir rendu ces années d'internat au PTB inoubliables.

Une dédicace toute particulière au COVID-19 qui a pimenté la préparation de cette thèse.

TABLE DES MATIERES

TABLE DES TABLEAUX.....	9
TABLE DES FIGURES.....	10
LISTE DES ABREVIATIONS	12
INTRODUCTION.....	13
PARTIE I : REVUE DE LA BIBLIOGRAPHIE.....	16
I) Point sur <i>Cryptosporidium</i> spp.....	16
1) Historique et taxonomie.....	16
2) Biologie du parasite	19
3) Epidémiologie.....	22
4) Clinique.....	26
5) Physiopathologie.....	27
6) Diagnostic	28
a) Microscopie	28
b) Recherche d'antigènes.....	31
c) Biologie moléculaire.....	31
7) Traitement et prévention	32
II) <i>Cryptosporidium</i> spp. à l'ère de la génomique.....	35
1) Le séquençage du génome complet	35
2) La caractérisation à l'échelle du sous-type	36
a) Objectifs du sous-typage	36
b) Cibles du sous-typage.....	37
c) Sous-typage de <i>Cryptosporidium</i> spp. par séquençage	39
c-1) Séquençage selon la technique Sanger	40
c-2) Sous-typage de <i>Cryptosporidium</i> spp. par approche NGS.....	41
c-2-a) Généralités sur le NGS	42
c-2-b) Principe du NGS.....	43
c-2-c) Applications générales du NGS.....	48
c-2-d) Le NGS appliqué à l'étude de <i>Cryptosporidium</i> sp.....	50
d) Applications du sous-typage.....	53
PARTIE II : PARTIE EXPERIMENTALE.....	56
I) Objectifs	56
II) Matériels et méthodes.....	57
1) Première partie : Comparaison de deux protocoles de préparation de librairie.....	57
2) Deuxième partie : Etude de mélanges connus d'ADN de <i>Cryptosporidium</i> sp. par NGS...	63

3) Troisième partie : Etude de mélanges d'ADN de <i>Cryptosporidium</i> sp. par PCR en temps réel et par séquençage selon la technique Sanger.....	68
4) Analyse bio-informatique	75
III) Résultats	76
1) Première partie : Comparaison des deux protocoles de préparation de librairie.....	77
2) Deuxième partie : Etude de mélanges d'ADN de différentes espèces et/ou sous-types de <i>Cryptosporidium</i> par NGS.	80
3) Troisième partie : Etude de mélanges d'ADN de <i>Cryptosporidium</i> sp. par PCR en temps réel ciblant le gène codant la sous-unité de l'ARN ribosomal et par séquençage selon la technique Sanger.	84
V) Perspectives	103
BIBLIOGRAPHIE	106
ANNEXES	114

TABLE DES TABLEAUX

Tableau 1 : Liste des espèces valides de <i>Cryptosporidium</i> spp.	18
Tableau 2 : Molécules testées en essais cliniques pour le traitement de la cryptosporidiose.....	34
Tableau 3 : Liste des familles de sous-types existantes	38
Tableau 4 : Principales technologies de NGS existantes et leurs caractéristiques	42
Tableau 5 : Récapitulatif des espèces et sous-types correspondants aux échantillons choisis...	58
Tableau 6 : Récapitulatif des espèces et sous-types correspondant aux échantillons choisis.....	64
Tableau 7 : Schéma de constitution des mélanges 1 à 8 du second essai.	66
Tableau 8 : Proportions en ADN de <i>C. parvum</i> (1770) et <i>C. hominis</i> (1772) dans les mélanges 1 à 8	67
Tableau 10 : Réactifs nécessaires à la PCR en temps réel ciblant le gène codant la sous unité 18S de l'ARN ribosomal.....	71
Tableau 11 : Schéma de constitution des mélanges d'ADN (A à G) pour le troisième essai.	73
Tableau 12 : Présentation des résultats obtenus avec le protocole de préparation de librairie Nextera XT (Illumina) :	77
Tableau 13 : Présentation des résultats obtenus avec le protocole de préparation de librairie 16S Metagenomics Sequencing Library Preparation (Illumina).	79
Tableau 14 : Présentation des résultats obtenus par NGS sur les mélanges 1 à 8 lors du second essai.....	81
Tableau 15 : Présentation des résultats obtenus par NGS sur les échantillons 1791, 2094 et 2369.....	83
Tableau 16 : Présentation des résultats obtenus par PCR en temps réel ciblant le gène codant la sous-unité 18S de l'ARN ribosomal.....	85
Tableau 17 : Schéma de construction des mélanges envisagé après constatation de la présence d'un inhibiteur de PCR dans l'échantillon 2055 pur.....	89
Tableau 18 : Résultats obtenus par PCR en temps réel ciblant le gène codant la sous-unité 18S de l'ARN ribosomal sur les mélanges A à G du troisième essai.....	93
Tableau 19 : Résultats du séquençage selon la technique Sanger des mélanges A à G et des échantillons de Divonne les Bains (D23, D24, D26, D45) lors du troisième essai.....	95

TABLE DES FIGURES

Figure 1 : Représentation du cycle biologique de <i>Cryptosporidium</i> spp. au sein de l'hôte (14)	21
Figure 2 : Oocystes de <i>Cryptosporidium</i> spp. après coloration de Heine (grossissement x1000 microscope à contraste de phase) et coloration de Ziehl-Neelsen modifiée (grossissement x1000 microscope optique).	30
Figure 3 : Schéma reprenant le principe du séquençage par la technique Sanger.	41
Figure 4 : Fragment d'ADN obtenu à l'issue de la préparation de librairie pour NGS	44
Figure 5 : Schéma représentant le principe de la préparation de librairie à l'aide du protocole Nextera XT (Illumina)	45
Figure 6 : Schéma représentant le principe de la préparation de librairie à l'aide du protocole 16S Metagenomics Sequencing Library Preparation (Illumina)	46
Figure 7 : Etapes du séquençage par technique NGS	48
Figure 8 : Schéma présentant les étapes de séquençage par synthèse selon la technologie Illumina®	50
Figure 9 : Etapes du protocole de préparation de librairie pour séquençage NGS Nextera XT (Illumina)	60
Figure 10 : Etapes du protocole de préparation de librairie pour séquençage NGS : 16S Metagenomic Sequencing Library Preparation (Illumina).	63
Figure 11 : Etapes expérimentales de la partie II de Matériels et Méthodes	68
Figure 12 : Schéma montrant l'intérêt de la PCR en temps réel ciblant le gène codant la sous-unité 18S de l'ARN ribosomal pour la validation des échantillons	72
Figure 13 : Schéma illustrant le déroulement de la troisième partie de nos essais	74
Figure 16 : Exemple de fichier .ASV obtenu après séquençage NGS	76
Figure 17 : Migration sur gel d'agarose 2% après PCR « overhang » ciblant le gène codant la protéine gp60 sur les échantillons 1 à 8, 1791, 2094 et 2369.	80
Figure 18 : Courbes d'amplification obtenues par PCR en temps réel réalisée sur les échantillons D23, D24, D26, D45, 1572 et 2055.	86
Figure 19 : Pics de fusion (obtenus à partir des courbes de fusion) issus de la PCR en temps réel réalisée sur les échantillons D23, D24, D26, D45, 1572 et 2055.	87

Figure 20 : Migration des produits de la PCR en point final ciblant le gène codant la protéine gp60 sur gel d'agarose 2%.....	88
Figure 21 : Migration sur gel d'agarose 4% des produits de PCR en point final ciblant le gène codant la protéine gp60 appliquée à l'échantillon 2055 dilué au 1/5e et à l'échantillon 1572 dilué au 1/5000e.....	90
Figure 22 : Courbes d'amplification issues de la PCR en temps réel appliquée aux mélanges d'ADN de <i>C. parvum</i> et <i>C. hominis</i> (A à G).	91
Figure 23 : Pics de fusion issus de la PCR en temps réel appliquée aux mélanges d'ADN de <i>C. parvum</i> et <i>C. hominis</i> (A à G).....	92
Figure 24 : Migration sur gel d'agarose 4% des produits de PCR en point final ciblant le gène codant la protéine gp60 appliquée aux mélanges d'ADN de <i>C. parvum</i> et <i>C. hominis</i> (A à G).	94
Figure 25 : Chromatogramme obtenu par séquençage par la technique Sanger de l'échantillon G en anti-sens (en reverse complément).....	96
Figure 26 : Chromatogramme obtenu par séquençage par la technique Sanger de l'échantillon G en sens.....	96
Figure 27 : Séquence obtenue par séquençage par la technique Sanger de l'échantillon G en sens.....	97
Figure 28 : Séquence obtenue par séquençage par la technique Sanger de l'échantillon G en anti-sens.....	97
Figure 29 : Migration sur gel d'agarose 4% des produits de PCR en point final ciblant le gène codant la protéine gp60 pour les échantillons de Divonne les Bains	97

LISTE DES ABREVIATIONS

ADN : Acide Désoxyribonucléique
ANOFEL : Association française des enseignants et praticiens hospitaliers titulaires de parasitologie et mycologie médicale
ARNr : Acide Ribonucléique ribosomal
C. andersoni : *Cryptosporidium andersoni*
C. canis : *Cryptosporidium canis*
C. cuniculus : *Cryptosporidium cuniculus*
CDC : Center for Disease Control
C. erinacei : *Cryptosporidium erinacei*
C. felis : *Cryptosporidium felis*
C. hominis : *Cryptosporidium hominis*
C.H.U. : Centre Hospitalo-Universitaire
C. meleagridis : *Cryptosporidium meleagridis*
C. muris : *Cryptosporidium muris*
CNR : Centre National de Référence
CNR-LE : Centre National de Référence – Laboratoire Expert
C. parvum : *Cryptosporidium parvum*
C_T : Cycle Treshold
C. ubiquitum : *Cryptosporidium ubiquitum*
ELISA : Enzyme Linked Immunosorbent Assay (Dosage immuno-enzymatique)
FDA : Food and Drug Administration
gp15 : glycoprotéine 15
gp40 : glycoprotéine 40
gp60 : glycoprotéine 60
HSP : Heat Shock Protein (Protéine de choc thermique)
ID50 : Dose infectieuse 50
IFN : Interféron
kDa : kiloDalton
Mb : Mégabase
MLST : Multi-Locus Sequence Typing (Typage multi-loci)
NGS : Next Generation Sequencing (Séquençage nouvelle génération)
OMS : Organisation Mondiale de la Santé
Pb : Paire de bases
PCR : Polymerase Chain Reaction (réaction de polymérisation en chaine)
RT-PCR : Real-Time Polymerase Chain Reaction (PCR en temps réel)
SIDA : Syndrome de l'Immunodéficience Acquise
T_m : Melting temperature (température de fusion)
VIH : Virus de l'Immunodéficience Humaine
WGS : Whole Genome Sequencing (séquençage du génome entier)

INTRODUCTION

Après une émergence bruyante dans les années 80 au cours de la pandémie SIDA (1,2,3,4), l'implication de *Cryptosporidium* spp. dans de nombreuses épidémies d'origine hydrique et/ou alimentaire l'a définitivement propulsée en tête des préoccupations internationales en matière de santé publique et d'environnement (1,2). De même, l'augmentation de la notification des maladies diarrhéiques dans le monde, avec 1.3 millions de décès en 2015, a montré que la cryptosporidiose était la quatrième cause de décès chez les enfants de moins de cinq ans (5,6). Cette année-là, la mortalité par cryptosporidiose représentait alors 12% de la mortalité globale de cette jeune population (5).

Cryptosporidium spp. est un parasite entérique ubiquitaire, qui affecte un large panel d'hôtes, incluant l'Homme et les animaux. Il est en effet retrouvé chez plus de 150 espèces de mammifères, 30 espèces d'oiseaux, 57 espèces de reptiles, 9 espèces de poissons et 2 espèces d'amphibiens (1,7,8,9). On le notifie dans 95 pays, et ce sur tous les continents et sous toutes les latitudes (1).

C'est un parasite hydrique majeur, comptant parmi les plus répandus dans le monde. Entre 2004 et 2016, *Cryptosporidium* sp. était ainsi responsable de 60% des épidémies parasitaires d'origine hydrique au niveau mondial (10,11). On peut par exemple citer la Thaïlande, où 6% des eaux océaniques et 11% des eaux de rivières sont contaminées par *Cryptosporidium* sp. (12). Il représente une problématique mondiale importante dans l'industrie de l'eau, du fait de sa prévalence que l'on vient de décrire, mais aussi de sa résistance aux désinfectants usuellement employés (hypochlorite de sodium, peroxyde d'hydrogène à doses usuelles) (13).

Le risque de contamination est d'autant plus élevé que les réservoirs sont vastes et que les oocystes, infectants par voie orale dès leur émission, sont très résistants dans l'environnement (6,9). En conséquence, cette parasitose se transmet non seulement par voie directe, mais également par voie indirecte, médiée par l'eau et les aliments. Outre sa large prévalence et sa résistance, c'est sa gravité chez les sujets immunodéprimés qui en fait une parasitose au cœur des préoccupations médicales et épidémiologiques (27). Les enfants, les

patients porteurs du VIH, les patients sous traitement immunosuppresseur ou atteints d'hémopathies, les patients porteurs d'un déficit immunitaire congénital sont des populations particulièrement à risque de développer des formes sévères et prolongées de cryptosporidiose (2). La clinique se montre alors intense et réfractaire, pouvant se compliquer par des phénomènes de malabsorption ou des formes extra-intestinales, menaçant potentiellement le pronostic vital (14).

Bien que *Cryptosporidium* sp. soit considéré comme l'une des causes les plus communes de diarrhée parasitaire, les connaissances autour de ce parasite restent à l'heure actuelle incomplète (14). Les thérapeutiques sont encore à ce jour limitées, particulièrement pour les populations immunodéprimées, faisant de la prévention et de la connaissance de la dynamique de transmission, des enjeux majeurs de contrôle de la pathologie (6,14). L'absence de culture axénique a largement freiné les travaux de recherche visant à mieux connaître la biologie du parasite (3,96). C'est la biologie moléculaire qui a su trouver une place prépondérante dans le diagnostic et l'étude épidémiologique de *Cryptosporidium* sp., particulièrement depuis le séquençage du génome de *C. parvum* et *C. hominis* il y a plus d'une décennie (69,70). Les vastes réservoirs animaux, la résistance dans l'environnement de la forme infectieuse, et les lacunes thérapeutiques actuelles sont autant de facteurs qui rendent difficile le contrôle épidémiologique de cette parasitose (1,3). La maîtrise de l'impact de la cryptosporidiose chez l'Homme doit donc passer par des stratégies de prévention et de contrôle solides, qui doivent se baser sur une bonne compréhension de l'épidémiologie de *Cryptosporidium* sp. (1,2,15). Dans ce sens, les études épidémiologiques ont depuis ces dernières années largement recours aux outils de biologie moléculaire, et ce en utilisant des marqueurs génétiques divers. Actuellement, le marqueur génétique le plus employé est le gène codant une glycoprotéine de 60kDa : gp60 (16,17). Ce marqueur est hautement polymorphique et est donc un outil de choix pour l'étude des sous-types de *Cryptosporidium* sp., permettant de répondre aux questions épidémiologiques telles que les dynamiques de transmission, les phénomènes d'hypervirulence, les sources de contamination, ... (16,18,19,94)

Les techniques moléculaires utilisées à ce jour reposent principalement sur de la PCR conventionnelle suivie de séquençage par la technique Sanger (5,16,18). Cette combinaison permet de détecter le génotype majoritaire au sein d'un isolat, mais rarement de caractériser une infection mixte, ce qui est un facteur limitant dans l'étude des épidémies (15,18,20,96). Le

séquençage haut débit ou NGS, une technique permettant le séquençage massif en parallèle de nombreux échantillons, peut nettement améliorer notre compréhension de la taxonomie, de la transmission, et surtout faciliter l'étude des épidémies, notamment les épidémies mixtes (20,96). Cette approche a en effet été employée récemment dans l'étude d'échantillons environnementaux, pour lesquels elle a su montrer une utilité en détectant la présence de plusieurs variants en sein d'un unique isolat. L'objectif du travail présenté ici est de mettre au point cette technique en pratique courante et d'évaluer son intérêt sur des échantillons cliniques.

Nous commencerons par une revue bibliographique qui précise les notions nécessaires à la compréhension de cette étude, puis dans un second temps je vous présenterai les résultats expérimentaux obtenus et les perspectives futures.

PARTIE I : REVUE DE LA BIBLIOGRAPHIE

I) Point sur *Cryptosporidium* spp.

1) Historique et taxonomie

C'est Ernest Tyzzer (1875-1965), parasitologue médical américain, qui décrit la première espèce de *Cryptosporidium* en 1907. Il s'agit de *C. muris*, alors découverte dans les glandes gastriques d'une espèce murine, *Mus musculus* (21). Quatre ans plus tard, Tyzzer rapporte l'existence d'une seconde espèce, *C. parvum*, qu'il distingue de la première espèce par (i) sa localisation dans les cellules intestinales de la souris, (ii) et la plus petite taille de ses oocystes (22). C'est ce même parasitologue qui décrit le premier le cycle du parasite, lui permettant d'établir en 1910 un nouveau genre proche des coccidies : *Cryptosporidium* (3). Bien qu'au fil des années notre compréhension du parasite ait évolué, les publications de Tyzzer ont défini la majorité de ce qui est connu actuellement sur la morphologie et le cycle de vie de *Cryptosporidium* spp. (3). En 1955, la découverte d'une nouvelle espèce, *C. meleagridis*, permet pour la première fois d'associer la présence de *Cryptosporidium* spp. à des signes cliniques chez l'animal (23).

Longtemps considéré comme rare et peu significatif, c'est près de soixante ans après sa découverte, dans les années 1970, que l'on reconnaît le rôle pathologique de *Cryptosporidium* spp. en étudiant des épidémies diarrhéiques mortelles chez des veaux (24,25,26). Les premiers cas humains de cryptosporidiose sont finalement rapportés en 1976 chez des patients présentant des diarrhées aqueuses sévères suite à des travaux collaboratifs entre médecins et vétérinaires (1,3,4,24,27,28).

Quant au premier cas chez un patient porteur de VIH, il est décrit en 1982, suivi de la documentation d'une cinquantaine de cas dans cette population à risque l'année suivante (3,24).

Depuis la description du genre *Cryptosporidium* sp., sa classification taxonomique a fait l'objet de nombreux débats (29). Casemore *et al.* soulignent d'ailleurs en 1990 une confusion dans les publications de l'époque au sujet de la taxonomie du parasite (24). Il est clairement

établi qu'il fait partie du Phylum des Apicomplexa (2,14,30), des entités parasitaires dotées d'un complexe apical impliqué dans l'adhésion et l'invasion de la cellule hôte. C'est également le cas de *Plasmodium* et *Toxoplasma* (8). Néanmoins, *Cryptosporidium* sp. a dès sa découverte été considéré comme une coccidie atypique (31). Le nom de ce genre est d'ailleurs issu de l'absence de sporocystes au sein de l'oocyste, particularité de son cycle de développement qui le distingue des autres coccidies. La microscopie électronique permettait déjà à Tyzzer de constater que l'oocyste était capable de sporuler en étant attaché à la cellule hôte, permettant une autoinfection, le séparant plus encore des autres coccidies. Plus tard, la mise en lumière de sa position intracellulaire particulière et de sa résistance aux traitements classiques des coccidies n'a fait que confirmer les différences déjà mises en évidence (3). L'étude génomique et biochimique de *Cryptosporidium* sp. a permis de constater l'absence de gènes et de voies métaboliques normalement présents chez les membres des Apicomplexa. L'absence d'apicoplaste et de gènes essentiels à la synthèse de protéines mitochondriales l'éloignent un peu plus des coccidies et le rapproche des grégarines (14,31,61). Plus récemment, les techniques de séquençage de génome entier ont conforté cette proximité et il a été proposé de classer *Cryptosporidium* sp. dans la classe des *Gregarinomorpha* et la sous-classe des *Cryptogregarina* (31,32). De plus amples recherches sur ces stades extracellulaires sont nécessaires afin de confirmer ces hypothèses (4,33).

Les règles de taxonomie traditionnellement utilisées pour définir les différentes espèces de *Cryptosporidium* sp. se sont au fil des études avérées insuffisantes (1,9). Ces caractéristiques étaient fondées sur la notion de spécificité d'hôte, la morphologie des oocystes, et la localisation des stades intracellulaires parasitaires (3,34). Ces notions ont entraîné une explosion de la description de nouvelles espèces, celles-ci étant nommées en fonction de l'hôte dans lequel elles étaient découvertes (34). La théorie de la spécificité d'espèce a été contredite dans les années 80 au décours de multiples études de transmission croisée, mettant en évidence qu'une espèce peut être transmise avec succès d'une espèce de mammifère à une autre (1,35). On peut citer en exemple l'espèce *C. ubiquitum*, qui présente une large diversité d'hôtes, tels que les ruminants, les rongeurs, les cerfs, les marmottes et l'Homme (5,36). La morphologie des oocystes est aussi un critère taxonomique majeur au sein du Phylum Apicomplexa. Néanmoins, le fait que les oocystes d'espèces distinctes puissent présenter une morphologie similaire, à l'origine de confusions, explique également l'insuffisance voire l'inexactitude des critères de classification jusqu'alors utilisés (34,37). Il a donc fallu créer des

outils plus performants afin de caractériser les différentes espèces du genre *Cryptosporidium* sp. avec plus d'exactitude. Dans les années 90, de premiers outils comme le Western Blot et le Southern Blot (3), ainsi que la première PCR développée par Laxer et *al.* en 1991 (3) ont permis d'observer une large hétérogénéité génétique qui est devenue une clé de l'amélioration de la classification taxonomique. Les outils de biologie moléculaire ont alors trouvé une place prépondérante dans la classification et la taxonomie. (3) Ils ont notamment permis la reconnaissance des espèces *C. hominis*, *C. canis* et *C. suis* comme des espèces à part entière, longtemps considérées comme des génotypes humains, canins et porcins de *C. parvum* respectivement (1,4). C'est en 2002 que le génotype 1 ou H (humain) de *C. parvum* était officiellement reconnu comme *C. hominis* (4).

A l'issue d'une session tenue à l'Institut Pasteur en 2004, intitulée « La taxonomie du genre *Cryptosporidium* spp. », un consensus est publié réunissant les conditions essentielles à la notification d'une nouvelle espèce de *Cryptosporidium* sp. Ces conditions englobent 1) une étude morphométrique des oocystes, 2) une caractérisation génétique du parasite, 3) L'analyse de la spécificité d'hôte naturelle ou expérimentale, 4) La conformité aux règles de l'«International code of zoological nomenclature » (34).

Le tableau 1 reprend la liste des espèces considérées comme valides selon les critères cités.

Tableau 1 : Liste des espèces valides de *Cryptosporidium* spp. (5)

Espèces	Hôtes	Description
<i>C. andersoni</i>	Bovins	Lindsay et <i>al.</i> 2000
<i>C. baileyi</i>	Oiseaux	Current et <i>al.</i> 1986
<i>C. bovis</i>	Bovins	Fayer et <i>al.</i> 2006
<i>C. canis</i>	Canidés	Fayer et <i>al.</i> 2001
<i>C. cuniculus</i>	Lapins	Inman et <i>al.</i> 1976
<i>C. erinacei</i>	Chevaux	Kvac et <i>al.</i> 2014
<i>C. fayeri</i>	Kangourous	Ryan et <i>al.</i> 2008
<i>C. felis</i>	Félidés	Iseki et <i>al.</i> 1979
<i>C. fragile</i>	Amphibiens	Jirku et <i>al.</i> 2008
<i>C. galli</i>	Oiseaux	Ryan et <i>al.</i> 2003
<i>C. hominis</i>	Humains	Morgan et <i>al.</i> 2002
<i>C. huwi</i>	Poissons	Ryan et <i>al.</i> 2015
<i>C. macropodum</i>	Marsupiaux	Power et <i>al.</i> 2008

<i>C. meleagridis</i>	Oiseaux	Slavin et al. 1955
<i>C. molnari</i>	Poissons	Alvarez et al. 2002
<i>C. muris</i>	Rongeurs	Tyzzler et al. 1907
<i>C. parvum</i>	Humains, bovins	Tyzzler et al. 1912
<i>C. proliferans</i>	Rongeurs	Kvac et al. 2016
<i>C. ryanae</i>	Bovins	Fayer et al. 2008
<i>C. rubeyi</i>	Ecureuils	Li et al. 2015
<i>C. scophthalmi</i>	Poissons	Alvarez et al. 2004
<i>C. scrofarum</i>	Cochons	Kvac et al. 2013
<i>C. serpentis</i>	Reptiles	Levine et al. 1980
<i>C. suis</i>	Cochons	Ryan et al. 2004
<i>C. tyzzleri</i>	Rongeurs	Ren et al. 2012
<i>C. ubiquitum</i>	Rongeurs, primates	Fayer et al. 2010
<i>C. varanii</i>	Reptiles	Pvalasek et al. 1995
<i>C. viatorum</i>	Humains	Elwin et al. 2012
<i>C. wrairi</i>	Cochons d'Inde	Vetterling et al. 1971
<i>C. xiaoi</i>	Ovins, caprins	Fayer et al. 2009

L'éclaircissement de la classification taxonomique est utile à la compréhension de la biologie du parasite, de son importance en santé publique, et de ses dynamiques de transmission (38).

2) Biologie du parasite (14,39)

Cryptosporidium spp. se multiplie dans le tractus gastro-intestinal de l'hôte selon un cycle monoxène mais complexe, comprenant deux stades asexués et un stade sexué. On parle respectivement de schizogonie et de gamogonie. Le cycle débute par l'ingestion ou l'inhalation d'oocystes sporulés de 5µm, infectants dès leur émission dans l'environnement, où ils sont capables de résister plusieurs mois.

Au cours d'une phase d'excystation dans la lumière intestinale, l'oocyste relargue par une suture de la paroi quatre sporozoïtes mobiles qui pénètrent la couche de mucus et s'attachent au bord luminal des entérocytes. Il semblerait que l'excystation soit un phénomène influencé par différents facteurs comme la température et le pH environnants, la présence d'enzymes pancréatiques, de sels biliaries, ainsi que de molécules sécrétées par le parasite.

Comme pour les autres membres du phylum Apicomplexa, la mobilité des sporozoïtes est médiée par le complexe apical du parasite, via un mécanisme actine-myosine-tubuline dépendant. Cela est primordial à l'initiation du cycle au niveau de la cellule hôte, en permettant l'adhérence et l'invasion.

Au cours de l'invasion, le parasite est englobé par la membrane cytoplasmique de la cellule, formant une vacuole parasitophore renfermant le stade intracellulaire : le trophozoïte. La localisation apicale dans la cellule de l'épithélium intestinal fait de *Cryptosporidium* spp. un parasite intracellulaire, mais extracytoplasmique du fait de la présence d'une couche dense aux électrons fabriquée par la cellule hôte qui le sépare du cytoplasme. Au sein de la vacuole parasitophore, le trophozoïte est protégé du milieu intestinal agressif et est nourri par l'intermédiaire d'un organelle nourricier.

Il peut ainsi commencer son cycle de multiplication asexuée. Il subit une succession de divisions mitotiques, qui permettent d'obtenir des mérontes de type I, contenant huit mérozoïtes de type I. Les mérozoïtes de type I, de structure similaire aux sporozoïtes, sortent de la vacuole parasitophore et sont libérés dans l'espace intraluminal du tube digestif. Ils peuvent ensuite envahir les entérocytes voisins, permettant d'augmenter la charge parasitaire. Cette capacité d'amplification de la charge parasitaire explique la faible dose infectante du genre *Cryptosporidium* sp. Au sein de cette nouvelle cellule, le mérozoïte de type I peut subir une nouvelle multiplication asexuée et donner naissance à de nouveaux mérontes de type I. Le mérozoïte de type I peut également suivre une autre voie évolutive, et dériver en un méronite de type II, qui contient quatre mérozoïtes de type II. Ces mérozoïtes de type II initient la phase de multiplication sexuée, et ce en produisant soit un macrogamonte (femelle), soit un microgamonte (mâle). Au moins seize microgamètes sont libérés du microgamonte et chacun peut fertiliser un macrogamète issu du macrogamonte, afin de former un zygote, qui se différencie lui-même en oocyste lors de la sporogonie. La sporogonie permet en fait de faire sporuler quatre sporozoïtes dans l'oocyste, directement au sein de l'hôte, on parle de sporulation in situ.

Les oocystes sporulés sont relargués dans la lumière intestinale. Il résulte de ce cycle monoxène deux types d'oocystes, qui diffèrent par la composition de leur paroi :

- Les oocystes à paroi fine, qui représentent 20% des oocystes produits, sont susceptibles d'excyster dès leur détachement de l'épithélium intestinal, et donc d'entraîner un cycle d'autoinfection.

- Les oocystes à paroi épaisse quant à eux, qui représentent 80% des oocystes, sont excrétés dans les selles et plus largement dans l'environnement.

En résumé, le cycle asexué permet une prolifération du parasite au sein de l'hôte, tandis que le cycle sexué assure l'émission d'oocystes sporulés dans l'environnement. La durée de ce cycle varie de 1 à 22 jours chez l'Homme (1).

Le cycle particulier que l'on vient de décrire, avec la localisation extracytoplasmique du trophozoïte, ainsi que la possibilité d'auto-infection, participe à faire de *Cryptosporidium* spp. une coccidie atypique.

Bien que la culture et l'étude *in vitro* de *Cryptosporidium* spp. soient difficiles, des stades de multiplication extracellulaire ont été décrits au décours du cycle, ce qui rapproche ce genre de celui des grégarines. Ces formes extracellulaires sont encore insuffisamment étudiées et il reste notamment à déterminer leur rôle en physiopathologie.

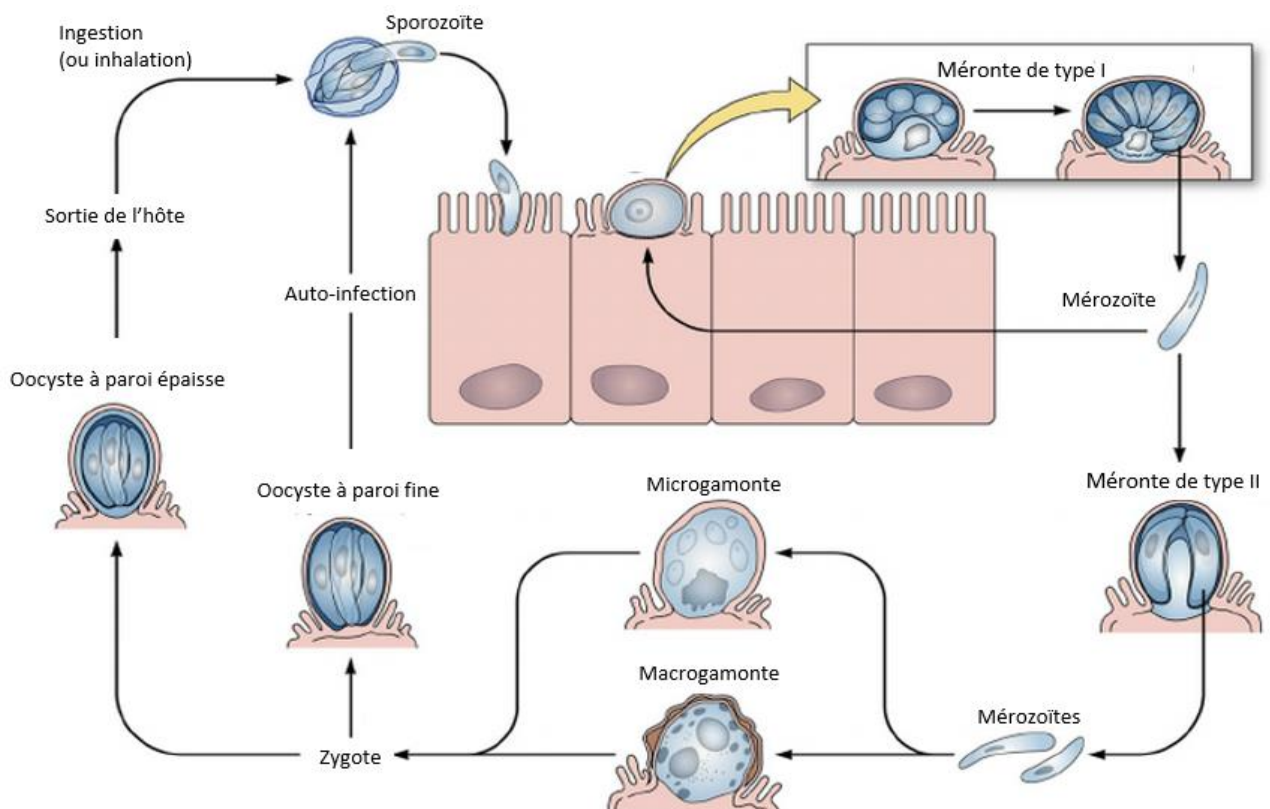


Figure 1 : Représentation du cycle biologique de *Cryptosporidium* spp. au sein de l'hôte (14)

3) Epidémiologie

La cryptosporidiose est contractée par ingestion ou plus rarement inhalation d'oocystes. Ces derniers sont directement infectants dès leur émission et émis en grande quantité (jusque 10^9 oocystes / selle (1)). De plus, la dose infectante peut être faible (moins de 100 oocystes (1,2)) et la résistance du parasite dans l'environnement très importante. Ainsi, la contamination féco-orale peut survenir selon différentes voies : les contacts directs personne à personne, ou animal à personne, et les contacts indirects avec des aliments ou de l'eau souillés par des oocystes, responsables de cas sporadiques ou épidémiques. (1,2,8,13,18,30,38).

On définit ainsi les facteurs de risque de développer une cryptosporidiose comme : (1,2,14,30)

- Les voyages en zone d'endémie ;
- Le contact avec une personne atteinte de cryptosporidiose ;
- Le contact avec des animaux de la ferme ;
- Le contact avec de jeunes enfants âgés de moins de cinq ans ;
- Le contact avec de l'eau de consommation ou de l'eau de baignade.

La transmission par voie hydrique est une voie de dissémination majeure, dans les pays en voie de développement comme dans les pays industrialisés (9,12,40). Entre 2004 et 2016, la cryptosporidiose représentait 60% des protozooses intestinales médiées par l'eau en Australie, en Europe et aux Etats Unis (10,11,31). Quant au Royaume-Uni, elle y représentait 92% des épidémies de gastro-entérites infectieuses associées à un contact avec une eau de piscine ou un plan d'eau entre 1988 et 2012 (1). Elle représente encore aujourd'hui la cause la plus fréquente de maladie infectieuse médiée par l'eau au Royaume-Uni (13,27). La voie hydrique est également à l'origine d'épidémies historiques. En 1989 à Oxfordshire, une épidémie d'origine hydrique touchait 5000 personnes, entraînant la création d'un groupe d'experts impliqués dans la surveillance de *Cryptosporidium* spp. dans les réseaux d'eau (3) ainsi que la mise en place de membranes de filtration dans ces mêmes réseaux (11). Quatre ans plus tard, en 1993, une épidémie massive impliquait 400 000 personnes à Milwaukee aux Etats-Unis. Une fois encore, la reconnaissance de l'impact de la cryptosporidiose en population générale, à travers une épidémie massive, imposait de se pencher sur des techniques de détection et de destruction de *Cryptosporidium* spp. dans l'eau (4). Ce protozoaire a en fait eu un impact

majeur sur l'industrie de l'eau et sa régulation (4,38,41). En 2011, l'OMS éditait de nouvelles recommandations quant à la qualité des eaux de boisson, en prenant en compte *Cryptosporidium* spp. comme pathogène de référence à surveiller (38,42). La résistance des cryptosporidies aux désinfectants usuellement employés a imposé de rechercher des alternatives en matière de désinfection des eaux. Aujourd'hui la destruction hydrique de *Cryptosporidium* spp. dans les réseaux hydriques repose principalement sur l'exposition de ces eaux à l'ozone et aux UV (1,2,3,4,5,38). Des épidémies d'origine alimentaire ont également été observées. C'est le cas par exemple en Angleterre en 2012 suite à la consommation de salade en sachet contaminée, ou plus récemment à Nantes en 2019 dans un collège suite à la consommation de fromage à base de lait non pasteurisé (1,4,98).

La cryptosporidiose a longtemps été associée exclusivement aux personnes immunodéprimées. Ainsi, en 1980 on retrouvait dans une publication de Bird *et al.* : « Lorsque les mécanismes immunitaires fonctionnent normalement et qu'il n'y a pas d'autres troubles intestinaux, la cryptosporidiose ne semble pas poser de problème et peut être considérée comme une parasitose opportuniste. » (43) La pandémie SIDA est en partie responsable de cette croyance erronée. En effet, après la description du premier cas chez un patient porteur de VIH en 1982, la cryptosporidiose connaît une nette augmentation de sa prévalence au cours de la pandémie SIDA. Le lien établi entre cryptosporidiose et SIDA est alors si fort que cette maladie infectieuse est inscrite sur la liste des infections opportunistes définissant le stade SIDA par le CDC (Center for Disease Control) d'Atlanta en 1993 (3). L'avènement des trithérapies dans les pays industrialisés a permis une large diminution de la prévalence de la cryptosporidiose chez le patient sidéen (1,5,30). Mais cette prévalence reste importante dans les pays en voie de développement chez les patients porteurs de VIH (4,30). Plus tard, il a été reconnu que les immunocompétents pouvaient également souffrir de signes cliniques en lien avec *Cryptosporidium* spp. La population immunodéprimée n'en reste pas moins particulièrement à risque de développer des formes plus sévères. En effet, les facteurs de risque pour les formes graves sont le VIH, les hémopathies, les transplantations, les traitements par immunosuppresseurs, ... autant de causes qui entraînent une diminution significative de l'immunité et particulièrement du taux sanguin de lymphocytes T CD4 (1,2,4,8,38). On estime qu'il existe un risque particulièrement élevé de complications quand le taux de CD4 passe sous le seuil des 100/mm³ (1,14,27,44).

Dès les années 80 des études de population ont montré une répartition bimodale en fonction de l'âge, avec un pic d'incidence dans la population des 1-5 ans (1,2,5,8,14,30), ainsi qu'un second pic, moins important, dans la fourchette 25-40 ans (5). Ces chiffres élevés peuvent s'expliquer chez cette jeune population par une hygiène précaire, un système immunitaire immature, voire une malnutrition dans les pays en voie de développement (14). Dans ces derniers, la cryptosporidiose est d'ailleurs une cause importante de morbi-mortalité chez l'enfant, notamment du fait d'un lien réciproque avec la malnutrition (1,45,89). La maladie se place par ailleurs en seconde place des causes de diarrhées mortelles chez le nourrisson, après le rotavirus (31,71). Le second pic quant à lui, concerne sans doute des parents, particulièrement des femmes (2,13), et confirme que le contact avec de jeunes enfants est un facteur de risque de contamination. L'incidence de la cryptosporidiose est donc variable selon l'âge, le statut immunitaire, mais également selon les saisons (2,3,5,8). Ce caractère saisonnier se caractérise par une dépendance aux fluctuations de températures et de précipitations (1,46). En France, on observe un pic annuel en fin d'été et au début de l'automne (1,2). La cryptosporidiose reste malgré tout une maladie largement sous-estimée. L'OMS classait d'ailleurs *Cryptosporidium* spp. parmi les « parasites négligés » en 2004 (6,38,47). Cette sous-estimation s'explique encore à l'heure actuelle par (i) le caractère asymptomatique ou paucisymptomatique chez les immunocompétents qui ne pousse pas à consulter un avis médical, (ii) l'absence de recherche systématique du parasite dans les selles par tous les laboratoires, ainsi que (iii) le fait qu'il ne s'agisse pas d'une maladie à déclaration obligatoire (1,2,30).

La cryptosporidiose est une maladie cosmopolite, retrouvée sur tous les continents, et causée par un parasite considéré comme l'un des entéropathogènes les plus importants dans le monde. Aujourd'hui, on estime entre 0.5% et 2% la prévalence dans les pays industrialisés et elle peut dépasser 10% dans les pays en voie de développement (1,3). Parmi la vingtaine d'espèces validées par les règles taxonomiques, cinq sont responsables de la majorité des cas humains dénombrés dans le monde : *C. hominis*, *C. parvum*, *C. canis*, *C. meleagridis* et *C. felis* (1,6,14,38). Les espèces *C. hominis* et *C. parvum* se distinguent par leur fréquence élevée, dont la proportion relative varie selon les régions. En France et plus largement en Europe, *C. parvum* est l'espèce majoritairement impliquée, à l'inverse des USA et du Japon où l'on retrouve principalement *C. hominis* (1). Dans certains pays comme l'Inde et le Pérou, où les conditions d'hygiène sont précaires, on retrouve une large majorité de cas liés à *C. hominis* (1,2). *C.*

hominis a pour seul réservoir l'Homme, faisant de cette espèce anthroponotique une source importante de contamination interhumaine, et notamment d'épidémies dans les collectivités et les familles (3,41). Quant à l'espèce *C. parvum*, elle est capable d'être transmise de l'animal à l'Homme. L'existence de cette voie zoonotique a été initialement démontrée en étudiant des infections chez des vétérinaires s'occupant de veaux infectés, en soulignant qu'ils étaient bien atteints par la même souche (6,14).

En France, plusieurs épidémies ont été décrites. La première à Sète en 1998 touchant 150 personnes, la seconde à Dracy le Fort (Saône et Loire) en 2001 impactant 480 personnes, la troisième en 2003 à Divonne les Bains avec 800 cas dénombrés, toutes trois d'origine hydrique (1,99). La dernière importante épidémie en lien avec une contamination du réseau d'eau date de 2019 et a eu lieu à Grasse, causant plusieurs milliers de cas dont 155 ont pu être confirmés au laboratoire du CHU de Dijon (48). En novembre 2017, une épidémie touchant 150 personnes a eu lieu dans un collège de Loire-Atlantique, mettant en évidence la consommation d'un fromage au lait non pasteurisé contaminé par *C. parvum* (98). Les rapports d'investigation de l'Agence Française de Sécurité Sanitaire des Aliments (AFSSA) concernant les trois premières épidémies ont permis de mettre en lumière la nécessité d'améliorer le diagnostic, l'étude épidémiologique et l'estimation du risque en France (49). C'est dans cette optique qu'a été créé en 2004 un réseau de laboratoires, Crypto-ANOFEL sous l'égide d'ANOFEL (Association française des enseignants et praticiens hospitaliers titulaires de parasitologie et mycologie médicale) et avec le soutien de l'InVS (Institut de Veille Sanitaire, aujourd'hui appelé Santé Publique France), comptant 53 laboratoires membres (1,5). En 2017, ce réseau a été consolidé par la création d'un CNR-LE cryptosporidiose par Santé Publique France, dont les missions sont l'expertise, le conseil, la surveillance et l'alerte de la cryptosporidiose en France. En 2018, le rapport d'activité du CNR-LE Cryptosporidiose en France rapportait 323 cas en 2017, dont 3 décès. La majorité des cas rapportés concernait des enfants de moins de cinq ans, et suivait une répartition estivale, en rapport avec une plus grande exposition aux eaux de boisson et de baignade, ainsi qu'aux crudités mal lavées en été. 36% des patients étaient immunodéprimés, avec comme principale cause d'immunodépression la transplantation d'organes solides. *C. parvum* était retrouvé dans 55% des cas et *C. hominis* dans 42% des cas. Les enquêtes épidémiologiques menées ont permis de mettre en lumière que 24% des patients ont pu contracter l'infection lors d'une baignade, 17% ont contracté leur infection à lors d'un séjour à l'étranger, 14% étaient patient contact d'une personne contaminée, 11% étaient en

contact avec des animaux d'élevage potentiellement infectés, et 8% se sont contaminés en ingérant un aliment souillé (85).

4) Clinique

Chez l'Homme, la cryptosporidiose est principalement responsable d'atteintes du jéjunum et de l'iléon, à l'origine du symptôme principal de la maladie : la diarrhée aqueuse (1,2,6,38). Cette diarrhée peut être profuse et prolongée, consistant en 3 à 6 selles par jour. On peut noter la présence de mucus dans les selles, tandis que la présence d'hématies et de leucocytes est plus rare (1,6).

La diarrhée ainsi que les douleurs abdominales qui l'accompagnent sont généralement la cause de la consultation médicale (14). On peut également noter la présence fréquente de nausées, de vomissements, d'anorexie, de perte de poids ainsi que d'une possible fièvre modérée (6,8,30,38). Plus occasionnellement, des myalgies et une sensation de malaise peuvent être retrouvées (8,38). Les symptômes apparaissent en moyenne sept jours après l'ingestion des oocystes (1,6,38,50). La durée et la sévérité des symptômes dépendent d'une variété de caractéristiques liées au parasite et aux facteurs de l'hôte. Parmi les facteurs d'hôte interviennent le statut immunitaire, l'état nutritionnel et la fréquence d'exposition au parasite (2,6,14). On distingue ainsi classiquement les atteintes de l'immunocompétent et de l'immunodéprimé.

Un patient immunocompétent présente en général un tableau asymptomatique ou paucisymptomatique (30), limité à une gastroentérite spontanément résolutive en neuf jours en moyenne, bien qu'il existe également de plus rares tableaux prolongés notamment chez le jeune enfant (6,50).

Le tableau clinique du patient immunodéprimé est marqué par des symptômes plus sévères et prolongés et un potentiel de récurrence, ayant un impact sur la morbi-mortalité du patient (1,2,14,27). La cryptosporidiose est en fait une pathologie grave de l'immunodéprimé. Elle entraîne alors des tableaux de diarrhées chroniques, qui durent plusieurs mois voire années, et de malabsorption qui peuvent menacer le pronostic vital. Chez les sujets profondément immunodéprimés, une dissémination extra intestinale peut être observée, responsable d'atteintes biliaires, pancréatiques ou respiratoires (1,2,3,8,14,27). Des complications articulaires ont pu être observées chez adultes et enfants pourtant immunocompétents, ainsi que des phénomènes de céphalées et douleurs oculaires récurrentes

(4,6,30,50,51), particulièrement avec *C. hominis* suggérant une plus grande sévérité de cette espèce (2).

5) Physiopathologie

La majorité de ce qui est connu en termes de physiopathologie résulte de travaux sur des modèles d'animaux immunodéprimés ou néonataux, et sur des modèles *in vitro* de cultures cellulaires épithéliales (8). Les mécanismes physiopathologiques par lesquels le parasite induit des diarrhées ne sont pas encore clairement établis. Les trois principales pistes explorées en recherche sont : (6,8)

- La malabsorption résultant en une diarrhée osmotique ;
- La génération induite par le parasite de produits inflammatoires et de neurotransmetteurs sécrétagogues ;
- La sécrétion d'une entérotoxine.

Le parasite entraîne l'afflux de cellules inflammatoires au niveau de la muqueuse intestinale, aboutissant à terme à une atrophie des villosités intestinales ainsi qu'à une hyperplasie cryptique. Des modèles animaux ont également permis la mise en évidence d'une baisse des échanges ioniques à travers la muqueuse de l'iléon, à l'origine des phénomènes de malabsorption (52). Vingt-cinq facteurs de virulence ont été par ailleurs identifiés par des techniques immunologiques ou moléculaires, intervenant à des étapes clés du cycle biologique du parasite, comme l'attachement à la cellule cible et l'internalisation. Parmi ceux-ci on retrouve des phospholipases, des protéases ainsi que des glycoprotéines de surface (14). La forme trophozoïte dans les cellules intestinales provoque une augmentation de l'expression de NF- κ B, inhibant l'apoptose cellulaire et permettant ainsi au cycle de continuer (1). Les récepteurs Toll-Like (TLR), qui sont des acteurs majeurs de reconnaissance des pathogènes, auraient eux aussi un rôle probable dans la physiopathologie de la cryptosporidiose (53). La gravité accrue chez l'immunodéprimé et la résolution des signes par la restauration immunitaire soulignent le rôle prépondérant de l'immunité dans la physiopathologie de la cryptosporidiose. L'immunité innée a un rôle important dans le contrôle de l'infection, particulièrement chez les enfants. Parmi les effecteurs de cette immunité innée on retrouve les cellules NK (natural killer) qui induisent l'apoptose cellulaire. La lectine liant le mannose, un

effecteur moléculaire de l'immunité innée, empêche quant à elle l'adhésion du parasite à la cellule hôte (54). L'immunité adaptative n'en reste pas moins essentielle à l'élimination du parasite. Un déficit immunitaire acquis ou congénital en cellules T entraîne une plus grande sensibilité à la cryptosporidiose (55). Les lymphocytes T CD4 sont indispensables à la clairance du parasite, et ce par la voie Th1 sous l'influence de l'effecteur moléculaire IFNgamma (2).

6) Diagnostic

La symptomatologie de la cryptosporidiose est commune à de nombreuses maladies. La diarrhée, l'asthénie, ainsi que les complications biliaires et respiratoires ne sont pas pathognomoniques de la cryptosporidiose. Le diagnostic de cette parasitose ne peut donc pas être posé sur la seule observation clinique du patient. La recherche biologique du parasite est indispensable au diagnostic. Depuis la reconnaissance irrévocable de l'existence de l'oocyste en 1978 (3), le diagnostic est en premier lieu microscopique avec la mise en évidence de cette entité. Cette recherche a lieu principalement dans les selles, mais peut également être réalisées sur d'autres natures de prélèvements tels que des liquides biliaires, des biopsies duodénales voire des prélèvements respiratoires en cas de suspicion d'atteinte de l'arbre pulmonaire. Des techniques immunologiques et de biologie moléculaire font également partie intégrante de la démarche diagnostique de la cryptosporidiose.

Une prescription spécifique est nécessaire à la recherche de *Cryptosporidium* sp. car les oocystes sont peu visibles à l'état frais et avec les techniques usuelles d'examen parasitologique des selles. Cela nécessite donc l'utilisation de techniques adaptées à la recherche spécifique des oocystes de *Cryptosporidium* sp.

a) Microscopie

La technique la plus utilisée en France pour la recherche de cryptosporidies est la coloration de Ziehl-Neelsen modifiée (1,2). Celle-ci utilise le caractère acido-résistant des oocystes, qui retiennent la fuschine lors de l'étape de décoloration, leur permettant de se démarquer du reste des constituants de la selle : les oocystes apparaissent en fuchsia sur fond vert (1). Il est préférable que cette coloration soit mise en œuvre après le recours à un procédé de concentration des selles (Technique de flottaison ou technique diphasique comme la

technique Bailanger par exemple). En effet, une concentration préalable permet de limiter les artéfacts à la lecture et d'améliorer la sensibilité de l'examen. D'une taille de 5-6 μ m, les oocystes sont de forme arrondie, avec une paroi épaisse et un contenu granuleux. Ils présentent fréquemment une vacuole qui apparaît plus claire. Bien que très spécifique, cette coloration rapporte une sensibilité variable, de 37% à 90% selon les études (56). Cette technique de coloration permet également la mise en évidence inconstante d'oocystes de plus grosse taille (8-9 μ m), les oocystes de *Cyclospora* spp. qui ne doivent pas être confondus. Notons que la technique de Ziehl-Neelsen classique ne permet pas la mise en évidence des cryptosporidies (1,2).

La coloration à l'auramine-phénol peut également être utilisée pour la recherche microscopique de cryptosporidies. L'auramine est un colorant fluorescent jaune, classiquement utilisé dans le marquage des bacilles acido-alcool-résistants. La lecture d'un frottis teinté avec ce colorant nécessite donc l'emploi d'un microscope à fluorescence, avec une longueur d'onde d'excitation de 350nm et une longueur d'onde d'émission de 450nm (57). Cette technique est très sensible, mais présente plusieurs inconvénients que sont (i) la manipulation de produits toxiques et (ii) une spécificité affaiblie par des artéfacts de fluorescence (56).

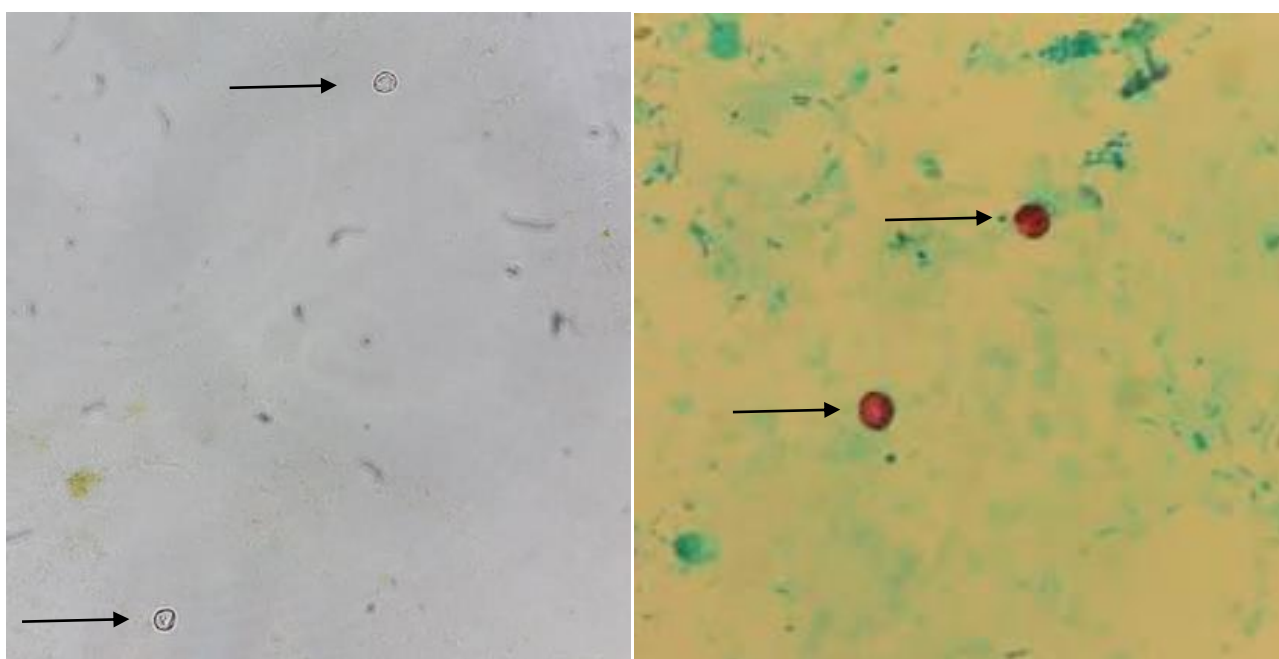
La lecture microscopique en contraste de phase est un outil disponible pour le diagnostic de cryptosporidiose. Elle est mise en œuvre soit directement à partir du prélèvement, soit par l'intermédiaire de procédés en coloration négative, comme celle de Heine. Cela permet la coloration de tous les éléments dans les selles sauf les oocystes de *Cryptosporidium* sp., qui apparaissent alors réfringents au microscope à contraste de phase. Simple à mettre en œuvre et fiable, cette technique nécessite toutefois une lecture rapide de la lame par un œil expert, au risque de perdre le contraste en une quinzaine de minutes (58).

Enfin, il est possible de révéler la présence d'oocystes par immunofluorescence directe ou indirecte, et ce grâce à des anticorps monoclonaux dirigés contre des épitopes de surface, couplés ou non à des fluorochromes (95). Les oocystes apparaissent alors fluorescents sur fond noir. Il s'agit d'une technique présentant de bonnes performances analytiques, avec une excellente spécificité et une sensibilité décrites entre 80% et 100% selon les études (56). Des performances similaires à la coloration à l'auramine ont été décrites pour des échantillons contenant un grand nombre d'oocystes, mais une meilleure sensibilité est observée quand le nombre d'oocystes diminue (57). Cette technique nécessite un équipement spécialisé ainsi

qu'une lecture par un opérateur expérimenté et implique donc des coûts de mise en œuvre plus élevés que les autres techniques microscopiques.

En conclusion, il est important de répéter l'examen microscopique sur trois selles prélevées à plusieurs jours d'intervalle, afin d'augmenter la sensibilité de détection. En effet, un examen microscopique négatif ne peut conclure à l'absence de cryptosporidies sur une seule selle, cela s'expliquant par l'émission intermittente d'oocystes dans les selles et par la sensibilité médiocre de l'examen microscopique. En plus de parfois souffrir d'un manque de sensibilité et/ou de spécificité, les techniques microscopiques ne permettent pas de réaliser un diagnostic d'espèce (30).

Remarque : l'examen anatomopathologique de biopsies intestinales par une coloration à l'hématoxyline et à l'éosine peut mettre en évidence la présence des formes endogènes au niveau du pôle apical des cellules intestinales. On parle « d'aspect en collier de perles ». On constate alors l'altération de la bordure en brosse (6).



Les oocystes sont objectivés par des flèches sur les images. Photographies réalisées au laboratoire de parasitologie du CHU de Dijon.

Figure 2 : Oocystes de *Cryptosporidium* spp. après coloration de Heine (grossissement x1000 microscope à contraste de phase) et coloration de Ziehl-Neelsen modifiée (grossissement x1000 microscope optique).

b) Recherche d'antigènes

Les kits de recherche d'antigènes pariétaux dans les selles sont nombreux à avoir été développés, et on retrouve parmi eux des techniques immunochromatographiques et immunoenzymatiques (ELISA) (6). La spécificité des trousse immunochromatographiques est excellente (> 95%) (5), mais une étude de leur sensibilité par le réseau Crypto-ANOFELa permis de mettre en lumière un défaut de sensibilité. Celle-ci varie de 70% à 90% concernant quatre trousse commercialisées en France, pour la détection des espèces *C. parvum* et *C. hominis*. Cette sensibilité chute à moins de 30% pour les autres espèces (59). Cela pose problème dans les pays où les espèces autres que *C. parvum* et *C. hominis* sont plus répandues. Quant aux trousse utilisant une technique ELISA, elles présentent une bonne sensibilité et une bonne spécificité, mais sont surtout réservées à l'étude de séries importantes (60).

c) Biologie moléculaire

Depuis la première PCR étudiant *Cryptosporidium* sp. en 1991, la biologie moléculaire s'est fait une place de choix dans le diagnostic de cryptosporidiose. A la fois plus sensibles et plus spécifiques que le diagnostic microscopique et immunologique (56), les techniques moléculaires présentent également l'avantage pour certaines d'entre elles de pouvoir identifier l'espèce de *Cryptosporidium* en cause, élément indispensable aux études épidémiologiques (1,2,30). Différentes techniques de PCR ont été mises en œuvre ces dernières décennies telles que la PCR nichée, la PCR quantitative ou la PCR en point final (1,5). La PCR quantitative ou q-PCR, encore appelée PCR en temps réel, présente l'avantage de donner une notion de la quantité relative d'ADN présente dans l'échantillon. Associée à l'utilisation de sondes d'hybridation, elle permet l'identification des espèces par le biais de courbes de fusion, faisant de cette PCR une technique de choix. Le diagnostic de la cryptosporidiose par techniques de biologie moléculaire a longtemps été réservé aux laboratoires spécialisés ou référents, avec comme principaux freins le prix de ces techniques et la difficulté de leur mise en œuvre (nécessité de locaux et d'équipements spécifiques), particulièrement dans les pays en voie de développement (1,2). Le développement plus récent de trousse commerciales a permis un premier pas vers une plus large utilisation de ces techniques.

Les principaux gènes de *Cryptosporidium* spp. ciblés par la biologie moléculaire sont le gène codant la sous-unité 18 de l'ARN ribosomal (ARNr 18SSU), le gène codant la HSP70 (Heat

Shock Protein 70), le gène codant l'actine, ainsi que le gène codant la glycoprotéine 60 (gp60), une glycoprotéine majeure de la paroi parasitaire (41). Les gènes de sous-unité 18S de l'ARN ribosomal, de l'actine et de la HSP70 présentent une faible variabilité intra-espèce mais sont polymorphiques d'une espèce à l'autre, ce qui en fait de bons outils d'étude d'espèces (1). Le gène codant pour la sous-unité 18S de l'ARN ribosomal est le plus largement utilisé pour l'identification d'espèces, comme en atteste le fait qu'on retrouve son utilisation dans 86% des publications concernant le diagnostic d'espèce de *Cryptosporidium* sp. par biologie moléculaire (5,18,94). Le gène codant la protéine gp60, est extrêmement polymorphique au sein d'une même espèce, ce qui explique son intérêt pour la différenciation de sous-types au sein d'une même espèce, et donc son intérêt épidémiologique majeur (18).

7) Traitement et prévention

Bien que la cryptosporidiose ait été décrite pour la première fois il y a plus d'un siècle, des traitements efficaces font toujours défaut (1,30,45). La localisation particulière de la vacuole parasitophore et l'existence de mécanismes métaboliques uniques à *Cryptosporidium* sp. sont probablement en cause dans l'inefficacité des traitements antiprotozoaires classiques sur *Cryptosporidium* sp. (61). Chez l'immunocompétent, les symptômes étant spontanément résolutifs, le traitement se limite souvent à un traitement de support (8,38). Ce traitement purement symptomatique vise à limiter les éventuelles complications d'une diarrhée importante, à savoir la déshydratation et les désordres électrolytiques, principalement chez l'enfant (38). C'est chez l'immunodéprimé que ce manque d'options curatives pose des difficultés thérapeutiques. *Cryptosporidium* spp. étant résistant aux traitements classiques des coccidies (61), il a fallu étudier d'autres classes de molécules afin de trouver des solutions thérapeutiques. Plusieurs molécules ont ainsi été testées comme l'azithromycine et la roxithromycine mais elles ne se sont pas révélées suffisamment efficaces, notamment concernant la clairance du parasite (8,27). La paromomycine, un antibiotique de la classe des aminoglycosides, a montré de bons résultats concernant la durée et l'intensité des symptômes, mais des rechutes ont été observées lors de l'usage de cette molécule (1,4). Cette molécule n'est pas disponible en France et est particulièrement utilisée en médecine vétérinaire (2,4).

Après avoir accumulé pendant des années des preuves d'efficacité et de sécurité chez l'immunocompétent, le Nitazoxanide a, en 2004, obtenu l'autorisation de la FDA (Food and Drug Administration) pour le traitement des adultes et des enfants âgés de plus d'un an (2,13,31,45). Il est aujourd'hui disponible en France sous condition d'obtention d'une ATU (attestation temporaire d'utilisation) nominative, destinée aux patients immunodéprimés (1,2). Le Nitazoxanide a démontré chez l'immunocompétent une efficacité sur l'intensité et la durée des signes cliniques, ainsi qu'une nette amélioration sur les complications à long terme (27,90,91,92). Il permet en effet, chez un patient avec un statut immunitaire fonctionnel, un arrêt des diarrhées dans 80% des cas, et une disparition du parasite dans plus de 60% des cas (2). Néanmoins, chez l'immunodéprimé, les effets du Nitazoxanide ne semblent pas pleinement satisfaisants et dépendent du degré d'immunosuppression (62). Le principal facteur de guérison d'un patient reste à ce jour la restauration de ses capacités immunitaires (8,63). Ainsi chez le patient porteur de VIH, c'est la mise sous antirétroviraux qui conditionne en grande partie la disparition des symptômes associés à la cryptosporidiose et la clairance du parasite (1,2,64).

Connaissant la gravité des atteintes chez les personnes immunodéprimées, le développement de nouvelles molécules reste indispensable, sans oublier le caractère diffusible indispensable aux atteintes extra-intestinales existantes. Des traitements continuent ainsi sans cesse d'être étudiés pour leur potentiel anti-cryptosporidien. C'est le cas ces dernières années de la clofazimine (antibactérien antilépreux) qui a été étudiée en essais cliniques sur des patients porteurs de VIH et a montré une bonne efficacité et une bonne tolérance (65). Un dérivé de la pipérazine (antihelminthique) a également été étudié dans un modèle murin et montrerait des effets parasitocides améliorant des formes cliniques aiguës et chroniques (66). Un autre exemple est l'étude de l'effet des statines (hypolipémiants) dans un modèle murin, qui montreraient un effet synergique avec le Nitazoxanide en privant le parasite de lipides essentiels à son métabolisme (67). Plus récemment, une équipe de l'INSERM a obtenu des résultats concluants in vivo sur une molécule nommée AN3661, ciblant la maturation des ARN messagers du parasite, montrant une bonne efficacité lors d'essais précliniques sur un modèle murin (101). Pour l'heure, le Nitazoxanide reste la molécule de choix, étant celle qui a montré les meilleurs résultats lors d'essais cliniques. L'amélioration des connaissances génétiques du parasite pourra sans doute orienter la recherche en termes de cibles thérapeutiques (30). Le

tableau 2 regroupe quelques-unes des molécules testées pour le traitement de la cryptosporidiose (102,103).

Tableau 2 : Molécules testées en essais cliniques pour le traitement de la cryptosporidiose.

	Molécules testées	Niveau de recherche	Résultats obtenus
Thérapeutiques ayant subi des essais cliniques	Nitazoxanide	Approuvé par la FDA et ATU nominative	Pas d'effet chez les patients très immunodéprimés
	Paromomycine	Multiplés essais cliniques	Effets variés selon les études, rechutes fréquentes
	Azithromycine	Un essai clinique publié	Réponse clinique mais persistance des oocystes
	Rifamycine	Plusieurs essais cliniques	Fonctionne chez les patients porteurs du VIH
	Roxithromycine	Multiplés essais cliniques	Réponse clinique mais persistance des oocystes
	Anticorps anti CD40	Un essai clinique publié	Rechutes après arrêt du traitement

Les options thérapeutiques limitées laissent une place d'autant plus grande à la prévention, individuelle et collective. La protection individuelle passe en premier lieu par la lutte contre le péril fécal. Il s'agit donc d'avoir une bonne hygiène des mains, de ne pas consommer de fruits et légumes crus mal lavés, de ne pas consommer d'eau non stérilisée en zone de forte endémicité, et de ne pas entrer en contact avec des personnes ou animaux susceptibles d'être contaminés (100). Il convient de respecter ces règles en population générale, mais surtout en population à risque. En 2013, l'OMS insistait sur l'importance de disposer d'une eau de boisson sans risque sanitaire, principalement pour les patients immunodéprimés, et notamment porteurs du VIH (68). Quant à la protection collective, elle doit prendre en compte la résistance des oocystes dans l'environnement, et surtout leur résistance aux désinfectants usuellement utilisés. Elle s'élabore donc autour d'un contrôle strict des zones de captage d'eau, de l'intégrité des points de distribution d'eau destinés à la consommation humaine, et de l'utilisation de techniques de décontamination efficaces de l'eau domestique. La filtration des réseaux et le dénombrement des oocystes de *Cryptosporidium*

spp. dans l'eau sont des outils de contrôle sanitaire, adoptés par de nombreux pays suite à des épidémies majeures comme celle de Milwaukee, et qui ont su montrer leur efficacité (4,38,41). Il est également primordial d'informer les professions à risque tels que les personnels de crèche, les professionnels de santé, les vétérinaires et les agriculteurs quant aux bons gestes à adopter (2).

II) *Cryptosporidium* spp. à l'ère de la génomique

1) Le séquençage du génome complet

La recherche génomique sur *Cryptosporidium* spp. a débuté à la fin des années 90 aux Etats-Unis, avec la mise en place d'un consortium d'universités, dont l'objectif était de pallier le manque de thérapeutiques et d'outils épidémiologiques, en mettant la biologie moléculaire au service de la recherche (5). C'est dans cette optique qu'ont été séquencés pour la première fois les génomes de *C. parvum* et *C. hominis* en 2004 (4). Ce sont les isolats *C. parvum* IOWA et *C. hominis* TU502 qui ont ainsi été séquencés par technique Sanger, sur la base d'ADN extrait d'oocystes issus de veaux « germ free » infectés par *Cryptosporidium* (5,69,70). Les génomes de *C. parvum* et *C. hominis*, d'une longueur totale similaire de 9Mb (4,31,71), s'organisent en huit chromosomes dont la taille varie entre 0.88 et 1.34Mb (4,31,61,71). Ces génomes compacts se sont révélés très similaires, avec 97% de similitude de leurs séquences nucléotidiques, et une organisation identique de leurs loci (72). *Cryptosporidium* sp. est dépourvu de certains gènes typiques des Apicomplexa : c'est le cas de gènes impliqués dans la synthèse de protéines mitochondriales, dans des fonctions oxydatives, dans la synthèse d'acides aminés et de sucres. (14,31,86). La petite taille du génome par rapport à celui des autres Apicomplexes (Par exemple *Toxoplasma gondii* : 65Mb) explique pourquoi *Cryptosporidium* sp. est en grande dépendance métabolique vis-à-vis de son hôte, et produit de l'énergie principalement par glycolyse (31,61).

Malgré leur proximité en termes de séquences nucléotidiques, *C. parvum* et *C. hominis* possèdent de nombreux traits phénotypiques distincts, comme le panel d'hôtes et la virulence (5). Il a longtemps été pensé que les différences phénotypiques majeures entre ces deux espèces reposaient sur des modifications mineures (notamment dans la longueur des mini- et

micro-satellites et dans la régulation des gènes) (5,71). Mais récemment, en utilisant des techniques de séquençage de génome entier (WGS), différentes insertions et délétions importantes ont été mises en évidence dans le génome de *C. parvum* qui n'étaient pas présentes chez *C. hominis*. De même, des familles multigéniques spécifiques de *C. parvum* ont été découvertes. Sachant que des gènes membres d'une famille multigénique jouent un rôle prépondérant dans les fonctions biologiques, cela pourrait expliquer les différences phénotypiques et notamment le réservoir d'hôtes plus vaste de *C. parvum* par rapport à *C. hominis* (71,72).

D'autres espèces ont vu leur génome séquencé plus récemment. C'est le cas de *C. meleagridis* et *C. andersoni*. Les génomes des espèces de *Cryptosporidium* spp. étudiées sont à l'heure actuelle disponibles sur la base de données CryptoDB (73). Les techniques de séquençage de génome entier appliquées à *Cryptosporidium* sp. ont largement facilité le développement du génotypage, du sous-typage et des outils comme le typage multi-locus (MLST) pour réussir à caractériser la transmission de *C. hominis* et *C. parvum* (5,74).

2) La caractérisation à l'échelle du sous-type

a) Objectifs du sous-typage

Les connaissances actuelles sur la taxonomie et l'épidémiologie de *Cryptosporidium* sp. sont basées sur des outils de biologie moléculaire, qui caractérisent le parasite à l'échelle de l'espèce et du sous-type. Les thérapeutiques étant encore limitées, la compréhension épidémiologique du parasite est surtout un enjeu majeur pour assurer le contrôle de la transmission de la pathologie. La caractérisation des espèces de *Cryptosporidium* sp. au rang du sous-type permet de répondre à ces enjeux. Le sous-typage s'inscrit aujourd'hui dans la routine des laboratoires spécialisés et permet la caractérisation de la transmission de *Cryptosporidium* sp. chez l'Homme et l'animal. Il permet une meilleure compréhension de l'importance de *Cryptosporidium* sp. en santé publique, en précisant les sources de contamination, les dynamiques de transmission, les disparités géographiques et de virulence (1,5,18). A l'heure actuelle, le sous-typage a principalement recours à l'étude d'un seul locus en séquençant partiellement le gène codant la gp60 (75,78). Il est le marqueur le plus étudié dans la plupart des publications à visée épidémiologique (75,78).

b) Cibles du sous-typage

Le sous-typage repose sur l'analyse d'un ou plusieurs loci hautement polymorphiques. Des mini- et micro-satellites et d'autres cibles hautement polymorphiques ont pu être mis en évidence ces dernières années à partir du séquençage du génome entier de *Cryptosporidium* sp., faisant de ceux-ci des marqueurs de choix pour le sous-typage (69,70).

C'est le séquençage de l'ADN du gène codant pour la gp60 qui est le marqueur de sous-typage de *Cryptosporidium* sp. le plus employé à l'heure actuelle, du fait de son haut degré de polymorphisme (1,2,5,18,41,76). En juin 2019, au cours de l'IGCC (International *Giardia* and *Cryptosporidium* Conference) qui s'est tenu à Rouen, le rôle encore primordial de ce marqueur comme outil épidémiologique a été consolidé (44). En effet, à la différence du gène codant la sous-unité 18S de l'ARN ribosomal, gène très conservé au sein d'une espèce (ce qui en fait un bon outil pour l'identification des espèces), le gène codant la protéine gp60 est très variable, à l'origine de variations intraspécifiques qui en font un bon outil de sous-typage. La glycoprotéine gp60 est aussi connue sous le nom de gp15/40 du fait que cette glycoprotéine de 60kDa est un précurseur protéique, scindé par une sérine protéase en deux glycoprotéines : gp15 (15kDa) et gp40 (40kDa) (5,41,76). Ces deux glycoprotéines sont présentes au pôle apical des formes invasives du parasite, et sont donc directement exposées à la réponse immunitaire neutralisante de l'hôte. La pression de sélection immunitaire exercée sur ces deux protéines pourrait donc expliquer le haut degré de polymorphisme de la protéine gp60 (18,76).

Le gène codant la protéine gp60 contient une séquence microsatellite. Il comprend en effet des répétitions en tandem de trinuécléotides codant pour la sérine (TCA, TCG et TCT) à son extrémité 5'. Des variations dans le nombre de répétitions existent, ainsi que des différences de séquences nucléotidiques dans les régions non répétées (18). Ces variations ont permis de classer *C. parvum* et *C. hominis* dans différentes familles de sous-types. On retrouve ainsi en marge des espèces validées, 11 familles de sous-types pour *C. parvum* et 6 familles de sous-types pour *C. hominis* (1,76). Au sein de chaque famille, les sous-types diffèrent principalement dans le nombre de répétitions des tandems de trinuécléotides (TCA, TCG et TCT) (18,76,77). Le nom du sous-type commence par le nom de la famille de sous-type (Ia, Ib, Ic, ...), suivi du nombre de répétitions des tandems TCA, TCG et TCT représentés par les lettres A, G et T respectivement. Ainsi, le nom du sous-type prédominant de *C. hominis*, IbA10G2, indique que

celui-ci appartient à la famille de sous-type Ib et qu'il comporte dix copies du tandem TCA, et deux copies du tandem TCG à l'extrémité 5' du gène codant la protéine gp60 (18).

Tableau 3 : Liste des familles de sous-types existantes : (18)

Espèces	Familles de sous-types	Tandems majoritaires répétés	Exemples de sous-type décrits
<i>C. hominis</i>	Ia	TCA	IaA23R4
	Ib	TCA, TCG, TCT	IbA10G2
	Id	TCA, TCG	IdA16
	Ie	TCA, TCG, TCT	IeA11G3T3
	If	TCA, TCG	IfA19G1
	Ig	TCA	IgA24
	<i>C. parvum</i>	IIa	TCA, TCG
IIb		TCA	IIbA14
IIc		TCA, TCG	IIcA5G3a
IId		TCA, TCG	IIaA18G1
IIe		TCA, TCG	IIeA12G1
IIf		TCA	IIfA6
IIg		TCA	IIgA9
IIh		TCA, TCG	IIhA7G4
IIi		TCA	IIiA10
IIk		TCA	IIkA14
III		TCA	IIIA18

Les sous-types en gras représentent les sous-types majoritairement représentés en pathologie humaine.

Le recours à l'étude de plusieurs loci tient compte du caractère sexué d'une phase de la reproduction de *Cryptosporidium* sp. La reproduction sexuée entraîne en effet de potentiels réarrangements qui risquent d'être ignorés par l'étude d'un locus unique (74). Le typage d'un seul locus peut en effet ne pas suffire quand il s'agit d'investiguer la structure d'une population de *Cryptosporidium* sp. (74,78), particulièrement quand celle-ci est dense et donc plus sujette aux réarrangements génétiques (17). Le typage multi-locus trouve donc son intérêt en zone endémique, dans l'étude de la dynamique de transmission, de structure de population et d'élucidation des mécanismes d'émergence de souches virulentes, et ce en augmentant la résolution par rapport à l'étude d'un simple locus (78). Bien que de plus en plus utilisé dans les publications, le typage multi-locus ne s'inscrit pas dans la routine des laboratoires, souffrant encore à l'heure actuelle d'un manque de standardisation des protocoles (75). Le choix des

marqueurs et la manière de les choisir restent des questions en suspens qui freinent le séquençage multi-locus et laissent une place de choix au gène codant la protéine gp60 (17,74).

c) Sous-typage de *Cryptosporidium* spp. par séquençage

Le développement du séquençage de l'ADN dans les années 70 a été un enjeu majeur pour la recherche scientifique, avec pour objectifs de mieux comprendre et appréhender les processus physiologiques et pathologiques du vivant. La première génération d'outils de séquençage a vu le jour avec le séquençage selon la technique Sanger, développé par Frédérick Sanger (79,88). Le séquençage selon la technique Sanger s'est amélioré au cours du temps, notamment en remplaçant les radiomarqueurs par des marqueurs fluorescents, et s'est nettement automatisé, avec notamment le remplacement des systèmes de migration des fragments d'acides nucléiques fluorescents en gel par des systèmes capillaires permettant une migration des acides nucléiques fluoromarkés et une meilleure intégration des différentes étapes du séquençage en un seul automate. Plus tard, dans les années 90, a vu le jour la technique de séquençage par pyroséquençage, plus rapide et à moindre coût par rapport à la méthode de Sanger (105). Enfin, dans les années 2000 est apparue une nouvelle génération de séquençage appelée séquençage haut débit ou séquençage nouvelle génération (NGS : « Next Generation Sequencing »). Cette nouvelle génération de séquençage a permis de répondre à la problématique du temps de séquençage, permettant de traiter des millions de séquences en parallèle. Elle a également permis d'améliorer la résolution de séquençage. Il y'a ainsi eu une évolution des années 80 où les outils de séquençage permettaient le séquençage de fragments de 800 à 1000 nucléotides en quelques jours avec des techniques lourdes, jusqu'à aujourd'hui avec des techniques modernes et simplifiées, capables de séquencer des millions de molécules nucléotidiques dans des temps relativement courts (80,105).

Qu'il s'agisse de l'étude d'un ou plusieurs loci, la majorité des publications traitant de *Cryptosporidium* sp. repose sur de la PCR conventionnelle suivie de séquençage par la technique Sanger. Il en est de même dans les laboratoires spécialisés ayant recours au sous-typage de *Cryptosporidium* sp. dans leur activité de routine.

c-1) Séquençage selon la technique Sanger

La technique de séquençage selon la technique Sanger repose sur une méthode par synthèse enzymatique. La polymérisation de l'ADN est initiée par une amorce complémentaire et l'élongation a lieu grâce à une ADN polymérase. Il y'a donc addition de désoxyribonucléotides (dATP, dCTP, dGTP, dTTP) afin d'allonger le brin d'ADN. Néanmoins, afin d'assurer la réaction de séquençage, des didésoxyribonucléotides (ddATP, ddCTP, ddGTP, ddTTP) marqués par un fluorophore et en faible concentration sont incorporés dans la séquence en cours de synthèse. Une fois fixés, ils empêchent la poursuite de l'élongation. On obtient ainsi de manière statistique un arrêt de l'élongation à toutes les positions possibles de la séquence complète, et par conséquent des brins d'ADN de longueur croissante. Ces derniers sont séparés sur gel de polyacrylamide et classés en fonction de leur taille. La détection de fluorophores associés aux didésoxyribonucléotide incorporés permet d'identifier le didésoxyribonucléotide incorporé au bout de chaque fragment de taille croissante, et cela permet finalement de reconstituer l'ensemble de la séquence étudiée (79,88,104). Les différentes étapes de la méthode sont résumées dans la figure 3.

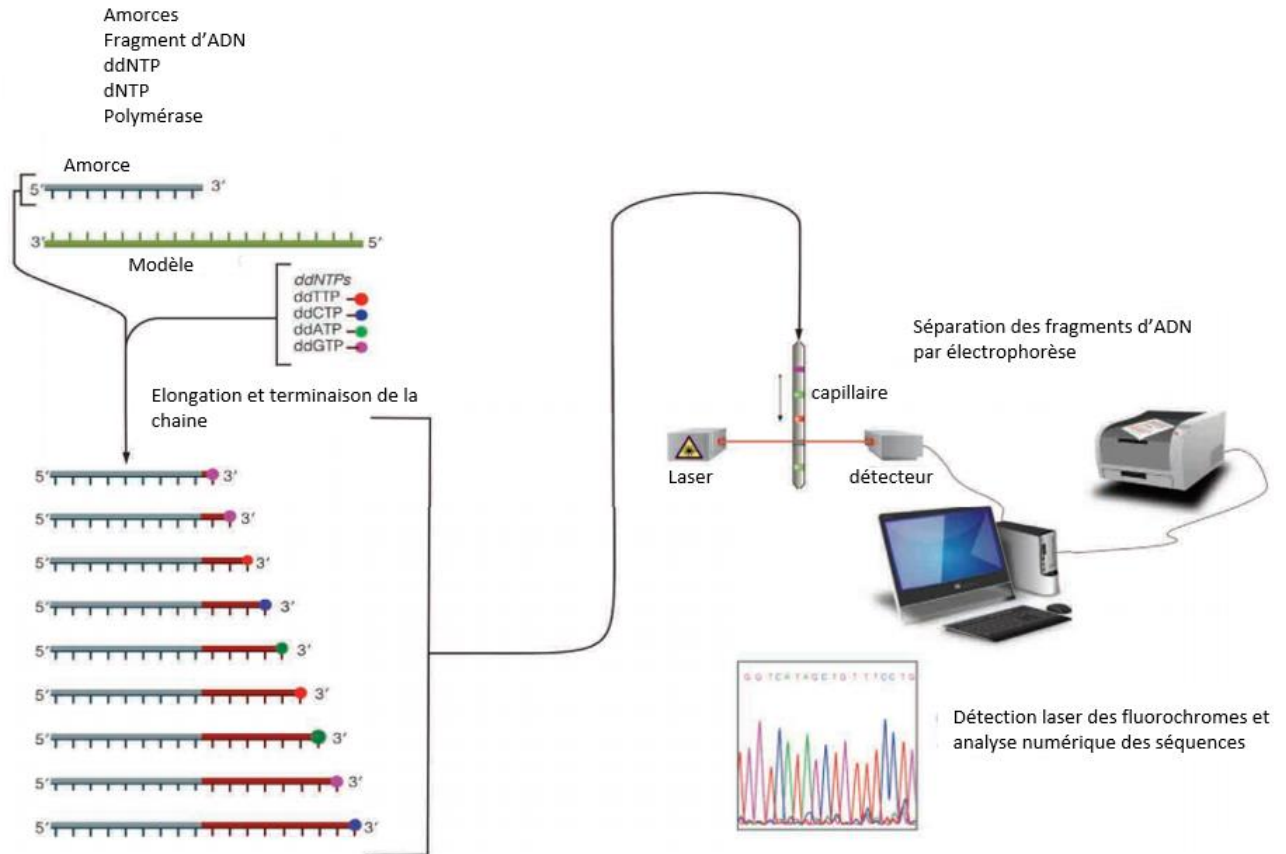


Figure 3 : Schéma reprenant le principe du séquençage par la technique Sanger. (79)

Bien que robustes, ces techniques ne détectent que les variants majoritaires présents au sein de matrices complexes (par exemple, les selles) pouvant contenir un mélange de plusieurs isolats/sous-types. Ainsi, un mélange d'ADN peut être ignoré par ces techniques, dans un échantillon environnemental ou un échantillon clinique (15,20,82).

c-2) Sous-typage de *Cryptosporidium* spp. par approche NGS

Ces problématiques d'une probable sous-estimation du nombre d'infections mixtes et/ou de l'impossibilité d'étudier des mélanges d'isolats/sous-types dans des échantillons environnementaux complexes, notamment hydriques, ont incité les communautés médicales et scientifiques à développer des outils qui permettraient de les contourner. A ce jour, la caractérisation de *Cryptosporidium* sp. dans des échantillons environnementaux par technique NGS semble pouvoir répondre à ces questions (15,82,83,84,87,96).

c-2-a) Généralités sur le NGS

Le NGS (« Next Generation Sequencing » ou séquençage nouvelle génération), aussi appelé séquençage haut débit, correspond à un ensemble de méthodes de séquençage apparues dans les années 2000. Le NGS a recours à des plateformes capables de réaliser le séquençage massif de milliers voire millions de bases en parallèle, apportant entre autres un gain de temps massif par rapport aux techniques de séquençage telles que Sanger. Cette technique se démocratisant dans les laboratoires, des automates plus compacts et donc plus accessibles ont été créés. Certes moins productifs, ces automates sont amplement suffisants pour l'étude de génomes microbiens, par exemple, du fait qu'ils sont 1000 fois plus petit que le génome humain (80).

Il existe de multiples technologies de NGS commercialisées telles que Illumina® (Solexa), Roche 454, SOLiD (Roche Diagnostics), qui diffèrent notamment par la méthode de séquençage utilisée (105). Le tableau 4 regroupe les principales technologies commercialisées.

Tableau 4 : Principales technologies de NGS existantes et leurs caractéristiques (105).

Technologie NGS	Plateforme de séquençage	Méthode de séquençage	Durée de séquençage	Capacité de séquençage
Ion Torrent	Ion Torrent PGM, Ion Proton	Séquençage par ligation (ion semi-conducteur)	2 heures	1 000Mb
Roche 454	Roche 454	Pyroséquençage	20 heures	900Mb
Illumina	MiSeq, HiSeq, TruSeq, NextSeq	Séquençage par synthèse	1-14 jours	200 000Mb
SOLiD	ABI SOLiD, SOLiD 5500, SOLiD 5500xl	Séquençage par ligation	8-12 jours	80 000Mb

c-2-b) Principe du NGS

Quelle que soit la technologie choisie, le NGS présente des étapes communes que sont (105) :

- La collecte des échantillons (sang, tissu, selles, ...) ;
- L'extraction et la purification des acides nucléiques (ADN ou ARN) ;
- La génération d'une banque d'ADN/ARN rendue compatible avec le NGS grâce à la fixation d'adaptateurs sur les fragments d'ADN/ARN étudiés, et l'amplification de cette banque : on parle de préparation de la librairie ;
- Le chargement de la librairie ;
- La synthèse de clusters ;
- Le séquençage.

Dans la suite de l'exposé, nous nous focaliserons sur le séquençage de fragments d'ADN.

Préparation de la librairie

Une étape clé du séquençage par NGS est la construction de la librairie qui correspond au mélange de fragments d'ADN de tous les échantillons destinés à être séquencés par l'automate.

L'ADN préalablement extrait et purifié est fragmenté de manière aléatoire, de façon différente selon le protocole de préparation de librairie. En fonction de l'approche utilisée, cette fragmentation peut se faire par (104,105) :

- Action mécanique (sonication par des ultrasons par exemple) ;
- Action chimique ;
- Action enzymatique (fragmentation et marquage simultanés : on parle alors de tagmentation) ;

Les fragments d'ADN ainsi obtenus doivent être rendus compatibles avec le séquençage NGS. Cela passe par l'ajout d'adaptateurs et d'index aux extrémités des fragments d'ADN. L'index correspond à un code-barre qui permet d'associer le brin sur lequel il est fixé à son échantillon d'origine. Ainsi, quand plusieurs échantillons sont mélangés dans une même librairie, il est facile de savoir à quels échantillons appartiennent les différents fragments d'ADN. Les adaptateurs sont quant à eux des séquences indispensables aux réactions d'amplification et de séquençage ultérieures. En effet, ces adaptateurs permettent aux brins d'ADN de se fixer

spécifiquement à des oligonucléotides présents en surface de la flow cell. Cette dernière est une plaque en verre à la surface de laquelle ont lieu les réactions d'amplification et de séquençage (104,105).

La fabrication de la librairie dépend de la technique choisie, mais le résultat final est le même, et celui-ci est présenté dans la figure 4.

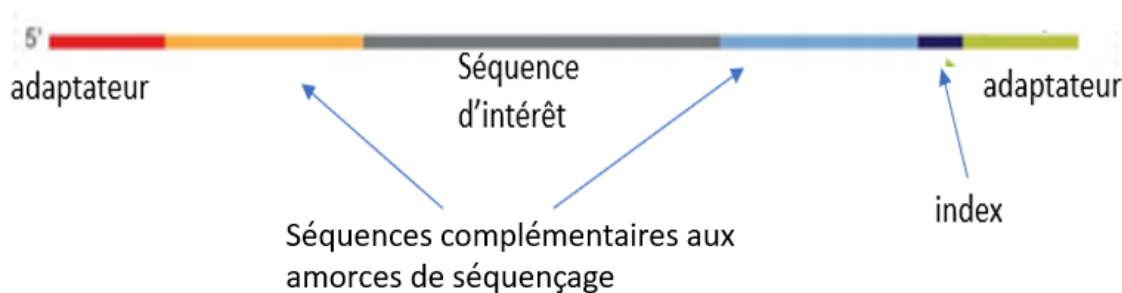


Figure 4 : Fragment d'ADN obtenu à l'issue de la préparation de librairie pour NGS (106)

Nous pouvons prendre l'exemple de deux protocoles de préparation de librairie de principe différent, qui sont deux protocoles auxquels nous avons eu recours dans ce travail.

Le protocole de préparation de librairie Nextera XT (Illumina) utilise une méthode de fragmentation enzymatique de l'ADN, appelée tagmentation, par l'intermédiaire de tagmentases (106). Ces tagmentases, aussi appelées transposomes, sont des enzymes capables de simultanément fragmenter l'ADN (toutes les 300pb environ) et le marquer par ajout d'une amorce à chaque extrémité des fragments d'ADN générés. Ces amorces sont indispensables pour rendre ces fragments compatibles avec le séquençage haut débit, car elles sont complémentaires des amorces de séquençage. Au cours d'une PCR dite d'indexation, un index (code-barre) et un adaptateur (fixation à la flow cell) viennent se lier à chaque extrémité de l'ADN déjà « tagmenté ». La figure 5 schématise le principe de ce protocole. Le résultat final de ce protocole montre que l'on obtient la même chose qu'en figure 4, à savoir un brin d'ADN avec à ses extrémités des adaptateurs, index et amorces de séquençage.

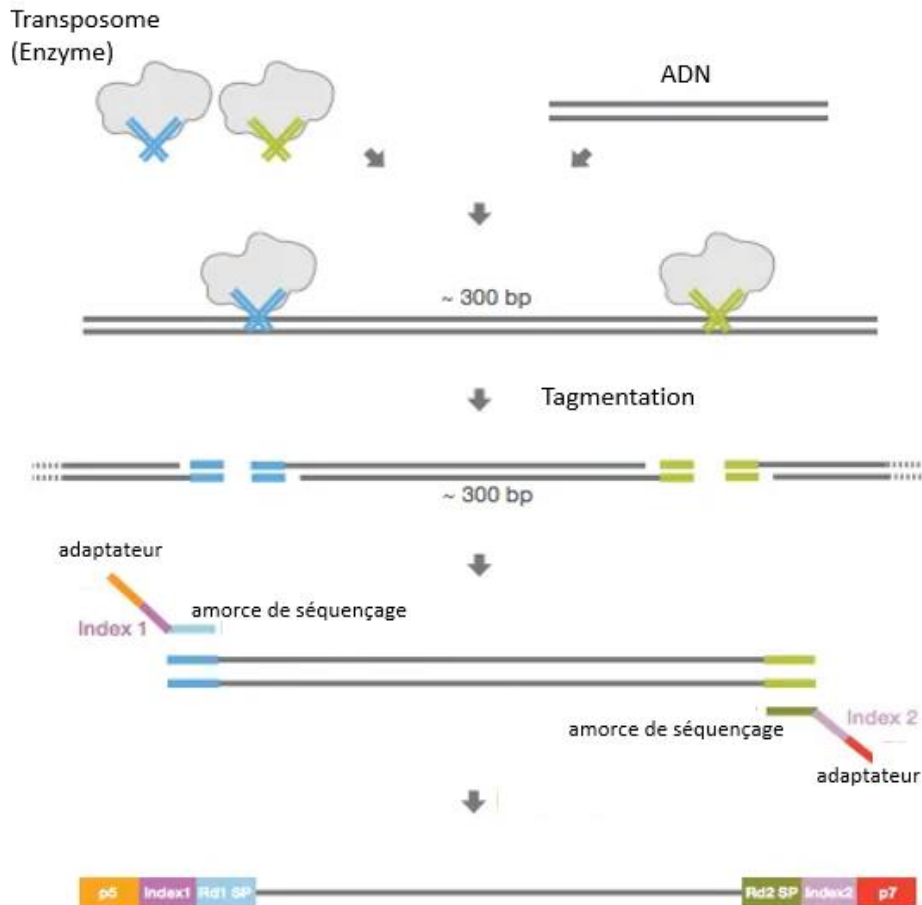


Figure 5 : Schéma représentant le principe de la préparation de librairie à l'aide du protocole Nextera XT (Illumina) (106).

- 1) Fixation des transposomes ou tagmentases sur les séquences d'ADN ;
- 2) Tagmentation toutes les 300pb environ : coupure enzymatique et fixation d'amorces de séquençage (en bleu et vert) ;
- 3) Fixation des index (rose et violet) et adaptateurs (orange et rouge) lors d'une PCR d'indexation avec obtention d'un fragment prêt pour le séquençage.

Un second protocole de préparation de librairie, de principe différent, est le protocole 16S Metagenomics Sequencing Library Preparation (Illumina) (107). Il s'agit du protocole employé dans la plupart des publications traitant du NGS appliqué à l'étude de *Cryptosporidium* sp. (15,82,83). Même si très similaire au protocole Nextera XT (Illumina), il comprend une différence majeure qui réside dans l'absence de tagmentation, nécessitant d'ajouter des amorces de séquençage d'une autre manière. Des amorces spécifiques de la séquence d'intérêt sont incrémentées d'adaptateurs dits « overhang » qui rendent les amorces compatibles avec le NGS en les rendant complémentaires de l'amorce de séquençage. Les amorces ainsi modifiées sont fixées à la région d'intérêt lors d'une réaction d'amplification. Puis au cours d'une PCR d'indexation, les adaptateurs et index sont à leur tour fixés. Les principales étapes de ce

protocole sont schématisées dans la figure 6. Le résultat final est identique à celui observé en figure 4.

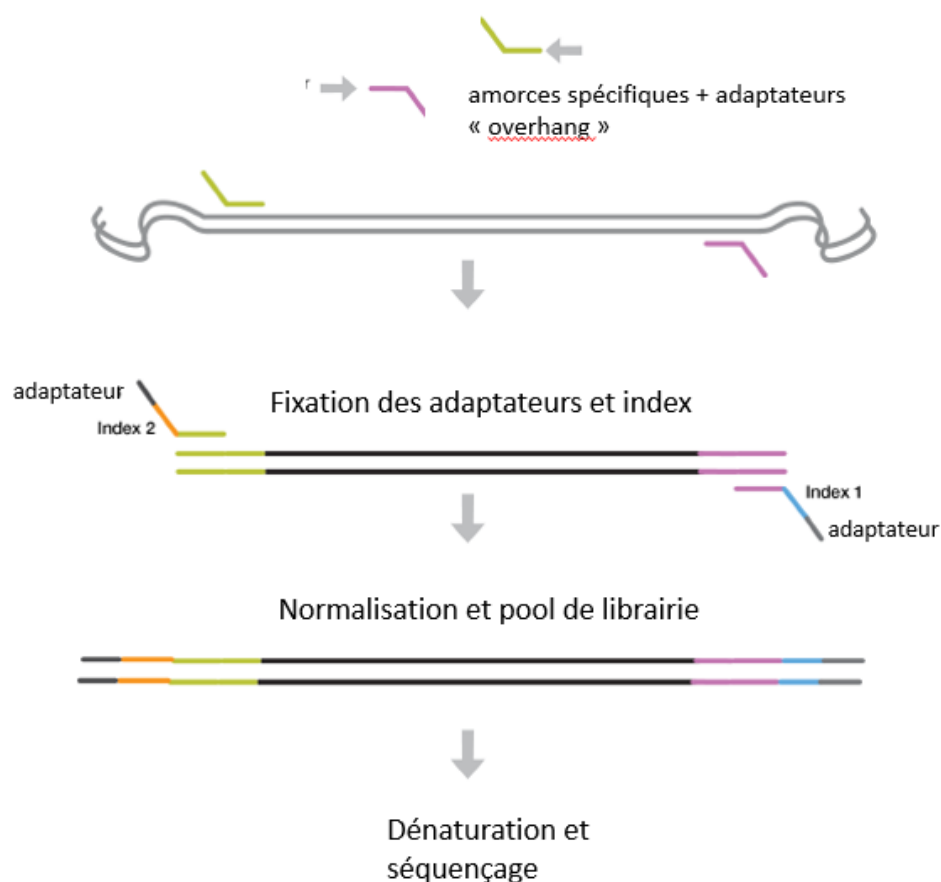


Figure 6 : Schéma représentant le principe de la préparation de librairie à l'aide du protocole 16S Metagenomics Sequencing Library Preparation (Illumina) (107).

Génération de clusters

Avant d'être chargée dans le séquenceur, la librairie est dénaturée afin d'obtenir de l'ADN simple brin. La librairie est ensuite injectée sur la flow cell et les différents fragments d'ADN vont venir se lier aux oligonucléotides en surface par l'intermédiaire de leurs adaptateurs. La flow cell est une lame de verre contenant de petits canaux, à travers lesquels les polymérases, les dNTP et les tampons peuvent circuler. Le verre à l'intérieur des canaux est incrémenté d'oligonucléotides complémentaires aux adaptateurs préalablement fixés aux brins d'ADN de la librairie (88,97). Chaque fragment d'ADN simple brin ainsi fixé est élongé selon le principe de PCR, permettant la formation d'un brin complémentaire. Une étape de dénaturation permet le détachement du brin d'origine, ne laissant que le brin issu de

l'élongation en surface de la flow cell, fixé de façon covalente. Ce brin se replie ensuite en direction de la flow cell, afin que l'adaptateur fixé sur son extrémité libre se fixe de manière aléatoire à un oligonucléotide sur la flow cell. On parle de pont. Une nouvelle réaction d'amplification a lieu, on obtient ainsi un pont fait d'ADN double brin. Une étape de dénaturation permet de séparer et redresser les deux brins complémentaires, qui restent chacun fixés à la flow cell par l'intermédiaire d'un de leurs adaptateurs. Les étapes de formation de ponts et d'amplification se répètent. Pour un même fragment d'ADN initial, on obtient ainsi une multitude de molécules d'ADN dans une même région de la flow cell, on parle de clusters. Dans ces clusters sont retrouvés des brins d'ADN dans les deux sens. Une digestion enzymatique permet d'éliminer un des deux brins et de ne conserver que des molécules d'ADN dans le même sens (104,105) (Figure 7).

Réaction de séquençage

Afin d'amorcer la réaction de séquençage, une amorce complémentaire se fixe à l'amorce d'un brin d'ADN. Le séquençage consiste ensuite en la fixation de dNTP fluorescents dits terminateurs. Le premier dNTP complémentaire se fixe au niveau de la première base du brin d'intérêt. Le fluorochrome du dNTP est excité et le signal est détecté. La notion de « terminateur » implique que ce dNTP doit être clivé afin qu'un autre dNTP puisse être incorporé à la suite. La séquence est ainsi reconstituée par fixation successive des différents dNTP tout au long de la séquence d'intérêt. Afin de permettre la lecture des brins d'ADN dans l'autre sens, ils sont retournés en formant des ponts à la surface de la flow cell : on parle de séquençage « paired-end » (105).

Les séquences obtenues dans les deux sens sont comparées entre elles afin d'aligner les parties qui se chevauchent. Enfin, les différentes séquences appartenant à un même échantillon sont regroupées, grâce aux index. Cela permet d'obtenir des séquences finales de plusieurs millions de bases pour chaque échantillon. A la fin du séquençage, ces séquences obtenues sont comparées à des séquences de référence (104).

La figure 7 résume les étapes qui viennent d'être décrites.

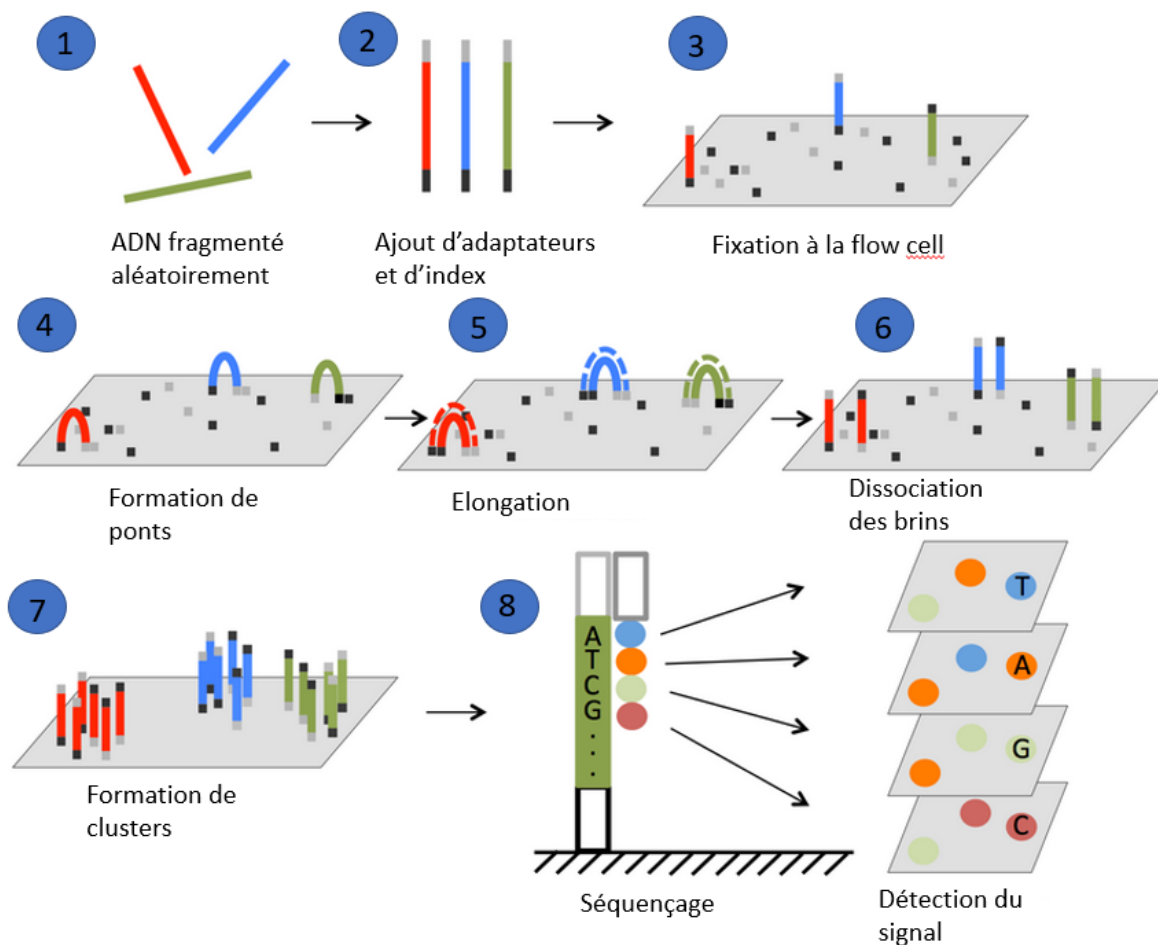


Figure 7 : Etapes du séquençage par technique NGS (108).

1 – Fragmentation aléatoire de l'ADN préalablement extrait ; 2 – Ajout d'adaptateurs (fixation sur la flow cell) et d'index (identification des échantillons) ; 3 – Fixation à la flow cell par l'intermédiaire d'une interaction entre les adaptateurs et des oligonucléotides en surface de la flow cell ; 4 – Formation de ponts possible grâce au repli du fragment d'ADN sur lui-même, qui fixe son adaptateur libre sur la flow cell ; 5 – Elongation et formation d'un brin complémentaire ; 6 – Dissociation des brins d'ADN par dénaturation ; 7 – La répétition des étapes précédentes aboutit à la formation de clusters ; 8 – Réaction de séquençage par l'intermédiaire de dNTP fluorescents terminateurs dont on détecte le signal pour reconstituer la séquence.

c-2-c) Applications générales du NGS

Le NGS peut permettre le séquençage de diverses cibles que sont l'ARN (RNA-seq), l'ADN (DNA-seq) ou encore la chromatine (ChIP-seq) (105,109). Le RNA-seq est une technique permettant le séquençage de tous les ARNm et les petits ARN. Cela permet de détecter, quantifier des ARN, voire découvrir des ARN jusqu'alors inconnus, et d'en étudier les implications dans divers mécanismes de régulation (oncogénétique, épigénétique, ...). Le ChIP-seq est une technique de NGS permettant de cartographier des sites de fixation de facteurs de transcription, mettre en évidence des interactions protéines/ADN ou encore d'étudier la

structure de la chromatine. Le séquençage haut débit peut également se faire à l'échelle de l'ADN (DNA-seq), et se divise alors en trois catégories : séquençage ciblé, étude d'exome ou de génome. Le séquençage du génome entier permet l'analyse approfondie voire la découverte d'organismes. Le séquençage des exomes offre la possibilité de séquencer toutes les parties codantes des gènes et est très utilisé dans l'étude des maladies génétique. Enfin le séquençage ciblé permet le séquençage de zones spécifiques du génome, offrant des informations plus spécifiques et plus simples à analyser (109). Notons que le séquençage ciblé (DNA-seq) est l'approche NGS à laquelle nous avons eu recours dans ce travail, par l'intermédiaire du séquençage du gène codant la protéine gp60 de *Cryptosporidium* sp.

L'accessibilité accrue au NGS a permis ces dernières années un large essor de l'étude génomique en microbiologie, expliquant par exemple que 98% des génomes bactériens actuellement séquencés l'ont été durant ces sept dernières années (80). Les approches de génomique qui ont pu être développées permettent l'investigation approfondie des souches microbiennes, de leur virulence, leur pathogénicité, leur résistance aux traitements usuels. Enfin, grâce à des approches dites métagénomiques, il est possible d'étudier la diversité microbiologique de n'importe quel échantillon, et de répondre à des problématiques actuelles telles que la composition des microbiotes (digestifs, pulmonaires, ...). Diverses plateformes de séquençage à haut débit ont vu le jour et ont pour principaux axes le développement d'approches diagnostiques (pour n'importe quel pathogène viral, fongique, bactérien ou parasitaire), et le développement d'approches de caractérisation (identification de mutations, de gènes d'intérêt, classification, type moléculaire) (81).

Nous nous intéresserons particulièrement à la technologie Illumina, qui est celle que nous avons utilisée dans ce travail. La technologie Illumina repose sur une méthode de séquençage par synthèse, ce qui signifie qu'elle permet d'identifier les bases simultanément à leur incorporation dans la chaîne d'acides nucléiques en cours de synthèse (105,109). Chaque base ainsi incorporée émet un signal fluorescent utilisé pour déterminer la séquence d'ADN. La figure 8 permet de schématiser l'étape du séquençage selon la technologie Illumina.

Les premières étapes du protocole sont variables selon le protocole de préparation de librairie choisie. L'ADN peut subir ou non une fragmentation initiale. Dans les deux cas, l'ajout

d'adaptateurs et d'index aux deux extrémités des fragments d'ADN formés est indispensable. La figure 7 schématise les différentes étapes du séquençage NGS selon la technologie Illumina.

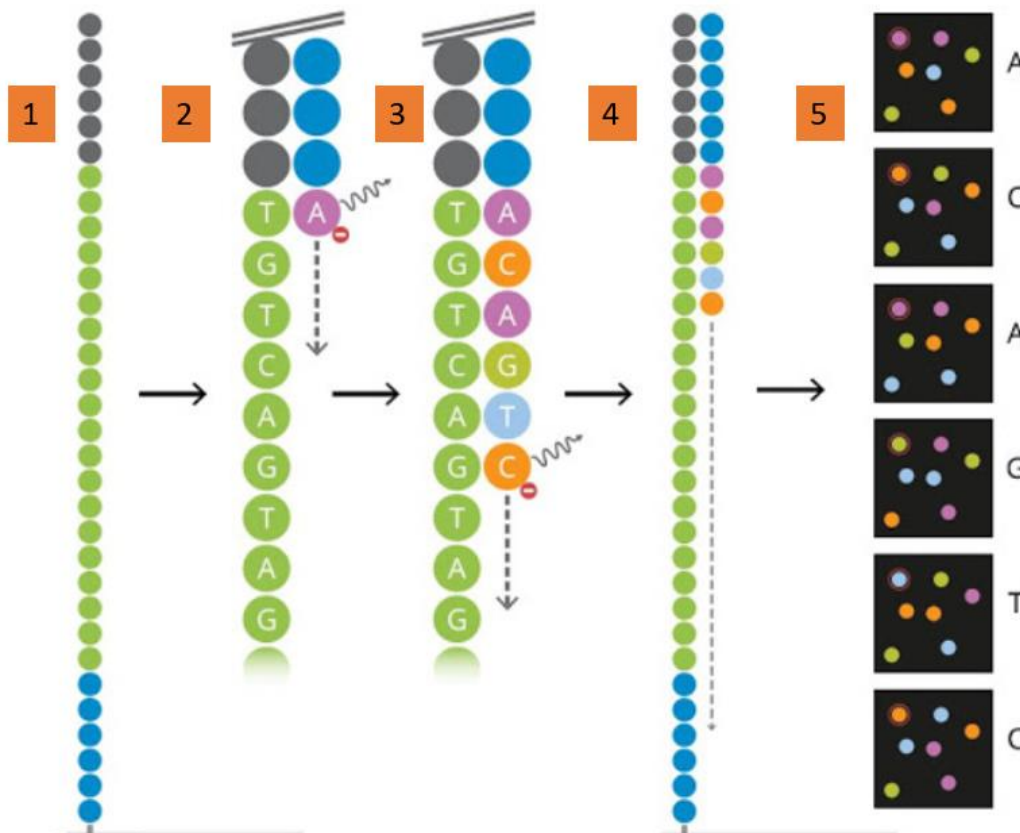


Figure 8 : Schéma présentant les étapes de séquençage par synthèse selon la technologie Illumina® (109).

- 1) Un fragment d'ADN simple brin issu de la librairie dénaturée est chargé dans le séquenceur. La séquence d'intérêt est représentée en vert. L'extrémité grise correspond à l'amorce fixée pendant la préparation de la librairie. Elle est complémentaire de l'amorce qui permettra de débiter la réaction de séquençage.
- 2) La séquence bleue qui vient se fixer sur le fragment d'ADN correspond à l'amorce de séquençage. Elle vient se fixer spécifiquement à l'amorce grise du fragment et permet de débiter l'élongation du brin complémentaire.
- 3) A chaque position vient se fixer un dNTP fluorescent terminateur complémentaire de la base correspondante sur le brin d'intérêt. Ainsi en première position se fixe un dATP (complémentaire du T) terminateur. Le fluorophore spécifique du dATP qui y est fixé est excité par un laser et le signal émis est recueilli. Le dATP terminateur est ensuite clivé pour permettre la fixation d'un nouveau dNTP fluorescent terminateur.
- 4) L'élongation peut continuer de la même manière et les signaux sont détectés à chaque incorporation de bases.
- 5) Les signaux sont interprétés et permettent de reconstituer la séquence d'intérêt représentée en vert.

c-2-d) Le NGS appliqué à l'étude de *Cryptosporidium* sp.

Le NGS offre l'avantage par rapport à la méthode Sanger, de pouvoir identifier efficacement des génotypes de faible abondance au sein d'échantillons mixtes (15) grâce au séquençage massif de molécules d'ADN en parallèle. Différentes équipes de recherche ont

abordé ces dernières années la possibilité d'appliquer l'approche NGS pour l'identification et le sous-typage de *Cryptosporidium* sp. (15,82,83,84,87,96).

Paparini et *al.* soulignaient en 2015 l'importance d'avoir à notre disposition des techniques fiables, rapides et sensibles pour mieux appréhender les épidémies (37). Ils ont réalisé le séquençage haut débit de 25 échantillons de selles humaines et animales, positives à *Cryptosporidium* sp. L'ADN de *Cryptosporidium* sp. a été amplifié en ciblant le gène codant la petite sous-unité 18S de l'ARN ribosomal et le gène codant l'actine, puis séquencé en utilisant ces mêmes cibles, en ayant recours à la technologie de séquençage Ion Torrent (ou séquençage par ion semi-conducteur, ThermoFischer). En utilisant cette approche, les séquences obtenues lors du séquençage n'existant qu'en une seule copie étaient exclues de toute interprétation, considérées comme des erreurs de PCR/séquençage. Les séquences retenues étaient assignées à une séquence de référence en comparant à une base de données « maison » à l'aide du logiciel BLAST. Cette technique a permis de mettre ici en évidence des mélanges au sein de différents échantillons (2 échantillons sur 25), là où le séquençage par la technique Sanger avait échoué à le faire. Notons que dans cette publication, environ 4% de séquences non spécifiques de *Cryptosporidium* sp. ont été retrouvées à l'issue du séquençage, lorsque la cible était le gène codant l'actine.

L'équipe australienne de Zahedi et *al.* s'est quant à elle intéressée à l'approche NGS appliquée à *Cryptosporidium* sp. à plusieurs reprises.

En 2017, une première étude avait pour but de mettre en évidence l'existence de co-infections, car celles-ci peuvent avoir des impacts sur le comportement du parasite au sein de l'hôte, en termes de virulence, résistance aux traitements, etc ... Ainsi 41 échantillons de fèces d'animaux ont été recueillis dans deux régions australiennes, tous positifs à une espèce de *Cryptosporidium* sp. incluant *C. hominis* (11 échantillons), *C. parvum* (22 échantillons) et *C. cuniculus* (8 échantillons). La détection de la présence d'ADN de *Cryptosporidium* sp. et l'identification des sous-types dans ces échantillons étaient réalisées par amplification par PCR du gène codant la protéine gp60 suivie d'un séquençage des produits d'amplification par NGS sur MiSeq (Illumina), en comparaison du séquençage par la technique Sanger. Les séquences issues du séquençage haut débit n'étant représentées qu'en une ou deux copies identiques étaient supprimées, ainsi que celles dont le nombre de copies représentait moins de 0.01% du nombre total de séquences. Les conditions d'appariement aux séquences de référence étaient

strictes, fixées à 100% d'homologie de séquence et 100% de recouvrement. Les séquences étaient comparées à une base de données contenant 131 séquences de sous-types de *Cryptosporidium* sp. issues de GenBank. Pour les 41 échantillons, les résultats du séquençage par NGS étaient similaires à ceux obtenus par séquençage Sanger en termes de sous-typage. En effet, le sous-type identifié pour chaque échantillon par Sanger était également le sous-type retrouvé comme majoritaire par NGS. Cependant, les résultats obtenus par NGS mettaient en évidence des sous-types additionnels pour plusieurs échantillons. Ainsi, pour les échantillons positifs à *C. parvum*, 2 à 4 sous-types cohabitant au sein d'un même isolat étaient retrouvés pour certains échantillons, chaque sous-type représentant entre 0.4% et 31% du nombre total de séquences. De même pour *C. cuniculus*, 3 sous-types étaient retrouvés au sein de certains échantillons pour lesquels un seul sous-type avait été identifié par technique Sanger, et ce à plusieurs reprises, chaque sous-type représentant entre 0.4% et 6.7% du nombre total de séquences obtenues. Seuls les échantillons de *C. hominis* ne présentaient pas de sous-type additionnel à l'issue du séquençage haut débit.

En 2018, ils ont de nouveau étudié *Cryptosporidium* sp. à l'aide de l'approche NGS mais d'un autre point de vue. Ils ont ainsi travaillé sur la présence de *Cryptosporidium* spp. dans les eaux usées (87). Le but était de cartographier la diversité de *Cryptosporidium* spp. dans ces eaux, en focalisant leur étude sur les espèces potentiellement pathogènes pour l'Homme, afin de disposer d'outils qui permettraient (i) de mieux appréhender les sources de contamination hydriques, et (ii) de réfléchir à des normes qualité dans l'optique d'un futur recyclage des eaux usées. 730 échantillons d'eaux usées ont été récoltés dans trois régions distinctes en Australie sur une période d'un an. L'ADN de *Cryptosporidium* sp. a été amplifié et séquencé en haut débit en ciblant le gène de la petite sous-unité 18S de l'ARN ribosomal, et ce en ayant recours à la technologie Illumina sur un automate MiSeq (Illumina). Les séquences obtenues lors du séquençage devaient présenter un minimum de 100 copies répliquées pour être conservées et assignées. Les séquences restantes devaient présenter une correspondance à 99% en termes de similitude de bases et 98% de recouvrement avec la séquence de référence. Les résultats de cette étude ont mis en évidence des mélanges entre 2 et 8 sous-types de *Cryptosporidium* spp. au sein d'échantillons de diverses régions australiennes.

d) Applications du sous-typage

Nous nous intéresserons pour la suite uniquement au sous-typage par séquençage partiel du gène codant la protéine gp60, car il s'agit du marqueur que nous étudierons par technique NGS dans ce travail. La caractérisation à l'échelle du sous-type par séquençage du gène codant la protéine gp60 a permis de mettre en évidence des différences intra-espèces, importantes pour la compréhension épidémiologique de *Cryptosporidium* sp.

Le sous-typage apporte des informations importantes en termes de distribution géographique du parasite. En effet, alors qu'il était admis que la distribution relative des deux principales espèces, *C. hominis* et *C. parvum*, variait en fonction des zones géographiques, l'étude des variations intraspécifiques de la protéine gp60 a permis d'en préciser les sous-types (5,71). Ainsi, en Europe et en Australie, *C. hominis* est principalement représenté par la famille de sous-type Ib (41,76), et plus particulièrement par IbA10G2 et IbA9G3 (1,18,30). A noter que le sous-type IbA10G2 est le sous-type de *C. hominis* responsable de l'épidémie massive de Milwaukee, qui a contaminé 400000 personnes (1,18). Ce sous-type est encore à l'heure actuelle responsable de 50% des contaminations d'origine hydrique aux Etats-Unis et au Canada (2). Il est également responsable d'environ 80% des cas humains d'infection en Europe, et de la quasi-totalité des épidémies européennes à *C. hominis* (4). Concernant *C. parvum*, la famille de sous-type prépondérante chez l'Homme en Europe, en Australie et aux Etats-Unis est la famille IIa (18,41,76,94).

En France, en 2018, *C. parvum* reste l'espèce majoritaire, comme les années précédentes. Le génotype IIaA15G2R1 est largement prédominant (40% des isolats). Concernant *C. hominis*, le génotype dominant demeure le génotype IbA10G2 (50% des isolats), ce qui est le reflet de la prédominance mondiale de ce génotype. Un second génotype est toutefois apparu en proportion importante en 2018, à savoir le génotype IaA22R2 (85).

En plus de mieux comprendre les disparités géographiques, l'étude des sous-types a également permis de préciser des dynamiques de transmission, notamment pour l'espèce *C. parvum* que l'on pensait jusque-là uniquement zoonotique. Pour préciser ces dynamiques de transmission, il faut comparer les sous-types présents dans une même région, chez l'animal et chez l'Homme. Par exemple, le sous-type de *C. parvum* prédominant chez les veaux au Portugal et en Slovénie est IIaA15G2R1, et c'est également le plus fréquemment retrouvé chez l'Homme dans ces pays. De même en Irlande du Nord où l'Homme se contamine principalement avec le

sous-type le plus représenté chez les veaux, à savoir le sous-type IIaA18G3R1 (18). Ces données épidémiologiques sont intéressantes pour souligner le caractère zoonotique encore prépondérant lors de la transmission de certains sous-types de *C. parvum*, que sont notamment deux sous-types importants : IIa et IIId. Cependant, le sous-typage de *C. parvum* a permis de mettre en évidence une famille de sous-types, qui malgré de nombreuses études n'a jamais été retrouvée chez l'animal. Il s'agit de la famille IIc (5), qui est notamment retrouvée dans les régions urbanisées des Etats-Unis, où elle devient majoritaire par rapport à la famille IIa, certainement du fait de son caractère anthroponotique pur (18).

La structure génétique de population de *Cryptosporidium* sp. dans les différentes régions du monde fait également l'objet de discussions grâce à l'avènement du sous-typage. Plusieurs études récentes ont ainsi souligné le caractère clonal de l'expansion européenne du sous-type IbA10G2. A l'inverse de ce l'on retrouve dans certains états des Etats-Unis, où des recombinaisons génétiques auraient lieu au sein du sous-type IbA10G2, du fait qu'il soit en contact de nombreux autres sous-types dans ces états (72). De même dans les pays en voie de développement, où l'on peut identifier jusque quatre sous-types qui cohabitent en proportions similaires, principalement les sous-types Ia, Ib, Id et Ie. On peut citer l'Inde et le Pérou où une haute diversité de sous-types de *C. hominis* est retrouvée (5).

Le sous-typage est également important en termes de cliniques, car il est possible de relier les caractéristiques biologiques et la virulence clinique avec le sous-type (18). Bien qu'il soit déjà connu que *C. hominis* entraîne des tableaux cliniques plus sévères que *C. parvum*, le sous-typage a permis de mettre en évidence des différences de virulence à l'échelle du sous-type (8,30). La première étude qui soulignait des variations dans la virulence entre les différents sous-types de *Cryptosporidium* sp. date de 2007. Cama *et al.* mettaient alors en avant des disparités cliniques entre patients porteurs de VIH au Pérou, en fonction du sous-type de *C. hominis* qui les infectait. Ainsi, des patients atteints par une souche de *C. hominis* de sous-type Id présentaient une diarrhée plus sévère que des patients contaminés par une souche de sous-type Ib, tandis que des patients atteints par une souche de sous-type Ia restaient asymptomatiques (86). Par ailleurs, des études cliniques chez des volontaires sains ont montré que différents isolats de *C. parvum* (UCP, IOWA, TAMU, dont les noms sont issus des universités où elles ont été découvertes) montraient des ID50s (dose infectieuse 50 %) distinctes, indiquant des variations dans leur pathogénicité (93). Au sein de l'espèce *C. hominis*, le sous-

type IbA10G2 est le plus représenté à l'échelle mondiale, et il est également reconnu comme le plus virulent (1,2,17). Les mécanismes expliquant l'hypertransmissibilité et la virulence de ces sous-types particuliers ne sont pour autant pas clairement élucidés. Les outils moléculaires ont ouvert la voie pour tenter de les appréhender. En étudiant les séquences génétiques de sous-types virulents comme IbA10G2, des recombinaisons génétiques particulières autour de la région codant pour la protéine gp60 ont pu être mises en évidence. Ces recombinaisons étant les seules différences retrouvées dans les séquences de ces sous-types particuliers, cela suggère leur implication dans la transmissibilité accrue et la virulence de certains sous-types. En somme, lorsque des sous-types majeurs sont comparés en termes de virulence et de prévalence, une grande similitude dans l'organisation de leurs gènes et dans la composition de leurs séquences est observée. Leurs différences résident surtout dans des recombinaisons génétiques, observées notamment dans la région du gène codant la protéine gp60 situé sur le chromosome 6. Même dans un pays où la structure de population de *C. hominis* semble clonale, des recombinaisons génétiques au sein d'un sous-type pourraient être à l'origine d'un sous-type dominant. Les recombinaisons génétiques seraient alors la principale force motrice pour l'émergence et l'expansion des sous-types virulents, à l'origine d'épidémies massives (5,17,19). D'autres études dans les années à venir permettront sans doute d'en préciser les mécanismes et peut être de découvrir d'autres pistes génétiques explicatives.

PARTIE II : PARTIE EXPERIMENTALE

I) Objectifs

Ces dernières années, la littérature scientifique s'est enrichie de publications traitant de l'approche NGS appliquée à *Cryptosporidium* (15,20,21,37,82,83,84,87). Les études ont porté sur des échantillons provenant de stations d'eaux usées, de fèces d'animaux, ou encore des échantillons de selles d'origine humaine. Différents loci ont été étudiés pour l'identification et/ou le sous-typage de *Cryptosporidium* sp. : gène codant la protéine gp60 (15), gène codant la petite sous-unité 18S de l'ARN ribosomal (82,84), gène codant pour l'actine (37). Différentes approches de séquençage NGS ont été employées telles que la technologie Ion Torrent (ThermoFisher) (37) ou encore Illumina (Illumina) (15,87), et ce sur différents séquenceurs (HiSeq, TruSeq, MiSeq (Illumina)) et avec différentes techniques de préparation des bibliothèques (fragmentation par nébulisation, ajouts d'adaptateurs « overhangs », tagmentation). Ces études ont montré des résultats encourageants quant à la détection de plusieurs variants au sein d'un même échantillon. Néanmoins, il apparaît clairement qu'il n'y a pas encore de standardisation des techniques NGS appliquées à *Cryptosporidium* sp. De plus, en l'état actuel de nos connaissances, aucune publication n'a porté sur la création de mélanges connus d'ADN de *Cryptosporidium* sp., afin d'en étudier le séquençage par NGS. Or, la mise en application d'une telle approche sur des échantillons cliniques au laboratoire, notamment dans l'investigation d'épidémies, nécessite de telles études afin d'assurer des résultats fiables.

Le travail que nous avons réalisé avait pour objectif d'utiliser cette approche NGS pour la détection de sous-types de *Cryptosporidium* sp., et particulièrement la mise en évidence d'infections mixtes. Nous nous sommes basés sur la technologie mise au point par Zahedi et *al.* en 2017, portant sur la comparaison des techniques de séquençage Sanger et NGS par technologie Illumina sur des échantillons dont la composition était inconnue (15). Cette méthodologie était basée sur l'étude du gène codant la gp60 grâce à la technologie Illumina sur MiSeq (Illumina).

Les premiers essais réalisés ont eu pour but (i) d'utiliser le NGS pour la réalisation du sous-typage de *Cryptosporidium* sp. et (ii) de confirmer les capacités du NGS à détecter des mélanges de sous-types de *Cryptosporidium* sp. dans un même échantillon de selles, mais

surtout d'en préciser les limites, c'est-à-dire : (i) dans quelle proportion un variant minoritaire peut être détecté dans un mélange de différents variants présents en proportions différentes, (ii) quel seuil faut-il appliquer lors de l'analyse afin de distinguer une identification erronée correspondant à un bruit de fond analytique par rapport à l'identification correcte d'un génotype réellement présent au sein d'un échantillon de selles.

Des essais secondaires ont eu pour objectif de comparer les résultats obtenus par technique Sanger à ceux obtenus par NGS lors de l'étude de mélanges d'ADN de *Cryptosporidium* sp.

II) Matériels et méthodes

1) Première partie : Comparaison de deux protocoles de préparation de librairie

Dans un premier temps, nous nous sommes attachés à trouver un protocole NGS adapté à notre problématique, à savoir le sous-typage de *Cryptosporidium* sp. En effet, les publications concernant le sous-typage de *Cryptosporidium* sp. par approche NGS restent peu nombreuses, et il n'existe pas encore de standardisation des protocoles. Il était donc important de s'assurer que le séquençage NGS est capable d'identifier correctement une espèce de *Cryptosporidium* au sein d'un échantillon, et cela sans entraîner trop de bruit de fond analytique. Nous avons ainsi étudié des échantillons de composition connue, positifs à une seule espèce de *Cryptosporidium* sp. par séquençage NGS grâce à deux protocoles de préparation de librairie distincts. Le protocole Nextera XT (Illumina) est un protocole largement utilisé pour la technologie Illumina. Le protocole 16S Metagenomics Sequencing Preparation Library (Illumina) est le protocole qui a été utilisé dans la majorité des publications traitant du NGS appliqué à *Cryptosporidium* sp.

a) Choix des échantillons

Trois échantillons ont été choisis afin de comparer les deux protocoles de préparation de librairie. Ils étaient composés d'ADN extraits de selles de patients positives à *Cryptosporidium* sp. L'extraction avait été réalisée sur l'automate Nuclisens® Easymag® (Biomérieux) ayant recours à la technologie BOOM, basée sur l'utilisation de billes de silice paramagnétiques. Les trois échantillons sélectionnés étaient conservés à -80°C après recherche initiale de parasites

digestifs par PCR au laboratoire de parasitologie du CHU de Dijon, et étaient identifiés 1761, 1770 et 1772.

L'identification de l'espèce de *Cryptosporidium* sp. avait été préalablement réalisée par une PCR en temps réel sur un Light Cycler 2.0 (Roche), ciblant la petite sous-unité 18S de l'ARN ribosomal et utilisant des sondes d'hybridation.

L'identification des sous-types avait quant à elle été réalisée par séquençage selon la méthode Sanger des amplicons obtenus par une PCR en point final ciblant le gène de la gp60 (Protocole de la technique Sanger détaillé en annexe 1) au cours de l'activité de routine du laboratoire de parasitologie du C.H.U. de Dijon.

Nous avons donc en notre possession trois échantillons positifs à *Cryptosporidium* sp. dont nous connaissions l'espèce et le sous-type, et dont nous voulions vérifier l'identification correcte par NGS, et ce par comparaison de deux protocoles de préparation de librairie. Le tableau 5 reprend l'identification des trois échantillons 1761, 1770 et 1772.

Tableau 5 : Récapitulatif des espèces et sous-types correspondants aux échantillons choisis

Numéros d'échantillons	Espèces identifiées	C _T obtenus	Sous-types identifiés
1761	<i>C. parvum</i>	26,3	IICa5G3
1770	<i>C. parvum</i>	28,1	IlaA18G1R1
1772	<i>C. hominis</i>	24,2	IIBa10G2

b) Protocole de préparation de librairie Nextera XT (Illumina)

Le premier protocole de préparation de librairie appliqué à l'étude de *Cryptosporidium* sp. était le protocole Nextera XT (Illumina).

Pour débiter, une première amplification de l'ADN était nécessaire, par PCR en point final ciblant le gène codant la protéine gp60, de la même manière que pour le séquençage par la technique Sanger (annexe 1). L'amplification utilisait les amorces sens AL3532 et anti-sens LX0029, qui sont des amorces spécifiques du gène codant la protéine gp60, et une polymérase HotStarTaq® (Qiagen). La réaction de PCR a été réalisée sur un thermocycleur ABI Applied Biosystems™ 9700 (ThermoFisher) selon le même programme que décrit pour le séquençage selon la technique Sanger (Annexe 1).

L'ADN issu de cette amplification était quantifié précisément afin d'assurer une fragmentation de qualité, et ce avec une méthode basée sur la fluorométrie utilisant l'appareil Qubit® (ThermoFisher). Cet ADN était dilué à une concentration de 0.2ng/μL, permettant de travailler avec 1ng de quantité totale d'ADN (5μL de prise d'essai) comme recommandé par Illumina.

L'ADN était ensuite « tagmenté », c'est-à-dire à la fois coupé par voie enzymatique (tagmentase) et marqué par des amorces spécifiques adaptées au séquençage ultérieur. Pour cela, 5μL d'ADN à 0.2ng/μL ont été mélangés avec 10μL de tampon de tagmentation et 5μL de mix de tagmentation d'amplicon, fournis par le kit Nextera XT (Illumina). Après centrifugation, le mélangeur était chauffé 5 minutes à 55°C. La réaction de tagmentation a été neutralisée à l'aide de 5μL de tampon de neutralisation (fourni dans le kit Nextera XT) ajouté au mélange réactionnel.

Une PCR d'indexation a ensuite été réalisée grâce au kit Nextera XT Index Kit® (Illumina), suivant le protocole décrit par Illumina. Ainsi, 5μL de chaque index (soit 10μL) ont été mélangés à 15μL d'un mélange réactionnel fourni par le kit Nextera XT (Nextera PCR Master Mix), ainsi qu'à 25μL d'ADN tagmenté. Les cycles d'amplification ont été réalisés sur un thermocycleur (ThermoFisher) au cours du programme suivant :

- 72°C pendant 3 minutes
 - 95°C pendant 30 secondes
 - 95°C pendant 10 secondes
 - 55°C pendant 10 secondes
 - 72°C pendant 30 secondes
- } 12 cycles
- 72°C pendant 5 minutes
 - maintien à 10°C

Une fois marqués par les index et adaptateurs, les échantillons ont été purifiés à l'aide de billes Ampure XP® (Beckman). L'incubation a eu lieu sur une plaque magnétique et deux lavages successives ont été réalisés à l'aide d'éthanol à 80%.

L'ADN de chaque échantillon a ensuite été normalisé de façon à assurer une représentation équivalente de chaque échantillon au sein de la librairie poolée. Pour un

échantillon, 44µL d'additif de normalisation (Illumina) et 8µL de billes de normalisation (Illumina) étaient mélangés. Un volume de 45µL de ce mélange a ensuite été mis en contact de 20µL d'ADN purifié. Une incubation sur une plaque magnétique et plusieurs lavages à l'aide d'un tampon de lavage (Illumina) ont été effectués. Enfin, une élution de l'ADN a été réalisée à l'aide de NaOH 0.1N préparé extemporanément, et 30µL de surnageant (contenant l'ADN) ont été stockés dans 30µL de tampon de stockage (Illumina).

Les échantillons ont été mélangés à volume égal au sein d'une unique librairie poolée. 24µL de cette librairie ont été prélevés et mélangés à 576µL de tampon d'hybridation (Illumina) afin d'en assurer la dilution à 20pM, concentration recommandée par Illumina pour le chargement dans l'automate.

Cette librairie poolée a été dénaturée à l'aide d'un chauffage à 96°C pendant 2 minutes, puis chargée dans l'automate MiSeq (Illumina), dans lequel elle a été séquencée à l'aide d'une cartouche de réactifs V3 600 cycles (Illumina). Dans le run a été ajouté un contrôle interne représenté par du PhiX (Illumina), lui-même dénaturé et dilué de la même manière que la librairie.

La figure 9 résume les principales étapes de ce protocole de préparation de librairie.

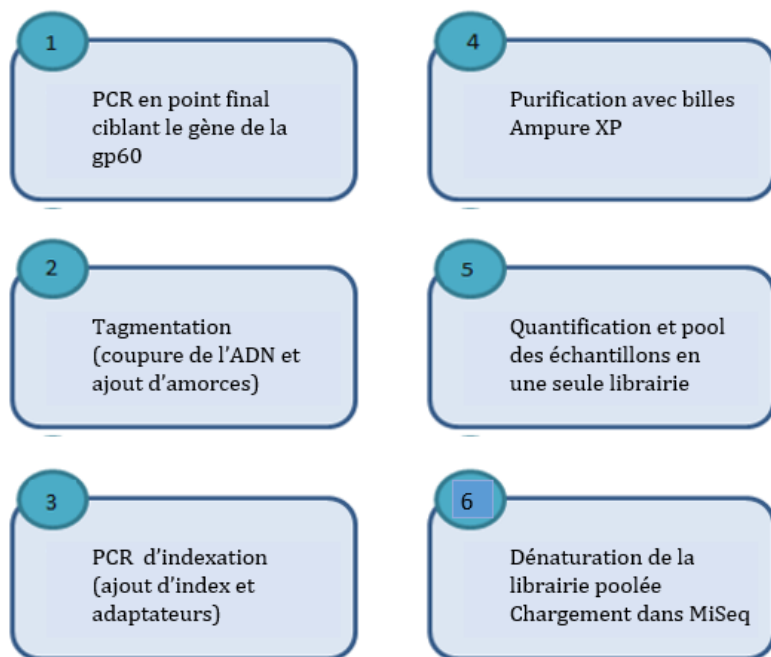


Figure 9 : Etapes du protocole de préparation de librairie pour séquençage NGS Nextera XT (Illumina)

c) Protocole de préparation de librairie 16S Metagenomics Sequencing Library Preparation (Illumina).

Le second protocole de préparation de librairie testé était le protocole 16S Metagenomic Sequencing Library Preparation (Illumina). Il s'agit du protocole employé dans la plupart des publications traitant du NGS appliqué au sous-typage de *Cryptosporidium* sp. (15,82,83).

La première amplification des extraits d'ADN de nos trois échantillons reposait sur une PCR en point final ciblant le gène codant la protéine gp60, identique à celle employée pour le séquençage par la technique Sanger ou pour le NGS avec le protocole Nextera XT (Illumina). Les amorces utilisées étaient donc AL3532 (sens) et LX0029 (anti-sens) (TIB Molbiol) et la Taq polymérase est une HotStarTaq® (Qiagen). L'amplification était réalisée sur un thermocycleur ABI Applied Biosystems™ 9700 (ThermoFisher), selon le même programme de PCR que pour le séquençage selon la méthode Sanger détaillé en annexe 1.

Une seconde amplification était nécessaire, par une PCR nichée en point final dirigée contre le gène codant la protéine gp60, quasiment identique à la première PCR. Cependant, pour être compatible avec un protocole de séquençage haut débit, cette seconde PCR comprenait une différence notable par rapport à la première amplification. Il s'agissait de l'ajout d'adaptateurs dits « overhang » sur les amorces spécifiques du gène codant la protéine gp60 : AL3532 et LX0029. Cela rendait possible la réaction de séquençage ultérieure. Ces adaptateurs « overhang » possédaient les séquences suivantes :

- Overhang sens : 5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG – {AL3532} – 3'
- Overhang anti-sens : 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG – {LX0029} – 3'

Les amplicons issus de la PCR nichée avec amorces modifiées par les séquences « overhang » étaient purifiés à l'aide de billes paramagnétiques AMPure XP® (Beckman), avec une incubation sur plaque magnétique et deux étapes de lavage à l'éthanol 80%. Un volume de 50µL de surnageant, contenant l'ADN purifié, a été récupéré pour chacun des échantillons.

L'étape suivante était une PCR dite d'indexation, consistant à fixer à l'extrémité de chaque fragment d'ADN un index et un adaptateur de séquençage issus du kit Nextera XT index Kit® (Illumina), et ce à l'aide d'une Taq polymérase KAPA HiFi® (KAPA Biosystems). Dans

chaque puits étaient déposés 5µL de chaque index (soit 10µL), 25µL de Taq polymérase, 10µL d'eau et 5µL d'ADN purifié, soit un volume total de 50µL. Le programme d'amplification était court :

- 95°C pendant 3 minutes
 - 95°C pendant 30 secondes
 - 55°C pendant 30 secondes
 - 72°C pendant 30 secondes
- } 8 cycles
- 72°C pendant 5 minutes
 - Maintien à 4°C

Cette PCR d'indexation était suivie d'une nouvelle étape de purification par les billes AMPure XP® (Beckman), à l'issue de laquelle 25µL de surnageant contenant l'ADN purifié était récupéré. Puis chaque échantillon amplifié et indexé était quantifié à l'aide du QuBit® (ThermoFisher). Il convenait de diluer chaque échantillon à une concentration de 4nM à l'aide de tampon Tris de pH 8.5 concentré à 10mM. Un volume de 5µL de chaque échantillon dilué à 4nM était enfin aliquoté et mélangé aux autres échantillons afin d'obtenir une unique librairie poolée de façon équimolaire.

La librairie ainsi constituée a été dénaturée par du NaOH 0.2N (5µL de la librairie + 5µL de NaOH) puis par la chaleur, en la chauffant à 96°C pendant 2 minutes. Elle a enfin été diluée à l'aide de tampon d'hybridation (Illumina), à une concentration finale de 4pM dans un volume total de 600µL, volume nécessaire à la prise d'essai du MiSeq (Illumina). Dans le run a été ajouté un contrôle interne représenté par du PhiX (Illumina), lui-même dénaturé et dilué de la même manière que la librairie. Le séquençage dans le MiSeq (Illumina) a été assuré par un kit de réactifs MiSeq V3 600 cycles (Illumina).

Les différentes étapes de ce protocole de préparation de librairie sont résumées dans la figure 10.

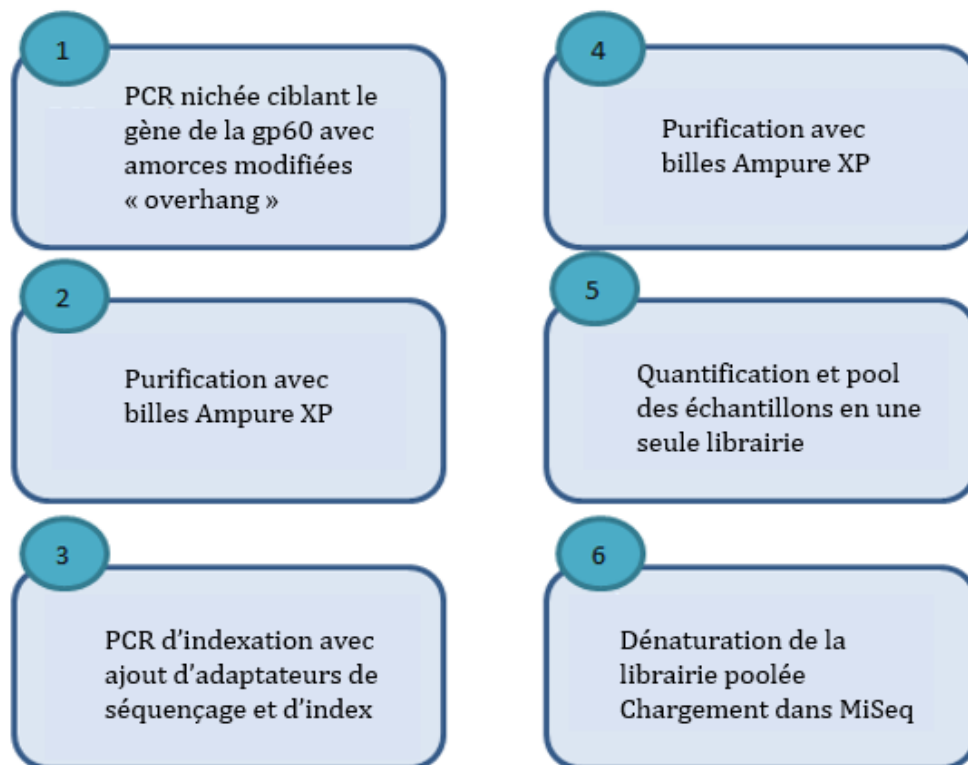


Figure 10 : Etapes du protocole de préparation de librairie pour séquençage NGS : 16S Metagenomic Sequencing Library Preparation (Illumina).

2) Deuxième partie : Etude de mélanges connus d'ADN de *Cryptosporidium* sp. par NGS

Après avoir choisi le protocole de préparation de librairie le plus adapté à notre problématique, nous avons constitué différents mélanges d'ADN de *Cryptosporidium* sp. afin d'en étudier le séquençage par NGS. Nos objectifs étaient de confirmer les capacités du NGS à détecter des variants minoritaires au sein de mélanges, et d'en préciser les limites.

a) Choix des échantillons

Sept échantillons ont été choisis afin d'étudier des mélanges d'ADN connus de *Cryptosporidium* sp., en suivant le protocole de préparation de librairie choisi à l'issue de la comparaison précédente.

Chaque échantillon correspondait à un ADN extrait de selles positives à *Cryptosporidium* sp., reçues pour génotypage au laboratoire du C.H.U. de Dijon, collaborateur du CNR-LE Cryptosporidioses. Les sept échantillons sélectionnés (1770, 1772, 1791, 2094, 2369, 2479, 2487) ont été conservés à -80°C.

L'identification d'espèce de *Cryptosporidium* sp. présente dans chaque échantillon avait été préalablement réalisée par PCR en temps réel ciblant le gène codant la sous-unité 18S de l'ARN ribosomal, avec sondes d'hybridation, sur un Light Cycler 2.0 (Roche).

L'identification des sous-types avait été opérée par séquençage selon la technique Sanger du gène codant la protéine gp60 (protocole détaillé en annexe 1). Les espèces et sous-types identifiés pour les 7 échantillons sont présentés dans le tableau 6.

Tableau 6 : Récapitulatif des espèces et sous-types correspondant aux échantillons choisis

Numéros d'échantillons	Espèces identifiées	C _T obtenus	Sous-types identifiés
1770	<i>C. parvum</i>	28,1	IlaA18G1R1
1772	<i>C. hominis</i>	24,2	IbA10G2
2479	<i>C. parvum</i>	26,8	IlaA20G1R1
2487	<i>C. parvum</i>	24,5	IlaA15G2R1
1791	<i>C. erinacei</i>	30.5	XIIIaA21R1
2094	<i>C. meleagridis</i>	25.7	IIIgA21G3R1
2369	<i>C. cuniculus</i>	26.6	VbA24

- C_T (Cycle Treshold ou Cycle seuil) : Nombre de cycles de PCR nécessaires à la détection du produit de PCR, obtenus ici lors de la PCR en temps réel dirigée contre le gène codant petite sous-unité 18S de l'ARN ribosomal.

b) Constitution des mélanges d'ADN de *Cryptosporidium* sp.

Les sept premiers mélanges (1 à 7) consistaient en un mélange d'ADN d'espèces distinctes, *C. hominis* IbA10G2 (1772) et *C. parvum* IlaA18G1R1 (1770) dans des proportions différentes. L'objectif était de réaliser une gamme de concentration pour déterminer jusqu'à quelle proportion il est possible de détecter un variant minoritaire par séquençage haut débit. Ces mélanges ont été réalisés directement à partir des extraits d'ADN de 1770 et 1772, n'ayant subi aucune amplification, de façon à être le plus proche possible de la composition réelle d'une selle positive à *Cryptosporidium* sp.

Le mélange 8 consistait en un mélange à parts égales de trois ADN d'une même espèce, *C. parvum*, mais de sous-types différents (1770, 2479 et 2487), permettant de vérifier que l'approche NGS permet de distinguer trois sous-types d'une même espèce très proches génétiquement au sein d'un même échantillon, chaque sous-type ne variant des autres que par

la répétition des tandems au sein du microsatellite (bases en rouge dans les séquences suivantes).

Les séquences de référence utilisées pour chaque échantillon sont issues de la banque de données GenBank :

➤ *C. parvum* IlaA15G2R1

TTAAAGGATGTTCTGTTGAGGGGTCATCATCGTCATCGTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAACATCAACCGTCGCACCAGCAAATAAGGCAAGAAGTGGAGAAGACGCAGAAGGCAGTCAA
GATTCTAGTGGTACTGAAGCTTCTGGTAGCCAGGGTTCTGAAGAGGAAGGTAGTGAAGACGATGGCC
AAACTAGTGCTGCTTCCCAACCCACTACTCCAGCTCAAAGTGAAGGCGCAACTACCGAAACCATAGAA
GCTACTCCAAAAGAAGAATGCGGCACTTCATTTGTAATGTGG

➤ *C. parvum* IlaA18G1R1

TTAAAGGATGTTCTGTTGAGGGGTCATCATCGTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAACATCAACCGTCGCACCAGCAAATAAGGCAAGAAGTGGAGAAGACGCAGAAGGC
AGTCAAGATTCTAGTGGTACTGAAGCTTCTGGTAGCCAGGGTTCTGAAGAGGAAGGTAGTGAAGACG
ATGGCCAAACTAGTGCTGCTTCCCAACCCACTACTCCAGCTCAAAGTGAAGGCGCAACTACCGAAACC
ATAGAAGCTACTCCAAAAGAAGAATGCGGCACTTCATTTGTAATGTGG

➤ *C. parvum* IlaA20G1R1

TTAAAGGATGTTCTGTTGAGGGGTCATCATCGTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAACATCAACCGTCGCACCAGCAAATAAGGCAAGAAGTGGAGAAGACGCA
GAAGGCAGTCAAGATTCTAGTGGTACTGAAGCTTCTGGTAGCCAGGGTTCTGAAGAGGAAGGTAGTG
AAGGCGATGGCCAAACTAGTGCTGCTTCCCAACCCACTACTCCAGCTCAAAGTGAAGGCGCAACTACC
GAAACCATAGAAGCTACTCCAAAAGAAGAATGCGGCACTTCATTTGTAATGTGG

Pour réaliser les mélanges d'ADN 1 à 8, nous nous sommes basés sur les C_T obtenus pour chaque échantillon lors de la PCR en temps réel ciblant le gène de la sous-unité de l'ARN ribosomal qui avait permis l'identification d'espèce. Pour rappel, l'échantillon 1770 avait un C_T de 28,1 et l'échantillon 1772 un C_T de 24,2.

En nous intéressant à la phase d'amplification exponentielle de la PCR en temps réel, nous avons pu déterminer la quantité d'ADN obtenue à l'issue de n cycles d'amplification Q_n , et ce à partir de la quantité d'ADN initiale Q_0 , grâce à la formule suivante :

$$Q_n = (1 + E)^n \cdot Q_0 \text{ où } E \text{ est l'efficacité de duplication de l'ADN.}$$

Au cours d'une réaction de PCR, il existe une valeur seuil à partir de laquelle sont déterminés les nombres de cycles d'amplification nécessaires à la détection des amplicons de chaque échantillon. Cette valeur seuil Q_s étant unique pour tous les échantillons traités au cours d'une même PCR, cela signifie que la quantité d'ADN initiale de chaque échantillon (Q_{01} , Q_{02} , Q_{03} , ...) est directement reliée à cette valeur seuil Q_s .

Ainsi, pour des échantillons 1, 2, ..., n traités au cours d'une même PCR on a :

$$Q_s = (1 + E)^{CT1} \cdot Q_{01}$$

$$Q_s = (1 + E)^{CT2} \cdot Q_{02}$$

$$Q_s = (1 + E)^{CTn} \cdot Q_{0n}$$

Il est donc possible de calculer le rapport entre les quantités d'ADN de deux échantillons à partir de leur différence de CT.

$$Q_{01} / Q_{02} = (1 + E)^{CT2 - CT1}$$

En estimant que la PCR a une efficacité de 100% et qu'elle est reproductible d'un échantillon à l'autre au cours de la PCR, nous obtenons : $Q_{01} / Q_{02} = 2^{CT2 - CT1}$

L'échantillon 1770 a été conservé pur et a été dilué au $1/5^e$, $1/10^e$ et $1/100^e$ dans différents tubes. L'échantillon 1772 a été conservé pur et dilué de la même manière dans différents tubes. Des mélanges volume à volume ont ensuite été réalisés, en prenant $2\mu\text{L}$ de 1770 (pur ou dilué) et $2\mu\text{L}$ de 1772 (pur ou dilué), permettant d'obtenir le schéma de constitution des mélanges exposé dans le tableau 7.

Tableau 7 : Schéma de constitution des mélanges 1 à 8 du second essai.

	1770	1772	2479	2487	1791	2094	2369
1	Pur	$1/100^e$					
2	Pur	$1/10^e$					
3	Pur	$1/5^e$					
4	Pur	Pur					
5	$1/5^e$	Pur					
6	$1/10^e$	Pur					
7	$1/100^e$	Pur					
8	Pur		Pur	Pur			
1791					Pur		
2094						Pur	
2369							Pur

Par exemple le mélange 2 était un mélange de $2\mu\text{L}$ d'ADN de 1770 pur et de $2\mu\text{L}$ d'ADN de 1772 dilué au $1/10^e$.

Sur la base de la formule énoncée précédemment nous avons calculé la proportion d'ADN de chaque échantillon dans les mélanges. Ces proportions sont indiquées dans le tableau 8.

Tableau 8 : Proportions en ADN de *C. parvum* (1770) et *C. hominis* (1772) dans les mélanges 1 à 8

	1770	1772	2479	2487	1791	2094	2369
1	85%	15%					
2	40%	60%					
3	25%	75%					
4	5%	95%					
5	2%	98%					
6	1%	99%					
7	0.1%	99.9%					
8	10%		25%	65%			
1791					100%		
2094						100%	
2369							100%

c) Application du protocole NGS

Une fois les mélanges 1 à 8 réalisés à partir des extraits d'ADN de *Cryptosporidium* sp., ces mélanges ainsi que les échantillons 1791, 2094 et 2369 ont été amplifiés par une première PCR en point final ciblant le gène codant la protéine gp60. Les produits de PCR ont été migrés sur gel d'agarose 2% afin de vérifier la présence d'une amplification significative.

Le protocole de préparation de librairie choisi à l'issue de la comparaison des deux protocoles a ensuite été appliqué de manière identique à ce qui a été décrit précédemment dans la partie I de Matériels et Méthodes.

La figure 11 résume les étapes de cette partie II de Matériels et Méthodes.

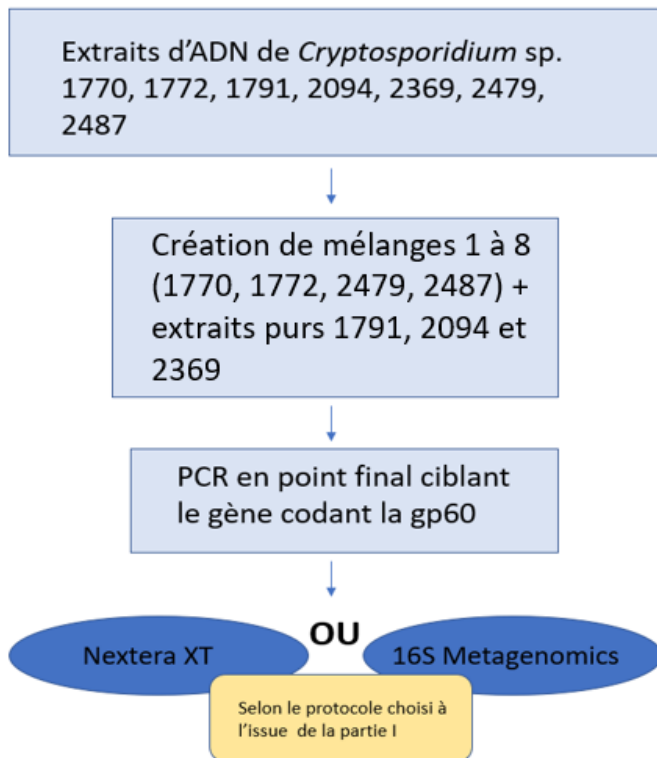


Figure 11 : Etapes expérimentales de la partie II de Matériels et Méthodes

3) Troisième partie : Etude de mélanges d'ADN de *Cryptosporidium* sp. par PCR en temps réel et par séquençage selon la technique Sanger.

Dans cette troisième partie, nous nous sommes attachés à étudier différents mélanges d'ADN de *Cryptosporidium* sp. avec les techniques de diagnostic actuellement déployées en routine au laboratoire de parasitologie du C.H.U. de Dijon. Les mélanges ont donc été traités comme le seraient de véritables échantillons cliniques à visée diagnostique, en étant d'abord identifiés au rang d'espèce par PCR en temps réel ciblant le gène de la sous-unité de l'ARN ribosomal, puis séquencés par séquençage Sanger pour déterminer leur sous-type.

Ils ont donc d'abord été amplifiés par PCR en temps réel afin de voir si, dès l'étape de l'identification d'espèce, il était possible de détecter les mélanges d'ADN. Ensuite, les différents échantillons ont été séquencés selon la technique Sanger afin de voir d'une part si cette technique permettait de détecter les différents variants au sein des mélanges, et d'autre part jusqu'à quelle proportion.

a) Choix des échantillons

a-1) Echantillons issus d'une épidémie de cryptosporidiose à Divonne

Une épidémie mixte de cryptosporidiose a été observée en 2003 dans la ville de Divonne les Bains, entraînant un nombre de malades estimé à 1100 cas (99). L'origine de cette épidémie était la contamination du réseau d'eau distribué à la population. Quatre extraits d'ADN, analysés à l'époque au laboratoire de parasitologie du C.H.U. de Dijon, ont été conservés et congelés à -80°C pendant ces deux dernières décennies.

Les échantillons correspondants numérotés 23, 24, 26, et 45, ont été appelés D23, D24, D26 et D45 pour la suite de ce travail. En 2003, ils avaient été identifiés comme suit au laboratoire de parasitologie du CHU de Dijon :

Echantillon D23 : mélange de *C. parvum* et *C. hominis*

Echantillon D24 : *C. hominis* seul

Echantillon D26 : mélange de *C. parvum* et *C. hominis*

Echantillon D45 : mélange de *C. parvum* et *C. hominis*

Il est important de préciser qu'en 2003, le diagnostic de ce mélange d'ADN avait été réalisé par clonage. En résumé, l'ADN extrait de selles était cloné dans un plasmide contenant un gène de résistance à l'ampicilline, puis transformé dans une bactérie (*Escherichia coli*), chaque bactérie contenant un seul plasmide incluant un « clone » d'ADN de *Cryptosporidium* sp. associé au gène de résistance à l'ampicilline. Après mise en culture de ces bactéries sur un milieu gélosé contenant de l'ampicilline, seules les bactéries contenant le plasmide (et donc le gène de résistance à l'ampicilline) étaient capables de croître. Ainsi, l'ADN de *Cryptosporidium* sp., obtenu à partir de 10 colonies de bactéries indépendantes, était séquencé selon la méthode Sanger afin d'investiguer la présence éventuelle de différents clones pouvant correspondre à différentes espèces et/ou sous-types de *Cryptosporidium* sp. A cette période, les génotypes I et II de *C. parvum* correspondaient aux actuels *C. hominis* et *C. parvum* respectivement. Le laboratoire avait donc rendu des selles positives aux génotypes I et II de *C. parvum* avec une proportion de 70% pour le génotype I et 30% pour le génotype II dans une des selles (99).

De plus, le sous-typage par séquençage du gène codant la protéine gp60 n'a pas été utilisé lors de l'investigation de cette épidémie. Enfin, à cette période, l'identification au rang d'espèce était obtenue par PCR en point final ciblant un fragment de l'ADN codant la petite sous-unité 18S de l'ARN ribosomique, qui ne donnait aucune information sur la quantité d'ADN de *Cryptosporidium* sp. présente dans les échantillons. A l'inverse, la technique de PCR en temps réel utilisée en routine actuellement au laboratoire permet une détermination semi-quantitative de l'ADN de *Cryptosporidium* sp. présente dans les échantillons par l'étude.

a-2) Echantillons fortement positifs à *Cryptosporidium* sp.

Ensuite, des échantillons fortement positifs (*i.e.* dont le CT était inférieur à 20) obtenus au laboratoire de parasitologie du C.H.U. de Dijon par PCR en temps réel ciblant le gène codant la sous-unité de l'ARN ribosomal, ont été sélectionnés pour créer les mélanges d'ADN mimant une infection mixte à *Cryptosporidium* sp. Ainsi, les échantillons 2055 (*C. hominis* IbA10G2 positif à 18C_T) et 1572 (*C. parvum* IlcA5G3 positif à 17C_T) ont été sélectionnés. Ces échantillons d'ADN étaient congelés au laboratoire de parasitologie à -20°C pour 2055 et -80°C 1572, en fonction de l'ancienneté de leur obtention.

a-3) Validation du choix des échantillons

Avant de passer à la réalisation des mélanges d'ADN de *Cryptosporidium* sp., les extraits d'ADN sélectionnés ont été traités individuellement afin d'en valider le choix. Ainsi, l'ADN de *Cryptosporidium* sp. contenu dans ces échantillons était amplifié par la technique de PCR en temps réel ciblant le gène de la sous-unité de l'ARN ribosomal actuellement utilisée au laboratoire de parasitologie du CHU de Dijon, afin (i) de vérifier la présence d'une amplification de l'ADN (donc de vérifier l'intégrité de l'ADN après une conservation par congélation), et (ii) de déterminer (pour les échantillons D23, D26 et D45) ou vérifier (pour les échantillons 1572 et 2055) le seuil de détection des produits d'amplification objectif par la valeur du C_T.

La PCR en temps réel utilisée au laboratoire de parasitologie du C.H.U. de Dijon permet la détection et l'identification des principales espèces de *Cryptosporidium* sp. Elle repose sur l'amplification par PCR d'une partie du gène codant pour la petite sous-unité de l'ARN ribosomal (18S ARNr), en utilisant les amorces Cry80F3 et Cry337R. La détection des produits

de PCR utilise des sondes d'hybridation CryAnch et CrySens. L'identification des espèces est réalisée par l'analyse des courbes de fusion réalisée à la fin de la PCR : la sonde CrySens est complémentaire d'une séquence d'ADN variable chez les différentes espèces, et la température de fusion de la sonde et de l'ADN sera variable en fonction de l'espèce. L'amplification est réalisée sur l'automate Light Cycler 2.0, les réactifs nécessaires à la réalisation de cette PCR sont présentés dans le tableau 10.

Tableau 9 : Réactifs nécessaires à la PCR en temps réel ciblant le gène codant la sous unité 18S de l'ARN ribosomal

Réactifs	Volumes pour un échantillon (µL)
Master Mix	2.0
Amorce Cry80F3	1.0
Amorce Cry337R	1.0
Sonde CryAnch	0.4
Sonde CrySens	0.4
MgCl ₂	1.6
UNG	0.5
Eau	8.1
Total	15.0

Deux types de courbes sont issus de cette PCR pour chaque échantillon :

- La courbe d'amplification, reflet de l'amplification en temps réel des produits de PCR en fonction du nombre de cycles d'amplification, permettant de déterminer le C_T.
- La courbe de fusion, objectivant l'évolution de la fluorescence en fonction de la température. Celle-ci permet de déterminer la température de fusion (T_m) définie comme étant le point d'inflexion de la courbe de fusion. Chaque espèce de *Cryptosporidium* sp. a une température de fusion (T_m) attendue. La détermination de cette T_m permet donc l'identification de l'espèce.

T_m attendus :

C. hominis = 61.5°C

C. parvum = 53.5°C

C. meleagridis = 57°C

C. felis = 50°C

La figure 12 permet de résumer cette partie qui nous a permis de valider le choix de nos échantillons, en vérifiant leur intégrité et la valeur de leur C_T .

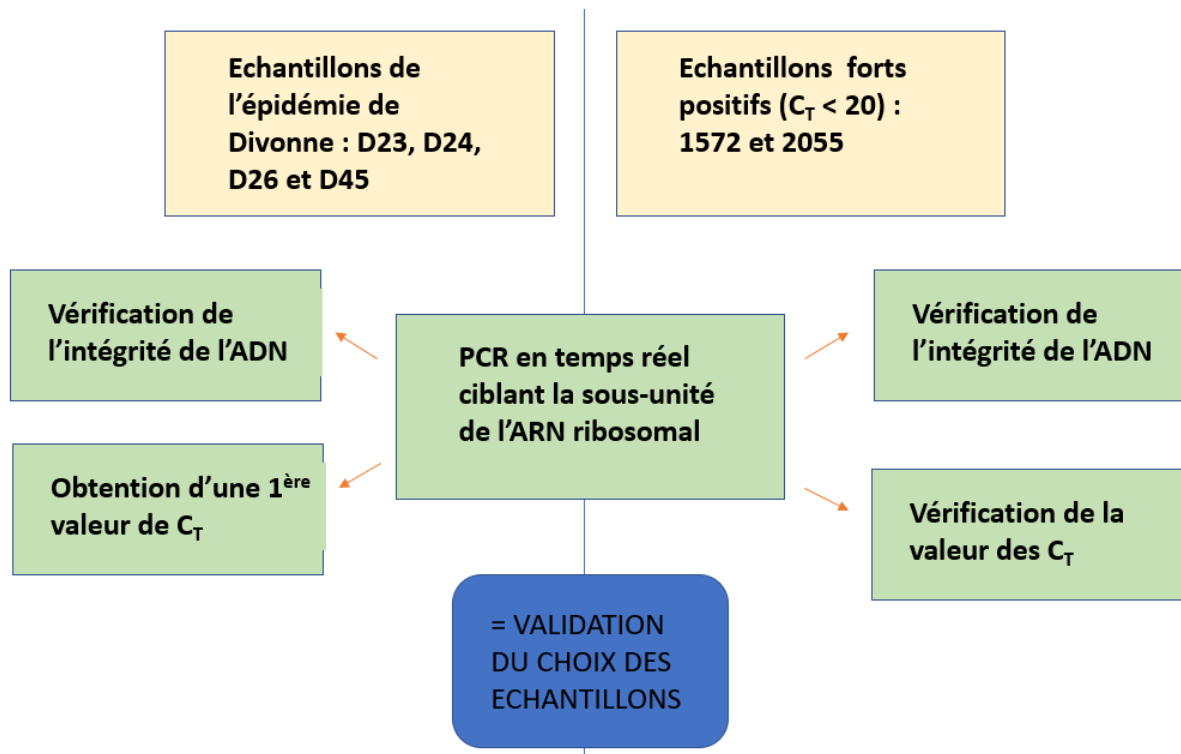


Figure 12 : Schéma montrant l'intérêt de la PCR en temps réel ciblant le gène codant la sous-unité 18S de l'ARN ribosomal pour la validation des échantillons

b) Schémas de dilution pour les mélanges d'ADN de *Cryptosporidium* sp.

Après nous être assurés de l'intégrité de nos extraits d'ADN grâce à la PCR en temps réel ciblant le gène codant la sous-unité de l'ARN ribosomal, le but était de réaliser des mélanges d'ADN de *C. hominis* (1572) et *C. parvum* (2055), avec l'objectif d'évaluer la capacité de la PCR en temps réel ciblant le gène codant la sous-unité de l'ARN ribosomal et du séquençage par la technique de Sanger à les détecter. Ces mélanges ont été réalisés directement à partir des extraits d'ADN de 1572 (*C. parvum*) et 2055 (*C. hominis*).

L'extrait d'ADN de 1572 a été conservé pur et a été dilué au 1/10^e, 1/100^e et 1/1000^e dans différents tubes. L'extrait d'ADN de 2055 a été conservé pur et dilué de la même manière dans différents tubes. Des mélanges volume à volume ont ensuite été réalisés, en prenant 5µL

de 1572 (pur ou dilué) et 5µL de 2055 (pur ou dilué), permettant d'obtenir le schéma de constitution des mélanges exposé dans le tableau 11.

Tableau 10 : Schéma de constitution des mélanges d'ADN (A à G) pour le troisième essai.

Mélanges	1572	2055
A	Pur	Pur
B	Pur	Dilué au 1/10 ^e
C	Pur	Dilué au 1/100 ^e
D	Pur	Dilué au 1/1000 ^e
E	Dilué au 1/10 ^e	Pur
F	Dilué au 1/100 ^e	Pur
G	Dilué au 1/1000 ^e	Pur

Dans le mélange B, 5µL d'ADN de 1572 pur étaient mélangés avec 5µL d'ADN de 2055 dilué au 1/10^e.

Avant de réaliser les mélanges d'ADN, nous avons vérifié la présence d'une amplification pour chaque dilution (1/10^e, 1/100^e et 1/1000^e) des échantillons 1572 et 2055 isolément. Ainsi, les extraits dilués de 1572 et 2055 ont été amplifiés individuellement par PCR en point final ciblant le gène codant la protéine gp60, puis migrés sur gel d'agarose 2% afin de vérifier la présence d'une bande signant une amplification significative. Cette étape permettait ainsi de vérifier que de l'ADN de *Cryptosporidium* sp. était détectable par PCR aux dilutions utilisées pour chacun des extraits, et ainsi de valider notre schéma de dilutions avant réalisation des mélanges.

c) PCR en temps réel sur les mélanges d'ADN de différentes espèces de *Cryptosporidium* sp.

Comme indiqué dans la figure 13, les mélanges d'ADN de *C. parvum* (1572) et *C. hominis* (2055) constitués directement à partir des extraits d'ADN étaient destinés à être amplifiés d'une part par PCR en temps réel ciblant le gène codant la sous-unité 18S de l'ARN ribosomal, et d'autre part par PCR en point final ciblant le gène codant la protéine gp60.

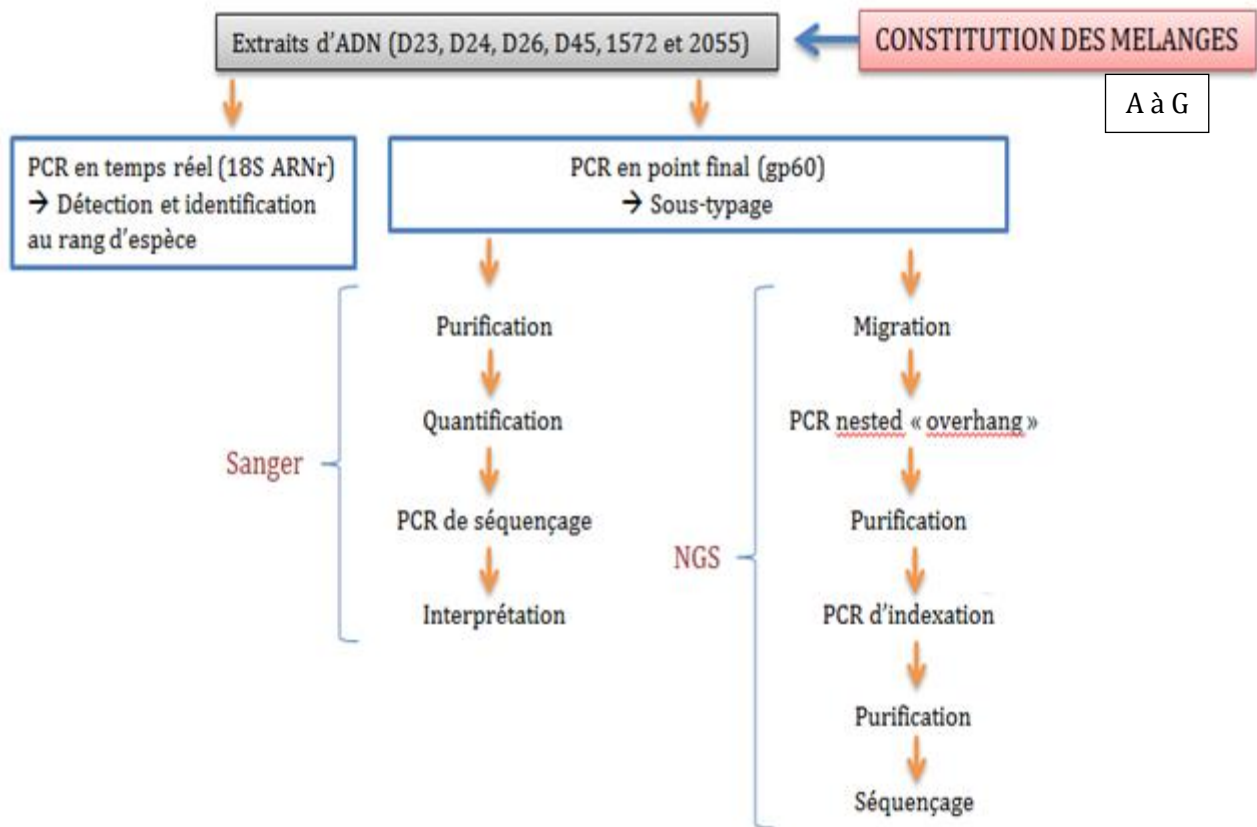


Figure 13 : Schéma illustrant le déroulement de la troisième partie de nos essais

Les mélanges A à G ont ainsi été amplifiés par PCR en temps réel ciblant le gène codant la sous unité 18S de l'ARN ribosomal de *Cryptosporidium*, afin d'en étudier les courbes d'amplification et de fusion et ainsi évaluer jusqu'à quelle proportion un mélange peut être détecté par cette technique. Notons qu'un mélange dans un échantillon se traduirait par un double pic sur sa courbe de fusion.

La PCR en temps réel ciblant la sous-unité de l'ARN ribosomal a été réalisée dans les mêmes conditions que celle utilisée pour conforter notre choix d'échantillons en partie a-3) (Page 70).

d) PCR en point final ciblant le gène codant la gp60

Comme l'illustre la figure 13, les quatre échantillons d'ADN issus de l'épidémie de Divonne les Bains (D23, D24, D26 et D45) et les sept mélanges d'ADN (A à G) ont également subi une amplification par PCR en point final ciblant le gène codant la protéine gp60. Cette PCR a permis de générer des amplicons pour ces onze échantillons, permettant de réaliser ensuite

une PCR de séquençage par technique Sanger. La PCR en point final ciblant le gène codant la protéine gp60 était la même que celle utilisée dans nos premiers essais (décrite en annexe 1).

e) Séquençage par la technique Sanger appliquée aux mélanges d'ADN

Les échantillons D23, D24, D26, D45, 1572, 2055 ainsi que les mélanges A à G ont ainsi été amplifiés par PCR en point final ciblant le gène codant la protéine gp60.

Les produits de PCR ont subi une migration sur un gel d'agarose à 4%. Les échantillons pour lesquels ont été obtenues plusieurs bandes, un découpage de ces bandes par excision du gel d'agarose et une purification individuelle était nécessaire, et ce grâce au kit GeneJET gel extraction kit (ThermoFisher). Quant aux échantillons ne présentant qu'une bande unique en migration, leur purification a été réalisée sur colonne Amicon Ultra 30K (MerckMillipore).

Après quantification, les amplicons ainsi purifiés ont été séquencés par la technique Sanger. Ce séquençage a été réalisé comme décrit dans l'annexe 1.

4) Analyse bio-informatique

Les séquences obtenues par séquençage NGS ont été comparées à une base de données que nous avons construite, regroupant les séquences consensus des différents sous-types connus de *Cryptosporidium* sp. Cette base de données de référence était basée sur celle de la publication de Zahedi et al. (15) et comprenait 131 séquences (annexe 2) issues de la base GenBank. Les séquences des sous-types IIAA18G1R1 et IIAA20G1R1, ainsi que celles des espèces *C. cuniculus* et *C. erinacei* n'étant pas présentes dans la base de données construite par Zahedi et al. (15), nous avons réalisé des recherches supplémentaires sur GenBank afin de compléter au mieux la base de données. L'analyse bio-informatique a été réalisée à l'aide de la version 1.6 de la méthode DADA2. Pour cela, les données générées par séquençage sur MiSeq (Illumina) ont été préalablement regroupées sous forme de fichiers individuels FastQ et les nucléotides non biologiques ont été retirés (amorces, index et adaptateurs). La méthode DADA2 permet de contrôler la qualité des séquences obtenues (notamment en supprimant celles dont le score de Phred, un score de qualité de la séquence nucléotidique, est inférieur à 30), fusionner les séquences appariées et éliminer les chimères. Enfin, les séquences sont

assignées à la séquence de référence correspondante dans la base de données. Le produit final de la méthode DADA2 est un fichier .ASV qui regroupe le nombre de chaque variante de séquences obtenues et à quelles séquences de référence ces variantes correspondent.

Notons qu'à l'issue de la comparaison des séquences obtenues par NGS à celles de notre base de données, celles qui sortaient finalement en « non appariées » étaient comparées à la base de données BLAST afin d'essayer de leur associer une identification.

III) Résultats

L'analyse bio-informatique des données de séquençage a permis d'obtenir pour chaque échantillon un fichier sous format .ASV que nous avons étudié sous Excel. Chaque fichier Excel permet la comparaison des séquences de référence extraites de la base de données GenBank, avec les séquences retrouvées lors du séquençage. Les séquences retrouvées par le séquençage sont classées par groupes de séquences identiques, et sont associées à la séquence de référence correspondante. Pour illustration, pour un échantillon peut être retrouvé :

	A	B	C
1	Sous-types de référence	Séquences obtenues par séquençage	Nombre de copies de séquences identiques
2	<i>C. hominis</i> IbA10G2_AY262031	TCCGCTGTATTCTCAGCCCCAGCCGTTCCAC	150 000
3	<i>C. hominis</i> IbA10G2_AY262031	TCCGCTGTATTCTCAGCCCCAGCCGTTCCAC	25 000
4	<i>C. hominis</i> IbA20G2_FJ153240	TCCGCTGTATTCTCAGCCCCAGCCGTTCCAC	14 000
5	<i>C. hominis</i> IbA12G3_KF679725	TCCGCTGTATTCTCAGCCCCAGCCGTTCCAC	200

Figure 14 : Exemple de fichier .ASV obtenu après séquençage NGS

Pour cet échantillon, un groupe de 150000 séquences identiques obtenues au cours du séquençage correspondent parfaitement à la séquence de référence de *C. hominis* IbA10G2. Un second groupe de 25000 séquences identiques est également identifié comme *C. hominis* IbA10G2. Ces séquences diffèrent de celles du premier groupe par 1 ou 2 pb d'insertions ou délétions, mais n'empêche pas la correspondance fiable à la séquence de référence (>99% de similitude de séquences). Pour plus de clarté, ces groupes de séquences ne différant que de quelques bases mais correspondant finalement au même sous-type seront réintégrées en un seul groupe dans la suite du manuscrit. Ainsi, on peut dire ici que 175000 séquences obtenues au cours du séquençage correspondent à *C. hominis* IbA10G2. Un troisième groupe de 14000 séquences identiques est identifié comme *C. hominis* IbA20G2. Enfin, le quatrième groupe de

séquences identiques obtenues est identifié comme *C. hominis* IbA12G3 et ne représente que 0.1% du nombre total de séquences obtenues.

1) Première partie : Comparaison des deux protocoles de préparation de librairie

Pour chaque protocole de préparation de librairie, trois fichiers ont été obtenus pour les échantillons 1761, 1770 et 1772. Dans chaque fichier, les différentes séquences identifiées comme appartenant au même sous-type ont été regroupées. Les résultats sont présentés dans le tableau 12.

Tableau 11 : Présentation des résultats obtenus avec le protocole de préparation de librairie Nextera XT (Illumina) :

Les sous-types en rouge sont les sous-types majoritaires identifiés par séquençage NGS, en se basant sur le nombre de séquences appariées.

	Nombre de séquences totales	Nombre de séquences appariées	Espèces et sous types identifiés par NGS	Sous-types identifiés par Sanger
Echantillon 1761	159783	157721 (98.7%)	<i>C. parvum</i> IIcA5G3	<i>C. parvum</i> IIcA5G3
		2062 (1.3%)	35 autres sous-types	
Echantillon 1770	161026	87592 (54%)	<i>C. parvum</i> IIaA18G1R1	<i>C. parvum</i> IIaA18G1R1
		15203 (10%)	<i>C. parvum</i> IIcA5G3	
		13216 (9%)	<i>C. parvum</i> IIaA19G2R1	
		12703 (8%)	<i>C. parvum</i> IIaA16G2R1	
		11010 (7%)	<i>C. parvum</i> IIaA15G2R2	
		10566 (6%)	<i>C. parvum</i> IIaA15G2R1	
		9918 (5.5%)	<i>C. parvum</i> IIaA20G3R1	
		818 (0.5%)	78 autres sous-types	
Echantillon 1772	201304	146929 (73%)	<i>C. hominis</i> IbA10G2	<i>C. hominis</i> IbA10G2
		20471 (10%)	<i>C. hominis</i> IbA20G2	
		18843 (9%)	<i>C. hominis</i> IbA12G3	
		6820 (3%)	<i>C. hominis</i> IbA9G3	
		5205 (2%)	<i>C. hominis</i> IbA6G3	
		2185 (1%)	<i>C. hominis</i> IfA22G1	
		851 (0.4%)	25 autres sous-types	

Les sous-types en rouge sont les sous-types majoritaires identifiés par séquençage NGS, en se basant sur le nombre de séquences appariées.

Au total, 522113 séquences de qualité ont été obtenues par séquençage haut débit des trois échantillons.

Pour rappel, le sous-type de l'échantillon 1761 identifié par séquençage selon la technique Sanger est *C. parvum* IIcA5G3. Un nombre total de 159783 séquences ont été obtenues par séquençage NGS de cet échantillon. Parmi elles, 157721 séquences soit 98.7% correspondent à la séquence de référence de *C. parvum* IIcA5G3. Parmi les 2062 séquences restantes, des groupes de séquences représentant chacun moins de 1% du nombre total de séquences ont été retrouvés et été assignés à 35 sous-types différents additionnels.

Pour l'échantillon 1771, 54% des séquences obtenues par séquençage NGS ont montré un résultat de sous-typage identique au sous-typage par Sanger, à savoir *C. parvum* IIaA18G1R1. Néanmoins, d'autres groupes de séquences correspondant à des sous-types différents ont été retrouvés dans des proportions variant de 0.1% à 10% du nombre total de séquences obtenues. 818 séquences présentaient un nombre de copies inférieur à 1%, aboutissant à l'identification de 85 sous-types au total pour l'échantillon 1770.

Parmi les 201304 séquences totales obtenues pour l'échantillon 1772, 73% correspondaient à la séquence de référence de *C. hominis* IbA10G2. D'autres groupes de séquences correspondant à 30 sous-types additionnels sont représentés en proportion variable, entre 0.1% et 10%.

De même, trois fichiers ont été obtenus pour les trois échantillons préparés par le protocole de préparation de librairie 16S Metagenomics Sequencing Library Preparation (Illumina). L'analyse des différents résultats sont regroupés dans le tableau 13.

Tableau 12 : Présentation des résultats obtenus avec le protocole de préparation de librairie 16S Metagenomics Sequencing Library Preparation (Illumina).

	Nombre de séquences totales	Nombre de séquences appariées	Espèces et sous types identifiés par NGS	Sous-types identifiés par Sanger
Echantillon 1761	42421	42421 (100%)	<i>C. parvum</i> IIcA5G3	<i>C. parvum</i> IIcA5G3
Echantillon 1770	49986	49122 (98.3%)	<i>C. parvum</i> IIaA18G1R1	<i>C. parvum</i> IIaA18G1R1
		864 (1.7%)	<i>C. parvum</i> IaA11G2R1	
Echantillon 1772	30238	30238 (100%)	<i>C. hominis</i> IbA10G2	<i>C. hominis</i> IbA10G2

Les sous-types en rouge sont les sous-types majoritaires identifiés par séquençage NGS, en se basant sur le nombre de séquences appariées.

Le séquençage des trois échantillons a permis d'obtenir 122645 séquences de qualité.

100% des séquences issues du séquençage de l'échantillon 1761, soit 42421 séquences, correspondaient au sous-type *C. parvum* IIcA5G3. Il s'agit du même sous-type que celui retrouvé par séquençage par la technique Sanger.

Le séquençage de l'échantillon 1770 a quant à lui permis de retrouver le sous-type *C. parvum* IIaA18G1R1. En effet, 98.3% des séquences totales correspondaient à ce sous-type, identique à celui retrouvé par la technique Sanger. Notons qu'un sous-type différent a été identifié, *C. parvum* IIaA12G1R1, représentant 1.7% du nombre total de séquences.

Enfin, le séquençage de l'échantillon 1772 a identifié uniquement des séquences correspondant au sous-type IbA10G2, correspondant à celui identifié suite au séquençage par la technique Sanger.

A l'issue de ces deux essais, le protocole de préparation de librairie 16S Metagenomics Sequencing Library Preparation (Illumina) était plus en accord avec les résultats obtenus par séquençage selon la technique Sanger. En effet, un seul sous-type a été retrouvé avec ce protocole pour les échantillons 1761 et 1772, et deux sous-types pour l'échantillon 1770. Avec le protocole Nextera XT (Illumina), le nombre de séquences correspondant au sous-type *C.*

parvum IlaA18G1R1 n'ont atteint que 54% des séquences totales et 84% de sous-types additionnels étaient identifiés, avec des pourcentages de fréquence de séquences allant de 0.1% à 10%. La technique de préparation de bibliothèques utilisant le protocole 16S Metagenomics Sequencing Library Preparation (Illumina) semble ainsi mieux adaptée que le protocole Nextera XT (Illumina). Nos observations corroborent le fait que le protocole 16S Metagenomics Sequencing Library Preparation (Illumina) est celui utilisé dans la plupart des publications traitant de techniques NGS appliquées à *Cryptosporidium* sp. Pour ces deux raisons, le protocole 16S Metagenomics Sequencing Library Preparation (Illumina) a été utilisé dans la suite de nos travaux.

2) Deuxième partie : Etude de mélanges d'ADN de différentes espèces et/ou sous-types de *Cryptosporidium* par NGS.

Les échantillons 1791, 2369, 2094, ainsi que les mélanges d'ADN 1 à 8 avaient dans un premier temps été amplifiés par PCR en point final ciblant le gène codant la protéine gp60. Cette première PCR a été suivie par une seconde PCR nichée dite « overhang » ciblant le gène codant la protéine gp60, en accord avec le protocole de préparation de bibliothèque 16S Metagenomics Sequencing Library Preparation (Illumina) choisi précédemment. Les amplicons issus de ces deux amplifications ont été migrés sur gel d'agarose 2%, et cette migration est retrouvée dans la figure 17.



Figure 15 : Migration sur gel d'agarose 2% après PCR « overhang » ciblant le gène codant la protéine gp60 sur les échantillons 1 à 8, 1791, 2094 et 2369.

Il est à souligner que la migration sur un gel d'agarose à 2% n'était pas suffisamment discriminante pour mettre en évidence des bandes séparées dans les mélanges 1 à 8 où deux (voire trois) espèces pourtant distinctes sont présentes.

De plus, l'échantillon 1791 n'a pas présenté d'amplification significative après amplification, comme en atteste l'absence de bande d'amplification sur le gel. Cet échantillon a toutefois été conservé jusqu'à la fin de notre essai.

Les échantillons ont donc tous été séquencés dans le MiSeq (Illumina). Dans le tableau 14, nous présentons d'abord les résultats obtenus par NGS sur les mélanges d'ADN 1 à 8.

Tableau 13 : Présentation des résultats obtenus par NGS sur les mélanges 1 à 8 lors du second essai.

	Nombre de séquences totales	Nombre de séquences appariées	Espèces et sous types	Proportions dans l'échantillon
Mélange 1	152468	95211 (62%)	<i>C. parvum</i> IlaA18G1R1	85%
		55533 (36%)	<i>C. hominis</i> IbA10G2	15%
		1724 (2%)	<i>C. parvum</i> IlaA11G2R1	0%
Mélange 2	287924	194803 (67%)	<i>C. hominis</i> IbA10G2	60%
		89199 (31%)	<i>C. parvum</i> IlaA18G1R1	40%
		3922 (2%)	<i>C. parvum</i> IlaA11G2R1	0%
Mélange 3	263257	160645 (61%)	<i>C. hominis</i> IbA10G2	75%
		99188 (37%)	<i>C. parvum</i> IlaA18G1R1	25%
		3424 (2%)	<i>C. parvum</i> IlaA11G2R1	0%
Mélange 4	705914	507304 (72%)	<i>C. hominis</i> IbA10G2	95%
		185942 (26%)	<i>C. parvum</i> IlaA18G1R1	5%
		11665 (1.6%)	<i>C. parvum</i> IlaA11G2R1	0%
Mélange 5	281537	266154 (94%)	<i>C. hominis</i> IbA10G2	98%
		13162 (5%)	<i>C. parvum</i> IlaA18G1R1	2%
		2221 (1%)	<i>C. parvum</i> IlaA11G2R1	0%
Mélange 6	478854	478830 (100%)	<i>C. hominis</i> IbA10G2	99%
		0 (0%)	<i>C. parvum</i> IlaA18G1R1	1%
Mélange 7	121384	121384 (100%)	<i>C. hominis</i> IbA10G2	99.9%
		0 (0%)	<i>C. parvum</i> IlaA18G1R1	0.1%
Mélange 8	232804	202456 (87%)	<i>C. parvum</i> IlaA15G2R1	65%
		30348 (13%)	<i>C. parvum</i> IlaA20G1R1	25%
		0 (0%)	<i>C. parvum</i> IlaA18G1R1	10%

Le mélange 1 est composé de 85% d'ADN issu de l'échantillon 1770 (*C. parvum* IIAA18G1R1) et de 15% d'ADN issu de l'échantillon 1772 (*C. hominis* IBA10G2). Le nombre total de séquences obtenues lors du séquençage et retenues après analyse par la méthode DADA2 est de 152468. Parmi elles, 62% des séquences correspondent au sous-type *C. parvum* IIAA18G1R1 et 36% au sous-type *C. hominis* IBA10G2. Un autre sous-type a été identifié, *C. parvum* IIAA11G2R1, à hauteur de 2% des séquences totales.

Le mélange 2 est composé de 60% d'ADN de *C. hominis* IBA10G2 et de 40% d'ADN de *C. parvum* IIAA18G1R1. A l'issue du séquençage NGS de ce mélange, les deux espèces sont retrouvées et représentent 194803 séquences (67%) et 89199 séquences (31%) respectivement. Un sous-type supplémentaire, *C. parvum* IIAA12G1 est détecté à hauteur de 2% des séquences totales.

Le mélange 3 est composé de 25% d'échantillon 1770 (*C. parvum* IIAA18G1R1) et 75% d'échantillon 1772. Le sous-type majoritairement détecté à l'issue du séquençage est *C. hominis* IBA10G2, à 61%. Le sous-type *C. parvum* IIAA18G1R1 également présent dans le mélange est détecté à hauteur de 99188 séquences (37%) correspondant à la séquence de référence. Un sous-type supplémentaire, *C. parvum* IIAA12G1 a été identifié à hauteur de 2% des séquences totales.

Le mélange 4 a été préparé à l'aide de 95% d'ADN de *C. hominis* (1772) et 5% d'ADN de *C. parvum* (1770). Le séquençage NGS de ce mélange a permis d'identifier 507304 séquences correspondant à *C. hominis* IBA10G2 et 185942 séquences correspondant à *C. parvum* IIAA18G1R1. Un sous type supplémentaire, *C. parvum* IIAA18G2R1 a été retrouvé à hauteur de 1.6% des séquences totales.

Dans les mélanges 5, 6 et 7, l'ADN de *C. parvum* IIAA18G1R1 est présent à 2%, 1% et 0.1% respectivement. Le séquençage haut débit du mélange 5 a permis d'identifier des séquences de ce sous-type, ce qui n'est pas le cas pour les mélanges 6 et 7 dans lesquels il n'a pas été mis en évidence. Dans ces trois mélanges, le sous-type majoritaire a été identifié avec succès.

Enfin, le mélange 8 est composé de trois sous-types de *C. parvum* génétiquement proches, en proportion de 10% (*C. parvum* IIaA18G1R1), 25% (*C. parvum* IIaA20G1R1) et 65% (*C. parvum* IIaA15G2R1). Seuls les deux sous-types majoritaires du mélange ont été détectés par séquençage. Parmi les 232804 séquences totales obtenues lors du séquençage, 86% des séquences ont été identifiées comme *C. parvum* IIaA15G2R1 et 13% comme du *C. parvum* IIaA20G1R1. Le sous-type *C. parvum* IIaA15G2R1, dont l'ADN était présent à 10% dans le mélange 8 n'a pas été détecté.

Tableau 14 : Présentation des résultats obtenus par NGS sur les échantillons 1791, 2094 et 2369.

	Nombre de séquences totales	Nombre de séquences appariées	Espèces et sous types	Proportion dans l'échantillon
1791	126	114 (90%)	<i>C. erinacei</i> XIIIaA21R1	100%
		12 (10%)	<i>C. hominis</i> IbA10G2	0%
2094	22247	20223 (90%)	<i>C. meleagridis</i> IIIgA21G3R1	100%
		2024 (10%)	Non appariées	0%
2369	142271	142271 (100%)	<i>C. cuniculus</i> VbA24	100%

Nous nous intéressons maintenant aux trois espèces plus rares que sont *C. erinacei* XIIIaA21R1 (1791), *C. meleagridis* IIIgA21G3R1 (2094) et *C. cuniculus* VbA24 (2369). Les résultats obtenus par séquençage NGS sont présentés dans le tableau 15.

Le séquençage de l'échantillon 1791 a rapporté 114 séquences correspondant à *C. erinacei* XIIIaA21R1 et représentant 90% des séquences obtenues. Cependant, le nombre de séquences générées lors du séquençage était faible, probablement dû à la très faible amplification de l'échantillon objectivée lors de la migration sur gel d'agarose (Figure 17 page 80). De plus, 10% des séquences obtenues ont permis d'identifier *C. hominis* IbA10G2, non identifié initialement dans l'échantillon. Ce résultat reste difficile à interpréter compte tenu du très faible nombre de séquences générées (12 séquences).

Le séquençage des échantillons 2094 et 2369 a permis d'identifier avec succès l'espèce et le sous-type initialement identifiés par séquençage selon la méthode Sanger. Cependant, concernant l'échantillon 2094, 10% des séquences (soit 2024 séquences) n'étaient pas appariées.

Il est néanmoins important de préciser que l'identification de ces espèces a nécessité de comparer les séquences obtenues par NGS à la base de données du logiciel BLAST, révélant les limites de la base de données initialement constituée pour ce travail. En effet, de nombreuses séquences étaient classées « non appariées » à l'issue du séquençage NGS, et c'est seulement après comparaison aux séquences incrémentées dans le logiciel BLAST qu'il nous a été possible de les identifier. Cela révèle qu'il est important de comparer les séquences non appariées à celles incrémentées dans le logiciel BLAST avant de conclure. Ici, 2024 séquences obtenues par séquençage de l'échantillon 2094 sont restées « non appariées » après recherche sur le logiciel BLAST.

3) Troisième partie : Etude de mélanges d'ADN de *Cryptosporidium* sp. par PCR en temps réel ciblant le gène codant la sous-unité de l'ARN ribosomal et par séquençage selon la technique Sanger.

Les échantillons choisis pour cette troisième partie étaient les échantillons D23, D24, D26, D45, 1572 et 2055. Une PCR en temps réel ciblant la sous-unité de l'ARN ribosomal a permis de valider notre choix d'échantillons. Elle a permis (i) de vérifier que l'ADN de *Cryptosporidium* sp. initialement détecté dans des extraits de selles étaient toujours amplifiables après conservation et (ii) évaluer de façon semi-quantitative pour la première fois la quantité d'ADN parasitaire des échantillons de Divonne les Bains (D23, D24, D26 et D45). Les résultats obtenus par cette PCR en temps réel sont présentés dans le tableau 16.

Tableau 15 : Présentation des résultats obtenus par PCR en temps réel ciblant le gène codant la sous-unité de l'ARN ribosomal

Echantillons	C _T obtenus	Identifications attendues	Identifications obtenues (T _m)
D23	X	<i>C. hominis</i> + <i>C. parvum</i>	<i>C. hominis</i> (61.6)
D24	23.6	<i>C. hominis</i>	<i>C. hominis</i> (60.9)
D26	20	<i>C. hominis</i> + <i>C. parvum</i>	<i>C. parvum</i> (53.4) + <i>C. hominis</i> (61.2)
D45	31	<i>C. hominis</i> + <i>C. parvum</i>	<i>C. parvum</i> (53.6)
1572	16.2	<i>C. parvum</i>	<i>C. parvum</i> (53.4)
2055	17.4	<i>C. hominis</i>	<i>C. hominis</i> (60.6)

Les courbes d'amplification et les courbes de fusion obtenues par PCR en temps réel sur les échantillons sont présentées dans les figures 18 et 19. En plus des échantillons de l'épidémie de Divonne, les échantillons 1572 et 2055 ont été testés par PCR en temps réel ciblant le gène de la sous-unité de l'ARN ribosomal afin de vérifier la stabilité après conservation de la quantité d'ADN de *Cryptosporidium* sp. Leur forte positivité a été de nouveau confirmée avec des valeurs de C_T similaires à celles initialement obtenues (*i.e.* C_T < 20 cycles), confirmant la possibilité de les utiliser pour réaliser les mélanges d'ADN dans les essais ultérieurs.

L'échantillon D23 a été très faiblement amplifié, mais avec une valeur de C_T au-delà de 45 cycles. Néanmoins, sa courbe de fusion a permis d'affirmer que l'échantillon contenait bien de l'ADN de *Cryptosporidium* sp. en faible quantité. A ce titre, cet échantillon a été conservé et utilisé dans la suite de notre travail comme échantillon faiblement positif.

Des mélanges d'ADN de *C. parvum* et *C. hominis* étaient attendus pour les échantillons D23, D26 et D45. Seuls les résultats de l'amplification par PCR de l'échantillon D26 ont permis de mettre en évidence l'existence d'un mélange de deux espèces de *Cryptosporidium* sp. La faible concentration en ADN des échantillons D23 et D45 (faible positif à 31 C_T) pourrait expliquer cette absence de détection des deux espèces de *Cryptosporidium* sp.

Amplification Curves

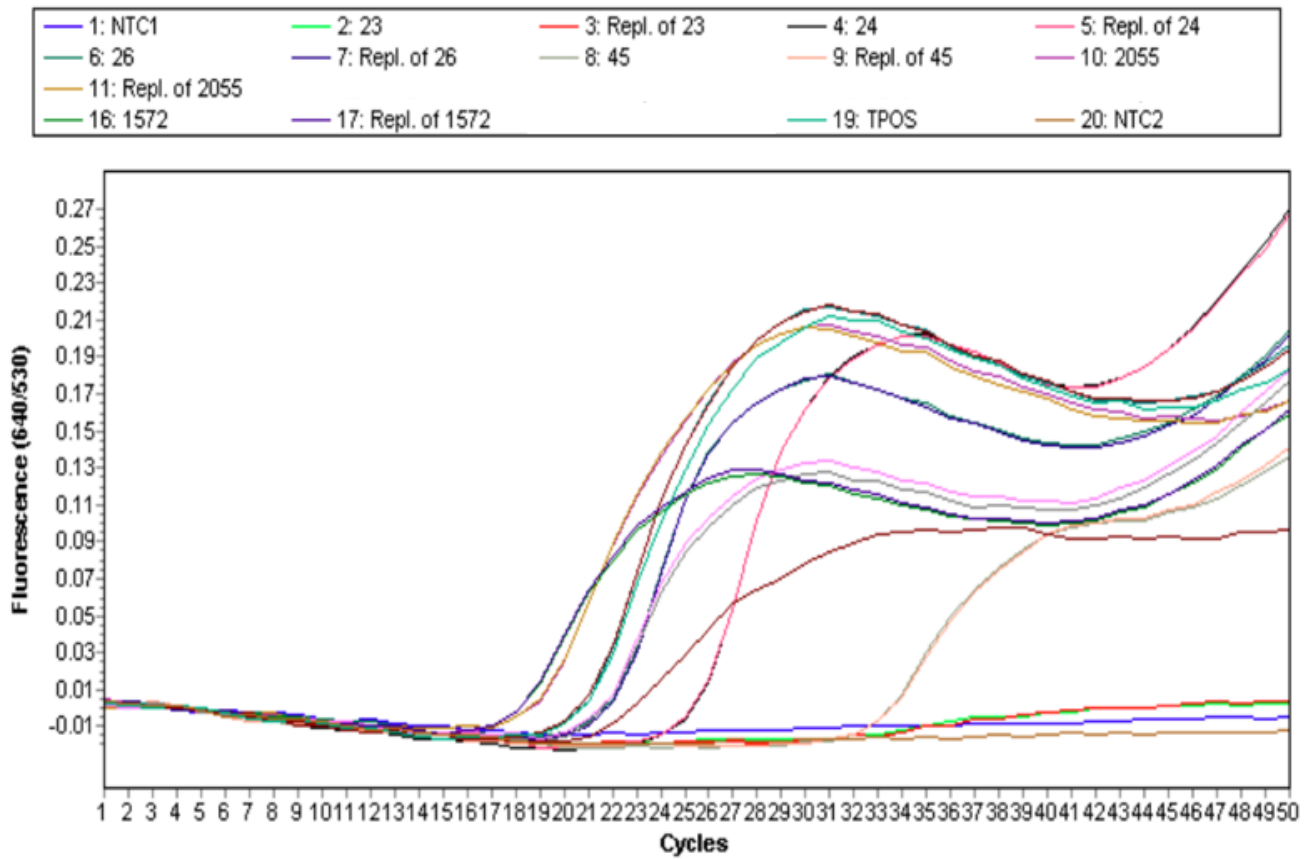


Figure 16 : Courbes d'amplification obtenues par PCR en temps réel réalisée sur les échantillons D23, D24, D26, D45, 1572 et 2055.

NTC1 et NTC2 = contrôles négatifs

26 = D26

Repl. of 26 = Répétition de l'échantillon D26

TPOS = témoin positif

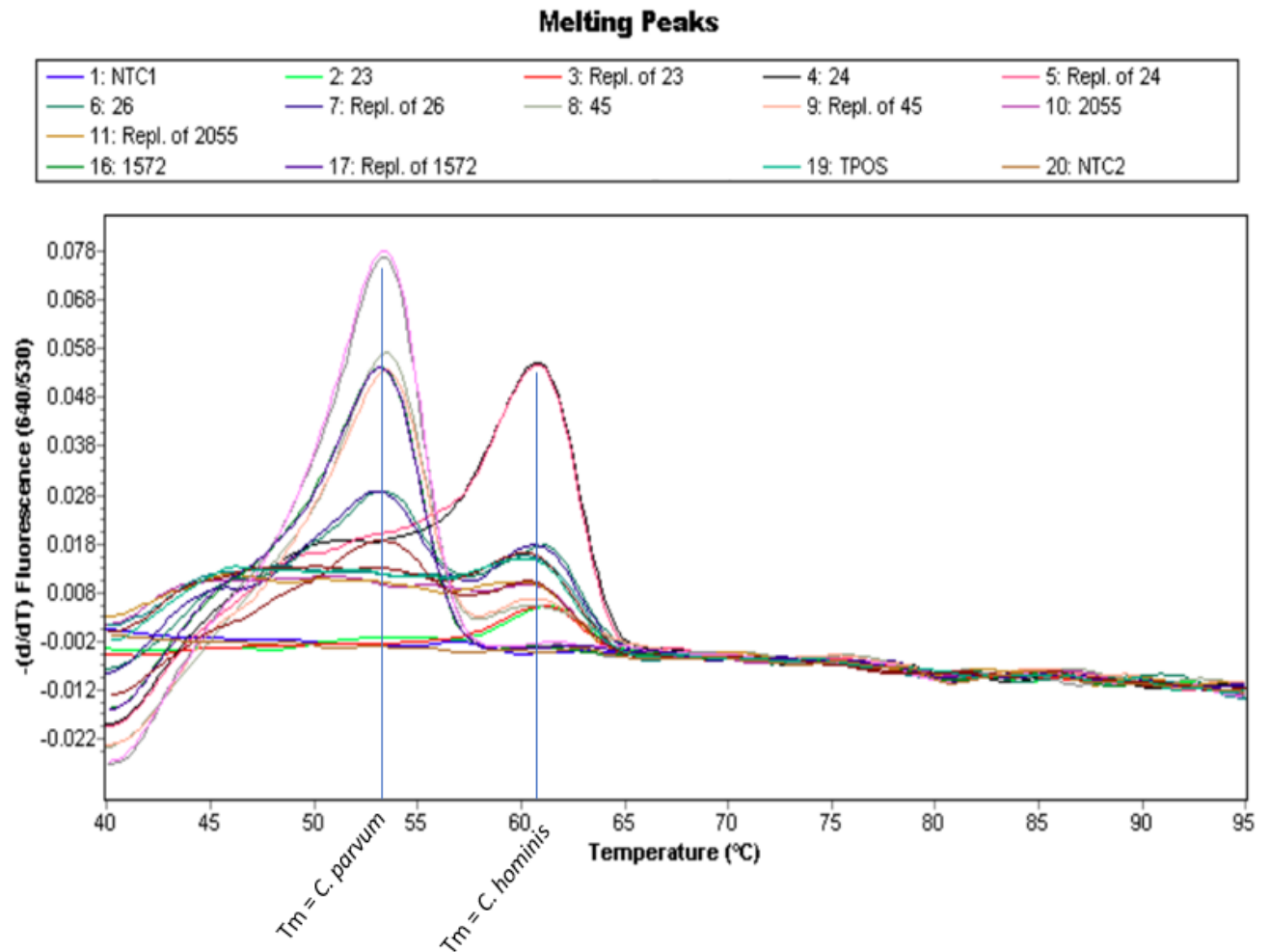


Figure 17 : Pics de fusion (obtenus à partir des courbes de fusion) issus de la PCR en temps réel réalisée sur les échantillons D23, D24, D26, D45, 1572 et 2055.

NTC1 et NTC2 = contrôles négatifs

26 = D26

Repl. of 26 = Répétition de l'échantillon D26

TPOS = témoin positif

Le choix des échantillons 1572 et 2055 ayant été validé par cette PCR en temps réel, les dilutions envisagées ont pu leur être appliquées afin de créer des mélanges d'ADN de deux espèces différentes : *C. parvum* et *C. hominis*.

Des dilutions au 1/10^e, 1/100^e et 1/1000^e ont été réalisés directement sur les extraits d'ADN des échantillons 1572 et 2055 (comme décrit dans le tableau 11 page 73). Les produits de dilution ainsi obtenus ont été amplifiés individuellement par PCR en point final ciblant le gène codant la protéine gp60. Les amplicons obtenus à l'issue de cette PCR ont été migrés sur

gel d'agarose à 2%. Cela a permis de s'assurer que le schéma de dilution élaboré était correct et permettait une amplification même dans les dilutions les plus fortes, correspondant aux concentrations d'ADN les plus faibles. Les résultats de migration sont présentés sur la figure 20.

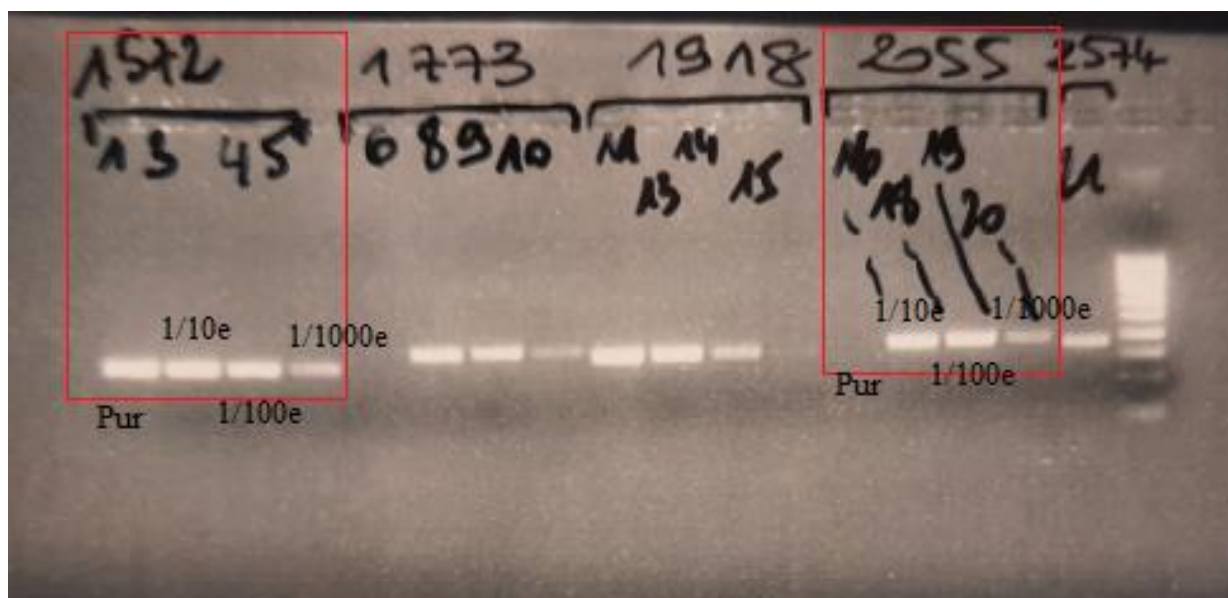


Figure 18 : Migration des produits de la PCR en point final ciblant le gène codant la protéine gp60 sur gel d'agarose 2%

Le cadre rouge à gauche met en évidence les bandes de migration des produits de PCR obtenus à partir de l'extrait d'ADN de 1572 dilué au 1/10e, 1/100e et 1/1000e - Le cadre rouge à droite met en évidence les bandes de migration des produits de PCR obtenus à partir de l'extrait d'ADN de 2055 dilué au 1/10e, 1/100e et 1/1000e.

Après migration des amplicons de la PCR en point final ciblant le gène codant la protéine gp60 sur gel d'agarose 2%, nous observons que l'échantillon 1572 a été amplifié avec succès par la PCR en point final ciblant le gène codant la protéine gp60, et ce pour toutes les dilutions, y compris les plus fortes (*i.e.* au 1/1000^e). Ces observations suggèrent que cette PCR en point final permet l'amplification d'ADN de *Cryptosporidium* sp. présent dans l'échantillon 1572, même à des concentrations faibles, et donc que cette PCR ne serait pas un facteur limitant pour la suite de nos essais.

Pour ce qui est de l'échantillon 2055, les produits de dilution au 1/10^e, 1/100^e et 1/1000^e ont été amplifiés avec succès. Néanmoins, l'échantillon 2055 n'a pas pu être amplifié pur, probablement en raison de la présence d'inhibiteurs de la PCR dans l'extrait d'ADN. Afin de lever l'effet de l'inhibiteur, nous avons dilué l'échantillon 1572 au 1/5^{ème}. Or, comme cela l'a

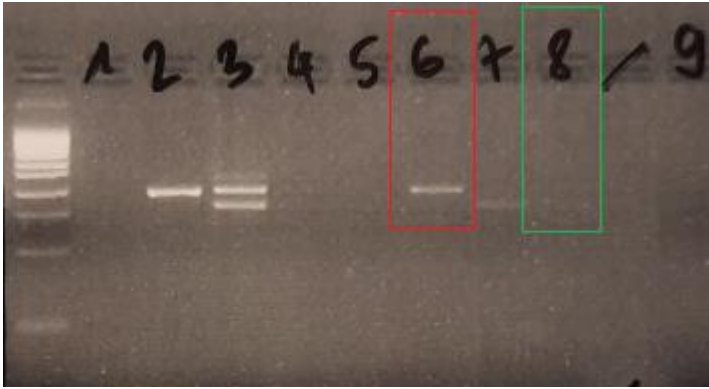
été décrit dans le tableau 11 (page 73), les mélanges A, E, F et G devaient contenir 5µL d'ADN de 2055 pur. La constitution de ces mélanges a donc été modifiée en tenant compte du facteur de dilution au 1/5^e de 2055, afin de conserver les proportions envisagées. Par exemple, dans le mélange A il était envisagé de mélanger 5µL d'ADN de 1572 pur et 5µL d'ADN de 2055 pur. L'ADN de 2055 devant être dilué au 1/5^e pour lever l'effet de l'inhibiteur de la PCR, l'ADN de 1572 devait être dilué au 1/5^e également pour respecter les proportions du mélange A. Le nouveau schéma des mélanges d'ADN de 1572 et 2055 est présenté dans le tableau 17.

Tableau 16 : Schéma de construction des mélanges envisagé après constatation de la présence d'un inhibiteur de PCR dans l'échantillon 2055 pur.

Mélanges	1572	2055
A	Dilué au 1/5 ^e	Dilué au 1/5 ^e
B	Pur	1/10 ^e
C	Pur	1/100 ^e
D	Pur	1/1000 ^e
E	Dilué au 1/50 ^e	Dilué au 1/5 ^e
F	Dilué au 1/500 ^e	Dilué au 1/5 ^e
G	Dilué au 1/5000 ^e	Dilué au 1/5 ^e

Les valeurs en rouge correspondent aux valeurs qui ont été modifiées en tenant compte de la dilution de l'échantillon 2055 au 1/5^e.

La dilution au 1/5^e de l'échantillon 2055 a donc entraîné une modification du schéma de constitution des mélanges A, E, F et G. L'ADN de l'échantillon 1572 a été dilué au 1/5000^e puis a été amplifié par PCR en point final ciblant le gène codant la protéine gp60. Le produit de PCR a été migré sur gel d'agarose à 4% afin de vérifier la présence d'une amplification par la PCR malgré sa faible concentration. La figure 21 montre le résultat de cette migration.



Echantillon 1572 dilué au 1/5000^e = puits n°8 – bande très discrète sur la photo
 Echantillon 2055 dilué au 1/5^e = puits n°6

Figure 19 : Migration sur gel d'agarose 4% des produits de PCR en point final ciblant le gène codant la protéine gp60 appliquée à l'échantillon 2055 dilué au 1/5^e et à l'échantillon 1572 dilué au 1/5000^e

Comme le montre la migration des produits de la PCR en point final ciblant le gène codant la protéine gp60 sur gel d'agarose, la dilution au 1/5^e de l'échantillon 1572 a permis de lever l'effet de l'inhibiteur de PCR. Cela est objectivé par la présence d'une bande de migration sur le gel. De plus, une bande discrète est apparue lors de la migration du produit de PCR de l'échantillon 1572 dilué au 1/5000^e. Cette constatation suggère que la PCR en point final permet l'amplification d'ADN de *Cryptosporidium* sp. présent dans l'échantillon 1572 dilué au 1/5000^e. Cela a permis de valider notre nouveau schéma de mélanges.

Les mélanges A à G constitués d'ADN de *C. parvum* (1572) et *C. hominis* (2055) ont donc pu être étudiés par PCR en temps réel (18S ARNr) et par séquençage par la technique Sanger (après amplification par PCR en point final ciblant le gène codant la protéine gp60).

Les résultats de la PCR en temps réel ciblant la sous-unité de l'ARN ribosomal appliquée aux mélanges d'ADN A à G sont exposés dans le tableau 18 et les figures 22 et 23.

Amplification Curves

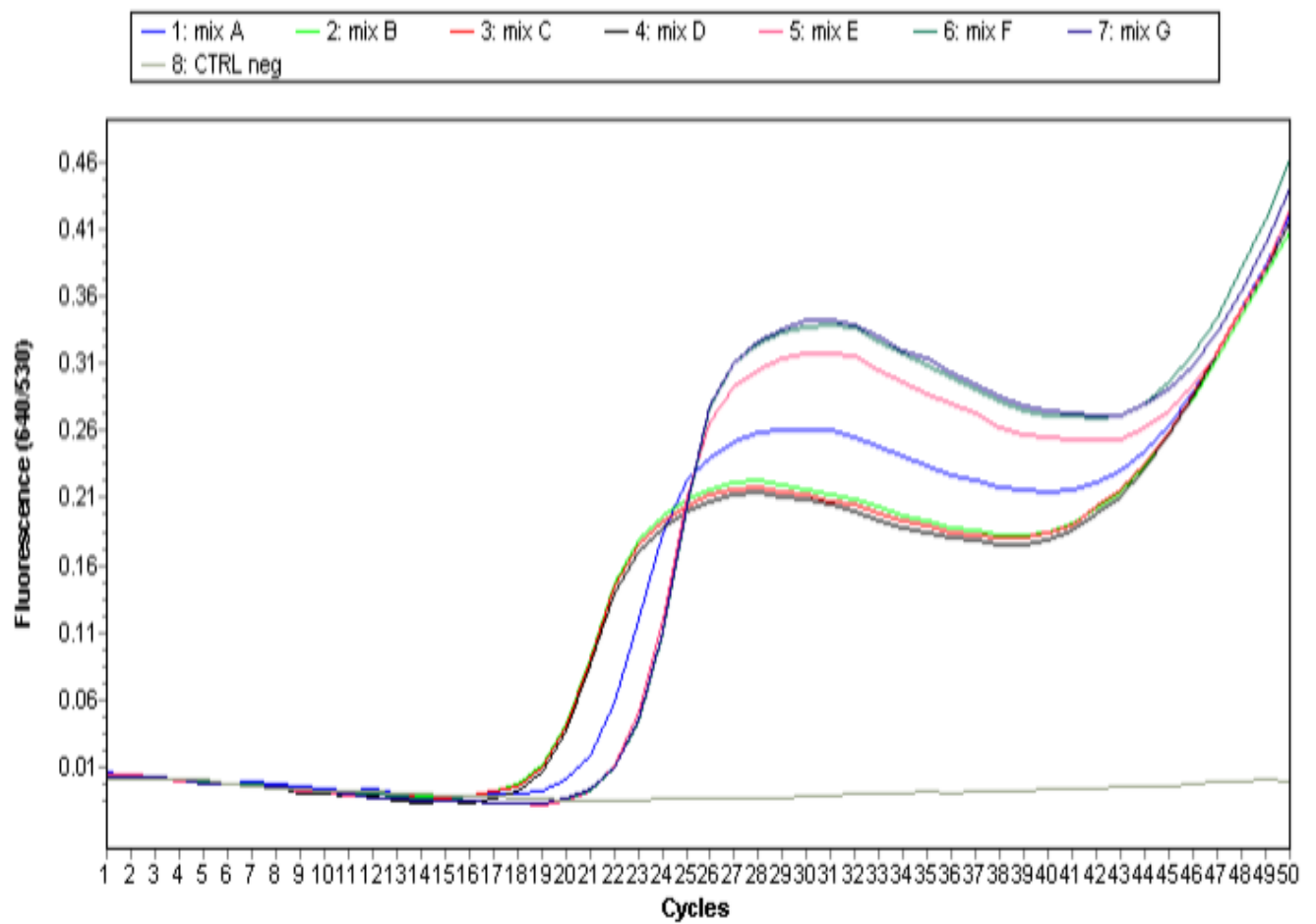


Figure 20 : Courbes d'amplification issues de la PCR en temps réel appliquée aux mélanges d'ADN de *C. parvum* et *C. hominis* (A à G).

Mix A = Mélange A

CTRL neg = Contrôle négatif

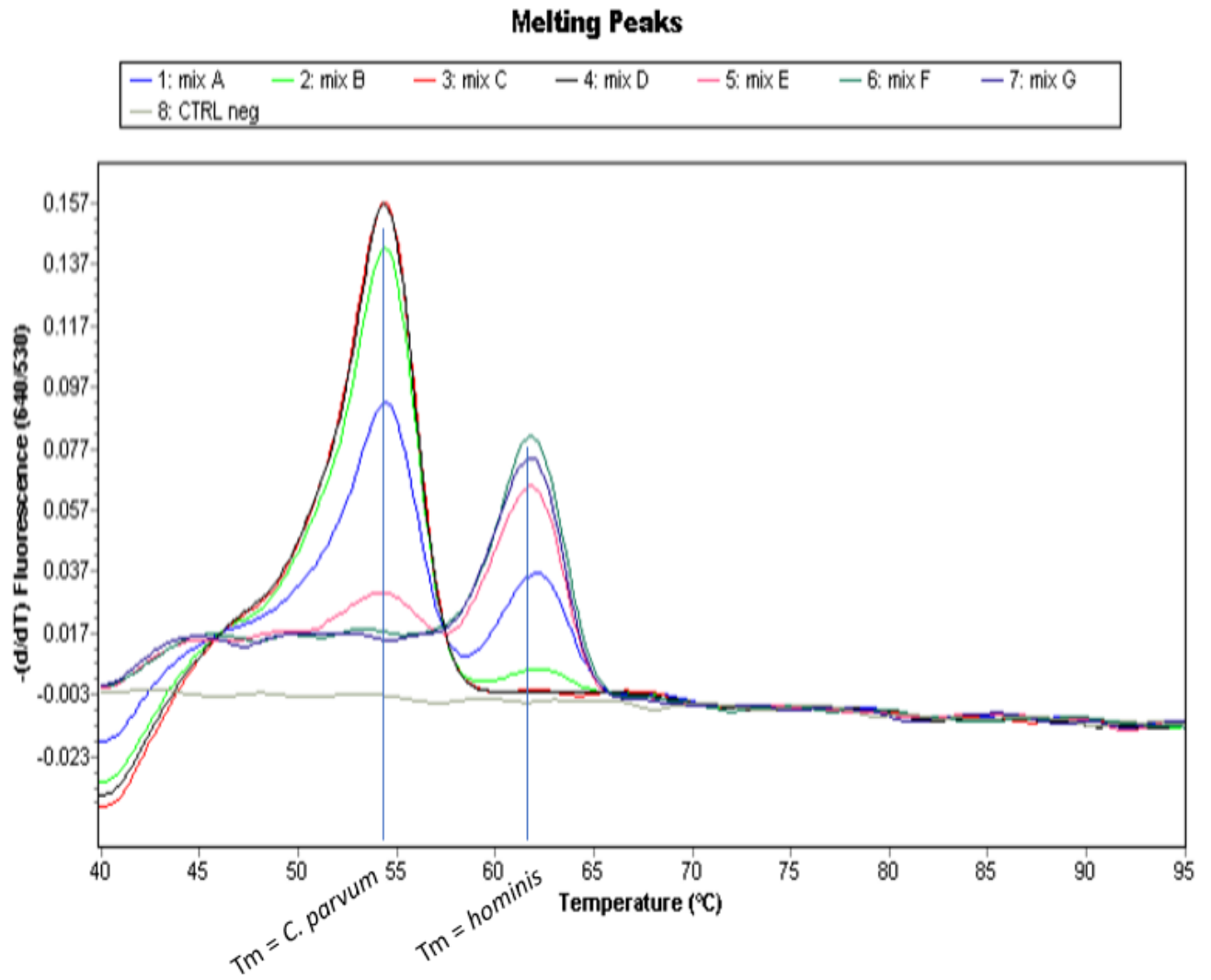


Figure 21 : Pics de fusion issus de la PCR en temps réel appliquée aux mélanges d'ADN de *C. parvum* et *C. hominis* (A à G).

Mix A = Mélange A

CTRL neg = Contrôle négatif

Tableau 17 : Résultats obtenus par PCR en temps réel ciblant le gène codant la sous-unité 18S de l'ARN ribosomal sur les mélanges A à G du troisième essai.

Mélanges	C _T obtenus	Composition des mélanges	Identifications obtenues (T _m)
A	19.03	<i>C. hominis</i> (50%) + <i>C. parvum</i> (50%)	<i>C. hominis</i> (61.9) + <i>C. parvum</i> (54.2)
B	17.10	<i>C. hominis</i> (10%) + <i>C. parvum</i> (90%)	<i>C. hominis</i> (61.9) + <i>C. parvum</i> (54.3)
C	17.09	<i>C. hominis</i> (1%) + <i>C. parvum</i> (99%)	<i>C. parvum</i> (54.4)
D	17.08	<i>C. hominis</i> (0.1%) + <i>C. parvum</i> (99.9%)	<i>C. parvum</i> (54.5)
E	20.28	<i>C. hominis</i> (90%) + <i>C. parvum</i> (10%)	<i>C. hominis</i> (61.8) + <i>C. parvum</i> (54.2)
F	20.57	<i>C. hominis</i> (99%) + <i>C. parvum</i> (1%)	<i>C. hominis</i> (61.9)
G	20.58	<i>C. hominis</i> (99.9%) + <i>C. parvum</i> (0.1%)	<i>C. hominis</i> (61.6)

Les mélanges d'ADN de *C. parvum* et *C. hominis* ont été détectés avec succès par la PCR en temps réel dans les échantillons A, B et E. Cela se traduit par un double pic sur la courbe de fusion et donc deux valeurs de T_m pour ces trois échantillons. Pour les mélanges C, D, E et G, seuls les variants majoritaires ont été identifiés.

Autrement dit, la PCR en temps réel est capable de détecter un variant d'espèce minoritaire dont l'ADN propre représente au moins 10% du mélange d'ADN d'espèces différentes de *Cryptosporidium* présentes dans l'échantillon. En effet, dans notre essai, lorsque le variant minoritaire était présent à 0.1% ou 1% de l'ADN total de *Cryptosporidium* sp., la PCR en temps réel ciblant le gène codant la sous-unité de l'ARN ribosomal permettait uniquement l'identification du variant majoritaire dans le mélange, à savoir *C. parvum* pour les mélanges C et D, et *C. hominis* pour les mélanges F et G.

Enfin, les mélanges d'ADN de *C. parvum* et *C. hominis* A à G, ainsi que les quatre échantillons issus de l'épidémie de Divonne les Bains (D23, D24, D26, D45) ont été amplifiés par PCR en point final ciblant le gène codant la protéine gp60 puis séquencés selon la technique Sanger. Une migration des produits de PCR obtenus après première amplification a été réalisée sur un gel d'agarose à 4%. L'image obtenue est présentée figure 24.



Figure 22 : Migration sur gel d'agarose 4% des produits de PCR en point final ciblant le gène codant la protéine gp60 appliquée aux mélanges d'ADN de *C. parvum* et *C. hominis* (A à G).

Pour les échantillons A, B et E, deux bandes ont été observées sur le gel d'agarose, ce qui a permis de les isoler par excision du gel, de les purifier et de les séquencer séparément. Quant aux échantillons C, D, F et G, ils ne présentaient qu'une seule bande en migration. C'est donc la totalité des produits de PCR qui a été purifiée et séquencée pour ces 4 échantillons.

Les résultats de séquençage selon la technique Sanger obtenus sont présentés dans le tableau 19. Un exemple de chromatogramme et de séquences déduites sont présentés figures 25 à 28.

Tableau 18 : Résultats du séquençage selon la technique Sanger des mélanges A à G et des échantillons de Divonne les Bains (D23, D24, D26, D45) lors du troisième essai.

Mélanges	Composition des mélanges	Résultats du séquençage Sanger
A	<i>C. hominis</i> (50%) + <i>C. parvum</i> (50%)	<i>C. hominis</i> IbA10G2 + <i>C. parvum</i> IIcA5G3
B	<i>C. hominis</i> (10%) + <i>C. parvum</i> (90%)	<i>C. hominis</i> IbA10G2 + <i>C. parvum</i> IIcA5G3
C	<i>C. hominis</i> (1%) + <i>C. parvum</i> (99%)	<i>C. parvum</i> IIcA5G3
D	<i>C. hominis</i> (0.1%) + <i>C. parvum</i> (99.9%)	<i>C. parvum</i> IIcA5G3
E	<i>C. hominis</i> (90%) + <i>C. parvum</i> (10%)	<i>C. hominis</i> IbA10G2 + <i>C. parvum</i> IIcA5G3
F	<i>C. hominis</i> (99%) + <i>C. parvum</i> (1%)	<i>C. hominis</i> IbA10G2
G	<i>C. hominis</i> (99.9%) + <i>C. parvum</i> (0.1%)	<i>C. hominis</i> IbA10G2
D23	<i>C. hominis</i> + <i>C. parvum</i>	X
D24	<i>C. hominis</i>	<i>C. hominis</i> IbA10G2
D26	<i>C. hominis</i> + <i>C. parvum</i>	<i>C. parvum</i> IIcA5G3 + <i>C. hominis</i> IbA10G2
D45	<i>C. hominis</i> + <i>C. parvum</i>	<i>C. parvum</i> IIcA5G3 + <i>C. hominis</i> IbA10G2

Les deux espèces *C. parvum* et *C. hominis* ont été identifiées avec succès dans les mélanges A, B et E. Seule l'espèce majoritaire a été identifiée dans les autres mélanges constitués de ces deux espèces.

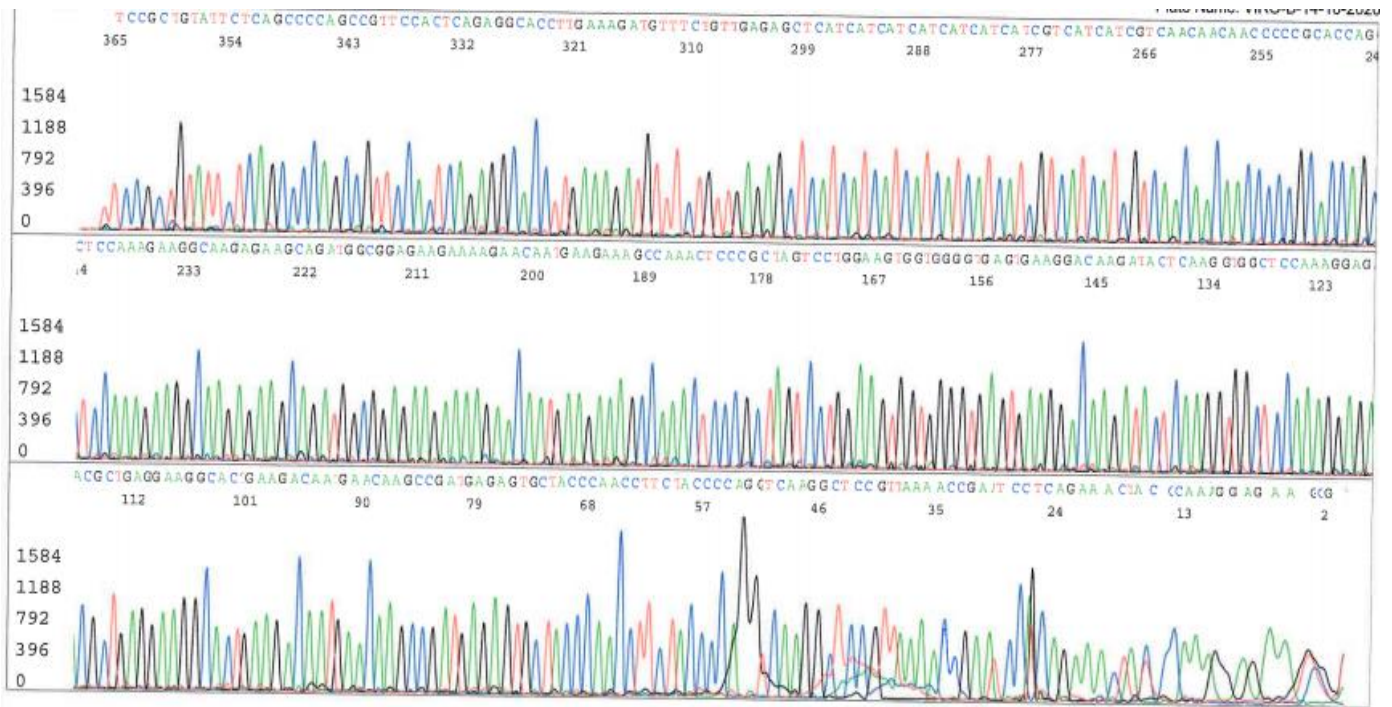


Figure 23 : Chromatogramme obtenu par le séquençage selon la technique Sanger de l'échantillon G en anti-sens (en reverse complément).

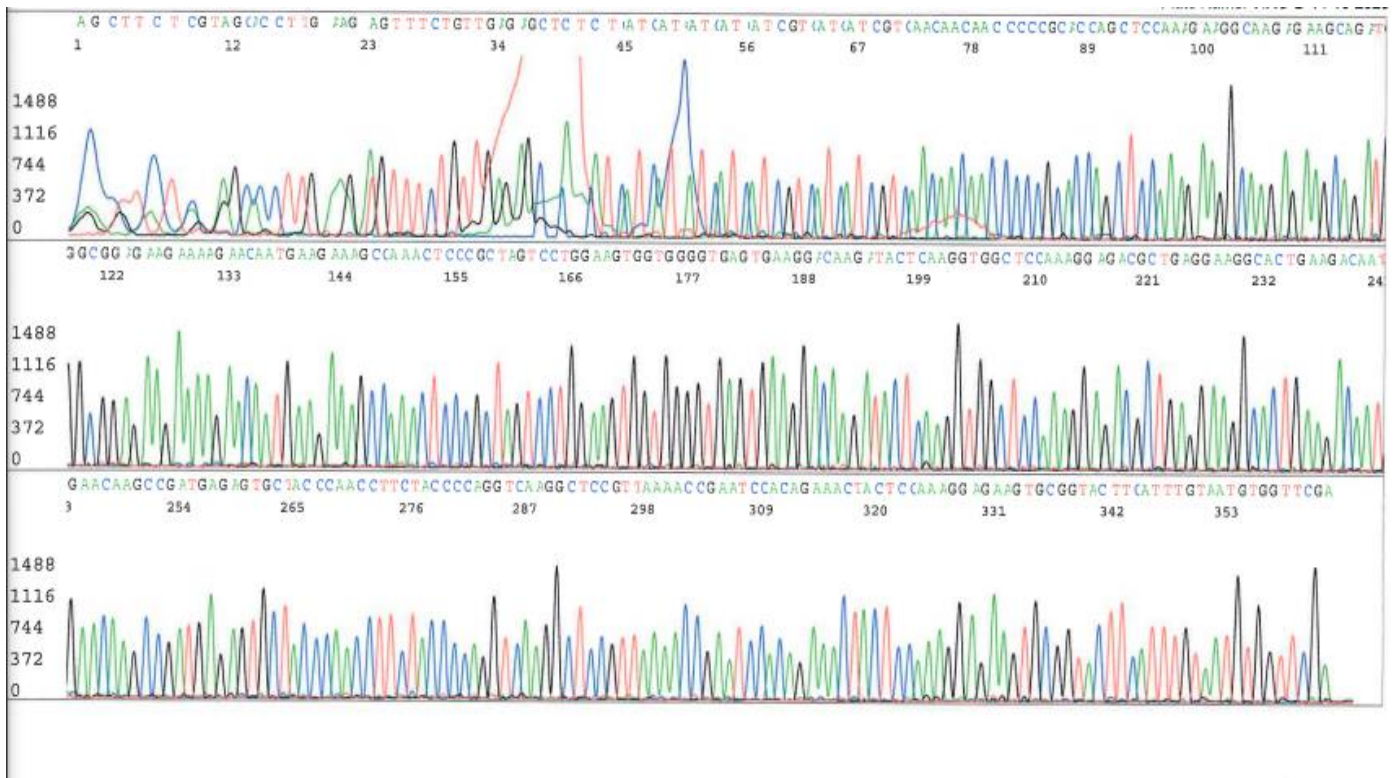


Figure 24 : Chromatogramme obtenu par séquençage par la technique Sanger de l'échantillon G en sens.

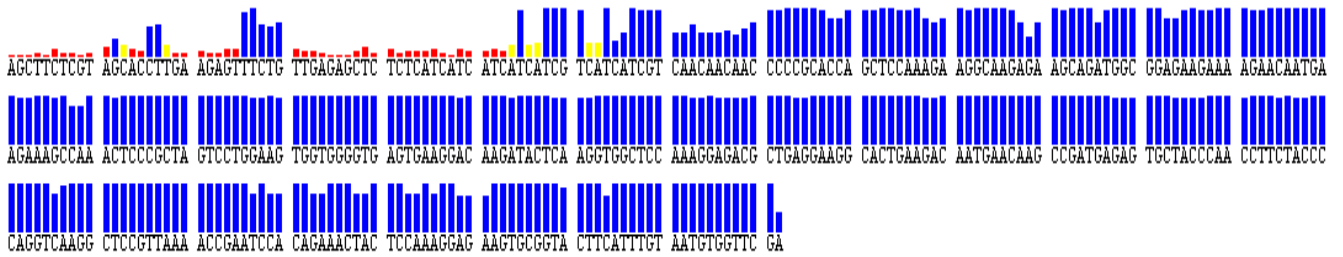


Figure 25 : Séquence obtenue par séquençage par la technique Sanger de l'échantillon G en sens

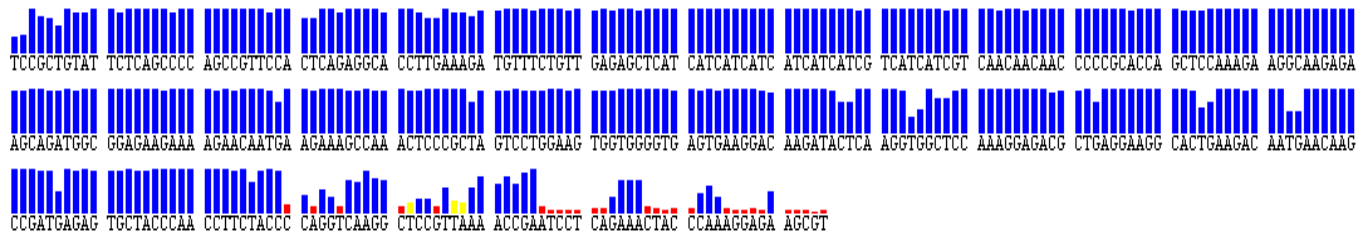
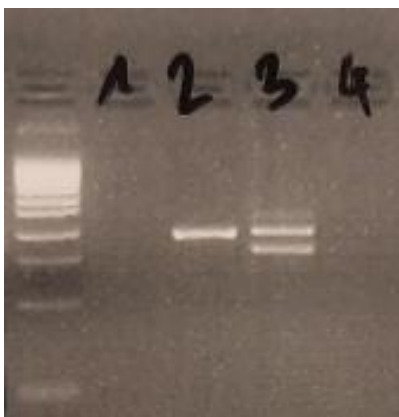


Figure 26 : Séquence obtenue par séquençage par la technique Sanger de l'échantillon G en anti-sens.



Puits n°1 = D23
 Puits n°2 = D24
 Puits n°3 = D26
 Puits n°4 = D45

Figure 27 : Migration sur gel d'agarose 4% des produits de PCR en point final ciblant le gène codant la protéine gp60 pour les échantillons de Divonne les Bains

Concernant les échantillons de Divonne les Bains (D23, D24, D26, D45), la première PCR en point final ciblant le gène codant la protéine gp60 n'a pas permis d'amplifier suffisamment l'échantillon D23 et D45. Cela est objectivé sur la figure 29 par l'absence de bande de migration concernant les amplicons de D23 et D45. Pour ces deux échantillons une seconde PCR nichée en point final identique à la première a donc été réalisée. Cette seconde PCR a permis d'amplifier avec succès l'échantillon D45 qui a ensuite pu être séquencé, mais pas l'échantillon D23 qui est resté négatif. Cet échantillon a néanmoins été passé en séquençage.

Pour l'échantillon D24, une seule bande de migration a été retrouvée sur gel d'agarose, en accord avec l'identification attendue d'une seule espèce de *Cryptosporidium*. Quant à l'échantillon D26, deux bandes ont été visualisées sur le gel d'agarose, qui ont été séparées par excision du gel, puis purifiées et séquencées séparément. La figure 29 présente les résultats de la migration sur gel des échantillons de Divonne les Bains après la première PCR.

Le séquençage par la technique Sanger de ces 4 échantillons a permis de retrouver les mélanges d'espèces *C. parvum* et *C. hominis* attendus pour les échantillons D26 et D45. Notons que les sous-types sont identiques entre les deux échantillons. Le séquençage de l'échantillon D24, identifié en 2003 comme positif à *C. hominis*, a permis de confirmer cette espèce et d'en préciser le sous-type : *C. hominis* IbA10G2. Enfin, le séquençage de l'échantillon 23 n'a pas permis d'identifier *Cryptosporidium* ce qui est en accord avec les résultats négatifs obtenus par la PCR nichée en point final ciblant la protéine gp60.

DISCUSSION

L'absence de thérapie pleinement efficace contre *Cryptosporidium* sp. donne à la connaissance épidémiologique de ce parasite une place prépondérante dans son contrôle. Les techniques de sous-typage s'inscrivent dans cet objectif. Ces techniques se basent principalement sur du séquençage Sanger, qui peut être mis en défaut pour la détection de mélanges de différentes espèces et/ou sous-types de *Cryptosporidium* sp. comme cela a pu être décrit à plusieurs reprises dans la littérature scientifique. Dans le but de dépasser ces limites, le séquençage haut débit a trouvé une place croissante dans l'étude du parasite et son sous-typage. En nous basant sur des publications ayant étudié le séquençage haut débit d'échantillons d'eaux usées, de selles d'animaux ou encore de selles humaines positifs à *Cryptosporidium* sp. (15,37,83,84,87), nous nous sommes intéressés à cette technique appliquée à l'étude de ce parasite, afin d'en confirmer les capacités et d'en préciser les limites.

I) Première partie : Définition du meilleur protocole

Le séquençage de *Cryptosporidium* par NGS est apparu récemment dans la littérature scientifique. Les protocoles entre les différentes publications souffrent pour l'heure d'un manque de standardisation. Différents protocoles de préparation de librairie ont été utilisés, différents gènes ont été ciblés pour le sous-typage, différents automates de NGS ont été employés (15,37,83,84,87). Nos premiers essais avaient donc pour objectif de trouver un protocole fiable, adapté à la problématique qu'est le sous-typage de *Cryptosporidium* sp. Pour cela, nous avons eu recours à trois échantillons positifs à une espèce de *Cryptosporidium* sp, préalablement sous-typés par séquençage selon la technique Sanger, ciblant le gène codant la gp60. Il est apparu à l'issue de la comparaison des deux protocoles une différence importante quant aux résultats obtenus sur nos échantillons.

Le meilleur accord avec les résultats obtenus par la technique Sanger est apparu avec le protocole 16S Metagenomics Sequencing Library Preparation (Illumina). En effet, les séquences correspondant au sous-type trouvé par Sanger représentaient en moyenne 99.3% des séquences totales obtenues par NGS pour nos trois échantillons. Cette moyenne était seulement de 75% avec le protocole Nextera XT (Illumina). De plus, il a été trouvé 150 sous-types additionnels erronés avec le protocole Nextera XT (Illumina) sur l'ensemble de nos échantillons, alors qu'il n'en a été trouvé qu'un avec le protocole 16S Metagenomics

Sequencing Library Preparation (Illumina). Ceci montrait donc une plus grande concordance des résultats du séquençage par la technique Sanger et du séquençage NGS utilisant le protocole 16S Metagenomics Sequencing Library Preparation (Illumina), et une moindre variabilité des résultats. Cette variabilité du protocole Nextera XT (Illumina) peut s'expliquer par le processus de tagmentation, qui consiste en une fragmentation aléatoire des fragments d'ADN. La taille minimale des fragments étudiés doit être de 300pb, nous sommes ici autour de 350pb, mais cela semble insuffisant. Statistiquement, la fragmentation aléatoire a certainement lieu dans la région microsatellite indispensable au sous-typage de *Cryptosporidium* sp., expliquant probablement la grande variabilité de nos résultats avec ce protocole. Cette étape était donc primordiale afin de choisir un protocole de préparation de librairie qui assure des résultats justes, avec un faible bruit de fond. A l'aide du protocole 16S Metagenomics Sequencing Library Preparation (Illumina), le séquençage haut débit de l'échantillon 1770, préalablement identifié comme *C. parvum* IIAA18G1R1 par séquençage selon la technique Sanger, a néanmoins identifié un sous-type supplémentaire : *C. parvum* IIAA11G2R1. Devant l'absence de bruit de fond concernant les autres échantillons (1761 et 1772), la question peut se poser d'un éventuel sous-type réellement présent dans l'échantillon et ignoré par sous-typage Sanger.

II) Deuxième partie : Identification de variants minoritaires au sein de mélanges d'ADN de *Cryptosporidium* par NGS

Dans un second temps, le protocole 16S Metagenomics Sequencing Library Preparation (Illumina) a été mis en application afin d'étudier huit mélanges d'ADN de *C. parvum* et *C. hominis* en proportions variables, dont nous connaissons la composition, ainsi que trois échantillons constitués d'espèces rares (*C. meleagridis*, *C. cuniculus* et *C. erinacei*). Cet essai avait pour principal objectif de confirmer que le séquençage haut débit peut mettre en évidence efficacement des variants minoritaires de *Cryptosporidium* sp. au sein de mélanges d'espèces, voire de sous-types de même espèce, comme cela a été rapporté dans la littérature (15,37,84). Cela a été le cas, vu que 75% des mélanges ont été mis en évidence avec succès par séquençage. Le séquençage NGS de 2 mélanges d'ADN sur les 8 n'a pas permis de mettre en évidence un quelconque variant minoritaire de *Cryptosporidium* sp., et ce quand ce variant était présent à 0.1% et 1%. La valeur de détection la plus basse d'un variant minoritaire a été de 2%. Un sous-type supplémentaire a été identifié par NGS dans 5 mélanges sur 8, il s'agit de *C.*

parvum IIAA11G2R1. Sa présence semble liée à l'échantillon 1770. En effet, les mélanges dans lesquels le séquençage NGS n'a pas permis de mettre en évidence l'ADN de l'échantillon 1770, le sous-type *C. parvum* IIAA11G2R1 était lui aussi absent. De plus, le sous-type *C. parvum* IIAA11G2R1 était le même que celui mis en évidence dans l'échantillon 1770 au cours de notre premier essai a de nouveau été retrouvé. Deux hypothèses peuvent ainsi être soulevées à partir de cette constatation :

- Soit le sous-type *C. parvum* IIAA11G2R1 est réellement présent dans l'échantillon 1770 et a été ignoré jusque-là par la technique Sanger, mais a été mis en évidence par NGS. Cela corroborerait des observations de précédentes études dans lesquels le NGS retrouve le sous-type majoritairement identifié par Sanger, mais également des sous-types additionnels en plus faible proportion (15,)

- Soit le sous-type *C. parvum* IIAA11G2R1 représente un artefact et est le reflet du bruit de fond de la technique. Dans tous nos échantillons, il a été identifié à hauteur de 1.7% en moyenne. Cela signifierait donc qu'un sous-type identifié à hauteur de 1.5% dans un échantillon par séquençage haut débit devrait être considéré comme du bruit de fond.

Un clonage de l'échantillon 1770 suivi du séquençage des clones obtenus permettrait de faire la lumière quant à la composition réelle de cet échantillon.

Les trois espèces rares ont été identifiées avec succès par séquençage NGS, ainsi que leur sous-type, identique à celui retrouvé par le séquençage selon la technique Sanger. Elles ont néanmoins nécessité de faire plus de recherches complémentaires sur le logiciel BLAST que pour *C. parvum* et *C. hominis*. Cela s'explique certainement par le fait que ce sont des espèces plus rarement décrites en pathologie humaine, et donc qu'il existe un plus faible nombre de séquences déposées sur les plateformes telles que GenBank, qui nous a servi à construire notre base de données. Cela a permis de révéler les insuffisances de notre base de données, mettant en lumière la nécessité d'incrémenter régulièrement cette dernière, et de systématiquement comparer nos séquences obtenues par NGS au logiciel BLAST avant de conclure lorsque celles-ci n'ont pas pu être appariées aux séquences de notre base de données.

Ce second essai comporte néanmoins des limites. Le seuil de détection de 2% d'un variant minoritaire par NGS peut être mis en cause. En effet, il est possible que les dilutions aient été trop fortes lorsque l'échantillon 1770 a été dilué au 1/100^e et au 1/1000^e afin qu'il soit présent à 1% et 0.1% dans le mélange. En effet, ces dilutions fortes appliquées à

l'échantillon 1770, dont le C_T pur était de 28 cycles, ont peut-être empêché son amplification par PCR en point final ciblant le gène codant la protéine gp60, pouvant expliquer qu'aucun ADN n'a été amplifié et donc détecté par NGS. La limite de détection ne serait alors pas due directement au NGS mais à la limite de la PCR initiale. Ce seuil de 2% nécessite donc d'être confirmé, et ce en utilisant des échantillons fortement positifs et en s'assurant que les dilutions appliquées n'empêchent pas l'amplification de l'ADN dès les premières étapes du NGS, consistant en une PCR en point final ciblant le gène codant la protéine gp60.

III) Troisième partie : Etude de mélanges d'ADN par les techniques déployées en routine au laboratoire

L'objectif de ce troisième essai était d'étudier différents échantillons positifs à *Cryptosporidium* sp. de la même manière qu'ils le seraient en routine au laboratoire. Les échantillons étaient des mélanges d'ADN de *C. parvum* et *C. hominis* en proportion variable et connue, ainsi que des échantillons issus d'une épidémie française identifiée avec un mélange de *C. hominis* et *C. parvum*. Il est ressorti de cet essai que les techniques de routine, à savoir la PCR en temps réel et le séquençage par la technique Sanger, sont capables d'identifier des mélanges d'ADN de *Cryptosporidium* sp. quand le variant minoritaire est présent à hauteur de 10%. Néanmoins, les variants étaient non détectés lorsqu'ils étaient présents à hauteur de 0.1% et 1%. Quant aux échantillons issus d'une épidémie à Divonne les Bains, ils ont été identifiés en accord avec ce qui avait été trouvé en 2003, à savoir des mélanges de deux espèces de *Cryptosporidium* sp. Il apparaît donc que les techniques actuellement déployées en routine sont capables d'identifier des mélanges d'ADN de *Cryptosporidium* sp.

Nous avons prévu de comparer le séquençage selon la technique Sanger et le NGS directement sur les mêmes échantillons, dans les mêmes conditions que celles qui ont été réalisées dans la troisième partie de nos essais. Cela a été réalisé mais le séquençage NGS sur l'automate MiSeq (Illumina) n'a pas fonctionné. Cela nous aurait permis de voir si, réalisés dans les mêmes conditions, le séquençage NGS montrait une supériorité quant au seuil de détection d'un variant minoritaire au sein d'un mélange par rapport au séquençage par la technique Sanger. De plus, cela aurait permis de voir si le séquençage haut débit de l'échantillon D23 aurait donné un résultat, là où le séquençage par la technique Sanger n'a pas pu en donner.

L'échec de cette expérimentation du NGS souligne la difficulté technique de cette approche de séquençage, qui demande du temps et des opérateurs expérimentés.

V) Perspectives

A l'issue de nos essais, le NGS apparaît comme une approche intéressante pour apporter des réponses satisfaisantes dans la caractérisation des co-infections au sein d'échantillons cliniques. Le séquençage massif en parallèle de nombreux échantillons en parallèle rend cette technique particulièrement intéressante dans l'étude d'épidémies. Il conviendrait donc à présent d'étudier à l'aide de cette technique des échantillons d'épidémies sous-typés au laboratoire par la technique Sanger, afin de comparer l'apport en termes de temps, de coût et surtout d'apport informatif épidémiologique. Des échantillons de l'épidémie récente de Grasse pourraient ainsi être analysés, ou encore des échantillons d'épidémies mixtes caractérisées.

A l'issue de ce travail, des pistes de réflexion sont apparues. Un lien entre la proportion d'ADN d'un sous-type au sein d'un mélange et le nombre de séquences appariées à ce sous-type semble apparaître. Néanmoins, n'ayant eu qu'un échantillon dans lequel *C. parvum* était majoritaire, il est indispensable de réaliser des travaux supplémentaires en étudiant de nouveaux mélanges connus d'ADN.

Enfin, il apparaît primordial de réaliser de nouveaux tests afin de pouvoir déterminer avec certitude un seuil d'exclusion fiable, qui permette d'ignorer les sous-types identifiés à tort, sans pour autant écarter faussement des sous-types bel et bien présents dans l'échantillon étudié. Cela est prévu prochainement, dans les mêmes conditions que celles utilisées pour évaluer le séquençage de mélanges d'ADN par technique Sanger.

MÉMOIRE DE D.E.S SOUTENU par Me. BAILLY Eloïse

CONCLUSIONS

Notre travail portait sur l'évaluation du séquençage haut débit (NGS) pour l'étude de *Cryptosporidium* spp. et plus particulièrement pour son sous-typage en ciblant le gène codant la glycoprotéine 60 (gp60). Ce travail s'appuyait sur des publications récentes démontrant la capacité du NGS à identifier des variants minoritaires au sein de mélanges d'ADN, situations où le séquençage par la technique Sanger, aujourd'hui largement déployé pour le sous-typage de *Cryptosporidium* spp., peut être mis en défaut.

Nos essais nous ont, dans un premier temps, permis de déterminer quel protocole de préparation de librairie était le plus adapté à notre problématique, afin d'assurer des résultats fiables pour la suite de nos essais. C'est le protocole 16S Metagenomics Preparation Library (Illumina) qui nous a permis de trouver des résultats moins variables que le protocole Nextera XT (Illumina) sur des échantillons positifs à une seule espèce (et un seul sous-type) de *Cryptosporidium* spp. Ce protocole a donc été utilisé pour la suite de nos essais.

Des mélanges d'ADN réalisés à l'aide d'échantillons positifs à *C. parvum* IIAA18G1R1 et *C. hominis* IBA10G2, ont été constitués de manière à connaître leur composition et principalement la proportion de chaque espèce dans le mélange. Les échantillons positifs à *C. parvum* et *C. hominis* avaient été préalablement identifiés individuellement par séquençage selon la technique Sanger. Comme rapporté dans la littérature, le séquençage haut débit a permis d'identifier les deux espèces de *Cryptosporidium* sp. présentes dans nos mélanges d'ADN, et également de retrouver les mêmes sous-types que ceux identifiés par séquençage selon la technique Sanger. Les deux espèces ont en effet été détectées dans 75%

de nos mélanges d'ADN, et ce jusqu'à une proportion de 2% concernant le variant minoritaire. Ce dernier était ignoré dans les mélanges où il était en proportion de 0.1% et 1%. Un sous-type supplémentaire de *C. parvum* (*C. parvum* IIaA11G2R1) a été identifié lors du séquençage haut débit des mélanges d'ADN. Cela a soulevé deux hypothèses : soit il s'agit d'un sous-type supplémentaire présent dans un de nos échantillons et qui a été ignoré jusqu'ici, soit il s'agit d'un artéfact de la technique NGS. Ces deux pistes nécessitent d'être explorées.

Enfin, nous nous sommes intéressés au séquençage par la technique Sanger de mélanges d'ADN de *C. parvum* IIcA5G3 et *C. hominis* IbA10G2 dont la proportion relative était connue. Ici, le variant minoritaire était détecté jusqu'à une proportion de 10% mais indétectable à 0.1% et 1%. Il semble donc que, comme cela a été décrit dans la littérature scientifique, le NGS soit plus en mesure de mettre en évidence des infections mixtes que le séquençage selon la technique Sanger.

De nouveaux travaux sont prévus prochainement afin de préciser les résultats que nous avons obtenus. En effet, il reste à déterminer plus précisément le seuil de détection des variants minoritaires par séquençage haut débit, et déterminer un seuil permettant de distinguer de façon fiable des artéfacts ou de réels sous-types présents au sein d'un isolat.

Le Directeur de thèse,



Le Président,



**Vu pour l'autorisation de
Soutenance**

**Dijon, le
Le Vice-Doyen,**



Dijon, le
Le Vice-Doyen,
de Santé
Scolaire
C. MARIE

BIBLIOGRAPHIE

- 1- Guyot K, Sarfati C, Derouin F. Actualités sur l'épidémiologie et le diagnostic de la cryptosporidiose. Feuillet de biologie. 2012 Jan;VolLIII-304
- 2- Gargala G. La cryptosporidiose, une cause trop méconnue de diarrhée aiguë. La Lettre de l'infectiologue. 2015 Dec. Tome III;6
- 3- Tzipori S, Widmer G. A hundred-year retrospective on cryptosporidiosis. Trends Parasitol. 2008 Apr;24(4):184-9.
- 4- Chalmers RM, Davies AP, Tyler K. *Cryptosporidium*. Microbiology (Reading). 2019 May;165(5):500-502
- 5- Khan A, Shaik JS, Grigg ME. Genomics and molecular epidemiology of *Cryptosporidium* species. Acta Trop. 2018 Aug;184:1-14
- 6- Chalmers RM, Davies AP. Minireview: clinical cryptosporidiosis. Exp Parasitol. 2010 Jan;124(1):138-46
- 7- Ripert C, Pajot FX, Vincendeau P, Esquerdo Gomez F. Epidemiologie des maladies parasitaires. 1996;Tome 1 : Cachan : EM Inter, 393 p. ISBN 2-7430-0076-3
- 8- Leitch GJ, He Q. Cryptosporidiosis-an overview. J Biomed Res. 2012 Jan;25(1):1-16
- 9- Fayer R. *Cryptosporidium*: a water-borne zoonotic parasite. Vet Parasitol. 2004 Dec 9;126(1-2):37-56
- 10- Baldursson S, Karanis P. Waterborne transmission of protozoan parasites: review of worldwide outbreaks - an update 2004-2010. Water Res. 2011 Dec 15;45(20):6603-14
- 11- Efstratiou A, Ongerth JE, Karanis P. Waterborne transmission of protozoan parasites: Review of worldwide outbreaks - An update 2011-2016. Water Res. 2017 May 1;114:14-22
- 12- Koompapong K, Sukthana Y. Seasonal variation and potential sources of *Cryptosporidium* contamination in surface waters of Chao Phraya River and Bang Pu Nature Reserve pier, Thailand. Southeast Asian J Trop Med Public Health. 2012 Jul;43(4):832-40
- 13- Swaffer B, Abbott H, King B, van der Linden L, Monis P. Understanding human infectious *Cryptosporidium* risk in drinking water supply catchments. Water Res. 2018 Jul 1;138:282-292
- 14- Bouzid M, Hunter PR, Chalmers RM, Tyler KM. *Cryptosporidium* pathogenicity and virulence. Clin Microbiol Rev. 2013 Jan;26(1):115-34
- 15- Zahedi A, Gofton AW, Jian F, Papparini A, Oskam C, Ball A, Robertson I, Ryan U. Next Generation Sequencing uncovers within-host differences in the genetic diversity of *Cryptosporidium* gp60 subtypes. Int J Parasitol. 2017 Sep;47(10-11):601-607

- 16- Widmer G, Sullivan S. Genomics and population biology of *Cryptosporidium* species. *Parasite Immunol.* 2012 Feb-Mar;34(2-3):61-71
- 17- Beser J, Hallström BM, Advani A, Andersson S, Östlund G, Winiecka-Krusnell J, Lebbad M, Alm E, Troell K, Arrighi RBG. Improving the genotyping resolution of *Cryptosporidium hominis* subtype IbA10G2 using one step PCR-based amplicon sequencing. *Infect Genet Evol.* 2017 Nov;55:297-304
- 18- Xiao L. Molecular epidemiology of cryptosporidiosis: an update. *Exp Parasitol.* 2010 Jan;124(1):80-9
- 19- Li N, Xiao L, Cama VA, et al. Genetic recombination and *Cryptosporidium hominis* virulent subtype IbA10G2. *Emerg Infect Dis.* 2013;19(10):1573-1582
- 20- Grinberg A, Widmer G. *Cryptosporidium* within-host genetic diversity: systematic bibliographical search and narrative overview. *Int J Parasitol.* 2016 Jul;46(8):465-71
- 21- Tyzzer EE. A sporozoan found in the peptic glands of the common mouse. *Proceedings of the Society for Experimental Biology and Medicine.* 1907;5(1):12-13
- 22- Tyzzer EE. *Cryptosporidium parvum* (sp. nov.), a coccidium found in the small intestine of the common mouse. *Arch Protistenkd.* 1912;26:394-412
- 23- Slavin D. *Cryptosporidium meleagridis* (sp. nov.) *Journal of Comparative Pathology and Therapeutics.* 1965;65:262-266
- 24- Casemore DP. Epidemiological aspects of human cryptosporidiosis. *Epidemiol Infect.* 1990 Feb;104(1):1-28
- 25- Tzipori S. Cryptosporidiosis in animals and humans. *Microbiol Rev.* 1983 Mar;47(1):84-96
- 26- Panciera J. Cryptosporidial infection in a calf. 1971 *Vet. Path.* 1971;8: 479-484
- 27- Hunter PR, Nichols G. Epidemiology and clinical features of *Cryptosporidium* infection in immunocompromised patients. *Clin Microbiol Rev.* 2002 Jan;15(1):145-54
- 28- Nime FA, Burek JD, Page DL, Holscher MA, Yardley JH. Acute enterocolitis in a human being infected with the protozoan *Cryptosporidium*. *Gastroenterology.* 1976 Apr;70(4):592-8
- 29- Zhu G, Keithly JS, Philippe H. What is the phylogenetic position of *Cryptosporidium*? *Int J Syst Evol Microbiol.* 2000 Jul;50 Pt 4:1673-1681
- 30- Cacciò SM, Chalmers RM. Human cryptosporidiosis in Europe. *Clin Microbiol Infect.* 2016 Jun;22(6):471-80
- 31- Ryan U, Hijjawi N. New developments in *Cryptosporidium* research. *Int J Parasitol.* 2015 May;45(6):367-73

- 32- Clode PL, Koh WH, Thompson RCA. Life without a Host Cell: What is *Cryptosporidium*? Trends Parasitol. 2015 Dec;31(12):614-624
- 33- Koh W, Thompson A, Edwards H, Monis P, Clode PL. Extracellular excystation and development of *Cryptosporidium*: tracing the fate of oocysts within *Pseudomonas* aquatic biofilm systems. BMC Microbiol. 2014 Nov 18;14:281
- 34- Xiao L, Fayer R, Ryan U, Upton SJ. *Cryptosporidium* taxonomy: recent advances and implications for public health. Clin Microbiol Rev. 2004 Jan;17(1):72-97
- 35- Tzipori S, Angus KW, Campbell I, Gray EW. *Cryptosporidium*: evidence for a single-species genus. Infect Immun. 1980 Dec;30(3):884-6.
- 36- Li N, Xiao L, Alderisio K, Elwin K, Cebelinski E, Chalmers R, Santin M, Fayer R, Kvac M, Ryan U, Sak B, Stanko M, Guo Y, Wang L, Zhang L, Cai J, Roellig D, Feng Y. Subtyping *Cryptosporidium ubiquitum*, a zoonotic pathogen emerging in humans. Emerg Infect Dis. 2014 Feb;20(2):217-24
- 37- Paparini A, Gofton A, Yang R, White N, Bunce M, Ryan UM. Comparison of Sanger and next generation sequencing performance for genotyping *Cryptosporidium* isolates at the 18S rRNA and actin loci. Exp Parasitol. 2015 Apr-May;151-152:21-7
- 38- Chalmers RM, Katzer F. Looking for *Cryptosporidium*: the application of advances in detection and diagnosis. Trends Parasitol. 2013 May;29(5):237-51
- 39- O'Hara SP, Chen XM. The cell biology of *Cryptosporidium* infection. Microbes Infect. 2011 Aug;13(8-9):721-30
- 40- Yan W, Alderisio K, Roellig DM, Elwin K, Chalmers RM, Yang F, Wang Y, Feng Y, Xiao L. Subtype analysis of zoonotic pathogen *Cryptosporidium* skunk genotype. Infect Genet Evol. 2017 Nov;55:20-25
- 41- Jex AR, Gasser RB. Genetic richness and diversity in *Cryptosporidium hominis* and *C. parvum* reveals major knowledge gaps and a need for the application of "next generation" technologies-research review. Biotechnol Adv. 2010 Jan-Feb;28(1):17-26
- 42- <https://www.who.int/publications/i/item/9789241549950>
- 43- Bird RG, Smith MD. Cryptosporidiosis in man: parasite life cycle and fine structural pathology. J Pathol. 1980 Nov;132(3):217-33
- 44- Buret AG, Cacciò SM, Favennec L, Svärd S. Update on Giardia: Highlights from the seventh International *Giardia* and *Cryptosporidium* Conference. Parasite. 2020;27:49
- 45- Bhalchandra S, Cardenas D, Ward HD. Recent Breakthroughs and Ongoing Limitations in *Cryptosporidium* Research. F1000Res. 2018 Sep 3;7:F1000 Faculty Rev-1380
- 46- Jagai JS, Castronovo DA, Monchak J, Naumova EN. Seasonality of cryptosporidiosis: A meta-analysis approach. Environ Res. 2009 May;109(4):465-78

- 47- Savioli L, Smith H, Thompson A. *Giardia* and *Cryptosporidium* join the 'Neglected Diseases Initiative'. Trends Parasitol. 2006 May;22(5):203-8
- 48- <https://www.paca.ars.sante.fr/cryptosporidiose-dans-les-alpes-maritimes-point-de-situation-2>
- 49- <https://www.anses.fr/fr/content/les-infections-%C3%A0-protozoaires-li%C3%A9es-aux-aliments-et-%C3%A0-l%E2%80%99eau%C2%A0-%C3%A9valuation-scientifique-des-0>
- 50- Hunter PR, Hughes S, Woodhouse S, Raj N, Syed Q, Chalmers RM, Verlander NQ, Goodacre J. Health sequelae of human cryptosporidiosis in immunocompetent patients. Clin Infect Dis. 2004 Aug 15;39(4):504-10
- 51- Ozgöl A, Tanyüksel M, Yazicioglu K, Arpacioğlu O. Sacroiliitis associated with *Cryptosporidium parvum* in an HLA-B27-negative patient. Rheumatology (Oxford). 1999 Mar;38(3):288-9
- 52- Barbot L, Topouchian A, Capet C, Magne D, Huneau J.F, Kapel N. *Cryptosporidium parvum* : étude fonctionnelle du syndrome de malabsorption intestinale. 2008
- 53- Borad A, Ward H. Human immune responses in cryptosporidiosis. Future Microbiol. 2010 Mar;5(3):507-19
- 54- McDonald V. Cryptosporidiosis: host immune responses and the prospects for effective immunotherapies. Expert Rev Anti Infect Ther. 2011 Nov;9(11):1077-86
- 55- Kutukculer N, Moratto D, Aydinok Y, Lougaris V, Aksoylar S, Plebani A, Genel F, Notarangelo LD. Disseminated *Cryptosporidium* infection in an infant with hyper-IgM syndrome caused by CD40 deficiency. J Pediatr. 2003 Feb;142(2):194-6
- 56- Khurana S, Sharma P, Sharma A, Malla N. Evaluation of Ziehl-Neelsen staining, auramine phenol staining, antigen detection enzyme linked immunosorbent assay and polymerase chain reaction, for the diagnosis of intestinal cryptosporidiosis. Trop Parasitol. 2012 Jan;2(1):20-3
- 57- Hanscheid T, Cristino JM, Salgado MJ. Screening of auramine-stained smears of all fecal samples is a rapid and inexpensive way to increase the detection of coccidial infections. Int J Infect Dis. 2008 Jan;12(1):47-50
- 58- Khanna V, Tilak K, Ghosh A, Mukhopadhyay C. Modified Negative Staining of Heine for Fast and Inexpensive Screening of *Cryptosporidium*, *Cyclospora*, and *Cystoisospora* spp. Int Sch Res Notices. 2014 Oct 20;2014:165424
- 59- Agnamey P, Sarfati C, Pinel C, Rabodoniriina M, Kapel N, Dutoit E, Garnaud C, Diouf M, Garin JF, Totet A, F; ANOFEL *Cryptosporidium* National Network. Evaluation of four commercial rapid immunochromatographic assays for detection of *Cryptosporidium* antigens in stool samples: a blind multicenter trial. J Clin Microbiol. 2011 Apr;49(4):1605-7

- 60- Garcia LS, Shimizu RY. Evaluation of nine immunoassay kits (enzyme immunoassay and direct fluorescence) for detection of *Giardia lamblia* and *Cryptosporidium parvum* in human fecal specimens. J Clin Microbiol. 1997 Jun;35(6):1526-9
- 61- Rider SD Jr, Zhu G. *Cryptosporidium*: genomic and biochemical features. Exp Parasitol. 2010 Jan;124(1):2-9
- 62- Amadi B, Mwiya M, Sianongo S, Payne L, Watuka A, Katubulushi M, Kelly P. High dose prolonged treatment with nitazoxanide is not effective for cryptosporidiosis in HIV positive Zambian children: a randomised controlled trial. BMC Infect Dis. 2009 Dec 2;9:195
- 63- Marcos LA, Gotuzzo E. Intestinal protozoan infections in the immunocompromised host. Curr Opin Infect Dis. 2013 Aug;26(4):295-301
- 64- Brady MT, Oleske JM, Williams PL, Elgie C, Mofenson LM, Dankner WM, Van Dyke RB; Pediatric AIDS Clinical Trials Group 219/219C Team. Declines in mortality rates and changes in causes of death in HIV-1-infected children during the HAART era. J Acquir Immune Defic Syndr. 2010 Jan;53(1):86-94
- 65- Zhang Y, Feng J, McManus SA, Lu HD, Ristroph KD, Cho EJ, Dobrijevic EL, Chan HK, Prud'homme RK. Design and Solidification of Fast-Releasing Clofazimine Nanoparticles for Treatment of Cryptosporidiosis. Mol Pharm. 2017 Oct 2;14(10):3480-3488
- 66- Jumani RS, Bessoff K, Love MS, Miller P, Stebbins EE, Teixeira JE, Campbell MA, Meyers MJ, Zambriski JA, Nunez V, Woods AK, McNamara CW, Huston CD. A Novel Piperazine-Based Drug Lead for Cryptosporidiosis from the Medicines for Malaria Venture Open-Access Malaria Box. Antimicrob Agents Chemother. 2018 Mar 27;62(4):e01505-17
- 67- Gargala G. Drug treatment and novel drug target against *Cryptosporidium*. Parasite. 2008 Sep;15(3):275-81
- 68- Peletz R, Mahin T, Elliott M, Montgomery M, Clasen T. Preventing cryptosporidiosis: the need for safe drinking water. Bull World Health Organ. 2013 Apr 1;91(4):238-8A
- 69- Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM, Serrano MG, Puiu D, Manque P, Akiyoshi D, Mackey AJ, Pearson WR, Dear PH, Bankier AT, Peterson DL, Abrahamsen MS, Kapur V, Tzipori S, Buck GA. The genome of *Cryptosporidium hominis*. Nature. 2004 Oct 28;431(7012):1107-12
- 70- Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, Deng M, Liu C, Widmer G, Tzipori S, Buck GA, Xu P, Bankier AT, Dear PH, Konfortov BA, Spriggs HF, Iyer L, Anantharaman V, Aravind L, Kapur V. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. Science. 2004 Apr 16;304(5669):441-5
- 71- Nader JL, Mathers TC, Ward BJ, Pachebat JA, Swain MT, Robinson G, Chalmers RM, Hunter PR, van Oosterhout C, Tyler KM. Evolutionary genomics of anthroponosis in *Cryptosporidium*. Nat Microbiol. 2019 May;4(5):826-836
- 72- Guo Y, Tang K, Rowe LA, Li N, Roellig DM, Knipe K, Frace M, Yang C, Feng Y, Xiao L. Comparative genomic analysis reveals occurrence of genetic recombination in virulent

Cryptosporidium hominis subtypes and telomeric gene duplications in *Cryptosporidium parvum*. BMC Genomics. 2015 Apr 18;16(1):320

73- Puiu D, Enomoto S, Buck GA, Abrahamsen MS, Kissinger JC. CryptoDB: the *Cryptosporidium* genome resource. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D329-31

74- Robinson G, Chalmers RM. Assessment of polymorphic genetic markers for multi-locus typing of *Cryptosporidium parvum* and *Cryptosporidium hominis*. Exp Parasitol. 2012 Oct;132(2):200-15

75- Chalmers RM, Pérez-Cordón G, Cacció SM, Klotz C, Robertson LJ; participants of the *Cryptosporidium* genotyping workshop (EURO-FBP). *Cryptosporidium* genotyping in Europe: The current status and processes for a harmonised multi-locus genotyping scheme. Exp Parasitol. 2018 Aug;191:25-30

76- Waldron LS, Power ML. Fluorescence analysis detects gp60 subtype diversity in *Cryptosporidium* infections. Infect Genet Evol. 2011 Aug;11(6):1388-95

77- Sulaiman IM, Hira PR, Zhou L, Al-Ali FM, Al-Shelahi FA, Shweiki HM, Iqbal J, Khalid N, Xiao L. Unique endemicity of cryptosporidiosis in children in Kuwait. J Clin Microbiol. 2005 Jun;43(6):2805-9

78- Xiao L, Feng Y. Molecular epidemiologic tools for waterborne pathogens *Cryptosporidium* spp. and *Giardia duodenalis*. Food Waterborne Parasitol. 2017 Sep 29;8-9:14-32

79- Muzzey D, Evans EA, Lieber C. Understanding the Basics of NGS: From Mechanism to Variant Calling. Curr Genet Med Rep. 2015;3(4):158-165

80- Diene SM, Bertelli C, Pillonel T, Schrenzel J, Greub G. Génomique et métagénomique bactériennes: applications cliniques et importance médicale [Bacterial genomics and metagenomics: clinical applications and medical relevance]. Rev Med Suisse. 2014 Nov 12;10(450):2155-61

81- Dortet L, Bonnin R, Naas T. Impact du séquençage d'ADN haut débit sur la surveillance des épidémies de bactérie multi-résistantes aux antibiotiques. Feuillet de Biologie. 2017 Jan 334

82- Zahedi A, Greay TL, Papparini A, Linge KL, Joll CA, Ryan UM. Identification of eukaryotic microorganisms with 18S rRNA next-generation sequencing in wastewater treatment plants, with a more targeted NGS approach required for *Cryptosporidium* detection. Water Res. 2019 Jul 1;158:301-312

83- Kaupke A, Gawor J, Rzeżutka A, Gromadka R. Identification of pig-specific *Cryptosporidium* species in mixed infections using Illumina sequencing technology. Exp Parasitol. 2017 Nov;182:22-25

84- Moreno Y, Moreno-Mesonero L, Amorós I, Pérez R, Morillo JA, Alonso JL. Multiple identification of most important waterborne protozoa in surface water used for irrigation purposes by 18S rRNA amplicon-based metagenomics. Int J Hyg Environ Health. 2018 Jan;221(1):102-111

- 85- <https://cnrcryptosporidioses.chu-rouen.fr/wp-content/uploads/sites/54/2019/10/rapport-2018-CNR-LE-Cryptosporidioses-def.pdf>
- 86- Cama VA, Bern C, Roberts J, Cabrera L, Sterling CR, Ortega Y, Gilman RH, Xiao L. *Cryptosporidium* species and subtypes and clinical manifestations in children, Peru. Emerg Infect Dis. 2008 Oct;14(10):1567-74
- 87- Zahedi A, Gofton AW, Greay T, Monis P, Oskam C, Ball A, Bath A, Watkinson A, Robertson I, Ryan U. Profiling the diversity of *Cryptosporidium* species and genotypes in wastewater treatment plants in Australia using next generation sequencing. Sci Total Environ. 2018 Dec 10;644:635-648
- 88- Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. J Appl Genet. 2011 Nov;52(4):413-35
- 89- Kotloff KL, Nataro JP. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet. 2013 Jul 20;382(9888):209-22
- 90- Rossignol JF, Ayoub A, Ayers MS. Treatment of diarrhea caused by *Cryptosporidium parvum*: a prospective randomized, double-blind, placebo-controlled study of Nitazoxanide. J Infect Dis. 2001 Jul 1;184(1):103-6
- 91- Rossignol JF, Cavier R. New derivatives of benzamido 5-nitrothiazoles. Chemical abstract. 1975;83:28216
- 92- Hoffman PS, Sisson G, Croxen MA, Welch K, Harman WD, Cremades N, Morash MG. Antiparasitic drug nitazoxanide inhibits the pyruvate oxidoreductases of *Helicobacter pylori*, selected anaerobic bacteria and parasites, and *Campylobacter jejuni*. Antimicrob Agents Chemother. 2007 Mar;51(3):868-76
- 93- Okhuysen PC, Chappell CL, Crabb JH, Sterling CR, DuPont HL. Virulence of three distinct *Cryptosporidium parvum* isolates for healthy adults. J Infect Dis. 1999 Oct;180(4):1275-81
- 94- Xiao L, Ryan UM. Cryptosporidiosis: an update in molecular epidemiology. Curr Opin Infect Dis. 2004 Oct;17(5):483-90
- 95- Stibbs HH, Ongerth JE. Immunofluorescence detection of *Cryptosporidium* oocysts in fecal smears. J Clin Microbiol. 1986 Oct;24(4):517-21
- 96- Grinberg A, Biggs PJ, Dukkipati VS, George TT. Extensive intra-host genetic diversity uncovered in *Cryptosporidium parvum* using Next Generation Sequencing. Infect Genet Evol. 2013 Apr;15:18-24
- 97- Behjati S, Tarpey PS. What is next generation sequencing? Arch Dis Child Educ Pract Ed. 2013 Dec;98(6):236-8
- 98- Loury P, Gross L, Dugast F, Favennec L, Dalle F, de Rougemont A, *et al.* Épidémie de cryptosporidiose dans un collège de l'ouest de la France, novembre 2017. Bull Epidemiol Hebd. 2019;(16):295-300

- 99- Gofti Laroche L, Schmitt M. Epidémie de gastro-entérites liée à la pollution du réseau de distribution d'eau potable de la commune de Divonne-les-Bains, Ain (01). Août-septembre 2003. Saint-Maurice : Institut de veille sanitaire; 2003. 49 p.
- 100- https://www.cdc.gov/parasites/crypto/gen_info/prevention-general-public.html
- 101- <https://presse.inserm.fr/epidemie-de-cryptosporidiose-linserm-etudie-des-pistes-therapeutiques/37456/>
- 102- Sparks H, Nair G, Castellanos-Gonzalez A, White AC Jr. Treatment of Cryptosporidium: What We Know, Gaps, and the Way Forward. *Curr Trop Med Rep*. 2015 Sep;2(3):181-187
- 103- https://www.infectiologie.org.tn/pdf_ppt_docs/cmi/05032010/anti_ccocid.pdf
- 104- http://genetique.snv.jussieu.fr/OLD%20SITE/Documents%202011/Epigenetique%202011/LeCRom_110310_SequencageHautDebit.pdf
- 105- <http://biochimej.univ-angers.fr/Page2/COURS/9ModulGenFoncVeg/5MethEtudGenFonc/1MethodeSEQUENGAGE/1SEQUENGAGE.htm>
- 106- https://emea.support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_nextera/nextera-xt/nextera-xt-library-prep-reference-guide-15031942-05.pdf
- 107- https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf
- 108- Yuan Lu. Next Generation Sequencing in Aquatic Models. *InTech Open*. 2016 Jan;(2):62-79
- 109- http://www.canceropole-gso.org/download/fichiers/3606/NGS+Canceropole+GSO+mai+2016+AGros_DIFF.pdf

ANNEXES

Annexe 1 : Protocole du séquençage par la technique Sanger.

La technique de sous-typage par séquençage Sanger repose sur l'amplification en point final d'un fragment du gène de la gp60, puis son séquençage. Notons que cette technique est réalisée au laboratoire de parasitologie du CHU de Dijon dans l'optique de son activité de laboratoire collaborateur du CNR-LE *Cryptosporidium*. L'amplification repose sur une PCR simple suivie d'une éventuelle PCR nichée si le résultat de la PCR simple s'avère négatif. Un résultat positif est caractérisé par la présence d'une bande d'amplification comprise entre 300 et 400pb sur gel à 2% d'agarose (taille variable en fonction de l'espèce et du nombre de répétitions au sein du microsatellite). Notons que les trois échantillons choisis pour notre étude (1761, 1770 et 1772) n'ont nécessité qu'une PCR simple afin d'assurer une amplification suffisante.

Les amorces utilisées sont identiques pour la PCR simple et la PCR nichée. Leurs séquences sont pour l'amorce sens AL3532 : 5' – TCCGCTGTATTCTCAGCC – 3' et pour l'amorce anti-sens LX0029 : 5' – CGAACCACATTACAAATGAAGT – 3' (TIB Molbiol). Quant à la Taq polymérase, il s'agit d'une HotStarTaq® (Qiagen). L'amplification est réalisée sur un thermocycleur ABI Applied Biosystems™ 9700 (ThermoFisher).

Le programme de PCR sur le thermocycleur s'organise comme suit :

- 95°C pendant 15 minutes
- 35 cycles de
 - o 94°C pendant 45 secondes
 - o 50°C pendant 45 secondes
 - o 72°C pendant 59 secondes
- 72°C pendant 10 minutes

L'enzyme étant une HotStarTaq®, elle n'a pas d'activité polymérase à température ambiante. Cela implique une activation nécessaire par la température (95°C pendant 15 minutes), et permet de limiter les amplifications non spécifiques.

Eau	9 μ L
HotStarTaq® Master Mix kit	12,5 μ L
Amorce sens AL3532	0,5 μ L
Amorce anti-sens LX0029	0,5 μ L
ADN	2,5 μ L
Volume réactionnel total	25μL

Les produits de PCR sont ensuite purifiés sur colonne Amicon® puis quantifiés par technique spectrophotométrique sur NanoDrop®. La réaction de séquençage qui suit est réalisée sur un thermocycleur ABI Applied Biosystems™ 9700 (ThermoFisher), selon le programme suivant :

- 96°C pendant 1 minute
- 25 cycles de
 - o 96°C pendant 10 secondes
 - o 50°C pendant 5 secondes
 - o 60°C pendant 4 minutes
- 4°C à l'infini

Eau	5 μ L
Tampon de séquençage 10X	2 μ L
Big Dye	4 μ L
Amorce sens ou anti-sens (AL3233 ou LX0029)	0,5 μ L
ADN quantifié et ajusté à 6 ng/ μ L	2,5 μ L
Volume réactionnel total	15μL

Deux mix de PCR de séquençage sont réalisés en parallèle : un premier pour une lecture sens (reverse), et un second mix pour une lecture anti-sens (forward). L'étape finale, à savoir la migration des produits de séquençage a lieu sur l'analyseur Applied Biosystems™ 3130xl (ThermoFisher). Les séquences obtenues sont analysées à l'aide de l'outil BLAST qui permet une comparaison à une base de données de séquences de référence.

Annexe 2 : Base de données des séquences de *Cryptosporidium* sp. issue de l'article de Zahedi et al. (15)

No.	<i>Cryptosporidium</i> species	Subtype	GenBank Accession no.	No.	<i>Cryptosporidium</i> species	Subtype	GenBank Accession no.
1	<i>C. fayeri</i>	IVaA11G3T1	FJ490060	67	<i>C. hominis</i>	IdA21	FJ707316
2	<i>C. fayeri</i>	IVaA9G4T1R1	HQ008934	68	<i>C. hominis</i>	IiA17	HM234173
3	<i>C. fayeri</i>	IVbA9G1T1	FJ490087	69	<i>C. hominis</i>	IaA13R8	JX088407
4	<i>C. fayeri</i>	IVcA8G1T1	FJ490069	70	<i>C. hominis</i>	IfA16G2	JX000570
5	<i>C. fayeri</i>	IVdA7G1T1	FJ490058	71	<i>C. hominis</i>	IbA9G3	DQ665688
6	<i>C. fayeri</i>	IVeA7G1T1	FJ490071	72	<i>C. hominis</i>	IbA20G2	FJ153240
7	<i>C. fayeri</i>	IVfA12G1T1	FJ490076	73	<i>C. hominis</i>	IfA12G1	EU161655
8	<i>C. parvum</i>	IiAa15G2R1	AY262034	74	<i>C. hominis</i>	IdA20	EU877245
9	<i>C. parvum</i>	IidA15G1	EF576977	75	<i>C. hominis</i>	IgA24	EF208067
10	<i>C. parvum</i>	IicA5G3j	GQ259136	76	<i>C. hominis</i>	IbA10G2R2	EU877240
11	<i>C. parvum</i>	IicA5G3b	AF164501	77	<i>C. hominis</i>	IdA16	DQ665692
12	<i>C. parvum</i>	IicA5G3c	AF164502	78	<i>C. hominis</i>	IdA15	DQ280498
13	<i>C. parvum</i>	IIfA6	AY738188	79	<i>C. hominis</i>	IfA12G2	JF681172
14	<i>C. parvum</i>	IiAa19G4R1	JF727803	80	<i>C. hominis</i>	IjA14	JF681174
15	<i>C. parvum</i>	IidA18G1	AY738194	81	<i>C. hominis</i>	IdA24	EU146132
16	<i>C. parvum</i>	IidA24G1	JF727809	82	<i>C. hominis</i>	IaA13R7	EU095261
17	<i>C. parvum</i>	IicA5G3e	HM234172	83	<i>C. hominis</i>	IaA11R4	EU095258
18	<i>C. parvum</i>	IicA5G3a	AF164491	84	<i>C. hominis</i>	IaA12R4	EU095259
19	<i>C. parvum</i>	IieA12G1	AY382675	85	<i>C. hominis</i>	IaA13R4	EU095260
20	<i>C. parvum</i>	IijA15G4	DQ648547	86	<i>C. hominis</i>	IaA14R6	EU095262
21	<i>C. parvum</i>	IInA8	FJ897787	87	<i>C. hominis</i>	IaA15R3	EU095263
22	<i>C. parvum</i>	IImA7G1	AY700401	88	<i>C. hominis</i>	IdA10	EU095264
23	<i>C. parvum</i>	IiAa21G3R1	JF727806	89	<i>C. hominis</i>	IkA7G4	KP314263
24	<i>C. parvum</i>	IicA5G3f	HM234171	90	<i>C. hominis</i>	IhA14G1	FJ971716
25	<i>C. parvum</i>	IihA7G4	AY873781	91	<i>C. hominis</i>	IaA14R7	JX000569
26	<i>C. parvum</i>	IigA9	AY873780	92	<i>C. erinacei</i>	XIIIaA20R10	KF055453
27	<i>C. parvum</i>	IiA10	AY873782	93	<i>C. wrairi</i>	VIIaA13T1	HM234185
28	<i>C. parvum</i>	IiAa14G2R1	JF727773	94	<i>C. wrairi</i>	VIIaA17T1	GQ121028
29	<i>C. parvum</i>	IicA5G3k	JF802123	95	<i>C. wrairi</i>	VIIaA16T1	HM234184
30	<i>C. parvum</i>	IiAa16G3R1	AB560744	96	<i>C. tyzzeri</i>	IXbA6	HM234176
31	<i>C. parvum</i>	IidA20G1a	AB560745	97	<i>C. tyzzeri</i>	IXaA6R3	GQ121030
32	<i>C. parvum</i>	IidA21G1a	AB560746	98	<i>C. tyzzeri</i>	IXaA6R2	HM234180
33	<i>C. parvum</i>	IidA14G1	KP793006	99	<i>C. cuniculus</i>	VbA23	KX375350
34	<i>C. parvum</i>	IiIA18R2	AM937006	100	<i>C. cuniculus</i>	VbA34	KU129016
35	<i>C. parvum</i>	IibA14	AF402285	101	<i>C. cuniculus</i>	VbA26	GU971638
36	<i>C. parvum</i>	IidA20G1	KU248815	102	<i>C. cuniculus</i>	VbA30	GU971632
37	<i>C. parvum</i>	IiAa18G3R1	DQ192508	103	<i>C. cuniculus</i>	VbA33	GU971635
38	<i>C. parvum</i>	IiAa15G2R2	DQ192501	104	<i>C. cuniculus</i>	VbA24	KF924253
39	<i>C. parvum</i>	IipA9	KC885904	105	<i>C. cuniculus</i>	VbA37	GU971639
40	<i>C. parvum</i>	IiAa20G3R2	KC307779	106	<i>C. cuniculus</i>	VbA19	FJ262733
41	<i>C. parvum</i>	IioA16G1	JN867335	107	<i>C. cuniculus</i>	VaA19	GU971642
42	<i>C. parvum</i>	IidA26G1	AB560742	108	<i>C. cuniculus</i>	VbA25	GU097642
43	<i>C. parvum</i>	IiAa16G2R1	DQ192505	109	<i>C. cuniculus</i>	VbA32	GU097641
44	<i>C. parvum</i>	IidA19G1	DQ280496	110	<i>C. cuniculus</i>	VaA18	GU971636
45	<i>C. parvum</i>	IidA17G1	DQ280495	111	<i>C. cuniculus</i>	VbA28	GU971631
46	<i>C. parvum</i>	IioA13G1	KC885906	112	<i>C. cuniculus</i>	VbA29	FJ262734

47	<i>C. parvum</i>	IIaA15G1R1	JF727794	113	<i>C. cuniculus</i>	VbA21	HQ397717
48	<i>C. parvum</i>	IIcA5G3c	EU095267	114	<i>C. cuniculus</i>	VbA22	GU971644
49	<i>C. parvum</i>	IIaA19G2R1	JQ362495	115	<i>C. cuniculus</i>	VbA20	GU971649
50	<i>C. parvum</i>	IIaA20G3R1	JQ362497	116	<i>C. cuniculus</i>	VaA31	KC157565
51	<i>C. parvum</i>	IIaA19G3R1	JQ362496	117	<i>C. cuniculus</i>	VaA9	GU971640
52	<i>C. parvum</i>	IIaA16G3R1	JQ362492	118	<i>C. cuniculus</i>	VaA11	GU971637
53	<i>C. parvum</i>	IIcA5G3	HQ149041	119	<i>C. cuniculus</i>	VaA21	GU971634
54	<i>C. parvum</i>	IIaA14G1R1	JX575598	120	<i>C. cuniculus</i>	VaA23	GU971650
55	<i>C. meleagridis</i>	IIIeA20G1	AB539721	121	<i>C. ubiquitum</i>	XIIa	JX412915
56	<i>C. meleagridis</i>	IIIcA6	AF401497	122	<i>C. ubiquitum</i>	XIIb	JX412926
57	<i>C. meleagridis</i>	IIIfA16G2	EU164813	123	<i>C. ubiquitum</i>	XIIc	JX412925
58	<i>C. meleagridis</i>	IIIaA6	DQ067570	124	<i>C. ubiquitum</i>	XIId	JX412922
59	<i>C. meleagridis</i>	IIIgA19G5	JX878614	125	<i>C. ubiquitum</i>	XIIe	KC204983
60	<i>C. meleagridis</i>	IIIaA24G3	AF401499	126	<i>C. ubiquitum</i>	XIIf	KC204983
61	<i>C. meleagridis</i>	IIIbA20G1	AB539720	127	Horse genotype	VIbA13	FJ435961
62	<i>C. hominis</i>	IbA10G2	AY262031	128	Horse genotype	VIaA11G3	FJ435960
63	<i>C. hominis</i>	IbA6G3	JF727785	129	Ferret genotype	VIIIaA5G2	GQ121029
64	<i>C. hominis</i>	IeA11G3T3	AY738184	130	Mink genotype	XaA5G1	HM234174
65	<i>C. hominis</i>	IfA22G1	AB560750	131	Opossum genotype I	XIaA4G1T1	HM234181
66	<i>C. hominis</i>	IbA12G3	KF679725				

Evaluation du séquençage haut débit pour l'étude de *Cryptosporidium* spp.

AUTEUR : Me. Eloïse BAILLY

RESUMÉ :

Cryptosporidium spp. est un parasite entérique d'importance mondiale, responsable de diarrhées pouvant être sévères et chroniques chez l'immunodéprimé. Une connaissance solide de l'épidémiologie de ce parasite est primordiale afin d'en améliorer le contrôle. Le sous-typage de *Cryptosporidium* spp. s'inscrit dans cette optique de compréhension épidémiologique, notamment grâce au séquençage du gène codant la glycoprotéine 60 (gp60). La plupart des publications traitant du sous-typage de *Cryptosporidium* spp. reposent actuellement sur l'utilisation de PCR conventionnelle suivie de séquençage selon la technique Sanger, qui peut être mis en défaut par des mélanges d'ADN se traduisant par des infections mixtes à *Cryptosporidium* spp. Afin de dépasser ces limites, la littérature scientifique s'est enrichie ces dernières années de publications étudiant le séquençage haut débit (NGS) appliqué à *Cryptosporidium* spp. et rapportant la capacité de cette technique à identifier efficacement des variants minoritaires au sein de mélanges d'ADN. En nous appuyant sur ces publications, et à l'aide de mélanges d'ADN de *Cryptosporidium* sp. dont la composition était connue, nous avons pu confirmer les capacités du NGS dans la caractérisation de mélanges d'ADN et avoir une idée du seuil de détection des variants minoritaires.

MOTS-CLÉS : *Cryptosporidium* spp. – sous-typage – NGS – Sanger – gp60