

REPUBLIQUE DU CAMEROUN

Paix-Travail-Patrie

\*\*\*\*\*

UNIVERSITE DE YAOUNDE I

\*\*\*\*\*

FACULTE DES SCIENCES

\*\*\*\*\*

CENTRE DE RECHERCHE ET DE FORMATION  
DOCTORALE, SCIENCES DE LA VIE, SANTE  
ET ENVIRONNEMENT

\*\*\*\*\*

UNITE DE RECHERCHE ET DE FORMATION  
DOCTORALE EN SCIENCES DE LA VIE

\*\*\*\*\*



REPUBLIC OF CAMEROON

Peace-Work-Fatherland

\*\*\*\*\*

THE UNIVERSITY OF YAOUNDE I

\*\*\*\*\*

FACULTY OF SCIENCE

\*\*\*\*\*

CENTRE FOR RESEARCH AND TRAINING IN  
GRADUATE STUDIES IN LIFE, HEALTH AND  
ENVIRONMENTAL SCIENCE

\*\*\*\*\*

RESEARCH AND DOCTORALE TRAINING UNIT IN  
LIFE SCIENCE

\*\*\*\*\*

DEPARTEMENT DE MICROBIOLOGIE

DEPARTMENT OF MICROBIOLOGY

## Diversité des virus à ARN Contaminant les Palourdes de la Sanaga *Galatea paradoxa* (Born, 1778)

‘THESE’

Présentée et soutenue en vue de l’obtention du **Doctorat/Ph.D.** en

**MICROBIOLOGIE**

Option : Microbiologie Alimentaire et Industrielle

Par :

**BONNY Patrice**

Matricule : 08R0122

Master en Biochimie

Devant le jury composé de :



**Président :** ETOA François-Xavier, Professeur, Université de Yaoundé 1

**Rapporteurs :** ESSIA NGANG Jean Justin, Professeur, Université de Yaoundé 1

LE GUYADER Françoise SOIZICK, Directeur de Recherches, IFREMER, Nantes

**Membres :** MBACHAM Wilfred, Professeur, Université de Yaoundé 1

KORO KORO Francioli, Maître de Conférences, Université de Douala

BOYOMO ONANA, Maître de Conférences, Université de Yaoundé 1

Année 2023

UNIVERSITÉ DE YAOUNDÉ I  
I  
FACULTE DES SCIENCES  
Département de Microbiologie  
BP 812 - Yaoundé  
Tél./fax (237) 222 23 44 96



THE UNIVERSITY OF YAOUNDE

FACULTY OF SCIENCE  
Department of Microbiology  
P.O. Box 812 Yaoundé  
Tel./fax (237) 222 23 44 96

### ATTESTATION DE CORRECTION

Nous soussignés, membres du Jury de la soutenance de Thèse de Doctorat/ Ph. D de Microbiologie (Option Microbiologie Alimentaire et Industrielle) de Monsieur **BONNY Patrice**, Matricule **08R0122**, soutenance autorisée par la correspondance N° **223258/UY1/VREPDIC/DAAC/DEPE/SSPE/CB-nsr** du 11 Novembre 2022, et défendue le 09 Janvier 2023 à 10 heures dans la salle Multimédia de la Faculté des Sciences de l'Université de Yaoundé I sous le Thème : « **Diversité des virus à ARN contaminant les palourdes de la Sanaga Galatea paradoxa (Born, 1778)** », attestons que les corrections exigées au candidat lors de cette évaluation ont été effectuées et que le dit document peut être déposé sous sa forme actuelle.

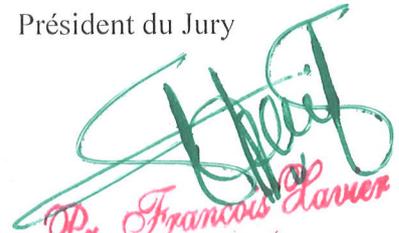
En foi de quoi la présente attestation lui est établie pour servir et valoir ce que de droit.

Fait à Yaoundé, le ... **14 NOV. 2023** .....

Examineurs

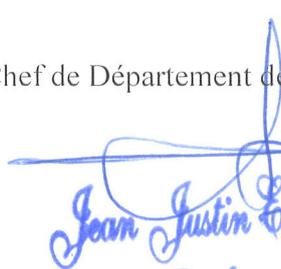
  
Koré Koré Fanebli  
  
Bongomo Ouana

Président du Jury

  
Pr. Francois Larmer  
Etoa

  
Prof. Abacham Wilfred Fon  
MS, DS, ScD (Harvard), FASI, FCAS, FAA  
Public Health Biotechnologist

Chef de Département de Microbiologie

  
Jean Justin  
Professeur



<b>&lt;UNIVERSITÉ DE YAOUNDÉ I</b> <b>Faculté des Sciences</b> Division de la Programmation et du Suivi des Activités Académiques		<b>THE UNIVERSITY OF YAOUNDE I</b> <b>Faculty of Science</b> Division of Programming and Follow-up of Academic Affairs
<b>LISTE DES ENSEIGNANTS PERMANENTS</b>		<b>LIST OF PERMANENT TEACHING STAFF</b>

**ANNÉE ACADEMIQUE 2022/2023**  
 (Par Département et par Grade)  
**DATE D'ACTUALISATION 31 MAI 2023**

**ADMINISTRATION**

**DOYEN** : TCHOUANKEU Jean- Claude, *Maître de Conférences*

**VICE-DOYEN / DPSAA**: ATCHADE Alex de Théodore, *Professeur*

**VICE-DOYEN / DSSE** : NYEGUE Maximilienne Ascension, *Professeur*

**VICE-DOYEN / DRC** : ABOSSOLO ANGUE Monique, *Maître de Conférences*

**Chef Division Administrative et Financière** : NDOYE FOE Florentine Marie Chantal, *Maître de Conférences*

**Chef Division des Affaires Académiques, de la Recherche et de la Scolarité DAARS** : AJEAGAH Gideon AGHAINDUM, *Professeur*

**1- DÉPARTEMENT DE BIOCHIMIE (BC) (43)**

N°	NOMS ET PRÉNOMS	GRADE	OBSERVATIONS
1.	BIGOGA DAIGA Jude	Professeur	En poste
2.	FEKAM BOYOM Fabrice	Professeur	En poste
3.	KANSCI Germain	Professeur	En poste
4.	MBACHAM FON Wilfred	Professeur	En poste
5.	MOUNDIPA FEWOU Paul	Professeur	<i>Chef de Département</i>
6.	NGUEFACK Julienne	Professeur	En poste
7.	NJAYOU Frédéric Nico	Professeur	En poste
8.	OBEN Julius ENYONG	Professeur	En poste

9.	ACHU Merci BIH	Maître de Conférences	En poste
10.	ATOGHO Barbara MMA	Maître de Conférences	En poste
11.	AZANTSA KINGUE GABIN BORIS	Maître de Conférences	En poste
12.	BELINGA née NDOYE FOE F. M. C.	Maître de Conférences	<i>Chef DAF / FS</i>
13.	DJUIDJE NGOUNOUE Marceline	Maître de Conférences	En poste
14.	DJUJKWO NKONGA Ruth Viviane	Maître de Conférences	En poste
15.	EFFA ONOMO Pierre	Maître de Conférences	<i>VD/FS/Univ Ebwa</i>
16.	EWANE Cécile Annie	Maître de Conférences	En poste
17.	KOTUE TAPTUE Charles	Maître de Conférences	En poste
18.	LUNGA Paul KEILAH	Maître de Conférences	En poste
19.	MBONG ANGIE M. Mary Anne	Maître de Conférences	En poste
20.	MOFOR née TEUGWA Clotilde	Maître de Conférences	<i>Doyen FS / UDs</i>
21.	NANA Louise épouse WAKAM	Maître de Conférences	En poste
22.	NGONDI Judith Laure	Maître de Conférences	En poste
23.	TCHANA KOUATCHOUA Angèle	Maître de Conférences	En poste

24.	AKINDEH MBUH NJI	Chargé de Cours	En poste
25.	BEBEE Fadimatou	Chargée de Cours	En poste
26.	BEBOY EDJENGUELE Sara Nathalie	Chargé de Cours	En poste
27.	DAKOLE DABOY Charles	Chargé de Cours	En poste
28.	DONGMO LEKAGNE Joseph Blaise	Chargé de Cours	En poste

29.	FONKOUA Martin	Chargé de Cours	En poste
30.	FOUPOUAPOUOGNIGNI Yacouba	Chargé de Cours	En poste
31.	KOUOH ELOMBO Ferdinand	Chargé de Cours	En poste
32.	MANANGA Marlyse Joséphine	Chargée de Cours	En poste
33.	OWONA AYISSI Vincent Brice	Chargé de Cours	En poste
34.	Palmer MASUMBE NETONGO	Chargé de Cours	En poste
35.	PECHANGOU NSANGO Sylvain	Chargé de Cours	En poste
36.	WILFRED ANGIE ABIA	Chargé de Cours	En poste

37.	BAKWU BASSOGOG Christian Bernard	Assistant	En Poste
38.	ELLA Fils Armand	Assistant	En Poste
39.	EYENGA Eliane Flore	Assistant	En Poste
40.	MADIESSE KEMGNE Eugenie Aimée	Assistant	En Poste
41.	MANJIA NJIKAM Jacqueline	Assistant	En Poste
42.	MBOUCHE FANMOE Marceline Joëlle	Assistant	En poste
43.	WOGUIA Alice Louise	Assistant	En Poste

## 2- DÉPARTEMENT DE BIOLOGIE ET PHYSIOLOGIE ANIMALES (BPA) (52)

1.	AJEAGAH Gideon AGHAINDUM	Professeur	<i>DAARS/FS</i>
2.	BILONG BILONG Charles-Félix	Professeur	<i>Chef de Département</i>
3.	DIMO Théophile	Professeur	En Poste
4.	DJIETO LORDON Champlain	Professeur	En Poste
5.	DZEUFUET DJOMENI Paul Désiré	Professeur	En Poste
6.	ESSOMBA née NTSAMA MBALA	Professeur	<i>CD et Vice Doyen/FMSB/UIYI</i>
7.	FOMENA Abraham	Professeur	En Poste
8.	KEKEUNOU Sévior	Professeur	En poste
9.	NJAMEN Dieudonné	Professeur	En poste
10.	NJIOKOU Flobert	Professeur	En Poste
11.	NOLA Moïse	Professeur	En poste
12.	TAN Paul VERNYUY	Professeur	En poste
13.	TCHUEM TCHUENTE Louis Albert	Professeur	<i>Inspecteur de service / Coord.Progr./MINSANTE</i>
14.	ZEBAZE TOGOUET Serge Hubert	Professeur	En poste

15.	ALENE Désirée Chantal	Maître de Conférences	<i>Vice Doyen/ Uté Ebwa</i>
16.	BILANDA Danielle Claude	Maître de Conférences	En poste
17.	DJIOGUE Séfirin	Maître de Conférences	En poste
18.	GOUNOUE KAMKUMO Raceline épouse FOTSING	Maître de Conférences	En poste
19.	JATSA BOUKENG Hermine épouse MEGAPTCHE	Maître de Conférences	En Poste
20.	LEKEUFACK FOLEFACK Guy B.	Maître de Conférences	En poste
21.	MAHOB Raymond Joseph	Maître de Conférences	En poste
22.	MBENOUN MASSE Paul Serge	Maître de Conférences	En poste
23.	MEGNEKOU Rosette	Maître de Conférences	En poste
24.	MOUNGANG Luciane Marlyse	Maître de Conférences	En poste
25.	NOAH EWOTI Olive Vivien	Maître de Conférences	En poste

26.	MONY Ruth épse NTONE	Maître de Conférences	En Poste
27.	NGUEGUIM TSOFAK Florence	Maître de Conférences	En poste
28.	NGUEMBOCK	Maître de Conférences	En poste
29.	TAMSA ARFAO Antoine	Maître de Conférences	En poste
30.	TOMBI Jeannette	Maître de Conférences	En poste

31.	ATSAMO Albert Donatien	Chargé de Cours	En poste
32.	BASSOCK BAYIHA Etienne Didier	Chargé de Cours	En poste
33.	ETEME ENAMA Serge	Chargé de Cours	En poste
34.	FEUGANG YOUNSSI François	Chargé de Cours	En poste
35.	FOKAM Alvine Christelle Epse KENGNE	Chargé de Cours	En poste
36.	GONWOUO NONO Legrand	Chargé de Cours	En poste
37.	KANDEDA KAVAYE Antoine	Chargé de Cours	En poste
38.	KOGA MANG DOBARA	Chargé de Cours	En poste
39.	LEME BANOCK Lucie	Chargé de Cours	En poste
40.	MAPON NSANGOU Indou	Chargé de Cours	En poste
41.	METCHI DONFACK MIREILLE FLAURE EPSE GHOUMO	Chargé de Cours	En poste
42.	MVEYO NDANKEU Yves Patrick	Chargé de Cours	En poste
43.	NGOUATEU KENFACK Omer Bébé	Chargé de Cours	En poste
44.	NJUA Clarisse YAFI	Chargée de Cours	<i>Chef Div. Uté Bamenda</i>
45.	NWANE Philippe Bienvenu	Chargé de Cours	En poste
46.	TADU Zephyrin	Chargé de Cours	En poste
47.	YEDE	Chargé de Cours	En poste
48.	YOUNOUSSA LAME	Chargé de Cours	En poste

49.	AMBADA NDZENGUE GEORGIA ELNA	Assistante	En poste
50.	KODJOM WANCHE Jacguy Joyce	Assistante	En poste
51.	NDENGUE Jean De Matha	Assistant	En poste
52.	ZEMO GAMO Franklin	Assistant	En poste

### 3- DÉPARTEMENT DE BIOLOGIE ET PHYSIOLOGIE VÉGÉTALES (BPV) (34)

1.	AMBANG Zachée	Professeur	<i>Chef de Département</i>
2.	DJOCGOUE Pierre François	Professeur	En poste
3.	MBOLO Marie	Professeur	En poste
4.	MOSSEBO Dominique Claude	Professeur	En poste
5.	YOUMBI Emmanuel	Professeur	En poste
6.	ZAPFACK Louis	Professeur	En poste

7.	ANGONI Hyacinthe	Maître de Conférences	En poste
8.	BIYE Elvire Hortense	Maître de Conférences	En poste
9.	MAHBOU SOMO TOUKAM. Gabriel	Maître de Conférences	En poste
10.	MALA Armand William	Maître de Conférences	En poste
11.	MBARGA BINDZI Marie Alain	Maître de Conférences	<i>DAAC /UDla</i>
12.	NDONGO BEKOLO	Maître de Conférences	En poste
13.	NGALLE Hermine BILLE	Maître de Conférences	En poste
14.	NGODO MELINGUI Jean Baptiste	Maître de Conférences	En poste
15.	NGONKEU MAGAPTCHE Eddy L.	Maître de Conférences	<i>CT / MINRESI</i>

16.	TONFACK Libert Brice	Maître de Conférences	En poste
17.	TSOATA Esaïe	Maître de Conférences	En poste
18.	ONANA JEAN MICHEL	Maître de Conférences	En poste

19.	DJEUANI Astride Carole	Chargé de Cours	En poste
20.	GONMADGE CHRISTELLE	Chargée de Cours	En poste
21.	MAFFO MAFFO Nicole Liliane	Chargé de Cours	En poste
22.	NNANGA MEBENGA Ruth Laure	Chargé de Cours	En poste
23.	NOUKEU KOUAKAM Armelle	Chargé de Cours	En poste
24.	NSOM ZAMBO EPSE PIAL ANNIE CLAUDE	Chargé de Cours	<i>En détachement/UNESCO MALI</i>
25.	GODSWILL NTSOMBOH NTSEFONG	Chargé de Cours	En poste
26.	KABELONG BANAHOU Louis-Paul-Roger	Chargé de Cours	En poste
27.	KONO Léon Dieudonné	Chargé de Cours	En poste
28.	LIBALAH Moses BAKONCK	Chargé de Cours	En poste
29.	LIKENG-LI-NGUE Benoit C	Chargé de Cours	En poste
30.	TAEDOUNG Evariste Hermann	Chargé de Cours	En poste
31.	TEMEGNE NONO Carine	Chargé de Cours	En poste
32.	MANGA NDJAGA JUDE	Assistant	En poste
33.	DIDA LONTSI Sylvere Landry	Assistant	En poste
34.	METSEBING Blondo-Pascal	Assistant	En poste

#### 4- DÉPARTEMENT DE CHIMIE INORGANIQUE (CI) (28)

1.	GHOYOMU Paul MINGO	Professeur	<i>Ministre Chargé de Mission PR</i>
2.	NANSEU NJIKI Charles Péguy	Professeur	En poste
3.	NDIFON Peter TEKE	Professeur	<i>CT MINRESI</i>
4.	NENWA Justin	Professeur	En poste
5.	NGAMENI Emmanuel	Professeur	<i>Doyen FS Univ.Ngaoundere</i>
6.	NGOMO Horace MANGA	Professeur	<i>Vice Chancellor/UB</i>
7.	NJOYA Dayirou	Professeur	En poste

8.	ACAYANKA Elie	Maître de Conférences	En poste
9.	EMADAK Alphonse	Maître de Conférences	En poste
10.	KAMGANG YOUNBI Georges	Maître de Conférences	En poste
11.	KEMMEGNE MBOUGUEM Jean C.	Maître de Conférences	En poste
12.	KENNE DEDZO GUSTAVE	Maître de Conférences	En poste
13.	MBEY Jean Aime	Maître de Conférences	En poste
14.	NDI NSAMI Julius	Maître de Conférences	<i>Chef de Département</i>
15.	NEBAH Née NDOIRI Bridget NDOYE	Maître de Conférences	<i>Sénatrice/SENAT</i>
16.	NJIOMOU C. épouse DJANGANG	Maître de Conférences	En poste
17.	NYAMEN Linda Dyorisse	Maître de Conférences	En poste
18.	PABOUDAM GBAMBIE AWAWOU	Maître de Conférences	En poste
19.	TCHAKOUTE KOUAMO Hervé	Maître de Conférences	En poste
20.	BELIBI BELIBI Placide Désiré	Maître de Conférences	<i>Chef Service/ ENS Bertoua</i>
21.	CHEUMANI YONA Arnaud M.	Maître de Conférences	En poste

22.	KOUOTOU DAOUDA	Maître de Conférences	En poste
23.	MAKON Thomas Beauregard	Chargé de Cours	En poste
24.	NCHIMI NONO KATIA	Chargée de Cours	En poste
25.	NJANKWA NJABONG N. Eric	Chargé de Cours	En poste
26.	PATOUOSSA ISSOFA	Chargé de Cours	En poste
27.	SIEWE Jean Mermoz	Chargé de Cours	En Poste
28.	BOYOM TATCHEMO Franck W.	Assistant	En Poste

#### 5- DÉPARTEMENT DE CHIMIE ORGANIQUE (CO) (37)

1.	Alex de Théodore ATCHADE	Professeur	<i>Vice-Doyen / DPSAA</i>
2.	DONGO Etienne	Professeur	<i>Vice-Doyen/FSE/UIYI</i>
3.	NGOUELA Silvère Augustin	Professeur	<i>Chef de Département UDS</i>
4.	PEGNYEMB Dieudonné Emmanuel	Professeur	<i>Directeur/ MINESUP/ Chef de Département</i>
5.	WANDJI Jean	Professeur	En poste
6.	MBAZOA née DJAMA Céline	Professeur	En poste

7.	AMBASSA Pantaléon	Maître de Conférences	En poste
8.	EYONG Kenneth OBEN	Maître de Conférences	En poste
9.	FOTSO WABO Ghislain	Maître de Conférences	En poste
10.	KAMTO Eutrophe Le Doux	Maître de Conférences	En poste
11.	KENMOGNE Marguerite	Maître de Conférences	En poste
12.	KEUMEDJIO Félix	Maître de Conférences	En poste
13.	KOUAM Jacques	Maître de Conférences	En poste
14.	MKOUNGA Pierre	Maître de Conférences	En poste
15.	MVOT AKAK CARINE	Maître de Conférences	En poste
16.	NGO MBING Joséphine	Maître de Conférences	<i>Chef de Cellule MINRESI</i>
17.	NGONO BIKOBO Dominique Serge	Maître de Conférences	<i>C.E.A/ MINESUP</i>
18.	NOTE LOUGBOT Olivier Placide	Maître de Conférences	<i>DAAC/Uté Bertoua</i>
19.	NOUNGOUE TCHAMO Diderot	Maître de Conférences	En poste
20.	TABOPDA KUATE Turibio	Maître de Conférences	En poste
21.	TAGATSING FOTSING Maurice	Maître de Conférences	En poste
22.	TCHOUANKEU Jean-Claude	Maître de Conférences	<i>Doyen /FS/ UYI</i>
23.	YANKEP Emmanuel	Maître de Conférences	En poste
24.	ZONDEGOUMBA Ernestine	Maître de Conférences	En poste

25.	MESSI Angélique Nicolas	Chargé de Cours	En poste
26.	NGNINTEDO Dominique	Chargé de Cours	En poste
27.	NGOMO Orléans	Chargée de Cours	En poste
28.	NONO NONO Éric Carly	Chargé de Cours	En poste
29.	OUAHOUE WACHE Blandine M.	Chargée de Cours	En poste
30.	OUETE NANTCHOUANG Judith Laure	Chargée de Cours	En poste
31.	SIELINOUE TEDJON Valérie	Chargé de Cours	En poste
32.	TCHAMGOUE Joseph	Chargé de Cours	En poste

33.	TSAFFACK Maurice	Chargé de Cours	En poste
34.	TSAMO TONTSA Armelle	Chargé de Cours	En poste
35.	TSEMEUGNE Joseph	Chargé de Cours	En poste

36.	MUNVERA MFIFEN Aristide	Assistant	En poste
37.	NDOGO ETEME Olivier	Assistant	En poste

#### 6- DÉPARTEMENT D'INFORMATIQUE (IN) (22)

1.	ATSA ETOUNDI Roger	Professeur	<i>Chef de Division MINESUP</i>
2.	FOUDA NDJODO Marcel Laurent	Professeur	<i>Inspecteur Général/ MINESUP</i>

3.	NDOUNDAM René	Maître de Conférences	En poste
4.	TSOPZE Norbert	Maître de Conférences	En poste

5.	ABESSOLO ALO'O Gislain	Chargé de Cours	<i>Chef de Cellule MINFOPRA</i>
6.	AMINO HALIDOU	Chargé de Cours	<i>Chef de Département</i>
7.	DJAM Xaviera YOUH - KIMBI	Chargé de Cours	En Poste
8.	DOMGA KOMGUEM Rodrigue	Chargé de Cours	En poste
9.	EBELE Serge Alain	Chargé de Cours	En poste
10.	HAMZA Adamou	Chargé de Cours	En poste
11.	JIOMEKONG AZANZI Fidel	Chargé de Cours	En poste
12.	KOUOKAM KOUOKAM E. A.	Chargé de Cours	En poste
13.	MELATAGIA YONTA Paulin	Chargé de Cours	En poste
14.	MESSI NGUELE Thomas	Chargé de Cours	En poste
15.	MONTHE DJIADEU Valery M.	Chargé de Cours	En poste
16.	NZEKON NZEKO'O ARMEL JACQUES	Chargé de Cours	En poste
17.	OLLE OLLE Daniel Claude Georges Delort	Chargé de Cours	<i>Sous-Directeur ENSET Ebolowa</i>
18.	TAPAMO Hyppolite	Chargé de Cours	En poste

19.	BAYEM Jacques Narcisse	Assistant	En poste
20.	EKODECK Stéphane Gaël Raymond	Assistant	En poste
21.	MAKEMBE. S . Oswald	Assistant	<i>Directeur CUTI</i>
22.	NKONDOCK. MI. BAHANACK.N.	Assistant	En poste

#### 7- DÉPARTEMENT DE MATHÉMATIQUES (MA) (33)

1.	AYISSI Raoult Domingo	Professeur	<i>Chef de Département</i>
2.	KIANPI Maurice	Maître de Conférences	En poste
3.	MBANG Joseph	Maître de Conférences	En poste
4.	MBEHOU Mohamed	Maître de Conférences	En poste
5.	MBELE BIDIMA Martin Ledoux	Maître de Conférences	En poste
6.	NOUNDJEU Pierre	Maître de Conférences	<i>Chef Service des Programmes &amp; Diplômes/FS/Uyi</i>

7.	TAKAM SOH Patrice	Maître de Conférences	En poste
8.	TCHAPNDA NJABO Sophonie B.	Maître de Conférences	<i>Directeur/AIMS Rwanda</i>
9.	TCHOUNDJA Edgar Landry	Maître de Conférences	En poste

10.	AGHOUKENG JIOFACK Jean Gérard	Chargé de Cours	<i>Chef Cellule MINEPAT</i>
11.	BOGSO ANTOINE Marie	Chargé de Cours	En poste
12.	CHENDJOU Gilbert	Chargé de Cours	En poste
13.	DJIADU NGAHA Michel	Chargé de Cours	En poste
14.	DOUANLA YONTA Herman	Chargé de Cours	En poste
15.	KIKI Maxime Armand	Chargé de Cours	En poste
16.	LOUMNGAM KAMGA Victor	Chargé de Cours	En poste
17.	MBAKOP Guy Merlin	Chargé de Cours	En poste
18.	MBATAKOU Salomon Joseph	Chargé de Cours	En poste
19.	MENGUE MENGUE David Joël	Chargé de Cours	<i>Chef Dpt /ENS Université d'Ebolowa</i>
20.	MBIAKOP Hilaire George	Chargé de Cours	En poste
21.	NGUEFACK Bernard	Chargé de Cours	En poste
22.	NIMPA PEFOUKEU Romain	Chargée de Cours	En poste
23.	OGADOA AMASSAYOGA	Chargée de Cours	En poste
24.	POLA DOUNDOU Emmanuel	Chargé de Cours	<i>En stage</i>
25.	TCHEUTIA Daniel Duviol	Chargé de Cours	En poste
26.	TETSADJIO TCHILEPECK M. Eric.	Chargé de Cours	En poste

27.	BITYE MVONDO Esther Claudine	Assistante	En poste
28.	FOKAM Jean Marcel	Assistant	En poste
29.	GUIDZAVAI KOUCHERE Albert	Assistant	En poste
30.	MANN MANYOMBE Martin Luther	Assistant	En poste
31.	MEFENZA NOUNTU Thiery	Assistant	En poste
32.	NYOUMBI DLEUNA Christelle	Assistant	En poste
33.	TENKEU JEUFACK Yannick Léa	Assistant	En poste

#### **8- DÉPARTEMENT DE MICROBIOLOGIE (MIB) (24)**

1.	ESSIA NGANG Jean Justin	Professeur	<i>Chef de Département</i>
2.	NYEGUE Maximilienne Ascension	Professeur	<i>VICE-DOYEN / DSSE</i>

3.	ASSAM ASSAM Jean Paul	Maître de Conférences	En poste
4.	BOUGNOM Blaise Pascal	Maître de Conférences	En poste
5.	BOYOMO ONANA	Maître de Conférences	En poste
6.	KOUTCHEU MABEKU Epse KOUAM Laure Brigitte	Maître de Conférences	En poste
7.	RIWOM Sara Honorine	Maître de Conférences	En poste
8.	NJIKI BIKOÏ Jacky	Maître de Conférences	En poste
9.	SADO KAMDEM Sylvain Leroy	Maître de Conférences	En poste

10.	ESSONO Damien Marie	Chargé de Cours	En poste
11.	LAMYE Glory MOH	Chargé de Cours	En poste
12.	MEYIN A EBONG Solange	Chargé de Cours	En poste
13.	MONI NDEDI Esther Del Florence	Chargé de Cours	En poste
14.	NKOUDOU ZE Nardis	Chargé de Cours	En poste
15.	TAMATCHO KWEYANG Blandine Pulchérie	Chargé de Cours	En poste
16.	TCHIKOUA Roger	Chargé de Cours	<i>Chef de Service de la Scolarité</i>
17.	TOBOLBAÏ Richard	Chargé de Cours	En poste

18.	NKOUÉ TONG Abraham	Assistant	En poste
19.	SAKE NGANE Carole Stéphanie	Assistant	En poste
20.	EZO'O MENGO Fabrice Téléfor	Assistant	En poste
21.	EHETH Jean Samuel	Assistant	En poste
22.	MAYI Marie Paule Audrey	Assistant	En poste
23.	NGOUE NAM Romial Joël	Assistant	En poste
24.	NJAPNDOUNKE Bilkissou	Assistant	En poste

### 9. DEPARTEMENT DE PYSIQUE(PHY) (43)

1.	BEN- BOLIE Germain Hubert	Professeur	En poste
2.	DJUIDJE KENMOE épouse ALOYEM	Professeur	En poste
3.	EKOBENA FOU DA Henri Paul	Professeur	<i>Vice-Recteur. Uté Ngaoundéré</i>
4.	ESSIMBI ZOBO Bernard	Professeur	En poste
5.	HONA Jacques	Professeur	En poste
6.	NANA ENGO Serge Guy	Professeur	En poste
7.	NANA NBENDJO Blaise	Professeur	En poste
8.	NDJAKA Jean Marie Bienvenu	Professeur	<i>Chef de Département</i>
9.	NJANDJOCK NOUCK Philippe	Professeur	En poste
10.	NOUAYOU Robert	Professeur	En poste
11.	SAIDOU	Professeur	<i>Chef de centre/IRGM/MINRESI</i>
12.	TABOD Charles TABOD	Professeur	<i>Doyen FSUniv/Bda</i>
13.	TCHAWOUA Clément	Professeur	En poste
14.	WOAFO Paul	Professeur	En poste
15.	ZEKENG Serge Sylvain	Professeur	En poste
16.	BIYA MOTTO Frédéric	Maître de Conférences	<i>DG/HYDRO Mekin</i>
17.	BODO Bertrand	Maître de Conférences	En poste
18.	ENYEGUE A NYAM épse BELINGA	Maître de Conférences	En poste
19.	EYEBE FOU DA Jean sire	Maître de Conférences	En poste
20.	FEWO Serge Ibraïd	Maître de Conférences	En poste
21.	MBINACK Clément	Maître de Conférences	En poste
22.	MBONO SAMBA Yves Christian U.	Maître de Conférences	En poste
23.	MEL' I Joelle Larissa	Maître de Conférences	En poste

24.	MVOGO ALAIN	Maître de Conférences	En poste
25.	NDOP Joseph	Maître de Conférences	En poste
26.	SIEWE SIEWE Martin	Maître de Conférences	En poste
27.	SIMO Elie	Maître de Conférences	En poste
28.	VONDOU Derbetini Appolinaire	Maître de Conférences	En poste
29.	WAKATA née BEYA Annie Sylvie	Maître de Conférences	<i>Directeur/ENS/UWI</i>
30.	WOULACHE Rosalie Laure	Maître de Conférence	<i>En stage depuis février 2023</i>
31.	ABDOURAHIMI	Chargé de Cours	En poste
32.	AYISSI EYEBE Guy François Valérie	Chargé de Cours	En poste
33.	CHAMANI Roméo	Chargé de Cours	En poste
34.	DJIOTANG TCHOTCHOU Lucie Angennes	Chargée de Cours	En poste
35.	EDONGUE HERVAIS	Chargé de Cours	En poste
36.	FOUEJIO David	Chargé de Cours	<i>Chef Cell. MINADER</i>
37.	KAMENI NEMATCHOUA Modeste	Chargé de Cours	En poste
38.	LAMARA Maurice	Chargé de Cours	En poste
39.	OTTOU ABE Martin Thierry	Chargé de Cours	Directeur Unité de production des réactifs/IMPM
40.	TEYOU NGOUPO Ariel	Chargé de Cours	En poste
41.	WANDJI NYAMSI William	Chargé de Cours	En poste
42.	NGA ONGODO Dieudonné	Assistant	En poste
43.	SOUFFO TAGUEU Merimé	Assistant	En poste

**10- DÉPARTEMENT DE SCIENCES DE LA TERRE (ST) (42)**

1.	BITOM Dieudonné-Lucien	Professeur	<i>Doyen / FASA /UDs</i>
2.	NDAM NGOUPAYOU Jules-Remy	Professeur	En poste
3.	NDJIGUI Paul-Désiré	Professeur	<i>Chef de Département</i>
4.	NGOS III Simon	Professeur	En poste
5.	NKOUMBOU Charles	Professeur	En poste
6.	NZENTI Jean-Paul	Professeur	En poste
7.	ONANA Vincent Laurent	Professeur	<i>Chef de Département/Uté. Eb.</i>
8.	YENE ATANGANA Joseph Q.	Professeur	<i>Chef Div. /MINTP</i>

9.	ABOSSOLO née ANGUE Monique	Maître de Conférences	<i>Vice-Doyen / DRC</i>
10.	BISSO Dieudonné	Maître de Conférences	En poste
11.	EKOMANE Emile	Maître de Conférences	<i>Chef Div./Uté Ebolowa</i>
12.	Elisé SABABA	Maitre de Conférences	En poste
13.	FUH Calistus Gentry	Maître de Conférences	<i>Sec. d'Etat/MINMIDT</i>
14.	GANNO Sylvestre	Maître de Conférences	En poste
15.	GHOGOMU Richard TANWI	Maître de Conférences	<i>Chef de Div. /Uté Bertoua</i>
16.	MBIDA YEM	Maitre de Conférences	En poste
17.	MOUNDI Amidou	Maître de Conférences	<i>CT/MINIMDT</i>
18.	NGO BIDJECK Louise Marie	Maître de Conférences	En poste

19.	NGUEUTCHOUA Gabriel	Maître de Conférences	<i>CEA/MINRESI</i>
20.	NJILAH Isaac KONFOR	Maître de Conférences	En poste
21.	NYECK Bruno	Maître de Conférences	En poste
22.	TCHAKOUNTE Jacqueline épouse NUMBEM	Maître de Conférences	<i>Chef. Cell /MINRESI</i>
23.	TCHOUANKOUE Jean-Pierre	Maître de Conférences	En poste
24.	TEMGA Jean Pierre	Maître de Conférences	En poste
25.	ZO'O ZAME Philémon	Maître de Conférences	<i>DG/ART</i>

26.	ANABA ONANA Achille Basile	Chargé de Cours	En poste
27.	BEKOA Etienne	Chargé de Cours	En poste
28.	ESSONO Jean	Chargé de Cours	En poste
29.	EYONG John TAKEM	Chargé de Cours	En poste
30.	MAMDEM TAMTO Lionelle Estelle, épouse BITOM	Chargée de Cours	En poste
31.	MBESSE Cécile Olive	Chargée de Cours	En poste
32.	METANG Victor	Chargé de Cours	En poste
33.	MINYEM Dieudonné	Chargé de Cours	<i>Chef Serv./Uté Maroua</i>
34.	NGO BELNOUN Rose Noël	Chargée de Cours	En poste
35.	NOMO NEGUE Emmanuel	Chargé de Cours	En poste
36.	NTSAMA ATANGANA Jacqueline	Chargée de Cours	En poste
37.	TCHAPTCHET TCHATO De P.	Chargé de Cours	En poste
38.	TEHNA Nathanaël	Chargé de Cours	En poste
39.	FEUMBA Roger	Chargé de Cours	En poste
40.	MBANGA NYOBE Jules	Chargé de Cours	En poste

41.	KOAH NA LEBOGO Serge Parfait	Assistant	En poste
42.	NGO'O ZE ARNAUD	Assistant	En poste
43.	TENE DJOUKAM Joëlle Flore, épouse KOUANKAP NONO	Assistante	En poste

### Répartition chiffrée des Enseignants de la Faculté des Sciences de l'Université de Yaoundé I

NOMBRE D'ENSEIGNANTS					
DÉPARTEMENT	Professeurs	Maîtres de Conférences	Chargés de Cours	Assistants	Total
BCH	8 (01)	15 (11)	13 (03)	7 (05)	<b>43 (20)</b>
BPA	14 (01)	16 (09)	18 (04)	4 (02)	<b>52 (16)</b>
BPV	6 (01)	12 (02)	13 (07)	3 (00)	<b>34 (10)</b>
CI	7 (01)	15 (04)	5 (01)	1 (00)	<b>28 (06)</b>
CO	6 (01)	18 (04)	11 (04)	2 (00)	<b>37 (09)</b>
IN	2 (00)	2 (00)	14 (01)	4 (00)	<b>22 (01)</b>
MAT	1 (00)	8 (00)	17 (01)	7 (02)	<b>33 (03)</b>
MIB	2 (01)	7 (03)	8 (04)	7 (02)	<b>24 (10)</b>
PHY	15 (01)	15 (04)	11 (01)	2 (00)	<b>43 (06)</b>
ST	8 (00)	17 (03)	15 (04)	3 (01)	<b>43 (08)</b>
<b>Total</b>	<b>69 (07)</b>	<b>125 (40)</b>	<b>125 (30)</b>	<b>40 (12)</b>	<b>359 (89)</b>

Soit un total de	<b>359 (89)</b> dont :
- Professeurs	<b>69 (07)</b>
- Maîtres de Conférences	<b>125 (40)</b>
- Chargés de Cours	<b>125 (30)</b>
- Assistants	<b>40 (12)</b>

( ) = Nombre de Femmes      **89**

**Dédicace**

A ma famille.

## REMERCIEMENTS

Mes remerciements vont à l'endroit de ceux qui ont participé à la réalisation de ce travail. Qu'il me soit permis de remercier particulièrement :

Le **Professeur ESSIA NGANG Jean Justin**, Chef de Département de Microbiologie de la Faculté des Sciences de l'Université de Yaoundé I, pour m'avoir fourni un cadre académique adéquat à la réalisation de ce travail. Par ailleurs, je vous remercie d'avoir co-dirigé ce travail de thèse. Vos orientations, conseils et encouragements divers tout au long de ce périple ont été déterminants dans la concrétisation de ce travail. Bien plus qu'un Directeur de thèse, vous vous êtes révélés être un second père pour moi. Pour tout ceci, je vous dis infiniment MERCI.

Le **Docteur LE GUYADER Françoise Soizick**, Chef du Laboratoire Santé, Environnement et Microbiologie (LSEM) de l'Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), Centre Atlantique de Nantes, et Directeur du Laboratoire de National de Référence pour la Microbiologie des Coquillages, pour avoir accepté de co-diriger cette thèse et de m'avoir ouvert les portes de votre laboratoire. Merci pour les moyens financiers, matériels et administratifs que vous avez mobilisés pour la réalisation de ce travail. Au cours de la première vague de la pandémie de COVID-19, alors même que le matériel était réquisitionné pour les tests de diagnostic dans les hôpitaux, et le confinement battait son plein, vous avez accordé une priorité à mes travaux, et avez obtenu que je fasse partie des premiers effectifs à reprendre le travail. Pour tout ceci, je vous dis infiniment MERCI.

Monsieur le **Directeur Général** de l'Institut de Recherches Médicales et d'Etudes des Plantes Médicinales (IMPM), le **Professeur ESSAME OYONO Jean-Louis**, pour les autorisations administratives ayant facilité ma mise en stage, ainsi que mes déplacements vers mon laboratoire d'accueil de l'IFREMER.

Monsieur le **Directeur du Centre Atlantique** de l'IFREMER de Nantes de m'avoir permis de réaliser mes stages au sein de votre structure.

Les **Membres du Jury** pour avoir accepté d'évaluer cette thèse en dépit de vos agendas chargés. Vos remarques et critiques permettront d'améliorer sa qualité scientifique.

Tous les enseignants du Département de Microbiologie de la Faculté des Sciences de l'Université de Yaoundé I qui n'ont ménagés aucun effort en participant aux différents

séminaires de suivi du déroulement des travaux de cette thèse. Vos remarques et conseils ont été déterminants pour l'amélioration de la qualité scientifique de ce travail.

Le **Docteur Marion DESDOUITS**, Chargé de Recherche au LSEM du Centre Atlantique de l'IFREMER, pour votre contribution à la conception de ce travail, ainsi que vos multiples suggestions. J'ai été impressionné par votre rigueur scientifique et votre grande maîtrise dans des domaines aussi complexes que variés de la biologie. J'espère pouvoir continuer à apprendre à vos côtés.

**Julien SCHEAFFER**, Ingénieur de Recherche au LSEM du Centre Atlantique de l'IFREMER, pour la mise en confiance, l'initiation et la formation aux analyses de biologie moléculaire, et en particulier à la préparation des librairies d'ADN pour le séquençage NGS.

**Alban BESNARD**, Bio-informaticien au LSEM du Centre Atlantique de l'IFREMER, pour votre contribution inestimable au traitement bio-informatique des données de séquençage, et mon initiation à cette discipline.

A tous les autres membres du LSEM du Centre Atlantique de l'IFREMER. Je pense notamment à Pascal, Sylvain, Jean-Côme, Cécile, Joanna, Françoise, Candice, Gabriel, Sarah, Jessica, Gabriel pour l'accueil chaleureux et l'intégration spontanée au sein de la belle équipe de recherche que vous constituez. L'attention et la délicatesse que vous avez portées à mon endroit m'ont permis de me sentir à la maison dès le premier jour. Je garde en mémoire les pauses café et les pique-niques, véritables moments d'échange et de communion. Merci de m'avoir accepté dans la « Meilleure équipe de recherche de France ».

Le **Docteur Bas Oude Munnink** et **David Nieuwenhuijse** du Département de Viroscience du Centre Médical de l'Université Erasmus de Rotterdam aux Pays-Bas, pour les discussions scientifiques ayant apportées une plus value dans le traitement bio-informatique des données du séquençage.

Aux **Docteurs NGOLONG NGEA Guillaume Legrand** et **NKOUDOU ZE Nardis** pour l'accompagnement, les conseils et les suggestions apportés.

Mes camarades doctorants, en particulier **MBU'U MBAMWI Cyrille**, pour l'amitié, les conseils.

Mes parents **DIBAGA Amos** et **ETONDE Alvine**, mes frères et sœurs **ELIMBI Richard**, **EBELLE Sylvie**, **ETONDE Julienne**, **DIBAGA Florence**, **Joseph Blanchard**

**DIBAGA, Alan Auriol DIBAGA** qui n'ont ménagé aucun effort pour m'apporter le soutien moral et financier, ainsi que la chaleur familiale nécessaires pour conduire ce travail.

**NTANKHEUH TCHINKOU Joséphine** pour le soutien moral et le réconfort. Dans les multiples moments difficiles tu as su me remonter le moral. Merci pour la patience dont tu as su faire montre durant ce très long périple. Je voudrais t'exprimer ici ma profonde gratitude.

Aux populations des villages Bolounga et Moulongo du Canton Malimba dans l'arrondissement de Mouanko, et particulièrement à la **Famille Jiengué** pour l'accueil, l'hébergement et la fourniture des échantillons.

# SOMMAIRE

Dédicace.....	xii
REMERCIEMENTS.....	xiii
RESUME.....	xviii
ABSTRACT .....	xix
LISTE DES ABREVIATIONS ET ACRONYMES .....	xx
LISTE DES TABLEAUX .....	Erreur ! Signet non défini.
LISTE DES FIGURES.....	xxiv
LISTE DES PHOTOS.....	xxvi
INTRODUCTION GENERALE.....	1
CHAPITRE I : REVUE DE LA LITTERATURE.....	4
1.1 Généralités sur les mollusques bivalves .....	4
1.1.1 Définition.....	4
1.1.2 Mode de nutrition .....	4
1.2 Importance économique et nutritionnelle des mollusques bivalves .....	5
1.3 Les virus d'origine alimentaire.....	7
1.3.1 Généralités sur les virus.....	7
1.3.2 Les virus d'origine alimentaire.....	7
1.3.3 Modes de contamination des aliments.....	10
1.4 Les principaux virus d'origine alimentaire transmis par les bivalves .....	12
1.4.1 Les norovirus .....	14
1.4.2 Le virus de l'hépatite A .....	18
1.4.3 Le virus de l'hépatite E.....	19
1.4.4 Les sapovirus .....	20
1.4.5 Les rotavirus .....	21
1.4.6 Les astrovirus.....	23
1.4.7 Autres virus émergents .....	24
1.5 Epidémies liées à la consommation des coquillages et les différents virus impliqués .....	27
1.6 Approche de détection et de quantification des virus dans les mollusques bivalves .....	28
1.6.1 Dissection et extraction des tissus digestifs.....	28
1.6.2 Elution virale .....	28
1.6.3 Extraction des acides nucléiques .....	29
1.6.4 Détection/quantification .....	30
1.7 Approche métagénomique de l'étude du virome des bivalves .....	31

1.7.1	Enjeux de l'étude métagénomique du virome .....	31
1.7.2	Analyse métagénomique de la contamination virale .....	32
<b>CHAPITRE II : MATERIEL ET METHODES .....</b>		<b>49</b>
2.1	Matériel.....	49
2.1.1	Matériel biologique.....	49
2.1.2	Réactifs .....	49
2.1.3	Appareillage .....	50
2.2	Méthodes .....	51
2.2.1	Description du site d'échantillonnage.....	51
2.2.3	Détection, quantification et génotypage des norovirus et le virus de l'hépatite A contaminant les palourdes de la Sanaga.....	52
2.2.4	Détermination de la diversité des virus à ARN présents dans les palourdes de la Sanaga .....	62
2.2.5	Identification des virus à ARN constituant un danger chez les humains.....	69
<b>CHAPITRE III : RESULTATS ET DISCUSSION .....</b>		<b>70</b>
3.1.	Résultats.....	70
3.1.1.	Contamination virale des palourdes de la Sanaga.....	70
3.1.2	Diversité des virus à ARN présents dans les palourdes de la Sanaga.....	80
3.1.3	Virus à ARN humains et les animaux pouvant avoir un intérêt en santé humaine .....	85
3.2.	Discussion.....	100
<b>CONCLUSION ET PERSPECTIVES .....</b>		<b>109</b>
<b>RECOMMANDATIONS.....</b>		<b>112</b>
<b>REFERENCES BIBLIOGRAPHIQUES.....</b>		<b>113</b>
<b>REFERENCES WEBOGRAPHIQUES: .....</b>		<b>139</b>
<b>ANNEXES.....</b>		<b>a</b>

## RESUME

De nombreuses pandémies récentes ont été reconnues comme des maladies virales. Si leurs origines restent souvent méconnues, l'environnement semble jouer un rôle important dans leur émergence. Ainsi, connaître la diversité des virus présents dans des échantillons alimentaires issus de l'environnement peut aider à comprendre les mécanismes d'émergence et de transmission des agents viraux. Ce travail décrit l'évaluation de la qualité virologique des palourdes de la Sanaga et la diversité des virus à ARN qu'elles hébergent et qui constituent les principaux virus pathogènes transmis par les aliments. Des palourdes collectées entre février 2018 et février 2020 à Bolounga et Mouloungo ont été testées par RT-qPCR pour la présence des norovirus (NoV) et du virus de l'hépatite A (VHA). Les régions codant la polymérase et la protéine de la capsid des NoV, ainsi que la jonction VP1/2A du VHA ont été amplifiées, séquencées et géotypées à l'aide de l'outil *Norovirus Typing Tool 2.0*. Des échantillons sélectionnés ont été analysés par une approche métagénomique. Les étapes de préparation des bibliothèques d'ADN comprenant une capture (VirCapSeq-VERT) des séquences d'intérêt, et de traitement bio-informatique comprenant l'élimination des séquences d'origine bactérienne et de l'hôte, ainsi qu'une déduplication avant l'assemblage *de novo* ont permis de cibler les virus à ARN des vertébrés. Il ressort que 80% des échantillons étaient contaminés par au moins un des virus, avec les NoV présents dans 69% des échantillons contre 36% pour le VHA. Nous avons observé une prévalence globale supérieure à Mouloungo (87%) par rapport à Bolounga (72%). Les NoV appartenaient aux géotypes GI.1, GII.3, GII.4 et GII.6 infectant les humains et précédemment impliqués dans des épisodes de gastroentérites. Le VHA était proche du géotype V infectant les simiens. L'analyse de la diversité a révélé que la fraction virale représentait 1,1 à 7,5% des *reads*, dont 68 à 87 % sont restés non attribués. Les séquences appartenant aux *Astroviridae* (16) étaient les plus abondantes, avec des génomes presque complets. Les séquences de *Picobirnaviridae* (14) étaient liées à des souches infectant les chauves-souris, les humains et d'autres hôtes. Les séquences d'*Hepeviridae* (04) étaient similaires à celles des souches détectées dans les éponges et aussi dans le porc. La plupart des séquences de *Caliciviridae* (08) et de *Picornaviridae* (11) étaient similaires à celles des souches infectant les chauves-souris, avec quelques-unes proches des NoV humains, du picornavirus ou du VHA appartenant au géotype V simien. Malgré le besoin d'améliorer la sensibilité de la méthode utilisée, cette étude a permis de décrire une grande diversité de séquences de virus à ARN humains et animaux présents dans les palourdes de la Sanaga. Ce qui met en évidence les risques sanitaires en couru par les consommateurs en l'absence d'un traitement adéquat.

**Mots clés :** Palourdes, virus entériques humains, virus de mammifères, NGS métagénomique.

## ABSTRACT

Many recent pandemics have been recognized as viral diseases. While their origins often remain unknown, the environment seems to play an important role in their emergence. Understanding the diversity of viruses present in food samples from the environment can help to understand the mechanisms of emergence and transmission of viral agents. This work describes the evaluation of the diversity of RNA viruses in Sanaga river clams, which constitute the main pathogenic viruses transmitted by food. Clams collected between February 2018 and February 2020 in Bolounga and Moulounga localities were tested by RT-qPCR for the presence of norovirus (NoV) and hepatitis A virus (HAV). Polymerase and capsid regions for NoV and junction VP1/2A for HAV of detected viruses were amplified, sequenced and genotyped using the Norovirus Typing Tool 2.0. Selected samples were analysed using a metagenomic approach in which a DNA library preparation steps included sequence capture (VirCapSeq-VERT) and bioinformatics processing with the removal of bacterial and host sequences, and deduplication prior to *de novo* assembly has made it possible to target vertebrate RNA viruses. About 80% of the samples were contaminated with at least one of the viruses, with NoV present in 69% of the samples against 36% for the HAV. We also observed a higher overall prevalence in Moulounga (87%) compared to Bolounga (72%). The detected NoV belonged to genotypes GI.1, GII.3, GII.4 and GII.6 infecting humans and previously identified in episodes of gastroenteritis. The HAV was like genotype V infecting non-human primates. Analysis of the diversity of viruses present showed that the viral fraction represented 1.1 to 7.5% of reads, of which 68 to 87% remained unassigned. *Contigs* assembled belonging to *Astroviridae* were the most abundant (16), with nearly complete genomes for bastroviruses. *Picobirnaviridae* sequences (14) were linked to strains infecting bats, humans or other hosts. *Hepeviridae* sequences (04) were mainly linked to strains detected in sponge and also in pigs. As for *Caliciviridae* (08) and *Picornaviridae* (11), most of the sequences were linked to strains infecting bats, with some sequences close to human NoV, picornavirus and HAV belonging to the genotype V infecting simian. Despite a need to improve the sensitivity the method used, this study made it possible to describe a great diversity of human and animal RNA virus sequences present in Sanaga clams. This highlights the health risks faced by consumers in the absence of adequate treatment.

**Key words:** Clams, human enteric viruses, mammalian viruses, metagenomic NGS.

## LISTE DES ABREVIATIONS ET ACRONYMES

<b>ADN</b>	: Acide Désoxy Nucléique
<b>ADNc</b>	: ADN complémentaire
<b>ADNdb/dsDNA</b>	: ADN double brin
<b>Ala</b>	: Alanine
<b>Arg</b>	: Arginine
<b>ARN</b>	: Acide RiboNucléique
<b>ARNr</b>	: ARN ribosomique
<b>BLAST</b>	: <i>Basic Local Alignment Search Tool</i>
<b>BLASTn</b>	: BLAST consistant à rechercher une correspondance de la séquence requête nucléotidique dans une base de données nucléotidiques
<b>BLASTx</b>	: BLAST consistant à rechercher une correspondance de la séquence requête nucléotidique dans une base de données protéiques
<b>Bootstrap</b>	: Nombre fois (pourcentage) où une branche apparait après un nombre fois ‘n’ de construction d’un arbre phylogénétique
<b>Contig</b>	: Séquence nucléotidique contigüe résultant de l’assemblage de courtes séquences ( <i>reads</i> ) issues du séquençage
<b>Ct (Cycle threshold)</b>	: Nombre de cycle de d’amplification (PCR) à l’issue duquel l’ADN cible est détecté
<b>DTT</b>	: Dithiothreitol
<b>dNTP</b>	: Di-Nucléotide Tri Phosphate
<b>EFSA</b>	: European Food Safety Agency
<b>E-value</b>	: Probabilité d’obtenir un résultat différent lors du BLAST dans une base données
<b>FAO</b>	: <i>Food and Agriculture Organization</i>
<b>Fasta</b>	: Format de fichier texte utilisé pour stocker des séquences biologiques de nature nucléique ou protéique
<b>Fastq</b>	: Format de fichier texte permettant de stocker à la fois des séquences biologiques (uniquement des séquences nucléiques) et les scores de qualité associés
<b>GenBank</b>	: Banque de séquences nucléotidiques contenue dans NCBI (National Center for Biotechnology Information)
<b>Gly</b>	: Glycine

<b>Gln</b>	:	Glutamine
<b>Glu</b>	:	Acide glutamique
<b>Heatmap</b>	:	Graphe permettant de représenter l'abondance relative de différentes familles virales dans un échantillon
<b>His</b>	:	Histidine
<b>IFREMER</b>	:	Institut Français de Recherche pour l'Exploitation de la Mer
<b>ISO</b>	:	<i>International Organization for Standardization</i>
<b>kb</b>	:	Kilo base
<b>kDa</b>	:	Kilo Dalton
<b>Krona</b>	:	Outil de visualisation qui permet une exploration intuitive des abondances relatives au sein des hiérarchies complexes des classifications métagénomiques
<b>Metavirome</b>	:	Ensemble des acides nucléiques viraux présents dans un échantillon
<b>NCBI</b>	:	National Center for Biotechnology Information
<b>NGS</b>	:	Next Generation Sequencing
<b>ORF</b>	:	<i>Open Reading Frame</i>
<b>Pb</b>	:	Paire de bases
<b>PCR</b>	:	<i>Polymerase Chain Reaction</i>
<b>Phred</b>	:	Score de qualité attribué à chaque séquence ( <i>read</i> ) par le séquenceur
<b>Phe</b>	:	Phénylalanine
<b>Read</b>	:	Courte séquence d'ADN issue du séquençage de seconde génération (Exemple Illumina)
<b>RefSeq</b>	:	Base de données des séquences de référence de NCBI
<b>RdRp</b>	:	ARN polymérase ARN dépendant
<b>RT-qPCR</b>	:	PCR quantitative reverse
<b>SARS-COV-2</b>	:	Coronavirus 2 du syndrome respiratoire aigu sévère
<b>Ser</b>	:	Serine
<b>SRAS</b>	:	Syndrome respiratoire aigu sévère
<b>ssRNA</b>	:	ARN simple brin
<b>TaxID</b>	:	Identifiant unique stable pour chaque taxon (pour une espèce, une famille, un ordre ou tout autre groupe) dans la base de données taxonomique du NCBI

<b>TCID<sub>50</sub></b>	:	<i>Median Tissue Culture Infectious Dose</i> (dilution d'une suspension virale nécessaire pour infecter 50% des puits d'une culture cellulaire)
<b>TOPO</b>	:	Kit de clonage
<b>Trp</b>	:	Tryptophane
<b>Tyr</b>	:	Tyrosine
<b>UTR</b>	:	<i>Untranslated Region</i> (Région Non Traduite)
<b>Val</b>	:	Valine
<b>VHA</b>	:	Virus de l'hépatite A
<b>VirCapSeq-</b>	:	Puce d'enrichissement permettant la capture des séquences de
<b>VERT</b>	:	génomés des virus des vertébrés
<b>VP</b>	:	Protéine virale
<b>VPg</b>	:	Protéine virale génomique
<b>WHO</b>	:	<i>World Health Organization</i> (Organisation Mondiale de la Santé)

## LISTE DES TABLEAUX

Tableau 1. Virus transmissibles via les aliments ou l'eau. Traduit de Greening et Cannon, 2016.....	9
Tableau 2. Estimation de la qualité des séquences obtenues à l'aide du score <i>Phred</i> . ....	45
Tableau 3. Composition du mélange réactionnel de la RT-qPCR .....	55
Tableau 4. Liste des différents couples d'amorces et des sondes utilisés pour la RT-qPCR.....	56
Tableau 5. Profil thermique de la RT-qPCR en une étape, utilisé pour l'efficacité d'extraction du mengovirus .....	56
Tableau 6. Profil thermique de la RT-qPCR en une étape, utilisé pour la détection/quantification des NoV GI, NoV GII, et le VHA .....	58
Tableau 7. Mélange réactionnel A de la rétrotranscription .....	59
Tableau 8. Liste des différents couples d'amorces et des sondes utilisés pour le typage .....	59
Tableau 9. Mélange réactionnel B de la rétrotranscription.....	60
Tableau 10. Mélange réactionnel de PCR classique (1 <sup>ère</sup> PCR).....	60
Tableau 11. Profil thermique de la PCR classique (1 <sup>ère</sup> PCR).....	60
Tableau 12. Mix de PCR semi-nichée (1/2 nichée) .....	61
Tableau 13. Tailles des différents produits de PCR attendus.....	61
Tableau 14. Mélange réactionnel I pour la synthèse des ADNc .....	64
Tableau 15. Mélange réactionnel II pour la synthèse des ADNc.....	65
Tableau 16. Mélange réactionnel pour la synthèse du second brin de l'ADNc.....	65
Tableau 17. Efficacité d'extraction, prévalence et concentrations moyennes géométriques des NoV et du VHA dans les échantillons de palourdes de la Sanaga. ....	71
Tableau 18. Concentrations des NoV et du VHA déterminées par RT-qPCR dans les palourdes la Sanaga. ....	72
Tableau 19. Séquences identifiées lors du génotypage des NoV et du VHA .....	74
Tableau 20. Résumé du nombre de <i>reads</i> et de contigs par échantillon à différentes étapes du <i>pipeline</i> du traitement bio-informatique. ....	82
Tableau 21. Quelques <i>contigs</i> d' <i>Astroviridae</i> identifiés dans les palourdes de la Sanaga.....	88
Tableau 22. Quelques <i>contigs</i> de <i>Caliciviridae</i> identifiées dans les palourdes Sanaga. ....	92
Tableau 23. Quelques <i>contigs</i> de <i>Picobirnaviridae</i> identifiées chez les palourdes Sanaga. ....	94
Tableau 24. Quelques <i>contigs</i> de <i>Picornaviridae</i> identifiées dans les palourdes de la Sanaga. ....	97
Tableau 25. Quelques <i>contigs</i> d' <i>Hepeviridae</i> identifiées dans les palourdes de la Sanaga. ....	99

## LISTE DES FIGURES

Figure 1. Anatomie interne du corps mou de la palourde du genre <i>Tapes</i> (Helm <i>et al.</i> , 2004).....	4
Figure 2. Synthèse du devenir des particules en suspension dans l'eau filtrée par le bivalve (Lassus <i>et al.</i> , 2002).....	5
Figure 3. Structure du génome des norovirus humains. (Alexis de Rougemont,.....	15
Figure 4. Classification phylogénétique des norovirus. ....	17
Figure 5. Structure du génome du virus de l'hépatite A.....	18
Figure 6. Structure du génome du virus de l'hépatite E (van Tong <i>et al.</i> , 2016). ....	19
Figure 7. Diversité génétique du virus de l'hépatite E (Debing <i>et al.</i> , 2016). ....	20
Figure 8. Structure du génome du sapovirus.....	21
Figure 9. Structure du génome du rotavirus.....	22
Figure 10. Structure du génome des astrovirus.....	23
Figure 11. Structure du génome des Entérovirus.....	24
Figure 12. Structure du génome de l'Aichivirus.....	25
Figure 13. Structure du génome du picobirnavirus (Ghosh et Malik, 2021). ....	26
Figure 14. Etapes de la préparation des échantillons selon l'approche Neto-Vir (Conceição-Neto <i>et al.</i> , 2015).....	35
Figure 15. Préparation des bibliothèques NGS par l'approche Nextera d'Illumina.....	39
Figure 16. Quelques technologies de séquençage de nouvelle génération appliquées à la recherche des virus dans les aliments (Desdouts <i>et al.</i> , 2020).....	41
Figure 17. Amplification des bibliothèques avant le séquençage de seconde génération. ....	43
Figure 18. Séquençage de troisième génération.....	44
Figure 19. Assemblage <i>de novo</i> des reads. ....	47
Figure 20. Localisation des points de prélèvement.....	53
Figure 21. <i>Pipeline</i> du traitement bio-informatique.....	68
Figure 22. Alignements de la séquence du génotype GI.1 de norovirus détecté dans les palourdes de la Sanaga avec la séquence partielle de la région codant la protéine VP1 de la capsid de la souche FJ383881.1.....	75
Figure 23. Alignements de la séquence du génotype GII.3 de norovirus détecté dans les palourdes de la Sanaga avec la séquence partielle de région codant la protéine majeure VP1 de la capsid de la souche MK764020.1. ....	76
Figure 24. Alignements de la séquence du génotype GII.6 de norovirus détecté dans les palourdes de la Sanaga avec la séquence partielle de la capsid de la souche MH260511.1. ....	76
Figure 25. Alignements de la séquence du génotype GII.4 de norovirus détecté dans les palourdes de la Sanaga avec la séquence partielle de la capsid de la souche MH218670.1. ....	77

<b>Figure 26. Alignements de la séquence de la souche du virus de l'hépatite A détecté dans les palourdes de la Sanaga (MT185676.1) avec la souche KP875241.1 détectée dans les eaux d'irrigation en Afrique du Sud.....</b>	<b>78</b>
<b>Figure 27. Arbre phylogénétique bayésien des séquences de virus de l'hépatite A.....</b>	<b>79</b>
<b>Figure 28. Abondance relative des <i>reads</i> par règne et par librairie.....</b>	<b>83</b>
<b>Figure 29. <i>Heatmap</i> de la diversité et de l'abondance relative virale. ....</b>	<b>85</b>
<b>Figure 30. Aperçu de la diversité des hôtes et des <i>reads</i> identifiés comme appartenant à quatre familles virales. ....</b>	<b>86</b>
<b>Figure 31. Arbres phylogénétiques de la famille des <i>Astroviridae</i> basés sur la protéine RdRp....</b>	<b>90</b>
<b>Figure 32. Arbre phylogénétique des séquences RdRp de picorbirnavirus.....</b>	<b>95</b>

## **LISTE DES PHOTOS**

Photo 1. <b>Echantillon de palourdes de la Sanaga</b> .....	<b>49</b>
Photo 2. <b>Extraction des tissus digestifs</b> .....	<b>54</b>
Photo 3. <b>Traitement des hydrolysats de TD aux ultrasons</b> .....	<b>63</b>
Photo 4. <b>Extraction et purification des AN à l'aide du système d'extraction semi-automatisé eGENE-UP™</b> .....	<b>64</b>

## INTRODUCTION GENERALE

De nombreuses pandémies humaines ont été identifiées comme ayant une source virale, et le plus souvent zoonotique. Si dans certains cas, l'origine a été liée à la modification des habitats fauniques naturels ou à la consommation d'aliments contaminés, plusieurs causes ou sources demeurent encore inconnues (Devaux *et al.*, 2019 ; Osterhauss *et al.*, 2020). Dans les différents scénarios envisagés, l'environnement semble jouer un rôle important car, il peut constituer le lien entre l'Homme et l'hôte originel. Ainsi, connaître la diversité virale dans des échantillons alimentaires issus de l'environnement peut constituer la première étape dans la compréhension des voies de transmissions zoonotiques et aider à leur prévention (Nieuwenhuijse *et al.*, 2020 ; Osterhauss *et al.*, 2020).

La demande en mollusques bivalves comestibles augmente dans le monde compte tenu de leur valeur nutritive et gustative (Smaal *et al.*, 2019). Se développant généralement sans l'intervention humaine, ils constituent une source substantielle de nutriments pour de nombreuses populations (Venugopal et Gopakumar, 2018). Cependant, ils sont identifiés depuis de nombreuses années comme des aliments à haut risque (Metcalf *et al.*, 1995). En effet, pour leurs activités physiologiques, ils filtrent de grands volumes d'eau, et peuvent concentrer divers types de contaminants issus de leur environnement, notamment des virus potentiellement dangereux pour les consommateurs. De nombreux virus entériques peuvent être détectés dans les tissus de ces organismes après leur exposition aux eaux usées humaines. Et lors d'épidémies, une grande diversité virale peut y être identifiée (Le Guyader *et al.*, 2008 ; Grodzki, *et al.*, 2012 ; Strubbia *et al.*, 2019). Dans certains cas, la grande diversité virale des mollusques bivalves rend difficile l'identification de l'agent étiologique exact, conduisant à un diagnostic erroné si l'approche de détection est sélective, d'où la nécessité d'une approche holistique de détection (Nieuwenhuijse et Koopmans, 2017).

Ces dernières années, le séquençage métagénomique a été de plus en plus utilisé pour l'étude des populations microbiennes dans différents types d'échantillons, dont les échantillons environnementaux. Cette approche permet la description de toutes les séquences d'acides nucléiques présentes dans un échantillon, et donc en théorie, permet l'identification de tous les organismes, y compris les virus (Dance, 2021). En tant qu'approche agnostique, il est très prometteur dans la surveillance et la détermination de la diversité virale dans les aliments et les échantillons environnementaux tels que les coquillages et les eaux usées (Desdouits *et al.*, 2020a ; Nieuwenhuijse *et al.*, 2020). Le séquençage métagénomique a été appliqué pour estimer

la diversité virale dans les eaux de différents océans et les eaux usées à travers le monde, montrant ainsi son potentiel dans la description des communautés virales (Desdouits *et al.*, 2020a ; Nieuwenhuijse *et al.*, 2020 ; Sunagawa *et al.*, 2020). Quelques études de métagénomique virale des mollusques bivalves ont été menées dans certains pays. Celles-ci se sont généralement intéressées à une seule famille de virus ayant un intérêt en santé humaine (Andrade *et al.*, 2015 ; Strubbia *et al.*, 2020, Desdouits *et al.*, 2020b) ou dans l'élevage ostréicole (Rosani et Gerdol, 2017 ; Kim *et al.*, 2020). Dès lors, il serait important d'étendre l'étude aux mollusques bivalves d'autres régions du monde tout en élargissant la diversité des virus recherchés.

Dans le Golf de Guinée, la palourde d'eau douce *Galatea paradoxa* (Born, 1778) présente dans plusieurs fleuves dont la Sanaga, constitue une source importante de protéines pour les populations locales ainsi qu'une ressource économique importante (Ajonina *et al.*, 2005). Malgré cette importance économique et nutritionnelle, cette palourde appelée *Behona* ou *Bissonda* par les populations locales est susceptible comme tous les mollusques bivalves d'accumuler divers contaminants, dont les virus présents dans leur environnement. Situé à quelques kilomètres de l'embouchure du fleuve Sanaga, leur gisement se trouve dans un parc national naturel riche en flore et faune sauvages (Ajonina *et al.*, 2005) et peuplé de plusieurs milliers d'âmes. Cette proximité suggère un apport en matières fécales aussi bien humaines qu'animales, et par conséquent, la contamination des eaux et des palourdes qui les filtrent par des virus pathogènes humains mais également par des virus animaux potentiellement zoonotiques. Ceci pourrait représenter un danger pour les populations riveraines qui consomment ces palourdes. Afin de prévoir et d'éviter des infections virales et l'émergence des zoonoses, il nous a donc semblé important de connaître la diversité des virus contaminant ces bivalves, ainsi que d'envisager leur utilisation comme sentinelles de la contamination environnementale.

## **Hypothèse**

Les palourdes de la Sanaga concentrent une grande diversité de virus à ARN humains et animaux.

## **Objectif général**

Evaluer la diversité des virus à ARN contaminant les palourdes de la Sanaga.

## **Objectifs spécifiques**

- Détecter, quantifier et géotyper les norovirus et le virus de l'hépatite A contaminant les palourdes de la Sanaga ;
- Déterminer la diversité des virus à ARN présents dans les palourdes de la Sanaga ;
- Identifier les virus à ARN constituant un danger chez les humains.

# CHAPITRE I : REVUE DE LA LITTERATURE

## 1.1 Généralités sur les mollusques bivalves

### 1.1.1 Définition

Les Bivalves ou Lamellibranches sont une classe constituée d'animaux comprimés latéralement dont les parties molles du corps sont partiellement ou intégralement couvertes par une coquille. Celle-ci est composée de deux valves liées par une charnière, d'où leur nom de Bivalve. Leurs branchies sont des organes bien développés qui assurent une double fonction : l'alimentation et la respiration. Certaines espèces vivent fixées sur les fonds grâce à un faisceau de filaments (le byssus) sécrétés par leur pied, d'autres sont mobiles. Toutes sont aquatiques, et la plupart marine (Helm *et al.*, 2004 ; Arapov *et al.*, 2010).

### 1.1.2 Mode de nutrition

Les bivalves sont des filtreurs des milieux aquatiques qui se nourrissent essentiellement de phytoplancton. Chez les juvéniles et les adultes, les cténidies ou branchies sont bien développés et assurent un double rôle : la nutrition et la respiration. Ces derniers sont recouverts de cils, filaments fins vibratiles, qui ensemble coordonnent les contractions induisant un courant d'eau. L'eau est aspirée par l'animal à travers l'ouverture ou siphon inhalant passant par les branchies et est refoulée dans le milieu environnant par le siphon exhalant. Les branchies captent le plancton ou autres particules de matière organique et l'enrobent dans un mucus. Des particules de nourriture chargées de mucus passent d'abord grâce à l'action des cils à travers des canaux spéciaux des filaments branchiaux jusqu'aux palpes labiaux dont le rôle est de diriger la nourriture vers la bouche. Les bivalves peuvent exercer une certaine sélection de leur nourriture, et périodiquement, les palpes rejettent de petites quantités de nourriture sous forme de pseudo-fèces qui sont expulsés à partir de la cavité du manteau, à la suite des battements vigoureux des deux valves (Arapov *et al.*, 2010 ; Rosa *et al.*, 2018).

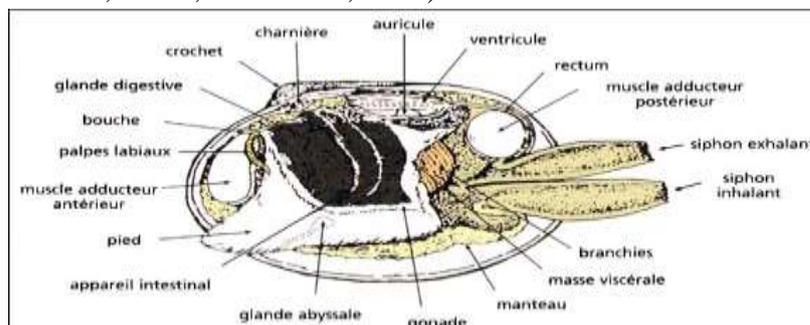
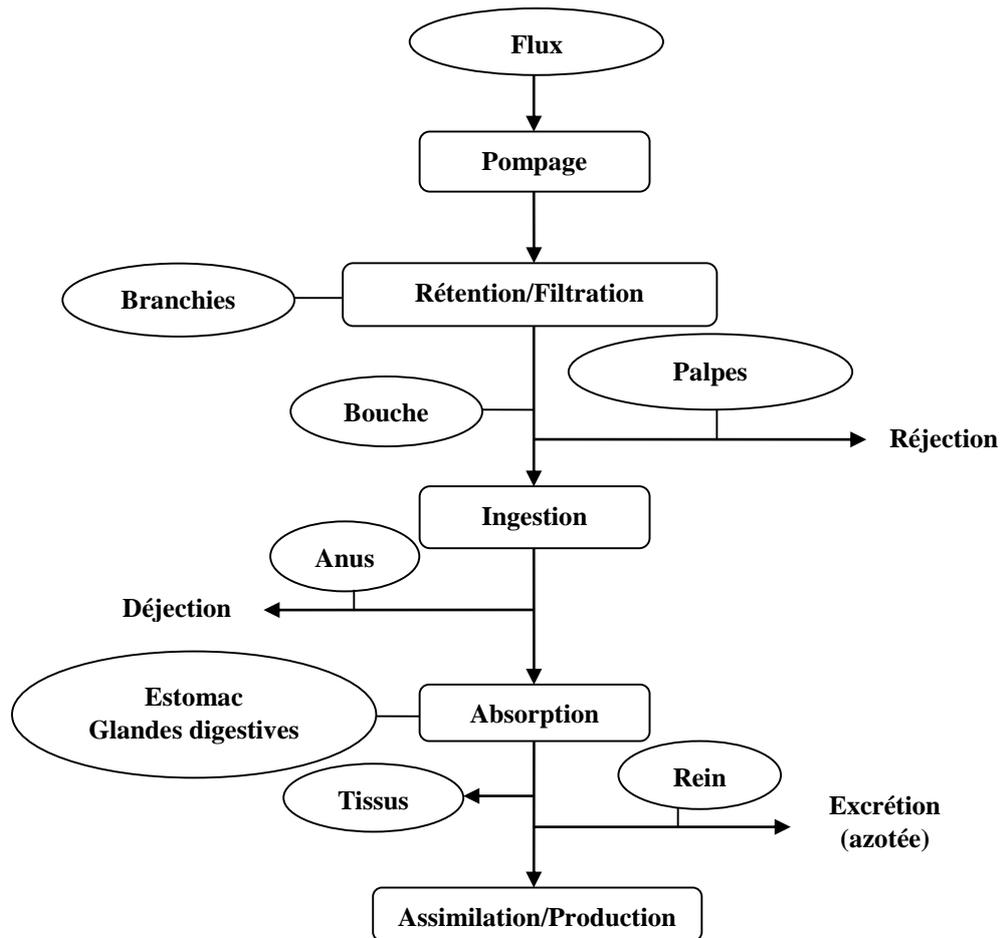


Figure 1. Anatomie interne du corps mou de la palourde du genre *Tapes* (Helm *et al.*, 2004).



**Figure 2. Synthèse du devenir des particules en suspension dans l'eau filtrée par le bivalve (Lassus *et al.*, 2002).**

## 1.2 Importance économique et nutritionnelle des mollusques bivalves

Sur le plan nutritionnel, l'intérêt envers les mollusques bivalves a récemment augmenté en raison de leurs qualités nutritionnelles. En effet, les coquillages et les mollusques bivalves en particulier sont connus pour leurs quantités appréciables de protéines digestibles, d'acides aminés essentiels, d'acides gras polyinsaturés à longue chaîne, de peptides bioactifs, de caroténoïdes, de vitamine B12 et autres vitamines, de minéraux, y compris le cuivre, le zinc, les phosphates inorganiques, le sodium, le potassium, le sélénium, l'iode, et aussi d'autres nutriments qui présentent des effets bénéfiques sur la santé des consommateurs (Venugopal et Gopakumar, 2018). Leur consommation de façon modérée est recommandée chez la femme enceinte pour améliorer le développement neurologique du fœtus durant la gestation et réduire les risques de survenue de maladies cardiovasculaires (EFSA, 2014).

La conchyliculture ou culture des mollusques bivalves est une composante majeure de l'aquaculture mondiale et revêt une importance économique sans cesse croissante. En effet, elle constitue 14% de la production marine globale. Entre 2010 et 2015, la production de mollusques bivalves destinée à la consommation humaine était estimée à 15 millions de tonnes par an, dont 89% provenant de l'élevage, et seulement 11% de gisements naturels. L'Asie et la Chine en particulier sont responsables de 85% de la production mondiale estimée à 23 milliards de dollars en 2015 (Smaal *et al.*, 2019).

La palourde *Galatea paradoxa* (Born, 1778) est une palourde d'eau douce rencontrée dans différents fleuves du Golf de Guinée dont la Volta au Ghana, la Cross River au Nigéria et la Sanaga au Cameroun (Etim et Brey, 1994). Au Cameroun, elle est récoltée dans le canton Malimba à 12 km de l'embouchure de la Sanaga, dans l'arrondissement de Mouanko, département de la Sanaga Maritime, région du Littoral où elle est appelée *Behona* ou *Bissonda* par les populations locales pour lesquelles elle constitue la principale source de protéines et de revenus. En effet, sa production estimée en 2005 à 8237 tonnes par an est au centre d'une importante activité économique qui génère environ 865000 Dollars US par an. Outre la partie molle commercialisée pour l'alimentation humaine, la coquille constitue une matière première des industries de production de provende et de céramique dans la ville de Douala, augmentant ainsi la valeur ajoutée de la coquille initialement considérée comme déchet (Ajonina *et al.*, 2005).

Leur classification est la suivante (MolluscaBase eds., 2022) :

- **Règne** : Animalia
- **Embranchement** : Mollusca
- **Classe** : Bivalvia
- **Sous-classe** : Autobranchia
- **Infraclasse** : Heteroconchia
- **Superordre** : Imparidentia
- **Ordre** : Cardiida
- **Superfamille** : Tellinoidea
- **Famille** : Donacidae
- **Genre** : *Galatea*
- **Espèce** : *Galatea paradoxa* (Born, 1778)

## **1.3 Les virus d'origine alimentaire**

### **1.3.1 Généralités sur les virus**

Les virus sont des parasites intracellulaires obligatoires nécessitant une cellule hôte dont ils utilisent le métabolisme et les constituants pour se répliquer. Encore appelée virion, la particule virale extracellulaire est caractérisée par sa structure qui est constituée d'un matériel génétique (acide nucléique ARN ou ADN simple brin ou double brin) entouré d'une coque protéique protectrice appelée capsid (Bosch *et al.*, 2018) ; l'ensemble acide nucléique-capsid formant une entité appelée nucléocapsid. Chez de nombreux virus, la nucléocapsid est entourée d'une enveloppe lipoglycoprotéique dérivée de la bicouche lipidique de la cellule hôte à laquelle sont associées des glycoprotéines virales. La présence ou l'absence de cette dernière permet de discriminer les virus en deux groupes : les virus enveloppés et les virus nus.

La nature et la structure du génome, la taille et la symétrie de la capsid, la présence ou l'absence d'enveloppe, ainsi que de leur mode de répllication, constituent les principaux éléments permettant la classification des virus (King *et al.*, 2012).

### **1.3.2 Les virus d'origine alimentaire**

Les virus peuvent être transmis de différentes façons : par voie respiratoire, à travers des gouttelettes (aérosols) générées par une personne infectée qui tousse, par voie féco-orale à travers l'ingestion d'aliments souillés par des rejets d'une personne infectée ou par les mains sales ou autres contacts avec des surfaces contaminées, par voie sexuelle, par contact avec des produits sanguins contaminés, par contact avec des animaux infectés (virus zoonotiques) ou par des vecteurs tels que les arthropodes (FAO et WHO, 2008 ; Louten, 2016).

Les virus dits « d'origine alimentaire » sont des virus qui présentent une phase de multiplication dans le tractus gastro-intestinal de leur hôte et qui sont par la suite excrétés à travers les fèces et/ou les vomis (Bosch *et al.*, 2018). Environ 26 familles de virus sont aujourd'hui connues pour leur capacité à infecter l'Homme (Siegel, 2018). Parmi ceux-ci, une dizaine peut se transmettre par voie alimentaire (Tableau 1) (Koopmans, 2012). On note particulièrement les norovirus, les sapovirus, les entérovirus, les adénovirus, les virus des hépatites A et E qui sont excrétés dans les fèces humaines et peuvent être à l'origine d'épidémies après la consommation d'aliments contaminés. Les maladies causées par ces derniers peuvent être regroupées en trois principaux types : les gastroentérites, les hépatites transmises par voie entérique, et les maladies affectant d'autres parties du corps à savoir les

yeux, le système respiratoire, le système nerveux central, et conduisant à la conjonctivite, la poliomyélite, la méningite et l'encéphalite (Greening et Cannon, 2016).

A l'exception des adénovirus dont le matériel génétique est constitué d'une molécule d'ADN, les virus d'origine alimentaire ou virus entériques ont un génome généralement constitué d'une ou plusieurs molécules d'ARN. Leur génome est protégé par une capsid non enveloppée : ce qui leur confère une stabilité aussi bien dans l'environnement et les aliments que dans le tractus gastro-intestinal (FAO et WHO, 2008 ; Greening et Cannon, 2016).

**Tableau 1. Virus transmissibles via les aliments ou l'eau. Traduit de Greening et Cannon, 2016.**

Genre ou espèce	Famille	Type d'acide nucléique	Enveloppe	Morphologie/Symétrie	Taille du virion (nm)	Cultivable <sup>a</sup>	Taille du génome (Kb)	Maladie
Adenovirus	<i>Adenoviridae</i>	ADNdb (+)	N	Icosahédrique	80-110	O <sup>a</sup>	28-45	Infections respiratoires, oculaires, et gastroentérite
Aichivirus	<i>Picornaviridae</i>	ARNsb (+)	N	Icosahédrique	20-40	O	8,2	Gastroentérite
Astrovirus	<i>Astroviridae</i>	ARNsb (+)	N	Icosahédrique	28-30	O <sup>a</sup>	6,8-7,8	Gastroentérite
Coronavirus	<i>Coronaviridae</i>	ARNsb (+)	O	Hélicoïdale	80-220	O <sup>a</sup>	27-32	Gastroentérite, infections respiratoires
Norovirus	<i>Caliciviridae</i>	ARNsb (+)	N	Icosahédrique	28-35	N	7,4-7,7	Gastroentérite épidémique
Sapovirus	<i>Caliciviridae</i>	ARNsb (+)	N	Icosahédrique	30-38	N	7,1-7,7	Gastroentérite
Hepatovirus : Virus de l'hépatite A	<i>Picornaviridae</i>	ARNsb (+)	N	Icosahédrique	27-32	O <sup>a</sup>	7,5	Inflammation du foie, hépatite
Hepevirus : Virus de l'hépatite E	<i>Hepaviridae</i>	ARNsb (+)	N	Icosahédrique	27-34	O	7,5	Inflammation du foie, hépatite
Rotavirus	<i>Reoviridae</i>	ARNdb (+), segmenté	N	Icosahédrique	60-80	O	16-27	Gastroentérite
Enterovirus	<i>Picornaviridae</i>	ARNdb (+)	N	Icosahédrique	28-30	O <sup>a</sup>	7,2-8,4	Poliomyélite, infections respiratoires, méningite, encéphalyte
Parvovirus	<i>Parvoviridae</i>	ADNsb (+)	N	Icosahédrique	18-26	N	5	Gastroentérite
Picobirnavirus	<i>Picobirnaviridae</i>	ARNdb (+), segmenté	N	Icosahédrique	33-41	N	3,8-4,5	Gastroentérite chez les animaux et ? les humains
Tick-borne encephalitis virus	<i>Flaviviridae</i>	ARNsb (+)	O	Icosahédrique	45-60	O	9,5-12,5	Tick-borne encéphalyte via le lait
Torovirus	<i>Coronaviridae</i>	ARNsb (+)	O	Hélicoïdale	100-150	O	27-32	Gastroentérite chez les animaux et ? les humains

**sb** : simple brin ; **db** : double brin; (+) : polarité positive ; **N**: non ; **O** : oui ; <sup>a</sup>: Toutes les souches ne sont pas cultivables

### 1.3.3 Modes de contamination des aliments

La contamination des aliments par des virus peut se produire :

- lors de la manipulation des aliments par des personnes infectées et ne respectant pas les règles d'hygiène ;
- lors de la production de l'aliment par contact avec des eaux usées ;
- et dans le cas des contamination zoonotique qui est plus rare et n'a été clairement objectivé que pour le virus de l'hépatite E (FAO et WHO, 2008 ; Bosch *et al.*, 2018 ; Desdouits *et al.*, 2020a).

#### a. Contamination lors de la manipulation des denrées alimentaires par des personnes infectées

Les aliments peuvent être contaminés par une personne en charge de leur production ou préparation et excréant des virus à travers les selles et/ou les vomissements. L'impact potentiel de tels événements de contamination dépend du type de produit et de sa préparation (Koopmans, 2012). Chez les personnes infectées par un virus d'origine alimentaire ou un virus entéro-pathogène, on observe généralement des titres viraux très élevés dans les selles allant jusqu'à  $10^{13}$  et  $10^{10}$  chez des patients souffrant respectivement de gastro-entérite ou d'hépatite (Bosch *et al.*, 2018). Par ailleurs, une récente étude a mis en évidence des titres variants de  $10^7$  à  $10^9$  GEq/mL (Equivalent génomique par mL) dans les vomissements de patients atteints de gastroentérites dues aux norovirus (Hagbom *et al.*, 2021). Les personnes infectées peuvent commencer à excréter des virus 12 heures après exposition comme observé avec les norovirus (WHO, 2008). L'excrétion peut se poursuivre jusqu'à plusieurs semaines selon le type de virus. Il est possible que des personnes infectées excrètent des virus avant de développer des symptômes et après disparition des signes cliniques. De plus, les infections asymptomatiques sont assez communes (WHO, 2008 ; Teunis *et al.*, 2015). La contamination virale résultant de la manipulation humaine peut survenir à n'importe quel stade du continuum ferme-assiette. Par exemple, les produits peuvent être manipulés par des mains pendant la récolte, l'emballage, la distribution et dans les commerces de détail, les restaurants ou à domicile (WHO, 2008 ; Koopmans, 2012).

Plusieurs épisodes d'infections virales d'origine alimentaire causées par les manipulateurs d'aliments ont été reportées à travers le monde cette dernière décennie. Les virus

majoritairement impliqués étant les norovirus et le virus de l'hépatite A liés à la consommation de fruits frais, des jus de fruits, des glaces, de la viande, des sushis contaminés par des manipulateurs infectés et asymptomatiques (Franck *et al.*, 2015 ; Chen *et al.*, 2016 ; Hernández *et al.*, 2019 ; Somura *et al.*, 2019 ; Kurup *et al.*, 2020).

**b. La contamination à la source lors de la production de l'aliment par contact avec des eaux usées**

La contamination à la source se produit lorsque les aliments sont contaminés pendant la production primaire (Koopmans, 2012), comme cela a été observé dans les produits frais tels que les baies, les oignons verts, la laitue, et les concombres à la suite de l'utilisation d'eaux usées contaminées pour l'irrigation agricole et le lavage des denrées. Cette voie de contamination est la plus fréquemment rapportée pour les mollusques bivalves (Fiorito *et al.*, 2019 ; Shin *et al.*, 2019 ; Werneck *et al.*, 2019 ; Bozhurt *et al.*, 2021, Zhang *et al.*, 2021). La contamination peut varier considérablement selon l'emplacement de la zone de production et la nature de la contamination des eaux usées. Les norovirus et le virus de l'hépatite A sont considérés comme des préoccupations prioritaires au regard du nombre d'épidémies identifiées et de personnes affectées (WHO, 2008 ; Bosch *et al.*, 2018 ; Velebit *et al.*, 2019 ; Desdouits *et al.*, 2021).

Contrairement aux contaminations associées aux manipulateurs d'aliments, les événements de contamination à la source peuvent impliquer plusieurs agents pathogènes qui peuvent être présents dans les eaux usées, y compris les virus animaux. Cette exposition simultanée à des mélanges de virus augmente la probabilité de recombinaison ou de réassortiment de génomes viraux lorsqu'une personne est infectée simultanément par plusieurs virus apparentés (Koopmans, 2012).

Depuis 2020, la COVID-19 qui sévit actuellement s'est rapidement propagé dans le monde entier pour devenir la crise sanitaire mondiale la plus importante depuis la pandémie de grippe de 1918 (Ceylan *et al.*, 2020). Même si les preuves indiquent une transmission du virus via des gouttelettes et aérosols infectés (OMS, 2020), sa présence dans les selles des patients infectés symptomatiques ou non, et par conséquent dans les eaux usées a conduit à divers travaux de surveillance de la contamination des environnements aquatiques en utilisant les mollusques bivalves comme sentinelles (Ali *et al.*, 2021 ; Dhama *et al.*, 2021 ; Desdouits *et al.*, 2021 ; Le Guernic *et al.*, 2021 ; Mateus *et al.*, 2021 ; Polo *et al.*, 2021 ; Sbaoui *et al.*, 2021 ; Wurtzer *et al.*, 2021).

### **c. Contamination des virus par voie zoonotique**

Une infection zoonotique d'origine alimentaire a lieu lorsque la chair, les organes ou tout autre produit issu d'un animal infecté est consommé (Koopmans, 2012). Contrairement aux bactéries (*Salmonella*, *Campylobacter*, etc.), la transmission alimentaire des virus zoonotiques est un événement rare, à l'exception du virus de l'hépatite E principalement transmis par cette voie (FAO et WHO, 2008).

Bien que ce mode de transmission soit moins courant chez les virus, le potentiel d'une telle transmission demeure préoccupant à chaque nouvelle épidémie (Koopmans, 2012). En effet, il existe aujourd'hui des évidences que le SRAS (Syndrome Respiratoire Aigu Sévère), le virus du monkeypox (varioloïde du singe, ou orthopoxvirose simienne), et le Nipah virus ont été initialement transmis à travers des incidents alimentaires (Koopmans, 2012).

## **1.4 Les principaux virus d'origine alimentaire transmis par les bivalves**

Malgré leur importance économique et nutritionnelle, les bivalves sont considérés comme des aliments à risques. En effet, du fait de leur mode d'alimentation et de respiration par filtration d'importants volumes d'eau, ces organismes sont capables de concentrer différents types de pathogènes de leur environnement, parmi lesquels les bactéries et les virus (Hatha *et al.*, 2005 ; Iwamoto *et al.*, 2010 ; Macaluso *et al.*, 2021), avec pour conséquence la survenue d'épidémies liées à la consommation de bivalves contaminés (Le Guyader et Pommepuy, 2002; Iwamoto *et al.*, 2010 ; Yu, *et al.*, 2016 ; Hardstaff *et al.*, 2018 ; Savini *et al.*, 2021).

Les virus à transmission alimentaire sont aujourd'hui considérés comme un réel problème émergent de sécurité sanitaire des aliments et de santé publique en général (Souza *et al.*, 2017). En effet, selon l'Autorité Européenne de la sécurité des Aliments (EFSA), les virus sont devenus pour la première fois en 2014 les principaux agents étiologiques responsables de 20% des épidémies transmises par voie alimentaire (EFSA, 2015). Ceci à cause de leur excrétion importante par les patients, leur persistance dans l'environnement due à leur extraordinaire stabilité en dehors de l'hôte, et leurs faibles doses infectantes (Romalde *et al.*, 2017).

Le premier lien entre les virus et les gastroentérites associées aux coquillages a été fait à l'été 1976-1977 au Royaume Uni. Aujourd'hui, les preuves épidémiologiques rassemblées suggèrent que les virus entériques humains sont aujourd'hui les pathogènes les plus communément transmis par les bivalves. En effet, par le passé, les mollusques bivalves étaient

majoritairement associés aux fièvres typhoïde et paratyphoïde. Celles-ci sont désormais rares dans les pays développés et sont remplacées par les gastro-entérites et les hépatites causées respectivement par les norovirus et le virus de l'hépatite A qui constituent les principaux agents infectieux transmis par les mollusques bivalves (Commission Européenne, 2017). Si ceci peut s'expliquer par la survie prolongée de ces virus non enveloppés dans l'environnement aquatique, et plus encore à l'intérieur des coquillages (Potasman *et al.*, 2002 ; Le Guyader *et al.*, 2012 ; McLeod *et al.*, 2017), plusieurs réglementations mises en place au niveau européen et américain justifient la réduction des infections bactériennes au profit des infections virales. Celles-ci portent principalement sur deux aspects :

– L'évaluation des sources et types de contaminations fécales (humaine et animale) à proximité des zones de production, associée à un suivi microbiologique basé sur l'utilisation de microorganismes indicateurs à l'instar d'*Escherichia coli* dans l'Union Européenne. Ceci fournit une évaluation du risque de contamination par des agents pathogènes bactériens (Commission Européenne, 2017).

– Le classement des zones de production en fonction du niveau de contamination de la chair (Union Européenne) ou de l'eau des sites de production (Etats Unis) qui détermine si les zones peuvent être utilisées pour la récolte et quel niveau de traitement post-récolte est nécessaire pour réduire le risque à un niveau considéré comme acceptable. Ainsi, trois classes ont été définies, à savoir :

- la classe A de faible contamination (80 % d'échantillons < 230 *Escherichia coli*/g de chair et de liquide intervalvaire, 20 % n'excédant pas 700 *E. coli*/g) ne nécessitant pas de traitement post-récolte,
- la classe B de contamination intermédiaire (90% d'échantillons < 4 600 NPP *E. coli*/100g de chair et intra-valvulaire liquide, 10 % d'échantillons excédant pas 46 000 NPP d'*E. coli*/100g de chair et intra-valvulaire liquide) nécessitant une purification, un reparcage dans une zone de meilleure qualité, ou une cuisson adéquate,
- la classe C fortement contaminée n'excédant pas 46 000 NPP d'*E. coli*/100 g de chair et intra-valvulaire liquide) nécessitant un reparcage ou une cuisson appropriée (Commission Européenne, 2017).

Si la purification qui est le traitement post-récolte le plus couramment utilisé a montré une efficacité dans l'amélioration de la qualité bactérienne des mollusques bivalves, réduisant ainsi la survenue des infections bactériennes, elle demeure toutefois peu efficace dans la

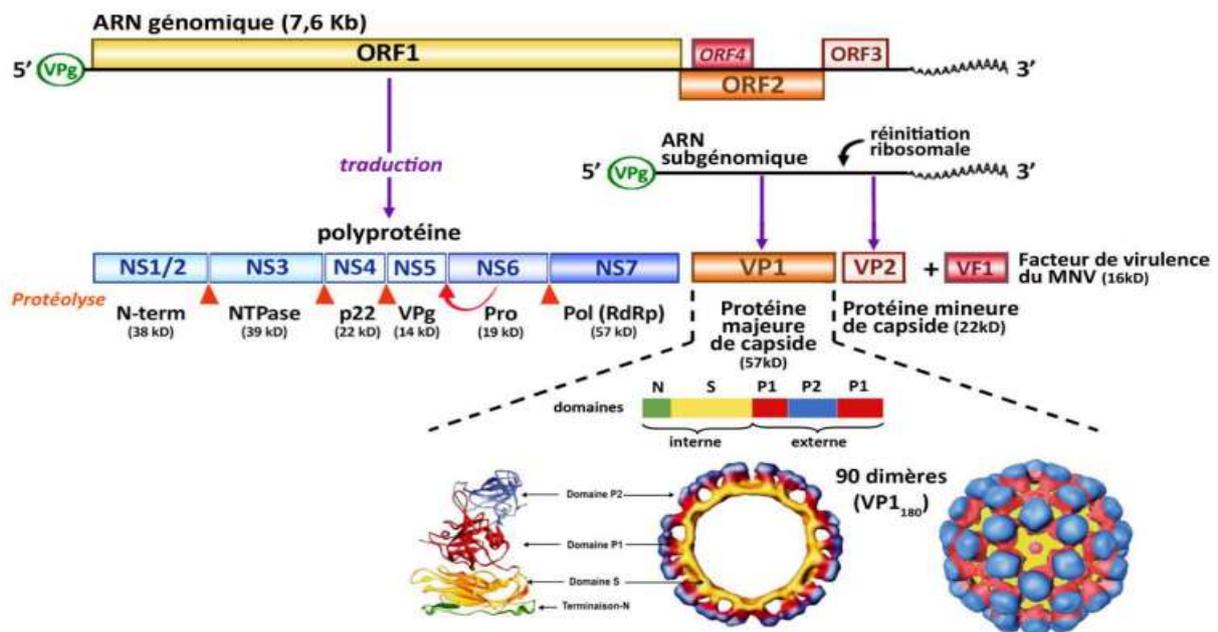
réduction des contaminants viraux, justifiant ainsi la persistance des infections virales (Le Guyader *et al.*, 2012 ; McLeod *et al.*, 2017).

### 1.4.1 Les norovirus

Les norovirus sont des virus non-enveloppés de taille comprise entre 28-35 nm appartenant à la famille des *Caliciviridae* au sein de laquelle ils constituent un groupe très diversifié. Leur génome est constitué d'une molécule d'ARN linéaire simple brin, de polarité positive et de taille comprise entre 7,4-7,7 kb, et flanquée d'une *UTR* (*Untranslated Region* : Région Non-Traduite) à l'extrémité 5' précédée d'une protéine génomique (VPg), et d'une *UTR* à l'extrémité 3' suivie d'une queue polyadénylée (poly(A)). Il est constitué de trois *ORF* (*Open Reading Frame* : Cadre Ouvert de Lecture) :

- l'ORF1 code une poly-protéine clivée en protéines non structurales (NS) impliquées dans la réplication : NS1/2 (N-term), NS3 (NTPase), NS4 (3A-like), NS5 (VPg), NS6 (protéase) et NS7 (polymérase) ;
- l'ORF2 code la principale protéine de capsid, VP1 ;
- l'ORF3 code la protéine mineure de la capsid, VP2, importante pour la stabilisation de la capsid et l'entrée virale (figure 3) (Thorne et Goodfellow, 2014 ; de Graaf *et al.*, 2016 ; Ford-Siltz *et al.*, 2019).

La capsid icosaédrique est constituée de 90 dimères de VP1, et se compose d'un domaine interne conservé qui constitue le noyau interne de la capsid, et d'un domaine externe hautement variable contenant des épitopes et des motifs protecteurs qui interagissent avec des facteurs d'attachement, tels que les antigènes des groupes d'histocompatibilité sanguins (HBGA), important pour initialiser l'infection virale (de Graaf *et al.*, 2016). Les protéines VP1 et VP2 de la capsid sont exprimées à partir d'un ARN sous-génomique qui est synthétisé pendant la réplication et qui comprend l'ORF2, l'ORF3, la région 3'-non traduite (3'-*UTR*), et la queue poly(A) terminale (figure 3) (Thorne et Goodfellow, 2014 ; Ford-Siltz *et al.*, 2019).

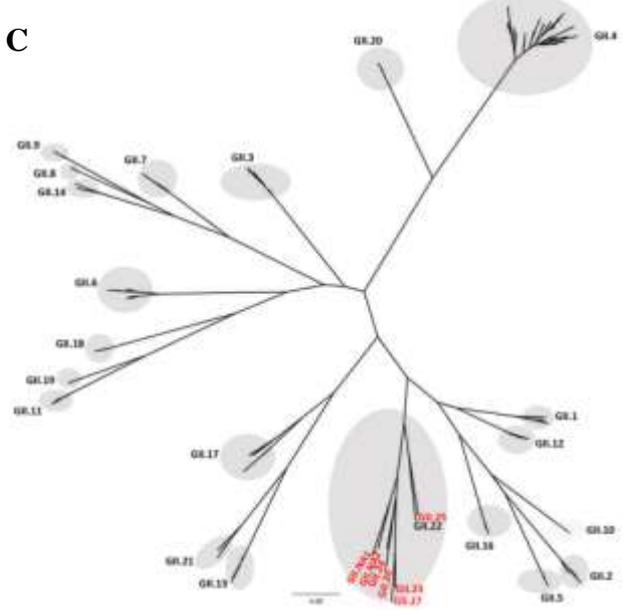
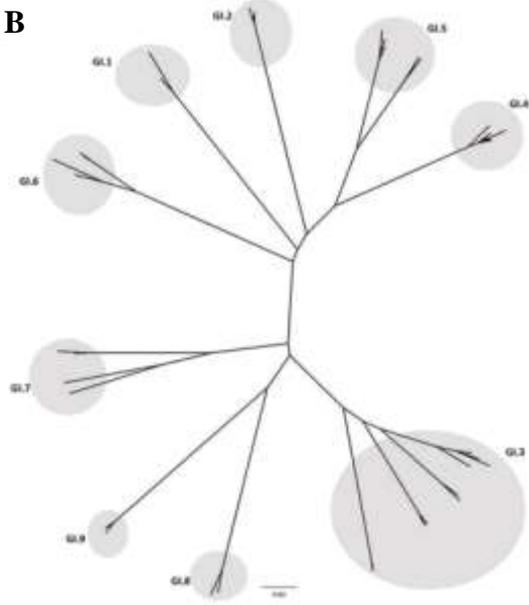
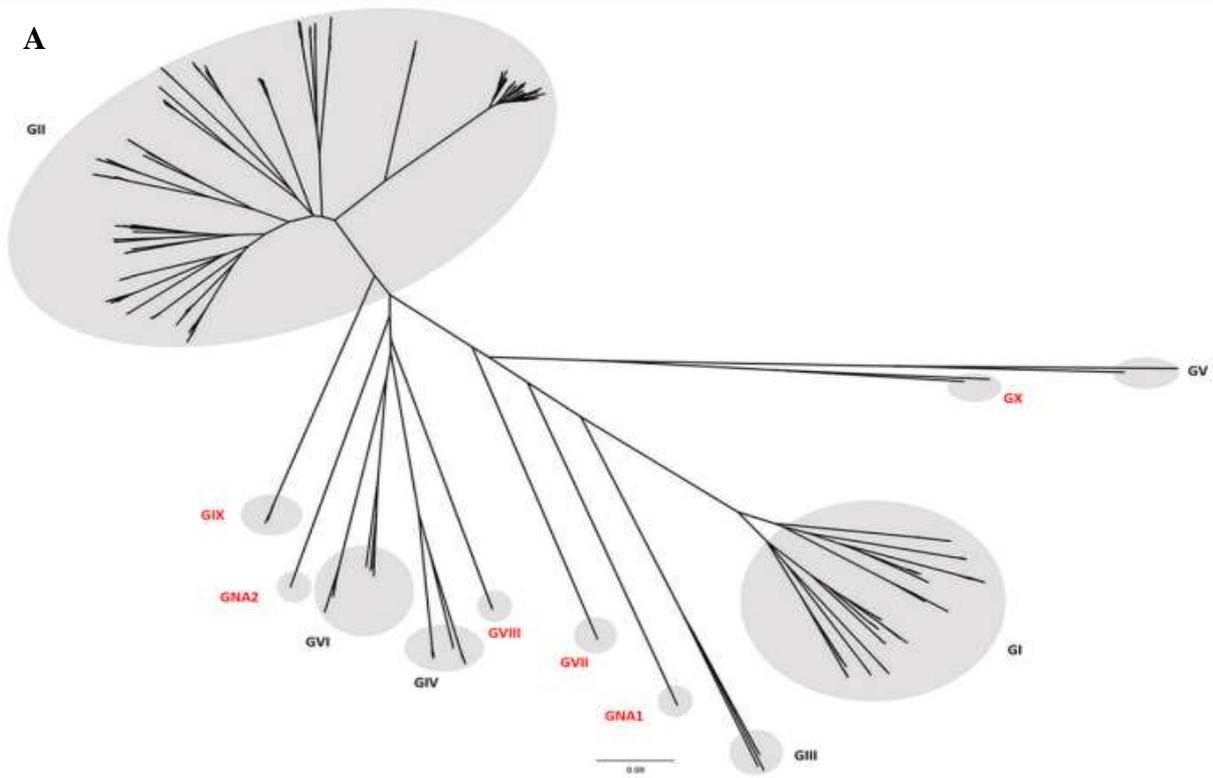


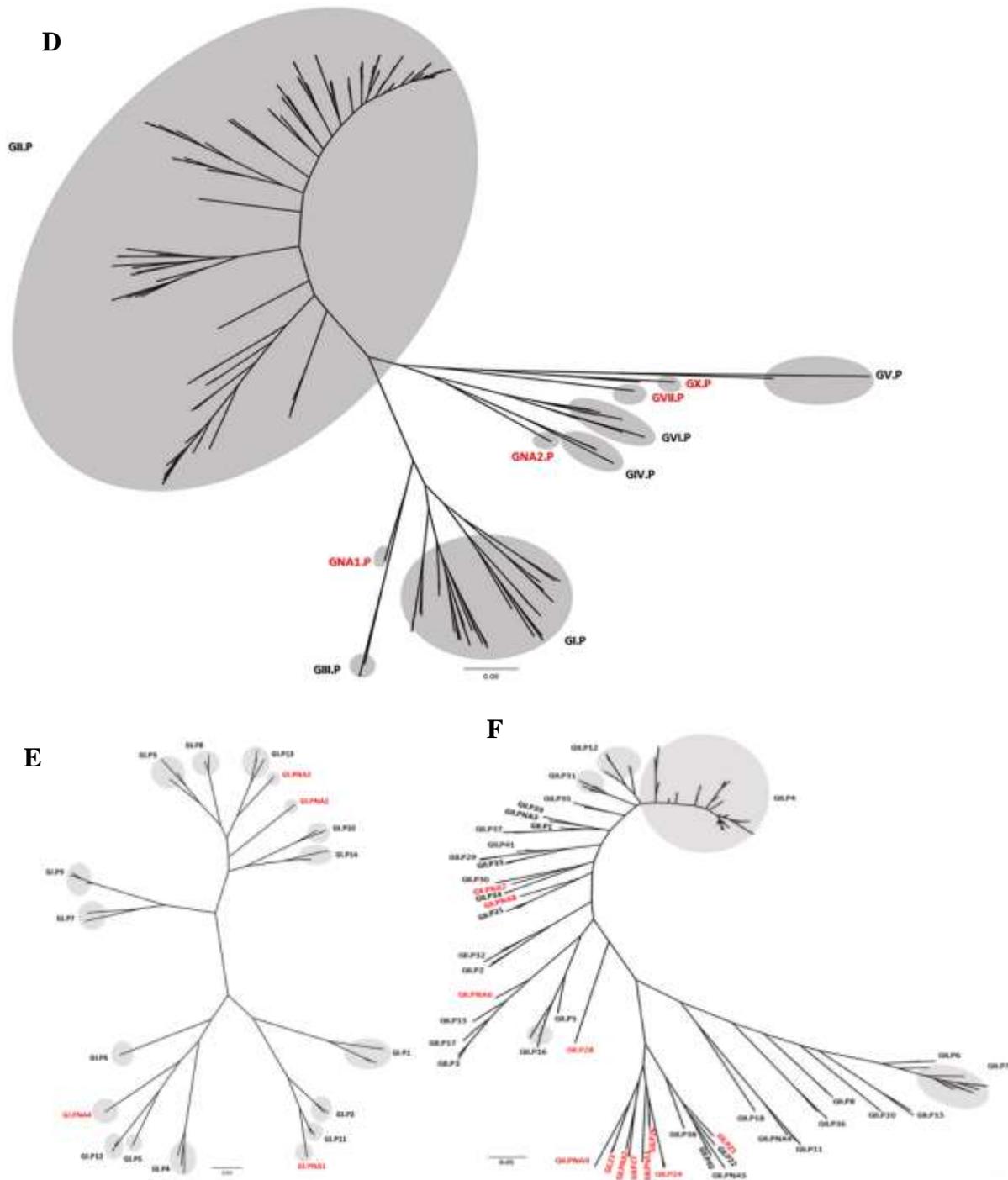
**Figure 3. Structure du génome des norovirus humains.** (Alexis de Rougemont, [https://www.sfm-microbiologie.org/wp-content/uploads/2019/02/VIRUS\\_NOROVIRUS.pdf](https://www.sfm-microbiologie.org/wp-content/uploads/2019/02/VIRUS_NOROVIRUS.pdf))

Phylogénétiquement, les norovirus peuvent être classés en différents génogroupes, puis en génotypes sur la base de la diversité du gène codant la protéine VP1 de la capside. Ces dernières années, plusieurs nouveaux norovirus ont été signalés, justifiant une mise à jour de leur classement. A ce jour, on dénombre 10 génogroupes de norovirus (GI à GX) (Figure 4A) avec des génogroupes infectant les humains (GI, GII, GIV, GVIII, et GIX), les bovins et caprins (GIII), les rongeurs (souris et rats) (GV), les porcs (GII), les canins (GIV, GVI, GVII), les félins (GIV, GVI), et les chauves-souris (GX). Ces génogroupes sont constitués de 49 génotypes (9 GI, 27 GII, 3 GIII, 2 GIV, 2 GV, 2 GVI et 1 génotype chacun pour GVII, GVIII, GIX et GX) dont ceux appartenant aux génogroupes GI et GII sont présentés dans les figures 4B et 4C respectivement. Les virus pour lesquels une seule séquence est actuellement disponible dans les bases de données publiques ont été classés en nouveaux génogroupes (GNA1 et GNA2) et génotypes (GII.NA1, GII.NA2 et GIV.NA1) provisoires (Chhabra *et al.*, 2019 ; Ford-Siltz *et al.*, 2019).

Sur la base de la diversité des nucléotides dans la région de l'ORF1 codant l'ARN polymérase ARN dépendante (RdRp), les norovirus peuvent également être classés phylogénétiquement en P(polymérase)-groupes et P-types. Ainsi, à ce jour on distingue 08 P-groupes (GI.P, GII.P, GIII.P, GIV.P, GV.P, GVI.P, GVII.P, et GX.P) subdivisés en 60 P-types (14 GI.P, 37 GII.P, 2 GIII.P, 1 GIV.P, 2 GV.P, 2 GVI.P, 1 GVII.P et 1 GX.P), ainsi que 02 P-

groupes provisoires (GNA1.P et GNA2.P) et 14 P-types provisoires (Figures 4D, 4E, 4F) (Chhabra *et al.*, 2019).



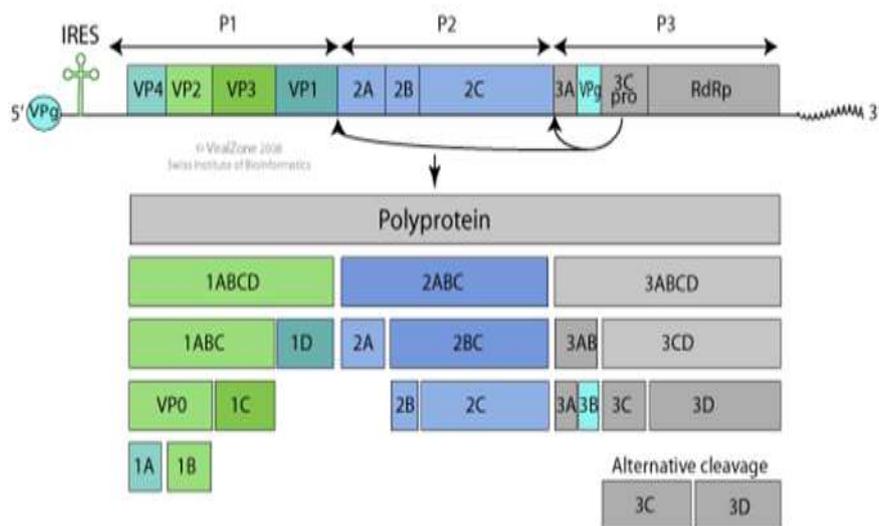


**Figure 4. Classification phylogénétique des norovirus.**

(A) Classification basée sur les séquences d'acides aminés de la VP1 en 10 génogroupes et 02 génogroupes non-attribués (NA), (B) génotypes de GI, (C) génotypes de GII. (D) Classification basée sur les séquences d'acides aminés de la RdRp des génogroupes, (E) P-types de GI, (F) P-types de GII. Les P-groupes et les P-types provisoires avec une seule séquence ou plusieurs séquences non identiques sont appelés non attribués (NA). Les génogroupes et génotypes nouvellement identifiés sont écrits en rouge. (Chhabra *et al.*, 2019).

## 1.4.2 Le virus de l'hépatite A

Responsable de l'hépatite A initialement connue sous le nom de jaunisse, le virus de l'hépatite A (VHA) ou Hepatovirus A appartient à la famille des *Picornaviridae* au sein de laquelle il forme avec huit autres espèces (Hepatovirus B à I), le genre *Hepatovirus* (Smith et Simmonds, 2018). C'est un virus nu de 27-32 nm de diamètre dont le génome est constitué d'une molécule d'ARN simple brin d'environ 7,5 kb à polarité positive possédant une seule *ORF* flanquée d'une *UTR* à extrémité 5' précédée d'une protéine génomique (VPg) servant d'amorce lors de la traduction, et d'une *UTR* à l'extrémité 3' suivie d'une queue poly(A). Après des modifications post-traductionnelles, la polyprotéine de 250 kDa issue de la traduction du génome est clivée en 11 protéines : 4 protéines structurales VP1 - VP4 (région P1) qui forment une capside virale à symétrie icosaédrique, et 7 protéines non structurales, 2A - 2C (région P2), et 3A - 3D (région P3) (Figure 5) (Wang *et al.*, 2015 ; Grenning et Cannon, 2016 ; Yu *et al.*, 2016).



**Figure 5. Structure du génome du virus de l'hépatite A**

(Viralzone/expasy, <https://viralzone.expasy.org/94>, consulté le 02/02/2022).

Sur le plan de la diversité, le VHA montre un degré modeste de diversité génétique, avec une moyenne de ~10 % de divergence de séquence de nucléotides (2,2 % de divergence de séquence d'acides aminés) parmi les séquences de régions codantes des génomes (presque) complets disponibles. Le regroupement phylogénétique des séquences a permis d'identifier six génotypes (gt I, II, III, IV, V, et VI). Avec les variants infectant les humains et appartenant aux gt I à III phylogénétiquement distincts des souches de VHA de primates non humains appartenant aux gt IV, V, et VI (Smith et Simmonds, 2018).

### 1.4.3 Le virus de l'hépatite E

Isolé et identifié pour la première en 1983 par Balaya et *al* chez des patients atteints d'hépatite non-A, non-B, le virus de l'hépatite E est un virus quasi-enveloppé de taille comprise entre 27-34 nm appartenant à la famille des *Hepeviridae* au sein de laquelle il constitue le genre *Orthohepevirus*. Son génome de 7,2 kb est constitué d'une molécule d'ARN simple brin linéaire de polarité positive. Il est composé de trois *ORFs* (*ORF1*, *ORF2*, et *ORF3*), des *UTRs* aux extrémités 5' et 3', et d'une queue polyadénylée. L'*ORF1* code pour des protéines non structurales, dont la méthyltransférase (Met), la cystéine protéase de type papaine (PCP : Papain-like cysteine protease), l'hélicase (Hel) et la RdRp impliquée dans la réplication ; l'*ORF2* code la protéine de la capsid virale ; et l'*ORF3* chevauchant partiellement l'*ORF2* et codant une petite phosphoprotéine multifonctionnelle de 113 ou 114 acides aminés qui agirait comme un adaptateur pour relier les voies de transduction intracellulaire, et réduire la réponse inflammatoire de l'hôte, interviendrait dans la sortie des virions des cellules infectées. Les *ORF2* et *ORF3* se chevauchant, leurs protéines sont traduites à partir d'un ARN subgénomique bicistronique de 2,2 kb (Figure 6) (van Tong *et al.*, 2016 ; Nagashima *et al.*, 2017).

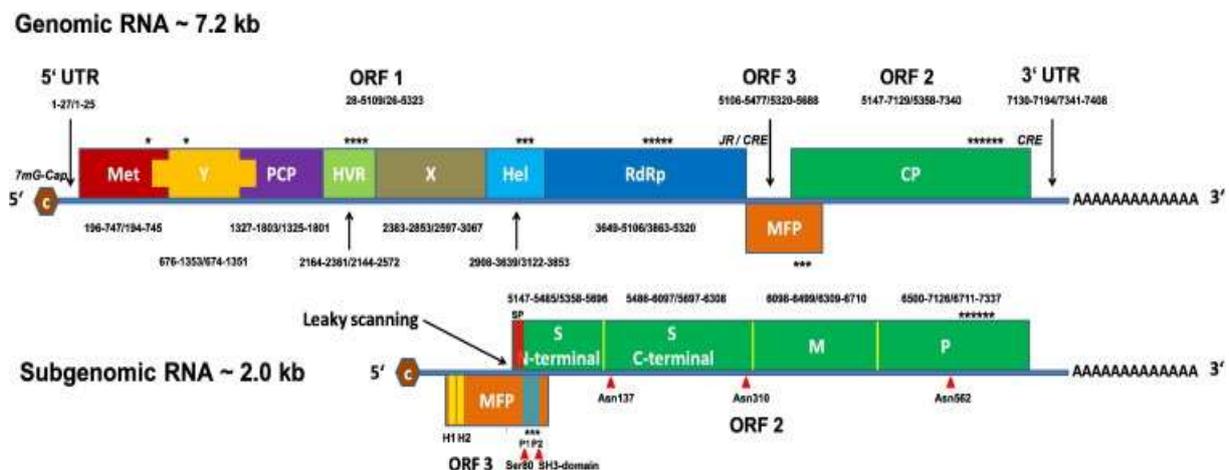


Figure 6. Structure du génome du virus de l'hépatite E (van Tong *et al.*, 2016).

Le virus de l'hépatite E présente une diversité génomique, et une variété de souches a été identifiée à travers le monde, aussi bien chez les humains que les espèces animales. Le genre *Orthohepevirus* comprend les souches de virus de l'hépatite E de mammifères et d'oiseaux et est divisé en quatre espèces (A, B, C et D) (Primadharsini *et al.*, 2019), dont l'espèce A constitué de sept génotypes (VHE-1 à VHE-7) parmi lesquels quatre (VHE-1 à VHE-4) sont capables d'infecter les humains tandis que les génotypes VHE-5 et VHE-6 n'ont été détectées jusqu'à présent que chez les animaux (sangliers) (van Tong *et al.*, 2016). Parmi les quatre génotypes

majeurs du virus de l'hépatite E capables d'infecter l'homme et appartenant à l'espèce *Orthohepevirus A*, les génotypes VHE-1 et VHE-2 sont limités à l'homme et hautement endémiques dans les régions en développement, y compris diverses régions d'Afrique, d'Asie, du Moyen-Orient et le Mexique, et sont transmis par voie fécale-orale. Tandis que le VHE-3 et le VHE-4 sont des virus zoonotiques avec une gamme d'hôtes élargie et sont la principale cause de cas sporadiques et autochtones d'hépatite E dans les pays développés (Figure 7). Les cas zoonotiques causés par les souches VHE-3 et VHE-4 sont principalement associés à des souches de porcs et de sangliers (Figure 7) (Primadharsini *et al.*, 2019 ; La Bella *et al.*, 2021).

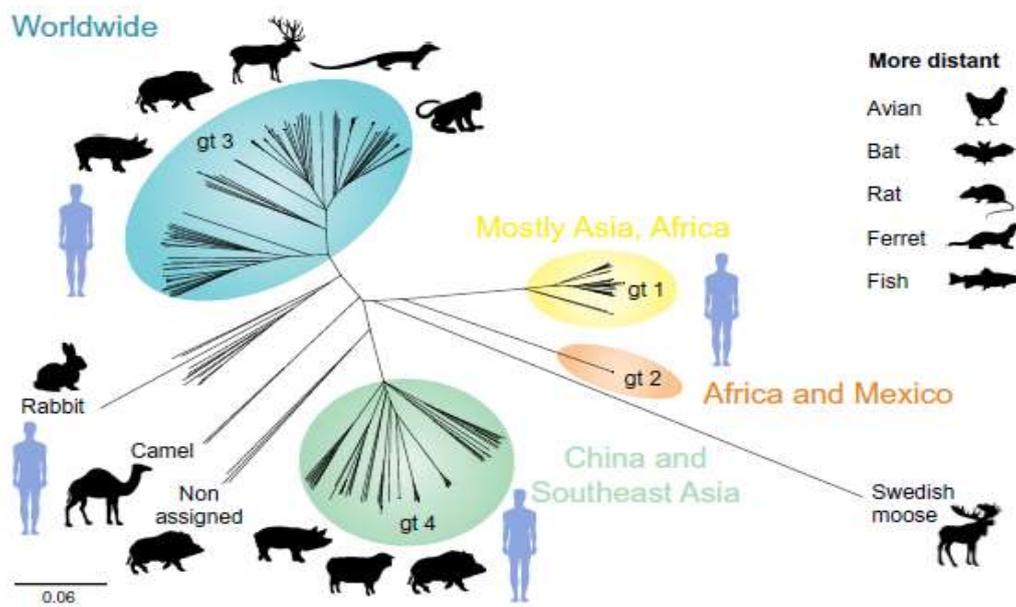
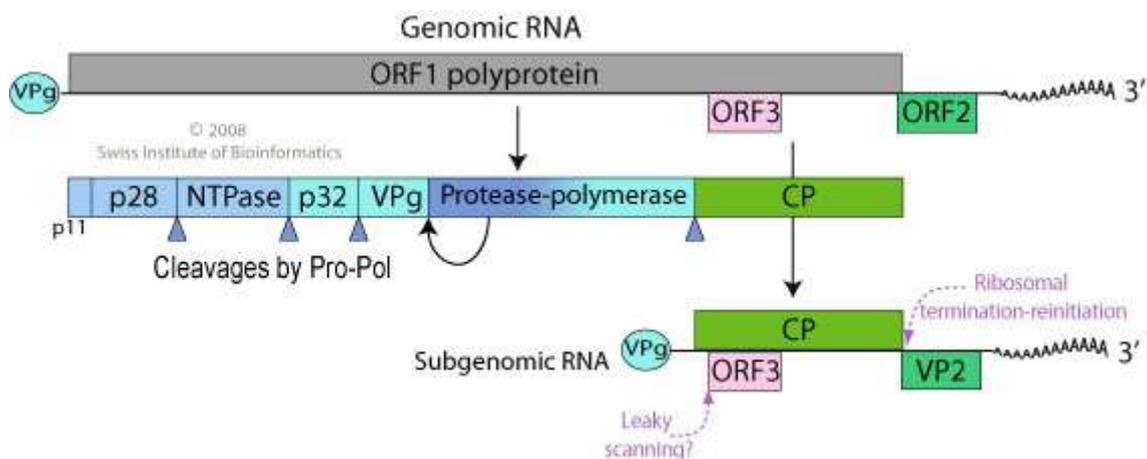


Figure 7. Diversité génétique du virus de l'hépatite E (Debing *et al.*, 2016).

#### 1.4.4 Les sapovirus

Les sapovirus sont des virus non enveloppés appartenant à la famille des *Caliciviridae*, et infectant les humains et les porcs. Le *Sapporo-like virus*, prototype de sapovirus infectant les humains chez lesquels il induit des gastroentérites a été identifié pour la première fois au Japon en 1977 dans la localité de Sapporo lors d'une épidémie de gastroentérite chez les enfants. Son génome dont la taille varie entre 7,1 et 7,7 kb est constitué d'une molécule d'ARN simple brin de sens positif et entouré d'une capsidie icosaédrique de 41-46 nm de diamètre. Les sapovirus sont divisés en cinq génogroupes (GI à GV) sur la base de la protéine VP1, avec les génogroupes GI, GII, GIV, GV infectant les humains. Neuf génogroupes supplémentaires (GVI à GXIV) ont récemment été proposés (Xue *et al.*, 2019).

Leur génome est constitué d'une protéine génomique (VPg) à l'extrémité 5', de deux ou trois *ORFs* respectivement pour le génogroupe GII et les génogroupes GI, GIV et GV, et d'une queue poly(A) à l'extrémité 3'. L'*ORF1* code une poly-protéine dont des modifications post-traductionnelles génèrent des protéines non structurales (la protéine génomique VPg, la protéase, la RdRp) et la protéine majeure de la capsid (CP). L'*ORF2* coderait la protéine mineure de la capsid (VP2) et l'*ORF3* une protéine dont la fonction demeure inconnue (Figure 8). Sur la base de la séquence de la région de l'*ORF1* codant la CP, les génogroupes GI et GII des sapovirus humains peuvent être divisés en sept génotypes chacun (GI.1 à GI.7 et GII.1 à GII.7), tandis que le génogroupe GIV n'a qu'un seul génotype (GIV.1) et le génogroupe GV deux (GV.1 et GV.2) (Oka *et al.*, 2015 ; Xue *et al.*, 2019).

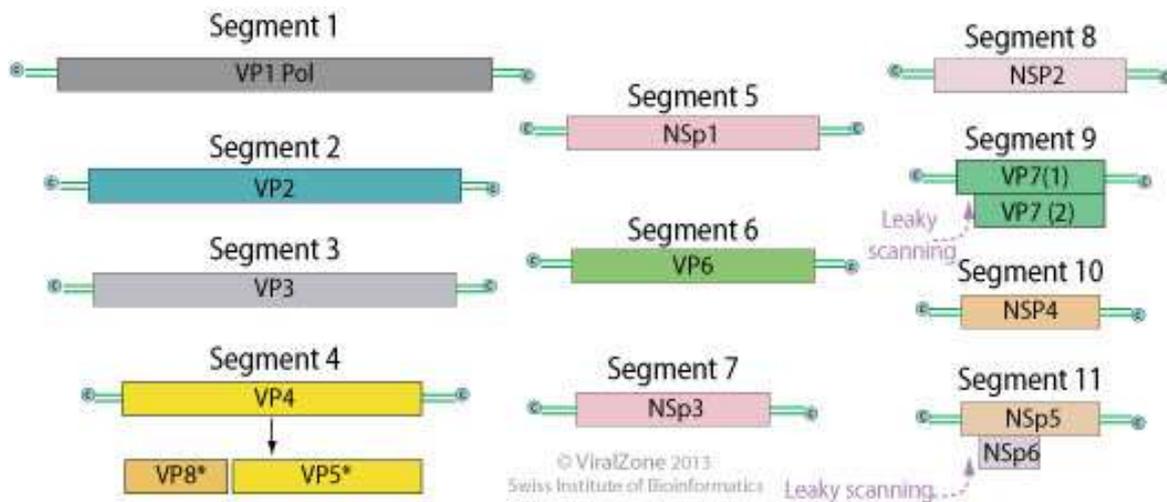


**Figure 8. Structure du génome du sapovirus**

(Viralzone/expasy <https://viralzone.expasy.org/196> , consulté le 02/02/2022).

### 1.4.5 Les rotavirus

Les rotavirus appartiennent à famille des *Reoviridae* au sein de laquelle ils constituent le genre *Rotavirus*. Ce sont des virus non enveloppés de 60-80 nm de diamètre protégés par une capsid icosaédrique (Greening et Cannon, 2016). Le génome des rotavirus est constitué de 11 segments d'ARN double brin de tailles comprises entre 667 à 3302 pb pour une taille globale de 16-27 kpb, et code 11 ou 12 protéines dont 6 protéines structurales (VP1, VP2, VP3, VP4, VP6, et VP7) et 5 ou 6 protéines non structurales (NSP1 à NSP5 ou NSP6). Chaque segment est constitué d'une *UTR* en 5' liée à une coiffe ( $m^7GpppG$ ), et d'une *UTR* en 3' et code une protéine spécifique, à l'exception du segment 11 codant les protéines NSP5 et NSP6 (figure 9) (Greening et Cannon, 2016 ; Crawford *et al.*, 2017).



**Figure 9. Structure du génome du rotavirus**

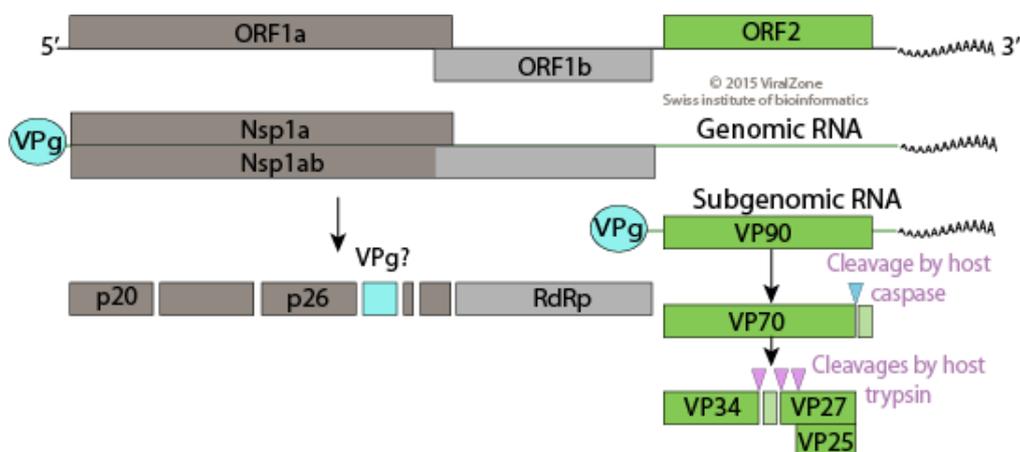
(Viralzone/expasy, <https://viralzone.expasy.org/107> , consulté le 02/02/2022).

Sur la base de la variation des réactions antigéniques et de la séquence de la protéine VP6, les rotavirus ont été regroupés en dix espèces (A-J). Considéré comme l'un des principaux virus responsables de gastroentérites chez l'homme, les rotavirus A sont classés en différents génotypes en fonction des séquences des segments 7 (génotypes G) et 4 (génotypes P). Ainsi à ce jour, 32 génotypes G et 47 génotypes P ont été identifiés. Toutefois, on observe une prédominance des génotypes G1, G2, G3, G4, G9, G12, P[4], P[6], et P[8], avec les souches G1P[8], G2P[4], G3P[8], G4P[8], G9P[8] et G12P[8] qui représentent plus de 90% des espèces de rotavirus A en circulation à travers le monde (Crawford *et al.*, 2017).

Les rotavirus, et plus particulièrement ceux du groupe A, sont les principaux agents étiologiques des diarrhées aiguës sévères chez les nourrissons et les jeunes enfants à travers le monde, et sont responsables d'environ 30 à 50 % des cas de diarrhées infectieuses aussi bien dans les pays développés que ceux en voie de développement (Tate *et al.*, 2016). Bien que la maladie touche toutes les tranches d'âge, elle est généralement considérée comme une infection bénigne chez les adultes. Toutefois, le développement des vaccins antirotavirus à partir de 2008 et leur introduction dans les programmes de vaccination de la plupart des pays a eu un impact manifeste à l'échelle mondiale, avec une réduction de 40% de la prévalence des infections à rotavirus (WHO, 2021).

### 1.4.6 Les astrovirus

Regroupés au sein de deux principaux genres, le genre Mamastrovirus (MAstVs) infectant les mammifères et le genre Avastrovirus (AAstVs) spécifique à la volaille, les *Astroviridae* sont des virus non enveloppés munis d'un génome constitué d'une molécule d'ARN simple brin à polarité positive, et protégé d'une capside rigide à symétrie icosahédrique d'environ 35 nm de diamètre. Ce dernier d'une longueur comprise entre 6,2 et 7,7 kb est constitué d'une *UTR* à son extrémité 5'-terminal, suivie de trois *ORFs* (*ORF1a*, *ORF1b*, *ORF2*), et enfin d'une chaîne poly(A). Situées à l'extrémité 5', les *ORFs* 1a et 1b codent ensemble une polyprotéine non-structurale dont les modifications post-traductionnelles aboutissent à la formation de la sérine protéase, la protéine génomique (VPg), et la RdRp. L'*ORF2* située quant à elle à l'extrémité 3' code les protéines de la capside exprimées à partir d'un ARN subgénomique (Figure 10) (Bosch *et al.*, 2014 ; Cortez *et al.*, 2017).



**Figure 10. Structure du génome des astrovirus**

(Viralzone/expasy, <https://viralzone.expasy.org/280>, consulté le 02/02/2022).

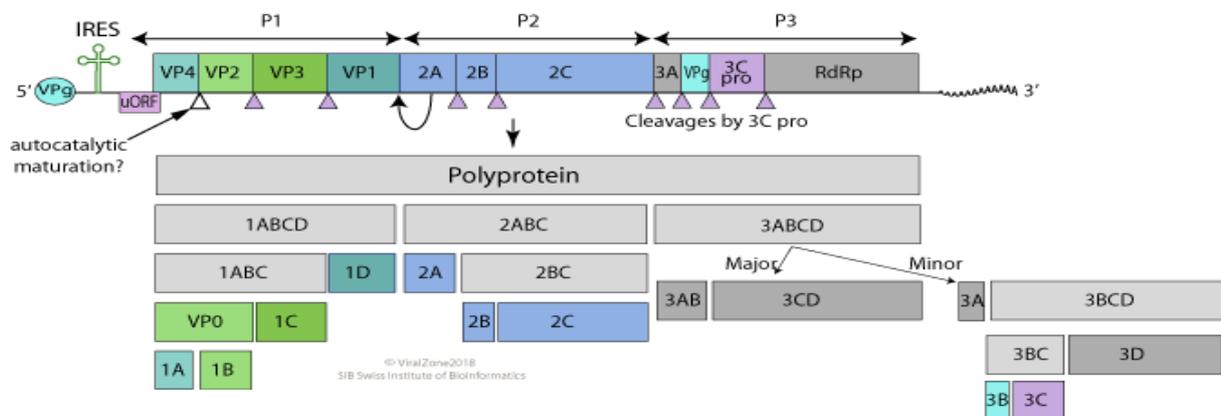
Les astrovirus humains affectent majoritairement les enfants de moins de deux ans, les personnes âgées, et les immunodéprimés, et seraient impliqués dans 20 % des cas de diarrhées sporadiques non bactériennes, et dans 0,5-15 % des épidémies de diarrhées. Les gastroentérites dues aux astrovirus sont généralement d'une sévérité moyenne, passagère et ne requièrent pas une hospitalisation. Les symptômes généralement observés sont des diarrhées liquides, et dans une moindre mesure les vomissements, les maux de tête, la fièvre, les douleurs abdominales, l'anorexie, et parfois des diarrhées chroniques chez les patients immunodéprimés (Karlsson et Schuitz-Cherry, 2013).

### 1.4.7 Autres virus émergents

D'autres virus transmis par voie féco-orale et trouvés dans les matières fécales humaines et animales comprennent entre autres les entérovirus, les aichivirus, et les picobirnavirus. Si pour certains de ces virus, la capacité à provoquer des maladies chez l'homme et/ou les animaux n'a toujours pas été prouvée, d'autres en revanche ont été à l'origine de quelques épidémies dans le passé. Bien que ces virus soient actuellement considérés comme des agents pathogènes d'une importance mineure en lien avec la consommation d'aliment, il est possible que certaines souches plus virulentes et/ou plus transmissibles puissent émerger et constituer un véritable problème de santé publique (Greening et Cannon, 2016).

#### a. Enterovirus

Les enterovirus sont des virus non enveloppés de 28-30 nm de symétrie icosaédrique dont le génome est constitué d'une molécule d'ARN simple brin, de sens positif, et de taille comprise entre 7,2 et 8,4 kb (Figure 11). Ils appartiennent à la grande famille des *Picornaviridae* et forment le genre *Enterovirus* constitué actuellement de 12 espèces : les Enterovirus A à H et J et les Rhinovirus A, B, et C. Les enterovirus comprennent entre autres les poliovirus, les coxsackie A et B virus, l'échovirus et les rhinovirus (Greening et Cannon, 2016).



**Figure 11. Structure du génome des Entérovirus**

(Viralzone/expasy, <https://viralzone.expasy.org/97>, consulté le 02/02/2022).

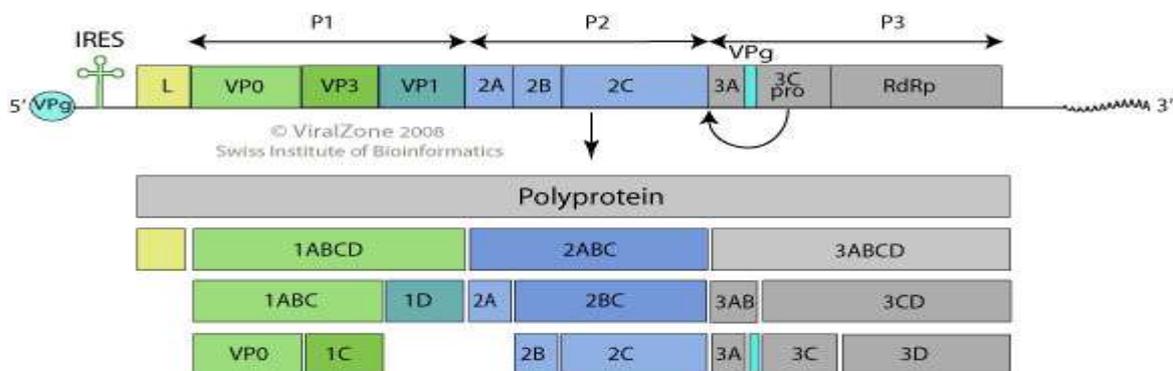
Chaque année, les enterovirus infectent des millions de personnes dans le monde. Ils se multiplient principalement dans le tractus gastro-intestinal, mais peuvent également le faire dans d'autres tissus tels que les nerfs et les muscles où ils sont responsables de méningite, d'encéphalite, de paralysie (poliomyélite), de rhume, et même de mort. Les enterovirus B sont la cause la plus fréquente de méningite dans le monde, qui touche principalement les jeunes enfants et est une cause importante d'hospitalisation. L'enterovirus D68 (EV-D68) et

l'enterovirus A71 (EV-A71) provoquent le plus souvent des maladies respiratoires et des maladies mains-pieds-bouche (HFMD) respectivement (Wollants *et al.*, 2019).

De nombreux enterovirus sont transmis par voie fécale-orale et sont excrétés dans les selles, mais ne provoquent généralement pas de gastro-entérite. On dénombre à ce jour peu d'épidémies d'origine alimentaire associées aux enterovirus malgré leur détection régulière dans l'environnement. En effet, les enterovirus, y compris les echovirus et les coxsackie virus A et B ont été isolés dans les eaux usées, les boues brutes et digérées, les eaux marines et douces, ainsi que les coquillages. Bien que les enterovirus soient parfois détectés dans les coquillages, et généralement en présence d'autres virus entériques, aucune infection liée à la consommation de coquillages contaminés par ces virus n'a été signalée (Greening et Cannon, 2016).

## b. Aichivirus

Anciennement appelé virus Aichi, c'est un virus d'environ 30 nm de diamètre appartenant au genre *Kobuvirus* de la famille des *Picornaviridae*. Il a été découvert au Japon en 1989 dans les selles des patients lors d'une épidémie de gastroentérite causée par les huîtres (D'Souza et Joshi, 2016). Son génome est constitué d'une molécule d'ARN simple brin de 8,28 kb (Figure 12) (Rivadulla et Romalde, 2020).



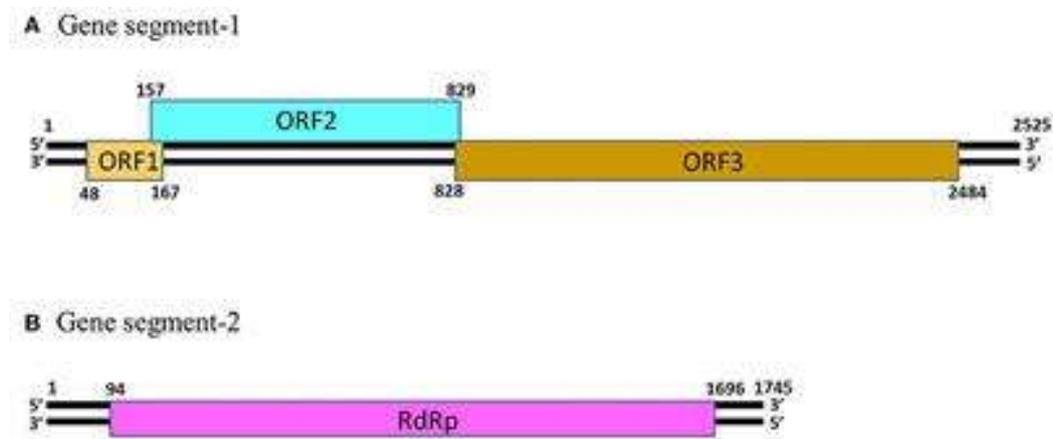
**Figure 12. Structure du génome de l'Aichivirus**

(Viralzone/expasy, <https://viralzone.expasy.org/652>, consulté le 02/02/2022).

Des aichivirus ont été détectés dans les selles, les eaux usées et les coquillages à travers le monde entier, mais semblent être plus souvent détectés dans les pays asiatiques. Toutefois, ils ont été détectés en présence de plusieurs autres virus entériques dans des échantillons cliniques et les huîtres associées à une épidémie de gastro-entérite (Guyader *et al.*, 2008). Leur importance en tant que pathogène à transmission alimentaire et hydrique demeure peu établie (Greening et Cannon, 2016).

### c. Picobirnavirus

Les picobirnavirus sont de petits virus non enveloppés dont la taille varie de 33 à 37 nm appartenant à la famille des *Picobirnaviridae* au sein de laquelle ils constituent le genre *Picobirnavirus*. Leur génome contient deux segments d'ARN double brin désignés segment-1 (taille ~2,2 à 2,7 kpb) et segment-2 (taille ~1,2 à 1,9 kpb). Le segment-1 est constitué de 2 ou 3 *ORFs*, désignés *ORF1*, *ORF2* et *ORF3*. L'*ORF3* code un précurseur de la protéine de la capsid qui subit un clivage autocatalytique pour générer la protéine de la capsid mature, et un peptide. L'*ORF2* code une protéine dont la fonction n'est pas encore connue. D'autre part, la fonctionnalité, ou encore la présence d'*ORF1* dans le segment 1 reste à élucider. Le segment-2 est constitué d'une seule *ORF* qui code la RdRp (Figure 13) (Ghosh et Malik, 2021).



**Figure 13. Structure du génome du picobirnavirus (Ghosh et Malik, 2021).**

Ces virus ont été détectés dans les excréments de nombreux mammifères et oiseaux et sont capables de provoquer des gastro-entérites chez une gamme d'animaux domestiques. Ils ont également été détectés chez les humains avec et sans diarrhée et sont à présent considérés comme agents pathogènes opportunistes pouvant provoquer des gastro-entérites chez les personnes immunodéprimées. Par ailleurs, ils ont été détectés dans les eaux usées et fluviales (Kashnikov *et al.*, 2020). Des études menées par Yinda *et al.*, en 2018 et 2019 ont permis de détecter ce virus aussi bien chez les chauves-souris que chez l'Homme au Cameroun. Leur rôle en tant que véritables agents pathogènes humains n'est pas établi, mais ils sont susceptibles d'être transmis par voie féco-orale et éventuellement au travers des aliments (Greening et Cannon, 2016).

## **1.5 Epidémies liées à la consommation des coquillages et les différents virus impliqués**

Les épidémies de gastro-entérite aiguë causées par la consommation de produits crus ou partiellement cuits tels que les huîtres, les moules, coques et palourdes contaminés par différents virus à transmission alimentaires sont fréquemment signalées dans le monde (Yang *et al.*, 2021). A titre d'exemple, de 2003 à 2017, on a noté 51 épidémies de norovirus humains dans le monde, dont 61 % étaient liés aux mollusques bivalves (Hardstaff *et al.*, 2018). Au début de l'année 2020, 1033 personnes en France ont été infectées, dont 21 hospitalisées, à la suite de la consommation de mollusques bivalves contaminés par ce virus. Cet épisode a entraîné la fermeture de 23 zones de production et plus de 400 entreprises ont été touchées (Gorji *et al.*, 2021).

A côté des norovirus qui sont les virus les plus fréquemment transmis par les mollusques bivalves, le virus de l'hépatite A vient en deuxième position en terme de prévalence, et est le pathogène viral le plus important véhiculé par ces organismes en terme de sévérité (Pintó *et al.*, 2009). Ce virus a été associé à de nombreuses épidémies d'hépatite liées à la consommation de mollusques bivalves contaminés. La plus importante remonte à 1988 en Chine avec environ 300000 personnes infectées (Halliday *et al.*, 1991). D'autres épidémies ont été associées à la consommation des huîtres en Australie (Conaty *et al.*, 2000), des moules en Italie (Crocchi *et al.*, 2000) et des palourdes en Espagne (Bosch *et al.*, 2001). La plupart de ces flambées étaient dues à la consommation de coquillages crus ou peu cuits récoltés dans des zones polluées par les eaux usées non traitées.

Connue pour sa transmission zoonotique, principalement à travers la consommation de la viande de porc, le virus de l'hépatite E est responsable de 20 millions de nouvelles infections chaque année, dont 3,3 millions aboutissent à une forme aiguë d'hépatite E et 70 000 décès (La Bella *et al.*, 2021). Plusieurs études menées au cours de cette décennie ont mis en évidence la présence du virus de l'hépatite E dans les mollusques bivalves dans différents pays (Gao *et al.*, 2015 ; Iaconelli *et al.*, 2015 ; La Rosa *et al.*, 2018 ; O'Hara *et al.*, 2018 ; Rivadulla *et al.*, 2019 ; La Bella *et al.*, 2021). Par ailleurs, en 2008, la consommation de coquillages aurait été à l'origine d'une infection d'hépatite E de 195 personnes à bord d'un bateau de croisière (Said *et al.*, 2009).

Plusieurs études ont mis en évidence la présence des sapovirus dans les mollusques bivalves dans plusieurs pays à travers le monde, notamment au Japon (Iizuka *et al.*, 2010), en

Espagne (Varela *et al.*, 2016), et en Italie (Fusco *et al.*, 2017) où ces derniers ont été associés à une épidémie de gastroentérite. En 2008, ces virus ont été détectés avec d'autres virus entériques dans des huîtres impliquées dans une épidémie en France (Le Guyader *et al.*, 2008).

Infectant majoritairement les jeunes enfants, la transmission des rotavirus se ferait principalement de personne-à-personne par l'intermédiaire des personnels soignants et des adultes en général. Les infections d'origine alimentaire demeurent peu connues. Toutefois, des épidémies associées aux aliments et à l'eau ont été signalées dans un certain nombre de pays (Greening et Cannon, 2016). Par ailleurs, les rotavirus, en particulier le rotavirus A ont été détectés dans les coquillages (Kittigul *et al.*, 2015 ; Marinho *et al.*, 2018 ; Ito *et al.*, 2021 ; Lyu *et al.*, 2021, Grodzki *et al.*, 2012).

## **1.6 Approche de détection et de quantification des virus dans les mollusques bivalves**

La détection et la quantification des virus dans les mollusques bivalves requièrent un ensemble d'opérations dont les principales sont :

### **1.6.1 Dissection et extraction des tissus digestifs**

L'élution des particules virales des tissus de mollusques bivalves demande une préparation en amont afin d'isoler les tissus cibles (branchies, manteaux, palpes labiaux) pour l'analyse. Lors de leur alimentation par filtration de l'eau, les bivalves concentrent des contaminants de diverses natures au niveau de leurs tissus digestifs (TD). Pour la détection des virus entériques humains, il a été démontré que ces tissus concentraient 90% de la contamination, facilitant ainsi leur détection (Atmar *et al.*, 1995). Ceci a été confirmé par les travaux de Maalouf (Maalouf *et al.*, 2010 ; Maalouf *et al.*, 2011) sur la détection des norovirus GI et GII dans les huîtres.

De manière générale, afin d'éviter les variabilités individuelles, 6 à 10 individus sont lavés et ouverts. Les TD sont disséqués à l'aide d'un scalpel stérile, coupés en petits morceaux (environ 1-2 mm), répartis en aliquotes de 2 g et congelés (-20°C) ou directement analysés (Le Guyader *et al.*, 2009).

### **1.6.2 Elution virale**

Deux approches sont généralement utilisées lors de l'élution des virus des TD des bivalves :

### **a. La digestion enzymatique**

Recommandée dans la méthode ISO/TS 15216-1 : 2017 pour la détection des norovirus et du virus de l'hépatite A dans les mollusques bivalves, cette approche utilise une protéase, la protéinase K, dont l'activité hydrolytique aboutit à la libération des virus de la matrice (Jothikumar *et al.*, 2006). Après l'action enzymatique favorisée par un traitement thermique adéquat, les hydrolysats tissulaires sont sédimentés par centrifugation et le surnageant contenant les particules virales est récupéré et utilisé pour l'extraction des acides nucléiques (ISO/TS 15216-1 : 2017).

### **b. Précipitation chimique**

Cette technique de concentration vise à réduire le volume de l'échantillon et à éliminer une partie de la matrice en limitant la perte de particules virales. Une phase primaire d'homogénéisation de l'échantillon (par exemple avec un potter ou un ultra-turax) est appliquée pour broyer les tissus, suivie par une étape d'éluion virale par variation de pH et/ou des conditions ioniques de l'échantillon. Elle utilise le plus souvent une précipitation chimique par le mélange Polyéthylène Glycol 6000/Chlorure de Sodium (PEG 6000/NaCl). Ce dernier forme un maillage où les particules virales seront retenues et précipitées à la suite d'une centrifugation réfrigérée. Le culot est ensuite remis en suspension dans une solution tampon (PBS ou glycine) pour l'extraction des acides nucléiques viraux (Atmar *et al.*, 1995 ; Le Guyader *et al.*, 2009). Bien que cette méthode nécessite plus d'étapes, avec un potentiel risque de perte de particules virales, elle présente l'avantage de ne pas endommager la capsid virale par l'utilisation d'une enzyme ou par un traitement thermique et permet la récupération des particules virales infectieuses (Polo *et al.*, 2018).

### **1.6.3 Extraction des acides nucléiques**

L'ARN viral peut être extrait de la particule virale par traitement thermique, enzymatique ou chimique (Bosch *et al.*, 2018).

La norme ISO/TS 15216-1 :2017 pour la détection et quantification des norovirus et du virus de l'hépatite A dans les matrices alimentaires, y compris les mollusques bivalves, décrit un protocole d'extraction des ARN commun aux différentes matrices alimentaires basé sur la dénaturation de la capsid à l'aide d'un réactif chaotropique (guanidinium thiocyanate) suivi de l'adsorption des ARN sur des billes de silice et de plusieurs étapes de lavage (Bosch *et al.*,

2018). Le développement des kits commerciaux basés sur cette technologie facilite les étapes d'extraction et de purification des acides nucléiques (Le Guyader *et al.*, 2009 ; Lowther *et al.*, 2019).

#### **1.6.4 Détection/quantification**

Grâce aux avancées dans les techniques de biologie moléculaire, la quantification des copies de génome présentes dans un échantillon est désormais une technique préconisée (ISO/TS 15216-1 : 2017), marquant un pas important entre l'analyse qualitative capable de seulement donner une information de présence/absence de l'agent pathogène ciblé et l'analyse quantitative. Par ailleurs, la quantification des virus constitue un point important pour l'investigation des épidémies, ainsi que la surveillance de routine car elle peut fournir des données intéressantes en vue de fixer des niveaux d'acceptation de la contamination virale des produits alimentaires. Cette approche quantitative a ainsi été utile pour réaliser une évaluation quantitative du risque lors d'une épidémie d'hépatite A liée à la consommation de tellines (bivalves) (Pintó *et al.*, 2009). Les données obtenues lors d'épidémies de gastroentérite liées à la consommation d'huîtres ont été comparées à celles obtenues lors d'études chez les volontaires. Une bonne corrélation a été observée, mais avec une très grande variabilité en fonction des souches de norovirus et des consommateurs (Teunis *et al.*, 2020).

La majorité des méthodes actuellement utilisées pour la détection des virus dans les aliments est basée sur la PCR. La RT-qPCR est une méthode sensible et spécifique disponible pour la détection des ARN dans des échantillons environnementaux où la concentration virale est généralement assez faible. Cette approche a obtenu une reconnaissance officielle à travers la norme ISO/TS 15216-1 pour la recherche et la quantification du virus de l'hépatite A et des norovirus dans les aliments et les surfaces alimentaires.

Après la rétrotranscription de l'ARN en ADN complémentaire (ADNc), l'enzyme Taq-polymérase amplifie la zone cible en effectuant un nombre défini de cycles d'amplifications. Une sonde marquée avec une molécule fluorescente permet de détecter le nombre de cycles auquel la fluorescence franchit le seuil de sensibilité (*Threshold*, en anglais) donnant la valeur de *Ct* (*Cycle threshold*). Cette valeur associée à une courbe standard générée à partir de quantités connues de la séquence cible représentée par un ARN ou un ADN synthétique ou transcrit *in vitro* permet de quantifier la concentration virale dans l'échantillon. Toutefois, la production et la quantification de réactifs standards par des laboratoires individuels peuvent

induire une variation inter-laboratoires dans la quantification. Ceci suggère donc l'utilisation de réactifs standards certifiés afin de réduire les variations (Bosch *et al.*, 2018).

## **1.7 Approche métagénomique de l'étude du virome des bivalves**

### **1.7.1 Enjeux de l'étude métagénomique du virome**

La détection des virus transmis par voie alimentaire et hydrique fait face à de nombreux défis : les fréquentes transmissions secondaires de personne à personne qui peuvent masquer l'origine initiale du virus, la grande diversité virale qui complique leur identification, et la non-sensibilisation des cliniciens face à la similitude des symptômes observés chez les patients. En effet, les symptômes cliniques des maladies virales d'origine alimentaire varient, allant de la diarrhée et des vomissements aux crampes abdominales et au malaise général. Ceci rend difficile l'identification de l'agent étiologique exact, conduisant à un diagnostic erroné si la démarche détective est sélective. Par ailleurs, il est important d'identifier la présence d'autres virus présentant des périodes d'incubation plus longues, comme le virus de l'hépatite A, afin d'avertir et de surveiller les consommateurs. Une approche globale de la détection des virus dans les aliments semble séduisante (Nieuwenhuijse et Koopmans, 2017 ; Desdouits *et al.*, 2020b).

La métagénomique est un terme utilisé pour décrire l'approche au cours de laquelle tous les acides nucléiques d'un échantillon sont séquencés. En virologie, elle fait référence à la tentative de récupération et d'identification des génomes complets et partiels de tous les virus présents dans un échantillon (Strubbia *et al.*, 2020). Elle présente un grand potentiel dans la surveillance des virus dans la chaîne alimentaire mondiale en raison de sa sensibilité, de sa large plage de détection et des informations détaillées sur les virus détectés (Aarestrup *et al.*, 2012). Le progrès des technologies du séquençage de nouvelle génération (NGS : *Next Generation Sequencing*), en particulier le NGS métagénomique (mNGS), ainsi que le développement rapide de la bio-informatique et des outils de calcul permettent une détection impartiale de tous les virus présents dans un échantillon. Ce qui montre un potentiel pour la détection en temps opportun d'étiologies infectieuses rares ou nouvelles, ainsi que pour la surveillance des virus d'origine alimentaire et hydrique. Offrant de nouvelles opportunités pour une surveillance des maladies infectieuses et émergentes d'origine alimentaire à grande échelle. Les exemples de séquençage métagénomique des virus dans les denrées alimentaires sont encore rares, mais quelques essais d'analyses métagénomiques ont permis de caractériser des virus de la famille des *Reoviridae* et des *Picobirnaviridae* sur de la laitue cultivée au champ (Aw *et al.*, 2014). Par

ailleurs, une étude récente a révélé l'efficacité de la technique mNGS dans l'étude de la diversité des virus à ARN et des papillomavirus dans des produits frais issus de l'agriculture en Espagne (Itarte *et al.*, 2021). Cette approche semble par ailleurs séduisante pour les coquillages, qui, comme décrit précédemment peuvent accumuler une large diversité de virus. En effet, on note quelques travaux rapportant un séquençage non ciblé du virome des mollusques (Andrade *et al.*, 2015 ; Rosani et Gerdol, 2017 ; Rosani *et al.*, 2019 ; Strubbia *et al.*, 2020 ; Desdouts *et al.*, 2020b ; Kim *et al.*, 2020 ; Tan *et al.*, 2021).

## **1.7.2 Analyse métagénomique de la contamination virale**

### **a. Préparation des échantillons**

Afin d'optimiser le séquençage métagénomique, des méthodes de préparation des échantillons visant à réduire le matériel génomique non viral ou à sélectionner spécifiquement des virus peuvent être utilisées (Nieuwenhuijse et Koopmans, 2017). Ces approches vont de différents protocoles d'extraction (Cotten *et al.*, 2014 ; Conceição-Neto *et al.*, 2015) à l'utilisation d'une puce spécifique de capture du virome (Briese *et al.*, 2015) ou d'anticorps pour capturer les particules virales (Oude Munnink *et al.*, 2013).

Plusieurs étapes importantes de traitement (Figure 14) permettent de réduire l'influence (sensibilité, formation des chimères) de la matrice et de la charge bactérienne, tout en éliminant les acides nucléiques libres. Il s'agit entre autres de :

- **L'homogénéisation**

Afin d'obtenir des résultats reproductibles, une bonne homogénéisation des solides biologiques est cruciale et minimise le colmatage ultérieur du filtre. Il a été démontré qu'une homogénéisation avec des billes de 2,8 mm de diamètre conduit à une destruction des particules virales peu importe la vitesse d'homogénéisation. De plus, les billes de 2,8 mm entraînaient une augmentation de 316 % et 350 % de la détection de l'ARNr 16S à des vitesses de 5 000 et 3 000 tr/min respectivement. Pour les bactéries, ces billes ont également montré des augmentations allant de 612% à 6184%. La réduction des particules virales était plus faible en utilisant une homogénéisation à 3 000 tr/min sans billes et n'a pas augmenté la quantité d'ADN/ARNr bactérien (Conceição-Neto *et al.*, 2015).

- **La centrifugation**

Elle est couramment utilisée pour précipiter des particules plus grosses, telles que des cellules et des débris cellulaires, tandis que les virus restent en solution. Une étude a montré

que la centrifugation à 100 g pendant 3 min entraînait des différences limitées par rapport au contrôle pour les virus, les bactéries et l'ARNr. Cependant, une centrifugation pendant 30 min à 100 g a entraîné une réduction accrue du mimivirus (94,6 %). Pour les bactéries, les réductions les plus élevées ont été observées pour les *Lactobacillus* (99,0 %) et les *Bifidobacterium* (96,0 %), alors que *E. coli* et les *Bacteroides* ont montré des réductions moins prononcées (80,6 % et 48 %, respectivement). La centrifugation à 17 000 g pendant 3 et 30 min a entraîné des réductions de plus de 99,99 % pour les *Bifidobacterium*, *E. coli* et les *Bacteroides*. La réduction de *Lactobacillus* était légèrement inférieure (99,7 % à 99,8 %), tandis que la réduction de l'ARNr était de 98,5 à 99,4 %. Pendant 3 min à 17000 g, les pertes pour les virus étaient négligeables, sauf pour le polyomavirus (43,3 %), l'herpèsvirus (65,3%). Ces résultats étaient beaucoup plus prononcés lorsque la centrifugation était effectuée à 17 000 g pendant 30 min (Conceição-Neto *et al.*, 2015).

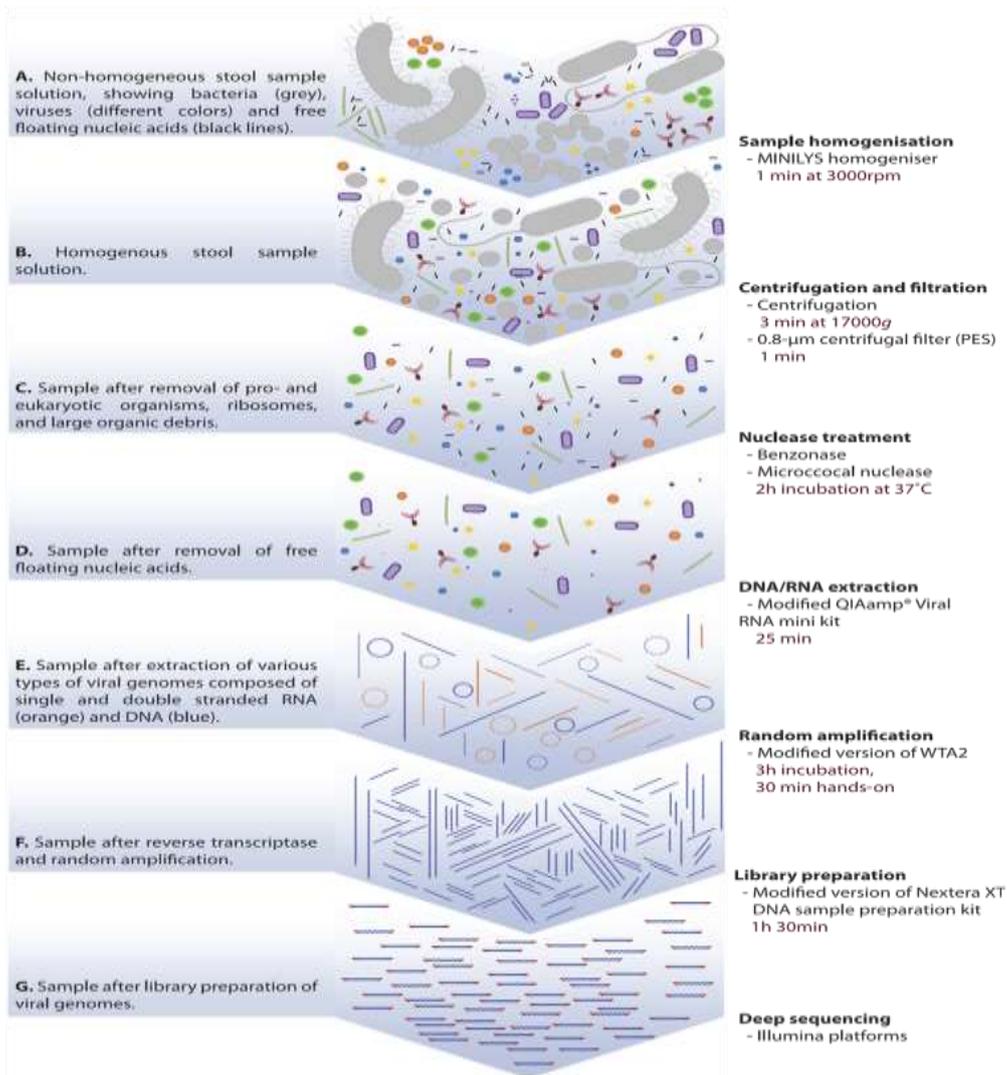
- **La filtration**

La filtration est une méthode efficace et largement utilisée pour enrichir les particules virales en éliminant les bactéries et les cellules hôtes (humaines/animales/végétales). Dans une étude menée par Conceição-Neto *et al.*, en 2015, un virome et une communauté bactérienne factices ont été filtrés avec un filtre centrifuge de 0,8 µm en polyéthersulfone (PES), un filtre en polycarbonate de 0,8 µm (PC), un filtre centrifuge de 0,45 µm en polyfluorure de polyvinylidène (PVDF) et un filtre centrifuge de 0,22 µm (PVDF), et comparé à un témoin non filtré. Les filtrations avec les deux filtres de 0,8 µm et celui de 0,45 µm ont montré des effets limités sur la plupart des virus, à l'exception du mimivirus, le plus gros virus du virome, pour lequel des diminutions de 81,7 %, 95,9 % et de 99,0 % ont été observées pour les filtres PES 0,8 µm, PC 0,8 µm et 0,45 µm, respectivement. La plus faible récupération (99,90% de perte) de ce virus ayant été observée avec le filtre de 0,22 µm. L'utilisation du filtre PC de 0,8 µm a entraîné une réduction de 99,1 % et de 99,2 % pour les *Lactobacillus* et les *Bifidobacterium*, respectivement, tandis que *E. coli* (80,7 %) et les *Bacteroides* (50,1 %) ont montré des réductions modestes. Une élimination beaucoup plus efficace des bactéries a été obtenue avec les filtres PES de 0,8 µm, 0,45 µm et 0,22 µm, éliminant 99,5 % à 99,90 % de toutes les bactéries présentes dans la communauté bactérienne. Indépendamment du filtre utilisé, l'ARNr a été éliminé moins efficacement que les bactéries (réduction de 30 à 75 %) (Conceição-Neto *et al.*, 2015).

- **Le traitement par le chloroforme**

Le chloroforme perturbe les membranes lipidiques et est souvent utilisé pour éliminer les bactéries. Cependant, certains virus sont enveloppés et leur capsid peut devenir instable après le retrait de l'enveloppe. Des tests ont rélevé que des incubations avec des pourcentages élevés de chloroforme (5 à 20 %) entraînaient des pertes très limitées des virus nus, notamment les circovirus et les parvovirus. Les réductions les plus élevées ont été observées avec le coronavirus enveloppé (81,4–89,9 %). Les bactéries Gram-positifs étaient sensibles à 1 % de chloroforme, entraînant une diminution de 97,6 à 99,4 % pour les *Bifidobacterium* et les *Lactobacillus* respectivement. Tandis que les bactéries à Gram négatif étaient moins affectées avec des réductions inférieures à 90 %. Cependant, des concentrations élevées de chloroforme (10 à 20 %) ont entraîné une réduction substantielle de toutes les espèces bactériennes (99,6 à 99,999 % ; 8,8 à 17,6 Ct). Par ailleurs l'ARNr a également été réduit de 83,5 à 94,9 % (Conceição-Neto *et al.*, 2015).

Ces différents traitements permettent d'enrichir l'échantillon en particules virales complètes. Toutefois, l'absence d'une méthode unique pour le séquençage haut débit complique l'analyse du virome, voire l'obtention de résultats reproductibles.



**Figure 14. Etapes de la préparation des échantillons selon l’approche Neto-Vir (Conceição-Neto *et al.*, 2015).**

### **b. Préparation des bibliothèques**

Le séquençage NGS nécessite la conversion des acides nucléiques initiaux en bibliothèques standards appropriées pouvant être chargées dans les séquenceurs (Figure 15). Il existe une grande variété de protocoles de préparation des bibliothèques NGS. Dans le cas général commun à presque tous les protocoles de préparation d'échantillons, le matériel de départ qui est une molécule d'ADN double brin (ADN génomique isolé ou ADNc) est relié à des adaptateurs contenant les éléments nécessaires à son immobilisation sur une surface solide et au séquençage. Par ailleurs, la qualité des données de séquençage dépendant fortement de la qualité du matériel séquencé, le processus de construction de la bibliothèque doit garantir une récupération moléculaire élevée des fragments originaux afin d'obtenir une bonne couverture

génomique (van Dijk *et al.*, 2014). Ainsi, la conversion des acides nucléiques initiaux en bibliothèques de séquençage nécessite plusieurs étapes dont les principales sont :

– **La synthèse de l'ADNc à partir de l'extrait d'ARN**

Des méthodes de séquençage direct des molécules d'ARN ont été reportées, mais ces dernières font encore l'objet d'une intense recherche/développement. Les technologies les plus développées et performantes ont été optimisées pour le séquençage de l'ADN. La transcription de l'ARN en ADNc est donc nécessaire et représente une étape critique lors de la préparation des bibliothèques (Boone *et al.*, 2018). Les ARNs extraits sont retro-transcrits en ADNc en utilisant une transcriptase inverse et des amorces constituées de courtes séquences aléatoires de quelques nucléotides. L'utilisation d'amorces aléatoires pendant cette étape est importante pour la métagénomique afin de permettre la rétrotranscription de tous les génomes présents dans l'échantillon. Parmi les amorces les plus souvent utilisées figurent :

- **Les hexamères random**

Commercialisés par *New England BioLabs* (NEB), le mélange optimisé et prêt à l'utilisation est composé d'amorces de 6 bases (d'où le nom d'hexamères) et d'oligo(dT)-ancrés de 23 résidus d'acide déoxythymidylique suivis de 02 résidus de nucléotides additionnels représentés par VN (V : dA, dC, ou dG et N : dA, dC, dG ou dT) qui facilitent la liaison avec la première base en amont de la queue poly(A) de l'ARNm (<https://international.neb.com/-/media/catalog/datacards-or-manuals/s1330datasheet-lot0041208.pdf?rev=33bbe02bd0684b90983f259786047360>). Bien que leur exacte composition ne soit pas connue, ces amorces aléatoires sont largement utilisées dans la synthèse du premier brin de l'ADNc.

- **Les hexamères non-ribosomiaux**

Ils sont sélectionnés à partir des hexamères random afin de limiter l'amplification de l'ARN ribosomal (ARNr) de l'hôte sans avoir un impact sur l'amplification des virus de mammifères (Endoh *et al.*, 2005). Le séquençage des molécules d'ARNr peut en effet réduire de façon importante la sensibilité globale du séquençage.

Le brin d'ADNc synthétisé sert ensuite à la synthèse du deuxième brin complémentaire grâce à l'action d'une ADN polymérase.

## – **Fragmentation**

La taille des fragments d'ADN dans la librairie finale est un paramètre clé pour la construction de la librairie NGS, car elle permet d'optimiser le séquençage. A titre d'illustration, pour un meilleur résultat de séquençage, les plateformes Illumina demandent des fragments de taille comprise entre 200 et 500 pb (Bronner *et al.*, 2013). Trois approches sont disponibles pour fragmenter les chaînes d'acides nucléiques : les approches physiques, enzymatiques et chimiques. La fragmentation de l'ADN est généralement effectuée par des méthodes physiques (cisaillement acoustique et sonication) ou des méthodes enzymatiques (cocktails d'endonucléases non spécifiques et réactions de tagmentation par une transposase) (Marine *et al.*, 2011).

Le cisaillement acoustique utilisant la technologie Acoustique Focalisée Adaptative (AFA) est une des méthodes les plus utilisées avec l'instrument Covaris (Covaris, Woburn, MA), et permet d'obtenir des fragments d'ADN dans la plage de 100 à 5 000 pb. Les méthodes enzymatiques comprennent la digestion par la DNase I ou la Fragmentase, un mélange de deux enzymes (New England Biolabs, Ipswich MA). Les comparaisons des librairies NGS construites avec le cisaillement/sonication acoustique par rapport à la Fragmentase ont révélé que les deux étaient efficaces. Cependant, la Fragmentase a produit un plus grand nombre d'artefacts par rapport aux méthodes physiques. Une méthode enzymatique alternative pour fragmenter l'ADN est la technologie de tagmentation Nextera d'Illumina (Illumina, San Diego, CA) dans laquelle une enzyme transposase fragmente et insère simultanément des séquences adaptatrices dans l'ADNdb. Cette méthode présente plusieurs avantages, notamment une réduction du temps de manipulation et de préparation des échantillons (Head *et al.*, 2014).

Indépendamment de la stratégie de fragmentation utilisée, la taille et la distribution des fragments doivent être vérifiées avant de passer à l'étape suivante. Cette analyse est réalisée par électrophorèse des fragments d'ADN, le plus souvent en utilisant un appareil de type *Bioanalyzer* (Agilent) qui permet une visualisation du profil des librairies en mesurant la concentration des fragments de différentes tailles (Lee et Abdullah, 2014 ; Aigrain *et al.*, 2016).

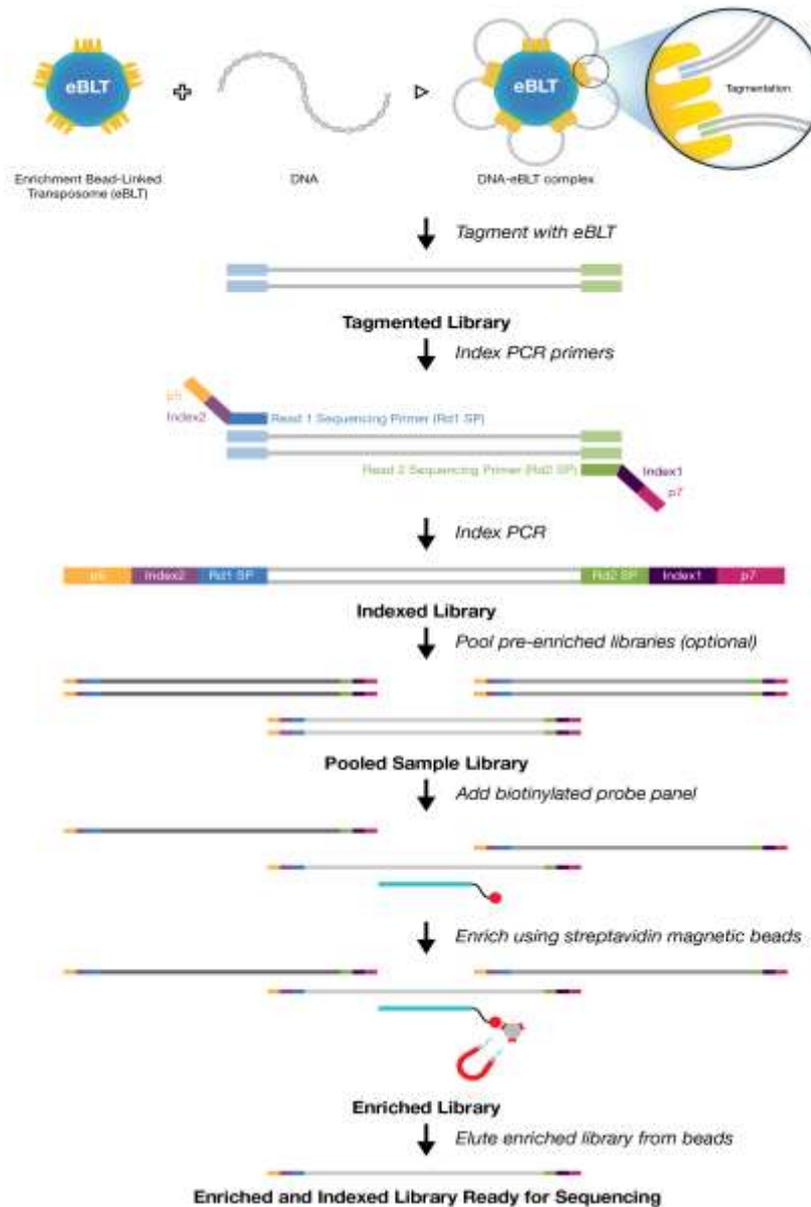
## – **Réparation (*End-repair*), (A)-tailing, ligation des adaptateurs et pré-amplification**

Une fois l'ADN de départ fragmenté, une réparation (*End-repair*) des extrémités des fragments générés et la phosphorylation des extrémités 5' sont nécessaires afin d'obtenir des bouts francs et permettre la ligation des adaptateurs associés à l'index. Ceci est réalisé grâce à

un mélange de trois enzymes : l'ADN polymérase T4, le grand fragment de Klenow, et la polynucléotide kinase T4 (Head *et al.*, 2014).

Afin d'éviter la ligation des fragments d'ADN entre eux, un nucléotide (généralement l'Adénine) est ajouté en position 3' du fragment d'ADN, étape appelée « *tailing* », à l'aide soit de la Taq polymérase, soit du fragment de Klenow (Head *et al.*, 2014). S'en suit alors la ligation des adaptateurs. Pendant cette réaction, le rapport adaptateur/fragment optimal est d'environ 10 : 1, calculé sur la base du nombre de copies ou de la molarité. En effet, trop d'adaptateur favorise la formation de dimères d'adaptateurs qui peuvent être difficiles à séparer et perturbent l'amplification ultérieure. Pour faciliter le multiplexage, différents adaptateurs à code-barres peuvent être utilisés avec chaque échantillon. Alternativement, des codes-barres peuvent être introduits à l'étape suivante d'amplification par PCR en utilisant différentes amorces de PCR à codes-barres pour amplifier différents échantillons. A cet effet, des réactifs de haute qualité avec des adaptateurs à code-barres et des amorces PCR sont disponibles dans des kits de nombreux fournisseurs (Head *et al.*, 2014).

Pour une meilleure qualité des bibliothèques, des opérations de lavage sont nécessaires. Celles-ci sont faites à l'aide de billes magnétiques ou de colonnes et peuvent être effectuées après la réparation finale et les réactions d'A-*tailing*. Toutefois, un nettoyage des bibliothèques avec des billes après la ligation des adaptateurs s'est révélé plus efficace quant à l'élimination de l'excès de dimères d'adaptateurs (Head *et al.*, 2014).



**Figure 15. Préparation des bibliothèques NGS par l'approche Nextera d'Illumina.**

La fragmentation de l'ADN est réalisée par une réaction enzymatique (Tagmentation). On note également un enrichissement de la bibliothèque à l'aide d'un panel de sondes biotinylées. (Illumina, 2022, <https://emea.support.illumina.com/downloads/illumina-dna-prep-with-enrichment-reference-guide-1000000048041.html> )

Une amplification préalable de chaque fragment de la bibliothèque est nécessaire afin d'augmenter le signal du séquençage. Lors du séquençage Illumina, l'amplification à pont (*bridge-PCR*) permet la fixation des adaptateurs sur la *flow cell* qui est un support en lame de verre creux à travers lequel s'écoulent les réactifs et l'ADN à séquencer (Figures 17B, C). À la suite de la formation d'un pont attaché, les fragments sont amplifiés formant des *clusters*

(Figure 17C). À chaque cycle, des étapes de synthèse et de dénaturation se succèdent pour produire, à la fin de la PCR, environ 1000 copies de chaque *cluster* (ensemble de séquences regroupées issues de l'amplification d'un fragment), avec plusieurs millions de *clusters* amplifiés simultanément. Le profil des librairies est ensuite visualisé et la concentration mesurée avant la réalisation du séquençage haut débit (Shendure et Ji, 2008).

### **c. Capture et enrichissement en séquences d'intérêt**

La faible charge virale généralement observée dans les aliments constitue un élément limitant pour leur identification. Afin d'optimiser la détection et la caractérisation des virus lors du séquençage NGS, des stratégies visant à augmenter la sensibilité se sont concentrées sur l'enrichissement des séquences virales à travers la centrifugation, la filtration, le traitement au chloroforme, et la soustraction des acides nucléiques de l'hôte et de l'ARNr par digestion enzymatique (Briese *et al.*, 2015 ; Conceição-Neto *et al.*, 2015). Bien que ces approches soient utiles, aucune n'a atteint la sensibilité requise. Afin de relever ce défi, des systèmes de sélection positive basée sur la capture des séquences d'ADN d'intérêt à l'aide des sondes nucléotidiques biotinylées ont été développés afin d'enrichir les librairies de séquences virales (Figure 15). Parmi ces derniers, la puce VirCapSeq-VERT développée en 2015 par Briese *et al.* est un système de capture du virome des vertébrés. Elle est utilisée lors du séquençage métagénomique NGS et permet d'augmenter la sensibilité de la détection des virus. Ce système utilise environ 2 millions de sondes oligonucléotidiques différents qui couvrent les génomes des membres des 207 taxons viraux connus pour infecter les vertébrés, y compris les humains. Selon ses concepteurs, l'utilisation de cette puce a entraîné une multiplication par 100 à 10000 du nombre de *reads* viraux à l'issue du séquençage de librairies issues d'extraits nucléotidiques sanguins et tissulaires. Ceci par rapport au séquençage Illumina conventionnel utilisant des procédures d'enrichissement viral établis (filtration, traitements aux nucléases et soustraction de l'ARNr, etc.) (Briese *et al.*, 2015).

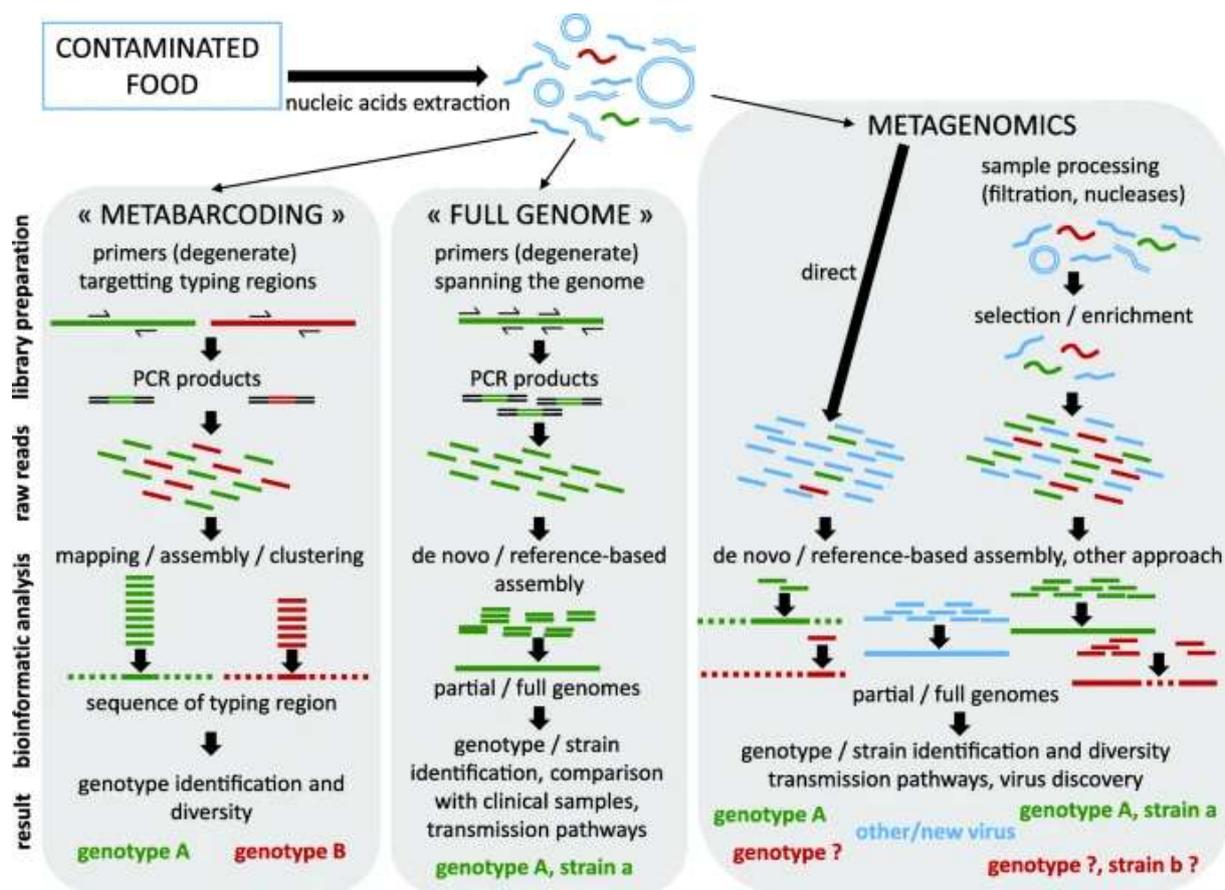


Figure 16. Quelques technologies de séquençage de nouvelle génération appliquées à la recherche des virus dans les aliments (Desdouts *et al.*, 2020).

#### d. Le séquençage : de génération en génération

L'enchaînement des résidus d'acides nucléiques dans les séquences polynucléotidiques détermine les propriétés héréditaires et biochimiques de l'ensemble des organismes terrestres. Par conséquent, la capacité de mesurer ou de déduire de telles séquences est impérative pour la recherche biologique. Au cours des cinquante dernières années, un grand nombre de chercheurs se sont appliqués à produire des techniques et des technologies pour faciliter cette approche, en séquençant des molécules d'ADN et d'ARN. Cette quête a connu d'énormes avancées, passant du séquençage de courts oligonucléotides à celui de longues chaînes polynucléotidiques constituées de millions de bases ; de la description d'un gène unique au séquençage rapide et disponible d'un génome entier (Heather et Chain, 2016).

##### – Séquençage de première génération

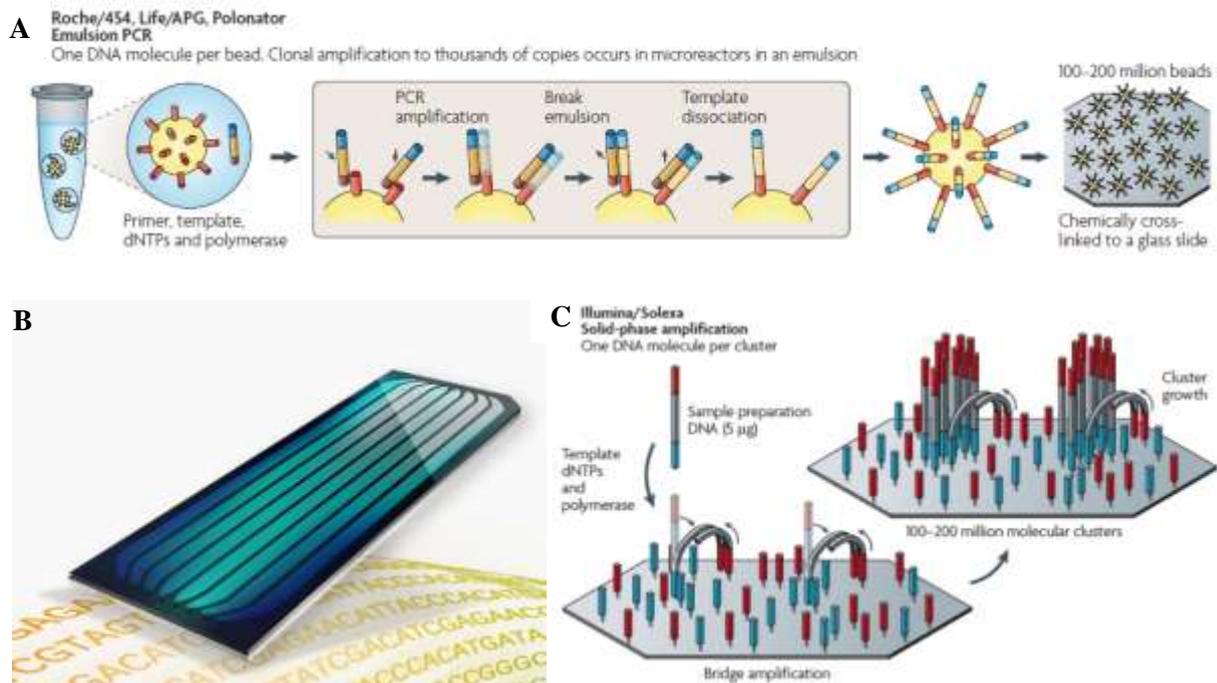
Les deux premières méthodes de séquençage de l'ADN sont apparues en 1977. L'une, connue sous le nom de séquençage par « clivage chimique » a été publiée en février par Maxam

et Gilbert en 1977. La seconde, connue sous le nom de séquençage en « terminaison de chaîne », séquençage « dideoxy » ou séquençage « Sanger », a été publiée en décembre par Sanger et ses collaborateurs (Gužvić, 2013). Les deux approches reposent sur la séparation d'un mélange de fragments d'ADN de différentes tailles sur des plaques de gel de polyacrylamide. En raison de deux inconvénients majeurs que sont l'utilisation de produits chimiques dangereux (par exemple, l'hydrazine qui est une neurotoxine), et la nécessité d'une grande quantité d'ADN, la première approche de séquençage n'a pas été largement utilisée. Le séquençage de Sanger, également connu sous le nom de séquençage de première génération, développé il y a près de 40 ans, est toujours considéré comme une bonne méthode et permet de séquencer de longues séquences nucléotidiques (jusqu'à 1000 pb) avec un taux d'erreur relativement faible. Contrairement à la méthode du « clivage chimique », ici, l'ADN a été synthétisé par extension de l'amorce oligonucléotidique médiée par l'ADN polymérase I (fragment de Klenow) ; l'ADNsb servant de matrice. En raison de sa simplicité et de sa fiabilité, cette approche est rapidement devenue la méthode de séquençage dominante. Son automatisation a permis la réalisation d'importants projets de séquençage. Notamment, le séquençage des génomes de la mouche (*Drosophila melanogaster*) et de l'Homme. De nos jours, il est fréquemment utilisé pour des reséquençages ciblés à petite échelle (Gužvić, 2013 ; Shendure *et al.*, 2017).

#### – Séquençage de seconde génération

Les séquenceurs dits de seconde génération permettent de séquencer en parallèle un très grand nombre de fragments d'ADN de courtes tailles (150 à 300 pb), et peuvent générer jusqu'à 20 milliards de séquences en simultané. Le séquençage de seconde génération englobe différentes technologies de séquençage utilisant chacune une méthode spécifique. Toutefois, elles sont toutes basées sur la même approche qui consiste après extraction, à découper l'ADN en fragments de tailles plus ou moins homogènes avant la préparation des librairies. Ceci aboutit à la fixation des index (ou barcodes) et des adaptateurs sur les extrémités de chaque fragment, permettant pour les premiers à identifier les échantillons d'origines différentes lors du séquençage simultané, et pour les seconds, la fixation des fragments sur le support de séquençage (*flowcell*) (Figure 17B), ainsi que l'initiation de la formation des clones (ou amplification) afin d'augmenter le signal (Figure 17C). L'amplification peut se dérouler en émulsion sur billes préalablement préparées afin de capturer les adaptateurs (Roche/454, Life/APG, Polonator) (Figure 17A), ou sur un support solide ou *flowcell* (Illumina/Solexa) (Figure 17C). Pendant le séquençage, des cycles de lavages et de lecture se succèdent (*wash & scan*) avec incorporation successive des nucléotides et émission d'un signal lumineux ou

chimique à chaque incorporation. La succession des signaux permet de lire les séquences d'ADN (*reads*) (Metzker, 2010).



**Figure 17. Amplification des bibliothèques avant le séquençage de seconde génération.**

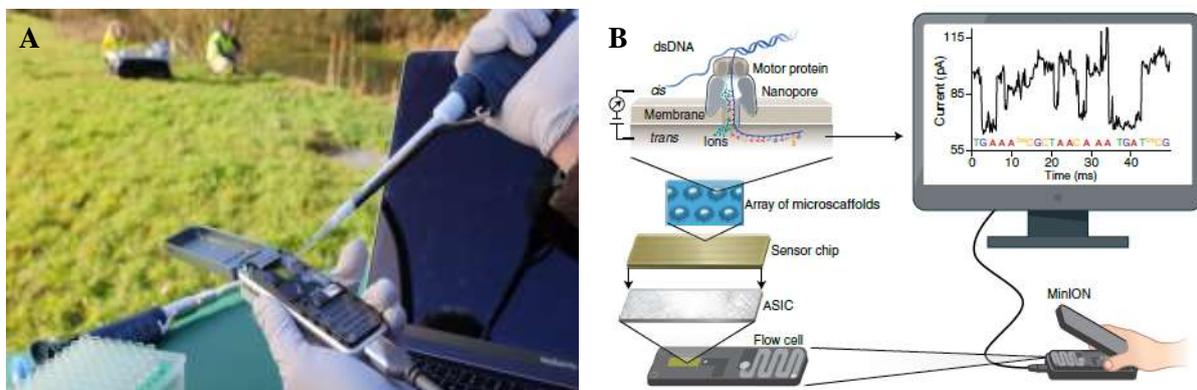
(A). Amplification en émulsion sur billes préalablement préparées afin de capturer les adaptateurs (Metzker, 2010). (B). Support solide d'amplification (ou *flowcell*) pour séquençage Illumina. (Illumina, <https://emea.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html>) (C). Amplification en « ponts » et constitution des *pools* de séquence sur *flowcell* (Metzker, 2010).

### – Séquençage de troisième génération

Également désigné « *long-read sequencing* », le séquençage de troisième génération a ces dernières années provoqué une nouvelle révolution dans le domaine de la génomique avec ses applications dans l'étude aussi bien des génomes, des transcriptomes, que des métagénomes. Avec une résolution sans précédent. En effet, alors que les techniques de seconde génération conduisent à la production de courtes séquences ou *reads*, ce dernier permet de générer des séquences de plusieurs dizaines de kilobases.

Les techniques de séquençage de troisième génération sont récemment devenues disponibles dans le commerce avec les séquenceurs distribués par *Pacific Biosystems* (PacBio) qui utilisent le séquençage en temps réel à molécule unique (SMRT), tandis qu'Oxford Nanopore Techniques (ONT) a développé un dispositif pour le séquençage nanopore.

Contrairement aux techniques de deuxième génération (454, Illumina, Ion Torrent), ces méthodes n'incluent pas d'étape d'amplification lors de la préparation de la librairie de séquençage. Permettant ainsi le séquençage d'une seule molécule (Bleidorn, 2016). S'agissant du séquençage Nanopore, on note un développement et une expansion accélérée avec la création en 2014 du séquenceur MinION (Figure 18A). La technologie repose sur un pore de protéine à l'échelle nanométrique, ou « nanopore », qui sert de biocapteur et est intégré dans une membrane polymère électriquement résistante. Dans une solution électrolytique, une tension constante est appliquée pour produire un courant ionique à travers le nanopore de telle sorte que des molécules d'ADN ou d'ARN simple brin chargées négativement soient entraînées à travers le nanopore du côté « *cis* » chargé négativement au côté « *trans* » chargé positivement. La vitesse de translocation est contrôlée par une protéine motrice qui fait passer progressivement la molécule d'acide nucléique à travers le nanopore. Les modifications du courant ionique au cours de la translocation correspondent à la séquence nucléotidique présente dans la région de détection et sont décodées à l'aide d'algorithmes de calcul. Ceci permet le séquençage en temps réel de molécules individuelles. En plus de contrôler la vitesse de translocation, la protéine motrice a une activité hélicase permettant de dérouler des duplex ADN double brin ou ARN-ADN en molécules simple brin qui traversent le nanopore (Figure 18B) (Wang *et al.*, 2021).



**Figure 18. Séquençage de troisième génération.**

(A). Séquenceur MinION de Nanopore. (Alex Ralph, 2021

<https://www.thetimes.co.uk/article/top-biotech-s-2bn-float-boosts-city-zkdc58jf2>) (B).

Principe du séquençage Nanopore (Wang *et al.*, 2021).

### e. Le traitement bio-informatique des données de séquençage

À la suite du nombre sans cesse croissant de *reads* (plusieurs centaines de milliers à plusieurs millions pour chaque librairie) générés avec l'optimisation des séquenceurs de nouvelle génération, le traitement et l'analyse des données issues du séquençage à haut débit posent des problèmes de calcul et requièrent des infrastructures ainsi que des connaissances en bio-informatique afin de sélectionner les outils les plus performants et les mieux adaptés tout en réduisant les temps d'analyse. Les outils d'analyse des données métagénomiques tels que UBLAST (Edgar, 2010), Kraken (Wood *et al.*, 2014), Kaiju (Menzel *et al.*, 2016) et Diamond (Buchfink *et al.*, 2014) disponibles en ligne permettent un traitement rapide et précis des données tout en nécessitant un outil informatique modeste (Nieuwenhuijse et Koopmans, 2017).

De manière générale, le traitement bio-informatique des données issues du séquençage de haut débit nécessite un certain nombre d'étapes :

#### – Le contrôle qualité et le prétraitement

La qualité des *reads* obtenus étant un élément déterminant dans le traitement des données issues du séquençage de haut débit, la première étape consiste à évaluer et ajuster la qualité des données, ainsi qu'à exclure les erreurs techniques. Ce prétraitement nécessite plusieurs étapes parmi lesquelles : la suppression des séquences des adaptateurs utilisés lors de la phase de préparation des librairies, l'élimination des *reads* de mauvaise qualité (caractérisés par un faible score *Phred* attribué par le séquenceur), ou la suppression des *reads* contenant des nucléotides non identifiés (Tableau 2) (Nooij *et al.*, 2017).

**Tableau 2. Estimation de la qualité des séquences obtenues à l'aide du score *Phred*.**

Score de qualité <i>Phred</i>	Probabilité d'une identification incorrecte	Précision de l'identification d'une base
10	1 pour 10	90 %
20	1 pour 100	99 %
30	1 pour 1000	99.9 %
40	1 pour 10000	99.99 %
50	1 pour 100000	99.999 %

#### – Filtrage des *reads* non-viraux

Ce traitement consiste à éliminer les séquences non virales, et permet d'accélérer la recherche dans les bases de données en réduisant le nombre de requêtes. Il permet également

de réduire les faux positifs, et empêche l'assemblage des séquences chimériques matrice-virus, bactéries-virus. Cependant, il peut également entraîner la perte de séquences virales. Ce traitement est rendu possible par des outils tels que BWA et Bowtie 2. Des logiciels tels que VIP permettent un filtrage en comparant d'abord les *reads* aux bases de données de la matrice, puis à celles des virus. Il supprime uniquement les *reads* qui sont les plus semblables à des séquences non-virales de référence dans le but d'atteindre une sensibilité élevée pour les virus, et potentiellement réduire les résultats faux positifs (Nooij *et al.*, 2018).

#### – **L'assemblage**

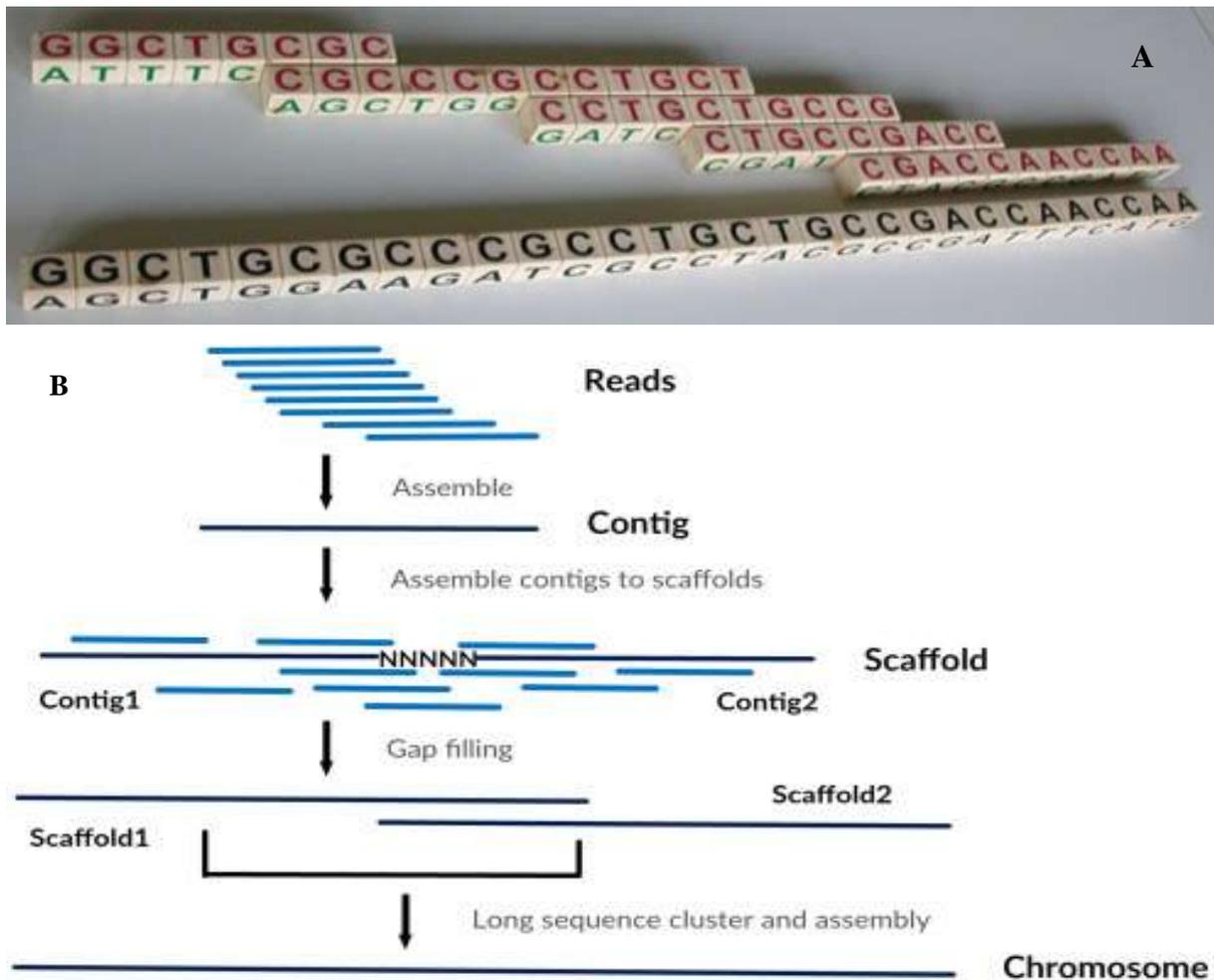
Avant la classification, les courts *reads* issus du séquençage sont assemblés en séquences contiguës plus longues (*contigs*) et consensuelles. Ceci permet d'éliminer les erreurs liées à chaque *read* et de réduire la quantité de données pour une analyse plus approfondie. Cette étape qui constitue un des défis majeurs de l'analyse métagénomique virale peut être faite selon deux approches :

- **L'alignement des *reads* à une séquence de référence**

Les *reads* sont comparés à des séquences de référence, permettant leur assemblage. Cette méthode présente un biais selon la séquence de référence choisie, et assemble en général peu de *reads*. Diminuant ainsi la qualité des informations sur les variations génétiques obtenues lors de l'assemblage (Lin *et al.*, 2017).

- **L'assemblage *de novo***

Les *reads* sont reliés ou assemblés entre eux grâce à la présence de chevauchements des nucléotides permettant la construction des *contigs* (séquence génomique générée par l'assemblage de courtes séquences issues du séquençage (*reads*)). Selon l'abondance de *reads*, chaque position du génome est couverte par plusieurs *reads* : ce nombre de *reads* constitue la couverture. L'importance et l'homogénéité de la couverture permet de différencier les « vrai *contigs* » et les faux positifs. L'assemblage des *contigs* entre eux permet d'obtenir une séquence plus longue (*scaffold*), en incluant dans la séquence des *gaps* (trou) induits par l'absence d'informations (Figure 19). L'estimation de la taille des *gaps* est possible en utilisant l'approche *Paired-End* qui se réfère à la possibilité de séquencer les deux extrémités du même fragment d'ADN, permettant un alignement des *reads* plus précis (Guo *et al.*, 2017 ; Nooij *et al.*, 2018)



**Figure 19. Assemblage *de novo* des reads.**

(A). Illustration de l'alignement nucléotidique. (Rayan Chikhi, 2018 [http://rayan.chikhi.name/pdf/big18\\_large\\_genome\\_assembly.pdf](http://rayan.chikhi.name/pdf/big18_large_genome_assembly.pdf).) (B). Assemblage des *reads* en *contigs*, *scaffolds*, et chromosomes (Guo *et al.*, 2017).

Dans les approches métagénomiques virales, l'assemblage *de novo* est souvent considéré comme la méthode de choix. En effet, les virus évoluant rapidement, des références adaptées ne sont pas toujours disponibles. De plus, les génomes viraux courts se traduisent généralement par une couverture de séquençage élevée, facilitant l'assemblage. Cependant, l'assemblage *de novo* est susceptible de générer des *contigs* erronés en reliant des *reads* contenant des erreurs techniques, telles que les erreurs liées au séquençage et celles se rapportant à la présence de séquences d'adaptateurs résiduelles. Une autre source de *contigs* erronés peut être la similarité des *reads* provenant d'organismes différents entraînant la formation de chimères. Ainsi, un assemblage *de novo* des *contigs* corrects dépend d'un contrôle qualité et d'un prétraitement strict, ainsi que d'un bon filtrage de la qualité et un regroupement

taxonomique : c'est-à-dire, d'un regroupement préalable des *reads* selon leurs taxons respectifs (Nooij *et al.*, 2018).

Les données de la métagénomique virale représentent un défi pour les outils d'assemblage. En effet, les régions répétées au sein des génomes viraux sont nombreuses, ainsi que les régions hypervariables associées à l'interaction avec l'hôte. De plus, le taux de mutation élevé entraîne une augmentation de la complexité du métagénome et une variation des souches conduisant à des assemblages fragmentés limitant l'analyse (Sutton *et al.*, 2019).

Une autre approche (Kaiju) évite l'étape d'assemblage en traduisant directement les *reads* en protéines pour les identifier dans une base de données protéiques (Menzel *et al.*, 2016). Une stratégie plus récente (Plass) a proposé la conversion des *reads* en protéines qui sont ensuite assemblées pour obtenir des séquences plus longues, réduisant la variabilité et facilitant l'identification génomique (Steinegger *et al.*, 2019).

#### **f. Recherche dans les bases de données**

Dans cette étape de recherche, les séquences obtenues (*reads* ou *contigs*) sont soumises à des bases de données de référence. Plusieurs outils sont disponibles pour l'alignement des séquences protéiques (Diamond) ou nucléotidiques (BWA ou Bowtie2) ou les deux (BLAST) à des bases de données, permettant ainsi l'assignation taxonomique et donc l'identification des séquences obtenues (Nooij *et al.*, 2018).

Tous ces outils de recherche reposent sur des bases de données de référence telles que NCBI GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), RefSeq (<https://www.ncbi.nlm.nih.gov/refseq/>). GenBank a l'avantage d'être une grande base de données fréquemment mise à jour avec les séquences de nombreux organismes. D'autres outils utilisent des bases de données spécifiques aux virus telles que GIB V (Hirahata *et al.*, 2007) ou ViPR (Pickett *et al.*, 2012), qui ont l'avantage d'une meilleure annotation et curation au détriment du nombre de séquences incluses (Nieuwenhuijse et Koopmans, 2017).

Toutefois, les données de ces bases conditionnent la qualité des résultats obtenus. En effet, avec l'avènement du séquençage haut débit, la vitesse à laquelle les séquences de nouveaux virus sont découvertes et ajoutées dans les bases de données augmente rapidement. Ceci pose d'une part le problème de l'importance de mises à jour récurrentes, et d'autre part celui de la qualité des séquences parfois incomplètes ou erronées soumises. De plus, l'homologie entre plusieurs génomes de référence peut conduire à une mauvaise assignation des séquences (Nieuwenhuijse et Koopmans, 2017).

## CHAPITRE II : MATERIEL ET METHODES

### 2.1 Matériel

#### 2.1.1 Matériel biologique

Le matériel biologique utilisé au cours de cette étude était constitué :

- Des échantillons de palourdes appartenant à l'espèce *Galatea paradoxa* (Born, 1778) qui se développe en amont de l'embouchure du fleuve Sanaga ont été collectés entre février 2018 et février 2020 dans les sites Bolounga et Moulongo dans le Canton Malimba de l'arrondissement de Mouanko (3°32'20.0"N 9°39'02.0"E), situés dans le département de la Sanaga Maritime, région du Littoral,
- Une suspension de mengovirus de concentration  $10^6$  TCID<sub>50</sub>/mL utilisée comme contrôle du processus d'éluotion virale,
- Les ARNs de référence des norovirus GI, GII, du virus de l'hépatite A et du mengovirus utilisés dans la réalisation des courbes standards de quantification virale,
- Le kit de clonage *TOPO TA Cloning Kit for Sequencing* constitué du vecteur plasmidique pCR4-TOPO-TA et du *One Shot TOP10 Chemically Competent Escherichia coli* (ThermoFischer).



**Photo 1.** Echantillon de palourdes de la Sanaga (Bonny Patrice, 2018).

#### 2.1.2 Réactifs

Un ensemble de réactifs a été utilisé. Il s'agissait entre autre de :

- La Protéinase K, 30 U/mg (Sigma-Aldrich, France/ThermoFischer Scientific),
- Les Superscript I, II et III (invitrogen, France),
- Le tampon de lyse au thiocyanate de guanidium,
- Le Kit d'extraction NucliSENS biomerieux,

- Différents couples d'amorces spécifiques aux norovirus (GI et GII), au virus de l'hépatite A et au mengovirus, ainsi que des amorces non spécifiques Random Hexamer (invitrogen, France),
- Le kit QIAprep Spin Miniprep (Qiagen) pour la purification des plasmides,
- L'Endonucléase Omnicleave™ (200 U/μL) pour la digestion des acides nucléiques en suspension,
- Le PolyEthylène Glycole (PEG),
- Le TURBO™ DNase (2U/μL) (invitrogen/Ambion, ThermoFischer Scientific, France).
- Le kit RNA Clean – Concentrator™-5 (Zymo Research, Irvine USA) pour la purification des ARN,
- Le kit NEBNext Ultra RNA Library (New England Biolabs, France) pour la préparation des bibliothèques Illumina,
- Le kit VirCapSeq-VERT capture panel (Roche) pour la capture et la concentration des séquences de virus des vertébrés.

### **2.1.3 Appareillage**

- Centrifugeuses,
- Agitateurs rotatifs,
- Portoirs magnétiques
- La plateforme d'extraction d'acides nucléiques NucliSens miniMAG (bioMérieux, France)
- Thermocycleur de modèle MX3000P *Quantitative PCR system* (Agilent Technologies France),
- Le séquenceur ABI 3730xl (Applied Biosystems),
- Le séquenceur Illumina NovaSeq 6000,
- Le sonicateur (Bandelin UD 2200),
- Le système d'extraction d'acides nucléiques semi-automatisé eGENE-UP™ (bioMerieux, France),
- Le Qubit™ 4 *Fluorometer* (invitrogen/ThermoFischer Scientific, France), pour la quantification des acides nucléiques,
- L'ultrasonicateur Covaris M220 pour la fragmentation des ADNc,
- Le bio-analyseur *Agilent 2100 Bioanalyser system* pour l'analyse de la qualité des bibliothèques d'ADN.

## 2.2 Méthodes

### 2.2.1 Description du site d'échantillonnage

L'échantillonnage a été effectué sur les points Bolounga et Moulongo à 12 km de l'embouchure du fleuve Sanaga dans le Parc National de Douala-Edéa, situé dans la Région du Littoral, Départements du Wouri et de la Sanaga Maritime, Arrondissements d'Edéa 1<sup>er</sup>, de Dizangué, de Mouanko et de Manoka, sur une superficie de 262 935 ha (DECRET N°2018/8399/PM du 11 Octobre 2018 portant mutation de la Réserve de Faune de Douala-Edéa en Parc National de Douala-Edéa.

<https://www.spm.gov.cm/site/sites/default/files/mutation%20de%20la%20reserve%20de%20la%20faune.pdf> ).

Situés dans le bassin de Kribi-Douala sur la côte atlantique et couvrant une grande partie de la plaine côtière camerounaise, ce site a un climat de type équatorial caractérisé par des pluies abondantes (3000-4000 mm) et des températures généralement élevées, avec une moyenne mensuelle de 24-29°C. Il est traversé par le fleuve Sanaga qui est le plus long fleuve du Cameroun, prenant sa source dans le sud de la Région de l'Adamaoua et parcourant 918 km jusqu'à l'océan Atlantique. Les 43 derniers km de ce fleuve sont situés dans le parc National de Douala-Edéa. Au sein de ce dernier, le fleuve a une largeur de 1 à 3 km, contenant une mosaïque de bancs de sable et d'îles végétalisées, de nombreux effluents et lacs avec un complexe de bancs de sable, de lagons peu profonds et d'îles couvertes de mangroves à l'entrée de la mer. Le débit d'eau du fleuve (débit moyen de 2072 m<sup>3</sup>/s avec un débit moyen de 473 m<sup>3</sup>/s en mars, le plus élevé d'août à novembre de 5700 m<sup>3</sup>/s) est si puissant que à marée haute l'influence de l'eau salée est limitée à moins de 1 km en amont de l'embouchure (Ajonina *et al.*, 2005).

Les points de Bolounga (Bol) et Moulongo (Mou) sont situés dans une importante zone humide englobant deux écosystèmes de base (marin et forêt humide), et caractérisé par une grande biodiversité. Les divers habitats offerts par le fleuve et les lacs associés abritent une grande diversité de flore et de faune. En effet, on note divers types de végétations : les forêts primaires pluvieuses ; les forêts secondaires ; les terres agricoles et mangroves ; entrecoupées de grands marécages et d'un réseau dense de rivières et de lacs abritant de fortes densités d'espèces de mammifères forestiers, notamment les sangliers, les potamochères, sitatunga (antilopes) et diverses espèces de primates comme le singe colobe noir, les chimpanzés, ainsi que les éléphants forêt. Il abrite également de nombreuses espèces d'oiseaux résidents et migrateurs. L'écosystème aquatique et le littoral atlantique sont riches en ressources aquatiques de renommée internationale d'importance pour la conservation, notamment les tortues de mer,

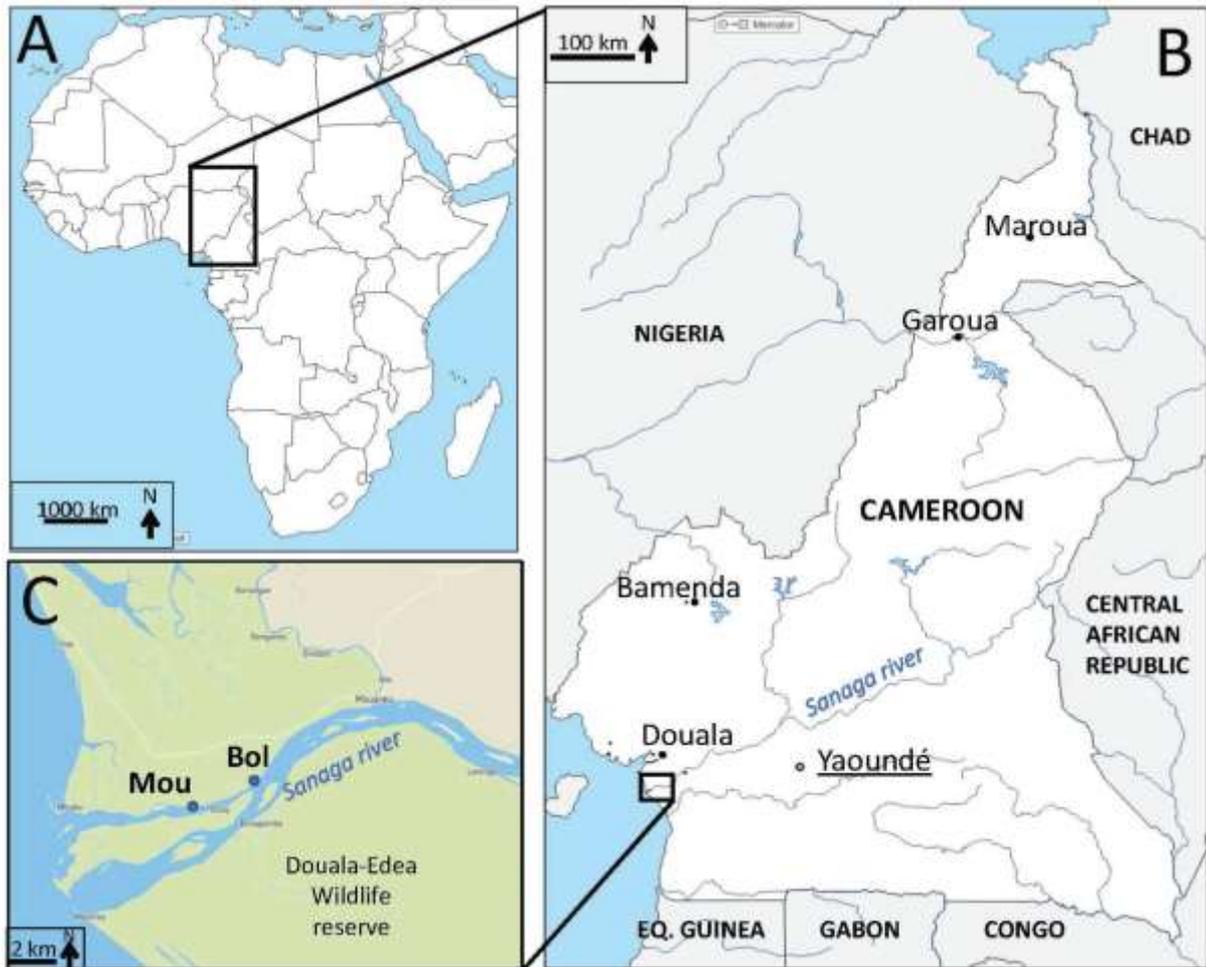
les dauphins, divers poissons, les crocodiles et les lamantins d'Afrique de l'Ouest. On note également l'établissement des populations humaines le long du fleuve et de ses îles, qui pêchent, cultivent et extraient du sable en fonction des inondations fluviales saisonnières et vivant principalement de l'exploitation des bivalves de décembre à juin, qui sont transformés en viande et en coquilles pour les industries environnantes de transformation de la céramique et aliments pour animaux (Ajonina *et al.*, 2005).

### **2.2.3 Détection, quantification et géotypage des norovirus et le virus de l'hépatite A contaminant les palourdes de la Sanaga**

Comme présenté dans l'introduction, le fleuve Sanaga serait soumis à une contamination d'origine humaine et animale. Nous nous sommes intéressés dans cette partie à la détection, la quantification et le géotypage des norovirus et du virus de l'hépatite A qui constituent les principaux virus à transmission d'origine alimentaire transmis par les mollusques bivalves.

#### **a. Echantillonnage des palourdes**

Un total de 675 palourdes de l'espèce *Galatea paradoxa* (Born 1778) constituées en 45 échantillons composés de 15 palourdes chacun a été collecté de février 2018 à février 2020, dont 23 échantillons à Bolounga (Bol) et 22 échantillons à Mouloungo (Mou). Ces deux points distants de 2,5 km et situés à 12 km de l'embouchure du fleuve Sanaga dans le Canton Malimba de l'arrondissement de Mouanko (3°32'20.0"N 9°39'02.0"E) constituent les principaux points de pêche des palourdes (figure 20). Les échantillons ont été transportés dans des glacières au laboratoire de microbiologie de la Faculté des Sciences de l'Université de Yaoundé 1 où ils ont été conservés à -20°C, puis acheminés dans les mêmes conditions au Laboratoire Santé, Environnement et Microbiologie de l'Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER) de Nantes (France).



**Figure 20. Localisation des points de prélèvement.**

- A : Carte de l'Afrique (d-maps.com) avec le Cameroun dans la région centrale.
- B : Carte (d-maps.com) du Cameroun (blanc) avec les frontières (traits noirs), les pays voisins (gris), la capitale (cercle noir, souligné) et les principales villes (points noirs), les principaux fleuves et lacs (bleu) dont le fleuve Sanaga. Eq. Guinea = Guinée Equatoriale.
- C : Zoom sur l'estuaire du fleuve Sanaga dans la réserve faunique Douala-Edéa (vert clair) avec localisation des points d'échantillonnage des palourdes : Bolounga (Bol), Moulounga (Mou).

### **b. Traitement des échantillons de palourdes**

Les différents échantillons de palourdes ont été lavés afin de débarrasser les coquilles de la vase et du sable, puis ouverts à l'aide d'un couteau stérile. Les tissus digestifs (TD) ont été disséqués, hâchés, homogénéisés, puis répartis en aliquotes de 2 g et immédiatement stockés à  $-20^{\circ}\text{C}$  jusqu'à l'analyse (Le Guyader *et al.*, 2009).



**Photo 2. Extraction des tissus digestifs (Bonny Patrice, 2020)**

### **c. Détection et quantification des NoV et du VHA**

La détection et quantification des NoV et du VHA ont été réalisées selon le protocole décrit dans la norme ISO/TS 15216-1 : 2017.

#### **– L'élution virale**

Deux (02) g de TD de chaque échantillon ont été introduits dans des tubes à centrifuger de 15 mL contenant chacun 200  $\mu$ L d'une solution de protéinase K (3 U/mL) à laquelle 1,8 mL d'eau stérile ont été ajoutés. Puis, 100  $\mu$ L d'une suspension de mengovirus (contrôle du processus d'élution) de concentration  $10^6$  TCID<sub>50</sub>/mL (*Median Tissue Culture Infectious Dose* : dilution d'une suspension virale nécessaire pour infecter 50% des puits d'une culture cellulaire) ont été introduits. Les tubes ont été homogénéisés (au vortex) à vitesse maximale pendant 10 à 15 secondes, et incubés à 37 °C pendant 15 min sous agitation (à 32 rotations/minute), puis à 60°C dans les mêmes conditions. Après cette double incubation, les tubes ont été centrifugés à 3000×g pendant 5 min à 4°C, et les surnageants transférés dans de nouveaux tubes de 15 ml pour l'extraction des acides nucléiques.

#### **– Extraction des acides nucléiques**

Dix (10) ml de tampon de lyse constitué de thiocyanate de guanidium ont été introduits dans les tubes contenant les surnageants des différents échantillons. Parallèlement, un contrôle de processus d'extraction a été préparé par ajout de 100  $\mu$ l de la suspension de mengovirus précédemment préparée dans un tube contenant 10 ml de tampon de lyse. Les tubes ont ensuite été homogénéisés et incubés à température ambiante pendant 10 min. Puis 140  $\mu$ l de silice

magnétique ont été introduits dans les différents tubes qui ont été incubés à température ambiante pendant 10 min avant d'être placés dans des portoirs magnétiques afin de collecter la silice magnétique. Les surnageants ont été éliminés et la silice magnétique récupérée et lavée à l'aide d'une rampe magnétique miniMAG et du kit d'extraction NucliSens (bioMérieux, France) selon les prescriptions du fabricant. Les acides nucléiques ont été recueillis dans 100 µL de tampon d'élution et conservés à -80 °C.

– **Détermination de l'efficacité d'extraction des acides nucléiques**

Il s'agissait d'évaluer la capacité de recouvrer les ARN du mengovirus ajouté aux échantillons lors de l'élution virale et de l'extraction des acides nucléiques.

A cet effet, 20 µL de mélange réactionnel de RT-qPCR (*one-step real-time* RT-PCR kit, invitrogen) (Tableau 3) ont été introduits dans les puits d'une plaque de PCR, puis 5 µL des extraits d'acides nucléiques des échantillons ont été ajoutés. Simultanément, un témoin négatif et une gamme étalon en quatre points de concentrations en ARN de mengovirus ont été constitués par ajout respectif de 5 µL d'eau stérile ou d'une série de dilutions décimales d'une suspension d'ARN de mengovirus à 20 µL de mélange réactionnel de RT-qPCR. Le couple d'amorces et la sonde décrits dans le tableau 4 ont été utilisés.

**Tableau 3. Composition du mélange réactionnel de la RT-qPCR**

Réactifs	Concentration finales (dans 25 µL)	Volumes (µL)
H <sub>2</sub> O	/	9,1
5X mélange réactionnel Ultrasense	1X	5
Amorce antisens (10 pmol/µL)	0,9 pmol/µL	2,25
Amorce sens (10 pmol/µL)	0,5 pmol/µL	1,25
Sonde (1 pmol/µL)	0,25 pmol/µL	0,6
Rox (50X)	1X	0,5
Enzyme SuperScript II (200U/µL)	10,4 U/µL	1,3
Extrait d'ARN	/	5
Total	/	25

**Tableau 4. Liste des différents couples d'amorces et des sondes utilisés pour la RT-qPCR**

Virus	Amorces	Séquences	Auteurs
Mengo	Mengo 110 (+) Mengo 209 (-) Mengo 147 (S)	gcg ggt cct gcc gaa agt gaa gta aca tat aga cag acg caca c FAM-ata aca tta ctg gcc gaa gc-MGB	Pinto <i>et al.</i> , 2009
NoV GI	QNIF4 (+) NV1LCR (-) NVGG1p (S)	cgc tgg atg cgn ttc cat cct tag acg cca tca tca tt tac FAM-tgg aca gga gay cgc rat ct-BHQ	Le Guyader <i>et al.</i> , 2008
NoV GII	QNIF2d (+) COG2R (-) QNIFs (S)	atg ttc agr tgg atg agr ttc tcw ga tcg acg cca tct tca ttc aca FAM-agc acg tgg gag ggc gat cg- BHQ	Le Guyader <i>et al.</i> , 2009
VHA	VHA68 (+) VHA240 (-) VHA150 (S)	tca ccg ccg ttt gcc tag gga gag ccc tgg aag aaa g FAM-cct gaa cct gca gga att aa-MGB	Costafreda <i>et al.</i> , 2006

(+) : Amorce sens ; (-) : Amorce antisens ; (S) : Sonde

La RT-qPCR en une étape a été réalisée à l'aide du thermocycleur MX3000P *Quantitative PCR system* (Agilent Technologies France) suivant le profil thermique décrit dans le tableau 5.

**Tableau 5. Profil thermique de la RT-qPCR en une étape, utilisé pour l'efficacité d'extraction du mengovirus**

Description de l'étape		Température et durée	Nombre de cycles
RT		55 °C, 15 min	1
Préchauffage		95 °C, 5 min	1
Amplification	Dénaturation	95 °C, 15 sec	45
	Hybridation	60 °C, 30 sec	
	Elongation	65 °C, 30 sec	

Le pourcentage de l'efficacité d'extraction a été déterminé par la formule :

$$\%eff = 10^{(Ct_u - b)/a} / 10^{Ct_{g1}/a}$$

Avec :

%*eff* : Pourcentage de l'efficacité d'extraction

C<sub>tu</sub> : Ct de l'échantillon utilisé (Ct : cycle seuil de détection de la fluorescence)

C<sub>g1</sub> : Ct du premier point de gamme

b : Ordonnée à l'origine de la droite d'étalonnage

a : Pente de la droite d'étalonnage

Seuls les extraits ayant un pourcentage d'extraction du mengovirus supérieur à 1% ont été retenus pour la quantification.

#### – Détection et quantification des NoV et du VHA

Afin de procéder à la détection et la quantification des NoV et du VHA, 20 µL du mélange réactionnel de RT-qPCR (Tableau 3) ont été introduits dans les puits d'une plaque de PCR, puis 5 µL des extraits d'acides nucléiques des différents échantillons ont été ajoutés. Simultanément, des témoins négatifs et des gammes étalons en quatre points de concentrations en ARN synthétiques des différents virus (NoV GI et GII ou VHA) ont respectivement été constitués par ajout de 5µL d'eau stérile ou des séries de dilutions décimales des suspensions d'ARN viraux à 20 µL de mélange réactionnel de RT-qPCR.

Les couples d'amorces correspondant (Tableau 4) ont été utilisés pour les différents virus recherchés. Le couple d'amorces utilisé pour la quantification du génogroupe I des NoV amplifiait un produit de 86 pb correspondant aux nucléotides 5291-5376 du virus Norwalk (numéro d'accès GenBank M87661), tandis que le couple d'amorces du génogroupe II amplifiait un produit de 89 pb correspondant aux nucléotides 5012-5100 du virus de Lordsdale (numéro d'accès GenBank X86557). La zone sélectionnée pour la détection des NoV était une région bien conservée à l'extrémité 5' de l'ORF2 codant la protéine majeure VP1 de la capsidie et adéquate pour la quantification de tous les génotypes des génogroupes I et II. Quant au VHA, le couple d'amorces utilisée amplifiait un produit de 173 pb correspondant aux nucléotides 68-240 de l'isolat VHA HM174 43c (numéro d'accès GenBank M59809). Région du génome caractérisée par une absence de variabilité et adéquate pour la quantification de tous les génotypes du VHA (ISO/TS 15216-1 : 2017).

La RT-qPCR en une étape a été réalisée à l'aide d'un thermocycleur MX3000P *Quantitative PCR system* (Agilent Technologies France) suivant les profils thermiques décrits dans le tableau 6.

**Tableau 6. Profil thermique de la RT-qPCR en une étape, utilisé pour la détection/quantification des NoV GI, NoV GII, et le VHA**

Description de l'étape		Température et durée		Nombre de cycles
RT		55 °C, 15 min		1
Préchauffage		95 °C, 5 min		1
Amplification	Dénaturation	95 °C, 15 sec		45
	Hybridation	60 °C, 30 sec (NoV GII, VHA)	55 °C, 30 sec (NoV GI)	
	Elongation	65 °C, 30 sec		

Les concentrations en copies d'ARN des différents virus ont été déterminées à l'aide de la courbe d'étalonnage :  $\text{copies/g} = (10^{(Ct-b)/a}) * 100 * (\text{volume total}/0,5) * (1/2)$  décrite en annexe 1.

#### **d. Génotypage des NoV et du VHA détectés**

Les NoV et le VHA sont caractérisés par une grande diversité génétique, avec plusieurs génotypes infectant des hôtes non humains à l'instar des félins, des canins, des primates, des bovins, et des chauves-souris. Aussi avons-nous entrepris d'effectuer un génotypage des NoV et VHA détectés dans les palourdes de la Sanaga afin de déterminer l'origine de la contamination virale de ces bivalves qui constituent une ressource alimentaire importante pour les populations.

##### **– Amplification des gènes cibles**

Elle a porté sur les régions du génome codant la protéine majeure (VP1) de la capsidie et l'ARN polymérase ARN dépendant (RdRp) des NoV GI et GII (figure 3), ainsi que la jonction VP1-2A des régions du génome codant la protéine VP1 de la capsidie et la protéine non structurale 2A du virus de l'hépatite A (figure 5).

- **Rétrotranscription (RT)**

Elle avait pour but de générer les ADNc des régions ciblées spécifiques aux différents virus recherchés à partir des extraits d'acides nucléiques.

Pour y parvenir, 7 µL de mélange réactionnel A (Tableau 7) constitué d'amorces RT présentés dans le tableau 8 ont été introduits dans les puits d'une barrette de PCR. Puis 5 µL de chaque extrait ont été ajoutés dans des puits respectifs. Les barrettes ont été placées dans des thermocycleurs et chauffées à 65 °C pendant 5 min, puis immédiatement placées dans de la

glace pendant 3 min afin de dénaturer les ARN et obtenir des molécules linéaires favorables à l'hybridation des amorces de la rétrotranscription. Ensuite, 8 µL de mélange réactionnel B (Tableau 9) ont été introduits dans les différents puits. Les barrettes ont enfin été placées dans le thermocycleur pour une rétrotranscription selon le profil thermique : 30 min (55 °C) ; 15 min (70 °C).

**Tableau 7. Mélange réactionnel A de la rétrotranscription**

Réactions	Concentrations finales	Volumes/réaction (µL)
H <sub>2</sub> O	/	6
Amorce RT (50µM)	2,5 µM	1

**Tableau 8. Liste des différents couples d'amorces et des sondes utilisés pour le typage**

Cibles	Virus	Amorces	Séquences	Auteurs
Capside	Norovirus GI	GISKR (RT)	cca acc car cca ttr tac a	Nishida <i>et al.</i> , 2003
		QNIF4 (1 <sup>ère</sup> PCR)	cgc tgg atg cgn ttc cat	Le Guyader <i>et al.</i> , 2009
		GISKF (1/2 nichée)	ctg ccc gaa tty gta aat ga	Nishida <i>et al.</i> , 2003
	Norovirus GII	GIISKR (RT)	ccr ccn gca trh ccr ttr tac at	Nishida <i>et al.</i> , 2003
		QNIF2d (1 <sup>ère</sup> PCR)	atg ttc agr tgg atr agr ttc tcw ga	Le Guyader <i>et al.</i> , 2009
		GIISKF (1/2 nichée)	cnt ggg agg geg atc gca gca a	Nishida <i>et al.</i> , 2003
Polymérase	Norovirus GI, GII	P110 (RT)	acd aty tca tca tca cca ta	Le Guyader <i>et al.</i> , 2000
	Norovirus GI	P36bp (1 <sup>ère</sup> PCR)	gat gcd gat tac aca gch tgg g	Jiang <i>et al.</i> , 1992
		NV4562 (1/2 nichée)	gat gcd gat tac aca gch tgg g	Yuen <i>et al.</i> , 2001
	Norovirus GII	NV4611bp (1 <sup>ère</sup> PCR)	cwg cag cmc tdg aaa tca tgg	Yuen <i>et al.</i> , 2001
		NV4692 (1/2 nichée)	gtg tgr tkg atg tgg gtg act t	Yuen <i>et al.</i> , 2001
VP1-2A	VHA	3381 (RT)	cca tt caa gag tcc aca cac t	Hutin <i>et al.</i> , 1999
		2870 (1 <sup>ère</sup> PCR)	gac aga ttc tac att tgg att ggt	
		2897 (1/2 nichée)	cta ttc aga ttg caa att aca at	

**Tableau 9. Mélange réactionnel B de la rétrotranscription**

Réactions	Concentrations finales	Volumes/réaction (µL)
First-Strand Buffer 5X	1	4
dNTP (10 mM)	0,5 mM	1
DTT (0,1M)	5 mM	1
RNaseOUT (40 unités/µL)	2 unités/µL	1
SuperScript III (200 unités/µL)	10 unités/µL	1
Total	/	8

- **Réalisation de la PCR classique (1<sup>ère</sup>PCR)**

A l'issue de la rétrotranscription, 30 µL de mélange réactionnel (Tableau 10) contenant les amorces de la 1<sup>ère</sup>PCR (Tableau 8) ont été ajoutés dans les différents puits. Ces derniers ont ensuite été placés dans un thermocycleur, et les ADNc amplifiés selon le profil thermique décrit dans le tableau 11.

**Tableau 10. Mélange réactionnel de PCR classique (1<sup>ère</sup>PCR)**

Réactifs	Concentrations finales	Volumes/réaction (µL)
H <sub>2</sub> O	/	21,3
10X PCR Buffer, Minus Mg	1	5
dNTP (10 mM)	0,2 mM	1
MgCl <sub>2</sub> (50 mM)	1,5 mM	1,5
Amorce PCR (10 µM)	0,2 µM	1
Platinum Taq DNA polymérase (250 unité/tube)	1 unité/tube	0,2
Total	/	30

**Tableau 11. Profil thermique de la PCR classique (1<sup>ère</sup>PCR)**

Description de l'étape		Température et durée	Nombre de cycles
Préchauffage		95 °C, 1 min	1
Amplification	Dénaturation	95 °C, 30 sec	40
	Hybridation	50 °C, 30 sec	
	Elongation	72 °C, 30 sec	
	Elongation finale	72 C, 7 min	1

- **PCR semi-nichée**

Deux (02)  $\mu\text{L}$  des différents amplicons de la PCR précédente ont respectivement été introduits dans des puits de PCR contenant  $48\mu\text{L}$  du mélange réactionnel de la PCR  $\frac{1}{2}$  nichée (Tableau 12) contenant les amorces de la RT et celles de la PCR  $\frac{1}{2}$  nichée (Tableau 8). L'ensemble a été brièvement homogénéisé. Les tubes de PCR ont ensuite été placés dans un thermocycleur, et l'amplification s'est faite selon le profil thermique présenté dans le tableau 11. Les produits de PCR ont été conservés à  $-20\text{ }^{\circ}\text{C}$  pour les analyses ultérieures.

Les tailles des produits attendus à l'issue de chacune des PCR sont consignées dans le tableau 13.

**Tableau 12. Mix de PCR semi-nichée ( $\frac{1}{2}$  nichée)**

Réactifs	Concentrations finales	Volumes/réaction ( $\mu\text{L}$ )
$\text{H}_2\text{O}$	/	38,3
10X PCR Buffer, Minus Mg	1	5
dNTP (10 mM)	0,2 mM	1
$\text{MgCl}_2$ (50 mM)	1,5 mM	1,5
Amorce RT (10 $\mu\text{M}$ )	0,2 mM	1
Amorce PCR $\frac{1}{2}$ nichée (10 $\mu\text{M}$ )	0,2 mM	1
Platinum Taq DNA polymérase (250 unité/tube)	1 unité/tube	0,2
Total	/	48

**Tableau 13. Tailles des différents produits de PCR attendus.**

PCR	Tailles des fragments (pb)				
	Capside		Polymérase		VP1-2A
	NoV GI	NoV GII	NoV GI	NoV GII	VHA
1 <sup>ère</sup>	380	377	397	273	512
$\frac{1}{2}$ nichée	329	343	322	192	392

– **Clonage, séquençage et analyse des séquences**

Des amplicons provenant d'échantillons positifs pour le ou les virus recherchés ont été clonés en utilisant le vecteur plasmidique pCR4-TOPO-TA et *One Shot® TOP10 Chemically Competent Escherichia coli* (ThermoFisher). La purification des plasmides a été effectuée en utilisant le kit QIAprep Spin Miniprep (Qiagen) selon les prescriptions du fabricant. Les

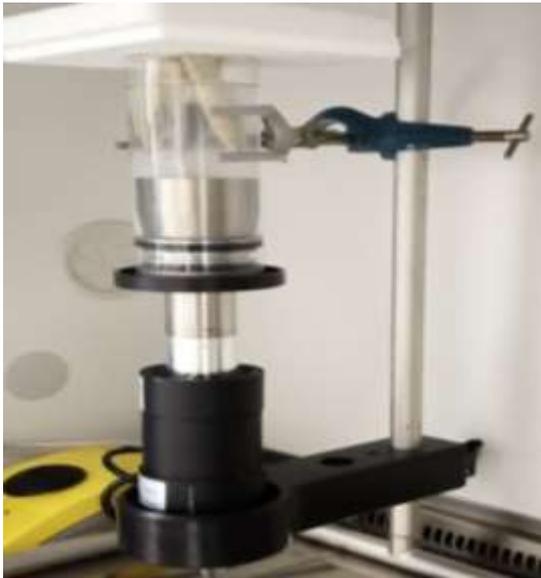
plasmides purifiés ont été séquencés à l'aide du séquenceur ABI 3730xl (Applied Biosystems), en utilisant les amorces spécifiques de la RT et de la PCR  $\frac{1}{2}$  nichée décrites dans le tableau 8. Les séquences ont été génotypées à l'aide des outils en ligne *Norovirus Typingtool* et *HAVNet Typingtool* développés par la *National Institute for Public Health and the Environment (RIVM)* des Pays-Bas (Kroneman et al. 2011 ; Kroneman et al., 2018), et identifiées par *BLASTn* sur la base de données nucléotidiques (nt) de NCBI bank.

#### **2.2.4 Détermination de la diversité des virus à ARN présents dans les palourdes de la Sanaga**

Le gisement des palourdes de la Sanaga se trouve dans une zone rurale située au sein d'une réserve naturelle. Dans cette aire protégée, se côtoie une abondante et grande diversité de populations animales et humaines favorable à une circulation inter-espèces de microorganismes, et à l'émergence de nouveaux agents pathogènes humains. Aussi nous sommes nous proposés d'analyser la diversité des virus à ARN contaminant les palourdes de la Sanaga par une approche métagénomique.

##### **a. Éluition virale**

Les tissus digestifs (TD) des échantillons sélectionnés sur la base de la détection des NoV et du VHA, ainsi que la représentativité des différents mois de l'année et des points d'échantillonnage ont été incubés dans 2 mL d'une solution de 30 U/mg de protéinase K (Sigma-Aldrich, France / Thermo Fischer Scientific) pendant 15 min à 37 °C sous agitation de 32 rotations/min, puis pendant 15 min à 60 °C. Pour chaque échantillon, le mélange a été soumis à une sonication (Bandelin UD 2200) à la puissance maximale (100%) pendant 3 min, puis centrifugé à 3000 × g pendant 5 min. Le surnageant a été mesuré et transféré dans des tubes de 15 mL. Une solution de pyrophosphate de sodium (1/10 du volume de surnageant) de 100 mM a été ajoutée dans les différents tubes, et les mélanges ont été incubés à 5 °C pendant 40 min sous agitation, puis centrifugés à 8000 × g pendant 20 min à 4 °C. Le surnageant (environ 3 mL) a été récupéré, mélangé à 1,5 mL de Polyéthylène Glycol (PEG) 24% (poids / vol) - Chlorure de Sodium (1,2 M) et incubé à 4 °C pendant 1 heure sous agitation. Après centrifugation à 10 000 × g pendant 20 min, le culot a été remis en suspension dans 2 mL de tampon glycine (0,05 M ; pH 9) préchauffés à 56 °C. La solution obtenue a été filtrée à l'aide de filtres en cellulose acétate de 5 ; 1,5 et 0,45 µm. Le filtrat a enfin été traité avec 20 µL d'Endonucléase OmniCleave™ (200 U / µL) et 200 µL de MgCl<sub>2</sub> (100 mM) pendant 1 heure à 37 °C.



**Photo 3. Sonication des hydrolysats de tissus digestifs (Bonny Patrice, 2020).**

#### **b. Extraction et purification des ARN viraux**

Dix (10) mL de tampon de lyse constitué de thiocyanate de guanidium ont été introduits dans les tubes contenant les filtrats des différents échantillons qui ont ensuite été homogénéisés et incubés à température ambiante pendant 10 min. Puis 140  $\mu\text{L}$  d'une suspension de silice magnétique ont été introduits dans les différents tubes qui ont ensuite été incubés à température ambiante pendant 10 min. La silice magnétique a ensuite été récupérée à l'aide d'un collecteur magnétique, et l'extraction des acides nucléiques a été réalisée à l'aide du kit NucliSens et du système de lavage semi-automatisé eGENE-UP<sup>TM</sup> (bioMérieux, France). Les extraits d'acides nucléiques ont ensuite subi un traitement à la DNase pendant 30 min à 37 ° C avec 12,5  $\mu\text{L}$  de TURBO<sup>TM</sup> DNase 2 U/ $\mu\text{L}$  (invitrogen / Ambion, Thermo Fischer Scientific, France). Après une purification supplémentaire à l'aide du kit RNA Clean - Concentrator<sup>TM</sup>-5 (Zymo Research, Irvine USA) selon les prescriptions du fabricant, les ARN ont été recueillis dans 100  $\mu\text{L}$  d'eau de qualité biologie moléculaire et conservés à -80 °C pour des analyses ultérieures.

Les extraits ont été testés par RT-qPCR en utilisant les amorces et les sondes ciblant la région conservée entre ORF1 et ORF2 des norovirus GII en utilisant le kit RT-qPCR *one-stape* (invitrogen) et le système de PCR en temps réel AriaMx (Agilent Technologies, France).



**Photo 4. Extraction et purification des acides nucléiques à l'aide du système d'extraction semi-automatisé eGENE-UP™ (Bonny Patrice, 2020).**

### c. Préparation des bibliothèques d'ADN et séquençage

#### – Génération des ADN complémentaires (ADNc) et optimisation de la synthèse d'ADNc double brins

Les extraits d'ARN ont été transcrits en triplets d'ADNc à l'aide de la Superscript II (invitrogen, France) et des amorces hexamères aléatoires *Random Hexamer* (invitrogen, France). A cet effet, 15 µL d'extrait d'ARN ont été introduits dans des tubes de PCR contenant 10 µL de mélange réactionnel I (Tableau 14). Après 2 min d'incubation à température ambiante, 20 µL de milieu réactionnel II (Tableau 15) ont été ajoutés dans les tubes qui ont ensuite été incubés suivant le profil thermique : 37 °C pendant 90 min, puis 70 °C pendant 20 min.

**Tableau 14. Mélange réactionnel I pour la synthèse des ADNc**

Réactifs	Volume par réaction (µL)
Hexamère non-ribosomique (100 µM)	4,6
10X <i>E. coli ligase buffer</i>	3
MgCl <sub>2</sub> (100 mM)	2,4
<b>Total</b>	<b>10</b>

**Tableau 15. Mélange réactionnel II pour la synthèse des ADNc**

Réactifs	Volume par réaction (µL)
10X E. coli <i>ligase buffer</i>	2
SuperScript II (200 U/µL)	1
dNTPs (25 mM pour chaque dNTP)	1
<i>Nuclease-free Water</i>	15
DTT	1
<b>Total</b>	<b>20</b>

Les ADNc ainsi produits ont été confirmés par PCR en temps réel (qPCR) à l'aide du mélange réactionnel, des amorces et sondes décrits respectivement dans les tableaux 3 et 4.

La synthèse du second brin d'ADNc a ensuite été effectuée à l'aide du kit *NEBNext Ultra RNA Library Kit* pour Illumina (New England Biolabs, France). A cet effet, 30 µL d'ADNc ont été transférés dans des tubes de PCR contenant 25 µL de mélange réactionnel (*Second Strand cDNA Synthesis Master mix*) (Tableau 16) et incubés à 16 °C pendant 1 heure. Les ADNc double brins ainsi synthétisés ont été quantifiés à l'aide du Qubit (Annexe 2).

**Tableau 16. Mélange réactionnel pour la synthèse du second brin de l'ADNc**

<i>Second strand cDNA Synthesis Master mix</i>	Volume/réaction (µL)
H <sub>2</sub> O	15,6
<i>Second Strand Synthesis Reaction Buffer (10X)</i>	6,2
<i>Second Strand Synthesis Enzyme Mix</i>	3,2
<b>Total</b>	<b>25</b>

#### – Préparation des librairies et *pools* de librairies

La préparation des librairies d'ADN (triplets de librairies par échantillon) pour le séquençage NGS Illumina a été faite selon le protocole décrit par Strubbia et *al.*, 2020. Pour cela, 50 µL d'ADNc ont préalablement été fragmentés par ultrasons (Ultrasonicator M220, Covaris) pendant 110 sec, puis les librairies ont été préparées à l'aide du kit KAPA Prep (Roche). Après fragmentation des ADNc, suivie d'une réparation des extrémités 3' et 5' de manière à générer des bouts francs, et fixation d'un résidu d'adénine à l'extrémité 3' (Annexe 3), des adaptateurs spécifiques (SeqCap Library Adapter) à chaque librairie ont été ligués à chaque extrémité des séquences d'ADN (Annexe 4). Plusieurs étapes de lavage à l'aide de la silice magnétique (*AmPure XP Beads*), d'un collecteur magnétique et de l'éthanol à 80° ont

permis de sélectionner les séquences d'ADN de tailles comprises entre 150 et 500 pb avec un *peak* à 320 pb (Annexe 5). Les librairies ont ensuite été amplifiées par PCR, lavées (Annexe 6) et quantifiées à l'aide du Qubit (Annexe 2). Une évaluation du profil de taille des séquences a été réalisée à l'aide d'un bio-analyseur (*Agilent 2100 Bioanalyzer system*).

Sept (07) *pools* d'un mélange équimassique de 10 librairies (à l'exception du *pool* VII constitué de 7 librairies) ont été constitués de manière à obtenir une masse d'acide nucléique de 1 µg.

#### – Capture des séquences d'intérêt et séquençage

Les séquences virales des différents *pools* ont été capturées et concentrées à l'aide du kit VirCapSeq-VERT capture panel (Roche) (Wylie *et al.*, 2015).

Dans les tubes de 1,5 mL contenant les différents *pools*, 5 µL de *COT Human DNA* (1mg/mL), 5 µL de *HyperCap Universal Blocking Oligos* et 10 µL de billes magnétiques AMPure préalablement ramenées à température ambiante ont été ajoutés. L'ensemble a été homogénéisé et incubé à température ambiante pendant 10 min afin de permettre la fixation des librairies aux billes. Les tubes ont ensuite été placés sur un portoir magnétique afin de capturer les billes. Après élimination des surnageants et lavage des billes avec 190 µL d'une solution d'éthanol à 80 %, 10,5 µL de mélange d'hybridation (Annexe 7) ont été ajoutés dans chaque tube. L'ensemble a été transféré dans des tubes de 200 µL, puis homogénéisé par vortex et laissé à température ambiante pendant 2 min. Les tubes ont ensuite été placés dans un portoir magnétique, et les surnageants transférés dans des tubes contenant 4,5 µL d'un mélange de sondes virales (SeqCap EZ Probe pool).

Le mélange a été homogénéisé et incubé à 95 °C (5 min), puis 47 °C (72 heures) pour permettre l'hybridation entre les sondes virales et les séquences des virus de vertébrés. Les séquences hybridées ont ensuite été capturées à l'aide des billes de capture spécifique du kit *HyperCap Bead Kit*. A cet effet, le mélange d'hybridation a été introduit dans des tubes contenant les billes de capture préalablement préparées selon les recommandations du fabricant. Puis, l'ensemble a été homogénéisé au vortex et incubé dans un thermocycleur à 47 °C pendant 15 min (température du couvercle réglé à 57 °C).

Les billes de capture ont ensuite été lavées à l'aide de différents tampons de lavage et amplifiées par PCR. Après un dernier lavage à l'aide des billes magnétiques et l'éthanol à 80°, les *pools* de librairies ont été récupérés dans 50 µL de tampon Tris-HCl (10 nM, pH 8,0). La quantité et la qualité des séquences dans les *pools* de librairies ont été évaluées respectivement à l'aide du Qubit et du Bio-analyseur afin de s'assurer que les concentrations sont comprises entre 6-13 nM et la distribution des tailles entre 150-500 pb avec un *peak* à 320 pb.

Le Séquençage à Haut Débit Illumina a été réalisé à l'Institut du Cerveau et de la Moelle de l'Hôpital de la Pitié Salpêtrière de Paris, France sur la plateforme Illumina NovaSeq 6000 à l'aide du kit *Illumina NovaSeq 6000 Reagent Kit* de façon à générer des *reads* de 2 x 250 pb (*Paired-End Sequencing*).

#### **d. Traitement bio-informatique des données de séquençage**

L'analyse bio-informatique des données de séquençage a été effectuée à l'aide de *nextflow* sur le serveur de l'Ifremer. Les triplets de fichiers *Fastq* issus du séquenceur ont été regroupés au début de l'analyse. L'analyse bio-informatique des données a débuté par une phase de prétraitement. A cet effet, un filtrage de la qualité des *reads* a été effectué à l'aide de *TrimGalore* avec un seuil de qualité fixé à 25, suivi de la suppression des adaptateurs effectuée à l'aide du même outil. Un second filtrage a ensuite été effectué afin d'éliminer les *reads* non-viraux. Pour cela, les *reads* ont été alignés sur la base de données *Silva RNA* à l'aide du logiciel *bowtie2* (Langmead *et al.*, 2012). Les *reads* s'alignant sur les séquences d'ARN ribosomiques de l'hôte ou bactériens ont été supprimés (Nooij *et al.*, 2018). La dernière étape du prétraitement a consisté en une déduplication en utilisant *cd-hit-dup* (Fu *et al.*, 2012). Ce traitement avait pour objectif d'éliminer les doublons de PCR afin de simplifier l'ensemble de données, de faciliter l'étape d'assemblage et de réduire le biais de la PCR lors du calcul de l'abondance.

Une analyse de type *Kaiju* a été effectuée sur les données brutes et les données post-déduplication en utilisant la base de données *nr\_euk* pour évaluer la composition des échantillons et l'impact du prétraitement (filtrage, suppression d'ARNr et déduplication).

Un assemblage *de novo* a été réalisé à l'aide de *metaSPAdes* avec les *kmer* 21, 33, 55, 77, 99 (Nurk *et al.*, 2017). Seuls les *contigs* de plus de 300 pb ont été retenus et identifiés à l'aide de *BLASTn* 2.6.0 (E-value 1E-5) sur la base de données nucléotidiques (nt) de NCBI (Camacho *et al.*, 2009) et de DIAMOND (E-value 1E-3) sur la base de données des protéines non redondantes (nr) de NCBI (Buchfink *et al.*, 2015). Lorsque les deux méthodes donnaient des résultats, le résultat du *BLASTn* était sélectionné.

Pour évaluer la couverture de chaque *contig*, les *reads* issus du prétraitement ont été alignés à l'aide du logiciel *bowtie2* sur les *contigs* assemblés. Les *reads* identiques ont été supprimés comme source de surestimation potentielle de l'abondance. L'identification taxonomique a été réalisée à l'aide de l'outil *Entrez direct* et le *taxid* (identité taxonomique) permettant d'extraire des informations à un niveau taxonomique défini. Les familles ont été

déterminées sur le fichier *krona* en prenant toutes les familles qui représentent plus de 0,01 % des *reads* viraux dans au moins l'un de nos échantillons.

La *Heatmap* a été réalisée à l'aide du logiciel R, et les *reads* par millions (abondance relative) ont été calculés à l'aide du nombre de *reads* par famille et du nombre total de *reads* issus des prétraitements.

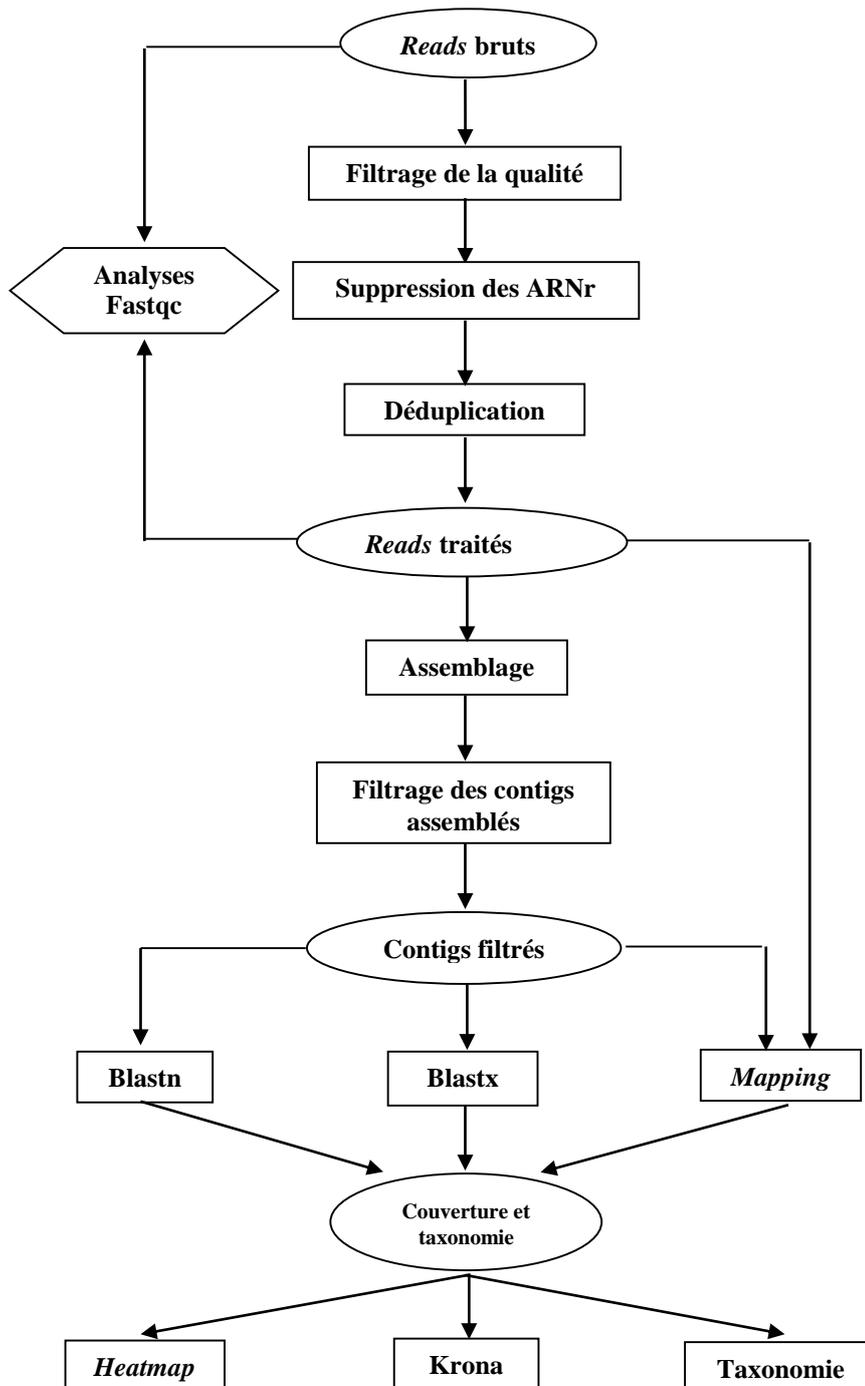


Figure 21. Pipeline du traitement bio-informatique.

## 2.2.5 Identification des virus à ARN constituant un danger chez les humains

Depuis la décennie 2000, on observe une augmentation significative des agents infectieux émergents et des risques de nouvelles pandémies comme illustré par la propagation du virus pathogène influenza H5N1 depuis 2003, H1N1 en 2009, H7N9 en 2013, le SRAS, le MERS, le chikungunya, la dengue, le Zika et l'Ebola (Osterhaus *et al.*, 2020). Par ailleurs, la COVID-19, première grande pandémie du nouveau millénaire a clairement démontré que des nouveaux agents pathogènes peuvent émerger à tout moment, de n'importe quel endroit, sans avertissement, et menacer la santé, le bien-être et les économies de toutes les sociétés. D'où la nécessité évidente pour les pays d'avoir la capacité de maintenir un système d'alerte et de réponse efficace permettant de détecter et de réagir rapidement aux flambées, et de partager des informations (Osterhaus *et al.*, 2020). Aussi nous sommes nous proposés dans cette dernière partie de rechercher et d'identifier les séquences de virus à ARN humains et animaux pouvant présenter un intérêt en santé humaine.

Après l'étape d'assemblage, les *contigs* viraux de plus de 300 nucléotides, assemblés à partir d'au moins cinq *reads* et ayant une *E-value*  $\leq 10^{-15}$  lors du BLAST ont été sélectionnés pour l'identification des virus susceptibles de constituer un danger chez les humains. Cette stratégie a été utilisée pour identifier des séquences appartenant aux familles *Astroviridae*, *Caliciviridae*, *Hepeviridae*, *Picobirnaviridae* et *Picornaviridae* aux seins desquelles se trouve la majorité des virus entéro-pathogéniques. Les séquences sélectionnées ont été annotées et déposées dans la base de données GenBank de NCBI. Une analyse phylogénétique basée sur la capsid ou et/ou la RdRp a enfin été effectuée pour les séquences appartenant à des virus caractérisés par une grande diversité génétique ou d'hôte. A cet effet, les séquences protéiques ont été alignées sur une longueur d'au moins 150 acides aminés à l'aide des outils *Muscle* et *Clustal Omega* implémentés dans *geneious Prime* version 2021.0.3. Les arbres phylogénétiques ont ensuite été construits à l'aide de l'outil en ligne *IQtree* avec un *bootstrap* de 1000. Les arbres ont été enracinés à partir du point moyen à l'aide de l'outil *FigTree v.1.4.3*. Seuls les *bootstrap* supérieurs à 70% ont été présentés sur les branches.

## **CHAPITRE III : RESULTATS ET DISCUSSION**

### **3.1. Résultats**

#### **3.1.1. Contamination virale des palourdes de la Sanaga**

##### **a. Prévalence des norovirus et du virus de l'hépatite A dans les palourdes de la Sanaga**

La qualité microbiologique des palourdes étant un aspect principal de notre étude, la première étape de ce travail consistait à détecter, quantifier et génotyper les NoV et le VHA qui sont les principaux virus transmis par les mollusques bivalves impliqués dans la transmission d'infections d'origine alimentaire.

La recherche de ces virus passant par l'extraction et l'analyse de leur matériel génétique, il était indispensable de se rassurer de la qualité de l'extraction de leur matériel génétique dans les tissus de nos palourdes à travers la validation de la méthodologie utilisée. L'évaluation de l'efficacité d'extraction des acides nucléiques a révélé des taux d'extraction compris entre 3 et 67% supérieurs au seuil de 1% (Tableau 17) et une absence d'inhibiteurs de la PCR.

De l'ensemble des échantillons analysés, les résultats obtenus et résumés dans le tableau 14 montrent que 80%, soit 36 échantillons sur les 45 analysés, étaient contaminés par au moins l'un des deux groupes de virus recherchés dans ces coquillages (NoV et VHA). Par ailleurs, 69% étaient contaminés par les NoV et 36% par le VHA, avec des titres viraux variant respectivement entre 0,9 – 4,69 unités logarithmiques de nombre de copie ARN/g (log copie ARN/g) et 0,85 – 2,97 log copie ARN/g. Les NoV du génogroupe II (GII) étaient près de deux fois plus souvent détectés que les NoV du génogroupe I (GI) ; 66% d'échantillons analysés étant contaminés par le GII et 34% par le GI (Tableau 18).

Une comparaison du degré de contamination des échantillons de palourdes dans l'espace et dans le temps montre globalement un pourcentage d'échantillons contaminés plus élevé dans le point de Moulongo que dans celui de Bolounga et ceci aussi bien au premier qu'au second semestre de l'année. En effet, environ 90% d'échantillons de Moulongo étaient contaminés aussi bien au premier (11/12) qu'au second semestre (9/10) contre près de 70% pour le point de Bolounga durant les mêmes périodes (9/13 et 7/10 respectivement au premier et second semestre) (Tableau 18). Quant à la contamination spécifique de ces échantillons par les différents virus recherchés, il apparaît sur le plan spatial des prévalences en NoV similaires entre les échantillons des deux points d'échantillonnage contrairement à celles en

VHA (Tableau 17) ; la prévalence de ce dernier dans les échantillons collectés dans le point de Moulongo étant près de deux fois supérieure à celle des échantillons de Bolounga.

Afin d'apprécier le degré de contamination de ces palourdes par les virus détectés, l'analyse quantitative de ces derniers a été réalisée à travers la détermination de leurs concentrations moyennes géométriques (*GMC : Geometric Mean Concentration*), calculée à partir des concentrations obtenues pour chaque échantillon positif et exprimée en copie d'ARN/g de tissu digestif. Ces données obtenues et présentées dans le tableau 17 confirment sur le plan spatial des concentrations moyennes plus importantes du VHA dans les échantillons de Moulongo. La tendance inverse est observée avec les NoV dont la concentration moyenne géométrique est deux fois plus importante dans les échantillons de Bolounga alors que le pourcentage d'échantillons positifs à ces virus était identique dans les deux sites.

**Tableau 17. Efficacité d'extraction, prévalence et concentrations moyennes géométriques des NoV et du VHA dans les échantillons de palourdes de la Sanaga.**

Site	Nb d'échantillons	Efficacité d'extraction*	Norovirus		VHA	
			Nb pos	GMC**	Nb	GMC
Bol	23	13 (3-67)	15 (65 %)	563	6 (26%)	69
Mou	22	9 (4-40)	16 (73 %)	297	10 (45%)	75
Total	45	11 (3-67)	31 (69 %)	405	16 (36%)	73

\* : efficacité d'extraction moyenne calculée pour les échantillons prélevés sur chaque site ou pour l'ensemble de l'étude exprimé en % [Moyenne (minimum-maximum)] ; \*\* **GMC** : concentration moyenne géométrique calculée à partir des concentrations obtenues pour chaque échantillon positif et exprimée en copies d'ARN/g de tissu digestif ; **Nb** : nombre ; **pos** : positif ; **VHA** : virus de l'hépatite A. **Bol** : Bolounga ; **Mou** : Moulongo

**Tableau 18. Concentrations des NoV et du VHA déterminées par RT-qPCR dans les palourdes la Sanaga.**

Sites	Années	Mois	Echantillons	Log <sub>10</sub> copies d'ARN/g		
				NoV GI	NoV GII	VHA
Bol	2018	Février	1	ND	ND	ND
			2	ND	ND	ND
			3	ND	ND	1,99
		Mai	4	ND	ND	ND
		Juillet	5	ND	2,15	1,89
		Août	6	ND	ND	ND
		Septembre	7	ND	4,244	1,23
			8	ND	4,694	ND
		Décembre	9	ND	ND	ND
			10	ND	2,708	ND
	2019	Janvier	11	3,327	ND	2,97
			12	3,149	ND	0,85
		Février	13	ND	ND	ND
		Avril	14	2,088	1,509	ND
		Juin	15	ND	1,794	ND
		Octobre	16	ND	2,097	ND
		Novembre	17	ND	1,694	ND
		Décembre	18	ND	ND	ND
	19		ND	1,436	2,14	
	2020	Janvier	20	ND	2,275	ND
21			1,871	1,326	ND	
Février		22	ND	2,225	ND	
		23	ND	2,523	ND	
Mou	2018	Février	24	ND	2,269	1,93
			25	ND	ND	2,29
		Mai	26	ND	ND	1,52
			27	ND	ND	2,41
		Juillet	28	2,139	ND	2,77
			29	ND	ND	ND
		Septembre	30	ND	ND	1,44
			31	2,621	ND	ND
		Décembre	32	3,091	ND	ND
			2019	Janvier	33	3,269
	34	1,926			1,655	2,37
	Mars	35		ND	ND	ND
	Avril	36		0,9	1,236	ND
	Juin	37		ND	1,993	2,17
	Octobre	38	ND	2,342	2,1	
	Novembre	39	ND	1,523	ND	
	Décembre	40	ND	2,287	ND	
		41	ND	2,415	ND	
	2020	Janvier	42	1,431	2,213	1,43
			43	ND	2,289	ND
		Février	44	1,931	2,728	ND
			45	1,572	2,361	ND

ND : Non détecté ; NoV : Norovirus ; G : Génogroupe ; VHA : Virus de l'hépatite A

## **b. Génotypes des NoV et du VHA présents dans les palourdes de la Sanaga**

Les sites de gisement et de pêche des palourdes de la Sanaga se trouvant dans une réserve faunique, il nous a semblé pertinent de typer les virus détectés par RT-qPCR afin de déterminer leur origine. Ce qui pourrait améliorer les mesures préventives de la propagation des infections liées à ces virus. C'est dans cette optique que les produits de PCR issus des échantillons positifs ont été purifiés, clonés, séquencés, puis génotypés. Les données du génotypage basé sur les régions codant respectivement la protéine majeure VP1 de la capsidie et l'ARN polymérase ARN dépendant (RdRp) pour les NoV, ainsi que la jonction VP1/P2A pour le VHA sont résumées dans le tableau 19.

Le génotypage basé sur la région codant la protéine majeure VP1 de la capsidie a permis d'identifier quatre génotypes de NoV ayant tous une origine humaine (Tableau 19). Parmi ceux-ci, on distingue le génotype GI.1, dont la séquence nucléotidique a présenté une similitude de 95,96% avec la souche FJ383881.1 détectée dans les selles d'humains collectées en Russie en 2008 (Tableau 19). L'alignement nucléotidique de la séquence détectée dans les palourdes de la Sanaga à celle détectée en Russie a révélé 23 substitutions dont seules les substitutions des résidus 323, 324, 326, 327, 328, 330, 331, et 333 (Figure 22A) ont induit la modification de la série d'acides aminés Trp-Val-Trp-Gln en Tyr-Ala-Gly-Gly (résidus 103 à 106) de la protéine codée (Figure 22B). Les trois autres génotypes correspondaient aux génotypes GII.3, GII.4 Sydney\_2012, et GII.6. Ils présentaient 97,09% à 98,54% de similitude nucléotidique avec les souches MK764020.1, MH260511.1 détectées aux Etats-Unis en 2014 et 2017 respectivement, et la souche MH218670.1 détectée au Royaume Uni dans les selles de patients humains (Tableau 19). Les alignements nucléotidiques des génotypes GII.3 et GII.6 (Figures 23A et 24A) avec leurs séquences homologues obtenues par *BLASTn* ont permis de détecter 11 et 7 substitutions dont seules les substitutions en positions 167 et 295 ont induit une modification d'acide aminé respectivement en position 43 (Arg/Gly) et 99 (Tyr/His) des protéines codées (Figures 23B et 24B). L'alignement nucléotidique du génotype GII.4 (Figure 25A) a mis en évidence 13 substitutions dont celles des positions 307 à 309 et 301 à 312 ont induit les mutations en positions 103 et 104 (Arg-Ala/Phe-Glu) de la protéine codée (Figure 25B).

**Tableau 19. Séquences identifiées lors du génotypage des NoV et du VHA**

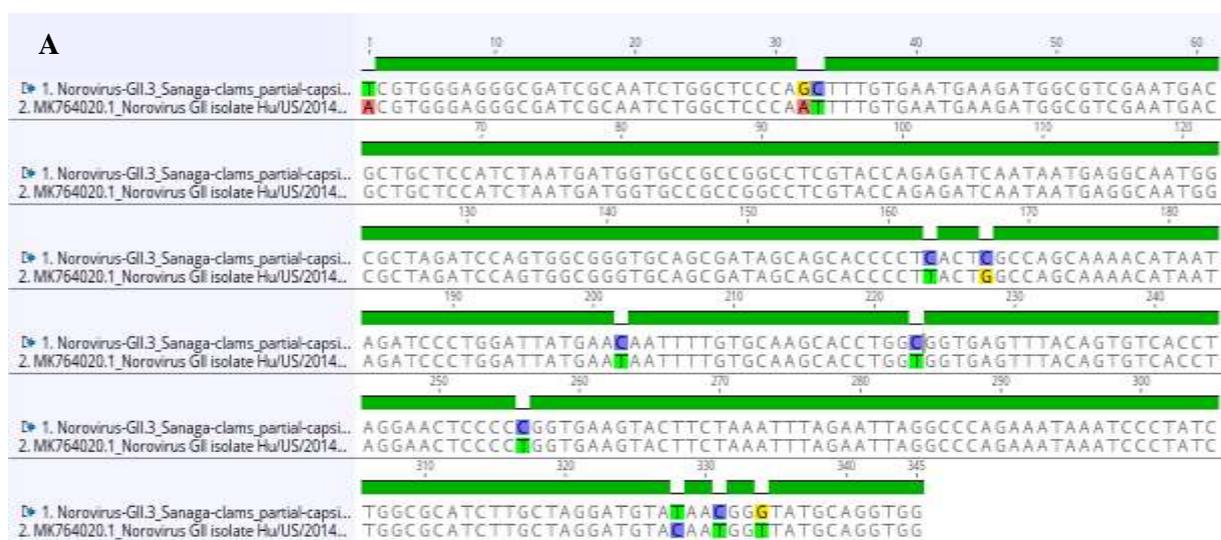
Point	Echantillon	Taille (nucléotides)	Génotype/ P-type	Hit	N° Accession	% identité	E-value
Bol	Janvier 19	334	GI.1	<i>Norovirus Hu/GI.1/2007/11580/Moscow/RUS capsid gene, partial cds (Russia)</i>	FJ383881.1	95.96%	1E-143
Bol	Septembre 18	345	GII.3	<i>Norovirus GII isolate Hu/US/2014/GII.P16-GII.3/Pittsylvania0388 nonstructural polyprotein (ORF1) gene, partial cds; and VP1 (ORF2) and VP2 (ORF3) genes, complete cds (USA)</i>	MK764020.1	97.09%	2E-161
Bol	Septembre 18	345	GII.6	<i>Norovirus GII isolate ST86 nonstructural polyprotein (ORF1) and VP1 (ORF2) genes, complete cds; and VP2 (ORF3) gene, partial cds (UK)</i>	MH260511.1	98.51%	3,00E-165
Bol	Décembre 18	362	GII.4 Sydney_2012	<i>Norovirus GII isolate NORO_212_17_12_2015, complete genome (USA)</i>	MH218670.1	98.54%	1,00E-168
Bol	Juillet 18	211	GII.P31	<i>Norovirus GII isolate NORO_180_01_09_2015, partial genome</i>	MH218648.1	95.54%	8,00E-84
Bol	Septembre 18	213	GII.P13	<i>Norovirus Hu/GII/1443/2006/Ghana RNA-dependent RNA polymerase gene, partial cds (Ghana)</i>	EU315742.1	90.16%	1,00E-62
Bol	Décembre 18	210	GII.P31	<i>Norovirus GII isolate 652 nonstructural polyprotein, RdRp region, (ORF1) gene, partial cds (Gabon)</i>	MW513414.1	98.44%	8,00E-89
Mou	Février 18	160	GII.P21	<i>Norovirus GII isolate 35 nonstructural polyprotein (ORF1) gene, partial cds; and VP1 (ORF2) and VP2 (ORF3) genes, complete cds (China)</i>	MN394543.1	97.32%	9,00E-63
Bol	Septembre 18	441	GV	<i>Hepatitis A virus strain KZN_Irr-20130530-2B polyprotein gene, partial cds (South Africa Sewage)</i>	KP875241.1	91.88%	4,00E-136

**E** : exponentielle



**Figure 22. Alignements de la séquence du génotype GI.1 de norovirus détecté dans les palourdes de la Sanaga avec la séquence partielle de la région codant la protéine VP1 de la capsid de la souche FJ383881.1.**

(A). Alignement des séquences de 334 résidus d’acides nucléiques mettant en évidence 23 substitutions. (B). Alignement des séquences protéiques de 106 résidus d’acides aminés mettant en évidence 04 substitutions aux extrémités 3’.





**Figure 23.** Alignements de la séquence du génotype GII.3 de norovirus détecté dans les palourdes de la Sanaga avec la séquence partielle de région codant la protéine majeure VP1 de la capsid de la souche MK764020.1.

**A** : Alignement des séquences de 345 résidus d'acides nucléiques mettant en évidence 11 substitutions. **B** : Alignement des séquences protéiques de 101 résidus d'acides aminés mettant en évidence 01 substitution en position 43.



**Figure 24.** Alignements de la séquence du génotype GII.6 de norovirus détecté dans les palourdes de la Sanaga avec la séquence partielle de la capsid de la souche MH260511.1.

(A). Alignement des séquences de 307 résidus d'acides nucléiques mettant en évidence 07 substitutions. (B). Alignement des séquences protéiques de 102 résidus d'acides aminés mettant en évidence 01 substitution en position 99 dans l'extrémité 3'.



**Figure 25. Alignements de la séquence du génotype GII.4 de norovirus détecté dans les palourdes de la Sanaga avec la séquence partielle de la capside de la souche MH218670.1.**

(A). Alignement des séquences de 314 résidus d'acides nucléiques mettant en évidence 07 substitutions. (B). Alignement des séquences protéiques de 104 résidus d'acides aminés mettant en évidence 02 substitutions en positions 103 et 104 dans l'extrémité 3'.

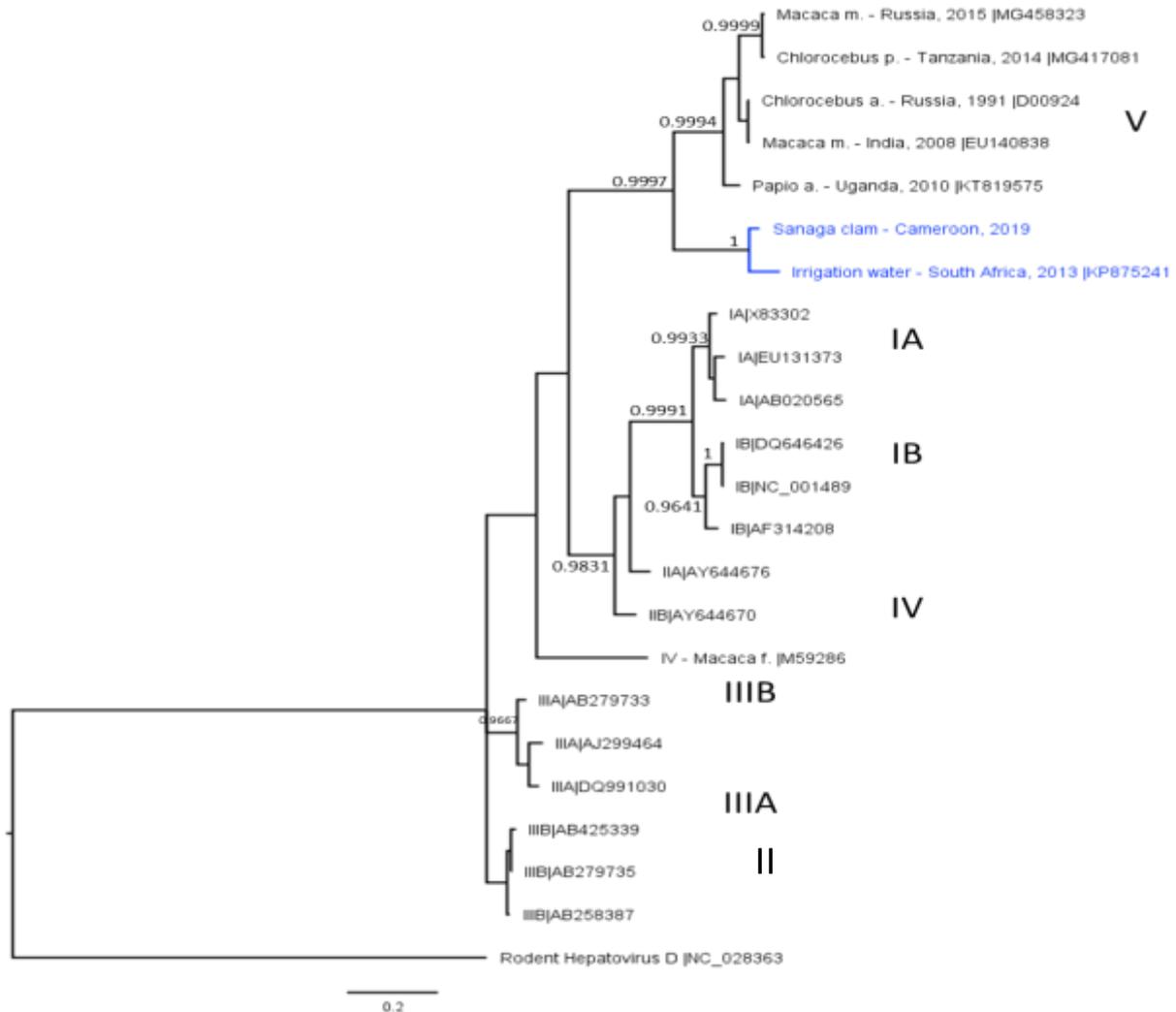
Le typage basé sur la région du génome codant la RdRp a permis d'identifier quatre séquences dont deux appartenant respectivement aux P-types GII.P13, GII.P21 et deux appartenant au P-type GII.P31 (Tableau 19). L'alignement *BLASTn* sur GenBank de ces dernières a révélé des similitudes de 90,16% à 98,44% avec les souches EU315742.1, MN394543.1, MH218648.1 et MW513414.1 respectivement isolées au Ghana, en Chine, et au Gabon (Tableau 19).

Le génotypage de la jonction VP1/P2A du génome du virus de l'hépatite A a permis d'identifier une séquence appartenant au génotype V (Tableau 19). Après alignement *BLASTn* sur GenBank, la séquence de la souche Sanaga (MT185676) a affiché une similitude nucléotidique de 91,88 % avec la souche KZN\_Irr-20130530 (KP875241.1) identifiée dans des échantillons d'eaux usées en Afrique du Sud (Tableau 19). L'alignement nucléotidique de la séquence de la souche du virus de l'hépatite A détectée dans les palourdes de la Sanaga avec son homologue détectée dans les eaux usées en Afrique du Sud (Figure 26A) a révélé 28 substitutions dont seule la substitution en position 130 a induit une mutation en position 44 (Gly/Ser) de la protéine codée (Figure 26B). L'analyse phylogénétique avec les séquences du virus de l'hépatite A de différents génotypes extraits de HAVNet et GenBank a montré le regroupement de ces deux souches extraites d'échantillons environnementaux en un clade étroitement lié au génotype V constitué des souches simiennes (Figure 27).



**Figure 26.** Alignements de la séquence de la souche du virus de l'hépatite A détecté dans les palourdes de la Sanaga (MT185676.1) avec la souche KP875241.1 détectée dans les eaux d'irrigation en Afrique du Sud.

(A). Alignement des séquences de 345 résidus d'acides nucléiques mettant en évidence 28 substitutions. (B). Alignement des séquences protéiques de 115 résidus d'acides aminés mettant en évidence 01 substitution en position 44.



**Figure 27. Arbre phylogénétique bayésien des séquences de virus de l'hépatite A.**

L'arbre phylogénétique est basé sur 441 nucléotides situés dans la région de jonction VP1-2A (position 2918 à 3359 par rapport à la séquence de référence du virus de l'hépatite A NC\_001489), à l'exception de MG417081 (315 nt), MG458323 (346 nt) et M59286 (334 nt) qui se terminent avant. Les séquences ont été alignées à l'aide de MUSCLE 3.8.425 implémenté dans Geneious (11.1.4). L'arbre a été construit à l'aide de Geneious, avec le modèle évolutif GTR. Les probabilités postérieures supérieures à 0,95 sont indiquées sur les branches correspondantes. La séquence de référence de l'hépatovirus D de rongeur a été utilisée comme groupe externe afin d'enraciner l'arbre. Les séquences du virus de l'hépatite A se regroupent par génotype (Gt) comme indiqué à droite, avec des virus humains appartenant aux Gt IA-B, IIA-B et IIIA-B. Le virus de l'hépatite A simien (l'espèce hôte est indiquée dans le nom du taxon) appartient au génotype IV ou V. Le virus de l'hépatite A des palourdes de la Sanaga ainsi qu'une

séquence identifiée dans un échantillon d'eau d'irrigation d'Afrique du Sud se regroupent dans un clade (bleu) séparé des séquences simiennes du génotype V.

### 3.1.2 Diversité des virus à ARN présents dans les palourdes de la Sanaga

La détection des NoV d'origine humaine et des virus d'origine animale, notamment d'une souche de VHA apparentée au génotype V d'origine simienne mettant en évidence la circulation de ces derniers dans cet environnement, il était intéressant de déterminer la diversité des virus présents dans ces organismes aquatiques qui constituent la principale ressource alimentaire des populations riveraines. Cette évaluation de la diversité devrait non seulement permettre d'identifier les potentiels virus à ARN pathogènes auxquels seraient exposés les consommateurs des palourdes, mais également d'envisager leur usage comme sentinelles de la contamination virale de l'environnement. Pour y parvenir, nous avons effectué une analyse du métagénome des virus à ARN des échantillons de palourdes de la Sanaga à l'aide d'une approche de NGS métagénomique, suivie du développement d'un *pipeline* de traitement bio-informatique.

#### a. Caractéristiques des données métagénomiques

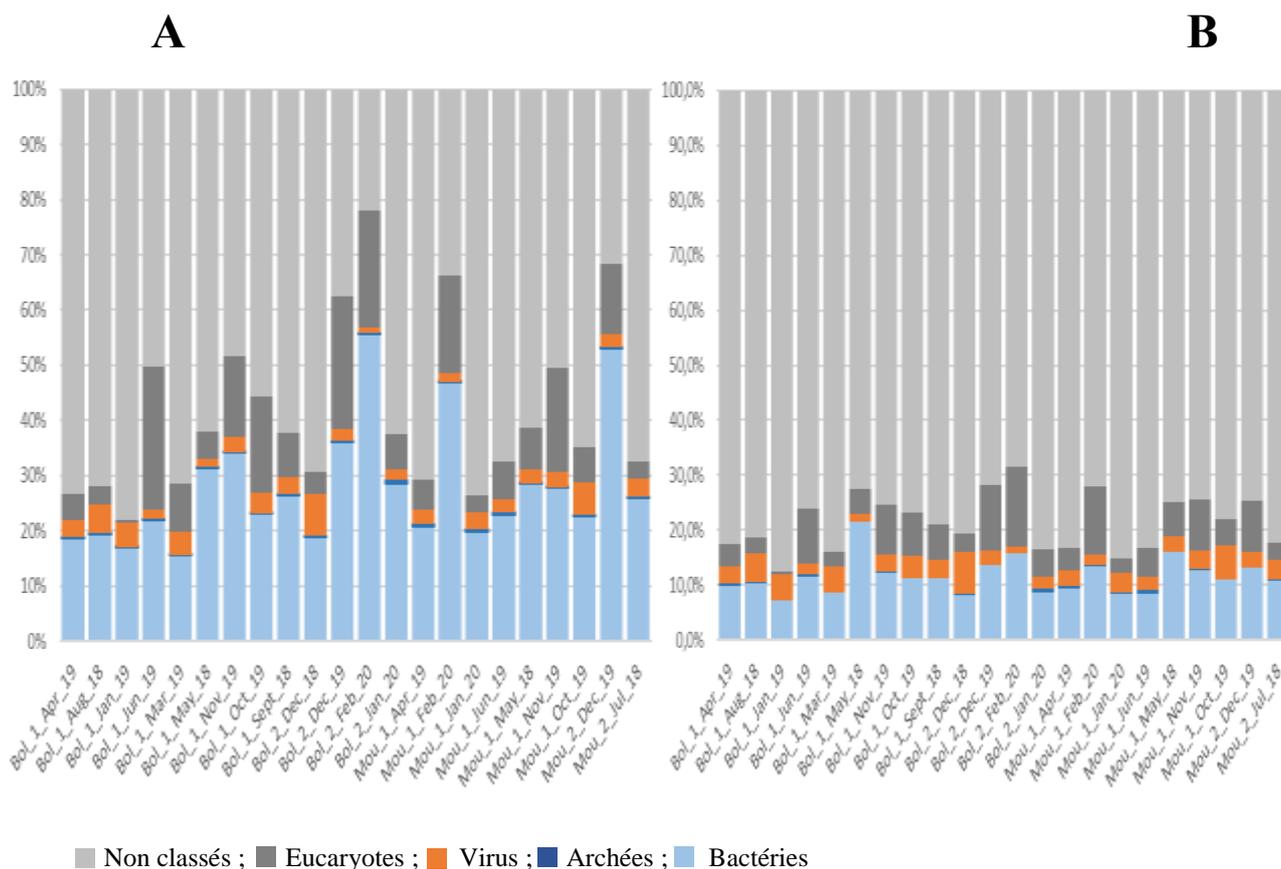
Vingt-deux (22) échantillons constitués de 330 sous échantillons (SE) de palourdes dont 13 (195 SE de palourdes) de Bolounga et 9 (135 SE de palourdes) de Mouloungo (Tableau 20), retenus sur la base de la présence des NoV et/ou du VHA dans leurs TD, de la valeur des *Ct*, ainsi que de la représentativité des points de récolte ont été soumis à une analyse métagénomique des virus à ARN.

Le séquençage Illumina NovaSeq a généré plus de 1,9 milliards de *reads* bruts variant de 42 à 177 millions par librairie de séquences d'ADN. Parmi ceux-ci, 20 % ont été éliminés à l'issue du filtrage de la qualité et de l'élimination des ARN ribosomiques (Tableau 20). Par ailleurs, afin de réduire la taille des données et optimiser le temps de calcul pour l'assemblage *de novo*, une déduplication a été réalisée. Ce traitement a permis d'éliminer jusqu'à 85 % des *reads* dans certaines librairies. De la vérification de l'impact de ce traitement de données sur la distribution des *reads* par une assignation taxonomique rapide à l'aide de l'outil en ligne Kaiju avant et après traitement (Figure 28), il ressort que les bactéries étaient les plus affectées avec 88 à 96 % des *reads* supprimés, suivi des eucaryotes avec une suppression de 77 à 95 % des *reads*. La suppression des *reads* de la fraction virale quant à elle était assez stable entre les différents échantillons et variait de 79 à 83 %. A l'issue de ce traitement, 13,8 % des *reads* initiaux (soit 264 millions de *reads*) ont été conservés (Tableau 20). Parmi ceux-ci, 68 à 87 %

n'ont pas pu être classés, 7,1 à 21,4 % ont été attribués aux bactéries, 0,7 à 10,2 % aux eucaryotes, 1,1 à 7,5 % aux virus, et 0,1 à 0,7 % aux archées (Figure 28.B). Toutes les analyses suivantes ont été effectuées sur des *reads* dédupliqués.

**Tableau 20. Résumé du nombre de *reads* et de contigs par échantillon à différentes étapes du *pipeline* du traitement bio-informatique.**

Point	Date	Total <i>reads</i> (millions)				Nombre de <i>contigs</i>			Virus	
		Brutes	Filtrés	Dédupliqués	<i>Mapped</i>	Assemblés	> 300 pb	BLAST	<i>Reads</i> (millions)	%
Bol	Mai 18	48	44	8	3	7890	6917	4340	0,9	12
Bol	Août 18	60	56	9	6	10200	8576	4148	4	45
Bol	Septembre 18	155	131	21	14	16145	14635	5416	7	33
Bol	Décembre 18	88	80	15	11	20339	15185	6608	8	52
Bol	Janvier 19	111	109	20	16	10044	8840	3087	15	77
Bol	Mars 19	71	66	12	10	9280	8233	3173	6	54
Bol	Avril 19	69	65	12	9	18067	13937	6332	4	30
Bol	Juin 19	86	64	11	9	10266	7548	2968	2	19
Bol	Octobre 19	47	36	7	5	12836	10215	4330	2	27
Bol	Novembre 19	44	32	6	4	12554	10751	4368	2	26
Bol	Décembre 19	42	27	5	4	9287	8173	3540	0,8	18
Bol	Janvier 20	129	102	18	15	18929	14789	6062	8	42
Bol	Février 20	177	95	14	12	11410	8487	3822	2	12
Mou	Mai 18	63	55	9	6	8919	8068	3982	3	28
Mou	Juillet 18	131	119	19	14	14143	12954	4972	9	50
Mou	Avril 19	70	62	12	9	15763	12640	5370	4	38
Mou	Juin 19	67	57	11	8	12655	10061	3737	4	40
Mou	Octobre 19	78	68	12	9	18287	14316	5776	4	35
Mou	Novembre 19	51	37	7	5	12444	10060	4407	1	21
Mou	Décembre 19	152	89	14	11	14108	11406	4847	3	24
Mou	Janvier 20	87	82	15	11	15704	11940	4046	5	31
Mou	Février 20	90	54	9	7	9475	7552	3145	2	19



**Figure 28. Abondance relative des reads par règne et par librairie.**

Les reads totaux obtenus après filtrage (A) ou après déduplication (B) ont été rapportés pour les bactéries, les archées, les virus, les eucaryotes, et les non classés.

### b. Assemblage du métagénome et assignation taxonomique

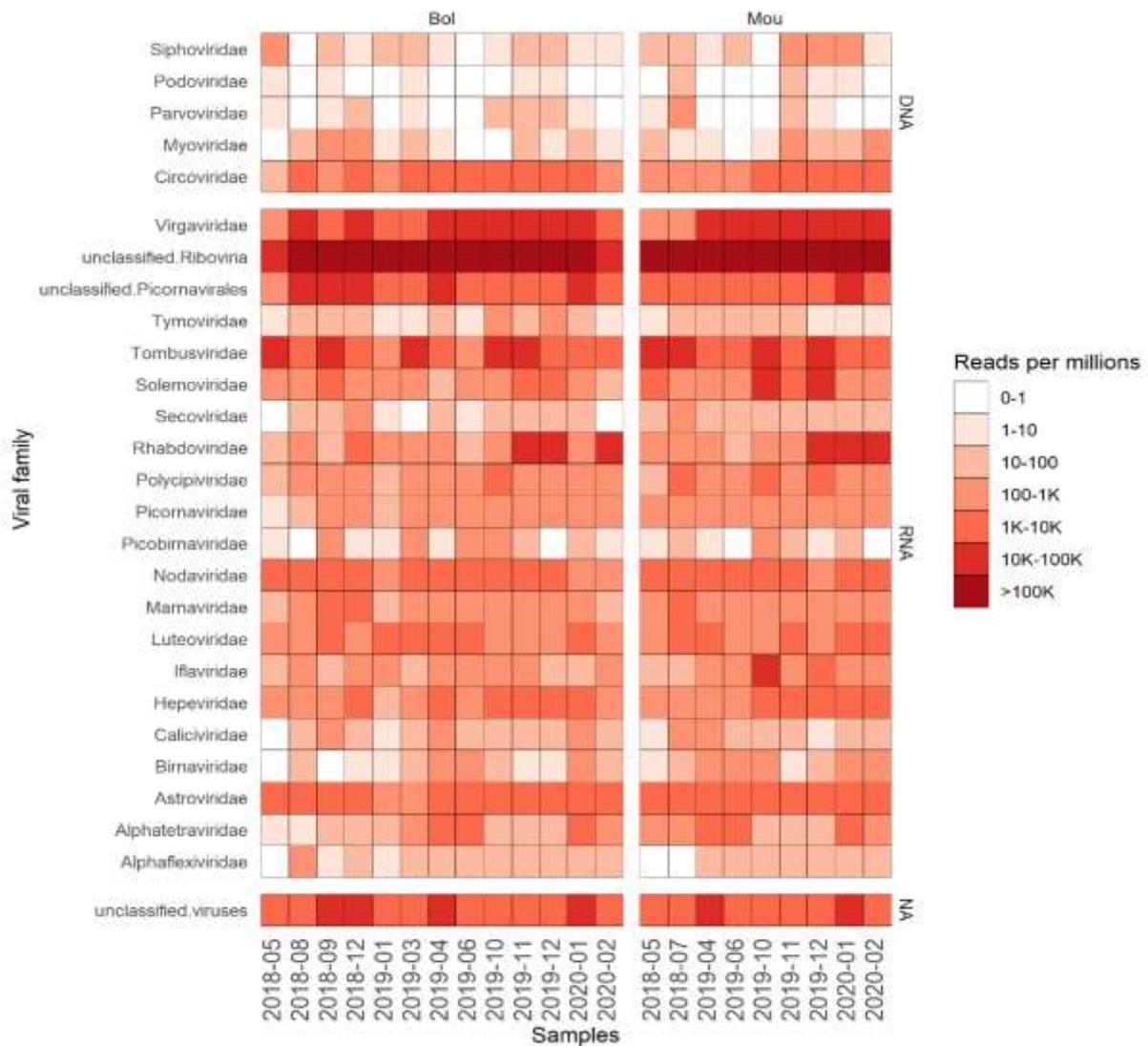
L'assemblage *MetaSPAdes* a permis d'assembler 268378 séquences *contigs* dont 204107 de tailles supérieures à 300 pb. Parmi ceux-ci, 98476 (48 %) ont été annotés par *BLASTn* ou *DIAMOND* (Tableau 20).

Un total de 198 millions de reads a été assigné, dont 95 millions identifiés comme étant des séquences virales. Le pourcentage de reads identifiés comme viraux dans les différents échantillons variait entre 12 et 77 % ; le pourcentage le plus élevé (77 %) ayant été observé avec l'échantillon collecté à Bolounga en janvier 2019 et les plus faibles observés avec des échantillons du même point, prélevés en mai 2018 et février 2020 (Tableau 20). Les reads viraux identifiés ont été classés en différentes familles virales sur la base des résultats *BLASTn*/*DIAMOND* (Figure 29). Ce qui a permis d'identifier une vingtaine de familles virales en moyenne dans les différents échantillons et représentant au moins 0,01 % de reads dans un des échantillons. Comme on pouvait s'y attendre à partir du protocole utilisé durant la

préparation de bibliothèques (traitement à la DNase et enrichissement en virus de vertébrés), peu de *reads* appartenant aux virus à ADN (0,6 %) ont été identifiés. Parmi ceux-ci, on distinguait peu de *reads* appartenant aux familles de bactériophages (0,03 %) dont les *Siphoviridae*, les *Podoviridae*, et les *Myoviridae*. Des *reads* appartenant à deux familles de virus à ADN infectant les vertébrés ont également été identifiés. Ils s'agissaient des *Circoviridae* fréquemment identifiés et les *Parvoviridae*.

Les résultats obtenus à partir du *BLASTn/DIAMOND* montrent que les *contigs* les plus abondants appartenaient aux virus à ARN non classés (96,6 % des *reads* viraux). Parmi les virus à ARN identifiés, nous avons noté une abondance de séquences de virus de plantes avec la famille des *Virgaviridae* fréquemment rencontrée, mais aussi celles *Solemoviridae*, des *Luteoviridae*, des *Tymoviridae*, des *Secoviridae*, et des *Alphaflexiviridae*. Des séquences virus hébergés par les invertébrés à l'instar des arthropodes dont les insectes et regroupées au sein des familles des *Nodaviridae*, des *Polycipiviridae*, *Iflaviridae*, *Alphatetraviridae* étaient également détectées. Il en est de même pour les séquences de virus de virus d'algues regroupées au sein de la famille des *Marnaviridae*. L'alignement des séquences *reads* sur les virus à ARN des mammifères était moins fréquent, et la famille des *Astroviridae* était la plus fréquemment rencontrée par rapport aux *Caliciviridae*, *Picornaviridae*, *Picobirnaviridae* et *Hepeviridae*. Des *contigs* appartenant aux *Rhabdoviridae*, famille au sein de laquelle se trouve le rhabdovirus ou virus de la rage ont également été détectés (Figure 29). Toutefois, ces derniers étaient similaires aux rhabdovirus identifiés dans les amphibiens.

Malgré la variabilité du pourcentage de *reads* identifiés comme étant des séquences virales notée entre les échantillons, aucune différence nette n'a été observée entre les différents échantillons analysés tant sur le plan spatial que temporel des points ou au cours de la période d'étude en ce qui concerne les principales familles de ribovirus identifiées. Ce constat ainsi que les observations précédentes (pourcentage élevé de *reads* de ribovirus non classés, faible nombre de *reads* correspondant aux familles de virus infectant les mammifères), nous ont conduit à combiner l'ensemble des *reads* pour améliorer les taux d'identification des séquences virales présentant un intérêt.



**Figure 29. Heatmap de la diversité et de l'abondance relative virale.**

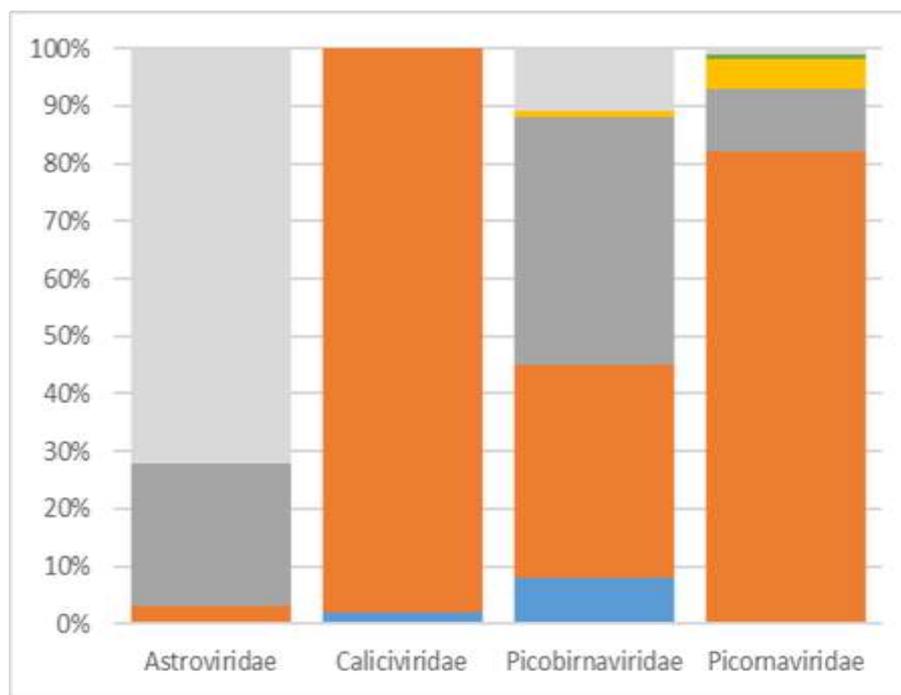
L'abondance des *reads* viraux après le contrôle qualité et la déduplication est présentée pour les deux sites (Bol : Bolounga et Mou : Moulongo). Les gradients de couleur représentent l'abondance relative des *reads* appartenant aux groupes taxonomiques indiqués, les cinq premières rangées correspondant aux virus à ADN, la dernière rangée aux virus non classés et les autres rangées aux virus à ARN.

### 3.1.3 Virus à ARN humains et les animaux pouvant avoir un intérêt en santé humaine

Afin d'identifier les virus humains et animaux présentant un intérêt en santé humaine, une attention a été portée aux séquences virales appartenant aux familles *Astroviridae*, *Caliciviridae*, *Hepeviridae*, *Picobirnaviridae* et *Picornaviridae* aux seins desquelles se trouve la majorité des virus entéro-pathogéniques humains, mais aussi animaux. En utilisant les critères de sélection définis en méthodologie, les *contigs* sélectionnés ont été identifiés à l'aide

du *BLASTn* ou de *DIAMOND* et l'hôte correspondant au meilleur résultat du BLAST a été considéré (Figure 30).

Il en ressort que seules les séquences de la famille des *Picobirnaviridae* étaient liées à une diversité d'hôtes avec environ le même pourcentage de séquences liées à des souches infectant des chauves-souris ou d'autres mammifères, et quelques autres à des souches infectant les humains ou n'ayant pas d'hôtes spécifiés. Les séquences identifiées comme appartenant aux *Hepeviridae* étaient globalement apparentées à des souches infectant les éponges (non illustré dans le graphe). Pour les familles des *Caliciviridae* et *Picornaviridae*, la plupart des séquences identifiées étaient liées à des souches infectant les chauves-souris. Toutefois quelques séquences de *Caliciviridae* étaient similaires à souches spécifiques aux humains. Les autres mammifères hôtes des *Astroviridae*, *Picobirnaviridae*, *Picornaviridae* et *Hepeviridae* étaient essentiellement les singes, les porcs, les lapins et les souris.



■ Humains ; ■ Chauves-souris ; ■ Autres mammifères ; ■ Oiseaux ; ■ Reptiles ; ■ Non précisé

**Figure 30. Aperçu de la diversité des hôtes et des reads identifiés comme appartenant à quatre familles virales.**

Pour chaque famille le nombre de *reads* a été exprimé en % se référant à l'hôte potentiellement infecté. Seules les *reads* assignés les plus abondants sont présentées ici.

## *Astroviridae*

Les *contigs* identifiés comme appartenant à la famille des *Astroviridae* étaient les plus abondants, avec 841 séquences assemblés à partir de 996887 *reads*.

Un grand nombre de ces *contigs* a été identifié comme appartenant au bastrovirus. Parmi ceux-ci, 620 détectés dans 21 échantillons étaient similaires à des séquences divergentes précédemment détectées dans des échantillons d'eaux usées au Brésil (Bastrovirus Brazil-sewage). Quatre *contigs* de plus de 4000 nucléotides dont trois génomes presque complets provenant du même échantillon ont été détectés. 117 *contigs* provenant de 20 échantillons ont été identifiés comme étant proches du Bastrovirus-*like* divergent détecté dans les excréments de chauve-souris au VietNam. Trois séquences (2311 à 5030 nucléotides) apparentées à la protéine non structurale du bastrovirus porcine PBastV\_USA 2017-1 ont également été détectées dans trois échantillons des deux sites (Tableau 21). Les arbres phylogénétiques basés sur une portion de la protéine VP1 de la capsid (Figure 31A) et de l'ARN polymérase ARN dépendant (Figure 31B) ont confirmé que les séquences de bastrovirus détectées dans les palourdes de la Sanaga se regroupaient dans un clade avec deux séquences divergentes identifiées au Brésil ou au VietNam.

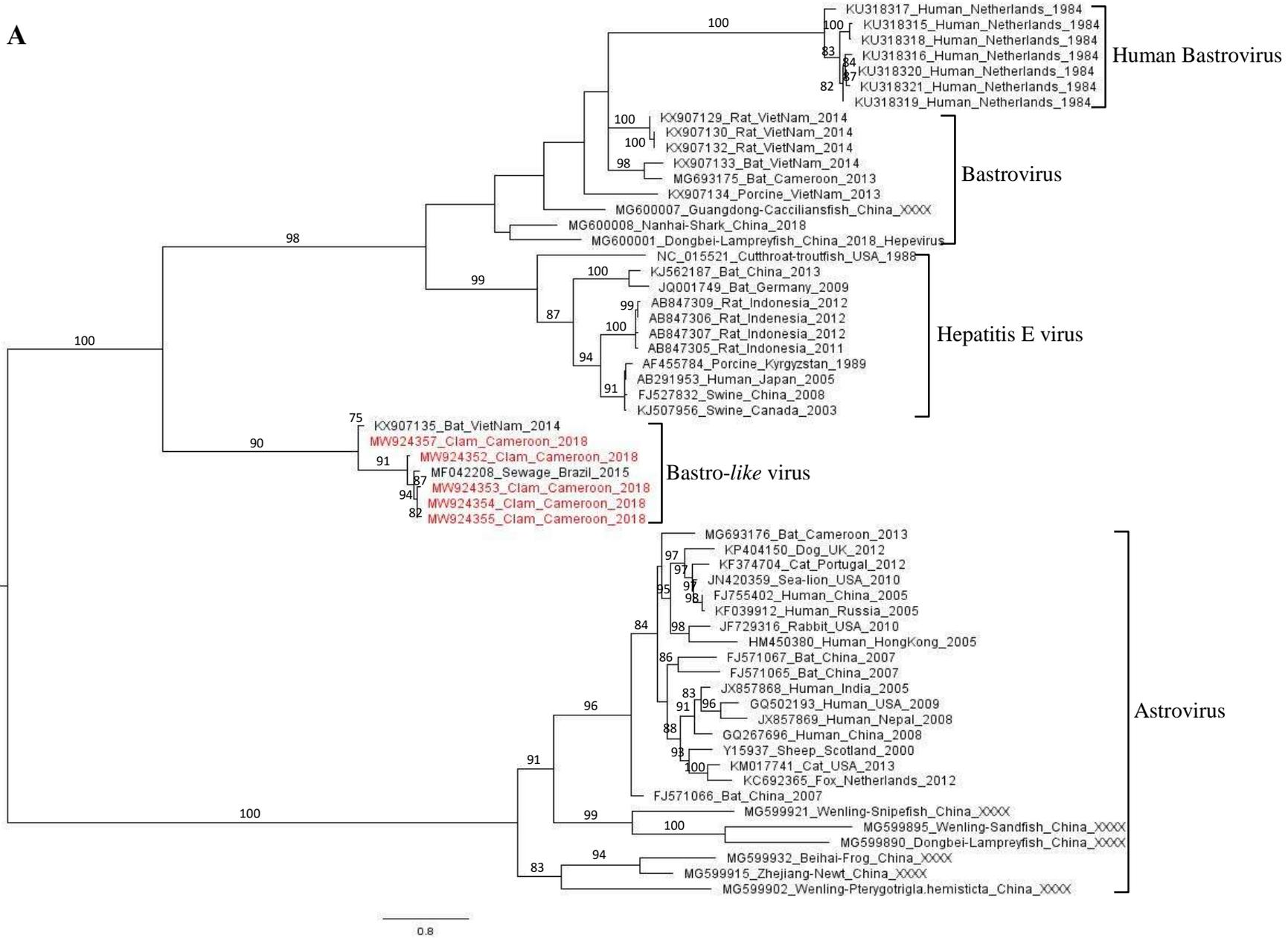
Concernant les astrovirus, 85 *contigs*, dont 65 allant de 1061 à 7084 nucléotides, détectés dans 21 échantillons présentaient de faibles similitudes avec des séquences référencées d'astrovirus détectées dans des échantillons de sol aux USA ; le *contig* le plus long (7084 nucléotides) ne s'alignant d'ailleurs que sur un tiers de la longueur. Deux séquences ont toutefois été identifiées comme étant similaires à un astrovirus infectant un oiseau (*European Roller*) (Tableau 21).

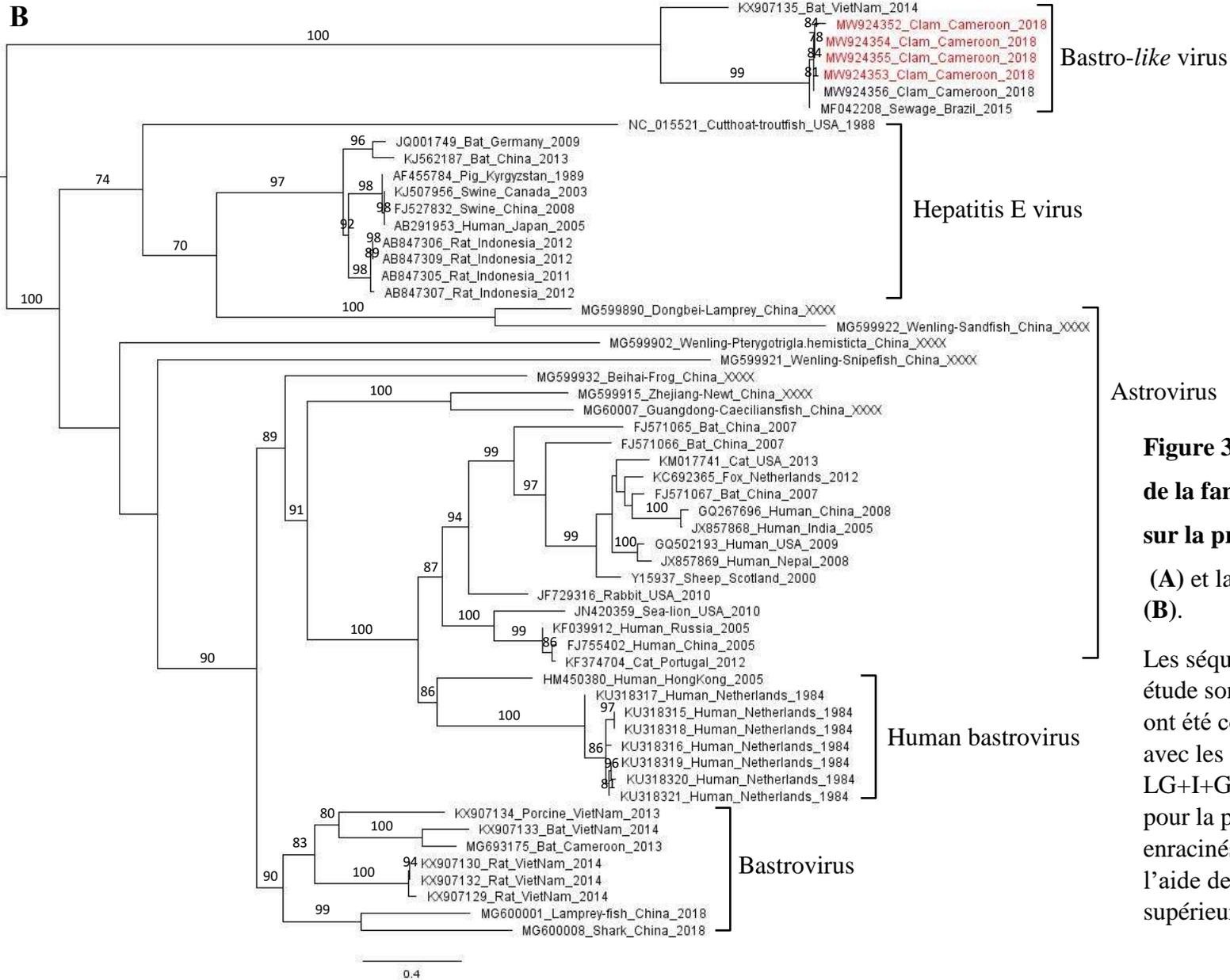
**Tableau 21. Quelques contigs d'*Astroviridae* identifiés dans les palourdes de la Sanaga.**

Site	Date	N° d'Accession*	Taille (nt)	Reads mappés	Match Nucléotidique	Identité (%)	E-value	Noms	Ref. séquence
Bol	Septembre 18	MW924352	5823	27103	4221	87,6	0	Bastrovirus Brazil-sewage	ASM79505.1
Bol	Septembre 18	MW924353	5819	22861	4221	95,2	0		
Bol	Septembre 18	MW924354	5728	17434	4221	95,4	0		
Mou	Juillet 18	MW924355	4312	6571	2901	94,5	0		
Mou	Juillet 18	MW924356	2390	8659	972	92,9	$9,6 \times 10^{-164}$		
Mou	Mai 18	MW924357	1571	991	1233 <sup>#</sup>	95	0	Bastrovirus-like virus, VietNam-Bat	KX907135.1
Mou	Juillet 18	MW924358	1360	1291	1287 <sup>#</sup>	82,2	0		
Mou	Juillet 18	MW924359	1097	1161	1097 <sup>#</sup>	79,8	0		
Bol	Décembre 18	MZ182250	389	99	376 <sup>#</sup>	84	$1,7 \times 10^{-95}$	Bastrovirus-like virus, US- Culex	NC_040647.1
Bol	Août 18	MZ182271	5030	1874	1044	29	$1,2 \times 10^{-27}$	Bastrovirus, US-Porc	QDA77022.1
Bol	Février 20	MZ182272	3325	70746	1044	29	$8,1 \times 10^{-28}$		
Mou	Mai 18	MZ291967	2311	7679	1062	27,7	$7,4 \times 10^{-28}$		
Bol	Octobre 19	MZ291968	404	36	390	56,9	$3,4 \times 10^{-37}$	Astrovirus, Hungary-european roller	QBZ38214.1
Mou	Novembre 19	MZ291969	361	142	324	53,7	$3,7 \times 10^{-27}$		
Bol	Janvier 19	MZ291970	3801	4413	798	31,2	$9,3 \times 10^{-20}$	<i>Astroviridae</i> sp. USA-soil	QDH87307.1
Mou	Juin 19	MZ291971	7084	9823	2316	25,4	$7,3 \times 10^{-34}$		

\*Ces numéros d'accession correspondent à des séquences identifiées dans cette étude et déposées dans GenBank. L'identification des contigs fait avec DIAMOND à partir des l'alignements protéiques, sauf pour certaines séquences notées avec # le BLAST nucléotidique a été utilisé. Les identités des séquences ont été exprimées en pourcentage avec les E-value. nt : nucléotides. Ref. séquence : numéro d'accession de la séquence la plus proche sur GenBank.

A





**Figure 31. Arbres phylogénétiques de la famille des *Astroviridae* basés sur la protéine RdRp**

(A) et la protéine VP1 de la capsid (B).

Les séquences identifiées dans cette étude sont écrites en rouge. Les arbres ont été construits à l'aide de *IQtree* avec les modèles de substitution LG+I+G4 pour la RdRp, et LG+G4 pour la protéine de la capsid, et enracinés à partir du point moyen à l'aide de *FigTree*. Seuls les *bootstrap* supérieurs à 70% sont présentés.

## *Caliciviridae*

Les *Caliciviridae* ont été retrouvés dans 21 des 22 échantillons analysés. En effet, 64 *contigs* assemblés (de taille allant de 352 à 1833 nucléotides) et représentant 20000 *reads* ont été identifiés comme des séquences des *Caliciviridae*, dont trois aux NoV humains et les autres aux calicivirus de chauves-souris (Tableau 22). Des trois *contigs* proches des NoV humains détectés dans 02 échantillons, deux partageaient environ 96 % d'identité nucléotidique avec une séquence du génome de l'isolat NHBGR59 appartenant au génotype GII.6 retrouvé dans les selles d'humains en Inde. Le troisième *contig* présentait quant à lui une forte similitude (97 % d'identité nucléotidique) avec une souche du génotype GII.2 détectée dans des selles d'humains au Japon (Tableau 22). Parmi les *contigs* présentant des similitudes avec les calicivirus de chauve-souris, les plus longs dont la taille variait de 1026 à 1833 nucléotides étaient proches d'une séquence de Bat calicivirus A10 identifiée chez la chauve-souris tricolore (*Perimyotis subflavus*) aux États-Unis et qui semble représenter un nouveau genre de *Caliciviridae* partageant des caractéristiques antigéniques avec des NoV humains (Kocher *et al.*, 2018).

**Tableau 22. Quelques contigs de *Caliciviridae* identifiées dans les palourdes Sanaga.**

Site	Date	N° d'Accession*	Taille (nt)	Reads mappés	Match Nucléotidique	Identité (%)	E-value	Noms	Réf. séquence
Bol	Septembre 18	MW854003	367	176	366 <sup>#</sup>	96,2	3.04 x 10 <sup>-167</sup>	<i>Norovirus GI.1[P6]6, India-humans</i>	KU870455.1
Bol	Septembre 18	MW853999	353	101	353 <sup>#</sup>	95,7	2.28 x 10 <sup>-158</sup>		
Mou	Décembre 19	MW854002	387	10	387 <sup>#</sup>	97.1	0	<i>Norovirus GI.1[Pe]2, Japan-humans</i>	LC209439.1
Bol	Janvier 20	MZ291972	1833	1003	699	87.1	1.10 x 10 <sup>-114</sup>	<i>Bat calicivirus A10, USA-bats</i>	MH259583.1
Bol	Avril 19	MZ291973	1328	846	1281	71.0	2.00 x 10 <sup>-155</sup>		
Bol	Septembre 18	MZ291974	1026	670	1026	51.3	6.70 x 10 <sup>-74</sup>		
Mou	Avril 19	MZ291975	1653	769	1389	70.8	1.10 x 10 <sup>-169</sup>		
Mou	Avril 19	MZ291976	1086	155	294	70.4	1.70 x 10 <sup>-30</sup>		

\*Le numéro d'accession correspond à la séquence identifiée dans cette étude et déposée dans GenBank. Les contigs ont été identifiés avec DIAMOND en utilisant des alignements protéiques, sauf pour certaines séquences notées avec # pour lesquelles le BLAST nucléotidique a été utilisé. L'identité de la séquence a été exprimée en pourcentage avec la E-value. nt : nucléotides. Réf. séquence : numéro d'accession de la séquence la plus proche sur GenBank.

## *Picobirnaviridae*

Un grand nombre de *reads* (21958) a été identifié comme appartenant à la famille des *Picobirnaviridae*. Ces derniers se sont réalignés sur 45 *contigs* de 354 à 2492 pb identifiés comme proches des souches de picobirnavirus isolés d'algues, d'humains, de chauves-souris et de singes.

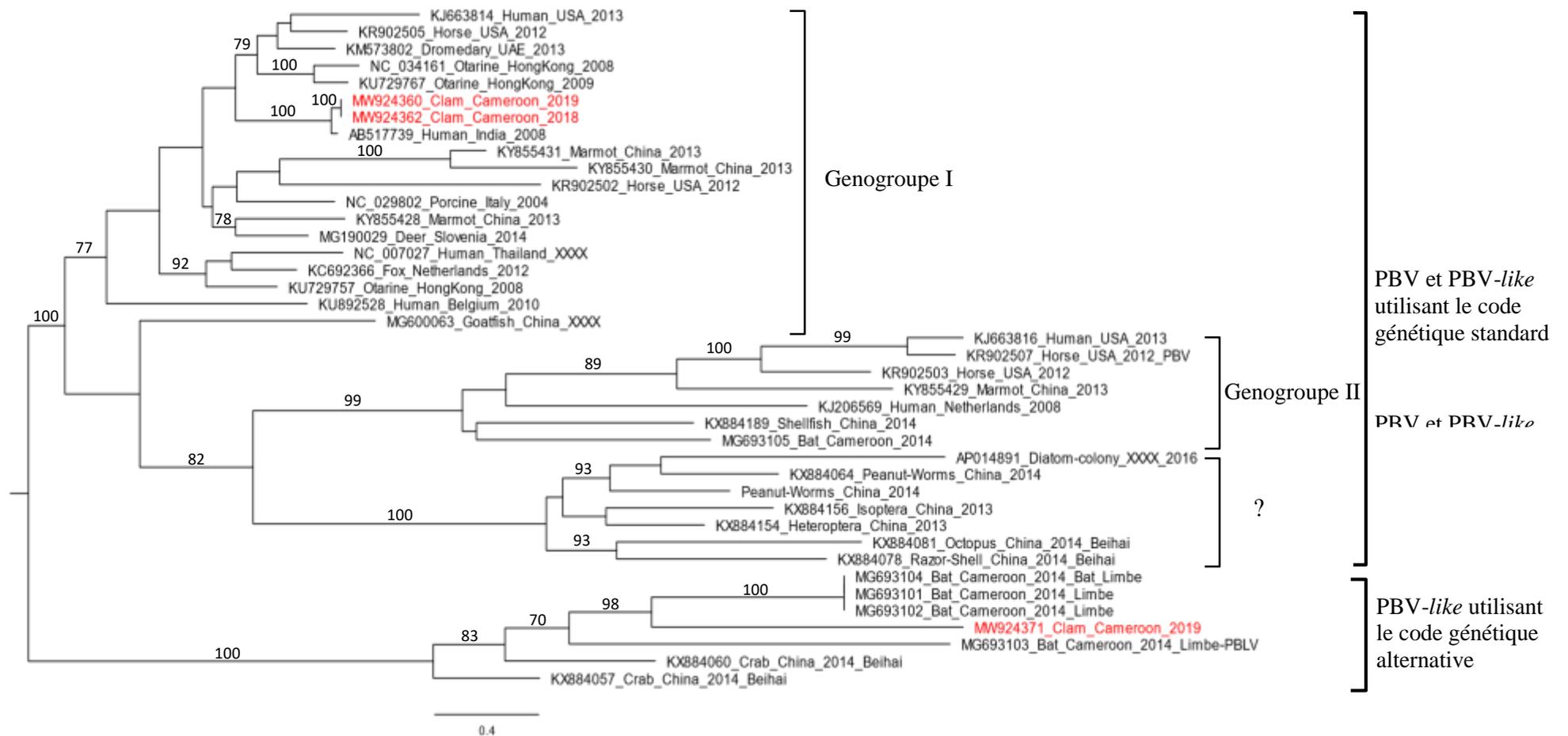
Huit *contigs* de tailles comprises entre 391 et 615 pb ont été identifiés comme étant étroitement liés aux séquences de picobirnavirus humains. Parmi ceux-ci, cinq *contigs* détectés sur le site de Bolounga présentaient une forte homologie nucléotidique avec une souche détectée dans les selles humaines en Inde (Tableau 23). Deux autres *contigs* étaient également liés à des souches humaines détectées respectivement dans les selles des patients en Belgique et au Cameroun. L'analyse phylogénétique de deux *contigs* identifiés aux picobirnavirus humains, basée sur l'ARN polymérase ARN dépendant (Figure 32) a révélé qu'ils se regroupent dans un clade avec un isolat indien au sein du génogroupe I.

Outre les picobirnavirus humains, trois *contigs* présentaient de fortes similitudes avec une souche simienne identifiée dans les excréments de gorilles (*Gorilla gorilla gorilla*) en République Démocratique du Congo (Duraisamy, 2018) et un avec une souche détectée chez les macaques (*Macaca mulatta*) aux Etats Unis (Krishnamurthy et Wang, 2018). Deux *contigs* de 2023 et 2029 pb détectés dans les deux sites en octobre 2019 présentaient une faible homologie avec le picobirna-like virus détecté dans des chauves-souris (*Eidolon helvum*) dans la localité de limbe au Cameroun (Yinda *et al.*, 2018). L'analyse phylogénétique du plus long *contig* basée sur la l'ARN polymérase ARN dépendant a montré le regroupement de cette séquence dans un clade constitué de picobirna-like virus non classifiés et utilisant un code génétique alternatif des mitochondries d'invertébrés (Figure 32).

**Tableau 23. Quelques contigs de *Picobirnaviridae* identifiées chez les palourdes Sanaga.**

Site	Date	N° d'Accession*	Taille (nt)	Reads mappées	Match Nucléotidique	Identité (%)	E-value	Noms	Réf. séquences
Bol	Septembre 18	MZ291977	399	260	388 <sup>#</sup>	94,1	2.00 x 10 <sup>-164</sup>	Human picobirnavirus, India-stool	AB517739.1
Bol	Décembre 18	MW924363	409	8	403 <sup>#</sup>	92,1	9.65 x 10 <sup>-158</sup>		
Bol	Janvier 19	MW924361	582	10	581 <sup>#</sup>	93,1	0		
Bol	Avril 19	MW924360	615	6	615 <sup>#</sup>	93,1	0		
Bol	Septembre 18	MW924362	645	905	639 <sup>#</sup>	93,9	0		
Bol	Janvier 19	MW924364	417	53	418 <sup>#</sup>	90,4	3.57 x 10 <sup>-152</sup>	Human picobirnavirus, India-stool	AB517731.1
Mou	Octobre 19	MW924365	539	72	303 <sup>#</sup>	88,1	3.05 x 10 <sup>-94</sup>	Human picobirnavirus, Belgium-stool	KU892528.1
Mou	Octobre 19	MW924366	429	63	429 <sup>#</sup>	92,8	9.95 x 10 <sup>-173</sup>	Human picobirnavirus, Cameroon-stool	MH933803.1
Bol	Septembre 18	MW924367	600	328	600 <sup>#</sup>	94,7	0	Picobirnavirus, RDC-gorilla feces	KY502865.1
Bol	Septembre 18	MW924368	600	286	600 <sup>#</sup>	96	0		
Bol	Septembre 18	MW924369	503	294	503 <sup>#</sup>	98,8	0		
Bol	Mai 18	MZ291978	378	18	203 <sup>#</sup>	90,6	2.79 x 10 <sup>-68</sup>	Macaque picobirnavirus, USA-macaque	MG010919.1
Bol	Octobre 19	MW924372	2023	2894	1725	33,7	8.90 x 10 <sup>-70</sup>	Picobirna-like virus, Cameroon-bats	AWV66962.1
Mou	Octobre 19	MW924371	2029	3215	1689	33,6	2.20 x 10 <sup>-68</sup>		

\*Ce numéro d'accèsion correspond à la séquence identifiée dans cette étude et déposée dans GenBank, les contigs ont été identifiés avec DIAMOND en utilisant des alignements protéiques, sauf pour certaines séquences notées avec # pour lesquelles le BLAST nucléotidique a été utilisé. L'identité de la séquence a été exprimée en pourcentage avec la E-value. nt : nucléotides. Réf. séquence : numéro d'accèsion de la séquence la plus proche sur GenBank.



**Figure 32. Arbre phylogénétique des séquences RdRp de picobirnavirus.**

Les séquences de référence ont été sélectionnées pour les trois principaux groupes de picobirnavirus, les séquences identifiées dans cette étude sont écrites en rouge. L'arbre a été construit à l'aide de IQtree avec le modèle de substitution rtREV+F+G4, et enracinés à partir du point moyen à l'aide de FigTree. Seuls les *bootstrap* supérieurs à 70 % sont présentés. PBV : picobirnavirus

## *Picornaviridae*

Trente-huit *contigs* assemblés à partir 69517 *reads*, et avec des tailles allant de 353 à 4631 nucléotides, ont été identifiés comme appartenant à la famille des *Picornaviridae*. Un *contig* de 353 nucléotides génétiquement proche du génogroupe V du VHA simien a été identifié dans un échantillon collecté à Bolounga en octobre 2019 (Tableau 24). Cette séquence partageait une identité nucléotidique de 91,73 % avec une portion du gène de la capsidie d'une souche qui serait d'origine simienne (Rachida *et al.*, 2016) détectée dans des d'eau d'irrigation en Afrique du Sud. Par ailleurs, elle présente 100% de similitude nucléotidique (sur une couverture de 18% de sa longueur) avec une séquence que nous avons précédemment identifiée lors du typage.

Les autres *contigs* apparentés aux *Picornaviridae* présentaient de faibles pourcentages de similitudes protéiques avec des séquences précédemment décrites. À l'exception d'un *contig* qui a montré une identité de 76 % avec les picornavirus de chauves-souris, tous les autres variaient de 26 à 38 % d'identité, cependant avec des *E-value* acceptables. Par ailleurs, ce faible niveau d'identité n'est pas corrélé au nombre de *reads* mappées. Les *contigs* proches du picornavirus de chauve-souris étaient les plus abondants (42 %) et étaient présents dans 68 % des échantillons. Avec des tailles allant de 382 à 4631 nucléotides, la plupart d'entre eux étaient plus proches au niveau protéique de la souche *Washington Bat* picornavirus (Tableau 24). D'autres *contigs* ont été identifiés comme étant proches du picornavirus porcin (453 à 1220 nucléotides), et présentaient de faibles similitudes avec le sapelovirus A, le sapelovirus 1 ou le kobuvirus, et quelques-uns avec le megrivirus infectant les poules.

**Tableau 24. Quelques contigs de *Picornaviridae* identifiées dans les palourdes de la Sanaga.**

Site	Date	N° d'Accession*	Taille (nt)	Read mappées	Match nucléotidique	Identité (%)	E-value	Noms	Ref. séquence
Bol	Octobre 19	MW924373	353	6	351 <sup>#</sup>	91,7	6,57 x 10 <sup>-134</sup>	Hepatitis A virus, South Africa-water	KP875240.1
Bol	Octobre 19	MZ291981	403	16	402	76,1	2,30 x 10 <sup>-54</sup>	Bat picornavirus, USA-bat ( <i>Washington Bat picornavirus</i> )	YP_009272812.1
Mou	Janvier 20	MZ291982	4631	2926	960	26,6	3,90 x 10 <sup>-20</sup>		
Bol	Mai 18	MZ291983	453	45	426	37,3	6,30 x 10 <sup>-16</sup>	Sapelovirus, USA-swine	ANY94479.1
Bol	Août 18	MZ291984	609	332	426	37,3	1,10 x 10 <sup>-15</sup>		
Mou	Avril 19	MZ291985	453	111	426	38,0	3,70 x 10 <sup>-16</sup>		
Mou	Novembre 19	MZ291986	784	146	735	31,0	7,50 x 10 <sup>-17</sup>	Porcine sapelovirus, Japan-swine	BBG56086.1
Mou	Avril 19	MZ291987	632	98	534	35,4	2,30 x 10 <sup>-16</sup>	Porcine kobuvirus, China-swine	AJW76498.1
Mou	Février 20	MZ291988	1220	247	525	36,0	5,80 x 10 <sup>-16</sup>		
Bol	Octobre 19	MZ291989	862	230	699	34,8	3,10 x 10 <sup>-24</sup>	Chicken megrivirus, Hungary-chicken	AHN10844.1
Bol	Décembre 19	MZ291990	938	18	648	35,2	1,70 x 10 <sup>-23</sup>		AYH53256.1

\*Ce numéro d'accession correspond à la séquence identifiée dans cette étude et déposée dans Genbank. Les contigs ont été identifiés avec DIAMOND en utilisant des alignements protéiques, sauf pour certaines séquences notées avec # pour lesquelles le BLAST nucléotidique a été utilisé. L'identité de la séquence a été exprimée en pourcentage avec la E-value. Réf. séquence : numéro d'accession de la séquence la plus proche sur GenBank.

## *Hepeviridae*

Un grand nombre de *contigs* (188) de longueurs variables (300 à 8223 nucléotides) représentant 390728 *reads* ont été identifiés comme appartenant à la famille des *Hepeviridae*, la plupart d'entre eux (92 %) étant identifiés comme proches du hepe-*like* virus infectant les éponges. Quatre *contigs* présentaient de fortes similitudes avec des séquences du virus de l'hépatite E de porc et de lapin, ainsi que de l'hepe-*like* virus de souris (Tableau 25).

**Tableau 25. Quelques *contigs* d'*Hepeviridae* identifiées dans les palourdes de la Sanaga.**

Site	Date	N° d'Accession*	Taille (nt)	Reads mappées	Match Nucléotidique	Identité (%)	E-value	Noms	Réf. séquence
Bol	Février 20	MZ476780	586	49	498	85%	$7,40 \times 10^{-17}$	Hepatitis E virus, China-swine	ACD44941.2
Bol	Décembre 18	MZ476781	572	46	570	100%	$1,30 \times 10^{-18}$	Hepatitis E virus, China-rabbit	BAM68889.1
Bol	Avril 19	MZ476782	558	34	498	89%	$1,10 \times 10^{-22}$	Hepe-like virus, USA-mice	AWB14594.1
Mou	Juin 19	MZ476783	611	96	537	88%	$4,50 \times 10^{-17}$		

\*Ce numéro d'accession correspond à la séquence identifiée dans cette étude et déposée dans GenBank. L'identification des contigs avec DIAMOND à partir des séquences protéiques. L'identité de la séquence a été exprimée en pourcentage avec la E-value. Ref. séquence : numéro d'accession de la séquence la plus proche sur GenBank.

### 3.2. Discussion

La détection des norovirus et du virus de l'hépatite A dans la majorité des échantillons de palourdes de la Sanaga analysés n'est pas surprenante. En effet, comme tous les bivalves, elles peuvent accumuler dans leurs tissus toutes sortes de particules présentes dans le milieu aquatique, y compris des virus libérés dans l'environnement à travers les selles des populations infectées en l'absence d'une station d'épuration des eaux usées efficace (Suffredini *et al.*, 2017, Haramoto *et al.*, 2018) comme c'est le cas sur le site de récolte de la Sanaga. Ce dernier étant situé dans une localité rurale, il ne dispose pas d'un réseau d'assainissement ou de traitement des déchets. Ceci favorise le déversement des matières fécales dans le lit du fleuve. La présence de ces virus dans ces palourdes confirme leur circulation dans cette localité. En effet, bien qu'aucune étude ne traite de la prévalence de ces virus dans la localité, de nombreuses études ont identifié les norovirus comme agent causal des diarrhées chez les enfants en Afrique (Mans *et al.*, 2016), ainsi que dans de nombreuses localités au Cameroun (Ayukekbong *et al.*, 2011 ; Ayukekbong *et al.*, 2014 ; Mugyia *et al.*, 2019). Plusieurs études ont par ailleurs identifié le virus de l'hépatite A comme étant un virus hautement endémique en Afrique subsaharienne avec des séroprévalences allant de 57 % chez les enfants et les adolescents à 95 % chez les adultes entre 2008 et 2018 (Jacobsen et Wiersma, 2010 ; Patterson *et al.*, 2019). Les concentrations de norovirus étaient similaires à celles obtenues par Strubbia *et al.*, en 2019 dans leur étude menée en France sur des huîtres naturellement contaminées. Sur le plan spatiale, les prévalences plus importantes à Moulongo pourraient se justifier par les situations géographiques des deux sites. En effet, Moulongo étant situé en aval de Bolounga, ce site recevrait les contaminations issues de Bolounga situé en amont.

La détection des norovirus et bien plus, du virus de l'hépatite A pour la première fois dans les palourdes de la Sanaga confirme les risques sanitaires que courent les populations en absence d'un traitement approprié aussi bien de ces bivalves que de l'eau de ce fleuve avant leur consommation. Elle pourrait en effet justifier l'une des origines des maladies diarrhéiques ainsi que la recrudescence des hépatites couramment appelées jaunisse par les populations de la localité qui vivent essentiellement de la pêche de la palourde. D'où l'intérêt d'une forte sensibilisation des populations au traitement préalable et adéquat aussi bien de cette eau que des palourdes avant leur consommation afin de réduire les risques de survenance des pathologies liées à ces virus.

Les virus entériques sont caractérisés par une grande diversité génétique, avec des génotypes spécifiques à différents hôtes. Ainsi, typer les virus détectés et en déduire leur origine nous a semblé important. Les norovirus constituent une cause majeure des gastroentérites virales sporadiques et épidémiques à travers le monde. Bien que leur impact sanitaire et économique dans les pays pauvres ou en voie de développement ne soit pas bien établi du fait du manque de données, l'épidémiologie moléculaire de ces virus entre 1997 et 2018 dans ces pays a montré que les génotypes les plus impliqués aussi bien dans les cas de gastroentérites que dans les infections asymptomatiques sont ceux appartenant au génogroupe II (GII). Avec une prédominance du génotype GII.4 suivit des génotypes GII.3 et GII.6. Le génotype GI.3 étant quant à lui le plus fréquent en ce qui concerne le génogroupe I (GI) (Mans, 2019). L'identification de ces génotypes infectant les humains dans les palourdes de la Sanaga, à l'exception du génotype GI.3 remplacé dans le cadre cette étude par le génotype GI.1, est donc en accord avec ces quelques données épidémiologiques disponibles. Ceci confirme les résultats d'une étude menée par Mugyia *et al.*, qui mettent en évidence la présence de ces derniers chez les enfants atteints de gastroentérite dans la ville de Yaoundé (Mugyia *et al.*, 2019). Par ailleurs, l'identification des P-types GII.[P31] et GII.[P21] a été reportée chez les enfants de moins de 5 ans présentant des symptômes de gastroentérites et admis dans des formations sanitaires dans la ville de Yaoundé et ses environs entre janvier 2010 et décembre 2013 (Mugyia *et al.*, 2020). Selon ces auteurs, ces deux types basés sur la polymérase représentaient respectivement 41,2% et 5,9% des GII détectés. Le P-type GII.[P31] détecté dans les palourdes de la Sanaga était très proches d'une séquence détectée au Gabon, pays frontalier. Les données actuelles ne fournissent pas d'informations sur la circulation au Cameroun du P-type GII.P[13] également détecté dans nos échantillons. Toutefois, la séquence détectée dans le cadre de cette étude a présenté une forte homologie avec une souche identifiée chez les enfants atteints de gastroentérite au Ghana. L'identification de ces P-types au sein des populations camerounaises et les similitudes observées entre les séquences détectées dans les palourdes de la Sanaga et celles identifiées chez des patients dans d'autres pays du golfe de Guinée pourraient traduire une forte circulation dans la sous-région.

Initialement classés sur la base du gène de la protéine VP1 de la capsid, une nouvelle approche de classification des norovirus intégrant également le typage du gène de l'ARN polymérase ARN dépendant a récemment été proposée. Aboutissant à la description d'un génotype et d'un P-type (Chhabra *et al.*, 2019). Le typage simultané de la protéine VP1 de la capsid et de la RdRp a été possible dans deux échantillons (Bolounga Septembre 2018 et

Bolounga Décembre 2018). Suggérant ainsi la présence du variant GII.4\_Sydney\_2012[P31] qui est le dernier variant pandémique de norovirus (Parra, 2019) et du type GII.6[P13]. A notre connaissance, notre étude est la première à associer le génotype GII.6 au P-type GII.[P13] qui pourrait constituer un nouveau recombinant encore inconnu. Toutefois, les palourdes et les échantillons environnementaux en général étant sujets à une contamination virale multiple, il ne faudrait pas exclure l'hypothèse d'un typage de souches distinctes du GII présentes dans les échantillons concernés. Nos résultats confirment la contamination des eaux du fleuve Sanaga par des rejets humains, et la contamination virale des palourdes de ce fleuve reflèterait l'épidémiologie moléculaire des virus circulants au sein des populations riveraines et de la sous-région.

Le virus de l'hépatite A est l'agent responsable des hépatites virales aiguës le plus fréquent à travers le monde. Sur la base de l'analyse de la jonction VP1/P2A, ce virus est classé en 6 génotypes (GI, GII, GIII, GIV, GV, et GVI) parmi lesquels trois génotypes (GI, GII, et GIII) sont responsables des infections chez les humains, tandis que les trois autres (GIV, GV, et GVI) sont d'origine simienne (Smith et Simmonds, 2018). La détection d'une souche apparentée au génotype V dans les palourdes de la Sanaga suggère une origine simienne. En effet, la proximité de la souche Sanaga avec des souches simiennes du virus de l'hépatite A et la présence documentée de primates non humains dans la zone d'échantillonnage (Ajonina *et al.*, 2005) laissent présager une contamination des eaux du fleuve par les matières fécales provenant des singes. La détection des norovirus humains et du virus de l'hépatite A apparenté aux souches infectant les primates non-humains démontre la contamination des palourdes de la Sanaga aussi bien par des virus humains qu'animaux. Ces organismes apparaissent ainsi aux vues de ces premières données comme des sentinelles fiables pour la détermination de la qualité des eaux du fleuve de la Sanaga. Ceci suggère qu'ils peuvent refléter la diversité des virus circulant dans les eaux du site de récolte en particulier et dans la localité en général, et par conséquent d'envisager leur usage dans la surveillance de la contamination virale de l'environnement.

La pandémie actuelle de COVID-19 a montré l'importance de surveiller la transmission possible des virus de la faune à l'homme. Ainsi, pouvoir évaluer la diversité des virus dans leurs réservoirs naturels pouvant représenter une population animale diversifiée, et prédire leur potentiel zoonotique constitue une étape critique dans la compréhension des dynamiques environnementales qui sous-tendent l'émergence de nouveaux pathogènes viraux (Osterhauss *et al.*, 2020). Dans cette optique, nous avons étudié la diversité des virus à ARN des palourdes

de la Sanaga afin d'identifier les virus pouvant présenter un intérêt en santé humaine et d'envisager l'utilisation des palourdes ou d'autres coquillages comme sentinelles de la contamination environnementale par des virus animaux et/ou humains. En effet, les palourdes, comme les autres mollusques bivalves filtrent de grands volumes d'eau, mais aussi en tant qu'animaux benthiques, elles sont en contact étroit avec des particules sédimentées telles que celles issues des matières fécales. Dans certaines régions où la collecte d'échantillons d'eau représentatifs peut être difficile, les mollusques bivalves peuvent être une bonne alternative.

Préserver la diversité génétique virale présente dans un échantillon reste un grand défi dans toutes les analyses métagénomiques, puisque les protocoles utilisés à ce jour contiennent des étapes qui peuvent sélectionner certaines particules virales au détriment d'autres. Il s'agit entre autres : (1) du biais d'échantillonnage, car la collecte d'échantillons dans l'environnement peut être source de variations, (2) des méthodes de concentration qui peuvent éliminer certains virus d'intérêt ou augmenter la présence d'inhibiteurs pouvant avoir un impact sur l'efficacité des enzymes utilisées pour la détection moléculaire, (3) de l'extraction des acides nucléiques et la préparation ultérieure des bibliothèques, y compris les traitements aux DNase ou RNase, ou encore (4) la dernière étape du traitement bio-informatique. De plus, la recherche de virus présents en faibles concentrations avec des génomes à ARN courts constitue une difficulté supplémentaire. Un certain nombre d'approches différentes peuvent aider à résoudre certains de ces problèmes (Briese *et al.*, 2015, Conceicao-Neto *et al.*, 2015 ; Kim *et al.*, 2017 ; Desdouits *et al.*, 2020). Des travaux précédents effectués par Strubbia *et al.* en 2020 ont proposé des approches méthodologiques visant à éliminer le plus possible les bactéries, les tissus de coquillages, ou les ADN et ARN libres, pour se concentrer sur les virus à ARN. Cela explique probablement la faible quantité de *reads* de virus à ADN et de phages dans notre étude, contrairement aux études métagénomiques agnostiques où ces virus constituaient la plus grande fraction virale (Bačnik *et al.*, 2020 ; Nieuwenhuijse *et al.*, 2020). Nous avons également effectué des triplets techniques d'ADNc et de bibliothèques, car cela augmente les chances de séquencer les souches virales présentes à de faibles concentrations, avec pour conséquence l'augmentation de la diversité (Kim *et al.*, 2017 ; Strubbia *et al.*, 2020). De plus, une étape d'enrichissement basée sur la capture de l'ADNc des virus des vertébrés à l'aide du panel de sonde VirCapSeq-VERT lors de la préparation des bibliothèques a été réalisée (Wylie *et al.*, 2015 ; Strubbia *et al.*, 2019). Enfin, compte tenu des faibles concentrations de norovirus et de virus de l'hépatite A mesurées par PCR quantitative, nous avons opté pour un séquençage Illumina sur la plateforme NovaSeq 6000 (pouvant générer un nombre record de 20 milliards de *reads*) afin

d'augmenter la profondeur par échantillon et par conséquent le nombre de *reads* viraux. En effet, de très grands nombres de *reads* par échantillon ont été obtenus. Toutefois, comme prévu, le nombre de *reads* viraux par échantillon est demeuré minoritaire par rapport au nombre de *reads* global obtenu. Cependant, l'assemblage *de novo* a été long et difficile pour ces ensembles de données très volumineux et diversifiés, certains échantillons ne produisant aucun *contig* assemblé malgré plusieurs jours de calcul. Il a été démontré que la déduplication des *reads* permet un assemblage plus rapide et plus facile en supprimant les *reads* répétitifs principalement générés par l'amplification requise pour atteindre des quantités minimales d'ADN nécessaires pour le séquençage (Nieuwenhuijse *et al.*, 2020). Ce traitement bio-informatique appliqué à nos données a considérablement réduit les fractions de *reads* bactériens et eucaryotes, augmenté le nombre de *reads* non classées, tandis que les fractions de *reads* viraux sont restées stables dans tous les échantillons. Après déduplication, l'assemblage *de novo* a donné des *contigs* assemblés dans tous les échantillons, avec des temps d'analyse bio-informatiques réduits.

Un autre problème majeur dans l'analyse métagénomique des échantillons environnementaux est la présence d'une grande quantité de séquences non identifiables appelées matière noire (*dark matter*), formellement définie comme un ensemble de séquences nucléotidiques qui ne peuvent pas être attribuées taxonomiquement, car ne pouvant s'aligner à aucune des séquences nucléotidiques ou protéiques présentes dans les bases de données (Krishnamurthy et Wang, 2017 ; Sunagawa *et al.*, 2020). Les échantillons dérivés de mélanges complexes tels que ceux provenant du tube digestif des mammifères ou de l'environnement peuvent inclure de nouveaux organismes non décrits dans les bases de données de référence ou non documentés (Connor *et al.*, 2019). Ainsi, la caractérisation de la fraction virale est particulièrement difficile car seule une petite fraction de la diversité virale de la Terre a été identifiée et mise à disposition dans des bases de données de référence (Connor *et al.*, 2019 ; Plyusnin *et al.*, 2020 ; Sunagawa *et al.*, 2020). Ceci est d'autant plus difficile lorsque l'on s'intéresse aux virus à ARN caractérisés par un taux de mutation élevé et donc, une grande diversité. Dans notre étude, jusqu'à 87 % des *reads* n'ont pu être identifiés et 96 % des *reads* d'ARN viraux n'ont pu être attribués à une famille virale. De telles observations précédemment reportées confirment que bien que les progrès du séquençage à haut débit (NGS) repoussent continuellement les limites des connaissances sur les séquences virales, certaines optimisations sont nécessaires pour augmenter les annotations virales (Bačnik *et al.*, 2020 ; Nieuwenhuijse *et al.*, 2020 ; Dance *et al.*, 2021). Il est donc important de continuer à améliorer les *pipelines* des

traitements bio-informatiques, ainsi que les bases de données génomiques afin de pouvoir rapidement analyser les données métagénomiques et pleinement identifier les organismes présents dans les échantillons (Nieuwenhuijse *et al.*, 2020 ; Plyusnin *et al.*, 2020 ; Dance *et al.*, 2021).

Afin d'identifier de manière fiable les séquences très diverses obtenues à la suite de l'assemblage *de novo*, chaque *contig* a été identifié par BLAST en tenant compte de la longueur du chevauchement avec la séquence correspondance de la base de données, du nombre de *reads* et de la *E-value* de l'alignement. La diversité virale dans notre étude a été décrite pour tous les échantillons ensemble plutôt que de comparer les deux sites ou les saisons, car les mêmes séquences ont été détectées dans les deux sites et considérant que le nombre d'échantillons et la durée de l'étude n'étaient pas être assez longs pour être capable de détecter certaines différences.

Le peu de séquences de virus à ADN à l'instar des bactériophages obtenu dans cette étude est contraire aux observations faites dans les études de métagénomique virale agnostique effectuée dans les eaux usées par Bacnik *et al.*, (2020) et Nieuwenhuijse *et al.*, (2020), et pour lesquelles ces virus représentaient une moyenne de jusqu'à 77% de séquences virales identifiées. Ces grandes différences pourraient s'expliquer par le fait que dans notre protocole, nous avons utilisé divers traitements tels que la filtration, le traitement à la DNase visant à se concentrer sur les virus à ARN. La plus grande fraction de séquences de virus à ARN identifiées appartenait aux virus de plantes avec une grande abondance de la famille des *Virgaviridae*. De telles observations ont également été faites par les précédents auteurs cités pour qui les virus des plantes constituaient la seconde fraction en terme d'abondance dans les eaux usées, dont les eaux usées issues d'une centaine de pays à travers le monde (Nieuwenhuijse *et al.*, 2020). Dans le contexte de notre étude, ce résultat se justifierait par l'abondant couvert végétal présent dans la localité qui se situe au cœur de la forêt équatoriale tel que reporté par Ajonina *et al.*, (2005). Nous ne nous sommes pas intéressés à la caractérisation approfondie de ces virus de végétaux. Toutefois selon certains auteurs (Nieuwenhuijse *et al.*, 2020), une attention devrait leur être portée à l'avenir. Notamment certains *Virgaviridae* infectant les légumes qui comme certains bactériophages pourraient, au regard de leur abondance dans l'environnement, constituer de bons indicateurs de pollutions fécales humaines.

Dans une approche « *One-Health* », nous nous sommes intéressés à l'identification de séquences de virus d'animaux, plus particulièrement les principales familles de virus de mammifères potentiellement liés à la santé humaine et connues pour être présentes dans les excréments humains et/ou animaux tels que les *Astroviridae*, *Caliciviridae*, *Hepeviridae*,

*Picobirnaviridae* et *Picornaviridae*. Peu de séquences de virus humains ont été identifiées par rapport à la diversité des autres séquences détectées. Cela peut être dû à la faible densité de population et à d'autres paramètres environnementaux tels que les courants ou l'évacuation des eaux usées qui n'ont pas été étudiés ici. Comme mentionné ci-dessus, tous les échantillons sélectionnés pour l'analyse métagénomique, à l'exception d'un seul, ont été détectés positifs pour le norovirus par RT-PCR en temps réel. Malgré ces premiers résultats de criblage, très peu de séquences apparentées ont été obtenues, ce qui signifierait que la sensibilité de l'approche métagénomique doit être améliorée. Dans une étude précédente, notre laboratoire d'analyse une grande diversité de norovirus a pu être obtenue dans des huîtres présentant des concentrations similaires à celles observées dans ces échantillons de palourdes (Strubbia *et al.*, 2019). Les résultats obtenus dans la présente étude pourraient s'expliquer par l'étape d'enrichissement qui n'aurait peut-être pas réussi à capturer toutes les souches présentes dans les extraits d'acides nucléiques, ou alors par la très grande diversité et l'abondance des séquences obtenues qui auraient pu empêcher l'assemblage de *contigs*. Les palourdes étant des organismes benthiques, elles pourraient concentrer des quantités plus élevées d'organismes très diversifiées par rapport aux huîtres. En effet, des études ont mis en évidence des abondances virales 600 à 2500 fois plus élevées à la surface des sédiments que dans la colonne d'eau (Kopylov *et al.*, 2020). Les données sur la prévalence des norovirus dans cette région sont encore rares et à notre connaissance, très peu d'études ont été publiées. D'après l'analyse d'un large panel de selle de tous les groupes d'âge dans la localité voisine de Limbe, peu de séquences de norovirus ont été identifiées par rapport à une diversité beaucoup plus élevée des séquences de sapovirus non détectées dans notre étude (Yinda *et al.*, 2019). Cette observation met en évidence la nécessité d'obtenir plus de données sur l'épidémiologie moléculaire de cet agent pathogène viral dans la population locale. Contrairement aux norovirus, de nombreuses séquences de picobirnavirus ont été identifiées. Mais, malgré plusieurs tentatives, seuls de courts *contigs* ont pu être assemblés. Qualifiés de virus opportunistes, leur rôle dans la survenue de gastro-entérite chez l'homme est encore peu étudié et le véritable hôte reste énigmatique (Ghosh et Malik, 2021). Cependant, une grande diversité de séquences de picobirnavirus a été observée dans l'analyse du virome intestinal des populations de la localité de Limbe au Cameroun. La plupart d'entre elles étant détectées au sein des tranches d'âge supérieures à 20 ans (Yinda *et al.*, 2019). Ces auteurs ont également évoqué la caractérisation au sein des populations d'une séquence du picobirna-like virus à 99% identique à une souche détectée dans une chauve-souris camerounaise. Mettant ainsi en évidence une circulation de virus entre le monde animal et les populations humaines.

La plupart des *contigs* identifiés dans cette étude comme provenant des virus de mammifères étaient étroitement liés à des virus infectant les chauves-souris. Nous avons caractérisé deux séquences de plus de deux mille nucléotides montrant des similitudes avec une souche de picobirnavirus précédemment décrite chez les roussettes camerounaises (Yinda *et al.*, 2018). De plus, une séquence de *bastro-like* virus était étroitement liée à une souche infectant les chauves-souris au Vietnam mais aussi au Cameroun (Yinda *et al.*, 2018). Plusieurs séquences liées au calicivirus de chauve-souris ont également été identifiées dans plusieurs échantillons. Tous ces *contigs* apparentés au calicivirus ont montré les meilleures homologies avec la souche A10 identifiée dans les excréments de chauve-souris collectées aux États-Unis (Kocher *et al.*, 2018). Le gène de la capsid de ce calicivirus A10 de chauve-souris était étroitement apparenté au norovirus du génogroupe V (GV) et phylogénétiquement proche des recovirus. Bien plus, cette souche présentait des similitudes avec les norovirus humains en termes de schéma de liaison aux antigènes du groupe majeur histocompatibilité sanguin humain. Suggérant le potentiel de ce virus à surmonter un obstacle important à la transmission inter-espèces (Kocher *et al.*, 2018). La transmission zoonotique potentielle est une question de longue date concernant les norovirus, avec des données rares signalées jusqu'à présent, soulignant ainsi la nécessité d'identifier leur réservoir potentiel (Villabruna *et al.*, 2019). Des séquences de picornavirus de chauve-souris ont également été identifiées dans plusieurs échantillons, confirmant une contamination de la zone étudiée et des eaux de ce fleuve par les excréments de chauves-souris. Ces mammifères sont considérés comme un réservoir majeur de virus, et plusieurs études ont rapporté la caractérisation de virus étroitement liés aux virus humains, confirmant leur rôle potentiel dans les événements zoonotiques (Hardmeier *et al.*, 2018 ; Yinda *et al.*, 2018). Il est donc important de considérer la contamination environnementale par ces animaux car elle peut constituer un facteur important de contamination humaine (Gortazar *et al.*, 2014 ; Kohl *et al.*, 2021).

Parmi les séquences de virus infectant les animaux détectées dans notre étude, celles des virus infectant les porcs arrivaient en deuxième position en termes de diversité. Nous avons détecté des sapelovirus et des kobuvirus de la famille des *Picornaviridae*, mais aussi des bastrovirus, des *bastro-like* virus et des virus de l'hépatite E. La plupart des épidémies d'hépatite E chez l'homme ont été liées à la consommation de porc, mais les données sur la contamination environnementale sont encore limitées. Compte tenu de la capacité de ce virus à franchir les barrières inter-espèces, la surveillance de la prévalence de ce dernier chez les porcs peut être intéressante (Sooryanarain et Meng, 2019).

Lors des tests de génotypage, nous avons séquencé une partie du génome d'un virus de l'hépatite A apparenté au génogroupe V qui comprend des virus simiens, et étroitement apparenté à un virus potentiellement simien identifié dans les eaux usées en Afrique du Sud. Nous avons de nouveau détecté cette séquence ainsi que quelques séquences de picobirnavirus simiens, confirmant la contamination de la zone par des excréments de singes. Le gisement de palourde du fleuve Sanaga étant situé dans une zone rurale abritant une réserve faunique caractérisée par une faune abondante et diversifiée, dont une importante population simienne (Ajonina *et al.*, 2005). Des transmissions zoonotiques de primates non-humains à l'homme ont été prouvées pour différentes familles virales (Devaux *et al.*, 2019 ; Di Cola *et al.*, 2021). Cependant, aucune des souches virales détectées dans les viromes des populations humaines de la localité de Limbe au Cameroun n'était liée aux singes dans l'étude menée en 2014 (Yinda *et al.*, 2019).

Nos résultats suggèrent que les conditions de propagation zoonotique des agents pathogènes sont réunies dans cette zone et que notre approche peut contribuer à sa surveillance (Osterhauss *et al.*, 2020 ; Williams *et al.*, 2021). En effet, malgré le besoin d'une meilleure sensibilité, nous rapportons ici une méthode qui a réussi à décrire une grande diversité de virus à ARN à partir d'échantillons de palourdes. En plus de leur rôle de sentinelle de la qualité microbienne de l'environnement, les coquillages constituent une source facile de protéines pour les populations locales aux faibles ressources économiques (Ajonina *et al.*, 2005 ; Venugopal *et Gopakumar*, 2018). Bien que le mode de consommation (bouilli, fumé) empêcherait probablement la transmission de ces virus par voie alimentaire, les premières étapes d'ouverture de la coquille et de collecte de la chair crue peuvent être à risque pour les travailleurs. Ainsi, il est important de sensibiliser et de fournir quelques recommandations aux populations locales.

Les approches métagénomiques sont des outils très prometteurs pour la description de la diversité virale et seront particulièrement intéressantes pour l'analyse d'échantillons alimentaires et environnementaux (Desdouits *et al.*, 2020). Elles sont particulièrement pertinentes pour l'analyse des coquillages. En effet, les mollusques bivalves, source de nutriments pour de nombreuses populations, peuvent également refléter une contamination microbienne causée par les animaux ou les humains environnants. Pouvoir décrire tous les contaminants viraux dans les mollusques bivalves, qu'ils soient considérés comme aliment ou comme sentinelle environnementale peut aider à anticiper certains événements de transmission zoonotique et alerter les autorités sanitaires d'une éventuelle émergence.

## CONCLUSION ET PERSPECTIVES

Les virus à d'origine alimentaire sont devenus les principaux agents étiologiques des infections d'origine alimentaire et constituent de ce fait un problème émergent de santé publique. Plusieurs aliments peuvent leur servir de véhicules parmi lesquels les plus importants sont les mollusques bivalves dont la contamination s'effectue principalement dans l'environnement aquatique. Ceci est à l'origine d'une grande diversité virale avec pour conséquence des diagnostics erronés lorsque les approches de détection sont ciblées. Dans ce contexte les approches agnostiques telles que la métagénomique pourraient constituer des alternatives prometteuses pour l'étude de la diversité virale. Dans ce travail, nous avons évalué la diversité des virus à ARN hébergés par les palourdes de la Sanaga. A cet effet, nous avons tout d'abord détecté et quantifié les norovirus et le virus de l'hépatite A qui constituent les principaux pathogènes viraux transmis par les bivalves, puis nous avons déterminé la diversité des virus à ARN hébergés par cette palourde, et enfin nous nous sommes intéressés à l'identification des virus à ARN pouvant constituer un danger chez les humains.

Il est ressorti de notre étude une forte prévalence de la contamination virale de ce mollusque avec la présence des norovirus et/ou du virus de l'hépatite A dans 80% des échantillons, et une prévalence de contamination plus importante à Moulongo. Par ailleurs, l'identification de plusieurs génotypes correspondant à des virus infectant l'homme ou les singes, notamment le variant pandémique GII.4 Sydney\_2012 du norovirus, ainsi qu'un virus de l'hépatite A apparenté au génotype V infectant les singes confirme la contamination des palourdes de la Sanaga par des virus entériques d'origines humaine et animale. Ce qui interpelle sur les potentiels risques d'infections virales des populations liées à la consommation de ces palourdes en l'absence des procédés de traitements adéquats.

Sur le plan de la diversité, il ressort que l'approche métagénomique optimisée utilisée a permis d'identifier une grande diversité virale dans les palourdes de la Sanaga. Celle-ci allant des bactériophages aux virus entéro-pathogènes humains, en passant par les virus d'algues, de plantes, d'invertébrés, d'oiseaux, et de plusieurs mammifères regroupés au sein d'une vingtaine de familles virales. Toutefois, une fraction majoritaire des séquences virales n'a pas pu être identifiée, soulignant la nécessité d'enrichir les banques génomiques disponibles.

S'agissant des virus à ARN susceptibles de constituer un danger en santé humaine+, à l'exception des séquences des picobirnavirus et de quelques-unes de norovirus similaires à des séquences identifiées chez des patients atteints de gastroentérites, peu de séquences correspondant aux virus entéro-pathogènes humains ont été détectées. La majorité des

séquences identifiées étant en effet proches des séquences de virus isolés chez les animaux, notamment les chauves-souris, les primates non humains et les porcs reconnus comme réservoirs d'une grande diversité de virus, avec des séquences similaires à celles de virus présentant des aptitudes à franchir les barrières inter-espèces pour infecter l'Homme. Au-delà de mettre en évidence pour la première fois des virus animaux dans les mollusques bivalves, ces résultats suggèrent d'une part que les conditions de propagation zoonotique des agents pathogènes sont réunies dans cette zone. D'autre part, les palourdes de la Sanaga peuvent constituer des sentinelles pour la surveillance de la contamination environnementale dans un contexte marqué par l'émergence de nouveaux agents pathogènes viraux issus de l'environnement.

### **Perspectives**

La réussite de l'approche métagénomique décrite, ainsi que la connaissance de la diversité et l'origine des virus détectés dans les palourdes de la Sanaga ouvrent de perspectives nouvelles qui mériteraient d'être explorées. Il s'agit entre autres de :

#### **1. Utiliser les palourdes de la Sanaga comme indicateurs de contamination environnementale**

Dans ce travail, l'approche métagénomique a permis d'identifier des séquences très proches de séquences virales précédemment détectées dans les populations humaines et dans une grande diversité d'hôtes animaux. Ceci trouve sa justification dans le fait que les mollusques bivalves sont des intégrateurs de la contamination de l'environnement. Nous envisageons dans la continuité de ce travail de rechercher la présence des virus détectés aux seins des populations humaines et animales de la localité, afin de proposer l'utilisation de cette palourde comme sentinelle pour le suivie des virus circulant dans cet environnement et la surveillance des potentiels phénomènes d'émergence.

#### **2. Evaluer l'infectiosité et les capacités zoonotiques de certains virus identifiés**

Plusieurs séquences ont présenté de fortes similitudes avec des séquences de virus précédemment détectées dans des hôtes (chauves-souris, primates non humains, porcs, rats) couramment impliqués dans des zoonoses d'origine virale. Dans un contexte mondiale marqué par l'émergence et la réémergence d'agents viraux zoonotiques, nous souhaitons réaliser à la suite de ce travail une étude sur l'infectiosité et les capacités zoonotiques de certains virus détectés. Ceci se fera à travers des tests d'infectiosité sur culture cellulaires, et l'étude des chemins de liaisons des déterminants antigéniques exprimés à la surface des particules virales

isolées par culture cellulaire ou des *Virus-Like Particules* (particules pseudo-virales dépourvues de génomes) obtenues par l'assemblage spontané de protéines de la capsid issues de l'expression des gènes correspondants dans un organisme recombinant.

### **3. Evaluer l'influence des procédés de traitement locaux (cuisson et fumage) sur la survie des virus présents dans ces palourdes**

Les palourdes de la Sanaga sont à la base de l'alimentation des populations pour lesquelles elles constituent la principale source de protéine. Elles sont généralement consommées après cuisson ou fumage, traitements susceptibles de détruire les différents pathogènes dont les virus. Toutefois, des épidémies d'hépatite A liées à la consommation de palourdes cuites ont été signalées dans le monde. Sur le plan local, il est connu de façon empirique que la consommation des palourdes de la Sanaga est souvent associée aux douleurs intestinales et aux diarrhées qui constituent des symptômes des gastroentérites. Ainsi, nous souhaitons à l'issue du présent travail, explorer l'impact de la cuisson et du fumage sur la désactivation des virus présents dans cette palourde afin de cerner le réel risque d'infections virales associé à la consommation de ces palourdes.

## RECOMMANDATIONS

- Sensibiliser les populations locales et les consommateurs de ces palourdes sur les dangers virologiques qui y sont associés, et sur la nécessité d'une cuisson poussée avant leur consommation.
- Sensibiliser les autorités publiques et sanitaires sur la nécessité d'effectuer un suivi de la qualité microbiologique et virologique en particulier de cette ressource afin de prévenir les risques de contamination à la suite de leur consommation.
- Encourager l'utilisation des palourdes comme indicateurs de la contamination de l'environnement par les virus d'origines humaine et animale afin de prévenir l'émergence de nouveaux agents pathogènes dont les agents viraux potentiellement zoonotiques.
- Mettre en place des unités de traitement des eaux usées le long du fleuve et encourager la construction des latrines aménagées dans toutes les habitations afin d'éviter le déversement des matières fécales dans le lit du fleuve.

## REFERENCES BIBLIOGRAPHIQUES

Aarestrup, F. M., Brown, E. W., Detter, C., Gerner-Smidt, P., Gilmour, M. W., Harmsen, D., Hendriksen, R. S., Hewson, R., Heymann, D. L., Johansson, K., Ijaz, K., Keim, P. S., Koopmans, M., Kroneman, A., Lo Fo Wong, D., Lund, O., Palm, D., Sawanpanyalert, P., Sobel, J., & Schlundt, J. (2012). Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. *Emerging Infectious Diseases*, 18(11), e1. <https://doi.org/10.3201/eid1811.120453>

Aigrain, L., Gu, Y. and Quail, M. A. (2016). Quantitation of next generation sequencing library preparation protocol efficiencies using droplet digital PCR assays - a systematic comparison of DNA library preparation kits for Illumina sequencing. *BMC Genomics*, 17(1), 1-11. <https://doi.org/10.1186/s12864-016-2757-4>

Ajonina, P. U., Ajonina, G. N., Jin, E., Mekongo, F., Ayissi, I. & Usongo, L. (2005). Gender roles and economics of exploitation, processing and marketing of bivalves and impacts on forest resources in the Sanaga Delta region of Douala-Edea Wildlife Reserve, Cameroon, *International Journal of Sustainable Development & World Ecology*, 12(2), 161-172, <http://dx.doi.org/10.1080/13504500509469627>

Ali, W., Zhang, H., Wang, Z., Chang, C., Javed, A., Ali, K., Du, W., Niazi, N. K., Mao, K., & Yang, Z. (2021). Occurrence of various viruses and recent evidence of SARS-CoV-2 in wastewater systems. *Journal of Hazardous Materials*, 414, 125439. <https://doi.org/10.1016/j.jhazmat.2021.125439>

Andrade, K. R., Boratto, P. P., Rodrigues, F. P., Silva, L. C., Dornas, F. P., Pilotto, M. R., La Scola, B., Almeida, G. M., Kroon, E. G., & Abrahão, J. S. (2015). Oysters as hot spots for mimivirus isolation. *Archives of Virology*, 160(2), 477-482. <https://doi.org/10.1007/s00705-014-2257-2>

Arapov, J., Ezgeta-Balié, D., Peharda, M., & Gladan, Z. N. (2010). Bivalve feeding-how and what they eat? *Ribarstvo*, 68, 105-116.

Atmar, R. L., Neill, F. H., Romalde, J. L., Le Guyader, F., Woodley, C. M., Metcalf, T. G., & Estes, M. K. (1995). Detection of Norwalk virus and hepatitis A virus in shellfish tissues

with the PCR. *Applied and Environmental Microbiology*, 61, 3014-3018. <https://doi.org/10.1128/aem.61.8.3014-3018.1995>

**Aw, T. G., Howe, A., & Rose, J. B. (2014).** Metagenomic approaches for direct and cell culture evaluation of the virological quality of wastewater. *Journal of Virological Methods*, 210, 15-21. <https://doi.org/10.1016/j.jviromet.2014.09.017>

**Ayukekbong, J. A., Andersson, M. E., Vansarla, G., Tah, F., Nkuo-Akenji, T., Lindh, M., & Bergström, T. (2014).** Monitoring of seasonality of norovirus and other enteric viruses in Cameroon by real-time PCR: An exploratory study. *Epidemiology and Infection*, 142(7), 1393-1402. <http://dx.doi.org/10.1017/S095026881300232X>

**Ayukekbong, J. A., Lindh, M., Nenonen, N., Tah, F., Nkuo-Akenji, T., & Bergström, T. (2011).** Enteric viruses in healthy children in Cameroon: Viral load and genotyping of norovirus strains. *Journal of Medical Virology*, 83, 2135-2142. <https://doi.org/10.1002/jmv.22243>

**Bačnik, K., Kutnjak, D., Pecman, A., Mehle, N., Tušek Žnidarič, M., Gutiérrez Aguirre, I., & Ravnikar, M. (2020).** Viromics and infectivity analysis reveal the release of infective plant viruses from wastewater into the environment. *Water Resources*. 177: 115628. <https://doi.org/10.1016/j.watres.2020.115628>

**Bleidorn, C. (2016).** Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. *Systematics and Biodiversity*, 14(1), 1-8. <https://doi.org/10.1080/14772000.2015.1099575>

**Boone, M., De Koker, A., & Callewaert, N. (2018).** Capturing the 'ome': the expanding molecular toolbox for RNA and DNA library construction. *Nucleic Acids Research*, 46(6), 2701-2721. <https://doi.org/10.1093/nar/gky167>

**Bosch, A., Gkogka, E., Le Guyader, F. S., Loisy-Hamon, F., Lee, A., van Lieshout, L., Marthi, B., Myrmel, M., Sansom, A., Schultz, A. C., Winkler, A., Zuber, S., & Phister, T. (2018).** Foodborne viruses: detection, risk assessment, and control options in food processing. *International Journal of Food Microbiology*, 285, 110-128. <https://doi.org/10.1016/j.ijfoodmicro.2018.06.001>

**Bosch, A., Pintó, R. M., & Guix, S. (2014).** Human Astroviruses. *Clinical Microbiology Reviews*, 27(4), 1048-1074. <https://doi.org/10.1128/CMR.00013-14>

**Bosch, A., Sánchez, G., Le Guyader, F., Vanaclocha, H., Haugarreau, L., & Pintó, R. M. (2001).** Human enteric viruses in Coquina clams associated with a large hepatitis A outbreak. *Water Science and Technology*, 43(12), 61-65. <https://doi.org/10.2166/wst.2001.0712>

**Bosch, A., Sánchez, G., Le Guyader, F., Vanaclocha, H., Haugarreau, L., & Pintó, R. M. (2015).** Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *mBio*, 6(5), e01491-15. <https://doi.org/10.1128/mBio.01491-15>

**Bozkurt, H., Phan-Thien, K.-Y., van Ogtrop, F., Bell, T., & McConchie, R., (2021).** Outbreaks, occurrence, and control of norovirus and hepatitis a virus contamination in berries: A review. *Critical Reviews in Food Science and Nutrition*, 61(1), 116-138. <https://doi.org/10.1080/10408398.2020.1719383>

**Briese, T., Kapoor, A., Mishra, N., Jain, K., Kumar, A., Jabado, et al. (2015).** Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *mBio*, 6(5), e01491-15. <https://doi.org/10.1128/mBio.01491-15>

**Bronner, I. F., Quail, M. A., Turner, D. J., & Swerdlow, H. (2014).** Improved Protocols for Illumina Sequencing. *Current Protocols in Human Genetics*, 80, 18.2.1–18.2.42. <https://doi.org/10.1002/0471142905.hg1802s79>

**Buchfink, B., Xie C., & Huson D. H. (2015).** Fast and sensitive protein alignment using diamond. *Nature Methods*, 12, 59-60. <https://doi.org/10.1038/nmeth.3176>

**Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009).** BLAST+: architecture and applications. *BMC Bioinformatics*, 10: 421. <https://doi.org/10.1186/1471-2105-10-421>

**Cannon, J. L., Bonifacio, J., Bucardo, F., Buesa, J., Bruggink, L., Chan, M. C., Fumian, T. M., Giri, S., Gonzalez, M. D., Hewitt, J., Lin, J. H., Mans, J., Muñoz, C., Pan, C. Y., Pang, X. L., Pietsch, C., Rahman, M., Sakon, N., Selvarangan, R., Browne, H., ... Vinjé, J. (2021).** Global Trends in Norovirus Genotype Distribution among Children with Acute Gastroenteritis. *Emerging Infectious Diseases*, 27(5), 1438-1445. <https://doi.org/10.3201/eid2705.204756>

**Costafreda, M. I., Bosch, A., Pintó, R. M. (2006).** Development, evaluation, and standardization of a Real-Time TaqMan Reverse Transcription-PCR Assay for quantification of hepatitis A virus in clinical and shellfish samples. *Applied and Environmental Microbiology*, 72(6), 3846- 3855. <https://doi.org/10.1128/AEM.02660-05>

**Ceylan, R. F., Ozkan, B., & Mulazimogullari, E., (2020).** Historical evidence for economic effects of COVID-19. *The European Journal of Health Economics*, 21, 817-823. <https://doi.org/10.1007/s10198-020-01206-8>

**Chen, M. Y., Chen, W. C., Chen, P. C. Hsu, S. W., & Lo, Y. C. (2016).** An outbreak of norovirus gastroenteritis associated with asymptomatic food handlers in Kinmen, Taiwan. *BMC Public Health*, 16, 372. <https://doi.org/10.1186/s12889-016-3046-5>

**Chhabra, P., de Graaf, M., Parra, G. I., Chan, M. C., Green, K., Martella, V., Wang, Q., White, P. A., Katayama, K., Vennema, H., Koopmans, M., & Vinjé, J. (2019).** Updated classification of norovirus genogroups and genotypes. *The Journal of General Virology*, 100(10), 1393-1406. <https://doi.org/10.1099/jgv.0.001318>

**Commission Européenne (2017).** Community Guide to the Principles of Good Practice for the Microbiological Classification and Monitoring of Bivalve Mollusc Production and Relaying Areas with regard to Regulation 854/2004. [https://ec.europa.eu/food/system/files/2018-12/biosafety\\_fh\\_guidance\\_community\\_guide\\_bivalve\\_mollusc\\_monitoring\\_en.pdf](https://ec.europa.eu/food/system/files/2018-12/biosafety_fh_guidance_community_guide_bivalve_mollusc_monitoring_en.pdf)

**Conaty, S., Bird, P., Bell, G., Kraa, E., Grohmann, G., & Mcanulty, J. (2000).** Hepatitis A in New South Wales, Australia, from consumption of oysters: The first reported outbreak. *Epidemiology and Infection*, 124(1), 121-130. <https://doi.org/10.1017/S0950268899003386>

**Conceição-Neto, N., Zeller, M., Lefrère, H., De Bruyn, P., Beller, L., Deboutte, W., Yinda, C. K., Lavigne, R., Maes, P., Van Ranst, M., Heylen, E., & Matthijssens, J. (2015).** Molecular approach to customize sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Scientific Reports*, 5, 16532. <https://doi.org/10.1038/srep16532>

**Connor, R., Brister, R., Buchmann, J. P., Deboutte, W., Edwards, R., Martí-Carreras, J., Tisza, M., Zalunin, V., Andrade-Martínez, J., Cantu, A., D'Amour, M., Efremov, A., Fleischmann, L., Forero-Junco, L., Garmeva, S., Giluso, M., Glickman, C., Henderson,**

**M., Kellman, B., Kristensen, D., ... Busby, B. (2019).** NCBI's Virus Discovery Hackathon: Engaging research communities to identify cloud infrastructure requirements. *Genes*, 10, 714. <https://doi.org/10.3390/genes10090714>

**Cortez, V., Meliopoulos, V. A., Karlsson, E. A., Hargest, V., Johnson, C., & Schultz-Cherry, S. (2017).** Astrovirus Biology and Pathogenesis. *Annual Review of Virology*, 4, 327-348. <https://doi.org/10.1146/annurev-virology-101416-041742>

**Costafreda, M. I., Bosch, A., & Pintó, R. M. (2006).** Development, evaluation, and standardization of a real-time TaqMan reverse transcription-PCR assay for quantification of hepatitis A virus in clinical and shellfish samples. *Applied and Environmental Microbiology*, 72(6), 3846–3855. <https://doi.org/10.1128/AEM.02660-05>

**Cotten, M., Oude Munnink, B., Canuti, M., Deijs, M., Watson, S. J., Kellam, P., & van der Hoek, L. (2014).** Full genome virus detection in fecal samples using sensitive nucleic acid preparation, deep sequencing, and a novel iterative sequence classification algorithm. *Plos One*, 9(4), e93269. <https://doi.org/10.1371/journal.pone.0093269>

**Coudray-Meunier, C., Fraise, A., Mokhtari, C., Martin-Latil, S., Roque-Afonso, A.-M., & Perelle, S. (2014).** Hepatitis A virus subgenotyping based on RT-qPCR assays. *BMC Microbiology*, 14, 296. <https://doi.org/10.1186/s12866-014-0296-1>

**Crawford, S. E., Ramani, S., Tate, J. E., Parashar, U. D., Svensson, L., Hagbom, M., Franco, M. A., Greenberg, H. B., O'Ryan, M., Kang, G., Desselberger, U., & Estes, M. K. (2017).** Rotavirus infection. *Nature Reviews Disease Primers*, 3, 17083. <https://doi.org/10.1038/nrdp.2017.83>

**Croci, L., De Medici, D., Scalfaro, C., Fiore, A., Divizia, M., Donia, D., Cosentino, A.M., Moretti, P. & Costantini, G. (2000).** Determination of enteroviruses, hepatitis A virus, bacteriophages and *Escherichia coli* in Adriatic Sea mussels. *Journal of Applied Microbiology*, 88, 293-298. <https://doi.org/10.1046/j.1365-2672.2000.00966.x>

**Dance, A. (2021).** The incredible diversity of viruses. *Nature Review*, 595, 22-25.

**da Silva, A. K., Le Saux, J. C., Parnaudeau, S., Pommepuy, M., Elimelech, M., & Le Guyader, F. S. (2007).** Evaluation of removal of noroviruses during wastewater treatment, using Real-Time Reverse Transcription-PCR: different behaviors of genogroups I and II.

*Applied and Environmental Microbiology*, 73(24), 7891-7897.  
<https://doi.org/10.1128/AEM.01428-07>

**Debing, Y., Moradpour, D., Neyts, J., & Gouttenoire, J. (2016).** Update on hepatitis E virology: Implications for clinical practice. *Journal of Hepatology*, 65 (1), 200-212.  
<https://doi.org/10.1016/j.jhep.2016.02.045>

**de Graaf, M., van Beek, J. & Koopmans, M. (2016).** Human norovirus transmission and evolution in a changing world. *Nature Reviews Microbiology*, 14, 421-433.  
<https://doi.org/10.1038/nrmicro.2016.48>

**Desdouits, M., de Graaf, M., Strubbia, S., Oude Munnink, B. B., Kroneman, A., Le Guyader, F. S., & Koopmans, M. (2020).** Novel opportunities for NGS-based one health surveillance of foodborne viruses. *One Health Outlook*, 2, 14. <https://doi.org/10.1186/s42522-020-00015-6>

**Desdouits, M., Piquet, J. C., Wacrenier, C., Le Menec, C., Parnaudeau, S., Jousse, S., Rocq, S., Bigault, L., Contrant, M., Garry, P., Chavanon, F., Gabellec, R., Lamort, L., Lebrun, L., Le Gall, P., Meteigner, C., Schmitt, A., Seugnet, J. L., Serais, O., Peltier, C., ... Le Guyader, F. S. (2021).** Can shellfish be used to monitor SARS-CoV-2 in the coastal environment?. *Science of The Total Environment*, 778, 146270.  
<https://doi.org/10.1016/j.scitotenv.2021.146270>

**Desdouits, M., Wacrenier, C., Ollivier, J., Schaeffer, J., & Guyader, F. (2020b).** A Targeted Metagenomics Approach to Study the Diversity of Norovirus GII in Shellfish Implicated in Outbreaks. *Viruses*, 12(9), 978. <https://doi.org/10.3390/v12090978>

**Devaux, C. A., Mediannikov, O., Medkour, H., & Raoult, D. (2019).** Infectious disease risk across the growing Human-nonhuman primate interface: a review of the evidence. *Frontiers in Public Health*, 7, 305. <https://doi.org/10.3389/fpubh.2019.00305>

**Dhama, K., Patel, S. K., Yattoo, M. I., Tiwari, R., Sharun, K., Dhama, J., Natesan, S., Malik, Y. S., Singh, K. P., & Harapan, H. (2021).** SARS-CoV-2 existence in sewage and wastewater: A global public health concern? *Journal of Environmental Management*, 280, 111825. <https://doi.org/10.1016/j.jenvman.2020.111825>

**Di Cola, G., Fantalli, A. X., Pisano, M. B., & Ré, V. E. (2021).** Foodborne transmission of hepatitis A and hepatitis E viruses: a literature review. *International Journal of Food Microbiology*, 338, 10898986. <https://doi.org/10.1016/j.ijfoodmicro.2020.108986>

**D'Souza, D. H., & Joshi, S. S. (2016).** Foodborne Viruses of Human Health Concern, Editor(s): Benjamin Caballero, Paul M. Finglas, Fidel Toldrá, *In Encyclopedia of Food and Health*, Academic Press, 87-93, ISBN 9780123849533. <https://doi.org/10.1016/B978-0-12-384947-2.00727-3>

**Duraisamy, R., Akiana, J., Davoust, B., Mediannikov, O., Michelle, C., Robert, C., Parra, H. J., Raoult, D., Biagini, P., & Desnues, C. (2018).** Detection of novel RNA viruses from free-living gorillas, Republic of the Congo: genetic diversity of picobirnaviruses. *Virus Genes*, 54, 256-271. <https://doi.org/10.1007/s11262-018-1543-6>

**Edgar, R. C. (2010).** Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), 2460-2461. <https://doi.org/10.1093/bioinformatics/btq461>

**EFSA 2014.** Scientific Opinion on health benefits of seafood (fish and shellfish) consumption in relation to health risks associated with exposure to methylmercury. *EFSA Journal* 2014, 12 (7), 3761, 80. <https://doi.org/10.2903/j.efsa.2014.3761>

**EFSA (2015).** The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2014. *EFSA Journal*, 13(12). <http://dx.doi.org/10.2903/j.efsa.2015.4329>

**Endoh, D., Mizutani, T., Kirisawa, R., Maki, Y., Saito, H., Kon, Y., Morikawa, S., & Hayashi, M. (2005).** Species-independent detection of RNA virus by representational difference analysis using non-ribosomal hexanucleotides for reverse transcription. *Nucleic Acids Research*, 33(6), e65. <https://doi.org/10.1093/nar/gni064>

**Etim, L., & Brey, T. (1994).** Growth, productivity, and significance for fishery of the bivalve *Egeria radiata* (Donacidae) in the Cross River, Nigeria. *Archive of Fishery and Marine Research*, 42(1), 63-75. <https://www.researchgate.net/publication/292007001>.

**FAO & WHO (2008).** Viruses in food: scientific advice to support risk management activities. *Microbiological Risk Assessment Series*. N°13. 6-7. <https://apps.who.int/iris/handle/10665/44030>

**Fiorito, F., Amoroso, M. G., Lambiase, S., Serpe, F. P., Bruno, T., Scaramuzzo, A., Maglio, P., Fusco, G., & Esposito, M. (2019).** A relationship between environmental pollutants and enteric viruses in mussels (*Mytilus galloprovincialis*). *Environmental Research*, 169, 156-162. <https://doi.org/10.1016/j.envres.2018.11.001>

**Ford-Siltz, L. A., Mullis, L., Sanad, Y. M., Tohma, K., Lepore, C. J., Azevedo, M., & Parra, G. I. (2019).** Genomics Analyses of GIV and GVI Noroviruses Reveal the Distinct Clustering of Human and Animal Viruses. *Viruses*, 11, 204. <https://doi.org/10.3390/v11030204>

**Franck, K. T., Lisby, M., Fonager, J., Schultz, A. C., Böttiger, B., Villif, A., Absalonsen, H., & Ethelberg, S. (2015).** Sources of calicivirus contamination in foodborne outbreaks in Denmark, 2005-2011--the role of the asymptomatic food handler. *The Journal of Infectious Diseases*, 211 (4), 563-570. <https://doi.org/10.1093/infdis/jiu479>

**Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012).** CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28, 3150-3152. <https://doi.org/10.1093/bioinformatics/bts565>

**Fusco, G., Di Bartolo, I., Cioffi, B. Ianiro, G. Palermo, P., & Monini, M. et al. (2017).** Prevalence of foodborne viruses in mussels in Southern Italy. *Food and Environmental Virology*, 9, 187-194. <https://doi.org/10.1007/s12560-016-9277-x>

**Fusco, G., Anastasio, A., Kingsley, D. H., Amoroso, M. G., Pepe, T., Fratamico, P. M., Cioffi, B., Rossi, R., La Rosa, G., & Boccia, F. (2019).** Detection of Hepatitis A Virus and Other Enteric Viruses in Shellfish Collected in the Gulf of Naples, Italy. *International Journal of Environmental Research and Public Health*, 16(14), 2588. <http://dx.doi.org/10.3390/ijerph16142588>

**Gao, S., Li, D., Zha, E., Zhou, T., Wang, S., & Yue, X. (2015).** Surveillance of hepatitis E virus contamination in shellfish in China. *International Journal of Environmental Research and Public Health*, 12(2), 2026-2036. <http://dx.doi.org/10.3390/ijerph120202026>

**Ghosh S., & Malik, Y. S. (2021).** The True Host/s of picobirnaviruses. *Frontiers in Veterinary Science*, 7, 615293. <https://doi.org/10.3389/fvets.2020.615293>

**Gorji, M. E., Tan, M. T. H. & Li, D. (2021).** Influence of fucosidase producing bifidobacteria on the HBGA antigenicity of oyster digestive tissue and the associated norovirus binding.

*International Journal of Food Microbiology*, 340:109058.  
<https://doi.org/10.1016/j.ijfoodmicro.2021.109058>

**Gortazar, C., Reperant, L. A., Kuiken, T., de la Fuente, J., Boadella, M., Martínez-Lopez, B., Ruiz-Fons, F., Estrada-Peña, A., Drosten, C., Medley, G., Ostfeld, R., Peterson, T., VerCauteren, K. C., Menge, C., Artois, M., Schultsz, C., Delahay, R., Serra-Cobo, J., Poulin, R., Keck, F., ... Mysterud, A. (2014).** Crossing the interspecies barriers: opening the door to zoonotic pathogens. *Plos Pathogens*, 10, e1004129.  
<https://doi.org/10.1371/journal.ppat.1004129>

**Greening, G. E., & Cannon, J. L. (2016).** Human and animal viruses in food (Including Taxonomy of Enteric Viruses). *Viruses in Foods*, 5-57. [https://doi.org/10.1007/978-3-319-30723-7\\_2](https://doi.org/10.1007/978-3-319-30723-7_2)

**Grodzki, M., Ollivier, J., Le Saux, J. C., Piquet, J. C., Noyer, M., & Le Guyader, F. S. (2012).** Impact of xynthia tempest on viral contamination of shellfish. *Applied and Environmental Microbiology*, 78(9), 3508-3511. <https://doi.org/10.1128/AEM.07604-11>

**Guo, Y., Dai, Y., Yu, H., Zhao, S., Samuels, D., C., & Shyr, Y. (2017).** Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. *Genomics*, 109(2), 83-90. <https://doi.org/10.1016/j.ygeno.2017.01.005>

**Gužvić, M. (2013).** The History of DNA Sequencing / ISTORIJAT SEKVENCIRANJA DNK. *Journal of Medical Biochemistry*, 32(4), 301-312. <https://doi.org/10.2478/jomb-2014-0004>

**Hagbom, M., Lin, J., Falkeborn, T., Serrander, L., Albert, J., Nordgren, J., & Sharma, S. (2021).** Replication in human intestinal enteroids of infectious Norovirus from vomit samples. *Emerging Infectious Diseases*, 27(8), 2212-2214.  
<https://doi.org/10.3201/eid2708.210011>

**Halliday, M. L., Kang, L. Y., & Zhou, T. K. (1991).** An epidemic of hepatitis A attributable to the ingestion of raw clams in Shanghai, China. *The Journal of Infectious Diseases*, 164(1), 852-859. <https://doi.org/10.1093/infdis/164.5.852>

**Haramoto, E., Kitajima, M., Hata, A., Torrey, J. R., Masago, Y., Sano, D., & Katayama, H. (2018).** A review on recent progress in the detection methods and prevalence of human

enteric viruses in water. *Water Research*, 135, 168-186.  
<https://doi.org/10.1016/j.watres.2018.02.004>

**Hardmeier, I., Aeberhard, N., Qi, W., Schoenbaechler, K., Kraettli, H., Hatt, J. M., Fraefel, C., & Kubacki, J. (2018).** Metagenomic analysis of fecal and tissue samples from 18 endemic bat species in Switzerland revealed a diverse virus composition including potentially zoonotic viruses. *Plos One*, 16, e025534. <https://doi.org/10.1371/journal.pone.0252534>

**Hardstaff, J. L., Clough, H. E., Lutje, V., McIntyre, K. M., Harris, J. P., Garner, P., & O'Brien, S. J. (2018).** Foodborne and food-handler norovirus outbreaks: A systematic review. *Foodborne Pathogens and Disease*, 15 (10):589-97. <https://doi.org/10.1089/fpd.2018.2452>

**Hatha A. A. M., Christi K. C., Singh R., & Kumar S. (2005).** Bacteriology of the freshwater bivalve clam *Batissa violacea* (Kai) sold in the Suva market. *The South Pacific Journal Natural Science*, 23, 48-50. <https://doi.org/10.1071/SP05009>

**Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014).** Library construction for next-generation sequencing: Overviews and challenges. *Biotechniques*, 56(2), 61-77.  
<https://doi.org/10.2144/000114133>

**Heather, J. M., & Chain, B. (2016).** The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107 (1), 1-8. <https://doi.org/10.1016/j.ygeno.2015.11.003>

**Helm, M. M., Bourne, N., Lovatelli, A. (2004).** Hatchery culture of bivalves: A practical manual. *FAO Fisheries Technical Paper*, 471, 177.  
<http://www.fao.org/3/y5720e/y5720e00.htm>

**Hernández, E., de Castro, V., Avellón, A., González, I., Muniozguren, N., Vázquez, S., & Muñoz-Chimeno, M. (2019).** Hepatitis A outbreak associated with a food handler in Bizkaia, 2017. *Enfermedades infecciosas y microbiología clinica* (English ed.), 37(9), 569-573.  
<https://doi.org/10.1016/j.eimce.2019.04.014>

**Hutin, Y. J., Pool, V., Cramer, E. H., Nainan, O. V., Weth, J., Williams, I. T., Goldstein, S. T., Gensheimer, K. F., Bell, B. P., Shapiro, C. N., Alter, M. J., & Margolis, H. S. (1999).** A multistate, foodborne outbreak of hepatitis A. National Hepatitis A Investigation Team. *The*

*New England Journal Of Medicine*, 340(8), 595-602.  
<https://doi.org/10.1056/NEJM199902253400802>

**Iaconelli, M., Purpari, G., Libera, S. D., Petricca, S., Guercio, A., & Ciccaglione, A. R. (2015).** Hepatitis A and E viruses in wastewaters, in river waters, and in bivalve molluscs in Italy. *Food and Environmental Virology*, 7, 316-324. <https://doi.org/10.1007/s12560-015-9207-3>

**Iizuka, S., Oka, T., Tabara, K., Omura, T., Katayama, K., Takeda, N., & Noda, M. (2010).** Detection of sapoviruses and noroviruses in an outbreak of gastroenteritis linked genetically to shellfish. *Journal of Medical Virology*, 82: 1247-1254. <https://doi.org/10.1002/jmv.21791>

**ISO 15216-1 (2017).** Microbiology of food and animal feed -horizontal method for determination of hepatitis A virus and norovirus in food using real-time RT-PCR, Part 1: method for quantification.

**Itarte, M., Martínez-Puchol, S., Forés, E., Hundesa, A., Timoneda, N., Bofill-Mas, S., Girones, R., & Rusiñol, M. (2021).** NGS techniques reveal a high diversity of RNA viral pathogens and Papillomaviruses in fresh produce and irrigation water. *Foods*, 10, 1820. <https://doi.org/10.3390/foods10081820>

**Ito, E., Pu, J., Miura, T., Kazama, S., Nishiyama, M., Ito, H., Konta, Y., Omura, T., & Watanabe, T. (2021).** Detection of Rotavirus Vaccine Strains in Oysters and Sewage and Their Relationship with the Gastroenteritis Epidemic. *Applied and Environmental Microbiology*, 87, e02547-20. <https://doi.org/10.1128/AEM.02547-20>

**Iwamoto, M., Ayers, T., Mahon, B. E., & Swerdlow, D. L. (2010).** Epidemiology of seafood-associated infections in the United States. *Clinical Microbiology Reviews*, 23(2), 399-411. <http://dx.doi.org/10.1128/cmr.00059-09>

**Jacobsen, K. H., & Wiersma, S. T. (2010).** Hepatitis A virus seroprevalence by age and world region, 1990 and 2005. *Vaccine*, 28(41), 6653-6657. <http://dx.doi.org/10.1016/j.vaccine.2010.08.037>

**Jiang, X., Wang, J., Graham, D. Y., & Estes, M. K. (1992).** Detection of Norwalk virus in stool by polymerase chain reaction. *Journal of clinical microbiology*, 30(10), 2529-2534. <https://doi.org/10.1128/jcm.30.10.2529-2534.1992>

**Jothikumar, N., Cromeans, T. L., Robertson, B. H., Meng, X. J., & Hill, V. R. (2006).** A broadly reactive one-step real-time RT-PCR assay for rapid and sensitive detection of hepatitis E virus. *Journal of Virological Methods*, 131(1), 65-71. <http://dx.doi.org/10.1016/j.jviromet.2005.07.004>

**Karlsson, E. A., & Schultz-Cherry, S. (2013).** Chapter 20-Astroviruses as Foodborne Infections, Editor(s): J. Glenn Morris, Morris E. Potter, *In Food Science and Technology, Foodborne Infections and Intoxications (Fourth Edition)*, Academic Press, 293-301, ISBN 9780124160415, <https://doi.org/10.1016/B978-0-12-416041-5.00020-2>.

**Kashnikov, A. Y., Epifanova, N. V., & Novikova, N. A. (2020).** Picobirnaviruses: prevalence, genetic diversity, detection methods. *Vavilovskii Zhurnal Genetiki i Seleksii*, 24(6), 661-672. <https://doi.org/10.18699/VJ20.660>

**Kim, O., Kagaya, Y., Tran, H. S., Minei, R., Tran, T., Duong, H., Le, B., Dang, L. T., Kinoshita, K., Ogura, A., & Yura, K. (2020).** A novel circular ssDNA virus of the phylum Cressnaviricota discovered in metagenomic data from otter clams (*Lutraria rhynchaena*). *Archives of virology*, 165(12), 2921–2926. <https://doi.org/10.1007/s00705-020-04819-9>

**Kim, Y., van Bonn, W., Aw, T. G., & Rose, J. B. (2017).** Aquarium viromes: viromes of human-managed aquatic systems. *Frontiers in Microbiology*, 8(1231). <https://doi.org/10.3389/fmicb.2017.01231>

**King, A. M. Q., Adams, M. J., Carstens, E. B., & Lefkowitz, E. J. (eds), (2012).** Virus taxonomy classification and nomenclature of viruses. *Ninth Report of the International Committee on Taxonomy of Viruses*. ISBN: 978-0-12-384684-6.

**Kittigul, L., Singhaboot, Y., Chavalitshewinkoon-Petmitr, P., Pombubpa, K., & Hirunpetcharat, C., (2015).** A comparison of virus concentration methods for molecular detection and characterization of rotavirus in bivalve shellfish species. *Food Microbiology*, 46, 161-167. <https://doi.org/10.1016/j.fm.2014.07.020>

**Kocher, J. F., Lindesmith, L. C., Debbink, K., Beall, A., Mallory, M. L., Yount, B. L., Graham, R. L., Huynh, J., Gates, J. E., Donaldson, E. F., & Baric, R. S. (2018).** Bat caliciviruses and human noroviruses are antigenically similar and have overlapping histo-blood group antigen binding profiles. *mBio*, 9: e00869-00818. <https://doi.org/10.1128/mBio.00869-18>

**Kohl, C., Nitsche, A., & Kurth, A. (2021).** Update on potentially zoonotic viruses of European bats. *Vaccines*, 9, 690. <https://doi.org/10.3390/vaccines9070690>

**Koopmans M. (2012).** Food-borne viruses from a global perspective. In: Institute of Medicine (US). Improving food safety through a One Health Approach: Workshop Summary. Washington (DC): *National Academies Press* (US). A9. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK114484/>

**Kopylov, A. I., Zobotkina, E. A., Romanenko, E. A., Kosolapov, D. B., & Sazhin, A. F. (2020).** Viruses in the water column and the sediment of the eastern part of the Laptev Sea. *Estuarine, Coastal and Shelf Science*, 242, 106836. <https://doi.org/10.1016/j.ecss.2020.106836>

**Krishnamurthy, S. R., & Wang, D. (2017).** Origins and challenges of viral dark matter. *Virus Research*, 239: 136-142. <https://doi.org/10.1016/j.virusres.2017.02.002>

**Krishnamurthy, S. R., & Wang, D. (2018).** Extensive conservation of prokaryotic ribosomal binding sites in known and novel picobirnaviruses. *Virology*, 516, 108-114. <https://doi.org/10.1016/j.virol.2018.01.006>

**Kroneman, A., de Souza, R., Verhoef, L., Koopmans, M. P. G., Vennema, H., & On Behalf of the HAVNetwork. (2018).** Usability of the international HAVNet hepatitis A virus database for geographical annotation, backtracing and outbreak detection. *Euro Surveillance*, 23(37), 1700802. <https://doi.org/10.2807/1560-7917.ES.2018.23.37.1700802>.

**Kroneman, A., Vennema, H., Deforche, K., v d Avoort, H., Peñaranda, S., Oberste, M. S., Vinjé, J., & Koopmans, M. (2011).** An automated genotyping tool for enteroviruses and noroviruses. *Journal of Clinical Virology*, 51(2), 121-125. <https://doi.org/10.1016/j.jcv.2011.03.006>.

**Kurup, K. K., Manickam, P., & Gurav, Y. (2020).** Infected food handlers led to an outbreak of hepatitis A in Ernakulam district, Kerala, Southern India, 2016. *Clinical Epidemiology and Global Health*, 8(1), 308-312. <https://doi.org/10.1016/j.cegh.2019.08.001>

**La Bella, G., Basanisi, M. G., Nobili, G., Terio, V., Suffredini, E., & La Salandra, G. (2021).** First Report of Hepatitis E Virus in Shellfish in Southeast Italy. *Applied Sciences*, 11(1), 43. <http://dx.doi.org/10.3390/app11010043>

**Langmead, B., & Salzberg, S. (2012).** Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357-359. <https://doi.org/10.1038/nmeth.1923>

**La Rosa, G., Proroga, Y., De Medici, D., Capuano, F., Iaconelli, M., Della Libera, S., & Suffredini, E. (2018).** First detection of hepatitis E virus in shellfish and in seawater from production areas in Southern Italy. *Food and Environmental Virology*, 10, 127-131. <https://doi.org/10.1007/s12560-017-9319-z>

**Lassus, P., Bardouil, M., Beliaeff, B., Bougrier, S., Gubbins M., Masselin, P., et al. (2002).** Coquillages contaminés par des phycotoxines paralysantes. Vers une procédure de détoxification. *Editions Ifremer*. <https://w3.ifremer.fr/archimer/recordview>

**Lee, L. J., and Abdullah, M. (2014).** Optimization of DNA shearing by sonication. *AsPac Journal of Molecular Biology and Biotechnology*, 22(3), 200-208. Retrieved from <https://pdfs.semanticscholar.org/b8bc/e77382a2c47c52bbbf38aa400a6bbcb4b6a.pdf>

**Lees, D. (2000).** Viruses and bivalve shellfish. *International Journal of Food Microbiology*, 59(1-2), 81-116. [https://doi.org/10.1016/S0168-1605\(00\)00248-8](https://doi.org/10.1016/S0168-1605(00)00248-8)

**Le Guyader, F. S., Atmar, R. L., & Le Pendu, J., (2012).** Transmission of viruses through shellfish: when specific ligands come into play. *Current Opinion in Virology*, 2(1), 103-110. <https://doi.org/10.1016/j.coviro.2011.10.029>

**Le Guyader, F., Haugarreau, L., Miossec, L., Dubois, E., & Pommepuy, M. (2000).** Three-year study to assess human enteric viruses in shellfish. *Applied and environmental microbiology*, 66(8), 3241-3248. <https://doi.org/10.1128/AEM.66.8.3241-3248.2000>

**Le Guyader, F. S., Le Saux, J. C., Ambert-Balay, K., Krol, J., Serais, O., Parnaudeau, S., Giraudon, H., Delmas, G., Pommepuy, M., Pothier, P., & Atmar, R. L. (2008).** Aichi virus, norovirus, astrovirus, enterovirus, and rotavirus involved in clinical cases from a French oyster-related gastroenteritis outbreak. *Journal of Clinical Microbiology*, 46(12), 4011-4017. <http://dx.doi.org/10.1128/JCM.01044-08>

**Le Guyader, F. S., Parnaudeau, S., Schaeffer, J., Bosch, A., Loisy, F., Pommepuy, M., & Atmar, R. L. (2009).** Detection and quantification of noroviruses in shellfish. *Applied and Environmental Microbiology*, 75(3), 618-624. <http://dx.doi.org/10.1128/aem.01507-08>

**Le Guyader, S., & Pommepuy, M. (2002).** Contamination virale des aliments : création d'un réseau Européen de surveillance et de recherche. *Revue Française des Laboratoires*, 348, 53-59. [https://doi.org/10.1016/S0338-9898\(02\)80308-0](https://doi.org/10.1016/S0338-9898(02)80308-0)

**Le Guernic, A., Palos Ladeiro, M., Boudaud, N., Do Nascimento, J., Gantzer, C., Inglard, J. C., Mouchel, J. M., Pochet, C., Moulin, L., Rocher, V., Waldman, P., Wurtzer, S., & Geffard, A. (2022).** First evidence of SARS-CoV-2 genome detection in zebra mussel (*Dreissena polymorpha*). *Journal of environmental management*, 301, 113866. <https://doi.org/10.1016/j.jenvman.2021.113866>

**Lin, Y. Y., Hsieh, C. H., Chen, J. H., Lu, X., Kao, J. H., Chen, P. J., Chen, D. S., & Wang, H. Y. (2017).** 'De novo assembly of highly polymorphic metagenomic data using in situ generated reference sequences and a novel BLAST-based assembly pipeline'. *BMC Bioinformatics*, 18(1), 1-10. <https://doi.org/10.1186/s12859-017-1630-z>

**Louten, J., (ed). (2016).** Chapter 5 - Virus transmission and epidemiology, *In Essential human virology. Academic Press*, 71-92. ISBN 9780128009475. <https://doi.org/10.1016/B978-0-12-800947-5.00005-3>

**Lowther, J. A., Bosch, A., Butot, S., Ollivier, J., Mäde, D., Rutjes, et al., (2019).** Validation of EN ISO method 15216 - Part 1 - Quantification of hepatitis A virus and norovirus in food matrices. *International Journal of Food Microbiology*, 288, 82-90. <https://doi.org/10.1016/j.ijfoodmicro.2017.11.014>

**Lyu, C., Lu, F., Shi, Z., & Wang, D., (2021).** Detection of group A rotavirus in oyster tissues by *in situ* capture RT-qPCR. *Food Control*, 127. <https://doi.org/10.1016/j.foodcont.2021.108161>

**Maalouf, H., Pommepuy, M., & Le Guyader, F. S. (2010).** Environmental conditions leading to shellfish contamination and related outbreaks. *Food and Environmental Virology*, 2(3), 136-145. <https://dx.doi.org/10.1007/s12560-010-9043-4>

**Maalouf, H., Schaeffer, J., Parnaudeau, S., Le Pendu, J., Atmar, R. L., Crawford, S. E., & Le Guyader, F. S. (2011).** Strain-dependent norovirus bioaccumulation in oysters. *Applied and Environmental Microbiology*, 77(10), 3189-3196. <https://doi.org/10.1128/AEM.03010-10>

- Macaluso, G., Guercio, A., Gucciardi, F., Di Bella, S., La Rosa, G., Suffredini, E., Randazzo, W., & Purpari, G. (2021).** Occurrence of Human Enteric Viruses in Shellfish along the Production and Distribution Chain in Sicily, Italy. *Foods*, 10(6), 1384. <http://dx.doi.org/10.3390/foods10061384>
- Mans, J. (2019).** Norovirus infections and disease in lower-middle and low-income countries, 1997-2018. *Viruses*, 11(4), 341. <https://doi.org/10.3390/v11040341>
- Mans, J., Armah, G. E., Steele, A. D., & Taylor, M. B. (2016).** Norovirus epidemiology in Africa: A Review. *Plos One*, 11(4), e0146280. <https://doi.org/10.1371/journal.pone.0146280>
- Marine, R., Polson, S. W., Ravel, J., Hatfull, G., Russell, D., Sullivan, M., Syed, F., Dumas, M., & Wommack, K. E. (2011).** Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. *Applied and Environmental Microbiology*, 77(22), 8071-8079. <https://doi.org/10.1128/AEM.05610-11>
- Marinho, A., Rocha, D., Kanai, Y. K., Alves, C. M., Costa, D. C., Sousa, A. H., Barros, B., Bonfim, M., & Mascarenhas, J. (2018).** Rotavirus analyses by SYBR Green real-time PCR and microbiological contamination in bivalves cultivated in coastal water of Amazonian Brazil. *Journal of Water and Health*, 16(6), 970-979. <https://doi.org/10.2166/wh.2018.130>
- Mateus, M., Miguel, R., LÍgia, P., & Alexandra, S. (2021).** Monitoring SARS-CoV-2 as a microbiological risk in shellfish aquaculture. *Frontiers in Marine Science*, 8, 473. <https://doi.org/10.3389/fmars.2021.669402>
- Maxam, A. M., & Gilbert, W. (1977).** A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74(2), 560-564. <https://doi.org/10.1073/pnas.74.2.560>
- McLeod, C., Polo, D., Le Saux, J.-C. & Le Guyader, F. S. (2017).** Depuration and Relaying: A Review on Potential Removal of Norovirus from Oysters. *Comprehensive Reviews in Food Science and Food Safety*, 16, 692-706. <https://doi.org/10.1111/1541-4337.12271>
- Menzel, P., Ng, K. L. & Krogh, A. (2016).** Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications*, 7, 1-9. <https://doi.org/10.1038/ncomms11257>

**Metcalf, T., Melnick, J. L., & Estes M. K. (1995).** Environmental Microbiology: from detection of virus in sewage and water by isolation to identification by molecular biology- a trip of over 50 years. *Annual Review of Microbiology*, 49, 461-487. <https://doi.org/10.1146/annurev.mi.49.100195.002333>

**Metzker, M. (2010).** Sequencing technologies - the next generation. *Nature Reviews Genetic*, 11, 31-46. <https://doi.org/10.1038/nrg2626>

**Mugyia, A. E., Ndze, V. N., Akoachere, J., Browne, H., Boula, A., Ndombo, P. K., Cannon, J. L., Vinjé, J., & Ndip, L. M. (2019).** Molecular epidemiology of noroviruses in children under 5 years of age with acute gastroenteritis in Yaoundé, Cameroon. *Journal of Medical Virology*, 91, 738-743. <https://doi.org/10.1002/jmv.25380>

**Mugyia, A. E., Ndze, V. N., Akoachere, J.-F. T. K., Boula, A., Koki Ndombo, P., & Ndip, L. M. (2020).** First cases of recombinants noroviruses in Cameroon. *Journal of Virology Research & Reports*, SRC/JVRR-112. [https://doi.org/10.47363/JVRR/2020\(1\)112](https://doi.org/10.47363/JVRR/2020(1)112)

**Nagashima, S., Takahashi, M., Kobayashi, T., Tanggis, Nishizawa, T., Nishiyama, T., Primadharsini, P. P., & Okamoto, H. (2017).** Characterization of the quasi-enveloped hepatitis E virus particles released by the cellular exosomal pathway. *Journal of Virology*, 91, e00822-17. <https://doi.org/10.1128/JVI.00822-17>

**Nieuwenhuijse, D. F., Oude Munnink, B. B., Phan, M., Global Sewage Surveillance project consortium, Munk, P., Venkatakrishnan, S., Aarestrup, F. M., Cotten, M., & Koopmans, M. (2020).** Setting a baseline for global urban virome surveillance in sewage. *Science Reports*, 10(13748). <https://doi.org/10.1038/s41598-020-69869-0>

**Nieuwenhuijse, D., & Koopmans, M. P. G. (2017).** Metagenomic sequencing for surveillance of food- and waterborne viral diseases. *Frontiers in Microbiology*, 8, 230. <https://doi.org/10.3389/fmicb.2017.00230>

**Nishida, T., Kimura, H., Saitoh, M., Shinohara, M., Kato, M., Fukuda, S., Munemura, T., Mikami, T., Kawamoto, A., Akiyama, M., Kato, Y., Nishi, K., Kozawa, K., & Nishio, O. (2003).** Detection, quantitation, and phylogenetic analysis of noroviruses in Japanese oysters. *Applied and Environmental Microbiology*, 69(10), 5782-5786. <https://doi.org/10.1128/AEM.69.10.5782-5786.2003>

- Nooij, S., Schmitz, D., Vennema, H., Kroneman, A., & Koopmans, M. P. G. (2018).** Overview of virus metagenomic classification methods and their biological applications. *Frontiers in Microbiology*, 9, 749. <https://doi.org/10.3389/fmicb.2018.00749>
- Nurk, S., Meleshko, D., Korobeynikov, A., & Tevzner, P. A. (2017).** metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, 27, 824-834. <https://doi.org/10.1101/gr.213959.116>
- O'Hara, Z., Crossan, C., Craft, J., & Scobie, L. (2018).** First report of the presence of hepatitis E virus in Scottish-harvested shellfish purchased at retail level. *Food and Environmental Virology*, 10, 217-221. <https://doi.org/10.1007/s12560-018-9337-5>
- Oka, T., Wang, Q., Katayama, K., & Saif, L. J. (2015).** Comprehensive review of human sapoviruses. *Clinical Microbiology Reviews*, 28, 32-53. <https://doi.org/10.1128/CMR.00011-14>.
- Osterhaus, A., Vanlangendonck, C., Barbeschi, M., Brusckke, C., Christensen, R., Daszak, P., de Groot, F., Doherty, P., Drury, P., Gmacz, S., Hamilton, K., Hart, J., Katz, R., Longuet, C., McLeay, J., Morelli, G., Schlundt, J., Smith, T., Suri, S., Umali, K., ... Wagenaar, J. A. (2020).** Make science evolve into a One Health approach to improve health and security: a white paper. *One Health Outlook*, 2, 6. <https://doi.org/10.1186/s42522-019-0009-7>
- Oude Munnink, B. B., Jazaeri Farsani, S. M., Deijns, M., Jonkers, J., Verhoeven, J. T., Ieven, M., Goossens, H., de Jong, M. D., Berkhout, B., Loens, K., Kellam, P., Bakker, M., Canuti, M., Cotten, M., & van der Hoek, L. (2013).** Autologous antibody capture to enrich immunogenic viruses for viral discovery. *Plos One*, 8(11), e78454. <https://doi.org/10.1371/journal.pone.0078454>
- Parra, G. I. (2019).** Emergence of norovirus strains: A tale of two genes. *Virus Evolution*, 5(2), vez048. <https://doi.org/10.1093/ve/vez048>
- Patterson, J., Abdullahi, L., Hussey, G. D., Muloiwa, R. & Kagina, B. M. (2019).** A systematic review of the epidemiology of hepatitis A in Africa. *BMC Infectious Diseases*, 19, 651. <https://doi.org/10.1186/s12879-019-4235-5>

- Pickett, B. E., Sadat, E. L., Zhang, Y., Noronha, J. M., Squires, R. B., Hunt, V., Liu, M., Kumar, S., Zaremba, S., Gu, Z., Zhou, L., Larson, C. N., Dietrich, J., Klem, E. B., & Scheuermann, R. H. (2012).** ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic acids research*, 40(Database issue), D593-D598. <https://doi.org/10.1093/nar/gkr859>
- Pintó, R. M., Costafreda, M. I., & Bosch, A. (2009).** Risk assessment in shellfish-borne outbreaks of hepatitis A. *Applied and Environmental Microbiology*, 75(23), 7350-7355. <https://doi.org/10.1128/AEM.01177-09>
- Plyusnin, I., Kant, R., Jääskeläinen, A. J., Sironen, T., Holm, L., Vapalahti, O., & Smura, T. (2020).** Novel NGS pipeline for virus discovery from a wide spectrum of hosts and sample types. *Virus Evolution*, 6 :,veaa091. <https://doi.org/10.1093/ve/veaa091>
- Polo, D., Lois, M., Fernandez-Nunez, M. T., & Romalde, J. L. (2021).** Detection of SARS-CoV-2 RNA in bivalve mollusks and marine sediments. *Science of Total Environment*, 786, 147534. <https://doi.org/10.1016/j.scitotenv.2021.147534>
- Polo, D., Schaeffer, J., Teunis, P., Buchet, V., & Le Guyader, F. S. (2018).** Infectivity and RNA persistence of a norovirus surrogate, the Tulane virus, in Oysters. *Frontiers in microbiology*, 9, 716. <https://doi.org/10.3389/fmicb.2018.00716>
- Polo, D., Varela, M., F., & Romalde, J. L. (2015).** Detection and quantification of hepatitis A virus and norovirus in Spanish authorized shellfish harvesting areas. *International Journal of Food Microbiology*, 193, 43-50. <https://doi.org/10.1016/j.ijfoodmicro.2014.10.007>
- Potasman, I., Paz, A., & Odeh, M. (2002).** Infectious outbreaks associated with bivalve shellfish consumption: A worldwide perspective. *Clinical Infectious Diseases*, 35, 921-928. <http://doi.org/10.1086/342330>
- Primadharsini, P. P., Nagashima, S., & Okamoto, H. (2019).** Genetic Variability and Evolution of Hepatitis E Virus. *Viruses*, 11(5), 456. <http://dx.doi.org/10.3390/v11050456>
- Rachida, S., Matsapola, P. N., Wolfaardt, M., & Taylor, M. B. (2016).** Genetic characterization of a novel hepatitis A virus strain in irrigation water in South Africa. *Journal of Medical Virology*, 88(4), 734-737. <https://doi.org/10.1002/jmv.24370>

- Rivadulla, E., Varela, M. F., Mesquita, J. R., Nascimento, M. S. J., & Romalde, J. L. (2019).** Detection of hepatitis E virus in shellfish harvesting areas from Galicia (Northwestern Spain). *Viruses*, 11(7), 618. <http://dx.doi.org/10.3390/v11070618>
- Rivadulla, E., & Romalde, J. L. (2020).** A comprehensive review on human Aichi Virus. *Virologica Sinica*, 35, 501-516. <https://doi.org/10.1007/s12250-020-00222-5>
- Romalde, J. L., Rivadulla, E., Varela, M. F., & Barja, J. L. (2017).** An overview of 20 years of studies on the prevalence of human enteric viruses in shellfish from Galicia, Spain. *Journal of Applied Microbiology*, 124(4), 943-957. <http://dx.doi.org/10.1111/jam.13614>
- Rosa, M., Ward, J. E., & Shumway, S. E. (2018).** Selective capture and ingestion of particles by suspension-feeding bivalve molluscs: A review. *Journal of Shellfish Research*, 37(4), 727-746. <https://doi.org/10.2983/035.037.0405>
- Rosani, U., & Gerdol, M. (2017).** A bioinformatics approach reveals seven nearly-complete RNA-virus genomes in bivalve RNA-seq data. *Virus Research*, 239, 33-42. <https://doi.org/10.1016/j.virusres.2016.10.009>
- Rosani, U., Shapiro, M., Venier, P., & Allam, B. (2019).** A Needle in A Haystack: Tracing Bivalve-Associated Viruses in High-Throughput Transcriptomic Data. *Viruses*, 11(3), 205. <http://dx.doi.org/10.3390/v11030205>
- Said, B., Ijaz, S., Kafatos, G., Booth, L., Thomas, H. L., Walsh, A., Ramsay, M., Morgan, D., & Hepatitis E Incident Investigation Team (2009).** Hepatitis E Outbreak on Cruise Ship. *Emerging Infectious Diseases*, 15(11), 1738-1744. <https://doi.org/10.3201/eid1511.091094>
- Savini, F., Giacometti, F., Tomasello, F., Pollesel, M., Piva, S., Serraino, A., & De Cesare, A. (2021).** Assessment of the Impact on Human Health of the Presence of Norovirus in Bivalve Molluscs: What Data Do We Miss?. *Foods*, 10(10), 2444. <https://doi.org/10.3390/foods10102444>
- Sbaoui, Y., Bennis, F., & Chegdani, F. (2021).** SARS-CoV-2 as enteric virus in wastewater: which risk on the environment and human behavior? *Microbiology Insights*, 41, 1-9. <https://doi.org/10.1177/1178636121999673>

- Shendure, J., Balasubramanian, S., Church, G., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston R. H. (2017).** DNA sequencing at 40: past, present and future. *Nature*, 550, 345-353. <https://doi.org/10.1038/nature24286>
- Shin, H., Park, H., Seo, D. J., Jung, S., Yeo, D., Wang, Z., Park, K. H., & Choi, C. (2019).** Foodborne Viruses Detected Sporadically in the Fresh Produce and Its Production Environment in South Korea. *Foodborne Pathogens and Disease*, 16(6), 411-420. <https://doi.org/10.1089/fpd.2018.2580>
- Siegel, R. D. (2018).** Classification of Human Viruses. *Principles and Practice of Pediatric Infectious Diseases*, 1044-1048.e1. <https://doi.org/10.1016/B978-0-323-40181-4.00201-2>
- Smaal, A. C., Ferreira, J. G., Grant, J., Petersen, J. K., & Strand Ø. (eds.), (2019).** Goods and Services of Marine Bivalves. <https://doi.org/10.1007/978-3-319-96776-9>
- Smith, D. B., & Simmonds, P. (2018).** Classification and Genomic Diversity of Enterically Transmitted Hepatitis Viruses. *Cold Spring Harbor perspectives in medicine*, 8(9), a031880. <https://doi.org/10.1101/cshperspect.a031880>
- Somura, Y., Nagano, M., Kimoto, K., Oda, M., Mori, K., Shinkai, T., & Sadamasu, K. (2019).** Detection of norovirus in food samples collected during suspected food-handler-involved foodborne outbreaks in Tokyo. *Letter in Applied Microbiology*, 69, 175-180. <https://doi.org/10.1111/lam.13189>
- Sooryanarain, H., & Meng, X-J. (2019).** Hepatitis E virus: reasons for emergence in humans. *Current Opinion in Virology*, 34: 10-17. <https://doi.org/10.1016/j.coviro.2018.11.006>
- Souza, D. S. M., Dominot, A.F. A., Moresco, V., & Barardi, C. R. M. (2017).** Presence of enteric viruses, bioaccumulation and stability in *Anomalocardia brasiliiana* clams (Gmelin, 1791). *International Journal of Food Microbiology*, 266, 363-371. <http://dx.doi.org/10.1016/j.ijfoodmicro.2017.08.004>
- Steinegger, M., Mirdita, M. & Söding, J. (2019).** Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold. *Nature Methods*, 16(7), 603-606. <https://doi.org/10.1038/s41592-019-0437-4> .

**Strubbia, S. (2019).** Norovirus et huître : Infectiosité et approche génomique. Thèse de Doctorat de l'Université de Nantes Comue Université de Bretagne Loire. Soutenue le 04/12/2019. <http://archive.bu.univ-nantes.fr/pollux/show.action?id=1c09c095-a1af-4640-b4b2-629c1644df27>

**Strubbia, S., Schaeffer, J., Besnard, A., Wacrenier, C., Le Mennec, C., Garry, P., Desdouits, M., & Le Guyader, F. S. (2020).** Metagenomic to evaluate norovirus genomic diversity in oysters: impact on hexamer selection and targeted capture-based enrichment. *International Journal of Food Microbiology*, 323: 108588. <https://doi.org/10.1016/j.ijfoodmicro.2020.108588>

**Strubbia, S., Schaeffer, J., Oude Munnink, B. B., Besnard, A., Phan, M., Nieuwenhuijse, D. F., de Graaf, M., Schapendonk, C., Wacrenier, C., Cotten, M., Koopmans, M., & Le Guyader, F. S. (2020).** Metavirome sequencing to evaluate norovirus diversity in sewage and related bioaccumulated oysters. *Frontiers in Microbiology*, 10, 3389. <https://doi.org/10.3389/fmicb.2019.02394>

**Suffredini, E., Proroga, Y., Di Pasquale, S., Di Maro, O., Losardo, M., Cozzi, L., Capuano, F., & De Medici, D. (2017).** Occurrence and trend of hepatitis A virus in bivalve mollusks production areas following a contamination event. *Food and Environmental Virology*, 9, 423-433. <https://doi.org/10.1007/s12560-017-9302-8>

**Sunagawa, S., Acinas, S. G., Bork, P., Bowler, C., Tara Oceans Coordinators, Eveillard, D., Gorsky, G., Guidi, L., Iudicone, D., Karsenti, E., Lombard, F., Ogata, H., Pesant, S., Sullivan, M. B., Wincker, P., & de Vargas, C. (2020).** Tara Oceans: towards global ocean ecosystems biology. *Nature Review*, 18. [https://doi.org/10.1038/590\\_s41579-020-0364-5](https://doi.org/10.1038/590_s41579-020-0364-5)

**Sutton, T., Clooney, A. G., Ryan, F. J., Ross, R. P., & Hill, C. (2019).** Choice of assembly software has a critical impact on virome characterisation. *Microbiome*, 7(1), 12. <https://doi.org/10.1186/s40168-019-0626-5>

**Tate, J. E., Burton, A. H., Boschi-Pinto, C., Parashar, U. D., & World Health Organization–Coordinated Global Rotavirus Surveillance Network (2016).** Global, regional, and national estimates of rotavirus mortality in children <5 years of age, 2000-2013. *Clinical Infectious Diseases*, 62(2), 96-105. <https://doi.org/10.1093/cid/civ1013>

- Tan, M., Ho, S. X., Chu, J., & Li, D. (2021).** Application of virome capture sequencing in shellfish sold at retail level in Singapore. *Letters in Applied Microbiology*, 73(4), 486-494. <https://doi.org/10.1111/lam.13540>
- Teunis, P., Le Guyader, F. S., Liu, P., Ollivier, J., & Moe, C. L. (2020).** Noroviruses are highly infectious but there is strong variation in host susceptibility and virus pathogenicity. *Epidemics*, 32, 100401. <https://doi.org/10.1016/j.epidem.2020.100401>
- Teunis, P. F., Sukhrie, F. H., Vennema, H., Bogerman, J., Beersma, M. F., & Koopmans, M. P. (2015).** Shedding of norovirus in symptomatic and asymptomatic infections. *Epidemiology and Infection*, 143(8), 1710-1717. <https://doi.org/10.1017/S095026881400274X>
- Thorne, L. G. & Goodfellow, I. G. (2014).** Norovirus gene expression and replication. *Journal of General Virology*, 95, 278-291. <https://doi.org/10.1099/vir.0.059634-0>
- van Dijk, E. L., Jaszczyszyn, Y., & Thermes, C. (2014).** Library preparation methods for next-generation sequencing: tone down the bias. *Experimental Cell Research*, 322(1), 12-20. <https://doi.org/10.1016/j.yexcr.2014.01.008>
- van Tong, H., Hoan, N. X., Wang, B., Wedemeyer, H., Bock, C. T., & Velavan, T. P. (2016).** Hepatitis E Virus Mutations: Functional and Clinical Relevance. *EBioMedicine*, 11, 31-42. <https://doi.org/10.1016/j.ebiom.2016.07.039>
- Varela, M. F., Polo, D., & Romalde, J. L. (2016).** Prevalence and genetic diversity of human sapoviruses in shellfish from commercial production areas in Galicia, Spain. *Applied and Environmental Virology*, 82(4), 1167-1172. <https://doi.org/10.1128/AEM.02578-15>
- Venugopal, V., & Gopakumar K. (2018).** Shellfish: nutritive value, health benefits, and consumer safety. *Comprehensive Review in Food Science and Food Safety*, 16, 1219-1242. <https://doi.org/10.1111/1541-4337.12312>
- Villabruna, N., Koopmans, M. P. G., & de Graaf, M. (2019).** Animals as reservoir for human norovirus. *Viruses*, 11, 478. <https://doi.org/10.3390/v11050478>

- Velebit, B., Djordjevic, V., Milojevic, L., Babic, M., Grkovic, N., Jankovic, V., & Yushina, Y. (2019).** The common foodborne viruses: A review. *IOP Conference Series: Earth and Environmental Science*, 333. <https://doi.org/10.1088/1755-1315/333/1/012110>
- Wang, H., Wang, X., Cao, J., Gao, Y., Zhou, W., & Bi, S. (2015).** Full-length genome characterization and quasispecies distribution of hepatitis A virus isolates in China. *Virology Reports*, 5, 29-46. <https://doi.org/10.1016/j.virep.2015.03.001>
- Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021).** Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, 39, 1348-1365. <https://doi.org/10.1038/s41587-021-01108-x>
- Werneck, L. M. C., Baptista, M., L., Miagostovich, M., P., & da Silva, E. E. (2019).** Dissemination of enteroviruses in the production chain of organic lettuce in Rio de Janeiro, Brazil. *MicrobiologyOpen*, 8, e653. <https://doi.org/10.1002/mbo3.653>
- Wood, D. E., & Salzberg, S. L. (2014).** Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, 15, R46. <https://doi.org/10.1186/gb-2014-15-3-r46>
- World Health Organization WHO (2008).** Viruses in food: Scientific advice to support risk management. <http://www.who.int/foodsafety/?/publications/?/micro/mra13/en/index.html>
- World Health Organization (WHO) (2020).** Modes of transmission of virus causing COVID-19: Implications for IPC Precaution Recommendations: Scientific Brief. 29 March 2020. World Health Organization. <https://apps.who.int/iris/handle/10665/331601>
- World Health Organization (WHO) (2021).** Rotavirus vaccines: WHO position paper-July 2021. *Weekly Epidemiological Record*, 96 (28), 301-319. World Health Organization. <https://apps.who.int/iris/handle/10665/342905>
- Williams, E. P., Spruill-Harrell, B. M., Taylor, M. K., Lee, J., Nywening, A. V., Yang, Z., Nichols, J. H., Camp, J. V., Owen, R. D., & Jonsson, C. B. (2021).** Common themes in zoonotic spillover and disease emergences: lessons learned from bat and rodent-borne RNA viruses. *Viruses*, 13, 1509. <https://doi.org/10.3390/v13081509>

- Wollants, E., Beller, L., Beuselinck, K., Bloemen, M., Lagrou, K., & Reynders, M., (2019).** A decade of enterovirus genetic diversity in Belgium. *Journal of Clinical Virology*, 121, 104205. <https://doi.org/10.1016/j.jcv.2019.104205>
- Wurtzer, S., Waldman, P., Ferrier-Rembert, A., Frenois-Veyrat, G., Mouchel, J. M., Boni, M., Maday, Y., OBEPINE consortium, Marechal, V., & Moulin, L. (2021).** Several forms of SARS-CoV-2 RNA can be detected in wastewaters: Implication for wastewater-based epidemiology and risk assessment. *Water Research*, 198, 117183. <https://doi.org/10.1016/j.watres.2021.117183>
- Wylie, T. N., Wylie, K. M., Herter, B. N., & Storch, G. A. (2015).** Enhanced virome sequencing using targeted sequence capture. *Genome Research*, 25, 1910-1920. <https://doi.org/10.1101/gr.191049.115>
- Xue, L., Cai, W., Zhang, L., Gao, J., Dong, R., Li, Y., Wu, H., Zhang, J., Zeng, H., Ye, Q., Ding, Y., & Wu, Q. (2019).** Prevalence and genetic diversity of human sapovirus associated with sporadic acute gastroenteritis in South China from 2013 to 2017. *Journal of Medical Virology*, 91, 1759-1764. <https://doi.org/10.1002/jmv.25511>
- Yang, M., Zhao, F., Tong, L., Wang, S., & Zhou, D. (2021).** Contamination, bioaccumulation mechanism, detection, and control of human norovirus in bivalve shellfish: A review. *Critical Reviews in Food Science and Nutrition*, 1-14. <https://doi.org/10.1080/10408398.2021.1937510>
- Yinda, C. K., Ghogomu, S. M., Conceição-Neto, N., Beller, L., Deboutte, W., Vanhulle, E., Maes, P., Van Ranst, M., & Matthijssens, J. (2018).** Cameroonian fruit bats harbor divergent viruses, including rotavirus H, bastroviruses, and picobirnaviruses using an alternative genetic code. *Virus Evolution*, 4(1), vey008. <https://doi.org/10.1093/ve/vey008>
- Yinda, C. K., Vanhulle, E., Conceição-Neto, N., Beller, L., Deboutte, W., Shi, C., Ghogomu, S. M., Maes, P., Van Ranst, M., & Matthijssens, J. (2019).** Gut virome analysis of Cameroonians reveals high diversity of enteric viruses, including potential interspecies transmitted viruses. *mSphere*, 4(1), e00585-18. <https://doi.org/10.1128/mSphere.00585-18>
- Yu, J. M., Li, L. L., Zhang, C. Y., Lu, S., Ao, Y. Y., Gao, H. C., Xie, Z. P., Xie, G. C., Sun, X. M., Pang, L. L., Xu, J. G., Lipkin, W. I., & Duan, Z. J. (2016).** A novel Hepatovirus identified in wild woodchuck *Marmota himalayana*. *Scientific Reports*, 6, 22361. <https://doi.org/10.1038/NDsrep22361>

**Yuen, L. K., Catton, M. G., Cox, B. J., Wright, P. J., & Marshall, J. A. (2001).** Heminested multiplex reverse transcription-PCR for detection and differentiation of Norwalk-like virus genogroups 1 and 2 in fecal samples. *Journal of clinical microbiology*, 39(7), 2690-2694. <https://doi.org/10.1128/JCM.39.7.2690-2694.2001>

**Zhang, H., Liu, D., Zhang, Z., Hewitt, J., Li, X., Hou, P., Wang, D., & Wu, Q. (2021).** Surveillance of human norovirus in oysters collected from production area in Shandong Province, China during 2017-2018. *Food Control*, 121, 107649. <https://doi.org/10.1016/j.foodcont.2020.107649>

## REFERENCES WEBOGRAPHIQUES:

[https://www.sfm-microbiologie.org/wp-content/uploads/2019/02/VIRUS\\_NOROVIRUS.pdf](https://www.sfm-microbiologie.org/wp-content/uploads/2019/02/VIRUS_NOROVIRUS.pdf)

ViralZone, SIB Swiss Institute of Bioinformatics (2008). Hepatovirus. <https://viralzone.expasy.org/94>. Consulté le 02 février 2022

ViralZone, SIB Swiss Institute of Bioinformatics (2008). Sapovirus. <https://viralzone.expasy.org/196>. Consulté le 02 février 2022

ViralZone, SIB Swiss Institute of Bioinformatics (2013). Rotavirus. <https://viralzone.expasy.org/107>. Consulté le 02 février 2022

ViralZone, SIB Swiss Institute of Bioinformatics. (2015). Mamastrovirus. <https://viralzone.expasy.org/281>. Consulté le 02 février 2022

ViralZone, SIB Swiss Institute of Bioinformatics. (2018). Enterovirus. <https://viralzone.expasy.org/97>. Consulté le 02 février 2022

ViralZone, SIB Swiss Institute of Bioinformatics. (2008). Kobuvirus. <https://viralzone.expasy.org/652>. Consulté le 02 février 2022

Random Primer Mix, BioLabs. <https://international.neb.com/-/media/catalog/datacards-or-manuals/s1330datasheet-lot0041208.pdf?rev=33bbe02bd0684b90983f259786047360>. Consulté le 02 février 2022

Illumina DNA Prep with Enrichment Reference Guide. <https://emea.support.illumina.com/downloads/illumina-dna-prep-with-enrichment-reference-guide-1000000048041.html>. Consulté le 02 février 2022

Support solide d'amplification (*flowcell*) pour le séquençage Illumina. <https://emea.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html>. Consulté le 02 février 2022

Alex Ralph, 2021. Séquenceur Oxford Nanopore. <https://www.thetimes.co.uk/article/top-biotech-s-2bn-float-boosts-city-zkdc58jf2>. Consulté le 02 février 2022

Chikhi, R. Large genome assembly. BiG talk, Lund University, 5 Décembre 2018. [http://rayan.chikhi.name/pdf/big18\\_large\\_genome\\_assembly.pdf](http://rayan.chikhi.name/pdf/big18_large_genome_assembly.pdf). Consulté le 02 février 2022

<https://www.ncbi.nlm.nih.gov/genbank/>

<https://www.ncbi.nlm.nih.gov/refseq/>

<https://www.spm.gov.cm/site/sites/default/files/mutation%20de%20la%20reserve%20de%20a%20faune.pdf>

## ANNEXES

### Annexe 1. Détermination du nombre de copies dans chaque échantillon

Soit l'équation de la droite :

- $Ct = \log(\text{copies}/\mu\text{L}) * a + b$
- $\text{copies}/\mu\text{L} = 10^{(Ct-b)/a}$

Rapporter aux 100  $\mu\text{L}$  d'extrait :

- $\text{copies/extrait} = (10^{(Ct-b)/a}) * 100$

Rapporter aux 500  $\mu\text{L}$  d'extrait par rapport au volume de surnageant :

- $\text{copies}/500\mu\text{L extrait} = (10^{(Ct-b)/a}) * 100 * (\text{volume total}/0,5)$

Rapporter aux 2 g analysés :

- $\text{copies}/\text{g} = (10^{(Ct-b)/a}) * 100 * (\text{volume de total}/0,5) * (1/2)$

### Annexe 2. Protocole de quantification des ADN à l'aide du Qubit

- Mélanger 1  $\mu\text{L}$  de réactif "*dsDNA BR Reagent*" à 199  $\mu\text{L}$  "*Qubit dsDNA BR Buffer*"
- Ajouter 10  $\mu\text{L}$  de standard Positif/Négatif à 190  $\mu\text{L}$  du mélange réactionnel préparé, et 1  $\mu\text{L}$  de chaque librairie à 199  $\mu\text{L}$  de mélange réactionnel et vortexer.
- Etalonner le lecteur Qubit par lecture des solutions standard. Puis mesurer les quantités d'ADN dans les différentes librairies.

### Annexe 3. Réparation des ADN fragmentés et fixation d'un résidu d'adényle à l'extrémité 3'

10  $\mu\text{L}$  d'*End Repair and A-tailing Master Mix* ont été ajoutés à 50  $\mu\text{L}$  d'échantillon d'ADN fragmenté. L'ensemble a été soigneusement mélangé et introduit dans un bain de glace, puis immédiatement incubé dans un thermocycleur avec couvercle chauffant suivant le programme thermique suivant :

- 20 °C, 30 min
- 65 °C, 30 min
- Maintien à 4 °C

End Repair and A-tailing Master Mix	Volume /librairie ( $\mu\text{L}$ )
KAPA End Repair & A-Tailing Buffer	7
KAPA End Repair & A-Tailing Enzyme Mix	3
<b>Total</b>	<b>10</b>

#### Annexe 4. Ligation des adaptateurs

Ajouter 5  $\mu\text{L}$  de l'adaptateur *SeqCap Library Adapter* dans chaque puit contenant 60  $\mu\text{L}$  d'ADN réparé obtenue en annexe 1. Ajouter ensuite 45  $\mu\text{L}$  *Ligation Master Mix*. Mélanger l'ensemble et incubé à 20 °C pendant 20 min.

	Volume/librairie ( $\mu\text{L}$ )
Ligation Master Mix	
PCR-grade water	5
KAPA Ligation Buffer	30
KAPA DNA Ligase	10
<b>Total</b>	<b>45</b>

#### Annexe 5. Sélection des tailles des librairies d'ADN

- a. Ajouter 32  $\mu\text{L}$  d'AMPure XP beads ramenés à température ambiante et remis en suspension à chaque puit contenant 50  $\mu\text{L}$  de suspension d'ADN liés à des adaptateurs, et bien mélanger.
- b. Incuber à température ambiante pendant 5 min pour permettre aux fragments de plus de 450 pb de se lier aux billes.
- c. Placer les puits sur un portoir magnétique pour capturer les billes, et laisser reposer jusqu'à clarification.
- d. Transférer avec précaution 80  $\mu\text{L}$  du surnageant contenant des fragments de inférieurs à ~ 450 pb dans un nouveau puits.
- e. Ajouter 10  $\mu\text{L}$  d'Agencourt AMPure XP beads ramenés à température ambiante, et remettre les billes en suspension en pipetant de haut en bas plusieurs fois.
- f. Incuber l'échantillon à température ambiante pendant 5 min pour permettre aux fragments de plus de 250 pb de se lier aux billes.
- g. Placer les puits sur un portoir magnétique pour capturer les billes. Incuber jusqu'à ce que le liquide soit clair. Puis retirez et jetez soigneusement la majeure partie du surnageant sans re-suspendre les billes en gardant les puits sur portoir.
- h. Ajoutez 200  $\mu\text{L}$  d'éthanol à 80 % fraîchement préparé, et incubé l'échantillon à température ambiante pendant  $\geq 30$  secondes.
- i. Retirez et jetez soigneusement l'éthanol en gardant les échantillons sur le portoir magnétique.
- j. Reprendre les étapes h et i.
- k. Laisser les billes sécher à température ambiante, suffisamment pour que tout l'éthanol s'évapore.

- l. Retirez les échantillons du collecteur de particules magnétiques et remettre les billes en suspension dans 23  $\mu\text{L}$  de tampon d'éluion (10 mM Tris-HCl, pH 8,0). Puis incubé à température ambiante pendant 2 min pour permettre à l'ADN d'éluier les billes.
- m. Placer les puits sur un portoir magnétique pour capturer les billes jusqu'à clarification du liquide.
- n. Transférer le surnageant clair ( $\sim 20 \mu\text{L}$ ) dans un nouveau tube/puits et procéder à l'amplification de la librairie.

#### **Annexe 6. Lavage des librairies d'ADN**

- a. Laissez les billes Agencourt AMPure XP à température ambiante pendant au moins 30 min avant utilisation.
- b. Homogénéiser les billes pendant 10 sec à l'aide d'un vortex.
- c. Ajouter 40  $\mu\text{L}$  de billes AMPure XP à 50  $\mu\text{L}$  la librairies amplifiées et vortexer pendant 10 sec. Puis, incubé à température ambiante pendant 5 min afin de permettre à l'ADN de se lier aux billes.
- d. Placer les puits contenant les librairies d'ADN liées aux billes dans un collecteur de particules magnétiques, et laissez la solution s'éclaircir.
- e. Retirer et jeter le surnageant en prenant soin de ne pas re-suspendre les billes.
- f. Ajouter 200  $\mu\text{L}$  d'éthanol à 80 % fraîchement préparé dans les puits contenant les librairies d'ADN liées aux billes, incubé à température ambiante pendant 30 sec, puis retirer et jeter l'éthanol.
- g. En gardant les puits sur le collecteur de particules magnétiques, ajoutez 200  $\mu\text{L}$  d'éthanol à 80 % fraîchement préparé, incubé à température ambiante pendant 30 sec, puis retirer et jeter l'éthanol.
- h. Laisser les billes sécher à température ambiante pendant 5 min, puis retirez les puits du collecteur de particules magnétiques.
- i. Remettre les librairies d'ADN en suspension en utilisant 53  $\mu\text{L}$  de tampon Tris-HCl 10 mM, pH 8,0 et incubé à température ambiante pendant 2 min.
- j. Remettre les puits dans le collecteur de particules magnétiques et laissez la solution s'éclaircir. Puis retirez 50  $\mu\text{L}$  de surnageant qui contient les librairies amplifiées et transférez-le dans un nouveau tube.
- k. Les librairie amplifiées et purifiées peuvent être stockées entre 2 et 8  $^{\circ}\text{C}$  pendant 1 à 2 semaines ou entre -15 et -25  $^{\circ}\text{C}$ .

**Annexe 7. Mélange d'hybridation**

<b>Hybridization Master-Mix</b>	<b>One reaction (µL)</b>
X Hybridization Buffer	7,5
Hybridization Component A	3
<b>Total</b>	<b>10,5</b>



# Contamination of Clams with Human Norovirus and a Novel Hepatitis A Virus in Cameroon

Patrice Bonny<sup>1,2,3</sup> · Marion Desdoutis<sup>1</sup> · Julien Schaeffer<sup>1</sup> · Pascal Garry<sup>1</sup> · Jean Justin Essia Ngang<sup>2</sup> · Françoise S. Le Guyader<sup>1,4</sup>

Received: 4 May 2020 / Accepted: 16 June 2020  
© The Author(s) 2020

## Abstract

Shellfish constitute an important protein source but may be contaminated by viruses from various origins. A study performed on clams collected in Cameroon showed a high prevalence of norovirus and hepatitis A virus. After sequencing, the hepatitis A virus showed similarities with the genotype V simian strains.

**Keywords** Shellfish · Hepatitis A virus · Molecular epidemiology

The need for a one-health approach to protect humans from emerging diseases is evident by the high number of microbial pathogens that become zoonotic following insect bites or consumption of contaminated meat. This is particularly well documented for infectious disease transmission from non-human primates to human (Devaux et al. 2019). Environmental issues are critical in such events, especially for RNA viruses that may be excreted at high concentrations by infected hosts and are very resistant outside their hosts (De Graaf et al. 2017). Transmission can occur directly by ingestion of contaminated waters or via contaminated food. For example, shellfish growing in coastal areas or rivers are known to concentrate microorganisms by their ability to filter large volume of waters and thus may favor the transmission of zoonotic strains to humans when consumed.

The demand for shellfish is increasing worldwide given their nutritive value and taste. In addition, shellfish grow without human intervention, are easily collected and constitute an important source of nutrients for many populations (Venugopal and Gopakumar 2017). In West-Africa, the Sanaga clams *Egeria radiata*, a freshwater clam, grow in several rivers, and constitute a substantial source of protein for the local population as well as an important economic income through the shell industry, such as the ceramic or animal feed processing industries (Ajonina et al. 2005). Up to now, no study has been conducted to evaluate the microbial contamination of these clams. Despite being mainly consumed following cooking, which may limit the risk of bacteria transmission, the evaluation of the presence of human enteric viruses implicated in cooked food-borne outbreaks was considered necessary (Pinto et al. 2009; Lunestad et al. 2016).

Fourteen samples of clams (*Egeria radiata*) were collected on the Sanaga river (Cameroon, west Africa, 12 km upstream of the Atlantic Ocean), between February 2018 and March 2019 from two sampling points (eight from site B and six from site M). Clams were immediately frozen and shipped to the Ifremer laboratory where they were dissected and the digestive tissues (DT) recovered. Viruses were eluted from clam DT and nucleic acids were extracted using the reference method (ISO 15216-1 2017). After validation of quality controls (extraction efficiency and inhibitor removal), reverse transcription real-time quantitative PCR was used to detect human norovirus and hepatitis A virus. Human norovirus were detected in ten samples (five from each sampling

---

GeneBank accession number: The hepatitis A sequence was deposited in Genbank and is available under the temporary accession number GenBank MT185676.

---

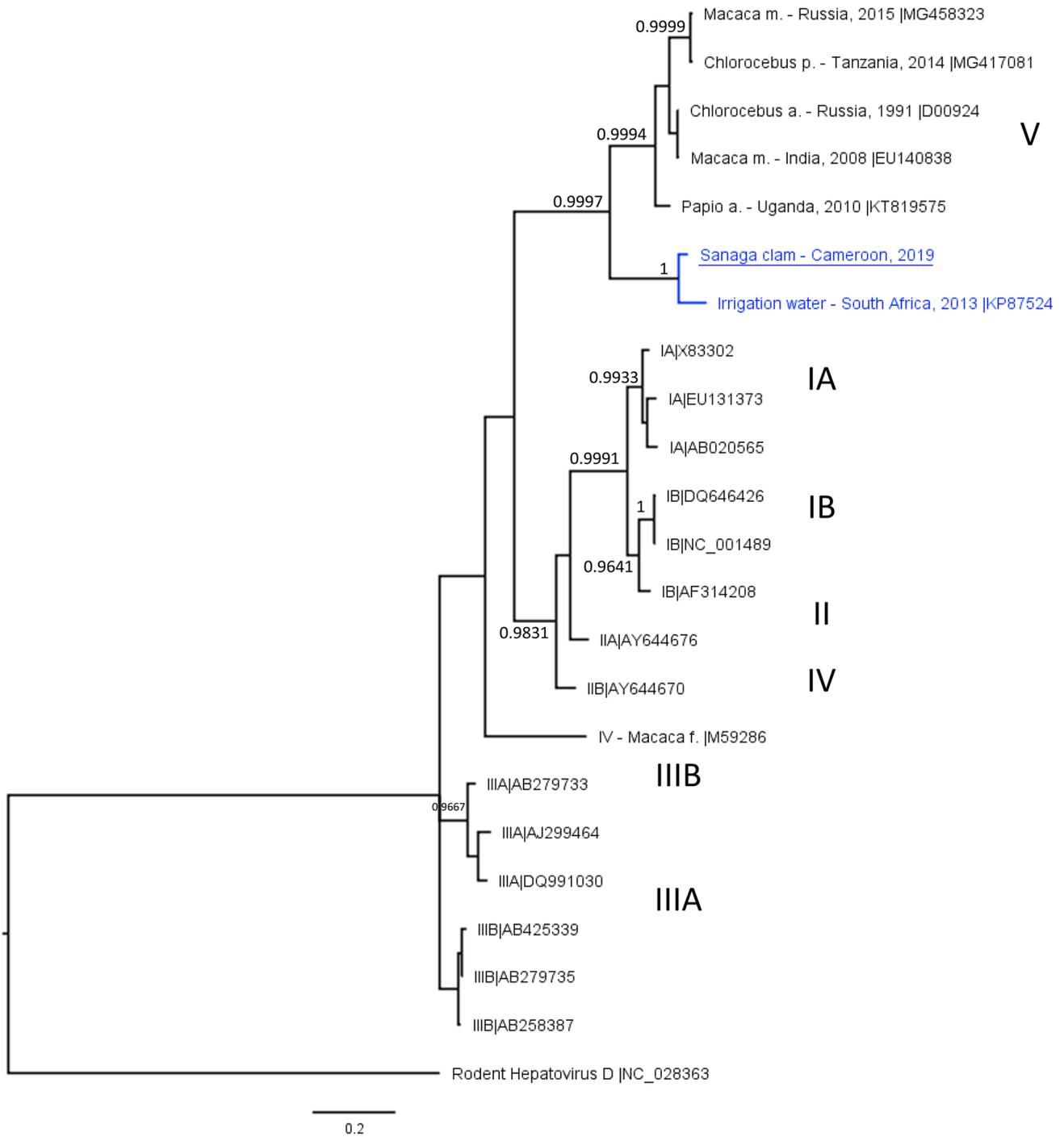
✉ Françoise S. Le Guyader  
soizick.le.guyader@ifremer.fr

<sup>1</sup> Laboratoire de Microbiologie, LSEM/SG2M, IFREMER, 44300 Nantes, France

<sup>2</sup> Département de Microbiologie, Université de Yaoundé I, B.P. 812, Yaoundé, Cameroun

<sup>3</sup> Centre de Recherche en Alimentation et Nutrition, IMPM, B.P. 6163, Yaoundé, Cameroun

<sup>4</sup> Laboratoire de Microbiologie, LSEM/SG2M, IFREMER, BP 21105, 44311 Nantes Cedex 03, France



**Fig. 1** Bayesian phylogeny of HAV sequence identified in the Sanaga clam samples. The tree is based on 441 nucleotides in the VP1-2A junction region (position 2918 to 3359 relative to the HAV reference sequence NC\_001489), except for MG417081 (315 nt), MG458323 (346 nt) and M59286 (334 nt) that end before. Sequences were aligned using MUSCLE 3.8.425 implemented in Geneious (11.1.4). The tree was constructed using Mr Bayes 2.3.6 plugin in Geneious, with the GTR evolutionary model, 2 millions generations sampled every 200 steps, and 20% burn-in. Posterior probabilities higher than

0.95 are indicated just above or below the corresponding branches. The rodent hepatovirus D reference sequence was used as an outgroup. HAV sequences cluster by genotype (Gt) as highlighted on the right, with virus from humans belonging to Gt IA-B, IIA-B, and IIIA-B. Simian HAV (host species is indicated in the taxon name) belong to genotype IV or V. HAV from Sanaga clams (underlined) (GenBank MT185676) clusters together with a sequence identified in an irrigation water sample from South Africa, in a clade (blue) separated from simian sequences in the genotype V

site) at low concentrations (ranging from below the limit of quantification up to 2500 RNAc/g of digestive tissues for one sample collected in September). Positive samples were amplified using the standard RT-PCR method with the same RT and Platinum *Taq* polymerase enzymes, and primers targeting the polymerase and the capsid regions (Le Guyader et al. 2008). Amplicons from positive samples were purified, cloned, sequenced, and genotyped using the Norovirus Typing Tool 2.0 (Kroneman et al. 2011). This approach allowed the identification of norovirus sequences such as GI.2, GII.3, GII.6, GII.4\_Sydney 2012 using primers targeting the capsid region or GII.P31, and GII.P21 when using primers targeting the polymerase region. HAV was detected in nine of the samples (four from site B and five from M). The VP1/P2A junction was amplified by PCR, cloned and sequenced, yielding a 441 bp sequence that was identified as HAV and assigned to the genotype V by the HAV Typing Tool (Kroneman et al. 2018). Using NCBI Blastn, the Sanaga strain (GenBank MT185676) displayed a highest similarity (91.88%) with a strain identified in irrigation water samples from South Africa (KZN\_Irr-20130530) (Rachida et al. 2016). Bayesian phylogenetic analysis with HAV sequences from different genotypes retrieved from HAVNet and GenBank shows the clustering of these two strains retrieved from environmental samples in a clade supported by high posterior probability and most closely related to simian strains of genotype V (Fig. 1).

The Sanaga strain and KZN\_Irr-20130530 exhibited between 79.71 and 82.71% similarities with the simian strains of genotype V, and below 79% with the other genotypes. This suggests, as discussed by Rachida et al., that this clade is too divergent to be considered a subgenotype of genotype V (10). The sequence of the VP3-VP1 cleavage site of KZN\_Irr-20130530 suggested a simian origin for this strain (Rachida et al. 2016). The proximity of the Sanaga strain with simian HAV strains and the documented presence of non-human primates in the sampling area suggest a simian origin, but does not exclude another host. Further investigations need to be done to identify human sewage discharge in this area or to compare with clinical cases.

Finding this virus in a food raises several concerns. If the strain originates from simian hosts, the risk of host jump to humans is enhanced by the high frequency of positive samples, the close relation between human and non-human primates, and the high stability of HAV. Indeed, some HAV strains appear highly resistant to heat and cooked coquina clams have been implicated in outbreaks (Pinto et al. 2009). Here this risk appears limited as Sanaga clams are usually consumed boiled or smoked. Nonetheless, the local population needs to be informed of this potential risk of virus transmission (Ajonina et al. 2005). The genomic plasticity of viruses, allowing selection of adaptative mutations may further facilitate their human-to-human transmission (De Graaf

et al. 2017). Alternatively, the presence of a new HAV strain in clams also contaminated with human norovirus raises the possibility of a human origin. Further investigations need to be implemented to investigate the origin and prevalence of this new HAV strain in the Sanaga region, and to evaluate the presence of other microbial contaminants in clams and other shellfish.

**Acknowledgements** This study was supported by the EU-H2020 grant No 643476 Compare. P.B. was supported by a founding from Ministry of Scientific Research and Innovation of Cameroon.

## Compliance with Ethical Standards

**Conflict of interest** None declared by the authors.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ajonina, P. U., Ajonima, G., Jin, E., Mekongo, F., Ayisse, I., & Usongo, L. (2005). Gender roles and economics of exploitation, processing and marketing of bivalves and impacts on forest resources in the Sanaga Delta region of Douala-Edea wildlife reserve, Cameroon. *International Journal of Sustainable Development & World Ecology*, 12(2), 161–172. <https://doi.org/10.1080/13504500509469627>.
- De Graaf, M., Beck, R., Caccio, S. M., Duim, B., Fraaij, P. L. A., Le Guyader, F. S., et al. (2017). Sustained fecal-oral human-to-human transmission following a zoonotic event. *Current Opinion in Virology*, 22, 1–6. <https://doi.org/10.1016/j.coviro.2016.11.001>.
- Devaux, C. A., Mediannikov, O., Medkour, H., & Raoult, D. (2019). Infectious disease risk across the growing Human-non human primate interface: A review of the evidence. *Frontiers in Public Health*, 7, 305. <https://doi.org/10.3389/fpubh.2019.00305>.
- ISO 15216-1. (2017). Microbiology of food and animal feed -horizontal method for determination of hepatitis A virus and norovirus in food using real-time RT-PCR, Part 1: method for quantification.
- Kroneman, A., de Souza, R., Verhoef, L., Koopmans, M. P. G., Vennema, H., & On Behalf of the HAVNetwork. (2018). Usability of the international HAVNet hepatitis A virus database for geographical annotation, backtracing and outbreak detection. *Euro Surveillance*, 23(37), 1700802. <https://doi.org/10.2807/1560-7917.ES.2018.23.37.1700802>.
- Kroneman, A., Vennema, H., Deforche, K., Avoort, H. V. D., Penaranda, S., Oberste, M. S., et al. (2011). An automated genotyping tool for enteroviruses and noroviruses. *Journal of Clinical Virology*, 51(2), 121–125. <https://doi.org/10.1016/j.jcv.2011.03.006>.

- Le Guyader, F. S., Le Saux, J.-C., Ambert-Balay, K., Krol, J., Serais, O., Parnaudeau, S., et al. (2008). Aichi virus, norovirus, astrovirus, enterovirus and rotavirus involved in clinical cases from a French oyster-related gastroenteritis outbreak. *Journal of Clinical Microbiology*, *46*, 4011–4017.
- Lunestad, B. T., Maage, A., Roiha, I. S., Myrmel, M., Svanevik, C. S., & Duinker, A. (2016). An outbreak of norovirus infection from shellfish soup due to unforeseen insufficient heating during preparation. *Food and Environmental Virology*, *8*(4), 231–234. <https://doi.org/10.1007/s12560-016-9245-5>.
- Pinto, R. M., Costafreda, M. I., & Bosch, A. (2009). Risk assessment in shellfish-borne outbreaks of hepatitis A. *Applied and Environmental Microbiology*, *75*(23), 7350–7355. <https://doi.org/10.1128/AEM.01177-09>.
- Rachida, S., Matsapola, P. N., Wolfaardt, M., & Taylor, M. B. (2016). Genetic characterization of a novel hepatitis A virus strain in irrigation water in South Africa. *Journal of Medical Virology*, *88*(4), 734–737. <https://doi.org/10.1002/jmv.24370>.
- Venugopal, V., & Gopakumar, K. (2017). Shellfish: Nutritive value, health benefits, and consumer safety. *Comprehensive Reviews in Food Science and Food Safety*, *16*(4), 1219–1242. <https://doi.org/10.1111/1541-4337.12312>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



# Human and Animal RNA Virus Diversity Detected by Metagenomics in Cameroonian Clams

Patrice Bonny<sup>1,2,3</sup>, Julien Schaeffer<sup>1</sup>, Alban Besnard<sup>1</sup>, Marion Desdouits<sup>1</sup>, Jean Justin Essia Ngang<sup>2</sup> and Françoise S. Le Guyader<sup>1\*</sup>

<sup>1</sup>Laboratoire de Microbiologie, LSEM/SG2M, Ifremer, Nantes, France, <sup>2</sup>Département de Microbiologie, Université de Yaoundé I, Yaoundé, Cameroon, <sup>3</sup>Centre de Recherche en Alimentation et Nutrition, IMPM, Yaoundé, Cameroon

## OPEN ACCESS

### Edited by:

Erna Geessien Kroon,  
Federal University of Minas Gerais,  
Brazil

### Reviewed by:

Jemma Geoghegan,  
University of Otago, New Zealand  
Jesus L. Romalde,  
University of Santiago  
de Compostela, Spain

### \*Correspondence:

Françoise S. Le Guyader  
soizick.le.guyader@ifremer.fr

### Specialty section:

This article was submitted to  
Virology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 03 September 2021

**Accepted:** 18 October 2021

**Published:** 30 November 2021

### Citation:

Bonny P, Schaeffer J, Besnard A, Desdouits M, Essia Ngang JJ and Le Guyader FS (2021) Human and Animal RNA Virus Diversity Detected by Metagenomics in Cameroonian Clams.  
*Front. Microbiol.* 12:770385.  
doi: 10.3389/fmicb.2021.770385

Many recent pandemics have been recognized as zoonotic viral diseases. While their origins remain frequently unknown, environmental contamination may play an important role in emergence. Thus, being able to describe the viral diversity in environmental samples contributes to understand the key issues in zoonotic transmission. This work describes the use of a metagenomic approach to assess the diversity of eukaryotic RNA viruses in river clams and identify sequences from human or potentially zoonotic viruses. Clam samples collected over 2 years were first screened for the presence of norovirus to verify human contamination. Selected samples were analyzed using metagenomics, including a capture of sequences from viral families infecting vertebrates (VirCapSeq-VERT) before Illumina NovaSeq sequencing. The bioinformatics analysis included pooling of data from triplicates, quality filtering, elimination of bacterial and host sequences, and a deduplication step before *de novo* assembly. After taxonomic assignment, the viral fraction represented 0.8–15% of reads with most sequences (68–87%) remaining un-assigned. Yet, several mammalian RNA viruses were identified. Contigs identified as belonging to the *Astroviridae* were the most abundant, with some nearly complete genomes of *Bastrovirus* identified. *Picobirnaviridae* sequences were related to strains infecting bats, and few others to strains infecting humans or other hosts. *Hepeviridae* sequences were mostly related to strains detected in sponge samples but also strains from swine samples. For *Caliciviridae* and *Picornaviridae*, most of identified sequences were related to strains infecting bats, with few sequences close to human norovirus, picornavirus, and genogroup V hepatitis A virus. Despite a need to improve the sensitivity of our method, this study describes a large diversity of RNA virus sequences from clam samples. To describe all viral contaminants in this type of food, and being able to identify the host infected by viral sequences detected, may help to understand some zoonotic transmission events and alert health authorities of possible emergence.

**Keywords:** shellfish, metagenomics, human enteric viruses, mammal viruses, zoonosis

## INTRODUCTION

In the last few years, Next Generation Sequencing (NGS) has been increasingly used to study microbial populations in environmental samples. This approach allows the description of all nucleic acid sequences present in a sample and thus, theoretically, enables the identification of all organisms including viruses (Dance, 2021). As a catch-all agnostic approach, it holds great promise for monitoring and determining viral diversity in food and environmental samples such as sewage and shellfish (Desdoutis et al., 2020; Nieuwenhuijse et al., 2020). Metagenomic sequencing has been applied to estimate viral diversity in different ocean waters, and sewage samples across the globe, showing its interest to describe viral communities (Nieuwenhuijse et al., 2020; Sunagawa et al., 2020).

Many recent human pandemics have been recognized as zoonotic viral diseases. If, in some cases, the origin has been linked to the modification of wildlife habitats or food consumption, many causes or sources are still unknown (Devaux et al., 2019; Osterhauss et al., 2020). Among main hypotheses, environment plays an important role as it can be the link between humans and the original host. Being able to describe the viral diversity in environmental samples may constitute the first step to understand the key issues in zoonotic transmission and help its prevention. However, the strategy of environmental sampling represents a major challenge. For human contaminations, sewage sampling is clearly a good approach, but when the animal or wildlife excreta has to be considered, other approaches are necessary (Nieuwenhuijse et al., 2020; Osterhauss et al., 2020).

The demand for shellfish is increasing worldwide given their nutritive value or taste. In addition, they grow without human intervention and constitute a substantial source of nutrients for many populations (Venugopal and Gopakumar, 2018). However, they have been identified for many years as high-risk food (Metcalf et al., 1995). Indeed, because of their mode of feeding by filtration of large volumes of water, they can concentrate various types of contaminants from their environment, including foodborne viruses. In line with this, many enteric viruses, such as noroviruses, enteroviruses, or hepatitis viruses, can be detected in shellfish tissues following their exposure to human sewage (La Rosa et al., 2017; Strubbia et al., 2019). In some cases, during shellfish-borne outbreaks, the high viral diversity in shellfish makes it difficult to identify the exact etiological agent, leading to an erroneous identification if the detection approach is too selective (Nieuwenhuijse and Koopmans, 2017), thereby explaining the need for a holistic approach when detecting viruses in shellfish. This also suggests that we can use shellfish as environmental sentinel, since they filter and concentrate the different pathogens present in the environment (Polo et al., 2021).

In West Africa, *Galatea paradoxa* is a freshwater clam found in several rivers including the Sanaga river and constitutes a substantial source of protein for local populations as well as a significant economic income (Ajonina et al., 2005). However, like all shellfish around the world, they are prone to accumulate various contaminants, including human enteric viruses. Our

previous work, on the same site, reported Sanaga clams contamination with norovirus and hepatitis A virus, suggesting contamination by human sewage (Bonny et al., 2020). Furthermore, a sequence of hepatitis A virus close to the genotype V known to infect non-human primates was also detected. These findings demonstrate the contamination of these clams by both human and animal viruses, suggesting that they can reflect the diversity of viruses circulating in the Sanaga River growing site.

To deeply investigate the viral contamination, we performed additional sampling and applied a metagenomic approach to assess the diversity of RNA viral sequences of strains that may infect vertebrates in these clams and identify potentially pathogenic viruses, both humans and potentially zoonotic.

## MATERIALS AND METHODS

### Sample Collection

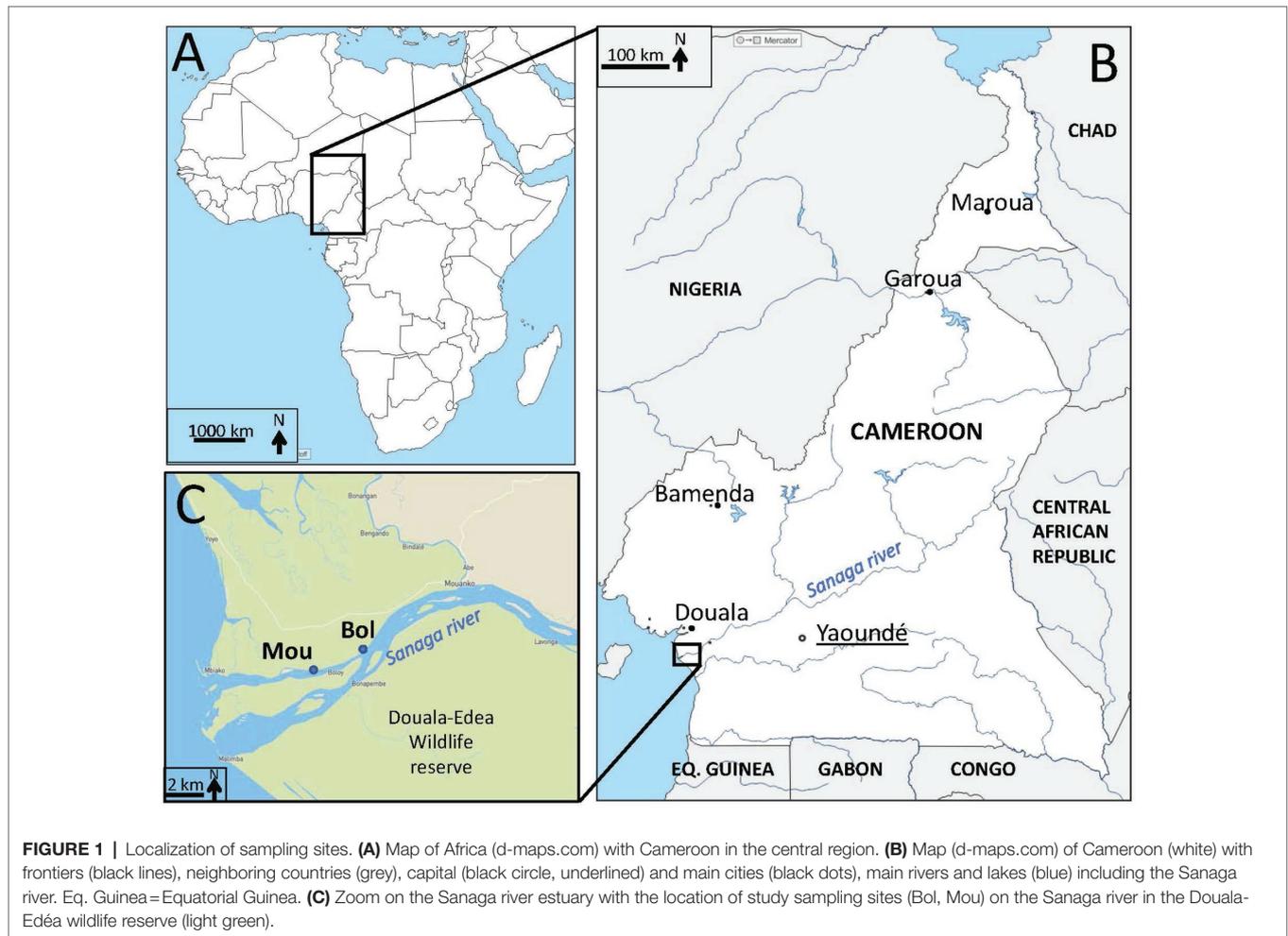
Two sampling points (Bol and Mou) were selected on the Sanaga river (Cameroon, Central Africa), 12 km upstream the Atlantic Ocean (Figure 1) and located in a subdivision of about 15,000 inhabitants (non-official estimation performed in 2016/personal communication). These two points were distant of about 2.5 km and located on the same side of the river. Naturally growing clams *Galatea paradoxa* were collected monthly in each point between February 2018 and February 2020 (except on one occasion for site Bol and three occasions for site Mou). Each sample was constituted of 10–15 alive individuals. After rapid cleaning under tap water, clams were frozen ( $-20^{\circ}\text{C}$ ) for further shipping to the Ifremer laboratory in France.

### Sample Preparation, Norovirus, and Hepatitis A Virus Detection

On arrival, samples were thawed, shucked, and dissected. Digestive tissues (DT) were recovered, pooled, chopped, distributed in 2 g aliquots, and immediately stored at  $-20^{\circ}\text{C}$  until analysis. Viruses were eluted from DT using proteinase K method (ISO, 15216-1, 2017), and nucleic acids were extracted using the NucliSENS kit (bioMérieux, France). After validation of quality controls (extraction efficiency and inhibitor removal), the quantitative detection of norovirus and hepatitis A virus was performed by RT-qPCR (ISO, 15216-1, 2017). All samples were analyzed in the triplicates of undiluted nucleic acids using a one-step real-time RT-PCR kit (Invitrogen) and a MX3000P Quantitative PCR system (Agilent Technologies France).

### Sample Preparation for Metagenomic Analysis

Samples positive for norovirus and hepatitis A virus were submitted to a novel extraction using an optimized protocol set up for metagenomics analysis (Strubbia et al., 2020). Briefly DT aliquots (2 g) were incubated with proteinase K solution (30 U/mg, Sigma-Aldrich, St-Quentin France) for 15 min at  $37^{\circ}\text{C}$  under agitation (32 rotations/min), and 15 min at  $60^{\circ}\text{C}$ . The mixture was sonicated (Bandelin UD 2200 with cup-horn



adaptor) for 3 min at maximum power (100%; cycles of 1 min with 1 min on ice) and centrifuged for 5 min at  $3,000\times g$ . Supernatant was measured, transferred to a 15-ml tube, mixed with sodium pyrophosphate (10 mM final concentration), and incubated for 40 min at  $4^{\circ}\text{C}$  under gentle agitation. After centrifugation for 20 min at  $8000\times g$ , supernatant (approximately 3 ml) was recovered, mixed with 1.5 ml of polyethylene glycol 6000 (PEG 24% wt/vol, Sigma-Aldrich)-sodium chloride (1.2 M), and incubated for 1 h at  $4^{\circ}\text{C}$ . After centrifugation for 20 min at  $10,000\times g$ , the pellet was resuspended in 2 ml of preheated ( $56^{\circ}\text{C}$ ) glycine buffer (0.05 M; pH 9), then filtered using 5, 1.5, and  $0.45\mu\text{m}$  acetate cellulose filters. Filtrate was treated with  $20\mu\text{l}$  of OmniCleave™ Endonuclease (Lucigen corporation;  $200\text{U}/\mu\text{l}$ ) and  $200\mu\text{l}$  of  $\text{MgCl}_2$  (100 mM) for 1 h at  $37^{\circ}\text{C}$ .

The extraction of nucleic acids was carried out using lysis buffer and the NucliSENS kit (bioMerieux) with the semi-automatized eGENE-UP™ system (bioMerieux). Nucleic acids were treated by TURBO DNase (25 U) for 30 min at  $37^{\circ}\text{C}$  (Ambion, Thermo Fischer Scientific, France). An additional RNA purification was carried out using the RNA Clean - Concentrator™-5 kit (Zymo Research, Irvine, United States). Nucleic acids were recovered in  $100\mu\text{l}$  of elution buffer and stored at  $-80^{\circ}\text{C}$ .

The detection of norovirus and hepatitis A virus was verified as described above, before libraries preparation.

## Library Preparation and Metagenomic Sequencing

Technical triplicates were performed for cDNA reverse-transcription of RNA extracts using the enzyme Superscript II (ThermoFisher, France) and random hexamer primers (ThermoFisher, France). After the production of cDNA using the second strand reaction buffer and synthesis enzyme mix (NEBNext Ultra RNA Library prep, New England Biolabs, France), a physical fragmentation (Ultrasonicator M220, Covaris) was carried out for 110 s (Strubbia et al., 2020).

Libraries were prepared using the KAPA Prep Kit (Roche, France). After ligation of the adapters, several libraries cleanup steps using AmPure XP Beads (Beckman Coulter, United States), and 80% ethanol were carried out to select fragments between 150 and 500 bp. The libraries were then quantified and pooled equimolarly into seven pools, after which viral enrichment was performed using the VirCapSeq-VERT capture panel (Roche, France) (Wylie et al., 2015).

The sequencing was carried out on the Illumina NovaSeq 6000 using NovaSeq reagent Kit to generate 2×250 base pair reads at the ICM platform (Institut Cerveau Moelle, Paris, France).

## Bioinformatic Pipeline

Bioinformatic analysis was performed using nextflow to monitor the run on the computing cluster. Fastq files from technical triplicates were pooled at the beginning of the analysis. TrimGalore was used for quality filtering and adapter removal, with a quality threshold of 25. Then, a second filtering of the non-target reads, in this case non-viral reads such as host or bacterial rRNA, was carried out by mapping clean reads using Bowtie2 (very-sensitive option) (Langmead and Salzberg, 2012) on the Silva database (Nooij et al., 2018). Finally, duplicate reads were binned using cd-hit-dup with a prefix length of 50 for each read (Fu et al., 2012), in order to simplify the data set, facilitate the assembly step, and reduce PCR bias during abundance calculation.

Taxonomic assignation of the reads was done using Kaiju with the nr\_euk database (February, 2020), on the raw data and on the data post deduplication to assess the composition of our samples and impact of preprocess steps (filtering, rRNA removal, and deduplication).

A *de novo* assembly was performed using metaSPAdes with kmer lengths 21, 33, 55, 77, and 99 (Nurk et al., 2017). Contigs longer than 300bp were identified using BLASTn 2.6.0 (e value 1E-5) on the nt NCBI bank (Camacho et al., 2009). In the absence of a match, the search was repeated using DIAMOND (e value 1E-3) on the nr NCBI bank (Buchfink et al., 2015). To evaluate the coverage of each contig, post-process reads were mapped using Bowtie2 on the metaSPAdes contigs. Multi-mapped reads were removed as a source of potential overestimation of the abundance. Taxonomic identification was done using Entrez direct tool, the taxid allowing to extract information at a defined taxonomic level. Families were determined on the krona file taking all families that account for more than 0.01% of the viral reads in any of our samples.

Reads per millions (rpm) were calculated as the number of reads per family divided by the total number of trimmed and deduplicated reads, using R, and heatmaps were generated using the ggplot R package. For the phylogenetic analysis, amino acid alignments were performed using Muscle and Clustal Omega. Phylogenetic trees were constructed using IQtree with a bootstrap of 1,000.

## RESULTS

### Norovirus and Hepatitis A Virus Detection in Samples

All 45 collected samples yielded extraction efficiencies above 1% and none presented PCR inhibitors (ISO, 15216-1, 2017). A total of 31 samples were positive for norovirus GI and/or GII (Table 1). The same number of samples was collected from each site, and no difference in terms of prevalence or genomic concentrations was detected for norovirus contamination

**TABLE 1** | Detection and quantification of norovirus and hepatitis A in collected samples.

Site	Nb of sample	Ext. eff.* (%)	Norovirus		Hepatitis A virus	
			Nb pos	GMC <sup>#</sup>	Nb pos	GMC <sup>#</sup>
Bol	23	13 (3–67)	15	563	6	69
Mou	22	9 (4–40)	16	297	10	75
Total	45	11 (3–67)	31	405	16	73

\*Ext eff: mean extraction efficiency calculated for samples collected on each site and total samples, expressed as %, with the range of extraction efficiency obtained into brackets ().

<sup>#</sup>GMC: geometric mean concentration calculated using concentrations obtained for each positive sample and expressed as cRNA/g of clam digestive tissue.

between sites (Table 1). Regarding hepatitis A virus contamination, only six samples were detected positive on site Bol compared to 10 samples on site Mou, with low concentrations very close to the limit of quantification.

To gain insight into the diversity of these viral pathogens, we selected 22 samples (13 from site Bol and nine from site Mou) collected during different months for metagenomic analyzes.

### Metagenomic Data

The Illumina sequencing generated approximately 1.9 billion raw reads ranging from 42 to 177 million per library. Of these, 20% were removed by quality filtering and rRNA removal (Table 2).

To reduce the data size and optimize computing time for *de novo* assembly, the dataset was simplified by removing duplicate reads (deduplication), resulting in up to 85% of eliminated reads for some libraries. To verify the possible impact on reads distribution between the different life kingdoms, we performed a fast taxonomic assignation using Kaiju before and after the deduplication step (Figure 2). The kingdom that was most affected by deduplication was Bacteria with 88–96% of removed reads and Eukaryota with a fraction of removed reads between 77 and 95%. The number of reads for the viral fraction was reduced by 79–83% following deduplication with few variations between samples. Eventually, 13.8% of initial reads (264 millions) were retained after the deduplication step. Of these, 68–87% were unclassified, 7.1–21.4% were assigned as bacteria, 0.7–10.2% as Eukaryote, 1.1–7.5% as virus, and 0.1–0.7% as Archaea (Figure 2B). All the following analyses were performed on deduplicated reads.

### Metagenome Assembly

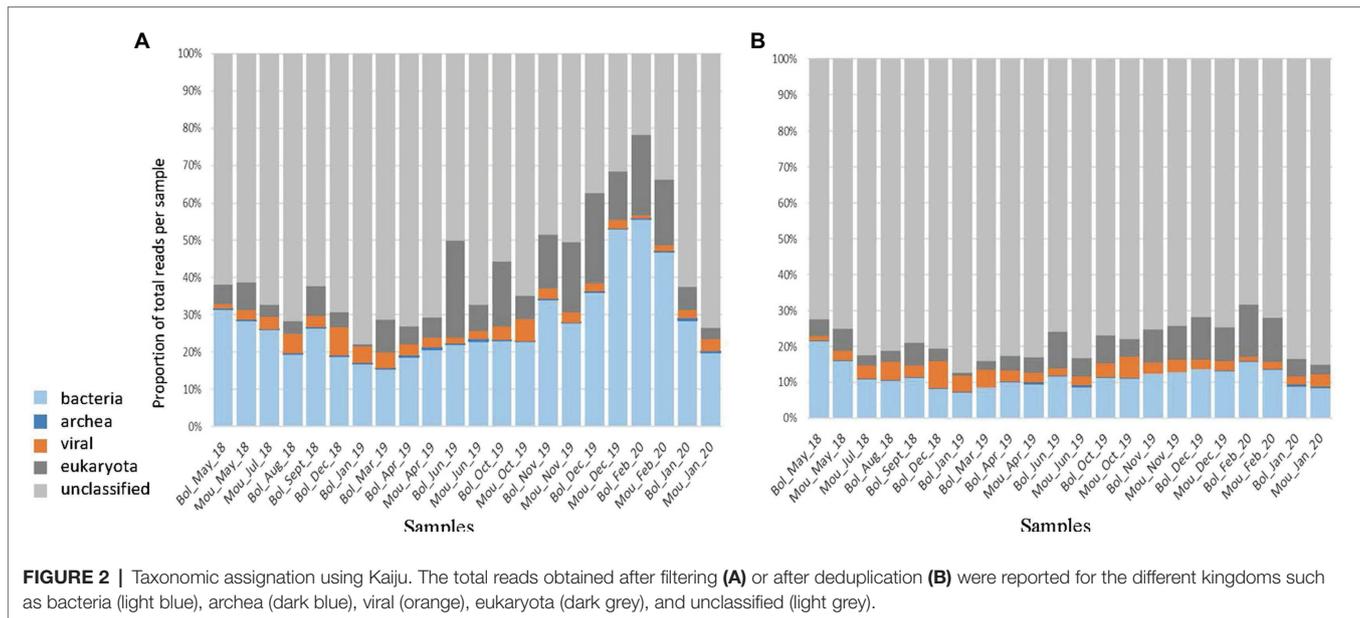
MetaSPAdes assembled 77% of deduplicated reads into a total of 268,378 contigs, of which 204,107 were longer than 300 base pairs. Of these, 98,476 (48%) contigs were annotated by either BLASTn or DIAMOND (Table 2).

A total of 198 millions of reads were assigned using BLAST, 95 millions of which being identified as viral sequences (Table 2). For each sample (three pooled libraries), the percentage of reads identified as viral sequences varied from

**TABLE 2** | Summary of read numbers and contigs at different steps of the pipeline per sample.

Site	Date	Total reads (millions)				Number of contigs			Virus	
		Raw	Filtered	Deduplicated	Mapped	Assembled	>300bp	BLAST	Reads (millions)	%*
Bol	May 18	48	44	8	3	7,890	6,917	4,340	0.9	12
Bol	August 18	60	56	9	6	10,200	8,576	4,148	4	45
Bol	September18	155	131	21	14	16,145	14,635	5,416	7	33
Bol	December 18	88	80	15	11	20,339	15,185	6,608	8	52
Bol	January 19	111	109	20	16	10,044	8,840	3,087	15	77
Bol	March 19	71	66	12	10	9,280	8,233	3,173	6	54
Bol	April 19	69	65	12	9	18,067	13,937	6,332	4	30
Bol	June 19	86	64	11	9	10,266	7,548	2,968	2	19
Bol	October 19	47	36	7	5	12,836	10,215	4,330	2	27
Bol	November 19	44	32	6	4	12,554	10,751	4,368	2	26
Bol	December 19	42	27	5	4	9,287	8,173	3,540	0.8	18
Bol	January 20	129	102	18	15	18,929	14,789	6,062	8	42
Bol	February 20	177	95	14	12	11,410	8,487	3,822	2	12
Mou	May 18	63	55	9	6	8,919	8,068	3,982	3	28
Mou	July 18	131	119	19	14	14,143	12,954	4,972	9	50
Mou	April19	70	62	12	9	15,763	12,640	5,370	4	38
Mou	June 19	67	57	11	8	12,655	10,061	3,737	4	40
Mou	October 19	78	68	12	9	18,287	14,316	5,776	4	35
Mou	November 19	51	37	7	5	12,444	10,060	4,407	1	21
Mou	December 19	152	89	14	11	14,108	11,406	4,847	3	24
Mou	January 20	87	82	15	11	15,704	11,940	4,046	5	31
Mou	February 20	90	54	9	7	9,475	7,552	3,145	2	19

\*The percentage was calculated by dividing the viral reads by the deduplicated reads (and x100).



12 to 54%, with only one sample reaching a higher percentage (77%; **Table 2**). These viral reads were further identified and classified into different viral families based on BLASTn/DIAMOND results (**Figure 3**). All samples displayed similar distribution of the reads regarding the different families selected here. As expected from our protocol for library preparation (DNase treatment and enrichment in viruses from vertebrates), few reads mapped to phage families (0.03%) or DNA viruses (0.6%), except for *Circoviridae* family that was frequently identified. These reads were not further analyzed as a bias may exist, and the purpose of this study was to identify RNA viruses. The most abundant contigs belonged to unclassified riboviruses (96.6% of viral reads). Viral families infecting vertebrates or invertebrates hosts (*Nodaviridae*) or plants (*Luteoviridae* – now *Solemoviridae*, *Tombusviridae*) were also frequently detected (**Figure 3**). Contigs related to a rhabdovirus previously detected in a sample from an amphibian were present in samples collected at the end study but (December 2019 to February 2020). Reads mapping to mammalian RNA viruses were less frequent, the *Astroviridae* family being the most frequently recognized. No clear difference was observed between the two sampling sites or over the studied period.

### Identification Vertebrate RNA Virus Sequences That May Infect Humans

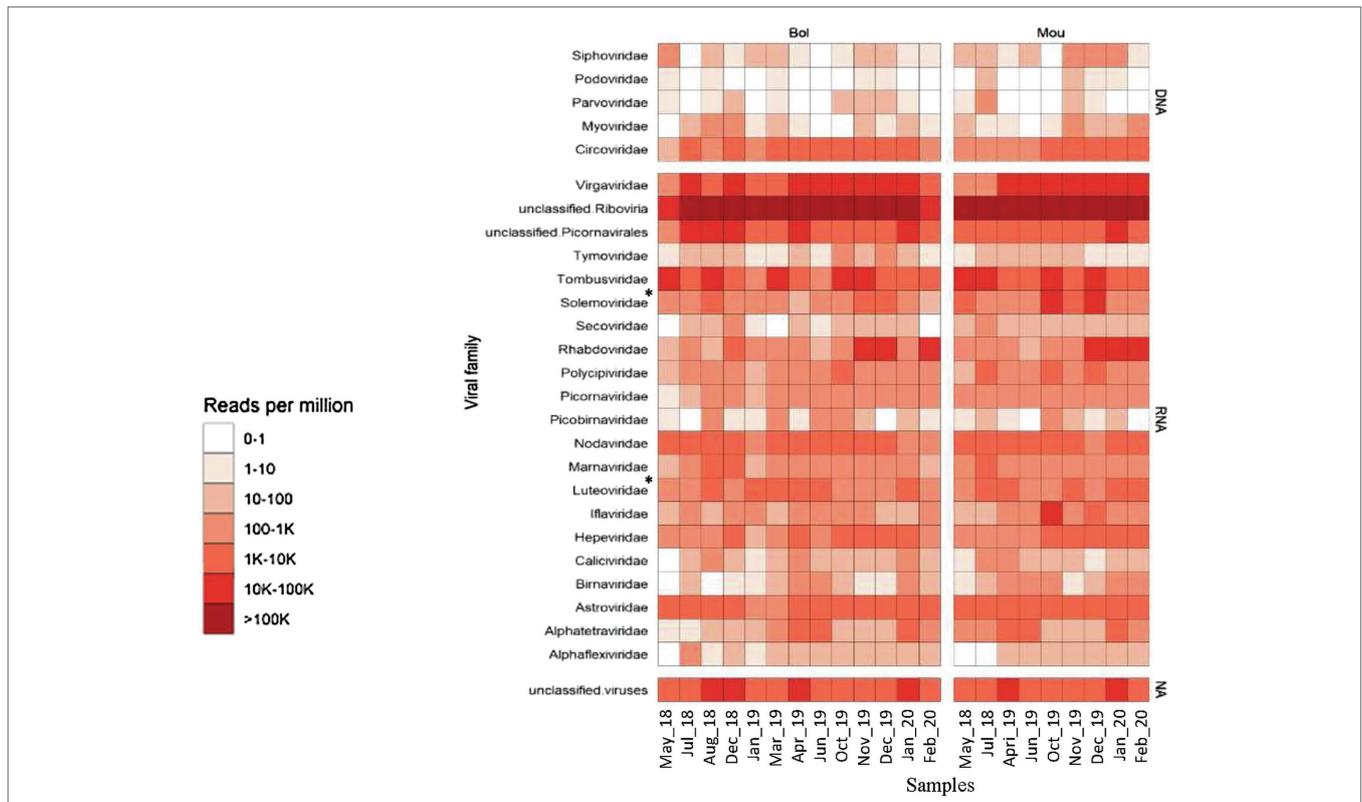
Viral contigs longer than 300 bases, with at least five reads mapping on the same area and an E-value threshold of  $10^{-15}$  for the BLAST, were selected to identify viral families that may be of interest for human disease. This strategy was used to identify sequences belonging to *Astroviridae*, *Caliciviridae*, *Hepeviridae*, *Picobirnaviridae*, and *Picornaviridae* families. Using these criteria, selected contigs were identified using either BLASTn or DIAMOND, and the host corresponding to the

best BLAST match was considered (**Figure 4**). Sequences identified as belonging to the *Picobirnaviridae* family showed the greatest host diversity with about the same proportion of sequences related to strains infecting bats or other mammal, and few sequences related to strains infecting humans or unspecified hosts. Sequences identified as belonging to *Caliciviridae* or *Picornaviridae* families were mainly related to strains detected in bat samples, while many sequences identified as *Astroviridae* were not clearly related to any host. Sequences identified as *Hepeviridae* were almost all related to strains detected in sponge samples (data not shown).

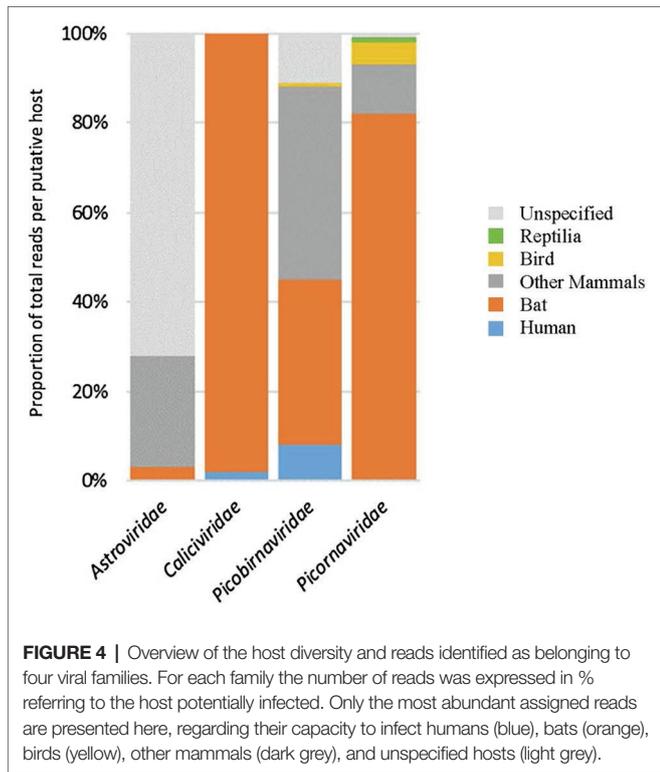
### *Astroviridae*

Contigs identified as belonging to the *Astroviridae* family were the most abundant with 841 sequences obtained representing 996,887 reads. A large number of these contigs (620) were similar to sequences previously identified as bastrovirus detected in sewage samples in Brazil (dos Anjos et al., 2017), including four sequences longer than 4,000 nucleotides (**Table 3**). One hundred seventeen contigs from 20 samples were found close to a bastro-like virus detected in stools of Vietnamese bats. Three sequences (2,311–5,030 bases) distantly related to non-structural polyprotein of a porcine bastrovirus were detected in three samples from both sites (**Table 3**). Regarding astrovirus, 85 sequences, of which 65 sequences ranging from 1,061 to 7,084 bases, were related to a sequence detected in a soil sample in the United States. Of note, the longest contig obtained (7,084 nucleotides) mapped only partially (1/3; **Table 3**). Two sequences were found very similar to an astrovirus infecting a bird (European roller).

Phylogenetic trees based on the RdRp (**Figure 5A**) and capsid proteins (**Figure 5B**) confirmed that these bastrovirus sequences clustered together in one clade with sequences identified in Brazil or Vietnam.



**FIGURE 3 |** Heatmap of the viral diversity. The read abundance after quality control and deduplication is presented for the two sites (Bol and Mou). Color gradients represent normalized relative abundance of reads expressed in reads per million (RPM) for each sample belonging to taxonomic groups indicated, with the top five rows corresponding to DNA viruses, the last row to unclassified viruses and middle rows to RNA viruses. \*Since 2021, *Luteoviridae* was assigned to *Solemoviridae*.



**FIGURE 4 |** Overview of the host diversity and reads identified as belonging to four viral families. For each family the number of reads was expressed in % referring to the host potentially infected. Only the most abundant assigned reads are presented here, regarding their capacity to infect humans (blue), bats (orange), birds (yellow), other mammals (dark grey), and unspecified hosts (light grey).

**Caliciviridae**

Initially, 21 of the 22 selected samples were positive for human noroviruses as detected by RT-PCR. Only 64 contigs (352–1,833 bases) were assigned as *Caliciviridae*. Importantly, three of them were very likely of human origin: two contigs shared around 96% nucleotide identity with a human norovirus GII.6 strain detected in a stool from China, and a third contig displayed 97% nucleotide identity with a human norovirus detected in Japan (Table 4). Several *Caliciviridae* contigs showed similarities with bat calciviruses, five were related to the bat calcivirus A10 (Table 4). The longest *Caliciviridae* contigs (1,026–1,833 bases) were close to a sequence identified in a tricolored bat (*Perimyotis subflavus*) in the United States.

**Picobirnaviridae**

A large number of reads (21,958) were identified as belonging to the *Picobirnaviridae* family. These reads were realigned on 45 contigs of 354–2,492 base pairs identified as close to algae, human, bat, and monkey picobirnavirus strains. Eight contigs ranging from 391 to 615 base pairs were identified as closely related to human picobirnavirus sequences. Among these, five contigs, detected in Bol site, exhibited high nucleotide homology with a virus detected in human stool in India (Table 5). Other contigs were also related to human strains, one being previously detected in Cameroonian stool. Phylogenetic analysis of two

**TABLE 3** | *Astroviridae* contigs identified in Sanaga clams.

Site	Date	Accession no.*	Length (nt)	Mapped reads	Nucleotide match	Identity (%)	E-value	Name	Ref. sequence
Bol	September 18	MW924352	5,823	27,103	4,221	87.6	0	Bastrovirus	ASM79505.1
Bol	September 18	MW924353	5,819	22,861	4,221	95.2	0	Brazil-sewage	
Bol	September 18	MW924354	5,728	17,434	4,221	95.4	0		
Mou	July 18	MW924355	4,312	6,571	2,901	94.5	0		
Mou	July 18	MW924356	2,390	8,659	972	92.9	$9.6 \times 10^{-164}$		ASM79506.1
Mou	May 18	MW924357	1,571	991	1,233 <sup>#</sup>	95.0	0	Bastrovirus-like virus	KX907135.1
Mou	July 18	MW924358	1,360	1,291	1,287 <sup>#</sup>	82.2	0	VietNam-Bat	
Mou	July 18	MW924359	1,097	1,161	1,097 <sup>#</sup>	79.8	0		
Bol	December 18	MZ182250	389	99	376 <sup>#</sup>	84.0	$1.7 \times 10^{-95}$	Bastrovirus-like virus, US-Culex	NC_040647.1
Bol	August 18	MZ182271	5,030	1,874	1,044	29.0	$1.2 \times 10^{-27}$	Bastrovirus	QDA77022.1
Bol	February 20	MZ182272	3,325	70,746	1,044	29.0	$8.1 \times 10^{-28}$	US-Porc	
Mou	May 18	MZ291967	2,311	7,679	1,062	27.7	$7.4 \times 10^{-28}$		
Bol	October 19	MZ291968	404	36	390	56.9	$3.4 \times 10^{-37}$	Astrovirus	QBZ38214.1
Mou	November 19	MZ291969	361	142	324	53.7	$3.7 \times 10^{-27}$	Hungary-european roller	
Bol	January 19	MZ291970	3,801	4,413	798	31.2	$9.3 \times 10^{-20}$	<i>Astroviridae</i> sp.	QDH87307.1
Mou	June 19	MZ291971	7,084	9,823	2,316	25.4	$7.3 \times 10^{-34}$	USA-soil	

\*This accession number corresponds to the sequence identified in this study and deposited in Genbank. Contig identification was done using DIAMOND except for some sequences noted with # for which BLASTn was used. Sequence identity is expressed as a percentage. E-value is from the BLAST tool.

of them based on RdRp sequences revealed clustering of both in genogroup I (Figure 6).

Beside human picobirnavirus, three contigs displayed similarities with a simian strain identified in gorilla stool in Democratic Republic of Congo and three contigs with strains detected in macaques (Duraisamy et al., 2018). Two contigs of 2,023–2,029 base pairs, detected in both sites in October 2019, were distantly related to a picobirna-like virus isolated from Cameroonian bats (*Eidolon helvum*; Yinda et al., 2018). Phylogenetic analysis of the longest one, based on RdRp sequences, showed clustering in a clade of uncharacterized picobirna-like viruses that use an alternative invertebrate mitochondrial genetic code (Figure 6).

### Picornaviridae

Thirty-eight contigs gathering 69,517 reads, with sizes ranging from 353 to 4,631 bases, were identified as belonging to the *Picornaviridae* family. A short contig of 353 bases genetically close to the genogroup V of the simian hepatitis A virus was identified in a sample collected in Bol in October 2019 (Table 6). This sequence shared a high nucleotide identity (91.7%) with the capsid gene of a simian strain detected in water samples in South Africa (Rachida et al., 2016). We previously sequenced a similar strain in a Sanaga clam sample collected in this area, albeit, with only 18% sequence overlap (Bonny et al., 2020).

With the exception of one contig displaying 76% amino-acid identity with its best DIAMOND match, the other contigs were more distantly related (26–38% amino-acid identity) to *Picornaviridae* sequences from the database. However, e-values remained in the acceptable range. Contigs related to bat picornavirus were the most abundant (42%) and were present in 68% of samples. With sizes ranging from 382 to 4,631 bases, most of them were closest at the protein level to the

Washington Bat picornavirus strain (Table 6). Other contigs were identified as being close to porcine picornavirus (453–1,220 bases) with similarities to sapelovirus A, sapelovirus 1 or kobuvirus, and few of them to chicken megrivirus.

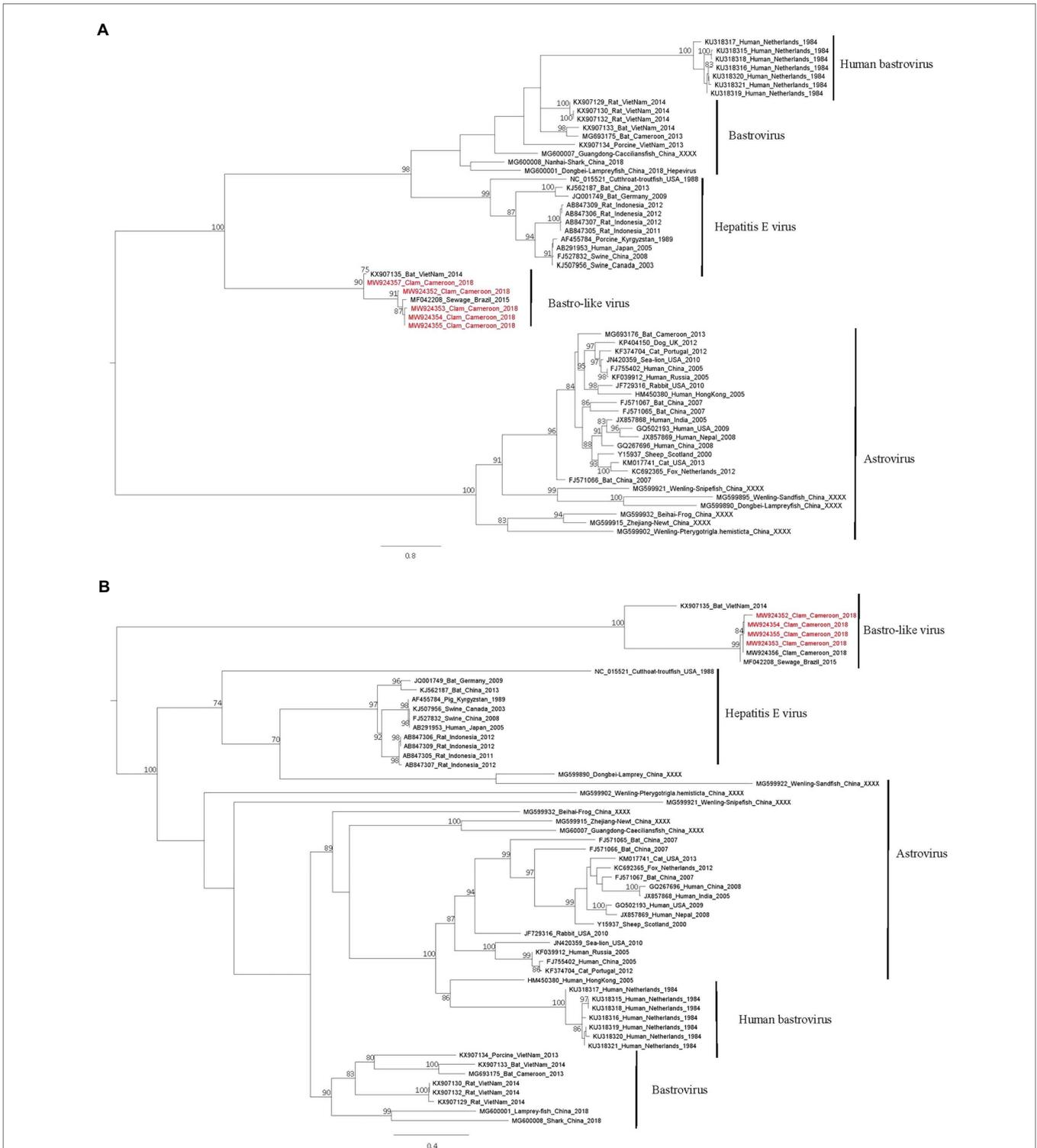
### Hepeviridae

A large number of contigs (188) of various lengths (300–8,223 nucleotides) representing 390,728 reads were identified as belonging to the *Hepeviridae* family. Most of them (92%) were identified as close to sponge hepe-like virus. Only four contigs were related to mammalian viral sequences (Table 7): two were close to sequences detected in swine feces and the two others to a strain detected in mice.

## DISCUSSION

The current COVID-19 pandemic highlights the importance of monitoring the possible transmission of viruses from wildlife to human. For this, the main challenge rests in detecting the viruses infecting diverse animal populations in their natural reservoirs and to predict their zoonotic potential (Osterhauss et al., 2020). Here, we propose to use clams as sentinels to detect potential environmental contamination with animal or human viruses, provided that identified sequences have been adequately related to their hosts. Clams, as other bivalve molluscan shellfish, filter large volumes of water. Furthermore, as benthic animals, they also are in close contact with sedimented particles such as the one originating from feces. In areas where collecting representative water samples is difficult, bivalve mollusks sampling may be a good alternative.

Preserving all viral genetic diversity presents in a sample is not possible yet since protocols used to date include steps



**FIGURE 5 |** Phylogenetic trees of the *Astroviridae* family based on RdRp (A) and capsid (B) proteins. Reference sequences were selected for the main genus of the *Astroviridae* family, sequences identified in this study are written in red. The tree was constructed using IQtree with the LG+I+G4 substitution model for the RdRp and the LG+G4 model for the capsid and ultrafast bootstrap analysis. Trees were mid-rooted using FigTree and bootstrap values greater than 70% are shown.

that may favor the selection of certain types of viral particles. These can be, for example, (1) sampling bias, as collecting samples from the environment may be source of variations,

(2) concentration methods that may eliminate some viruses of interest or increase the presence of inhibitors that may impact the efficacy of enzymes used for the molecular detection,

**TABLE 4** | *Caliciviridae* sequences identified in Sanaga clams.

Site	Date	Accession no*	Length (nt)	Mapped reads	Nucleotide match	Identity (%)	E-value	Names	Ref. sequence
Bol	September 18	MW854003	367	176	366#	96.2	$3.04 \times 10^{-167}$	Norovirus GII.[P6]6	KU870455.1
Bol	September 18	MW853999	353	101	353#	95.7	$2.28 \times 10^{-158}$		
Mou	December 19	MW854002	387	10	387#	97.1	0	Norovirus GII.[Pe]2	LC209439.1
Bol	January 20	MZ291972	1,833	1,003	699	87.1	$1.10 \times 10^{-114}$		
Bol	April 19	MZ291973	1,328	846	1,281	71.0	$2.00 \times 10^{-155}$		
Bol	September 18	MZ291974	1,026	670	1,026	51.3	$6.70 \times 10^{-74}$	Bat calicivirus A10	AWK23451.1
Mou	April 19	MZ291975	1,653	769	1,389	70.8	$1.10 \times 10^{-169}$		
Mou	April 19	MZ291976	1,086	155	294	70.4	$1.70 \times 10^{-30}$		

\*The accession number corresponds to the sequence identified in this study and deposited in Genbank. Contig identification was done using Diamonds except for some sequences noted with # for which BLASTn was used. Sequence identity is expressed as a percentage. E-value is from the BLAST tool.

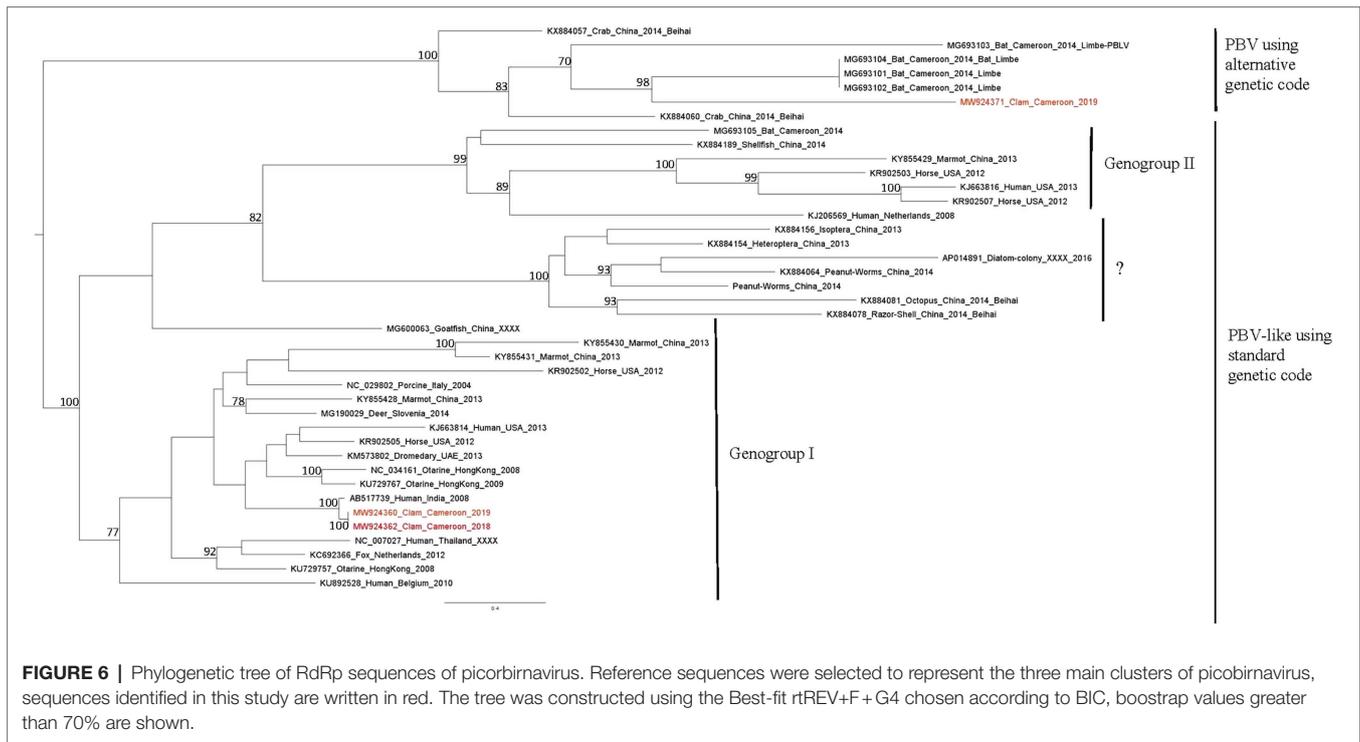
**TABLE 5** | *Picobirnaviridae* sequences identified in Sanaga clams.

Site	Date	Accession no*	Length (nt)	Mapped reads	Nucleotide match	Identity (%)	E-value	Names	Ref. sequence
Bol	September 18	MZ291977	399	260	388#	94.1	$2.00 \times 10^{-164}$		
Bol	December 18	MW924363	409	8	403#	92.1	$9.65 \times 10^{-158}$		
Bol	January 19	MW924361	582	10	581#	93.1	0	Human picobirnavirus, India-stool	AB517739.1
Bol	April 19	MW924360	615	6	615#	93.2	0		
Bol	September 18	MW924362	645	905	639#	93.9	0		
Bol	January 19	MW924364	417	53	418#	90.4	$3.57 \times 10^{-152}$	Human picobirnavirus, India-stool	AB517731.1
Mou	October 19	MW924365	539	72	303#	88.1	$3.05 \times 10^{-94}$	Human picobirnavirus, Belgium-stool	KU892528.1
Mou	October 19	MW924366	429	63	429#	92.8	$9.95 \times 10^{-173}$	Human Picobirnavirus, Cameroon-stool	MH933803.1
Bol	September 18	MW924367	600	328	600#	94.7	0		
Bol	September 18	MW924368	600	286	600#	96.0	0	Picobirnavirus, Congo Rep-gorilla feces	KY502865.1
Bol	September 18	MW924369	503	294	503#	98.8	0		
Bol	May 18	MZ291978	378	18	203#	90.6	$2.79 \times 10^{-68}$	Macaque picobirnavirus, USA-macaque	MG010919.1
Bol	October 19	MW924372	2,023	2,894	1,725	33.7	$8.90 \times 10^{-70}$	Picobirna-like virus, Cameroon-bats	AWV66962.1
Mou	October 19	MW924371	2,029	3,215	1,689	33.6	$2.20 \times 10^{-68}$		

\*This accession number corresponds to the sequence identified in this study and deposited in Genbank. Contig identification was done using Diamond except for some sequences noted with # for which BLASTn was used. Sequence identity is expressed as a percentage. E-value is from the BLAST tool.

(3) nucleic acid extractions and subsequent library preparation including DNase or RNase treatments, or (4) at the last step, the bioinformatic pipeline. Furthermore, searching for viruses with a short RNA genome present in low concentrations adds difficulties. A number of different approaches may help to solve some of these issues (Briese et al., 2015; Conceicao-Neto et al., 2015; Kim et al., 2017; Desdouts et al., 2020). In our previous work, we proposed methodological approaches that aimed to eliminate as much as possible bacteria, shellfish tissues, or free DNA or RNA (Strubbia et al., 2020), to focus on RNA viruses. This likely explains the low amount of DNA and phages viruses reads in our study unlike agnostic metagenomic studies where these viruses constituted the largest viral fraction (Bačnik et al., 2020; Nieuwenhuijse et al., 2020). We also performed

technical triplicates of cDNA synthesis and libraries as we observed that it increases the odds to recover sequences from viruses present at low concentrations (Strubbia et al., 2020). Generating technical or biological replicates help to increase virus diversity (Kim et al., 2017). Furthermore, we used an enrichment step based on capture of cDNA from vertebrate viruses using the VirCapSeq-VERT assay during libraries preparation (Wylie et al., 2015; Strubbia et al., 2019). Finally, considering the low concentrations of norovirus and hepatitis A virus as measured by quantitative PCR, we selected a NovaSeq run to increase the depth per sample and thus increase the number of viral reads. As anticipated, we observed that viral reads remained a minority among the huge amount of reads that were obtained. These very large and diverse datasets resulted



in long, difficult, and sometimes impossible *de novo* assembly. Read deduplication has been shown to allow a faster and easier assembly by removing repeated reads, mostly generated by the amplification required to reach minimal DNA amounts for sequencing (Nieuwenhuijse et al., 2020). Here, this bioinformatic treatment considerably reduced fractions of bacteria and eukaryote reads, increased the number of unclassified reads, while fractions of viral reads remained stable in all samples. After deduplication, *de novo* assembly yielded assembled contigs for all samples, with a reduced computing time.

Another major issue in the metagenomic analysis of environmental samples is the presence of a large amount of unclassified sequences referred to as dark matter, formally defined as any nucleotide sequence that cannot be taxonomically assigned by alignment to any reference nucleotide or amino acid sequence (Krishnamurthy and Wang, 2017; Sunagawa et al., 2020). Samples derived from complex mixtures like those from mammalian guts or environmental samples are known to include novel organisms not described in reference sets or otherwise undocumented (Connor et al., 2019). Characterization of the viral fraction is especially difficult as only a small portion of Earth's viral diversity has been identified and made available in reference sets (Connor et al., 2019; Plyusnin et al., 2020; Sunagawa et al., 2020). In our study, up to 87% of reads could not be identified, and 96% of RNA viral reads could not be assigned to any viral family. Such observations previously reported in other studies, confirmed that although advances in high-throughput sequencing expand the boundaries of viral sequence space, some developments are needed to increase viral annotations (Bačnik et al., 2020; Nieuwenhuijse et al., 2020; Dance, 2021).

In any case, it is important to continue improving bioinformatic pipelines and databases to be able to screen rapidly metagenomic data and fully identify the organisms in a sample (Nieuwenhuijse et al., 2020; Plyusnin et al., 2020; Dance, 2021).

To reliably identify the very diverse sequences obtained through *de novo* assembly, each contig was identified by BLAST considering the length of overlap with the database sequence match, the numbers of reads and the BLAST E-value. The viral diversity in our study is described for all samples together rather than comparing the two sites or the season as the same sequences were detected in both sites and considering that the number of samples or the duration of the study may not be long enough to be able to detect some differences.

Following a one-health approach, we focused on the identification of RNA mammalian viral sequences, and especially on the main viral families potentially related to human health and known to be present in human or animal excreta such as *Astroviridae*, *Caliciviridae*, *Hepeviridae*, *Picobirnaviridae*, and *Picornaviridae*. Of note, since we rely on the sequence only, host assignment in our study is based on the host identified for the best BLAST match, which can be quite distantly related to our sequences. We grouped putative hosts into large families like humans, swine, or bats (Figure 3), but this information should be considered as indicative, and not definitive, as some sequences may have been detected in samples that are not their replicating host, and hosts may be misidentified in databases (Cobbin et al., 2021).

Few sequences from human viruses were identified compared to the diversity of other sequences detected. This may be due to the low human population density and other environmental parameters such as currents or sewage disposal that were

**TABLE 6** | Picornaviridae sequences identified in Sanaga clams.

Site	Date	Accession no.*	Length (nt)	Mapped reads	Nucleotide match	Identity (%)	E-value	Names	Ref. sequence
Bol	October 19	MM924373	353	6	351#	91.7	$6.57 \times 10^{-134}$	Hepatitis A virus, South Africa-water	KP875240.1
Bol	October 19	MZ291981	403	16	402	76.1	$2.30 \times 10^{-54}$	Bat picornavirus	YP_009272812.1
Mou	January 20	MZ291982	4,631	2,926	960	26.6	$3.90 \times 10^{-20}$	USA-bat	
Bol	May 18	MZ291983	453	45	426	37.3	$6.30 \times 10^{-16}$		
Bol	August 18	MZ291984	609	332	426	37.3	$1.10 \times 10^{-15}$	Sapelovirus, USA-swine	ANY94479.1
Mou	April 19	MZ291985	453	111	426	38.0	$3.70 \times 10^{-16}$		
Mou	November 19	MZ291986	784	146	735	31.0	$7.50 \times 10^{-17}$	Porcine sapelovirus, Japan-swine	BBG56086.1
Mou	April 19	MZ291987	632	98	534	35.4	$2.30 \times 10^{-16}$	Porcine kobuvirus,	AJW76498.1
Mou	February 20	MZ291988	1,220	247	525	36.0	$5.80 \times 10^{-16}$	China-swine	
Bol	October 19	MZ291989	862	230	699	34.8	$3.10 \times 10^{-24}$	Chicken megivirus,	AHN10844.1
Bol	December 19	MZ291990	938	18	648	35.2	$1.70 \times 10^{-23}$	Hungary-chicken	AYH53256.1

\*This accession number corresponds to the sequence identified in this study and deposited in GenBank. Contig identification was done using Diamond except for some sequences noted with # for which BLASTn was used. Sequence identity is expressed as a percentage. E-value is from the BLAST tool.

not investigated here. As mentioned above, all but one of the samples selected for the metagenomic analysis were detected positive for norovirus by real-time RT-PCR. Despite this first screening results, very few related sequences were obtained, meaning that the sensitivity of the metagenomics technique needs to be further improved. In a previous study, we were able to identify a large diversity of norovirus in oysters displaying similar concentrations as these clam samples (Strubbia et al., 2019). In the current study, the enrichment step approach may have failed to catch all strains present in the nucleic acid extracts, or the large diversity of sequences obtained could have prevented contig assembly. Among the 40% of reads identified as norovirus and not assembled into contigs, most of them matched to the A-poly tail, suggesting that contig assembly was performant on this virus. Data on norovirus prevalence in this area are still scarce and to our knowledge only one study was published. Based on the analysis of a large panel of stool from all age groups, few sequences of norovirus were identified compared to a much higher diversity of sapovirus sequences that were not detected in our study (Yinda et al., 2019). This observation highlights the need to get more data on molecular epidemiology of viral pathogens in the local population. At variance to norovirus, many picobirnavirus sequences were identified but, despite several attempts, only short contigs could be assembled. Described as opportunistic viruses, their role in the occurrence of gastroenteritis in humans is still poorly studied and the true host remains enigmatic (Ghosh and Malik, 2021). However, a large diversity of picobirnavirus sequences was reported in the gut virome analysis of Cameroonians, most of them being detected in age groups above 20 (Yinda et al., 2019). The authors also mentioned the characterization of picobirna-like virus 99% identical to Cameroonian bat strain picobirna-like viruses.

Most of the contigs identified in this study as mammalian viruses were most closely related to viruses identified in bats. We characterized two sequences longer than 2,000 nucleotides showing similarities with a picobirnavirus strain previously described in Cameroonian fruit bats (Yinda et al., 2018). Moreover a bastrolike-virus sequence was closely related to a strain previously detected in bat samples in VietNam but also in Cameroon; however, with no insurance that the bats were infected (Yinda et al., 2018). Several sequences related to bat calicivirus were identified in up to five samples. All these contigs showed the best homologies with the same strain identified in a bat stool collected the United States (Kocher et al., 2018). The capsid gene of this bat calicivirus A10, was closely related to norovirus genogroup V (GV) and phylogenetically related to recoviruses. This strain displayed similarities with human noroviruses in terms of binding pattern to human histo-blood group antigens, suggesting the potential of this virus to overcome a significant barrier to cross-species transmission (Kocher et al., 2018). The potential zoonotic transmission is a long-standing question regarding norovirus, with scarce data reported so far, highlighting the need to identify the potential reservoir (Villabrana et al., 2019). A bat picornavirus sequence was

**TABLE 7** | *Hepeviridae* sequences identified in Sanaga clams.

Site	Date	Accession no*	Length (nt)	Mapped reads	Nucleotide match	Identity (%)	E-value	Names	Ref. sequence
Bol	February 20	MZ476780	586	49	498	85	$7.40 \times 10^{-17}$	Hepatitis E virus, China-swine	ACD44941.2
Bol	December 18	MZ476781	572	46	570	100	$1.30 \times 10^{-18}$	Hepatitis E virus, China-rabbit	BAM68889.1
Bol	April 19	MZ476782	558	34	498	89	$1.10 \times 10^{-22}$	Hepe-like virus, USA-mice	AWB14594.1
Mou	June 19	MZ476783	611	96	537	88	$4.50 \times 10^{-17}$		

\*This accession number corresponds to the sequence identified in this study and deposited in Genbank. Contig identification was done using Diamond except for some sequences noted with \* for which BLASTn was used. Sequence identity is expressed as a percentage. E-value is from the BLAST tool.

also identified in two samples. Taken together these observations suggest a repeated contamination of the studied area by bat feces. Bats are considered as a major reservoir of viruses, and several studies reported the characterization of viruses closely related to human viruses confirming their potential role for zoonotic events (Hardmeier et al., 2018; Yinda et al., 2018). Based on our study, we suggest to consider the environmental contamination by these animals as it may constitute a driver for human contamination (Gortazar et al., 2014; Kohl et al., 2021).

Among the animal virus sequences detected in this work, those infecting pigs were second in terms of diversity. We detected not only some sapelovirus and kobuvirus from the *Picornaviridae* family but also some bastrovirus, bastro-like virus, and some Hepatitis E virus. Considering the capacity of this virus to infect across species barriers, monitoring prevalence of this virus in pigs in the region may be of interest (Sooryanarain and Meng, 2019).

In a previous study conducted in this area, we sequenced a portion of the genome of a hepatitis A virus, related to genogroup V which include simian viruses, and most closely related to a potentially simian virus identified in surface waters (Bonny et al., 2020). We detected again this sequence and also some simian picobirnavirus sequences confirming the contamination of the area with macaque feces. The Sanaga river, a rural area, harbors a wildlife reserve (**Figure 1**) characterized by abundant and diverse fauna, including an important simian population. Zoonotic transmissions have been proven from non-human primates to humans for different viral families (Devaux et al., 2019; Di Cola et al., 2021). However, none of the human viral strains from Cameroon were related to apes in the study conducted in 2014 (Yinda et al., 2019).

Our approach considering the diversity of RNA virus sequences detected in this area may contribute to the surveillance of zoonotic events (Osterhauss et al., 2020; Williams et al., 2021). Indeed, despite a need for better sensitivity, we report here a method that succeeded to describe a large diversity of RNA viruses from clam samples. In addition to their role as a sentinel of the microbial environmental quality, shellfish constitute an easy source of proteins for local populations with low economic resources (Ajonina et al., 2005; Venugopal and Gopakumar, 2018). While the way of consumption (boiled and smoked) likely prevents foodborne transmission of these viruses, the first steps of opening the shell and collecting the raw flesh may be at risk for the

workers. Thus, it is important to share information and provide some recommendation to local populations.

Metagenomic approaches are very promising tools to describe viral diversity and will be of particular interest for food and environmental samples analysis (Desdoutis et al., 2020). It is especially relevant for shellfish analysis. Indeed, bivalve molluscan shellfish, a nutrient source for many populations, may also reflect microbial contamination shed by surrounding animal or humans. Being able to describe all viral contaminants in molluscan bivalve shellfish, whether considered as a food or environmental sentinel may contribute to the better understanding of some zoonotic transmission events and alert health authorities of possible emergence.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://doi.org/10.12770/5abfae46-aca4-49ad-b1bc-4fe5a1ba4454>.

## AUTHOR CONTRIBUTIONS

PB, JS, MD, and FG contributed to the conception and design of the study. PB, JS, and AB performed the bio-info analysis. PB, JS, AB, MD, JE, and FG contributed to data interpretation and manuscript preparation. All authors contributed to the article and approved the submitted version.

## FUNDING

This study was supported by the VEO European project (H2020, SC1-2019-874735). PB was supported by a founding from Ministry of Scientific Research and Innovation of Cameroon.

## ACKNOWLEDGMENTS

This work benefited from equipment and services from iGenSeq core facility, at ICM (Institut Cerveau Moelle, Paris, France). We thank B. Oude Munnink and D. Nieuwenhuijse for helpful scientific discussion.

## REFERENCES

- Ajonina, P. U., Ajonima, G., Jin, E., Mekongo, F., Ayisse, I., and Usongo, L. (2005). Gender roles and economics of exploitation, processing and marketing of bivalves and impacts on forest resources in the Sanaga Delta region of Douala-Edea wildlife reserve, Cameroon. *Int. J. Sustain. Dev. World Ecol.* 12, 161–172. doi: 10.1080/13504500509469627
- Bačnik, K., Kutnjak, D., Pecman, A., Mehle, N., Žnidarič, M. T., Aguirre, I. A., et al. (2020). Viromics and infectivity analysis reveal the release of infective plant viruses from wastewater into the environment. *Wat. Res.* 177:115628. doi: 10.1016/j.watres.2020.115628
- Bonny, P., Desdoutis, M., Schaeffer, J., Garry, P., Essia Nganag, J. J., and Le Guyader, F. S. (2020). Contamination of clams with human norovirus and a novel hepatitis A virus in Cameroon. *Food Environ. Virol.* 12, 274–277. doi: 10.1007/s12560-020-09432-2
- Briese, T., Kapoor, A., Mishra, N., Jain, K., Kumar, A., Jabado, O. J., et al. (2015). Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *MBio* 6, e01491–e01415. doi: 10.1128/mBio.01491-15
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using diamond. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Cobbin, J., Charon, J., Harvey, E., Holmes, E. C., and Mahar, J. (2021). Current challenges to virus discovery by meta-transcriptomics. *Curr. Opin. Virol.* 51, 48–55. doi: 10.1016/j.coviro.2021.09.007
- Conceicao-Neto, N., Zeller, M., Lefrere, H., de Bruyn, P., Beller, L., Deboutte, W., et al. (2015). Molecular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Sci. Rep.* 5:16532. doi: 10.1038/srep16532
- Connor, R., Brister, R., Buchmann, J. P., Deboutte, W., Edwards, R., Martí-Carreras, J., et al. (2019). NCBI's virus discovery Hackathon: engaging research communities to identify cloud infrastructure requirements. *Genes* 10:714. doi: 10.3390/genes10090714
- Dance, A. (2021). The incredible diversity of viruses. *Nature Rev.* 595, 22–25.
- Desdoutis, M., de Graaf, M., Strubbia, S., Oude Munnink, B. B., Kroneman, A., Le Guyader, F. S., et al. (2020). Novel opportunities for NGS-based one health surveillance of foodborne viruses. *One Health Outlook* 2:14. doi: 10.1186/s42522-020-00015-6
- Devaux, C. A., Mediannikov, O., Medkour, H., and Raoult, D. (2019). Infectious disease risk across the growing human-non human primate interface: a review of the evidence. *Front. Public Health* 7:305. doi: 10.3389/fpubh.2019.00305
- Di Cola, G., Fantalli, A. X., Pisano, M. B., and Ré, V. E. (2021). Foodborne transmission of hepatitis A and hepatitis E viruses: a literature review. *Int. J. Food Microbiol.* 338:10898986. doi: 10.1016/j.ijfoodmicro.2020.108986
- dos Anjos, K., Nagata, T., and Melo, F. L. D. (2017). Complete genome sequence of a novel bastrovirus isolated from raw sewage. *Genome Announc.* 5, e01010–e01017. doi: 10.1128/genomeA.01010-17
- Duraisamy, R., Akiana, J., and Mediannikov, Davoust B., O., Michelle, C., Robert, C. et al. (2018). Detection of novel RNA viruses from free-living gorillas, Republic of the Congo: genetic diversity of picobirnaviruses. *Virus Genes* 54: 256–271. doi: 10.1007/s11262-018-1543-6
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565
- Ghosh, S., and Malik, Y. S. (2021). The true host/s of picobirnaviruses. *Front. Vet. Sci.* 7:615293. doi: 10.3389/fvets.2020.615293
- Gortazar, C., Reperant, L. A., Kuiken, T., de la Fuente, J., Boadella, M., Martínez-Lopez, B., et al. (2014). Crossing the interspecies barriers: opening the door to zoonotic pathogens. *Plos Pathog.* 10:e1004129. doi: 10.1371/journal.ppat.1004129
- Hardmeier, I., Aeberhard, N., Qi, W., Schoenbaechler, K., Krattli, H., Hatt, J.-M., et al. (2018). Metagenomic analysis of fecal and tissue samples from 18 endemic bat species in Switzerland revealed a diverse virus composition including potentially zoonotic viruses. *PLoS One* 16:e025534. doi: 10.1371/journal.pone.025534
- ISO 15216-1. (2017). Microbiology of food and animal feed -horizontal method for determination of hepatitis A virus and norovirus in food using real-time RT-PCR, Part 1: method for quantification.
- Kim, Y., van Bonn, W., Aw, T. G., and Rose, J. B. (2017). Aquarium viromes: viromes of human-managed aquatic systems. *Front. Microbiol.* 8:1231. doi: 10.3389/fmicb.2017.01231
- Kocher, J. F., Lindesmith, L. C., Debbink, K., Beall, A., Mallory, M. L., Yount, B. L., et al. (2018). Bat caliciviruses and human noroviruses are antigenically similar and have overlapping histo-blood group antigen binding profiles. *MBio* 9, e00869–e00818. doi: 10.1128/mBio.00869-18
- Kohl, C., Nitsche, A., and Kurth, A. (2021). Update on potentially zoonotic viruses of European bats. *Vaccine* 9, 690. doi: 10.3390/vaccines9070690
- Krishnamurthy, S. R., and Wang, D. (2017). Origins and challenges of viral dark matter. *Virus Res.* 239, 136–142. doi: 10.1016/j.virusres.2017.02.002
- La Rosa, G., Sanseverino, I., Della Libera, S., Laconelli, M., Ferrero, V. E. V., Barra Caracciolo, A., et al. (2017). The impact of anthropogenic pressure on the virological quality of water from the Tiber river, Italy. *Lett. Appl. Microbiol.* 65, 298–305. doi: 10.1111/lam.12774
- Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Metcalf, T., Melnick, J. L., and Estes, M. K. (1995). Environmental microbiology: from detection of virus in sewage and water by isolation to identification by molecular biology- a trip of over 50 years. *Annu. Rev. Microbiol.* 49, 461–487. doi: 10.1146/annurev.mi.49.100195.002333
- Nieuwenhuijse, D., and Koopmans, M. P. G. (2017). Metagenomic sequencing for surveillance of food- and waterborne viral diseases. *Front. Microbiol.* 8:230. doi: 10.3389/fmicb.2017.00230
- Nieuwenhuijse, D. F., Oude Munnink, B., Phan, M. V. T., the Global sewage Surveillance projet consortiumMunk, P., Venkatakrishnan, S., et al. (2020). Setting a baseline for global urban virome surveillance in sewage. *Sci. Rep.* 10:13748. doi: 10.1038/s41598-020-69869-0
- Nooij, S., Schmitz, D., Vennema, H., Kroneman, A., and Koopmans, M. P. G. (2018). Overview of virus metagenomic classification methods and their biological applications. *Front. Microbiol.* 9:749. doi: 10.3389/fmicb.2018.00749
- Nurk, S., Meleshko, D., Korobeynikov, A., and Tevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27, 824–834. doi: 10.1101/gr.213959.116
- Osterhauss, A. D. M. E., Vanlangendonck, C., Barbeshi, M., Bruschie, M., Christensen, R., Daszak, P., et al. (2020). Make science evolve into a One Health approach to improve health and security: a white paper. *One Health Outlook* 2:6. doi: 10.1186/s42522-019-0009-7
- Plyusnin, I., Kant, R., Jaaskelainen, A. J., Sironen, T., Holm, L., Vapalahti, O., et al. (2020). Novel NGS pipeline for virus discovery form a wide spectrum of hosts and sample types. *Virus Evol.* 6:veaa091. doi: 10.1093/ve/veaa091
- Polo, D., Lois, M., Fernandez-Nunez, M. T., and Romalde, J. L. (2021). Detection of SARS-CoV-2 RNA in bivalve mollusks and marine sediments. *Sci. Total Environ.* 786:147534. doi: 10.1016/j.scitotenv.2021.147534
- Rachida, S., Matsapola, P. N., Wolfaardt, M., and Taylor, M. B. (2016). Genetic characterization of a novel hepatitis A virus strain in irrigation water in South Africa. *J. Med. Virol.* 88, 734–737. doi: 10.1002/jmv.24370
- Sooryanarain, H., and Meng, X.-J. (2019). Hepatitis E virus: reasons for emergence in humans. *Curr. Opin. Virol.* 34, 10–17. doi: 10.1016/j.coviro.2018.11.006
- Strubbia, S., Schaeffer, J., Besnard, A., Wacrenier, C., Le Mennec, C., Garry, P., et al. (2020). Metagenomic to evaluate norovirus genomic diversity in oysters: impact on hexamer selection and targeted capture-based enrichment. *Int. J. Food Microbiol.* 323:108588. doi: 10.1016/j.ijfoodmicro.2020.108588
- Strubbia, S., Schaeffer, J., Oude Munnink, B. B., Besnard, A., Phan, M. V. T., Nieuwenhuijse, D. F., et al. (2019). Metavirome sequencing to evaluate norovirus diversity in sewage and related bioaccumulated oysters. *Front. Microbiol.* 10:3389. doi: 10.3389/fmicb.2019.02394
- Sunagawa, S., Acina, S. G., Bork, P., Bowler, C., Coordinators, T. O., Eveillard, D., et al. (2020). Tara oceans: towards global ocean ecosystems biology. *Nature Rev. Microbiol.* 18. doi: 10.1038/s41579-020-0364-5

- Venugopal, V., and Gopakumar, K. (2018). Shellfish: nutritive value, health benefits, and consumer safety. *Compr. Rev. Food Sci. Food Saf.* 16, 1219–1242. doi: 10.1111/1541-4337.12312
- Villabruna, N., Koopmans, M. P. G., and de Graaf, M. (2019). Animals as reservoir for human norovirus. *Viruses* 11:478. doi: 10.3390/v11050478
- Williams, E. P., Sprill-Harrel, B. M., Taylor, M. K., Lee, J., Nywening, A. V., Yang, Z., et al. (2021). Common themes in zoonotic spillover and disease emergences: lessons learned from bat and rodent-borne RNA viruses. *Viruses* 13:1509. doi: 10.3390/v13081509
- Wylie, T. N., Wylie, K. M., Herter, B. N., and Storch, G. A. (2015). Enhanced virome sequencing using targeted sequence capture. *Genome Res.* 25, 1910–1920. doi: 10.1101/gr.191049.115
- Yinda, C. K., Ghogomu, S. M., Conceicao-Neto, N., Beller, L., Deboutte, W., Vanhulle, E., et al. (2018). Cameroonian fruit bats harbor divergent viruses, including rotavirus H, bastroviruses, and picobirnaviruses using an alternative genetic code. *Virus Evol.* 3:vey008. doi: 10.1093/ve/vey008
- Yinda, C. K., Vanhulle, E., Conceicao-Neto, N., Beller, L., Deboutte, W., Shi, C., et al. (2019). Gut microbiome of Cameroonians reveals high diversity of enteric viruses, including potential interspecies transmitted viruses. *mSphere* 4, e00585–e00518. doi: 10.1128/mSphere.00585-18
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2021 Bonny, Schaeffer, Besnard, Desdouits, Essia Ngang and Le Guyader. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



N° 001/19 /AVT/MINEPIA/SG/DREPIA-CE/SRSV

## AUTORISATION VÉTÉRINAIRE DE TRANSPORT DES ÉCHANTILLONS BIOLOGIQUES

*Dr. Bemwa Jonas* **Chef de Service**  
**des Services Vétérinaires**

Je soussigné..... **Médecin Vétérinaire**..... **des Services Vétérinaires**..... Vétérinaire officiel  
du..... **CAMEROUN**..... certifie avoir examiné **#500 #** échantillons  
de..... **Palaudes (*Egeria radiata*)**..... légalement  
récoltés par..... **Monsieur BONNY Patrice** à..... **MDUANKO**.....  
..... **SANAGA MARITIME - DOUALA - CAMEROUN**.....  
Pour..... **des recherches Scientifiques**..... et autorise leur  
expédition..... à..... **NANTE - FRANCE**..... **Laboratoire Santé Environnement**  
**et Microbiologie**..... par..... **AVION**..... le..... **15/03/2019**

### Déclaration du propriétaire

M..... **BONNY PATRICE**..... signature..... **Bonny**..... déclare que :

1. Ces échantillons ont été traités à base de..... **la glace à -20°C**.....  
..... **sans aucun produit chimique**..... et  
conditionnés..... **dans des glacières en polystyrène recouvertes de**  
**la mousse carbonique**..... pour éviter toutes contaminations extérieures ou effusion ;

2. Ces échantillons n'ont pas été en contact avec un produit à risque ;

3. L'expédition sera effectuée dans les conditions garantissant aucune  
contamination avec les agents pathogène après traitement et conditionnement.

**Mention spéciale : Aucune maladie notifiée à l'OIE suivant le « Code international  
des animaux terrestres et aquatiques » de l'OIE n'a été observée chez les  
animaux..... **aquatiques**..... de cette région, ainsi que dans un rayon de  
30 kilomètres autour de la région, pendant les trois (03) mois avant le début du  
prélèvement des échantillons, pour autant qu'elle puisse être transmise à cette  
espèce animale ou aux hommes.**

En foi de quoi, la présente autorisation vétérinaire est établie pour servir et valoir ce  
que de droit.

### Pièces jointes :

- Certificat Sanitaire Vétérinaire
- Attestation de Recherche.

Fait à..... **Yaoundé**..... le..... **27 FEV. 2019**

**LE VÉTÉRINAIRE OFFICIEL**



**Chief de Service**  
**des Services Vétérinaires**

*Dr. Bemwa Jonas*  
**Médecin Vétérinaire**

REPUBLIQUE DU CAMEROUN

Paix - Travail - Patrie

MINISTERE DE L'ELEVAGE, DES PECHEES  
ET DES INDUSTRIES ANIMALES



REPUBLIC OF CAMERCON  
Peace - Work - Fatherland

MINISTRY OF LIVESTOCK, FISHERIES  
AND ANIMAL INDUSTRIES

**CERTIFICAT SANITAIRE VETERINAIRE**  
**Cher de Service No 1286141**  
**VETERINAIRE CERTIFICATE**  
**des Services Vétérinaires**

Médecin Vétérinaire

Je soussigné (I, undersigned)<sup>(1)</sup> : .....  
Certifie avoir examiné (certify that I have today examined)<sup>(2)</sup> : MO. CATA d'échouantillages

de Pêcheuses  
Appartenant à (belonging to) : BONNY Patoke

Demeurant à (residing at) : GRAND du 27/02/2019

N° Autorisation d'exportation / d'Importation <sup>(3)</sup> : ..... of NANTÉ-FRANCE

N° of authorization to export  
Qui nous a déclaré les expédier à (who has declared to us that he is moving them to) : .....

Par (by) ..... le (on) .....  
Ces (these) .....

- Preses-à-la-consommation (fit for human consumption)<sup>(3)</sup>  
- Indemnes de toutes affections cliniquement décelables (free of all clinically detectable conditions)<sup>(3)</sup>

ont été reconnus (have been found)

A (done in) .....

1 Nom et fonction du technicien (name and function of technician)  
2 Nature, espèce et nombre d'animaux ou de denrées alimentaires examinées  
(Nature, species and number of animals of foodstuff inspected)  
3 Rayer la mention inutile (delete where not applicable)



**21 FEV 2019**  
**Cher de Service**  
**des Services Vétérinaires**  
Jonas  
Médecin Vétérinaire