Check for updates

EXTERNAL SCIENTIFIC REPORT

Ifremer Erasmus MC Cefas DTU

APPROVED: 5 May 2022

doi:10.2903/sp.efsa.2022.EN-7348

# Application of Next Generation Sequencing on Norovirus-contaminated oyster samples

Joanna Ollivier[a], James Lowther[b], Marion Desdouits[a], Julien Schaeffer[a], Candice Wacrenier[a], Bas B. Oude Munnink[d], Alban Besnard[a], Frederico Mota Batista[b], Tina Stapleton[b], Anna Charlotte Schultz[c,1], Frank Aarestrup[c], Marion Koopmans[d], Miranda de Graaf[d], Soizick Le Guyader[a]

[a] Ifremer, Microbiology laboratory, Nantes, France; [b] Centre for Environment, Fisheries and Aquaculture Science, Weymouth, UK; [c] National Food Institute, Technical University of Denmark, Lyngby, Denmark; [d] Erasmus University Medical Center, Rotterdam, the Netherlands

## Abstract

Bivalve molluscan shellfish (BMS) contamination with gastroenteritis viruses such as norovirus (NoV) is recognized as a significant public health risk worldwide. These foodborne epidemics are characterized by very low viral concentrations in the implicated foods, and by diverse viruses or multiple NoV strains originating from human sewage, resulting in different strains (co-)infecting the consumers. Next-generation sequencing (NGS) offers promising means to describe the diversity of strains present in BMS or to retrace transmission chains in outbreak settings, but their sensitivity and reproducibility remained to be assessed for this application. In this work, we evaluated the ability of three promising NGS methods to sequence diverse NoV in environmental or food BMS and human stool samples. Using laboratory-prepared samples of known NoV composition, we evaluated the sensitivity, reproducibility, repeatability and selectivity of metabarcoding, capture-based metagenomics and long amplicon sequencing, considering representative NoV strains from genogroup I and II, and the impact of the BMS matrix. The metabarcoding, with separate amplification of polymerase and capsid gene segments followed by Illumina sequencing, was the most sensitive method. It was applied to a selection of 212 BMS samples collected through the European Commission's NoV baseline survey (BLS), demonstrating a high diversity of NoV sequences found in the BMS which reflect the diversity of NoV strains circulating in the European human population. Besides, a capture-based metagenomics with enrichment of vertebrate viruses was applied on 20 of these BLS samples as well as 20 BMS linked to outbreaks and 10 related human stool samples. In BMS, it yielded NoV sequences compatible with the genomes identified in stool samples, but they were too short to allow definitive confirmation of the infection source. The present report describes NGS methods, including the bioinformatic pipelines, applicable to molecular epidemiology of NoV in BMS, their current limitations and expected outcomes.

© European Food Safety Authority, 2022

**Key words:** next-generation sequencing, norovirus, oyster, shellfish, food contamination

**Question number:** EFSA-Q-2019-00153

**Correspondence:** biohaw@efsa.europa.eu

[1] Anna Charlotte Schultz left the DTU and the consortium in February 2020.

**Disclaimer:** The present document has been produced and adopted by the bodies identified above as author(s). This task has been carried out exclusively by the author(s) in the context of a contract between the European Food Safety Authority and the author(s), awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**Acknowledgements:** We would like to thank the National testing laboratories involved in the European baseline survey for norovirus in oysters, for their help and cooperation in provision of oyster samples and nucleic acids. We gratefully acknowledge the authors from laboratories that submitted sequences to the NoroNet Database. We would further like to thank C. Schapendonk, D. Prasad, M. Olthof and R. Izquirdo-Lara for excellent technical assistance.

**Suggested citation:** Ollivier J, Lowther J, Desdouits M, Schaeffer J, Wacrenier C, Oude Munnink BB, Besnard A, Mota Batista F, Stapleton T, Schultz AC, Aarestrup F, Koopmans M, de Graaf M and Le Guyader S, 2022. Application of Next Generation Sequencing on Norovirus-contaminated oyster samples. EFSA supporting publication 2022:EN-7348. 144 pp. doi:10.2903/sp.efsa.2022.EN-7348

**ISSN:** 2397-8325

© European Food Safety Authority, 2022

Reproduction is authorised provided the source is acknowledged.

# Table of contents

Abstract ....... 1
1. Introduction ....... 9
1.1. Norovirus contamination in bivalve molluscan shellfish ....... 9
1.2. Background and Terms of Reference as provided by the requestor ....... 10
1.3 Interpretation of the Terms of Reference ....... 13
2. Data and Methodologies ....... 14
2.1. Data ....... 14
2.1.1. Samples and strains collection ....... 14
2.1.2. Strains used for the test sample sets (TSS) ....... 15
2.1.3. Samples from the European Baseline Survey (BLS) of norovirus in oysters ....... 18
2.1.4. Historical outbreak-related (OB) sample set ....... 22
2.2. Methodologies ....... 24
2.2.1. Overview of the three NGS methods ....... 24
2.2.2. Application of the NGS methods to the different sample sets ....... 27
2.2.3. Preparation of Test sample sets 1 and 2 ....... 29
2.2.4. Metabarcoding ....... 30
2.2.5. VirCapSeq metagenomics ....... 36
2.2.6. Long amplicon sequencing by Oxford Nanopore Technology ....... 41
3. Results ....... 42
3.1. Method comparison on test sample sets ....... 42
3.1.1. Preparation and characterization of TSS ....... 42
3.1.2. Metabarcoding ....... 46
3.1.3. VirCapSeq Metagenomics ....... 59
3.1.4. Long amplicon sequencing by Oxford Nanopore Technology ....... 70
3.2. Metabarcoding on BLS samples ....... 72
3.2.1. Generation of amplicons from samples in the main and reserve lists ....... 72
3.2.2. Molecular characterisation and genetic diversity of NoV strains in BLS samples set ....... 73
3.2.3. Diversity of genotypes and P-types detected in BLS samples and human outbreak samples submitted to NoroNet ....... 84
3.2.4. Concluding remarks for metabarcoding on BLS samples ....... 89
3.3. VirCapSeq Metagenomics on BLS samples ....... 90
3.3.1. Sequencing quality ....... 90
3.3.2. Norovirus detection and diversity by VirCapSeq metagenomics ....... 91
3.3.3. Comparison of metabarcoding and VirCapSeq metagenomics on BLS samples ....... 92
3.3.4. Concluding remarks for VirCapSeq metagenomics on BLS samples ....... 93
3.4. VirCapSeq Metagenomics on OB samples ....... 94
3.4.1. VirCapSeq on OB BMS samples ....... 94
3.4.2. VirCapSeq metagenomics on human stool samples ....... 96
3.4.3. Matching results between BMS and stool samples from the OB sample set ....... 98
4. Discussion and conclusions ....... 99
5. Additional supporting information ....... 107
References ....... 108
Abbreviations ....... 114
Appendix A - Validation of TSS1 libraries for the metabarcoding ....... 115
Appendix B - Preparation and validation of the TSS2 libraries for the metabarcoding ....... 116
B.1 RNA extraction and quantification at Ifremer ....... 116
B.2 Preparation and validation of TSS2 libraries at Ifremer ....... 116
B.3 Preparation and validation of TSS2 libraries at Cefas ....... 116
Appendix C - Preparation and validation of the TSS libraries for the VirCapSeq metagenomics approach ....... 118



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

C.1 VirCapSeq metagenomics on TSS1 samples........118
C.1.1 Preparation and validation of TSS1 libraries at Ifremer........118
C.1.2 Preparation and validation of TSS1 libraries at EMC........118
C.2 VirCapSeq metagenomics on TSS2 samples........119
C.2.1 Preparation and validation of TSS2 libraries at Ifremer........119
C.2.2 Preparation and validation of TSS2 libraries at EMC........121
Appendix D - Library preparation for long amplicon sequencing using Oxford Nanopore Technology on TSS samples........123
D.1 cDNA PCR and long amplicon PCR........123
Appendix E - Preparation of BLS samples for metabarcoding........124
Appendix F – BLS samples and OB shellfish samples library preparation and validation for VirCapSeq metagenomics........125



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

## List of Tables:

Table 1: Summary of the sample sets used in the project ........ 15
Table 2: List of selected NoV strains included in TSS1 and TSS2 ........ 18
Table 3: Representation of sampling periods in the initial selection of BLS samples ........ 20
Table 4: Number of additional BLS samples selected to ensure at least 5 per sampling period ........ 21
Table 5: Representation of sampling periods in the initial selection of BLS samples for metagenomics ........ 22
Table 6. BMS-related outbreak samples selected for the project ........ 24
Table 7: Overview of the three NGS methods proposed and their expected potential before the study ........ 27
Table 8: Methods applied to the different sample sets by the four partners of the consortium ........ 29
Table 9: Primers used for the metabarcoding approach. ........ 33
Table 10: Primers used for the long amplicon sequencing with Oxford Nanopore technology. ........ 42
Table 11: The mean concentrations of TSS1 samples ........ 43
Table 12: Oyster testing before bioaccumulation ........ 43
Table 13: Homogeneity test on contaminated oysters ........ 44
Table 14: Quantification of digestive tissues dilutions and selection of TSS2 samples for NGS ........ 46
Table 15: Number, maximum size and genotyping of NoV sequences (contigs) obtained for each sample of the TSS1 by both laboratories ........ 60
Table 16: Detection of the 10 expected strains in TSS1 in both datasets using 4 different pipelines. ........ 64
Table 17: Number, maximum size and genotyping of NoV sequences (contigs) obtained for each sample of the TSS1 by both laboratories ........ 66
Table 18: Detection of the 10 expected strains in TSS2 in both datasets using 4 different pipelines. ........ 69
Table 19: Genome Detective results of TSS1 and TSS2 samples sequenced with long amplicon sequencing by Oxford Nanopore Technology. ........ 71
Table 20. The BLS amplicons obtained for metabarcoding sequencing ........ 73
Table 21: Number of NoV clusters obtained per assay per NoV genogroup. ........ 74
Table 22: NoV genogroup and RdRp type assignation of RdRp Clusters identified ........ 75
Table 23: NoV genogroup and genotype assignation of VP1 Clusters identified. ........ 76
Table 24: Characteristics and identification of NoV contigs obtained after VirCapSeq metagenomics oF BLS samples. ........ 92
Table 25: Comparison of genotypes / P-types identified by metabarcoding and VirCapSeq metagenomics in BLS samples ........ 93
Table 26: Characteristics and identification of NoV contigs obtained through VirCapSeq metagenomics oF OB-BMS samples. ........ 96
Table 27: Characteristics and identification of NoV contigs obtained through VirCapSeq metagenomics oF OB-Stool samples. ........ 98



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 28: Comparison of NoV contigs obtained through VirCapSeq metagenomics oF OB-Stool and OB-BMS samples. .... 99
Table B 1: Quantification of RNA extracts from TSS2 samples. .... 116
Table C 1: Quantification of RNA extracts from TSS2 samples. .... 119



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

## List of Figures:

Figure 1: GENETIC DIVERSITY OF NOV STRAINS SELECTED FOR TEST SAMPLE SETS TSS1 AND TSS2 (adapted from De Graaf et al, 2016). 18
Figure 2: Schematic view of two NGS approaches, metabarcoding and metagenomics, when applied to the sequencing of a food for the sequencing and identification of a viral contamination. 25
Figure 3: IMPLEMENTATION OF THE NGS METHODS ON THE SAMPLE SETS. 28
Figure 4: RETESTING ALGORITHM FOR BLS SAMPLES. 34
Figure 5: Metagenomic pipeline at Ifremer. 39
Figure 6: METAGENOMIC PIPELINE AT EMC. 40
Figure 7: Mean norovirus concentrations calculated for two extraction methods on bioaccumulated oysters for tss2 preparation 45
Figure 8: Serial dilutions of contaminated digestive tissues (DT) 45
Figure 9: Metabarcoding reads quality and count for TSS1 samples 47
Figure 10: Relative abundancy of clusters identified as GI genotypes in TSS1 using three primer sets. 49
Figure 11: Relative abundancy of clusters identified as GII genotypes in TSS1 using three primer sets. 52
Figure 12: Mean number of raw and reads kept after frogs preprocess (clean reads) obtained for each set of primers by the three laboratories 53
Figure 13: Relative abundancy of NoV GI genotypes identified in TSS2 using three primer sets. 55
Figure 14: Relative abundancy of NoV GII genotypes identified in TSS2 using three primer sets. 57
Figure 15: Number of reads removed from raw reads and clean reads kept for analysis for TSS1 60
Figure 16: Proportion of reads assigned to identified GI or GII P-types and genotypes using ifremer's pipeline in Ifremer (left) and EMC (right) TSS1 data. 62
Figure 17: Number of reads removed from raw reads (red) and clean reads kept for analysis (blue) for TSS2 65
Figure 18: Proportion of reads assigned to identified GI or GII P-types and genotypes using ifremer's pipeline in Ifremer (left) and EMC (right) TSS2 data. 67
Figure 19: Distribution of NoV cluster numbers in BLS samples. 77
Figure 20: Distribution of NoV genotype and P-type numbers in BLS samples. 77
Figure 21: Proportion of sequenced samples from different countries containing different GI RdRp types. 79
Figure 22: Proportion of sequenced samples from different countries containing different GI genotypes. 79
Figure 23: Proportion of sequenced samples from different countries containing different GII RdRp types. 80
Figure 24: Proportion of sequenced samples from different countries containing different GII genotypes. 80



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Figure 25: Proportion of sequenced samples from different periods containing different GI RdRp types. 82
Figure 26: Proportion of sequenced samples from different periods containing different GI genotypes. 82
Figure 27: Proportion of sequenced samples from different periods containing different GII RdRp types. 83
Figure 28: Proportion of sequenced samples from different periods containing different GII genotypes. 83
Figure 29: Proportion of GII P-type and genotype diversity in BLS samples vs Noronet surveillance samples isolated between 2016 and 2018 85
Figure 30: Proportion of GII P-type and genotype diversity in Bls samples vs Noronet surveillance samples isolated between 2016 and 2018. 86
Figure 31: Maximum likelihood tree inferred with (A) GII.2 (246 nt) (B) GII.3 (307 nt) 88
Figure 32: Number of reads removed from raw reads and clean reads kept for analysis for BLS by vircapseq metagenomics. 91
Figure 33: Number of reads removed from raw reads (red) and clean reads kept for analysis (blue) for bls by vircapseq metagenomics. 95
Figure 34: Number of reads removed from raw reads (red) and clean reads kept for analysis (blue) for OB-stool by vircapseq metagenomics. 96
Figure 35: Proportion of reads assigned to identified Genotypes or P-types using EMC pipeline in Ifremer (left) for OB-stool 97
Figure A 1: Electrophoresis of TSS1 libraries in 1% agarose gels 115
Figure B 1: Electrophoresis of TSS2 libraries in 1% agarose gels at IFREMER 116
Figure B 2: Electrophoresis of TSS2 libraries in 1% agarose gels at Cefas 117
Figure C 1: Bioanalyzer profile of TSS1 samples, example for replicate 2 118
Figure C 2: Bioanalyzer profile of library pool I 118
Figure C 3: Bioanalyzer profile of TSS1 samples 119
Figure C 4: Bioanalyzer profiles of TSS2 samples of replicates 2 120
Figure C 5: Bioanalyzer profile of library pool II and III 120
Figure C 6: Bioanalyzer profile of library pools 121
Figure C 7: Bioanalyzer profile of TSS2 samples 121
Figure C 8: Bioanalyzer profile of library pool II and III 122
Figure D 1: ELECTROPHORESIS OF TSS1 AND TSS2 LIBRARIES using long amplicon PCR for Oxford Nanopore Technology sequencing IN 1% AGAROSE GELS. 123
Figure E 1: Electrophoresis of BLS GI NoV VP1 libraries in 1% agarose gels at Cefas 124
Figure F 1: BIOANALYZER PROFILE OF LIBRARY POOLS 125



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

# 1. Introduction

## 1.1. Norovirus contamination in bivalve molluscan shellfish

Contamination of Bivalve Molluscan Shellfish (BMS) with human pathogenic viruses through sewage discharge to the marine environment is recognized as a significant and increasing public health risk affecting the BMS production sector with frequent alerts and recalls. In 2011 EFSA's BIOHAZ Panel concluded that norovirus (NoV) and hepatitis A virus (HAV) in BMS were among the most frequently recognised causes of foodborne illness among all virus / food commodity combinations (EFSA, 2011). The diversity and the low copy numbers of the viral pathogens found in BMS complicate identification, containment and mitigation of foodborne outbreaks (McLeod et al., 2017). Pathogen sequencing is used for confirmation of virus presence and for epidemiological tracking and tracing, but based on partial genome sequences with limited discriminatory power, and without taking the source of contamination into account. **This project addresses these current limits in virus sequencing from food samples by using state-of-the-art next generation sequencing (NGS) methods for the identification and characterization of NoV in BMS and patient samples, both from routine monitoring and surveillance of BMS, and from human stool samples from linked outbreaks in Europe, respectively (Noronet database).** In Europe, the most significant risk to human health is associated with consumption of contaminated oyster, as these are the BMS species most commonly consumed raw (EFSA, 2012). Following these observations, a two year baseline survey of NoV contamination in oysters in the European Economic Area was undertaken (November 2016 - December 2018) (EFSA, 2016). The study targeted oysters collected in harvesting areas and in approved dispatch centers in European countries producing oysters, using quantitative (q, also termed real time) reverse transcriptase (RT) polymerase chain reaction (PCR) (qRT-PCR) for the detection and quantification of the NoV genome (ISO 15216-1). Three research institutes participating in the present project, Ifremer, Cefas and DTU, were involved in the European baseline survey and jointly contributed to more than 70 % of the total number of samples collected and analysed.

Norovirus exhibits tremendous diversity at the genomic level, with recently expanded classification to 10 known genogroups (GI – GX) and 49 genotypes. NoV from genogroups I (GI), II (GII) and (occasionally) IV (GIV), GVII and GIX infect humans. Inter-genotype recombination events frequently occurs at the junction between the polymerase (RdRp) and the capsid genes (VP1), generating a second layer of diversity (Kroneman et al., 2013, De Graaf et al., 2016, Chhabra et al. 2019). Together with the genetic drift, this contributes to the frequent emergence of novel NoV strains in the human population, as evidenced through Noronet, an international norovirus surveillance network coordinated by the consortium partner EMC. Noronet is an informal network of scientists working in public health institutes or universities sharing virological, epidemiological and molecular data on norovirus.

Viruses from different genotypes potentially vary in their ability to cause disease, and therefore proper risk assessment requires the genetic typing of positive samples detected by qRT-PCR. Up until now, the determination of the genotype(s) of NoV implicated in human outbreaks or detected in contaminated foods relied on the amplification, cloning and classical (dideoxy) consensus sequencing of selected segments of their genome, ideally located in RdRp and VP1 genes to account for possible recombination (Kroneman et al., 2013). This approach is specific but is labour- and time-consuming. In addition, BMS can be simultaneously contaminated with multiple strains and these techniques will often miss the low abundance strains, or strains for which the primers are mismatched (Symes et al., 2007, Le Guyader et al., 2008). Next Generation Sequencing methods have been developed to better characterize NoV diversity, especially in environmentally contaminated foods or waters, which are more likely to harbour many viral strains (Lowther et al., 2012; Le Mennec et al., 2017). A key challenge however is the sensitivity needed for implementation of NGS in food safety for NoV detection as NoV-contaminated foods may harbour very low concentrations of viral genomes, near the limit of detection of molecular techniques. The consortium partners have developed a range of approaches, from amplicon-based metabarcoding to virome capture sequencing of NoV or other human pathogens in different types of matrices (oysters, sewage or clinical samples) either individually or through their contribution to the



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

H2020 project COMPARE[2] (Strubbia et al., 2019a, 2019b; Desdouits et al., 2020). **In this work, we apply the most promising NGS methods to oyster samples collected during the European Baseline Survey of norovirus in oysters to evaluate their ability to detect and characterize NoV, and their use for surveillance, outbreak analysis and interpretation of (potential) public health effect**.

## 1.2. Background and Terms of Reference as provided by the requestor

The present Call is based on EFSA's 2018 Work Programme for grants and operational procurements as presented in Annex IX of the EFSA Programming Document 2018 – 2020, available on the EFSA's website[3].

This procurement is part of a series of EFSA activities related to whole genome sequencing (WGS), which include for example the EFSA WGS Umbrella project (under the EFSA Information Management project, IMP) and a self-task mandate of the Panel on Biological Hazards (BIOHAZ) for a scientific opinion on the application of NGS for risk assessment of food-borne pathogens. This activity is a continuation/follow-up of the following EFSA activities: (i) Procurement activity 'Closing gaps for performing a risk assessment on *Listeria monocytogenes* in ready-to-eat (RTE) foods: activity 3, the comparison of isolates from different compartments along the food chain, and from humans using whole genome sequencing (WGS) analysis', (ii) thematic grants on new approaches in identifying and characterizing microbiological hazards (ENGAGE and INNUENDO), (iii) grant 'Comparative genomics of quinolone-resistant *Campylobacter jejuni* of poultry origin from major poultry producing European countries – GENCAMP'. It also relates to the ongoing baseline survey on Norovirus in oysters in the EU.

The major identified transmission route of Norovirus to humans remains the person-to-person contact, followed by food- and waterborne transmission and by the environmental spread. Food-borne transmission can occur either when food handlers (symptomatic or asymptomatic carriers) contaminate food or even earlier during the primary production stages. For instance, shellfish grown along the coastlines can be contaminated by faecal discharge containing Norovirus and fresh produce can be irrigated with sewage-contaminated water.

In the European Union (EU), food-borne infections due to viruses accounted for 9.8% of the total number of reported food- and waterborne outbreaks of gastroenteritis in 2016. Among those viruses, Calicivirus, including Norovirus, caused the highest number of human cases, 11,993 cases, which was 24.0% of all cases caused by all outbreaks and was associated with the highest mean number of cases per outbreak (31.6). In 2016, Calicivirus, including Norovirus, were reported among the top-5 food-pathogen combinations (agent/food vehicle) causing 903 human cases in 6 food-borne outbreaks implicating vegetables and juices as well as 436 human cases in 36 food-borne outbreaks implicating crustacean, shellfish and molluscs[4].

Next generation sequencing has recently emerged as a new tool allowing the characterization of food-borne pathogens including viruses with an unheard precision, thus providing new insights in norovirus diversity and genetic evolution. NGS techniques are rapidly contributing to the understanding of multiple aspects such as the epidemic spread of norovirus, the discovery of emerging genotype variants and pathogenic determinants, and the source identification in food-borne outbreak investigation.

There is an urgent need for EFSA to acquire expertise on the tools available and on their application on food-borne virus molecular epidemiology. Despite the fact that this technology is very promising and is evolving quickly, in the EU there is limited experience with its application in public health. It is still not

---

[2] The COMPARE project was financed by the European Commission H2020 program under grant n° 643476 from 2015 to 2019.

[3] http://www.efsa.europa.eu/en/corporate/pub/amp1820

[4] EFSA (European Food Safety Authority) and ECDC (European Centre for Disease Prevention and Control), 2017. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2016. EFSA Journal 2017;15(12):5077, 228 pp.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

clear when and how this technology will be ready for routine surveillance activities. Thus, “proof of concept” projects are necessary and of high priority.

The recent EFSA Scientific Opinion of the BIOHAZ Panel on the public health risks associated with the hepatitis E virus (HEV) as a food-borne pathogen recommended that there is a need for harmonisation of typing, subtyping and strain comparison methodologies. Thresholds for definition of types, subtypes and ‘identical’ strains based on sequence comparisons should be defined. These recommendations are also applicable to the surveillance of other very important food-borne viruses like Caliciviruses, including noroviruses.

An EU-baseline survey on norovirus in oysters is currently ongoing. Member States (MSs) are collecting and storing viral RNA extracts that could be further characterized within this procurement using NGS. This characterisation would allow the assessment of the circulation of these viruses in Europe and support the investigation of outbreaks due to norovirus linked to oysters’ consumption. The need to proceed rapidly with the characterization of these viruses is related first to the low stability of the samples, and to the fact that MSs have only been asked to store the samples for a minimum of two years after the survey has been completed[5].

The aim of this procurement procedure is to conclude a direct contract for the execution of specific tasks over a clearly defined period as defined in the tender specifications.

The main objective is to make use of NGS to identify and characterise noroviruses from the relevant food sources (e.g. crustaceans, shellfish, molluscs, vegetables, fruits, and their products thereof), the environment, and human cases or asymptomatic carriers.

**EFSA specific objectives (ESO)**

The EFSA specific objectives of the contract resulting from the present procurement procedure are as follows:

**- Objective 1 (ESO1)**

To carry out the molecular characterisation of a selection of norovirus isolates from different sources (i.e. foods, environment and humans) employing NGS analysis.

The selected contractor is expected, as part of the project, to present first a list of the norovirus isolates from different sources, i.e. foods, environment and humans, that are accessible for NGS analysis. The list of norovirus isolates should include an explanation of the selection criteria for the listed isolates. All relevant and available metadata characterising further the list of proposed isolates should also be provided. Efforts should be made to ensure that the stored samples collected during the ongoing EU baseline survey of norovirus in oysters are also included in this procurement activity. This 2-year baseline survey will be concluded in December 2018. The financial decision from the European Commission for this baseline survey proposes to archive all samples collected to facilitate further research e.g. on NoV whole genome sequencing. It also indicates that: i) once the analytical result of samples has been confirmed the digestive glands, supernatant and RNA should be stored at -70°C or below; ii) these samples should be stored at a single location in each country, either the NRL or another laboratory with suitable facilities; iii) the samples should be stored for a minimum of two years after the baseline survey has been completed.

EFSA will facilitate access for the successful contractor to these isolates, however the successful contractor will not be authorised to use these isolates for any purpose outside the remit of this project and will have to destroy the RNA samples after completion of the project. The successful contractor shall treat all data confidentially and Member States will be anonymised in the contractor’s report. An

[5] EC (European Commission), 2016. Grant decision for an action. Sanitary survey for detection of Norovirus in oysters – SI2.738415



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

example of "Agreement for the transfer of Materials" to be signed between the Contractor and the NRLs or other laboratories before the shipment of the isolates can be found in Annex 7 of these tender specifications. All the costs related to the preparation and transport of these samples from the different storage locations to their own premises shall be covered by the contractor.

Further, norovirus isolates from other EU sources should also be included, i.e. from other foods, environment as well as human isolates from the same period of the baseline survey (i.e. 2016-2018). If necessary, the isolate collection can be complemented with other isolates/sequences from recent years. For each isolate relevant metadata should be available in order to fulfil the objective. The selected contractor is responsible for the identification of the origin (e.g. geography, laboratory) of these isolates. Consideration of the sources, including the number of isolates and related metadata, should be made with the overall aim of maximising the outcome of the objectives of the study.

The criteria employed for the selection of the isolates for NGS analysis should be clearly described in the offer. The minimum and/or maximum number of isolates to be included is left to the discretion of the tenderer, with an estimate of this number to be provided in the offer. The total number of isolates proposed should soundly represent isolates from different sources as stated above and humans as well as from various geographical regions in the EU. Both the selection criteria and the number of isolates proposed should be indicated in the offer and applied with the overall aim of maximising the outcome of the objectives of the study.

Although the primary responsibility to list and select norovirus isolates is with the selected contractor, the successful contractor should consult and agree with EFSA on the final selection of the set of isolates to be typed. The NGS should be carried out with state-of-the-art equipment and methodologies which are conform with current laboratory standards and that can be referred to or reported in a clear and concise manner. Robust annotation pipelines for the NGS data generated should be designed and/or implemented with the aim of getting a harmonised framework for subsequent data analysis.

**- Objective 2 (ESO2)**

To explore the genetic diversity and to expand the understanding of the possible patterns of Norovirus diversification within and between the different aforementioned sources.

**- Objective 3 (ESO3)**

To assess the epidemiological relationship between norovirus strains from different sources including the possible different genotype mixture within the same matrix and the potential viral recombination events.

**- Objective 4 (ESO4)**

To perform retrospective analysis of norovirus strains/isolates epidemiologically linked to food-borne outbreaks (if strains/isolates from outbreaks are available) in order to investigate the feasibility of NGS as tool for identification of sources and transmission pathways.

Strains from known norovirus food-borne outbreaks should be characterised employing NGS methods and analysed following the methodological frame employed under objective 1 above. Next, the available NGS data should be analysed for establishing and/or supporting links between the different strains. The outcome of this analysis should provide an evaluation on the advantages and limitations of employing NGS data for investigating outbreaks of food-borne Norovirus.

This contract was awarded by EFSA to:

Contractor: Institut Français de Recherche pour l'Exploitation de la MER (Ifremer), Centre for Environment, Fisheries and Aquaculture Science (Cefas), Technical University of Denmark (DTU) and Erasmus University Medical Center (EMC).

Contract title: Application of NGS (next generation sequencing) on Norovirus



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Contract number: OC/EFSA/BIOCONTAM/2018/01

## 1.3 Interpretation of the Terms of Reference

The consortium decided to focus this project on norovirus-contaminated oyster samples, as NoV-contaminated BMS (especially oysters) have been implicated in a high proportion of European NoV foodborne outbreaks. Indeed, the European Baseline Survey of Norovirus in oysters conducted by EFSA from 2016 to 2018 allowed the selection of oyster samples from all major production areas throughout Europe, a representative source of norovirus-contaminated foods that has no equivalent for other food matrices. Nucleic acid extraction protocols optimized for next-generation sequencing are highly dependent on the sample matrix, and the consortium members have expertise in the analysis of oyster samples. Analysis of this type of food raises specific challenges, due to the low levels of viral contamination. To answer the specific objectives of this tender, additional samples provided by the consortium consist of contaminated oyster samples and samples containing several norovirus genotypes for method optimization as well as NoV outbreak-related oyster samples and corresponding clinical samples from infected humans.

Importantly, NoV are very difficult to cultivate and therefore cannot be isolated or replicated from food or environmental samples (Ettayebi et al., 2016). Thus, the viral material used during this study are not passaged virus isolates. We will use the term "sample" when referring to oyster or stool samples containing both matrix and virus, with possibly different NoV strains in the same sample. As positive controls, we intend to use well-characterized NoV-positive stool samples, for which we determined the presence of a unique NoV genotype, which will be referred to as a "strain".
In previous work, the consortium members observed that successful NoV genome sequencing by NGS is highly dependent on the level of contamination, the quality of extracted nucleic acids, and the matrix itself (Strubbia et al., 2019a, Nooij et al., 2018). The consortium members developed several NGS methods that could be used to address the objectives of this project. Each method has its advantages and limitations and to maximize the outcome of the study the consortium has selected three different, most promising NGS methods, to be applied for NoV sequencing during this project (see section 2.2):

- A metabarcoding approach with deep sequencing of specific regions of the NoV genome, amplified by PCR, on the Illumina platform;
- An Oxford Nanopore sequencing based approach with deep sequencing of long regions of the NoV genome that are amplified by PCR;
- A VirCapSeq metagenomics approach with agnostic deep sequencing of RNA enriched in viral sequences by capture selection on the Illumina platform (Strubbia et al., 2019b, Wylie et al., 2018)

The consortium has also framed four objectives covering the specific objectives of the tender. These objectives are framed taking into account the NGS methods, their applications and the optimization and comparison steps that are needed before they are applicable for environmental and food samples.

**Consortium Objective 1 (CO1):** To perform molecular characterisation by (whole) genome sequencing of a selection of NoV positive samples from different sources, i.e. foods (in this project, BMS batches related to outbreaks), environment (wild and farmed oysters sampled in production areas and oysters batches from dispatch centres collected during EFSA baseline study) and humans (stool samples), using three currently available protocols for NGS analysis. Answering this objective will allow to answer the first and second specific objectives of EFSA (ESO1 and ESO2).

**Consortium Objective 2 (CO2):** To assess and compare the ability of the three NGS methods to characterize genetic diversity of NoVs. This additional objective is required since several methods will be used in parallel. Their respective outcomes and limitations remain to be assessed on norovirus-contaminated samples with low viral levels such as environmental and food samples.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**Consortium Objective 3 (CO3):** To assess how NGS outputs from the different methods can be used to determine the epidemiological relationship between NoV strains from different sources including the possible different genotype mixture within the same matrix and the potential viral recombination events. This objective covers EFSA's specific objective 3 (ESO 3).

**Consortium Objective 4 (CO4):** To perform retrospective analysis of samples from confirmed food-borne outbreaks in order to investigate the feasibility of NGS as tool for identification of sources and transmission pathways. This objective covers EFSA's specific objective 4 (ESO4).

Given the advantages and limits of developed NGS technologies, our consortium proposes to start with a well-controlled comparison of methods with a shared core of test sample sets, as currently there is no single preferred method that can answer the four objectives while maximizing the outcome of the study. **Based on our expertise in NGS applied to NoV and food or environmental contamination, we first optimize and compare three different approaches and then apply them to oyster and human samples to answer the objectives listed above.**

# 2. Data and Methodologies

## 2.1. Data

### 2.1.1. Samples and strains collection

To address the consortium objectives, we selected five sample sets, summarized in Table 1.

For the optimization and comparison of the three different NGS approaches (addressing CO1 and CO2), chosen by the consortium for sequencing norovirus from oysters, the use of artificial samples is necessary to control for the concentration and genetic diversity of NoV strains. Indeed, naturally contaminated oyster samples often harbor several different norovirus strains belonging to different genotypes. Their individual quantities cannot be measured independently since the current method for norovirus detection and quantification is specific for genogroups, but doesn't distinguish between genotypes. We prepared two artificially created **test samples sets (TSS)**: **TSS1** consists of a mix of nucleic acids from previously characterized NoV strains of known genotypes belonging to genogroups I (GI) and II (GII), and **TSS2**, consists of oysters contaminated by bioaccumulation (which mimics the natural contamination) with a subset of these NoV strains. Preparation of these sample sets is described in section 2.1.2 of this report.

As stated in the tender specifications, NoV-positive samples from the EU-wide baseline survey (BLS) on NoV in oysters, conducted between 2016 and 2018, should be included in this project to address EFSA specific objective 1, 2 and 3, and the corresponding consortium objectives 1 and 3 (CO1 and CO3). Therefore, a **third sample set** consists of oyster samples from the **BLS (BLS samples)**. Two hundred samples were selected for NGS analysis in this project based on their positivity for NoV GI and/or GII, the viral concentration, the country of origin, the sampling location and sampling date aiming to maximize both the diversity of samples and the likelihood to obtain NoV sequences. The criteria of sample selection from the BLS on NoV in oysters are specified in section 2.1.3 of the present report.

Another important objective in the tender is to determine the ability of NGS to identify infection sources and transmission pathways (CO3 and CO4). For this, we selected a **fourth sample set** consisting of pairs of human and BMS samples from sixteen confirmed historical food-borne outbreaks that occurred in France and in Denmark from 2012 to 2019, which includes the period during which the EU-wide baseline survey on NoV in oysters took place. The selection criteria for this **outbreak-related (OB) sample set** is presented in section 2.1.4 of the present report.

Moreover, application of NGS approaches on naturally contaminated samples (BLS samples and OB samples) serve to validate the protocols elaborated using TSS sample sets (addressing CO2).

A **fifth sample set**, addressing CO4, consist of human NoV sequences, submitted to the international norovirus surveillance network **Noronet**, is used to compare the diversity of norovirus genotypes in oysters to the diversity of those circulating in the European population during the same period of time.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

This sample set provides information on the number of subtypes detected in the same time period. Information on the database and on the submitting laboratory are described by van Beek et al. (2018).

**TABLE 1: SUMMARY OF THE SAMPLE SETS USED IN THE PROJECT**

| Sample set name | Description | Objectives | NGS technique performed |
|---|---|---|---|
| Test sample set 1 (TSS1) | Mix of nucleic acids from known norovirus GI and GII strains | CO1 and CO2 | Metabarcoding (3 samples)<br>VirCapSeq metagenomics (3 samples)<br>Oxford Nanopore sequencing (3 samples) |
| Test sample set 2 (TSS2) | Digestive tissue from oysters contaminated with a subset of the norovirus GI and GII strains included in TSS1 | CO1 and CO2 | Metabarcoding (6 samples)<br>VirCapSeq metagenomics (6 samples)<br>Oxford Nanopore sequencing (6 samples) |
| Baseline Study samples (BLS) | Digestive tissues, PK-supernatant or nucleic acids extracted from oyster samples collected during the BLS on Norovirus in oysters | CO1 and CO3 | Metabarcoding (212 samples*)<br>VirCapSeq metagenomics (20 samples) |
| Outbreak samples (OB) | Digestive tissues from BMS associated to confirmed norovirus outbreaks and nucleic acids from matched clinical human stool samples | CO3 and CO4 | VirCapSeq metagenomics (30 shellfish and 10 stool samples) |
| Noronet sequences | Human Norovirus sequences deposited on the Noronet platform by Noronet partners | CO3 and CO4 | Not applicable |

* initially, a goal of 200 samples was set, but in the process of library preparation, more samples were used and sent for sequencing, resulting in 212 BLS samples sequenced by metabarcoding.

### 2.1.2. Strains used for the test sample sets (TSS)

To characterise and compare the three different NGS methods, we use sample sets 1 (TSS1) and 2 (TSS2) (addressing CO2). A low limit of detection is needed because food samples often contain low levels of NoV contamination that can still result in food-borne outbreaks in humans. As the results of NoV surveillance can have large implications for the food industry, it is important that the methods are accurate and reproducible. Using artificial samples will enable us to compare the sequencing results with the theoretical initial composition, and to avoid spoiling rare natural samples with limited volumes.

NGS methods using primers to amplify a region of the viral genome may skew the final representation of each strain in a given sample, because primers do not bind to the genome of all strains with the



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

exact same affinity. Two of the three methods we implement here use primers to target NoV. To evaluate the accuracy and reproducibility of the methods, we need test samples with a controlled composition in terms of NoV concentration and proportion of the different strains. Thus, we prepared test sample set 1 (TSS1) by mixing equal quantities (in terms of NoV genome copies) of nucleic acids extracted from selected norovirus-positive human stool samples. This mix is serially diluted to reach concentrations of $10^3$, $10^2$ and $10^1$ genome copies of each norovirus strain per microlitre of extract, that are finally validated by qRT-PCR. Method accuracy is evaluated by comparing the yield in specific NoV sequences compared to total reads, for all NoV strains and for each genotype. All three methods are evaluated by conducting each analysis in triplicate. Laboratory-contaminated oysters are not favoured for this evaluation because different NoV strains may not be bioaccumulated by oysters with the same efficacy, which could add a bias in the sequence outputs that we cannot control (Maalouf et al., 2011).

Based on our preliminary results, we expect that the metabarcoding approach and long amplicon sequencing by Oxford Nanopore are applicable to samples with low concentrations of NoV genomes. The virome capture sequencing metagenomics approach using VirCapSeq selection may only be applicable to samples with intermediate to high concentration but should be the most informative for systematic surveillance as it is expected to capture the broadest diversity of strains. The method sensitivity is likely to be affected by the nature of the sample matrix, since shellfish extracts contain many nucleic acids from the shellfish itself, as well as from bacteria and protists, besides the virome. In addition, the oyster virome comprises not only NoV or other human viruses contaminating the environment, but mostly oyster viruses or bacteriophages. PCR inhibitors are also frequently observed in these complex extracts. Thus, to evaluate the sensitivity of the three NGS methods, we prepared test sample set 2 (TSS2), by oysters' bio-accumulation with selected human stool samples positive for NoV. We focus on the digestive tissues of the oysters, because it is known to concentrate most of the NoV contamination and is used in the standardized method for NoV and hepatitis A virus detection in shellfish (ISO 15216-1 2017). Digestive tissues from the laboratory-contaminated oysters are serially diluted in digestive tissues from oysters confirmed to be non-contaminated by norovirus, to obtain oyster samples with four levels of contamination: high, intermediate, medium-low and very-low. Ifremer has validated this approach in the frame of the European validation of the ISO method (ISO 15216-1 2017) for quantitative detection of NoV in shellfish (Lowther et al., 2019). This allows a realistic estimation of the detection limit one can expect when applying each NGS method on NoV-contaminated shellfish.

To build TSS1 and TSS2, Ifremer and EMC contributed with their bank of NoV strains, in the form of stool samples collected from infected human individuals for diagnostic purposes, of known genogroup and genotype, previously sequenced. TSS1 contains the 12 NoV strains listed in Table 2 of the section 2.1.2.2, namely 8 strains belonging to GII and 4 belonging to GI. TSS2 contains 6 NoV strains, i.e. 4 belonging to GII and 2 belonging to GI, marked with an "X" in the corresponding column of Table 2.

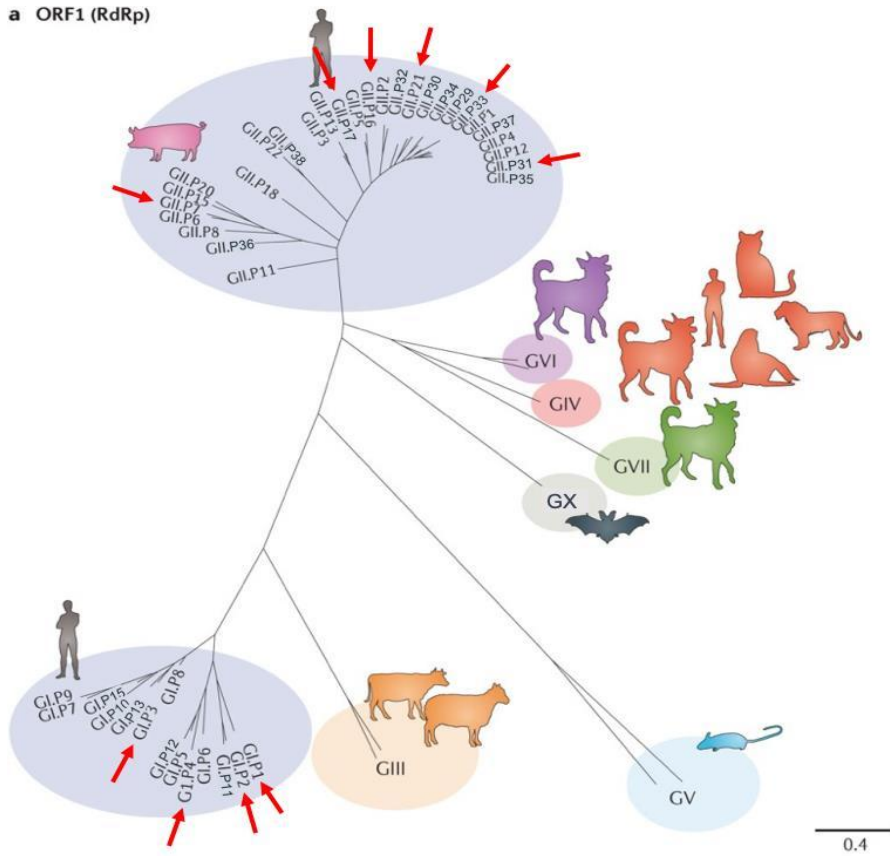
#### 2.1.2.1. Selection criterion for TSS1: Norovirus genetic diversity and frequency

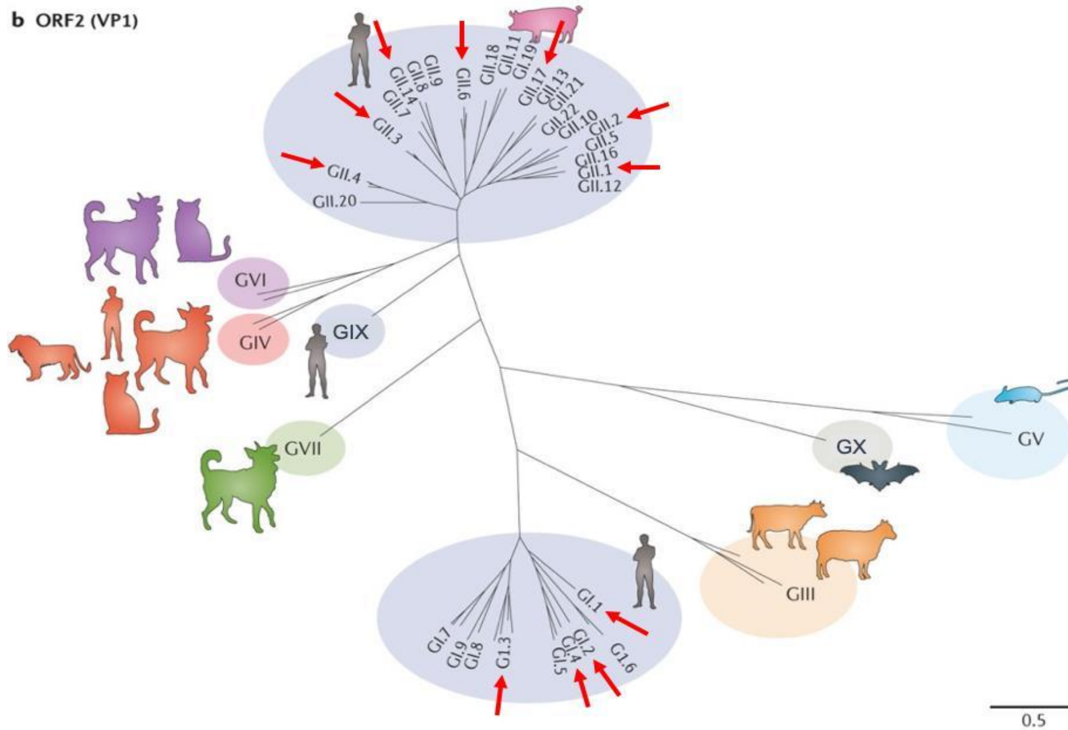
The first criterion to select NoV strains is to maximize the overall genetic diversity while limiting the number of samples used, and focusing on genotypes frequently detected in the human population during the recent years. We selected 8 NoV GII and 4 NoV GI strains, representing 7 different GII genotypes for VP1 and RdRp, and 4 different GI genotypes for VP1 and RdRp, covering phylogenetically distant NoV genotypes (Figure 1**,** red arrows). Strains belonging to NoV GII.4[P16], GII.4[P31] and GII.17[P17] genotypes are selected in priority due to their relative high frequency in the European human population since 2016 based on data collected within the informal surveillance network Noronet[6] (Noronet unpublished data, search date 19-01-2019). The other selected strains were detected in the human population between 2016 and 2018, albeit less frequently.

[6] https://www.erasmusmc.nl/en/research/projects/Noronet
https://www.rivm.nl/en/noronet



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.






The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

### FIGURE 1: GENETIC DIVERSITY OF NOV STRAINS SELECTED FOR TEST SAMPLE SETS TSS1 AND TSS2 (ADAPTED FROM DE GRAAF ET AL, 2016).

Norovirus phylogenetic tree for the RdRp (a) and the VP1 genes (b), showing high numbers of genotypes within each genogroup with their corresponding hosts (humans, swine, dogs, cats, lions, sea lions, rodents, bats, cattle and sheep). Strains selected for the TSS1 and TSS2 belong to the genogroups I and II and the position of their respective RdRp or VP1 genotypes in the phylogenetic trees is marked by a red arrow. Adapted from De Graaf et al, Nature Review Microbiology, 2016.

#### 2.1.2.2. Selection criterion for TSS2: sample quantity

For the preparation of TSS2, in addition to the previous selection criterion, we were limited by the quantity of human stool sample available. Indeed, for artificial contamination of oysters, our in-house protocol uses around $10^7$ NoV genome copies per strain, which ensures that oysters are efficiently and homogeneously contaminated, taking into account the loss of virus binding on oyster shells and tank walls (Maalouf et al., 2011). We selected 4 NoV GII and 2 NoV GI strains that fulfilled this criterion.

**TABLE 2: LIST OF SELECTED NOV STRAINS INCLUDED IN TSS1 AND TSS2**

| Genogroup | Strain name | Sampling year | RdRp P-type | VP1 genotype | TSS 1 | TSS 2 |
|---|---|---|---|---|---|---|
| GII | S582 | 2016 | GII.P17 | GII.17 | X | X |
| | S287 | 2008 | GII.P21 | GII.3 | X | X |
| | S596 | 2016 | GII.P33 | GII.1 | X | |
| | S570 | 2015 | GII.P31 | GII.4 | X | X |
| | S599 | 2017 | GII.P7 | GII.6 | X | |
| | ST19342 | 2018 | GII.P16 | GII.4 | X | |
| | ST28806 | 2018 | GII.P7 | GII.14 | X | |
| | S592 | 2017 | GII.P16 | GII.2 | X | X |
| GI | S505 | 2012 | GI.P4 | GI.4 | X | X |
| | S538 | 2014 | GI.P1 | GI.1 | X | X |
| | S55 | 2003 | GI.P3 | GI.3 | X | |
| | S318 | 2009 | GI.P2 | GI.2 | X | |

### 2.1.3. Samples from the European Baseline Survey (BLS) of norovirus in oysters

#### 2.1.3.1. Background

The European Baseline Survey (BLS) of norovirus in oysters (EFSA, 2016) concluded with the publication of the report on the Analysis of the European baseline survey of norovirus in oysters in July 2019 (EFSA, 2019). From November 2016 to October 2018, samples of oysters (flat oysters *Ostrea edulis*, Pacific oysters *Crassostrea gigas* and Portuguese oysters *Crassostrea angulata*) were collected every two months from sampling locations in two points in the food chain (production areas and dispatch centres) in 13 oyster-producing countries in the EU and EFTA (Ireland, United Kingdom, Norway, Sweden, Denmark, Germany, Netherlands, France, Spain, Portugal, Italy, Croatia, Greece). The samples were



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

tested for norovirus GI and GII using a qRT-PCR method compliant with ISO 15216-1:2017 “Microbiology of the food chain - Horizontal method for determination of hepatitis A virus and norovirus using qRT-PCR -- Part 1: Method for quantification”. Testing laboratories were required to store sample materials (i.e. residual digestive glands, proteinase K supernatant and RNA) for a minimum of two years after the baseline survey had been completed for possible future use in EU research projects. For this project, an initial selection of a main set of 200 BLS samples for metabarcoding NGS analysis was made as described below. In order to provide material for metabarcoding NGS, it was necessary to reamplify the selected samples using conventional nested RT-PCR. It was anticipated that at this stage some samples would give negative results and therefore no material would be available for metabarcoding NGS. In order to allow for metabarcoding NGS analysis on the full target number of at least 200 samples, a reserve set comprising an additional 100 samples was also selected.

#### 2.1.3.2. Selection principles for the main set (200 samples)

The following principles were applied to the selection process (in the order of priority given here):

i) Only samples that were positive for NoV GI and/or GII were considered for analysis

ii) Samples with the highest levels of norovirus (GI + GII combined) were preferred (provided that the principles iii) to vi) were followed)

iii) A minimum of 5 samples per sampling country were selected (or all positive samples if <5 were available for a specific sampling country)

iv) A minimum of 5 samples per two-month sampling period were selected

v) A maximum of 5 samples per sampling site (dispatch centre or production area) were selected

vi) A maximum of 50 samples per sampling country were selected.

The metabarcoding RT-PCR used in this project is not as sensitive as the qRT-PCR used for the analysis of the samples during the BLS. In addition, the extra freeze-thaw steps required for carrying out the metabarcoding will tend to decrease the viral load in selected samples. Therefore, it could not be assumed that all samples that tested positive during the BLS would also test positive at the reamplification stage, while negative samples from the BLS were highly unlikely to test positive by the metabarcoding RT-PCR. Principles i) and ii) ensured that the metabarcoding RT-PCR was targeted as much as possible at the samples that were most likely to provide a positive result and subsequent sequence data using the metabarcoding NGS. Principles iii) and iv) ensured a wide geographical and temporal spread amongst the targeted samples whilst principles v) and vi) ensured the sample selection was not unduly biased by targeting large numbers of samples from a single geographical origin, either at the country or site level.

Overall, out of a total of 379 sampling sites (207 dispatch centres and 172 production areas) included in the BLS, samples from 235 sampling sites (91 dispatch centres and 144 production areas) provided at least one positive result. Since there were more sites with positive results (235) than samples to be selected (200), it was not possible to establish an additional criterion with a minimum requirement for the number of samples per site.

The selection principles described in this section were not designed to produce a sample set that is statistically representative of the complete BLS sample set, but aimed to produce a set that was most likely to provide positive results using the metabarcoding RT-PCR, whilst ensuring samples from a broad variety of geographical and temporal origins were included.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

#### 2.1.3.3. Step-by-step description of the selection process for the main set (200 samples)

i) Selection was carried out using a set of 4309 samples (2129 samples from dispatch centres and 2180 from production areas), identical to the set of samples used for quantitative analysis in the report of the BLS (EFSA, 2019). During analysis of the BLS, this sample set had been generated by removing all samples collected outside the survey period (November 2016 to October 2018), all untested samples, all samples giving "not valid" results for one or more genogroups and all samples tested using diluted RNA for one or more genogroups from the entire set of all samples with metadata submitted to EFSA during the BLS.

ii) All negative samples i.e. those that gave "not detected" results for both genogroups, were removed from the set, leaving only 1057 samples that gave positive results for one or both genogroups (226 samples from dispatch centres and 831 production area samples).

iii) For each sampling country the 5 positive samples with the highest norovirus levels (based on the sum of GI + GII in copies/g) were selected for the main set. Out of 13 sampling countries, in one case no positive samples were found (therefore no samples from this country were included in the selection), while in another case only 4 positive samples were found (all 4 samples were selected at this stage). This resulted in a selection of 59 samples (11 countries x 5 samples, 1 country x 4 samples, 1 country x 0 samples).

iv) The 59 selected samples were tabulated according to the sampling period as shown in Table 3:

**Table 3: Representation of sampling periods in the initial selection of BLS samples**

| Year | Period | Number of samples selected |
|---|---|---|
| 2016 | Nov-Dec | 5 |
| 2017 | Jan-Feb | 14 |
| 2017 | Mar-Apr | 4 |
| 2017 | May-Jun | 1 |
| 2017 | Jul-Aug | 0 |
| 2017 | Sep-Oct | 2 |
| 2017 | Nov-Dec | 5 |
| 2018 | Jan-Feb | 13 |
| 2018 | Mar-Apr | 11 |
| 2018 | May-Jun | 1 |
| 2018 | Jul-Aug | 0 |
| 2018 | Sep-Oct | 3 |
| **TOTAL** | | **59** |

v) For those sampling periods with fewer than 5 samples in the selection (highlighted in red in Table 3), additional samples were selected targeting the highest contamination levels to increase the number of selected samples in these sampling periods up to 5 as follows (



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

vi) Table 4):



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**TABLE 4: NUMBER OF ADDITIONAL BLS SAMPLES SELECTED TO ENSURE AT LEAST 5 PER SAMPLING PERIOD**

| Year | Period | Number of additional samples selected |
|---|---|---|
| 2016 | Nov-Dec | - |
| 2017 | Jan-Feb | - |
| 2017 | Mar-Apr | 1 |
| 2017 | May-Jun | 4 |
| 2017 | Jul-Aug | 5 |
| 2017 | Sep-Oct | 3 |
| 2017 | Nov-Dec | - |
| 2018 | Jan-Feb | - |
| 2018 | Mar-Apr | - |
| 2018 | May-Jun | 4 |
| 2018 | Jul-Aug | 5 |
| 2018 | Sep-Oct | 2 |
| **TOTAL** | | **24** |

vii) The set of 83 selected samples (59 + 24) was then screened to check whether more than 5 samples from any given sampling site (production area or dispatch centre) had been selected. One case was detected with 6 samples from a single sampling site. In this case the least contaminated sample from the set of 6 was removed from the selection. The removed sample was then replaced in the selection by the next most contaminated sample from the same two-month sampling period. This sample was checked to ensure it did not originate from a sampling site from which 5 samples had already been selected.

viii) At this stage it was not necessary to carry out a specific check to identify if more than 50 samples had been selected per country; given the addition of 5 samples per sampling country at stage iii) followed by the addition of 24 samples in total at stage v) it was not possible for any sampling country to have contributed more than 29 samples at this stage to the selection for the main set of samples.

ix) The set of 974 non-selected positive samples (1057 – 83) was then ranked according to their GI + GII combined contamination level. Starting with the most contaminated sample, these were checked to see if their addition to the main set would result in either more than 5 samples from an individual sampling site, or more than 50 samples from a sampling country. In the case that either limit would be exceeded, the sample was removed from consideration; if neither limit would be exceeded the sample was added to the main set. This process was repeated with each successively less contaminated sample until a total of 117 samples had been added to complete the main set of 200 samples (83 positive samples selected in steps iii) to vi) plus 117 newly selected samples). This selection procedure produced a main set with a range of norovirus concentrations (GI + GII) from 13 to 24,566 copies/g.

### 2.1.3.4. Selection of the reserve set (100 samples)

Given that the reserve set was to be used only in cases where samples from the main set provided negative results with the metabarcoding RT-PCR, the overriding principles for its selection were those related to norovirus positivity and combined GI + GII contamination levels (i.e. principles i) and ii) as described in 2.1.3.2). The other principles (iii), iv), v) and vi)) related to maximum and minimum numbers of samples per country, site or sampling period were not applied; this meant that after replacement of samples from the main set providing negative results with samples from the reserve set,



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

it was possible that the minima and maxima outlined in these principles would no longer be met. The reserve set is therefore comprised of a selection of the 100 samples with the highest norovirus contamination levels from the set of 857 positive samples that were not selected for the main set (1057 - 200). The reserve set includes samples with a range of norovirus concentrations (GI + GII) from 327 to 9,075 copies/g.

#### 2.1.3.5. Selection criteria of BLS samples for metagenomics (20 samples)

The following principles were applied to the selection process of the BLS samples for metagenomics analysis:

i. Norovirus GI and/or GII positive samples
ii. High concentrations (GI and GII combined) and 4 samples (20%) at low concentrations (>200 gc/g of digestive tissue, DT, per genotype)
iii. Successful analysis by metabarcoding (sequences obtained)
iv. At least one aliquot of DT available as the re-extraction is required (extraction method for metagenomics is not the ISO method, as specified in section 2.2.5.1)
v. Samples from different sampling countries / sampling sites / sampling periods

Twenty-six samples were preselected in case of extraction failure or negative cDNA synthesis results to assure 20 BLS samples for metagenomics analysis. The preselected samples, collected in 8 different countries, were tabulated according to the sampling period as shown in Table 5.

**TABLE 5: REPRESENTATION OF SAMPLING PERIODS IN THE INITIAL SELECTION OF BLS SAMPLES FOR METAGENOMICS**

| Year | Period | Number of samples selected | |
|---|---|---|---|
| | | Main list | Reserve list |
| 2016 | Nov-Dec | 2 | 1 |
| 2017 | Jan-Feb | 4 | 1 |
| 2017 | Sep-Oct | 2 | 0 |
| 2017 | Nov-Dec | 0 | 1 |
| 2018 | Jan-Feb | 7 | 2 |
| 2018 | Mar-Apr | 3 | 1 |
| 2018 | Jul-Aug | 1 | 0 |
| 2018 | Sep-Oct | 1 | 0 |
| **TOTAL** | | **20** | **6** |

### 2.1.4. Historical outbreak-related (OB) sample set

The samples from norovirus outbreaks associated to BMS as food vehicle, included the clinical human stool samples from ill consumers and samples from the BMS batch linked to the outbreak and were selected according to the criteria listed below.

#### 2.1.4.1. Selection criterion OB 1: sample availability and NoV positivity both in humans and BMS

The first selection criterion was to consider countries participating to the baseline survey that could provide both epidemiological and clinical human data, as well as the associated patient's stool samples and implicated BMS samples from outbreaks. Only Denmark and France were able to provide all these details with matching patient stool and BMS samples both positive for NoV.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

#### 2.1.4.2. Selection criterion OB 2: period of time

To allow the comparison of NoV strains in these samples and those of baseline survey samples, outbreaks that occurred during the same time period as the baseline survey (Nov. 2016- Oct. 2018) were given priority. However, due to an insufficient amount of Danish outbreak material within this period and to allow assessing the performance of the sequencing methods on a greater variety of norovirus strains, outbreaks occurring back to January 2012 were included.

In France, over the baseline survey period, the Ifremer laboratory was informed of a total of 28 shellfish-borne NoV outbreaks. Overall, a total of 38 oyster samples from these 28 outbreaks were received for analysis. Among these, five were found under the detection limit for both genogroups. NoV GI was detected in two samples, NoV GII in 10 samples and both GI and GII in 21 samples. Among these 28 outbreaks, only 10 had epidemiological data available together with BMS and stool samples. Nine outbreaks were linked with oyster consumption and one with mussels (*M. edulis*). For the outbreak linked to the mussels as a food vehicle, an oyster sample from the same production area as the mussels leftover was also sampled at the restaurant.

In Denmark, only two oyster-borne NoV outbreaks occurring during the baseline survey period had matched oyster and patient stool samples available in addition to epidemiological data. Beside these, nine outbreaks occurring in the period from January 2012 to January 2016 had epidemiological data available together with oyster and stool samples. Among the remaining outbreaks occurring in Denmark since 2012, epidemiological data and oyster material were present, but no stool samples.

#### 2.1.4.3. Selection criterion OB 3: link between outbreak and BMS sample

BMS samples related to the confirmed outbreaks can be (1) left-overs from consumed meals, and thus directly implicated in the outbreak, (2) BMS belonging to the same batch implicated in the outbreaks, (3) BMS from a different batch of the same origin coming from the same producer, or (4) BMS coming from the same production area. This additional criterion aimed to prioritise available BMS samples based on their closeness to the batch implicated in the outbreaks. When available, a sample from the corresponding production area of the suspected implicated BMS was also considered and included in the analysis to compare NoV strains detected in production area and in the food implicated in the outbreak.

In total, according to the selection criteria and sufficient quantity of the available digestive tissues (at least 2g), thirty shellfish samples and ten stool samples implicated in sixteen outbreaks were available for this project, as shown in



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 6.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**TABLE 6.** BMS-RELATED OUTBREAK SAMPLES SELECTED FOR THE PROJECT

| Outbreak code | Outbreak date | Shellfish species * | Link with batch implicated in outbreak | Shellfish sample code | Stool sample code |
|---|---|---|---|---|---|
| OB-001 | 03/2016 | *C. gigas* | Restaurant, batch same origin | OB001_1_BMS | OB001_1_S<br>OB001_1_S |
| OB-002 | 01/2016 | *C. gigas* | Same production area | OB002_3_BMS | OB002_1_S<br>OB002_2_S |
| | | *C. gigas* | Restaurant, batch same origin | OB002_2_BMS | |
| | | *M. edulis* | Leftover | OB002_1_BMS | |
| OB-003 | 02/2016 | *C. gigas* | Same production area | OB003_1_BMS | No stool sample |
| OB-004 | 03/2016 | *C. gigas* | Same producer, same origin | OB004_1_BMS | OB004_1_S |
| | | *C. gigas* | Same production area | OB004_3_BMS | |
| | | *C. gigas* | Same production area | OB004_2_BMS | |
| OB-005 | 03/2016 | *C. gigas* | Same producer, same origin | OB005_1_BMS | OB005_1_S<br>OB005_2_S |
| | | *C. gigas* | Leftover | OB005_2_BMS | |
| OB-006 | 12/2017 | *C. gigas* | Leftover | OB006_1_BMS | No stool sample |
| OB-007 | 11/2017 | *C. gigas* | Leftover | OB007_1_BMS | No stool sample |
| | | *C. gigas* | Leftover | OB007_2_BMS | |
| OB-008 | 03/2016 | *C. gigas* | Leftover | OB008_1_BMS | OB008_1_S |
| OB-009 | 01/2016 | *C. gigas* | Leftover | OB009_1_BMS | No stool sample |
| OB-010 | 03/2014 | *C. gigas* | Restaurant | OB010_1_BMS | No stool sample |
| | | *C. gigas* | Restaurant | OB010_2_BMS | |
| OB-011 | 12/2012 | *C. gigas* | Leftover | OB011_1_BMS | No stool sample |
| | | *C. gigas* | Leftover | OB011_2_BMS | |
| | | *C. gigas* | Restaurant, 3 days older batch | OB011_3_BMS | |
| OB-012 | 02/2012 | *C. gigas* | Restaurant | OB012_1_BMS | No stool sample |
| | | *C. gigas* | Restaurant | OB012_2_BMS | |
| OB-013 | 02/2012 | *C. gigas* | Restaurant | OB013_1_BMS | No stool sample |
| | | *C. gigas* | Restaurant | OB013_2_BMS | |
| | | *C. gigas* | Restaurant | OB013_3_BMS | |
| OB-014 | 12/2015 | *C. gigas* | Leftover | OB014_1_BMS | OB014_1_S<br>OB014_2_S |
| | | | Same production area | OB014_2_BMS | |
| OB-015 | 03/2016 | *C. gigas* | Same producer, same origin | OB015_1_BMS | No stool sample |
| | | *C. gigas* | Same producing area | OB015_2_BMS | |
| OB-016 | 11/2016 | *C. gigas* | Restaurant, 4 day older batch | OB016_1_BMS | No stool sample |

** C. gigas – Crassostrea gigas; M. edulis - Mytilus edulis*

## 2.2. Methodologies

### 2.2.1. Overview of the three NGS methods

In order to fulfil the CO1 objective of the study, i.e. to use NGS to sequence NoV in oysters, three different NGS approaches developed by the consortium partners were employed. They belong to two different approaches, as presented in Figure 2, namely Metabarcoding and Metagenomics. This section provides an overview of the methods used and the rationale for choosing them.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

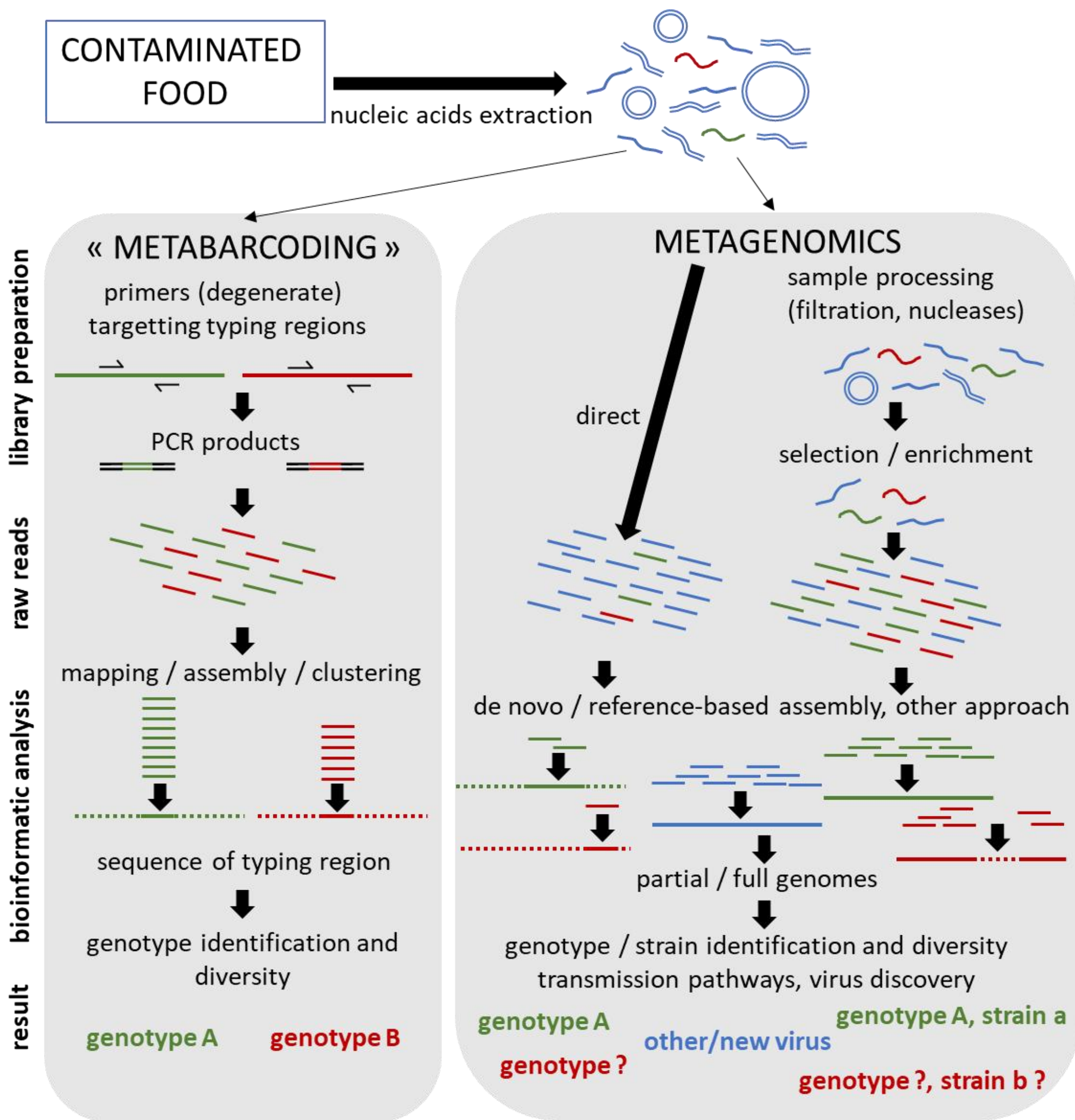



**FIGURE 2: SCHEMATIC VIEW OF TWO NGS APPROACHES, METABARCODING AND METAGENOMICS, WHEN APPLIED TO THE SEQUENCING OF A FOOD FOR THE SEQUENCING AND IDENTIFICATION OF A VIRAL CONTAMINATION.**

The different approaches comprise on the same key steps (left), ie. preparation of the nucleic acid library from the sample, generation of raw reads, bioinformatic analysis, and interpretation of results. In the case of metabarcoding, a selected, short region of the viral genome is targeted using degenerate primers and the resulting PCR products are deep-sequenced, which allows the identification of the viral species and possibly its genotype. In the case of metagenomics, the whole nucleic acid extract can be sequenced, or the sample can be prepared to select or enrich in viral nucleic acids, which are then deep sequenced and analysed to reconstruct viral genomes. These can be full genomes or partial ones depending on the initial amount of viral contaminant and the sequencing depth, and allow for the identification of this virus from the species to the strain if the sequence is long enough. Adapted from Desdouits et al, One Health Outlook, 2020.

The first method, a metabarcoding approach using the Illumina technology, has been developed to study the diversity of microbial communities (bacteria, archea or eukaryotes). For NoV, this method



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

takes advantage of the well-established primers used for NoV genotyping, that target specific, well-studied regions of the viral genome, namely portions of the genes coding for the polymerase (RdRp) and the capsid protein (VP1) (Imamura et al., 2017, Lun et al., 2018, Oshiki et al., 2018). Then, the amplified genome fragments, which may arise from different viruses present in the initial sample, are sequenced using the Illumina technology. Comparison of these sequences to established databases (NCBI, Norovirus Typing Tool 2.0) allows characterization of the diversity of NoV genotypes in the sample. Importantly, this approach can be applied to nucleic acids extracted from shellfish following the ISO 15216 method (ISO 15216-1 2017) for NoV detection, as already demonstrated by studies performed by Ifremer (Desdouits et al., 2020). This approach is easy to scale-up, affordable, and user-friendly pipelines are available for the bioinformatics analysis of Illumina reads (Escudié et al., 2018). However, this technique has some limitations. First, the use of primers and PCR amplification can introduce some bias, amplifying only already known sequences, and skewing the final proportion of each genotype, which remains a relative estimation, not an absolute one. More importantly, the genomic segments are short (200-500 bp), due to the constraints on primer selections and the limited size of Illumina reads (up to 300 nt). Such partial genomic information is enough to identify the genotypes of viral strains but may not provide enough resolution to differentiate between strains from the same genotype, or to identify accurately sources of contamination and transmission pathways. Finally, this approach, as previously set up by the consortium partners, gives access to both RdRp and VP1 sequences, but obtained separately, and thus cannot allow the identification of viral recombinants, which requires sequencing the full genome or long segments spanning both regions of the genome (Lun et al., 2018). To circumvent this problem, new primers were designed for the purpose of this project, enabling amplification of a longer fragment (400-500 nt) comprising a part of RdRp and a part of VP1 gene sequences (



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 9:) and thus allowing the typing of both genes in the same amplicon (dual typing).These primers were validated on NoV-contaminated stool and shellfish extracts, and are compared to previously existing primers in the frame of this project, to choose the most relevant approach. This metabarcoding method is proposed to respond to CO 1, CO 2 and partially to CO 3 of the present project.

The second method applied in the project is the VirCapSeq-VERT approach, which combines agnostic sequencing with a specific capture assay for vertebrate viruses, using the Illumina technology (Briese et al., 2015). The option to enrich for vertebrate viruses with the capture-based method showed a higher recovery of NoV genome sequences from naturally contaminated oysters (Strubbia et al., 2019b). This metagenomics approach offers the great advantage of unbiaised detection of NoV, including the detection of new NoV recombinants. Data can be analysed using several bioinformatic approaches, such as custom scripts or Genome Detective (https://www.genomedetective.com/) (for more details on bioinformatic analysis used in this study please refer to subsection 2.2.5.5). This approach is proposed to answer the CO objectives 1, 2, 3 and 4.

The third NGS method is another metabarcoding approach based on long amplicon sequencing which uses the Oxford Nanopore Technologies (



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 9:). This newest generation of sequencing technology allows the sequencing of long nucleotide fragments up to 500 kb long. This technology combines advantages from both short metabarcoding and metagenomics approaches. Amplicon sequencing allows sequencing of viral genome targets from samples with low levels of contamination. The use of long fragments allows identification of NoV strains and recombinants (the latter when analysing reads overlapping the ORF1-ORF2 junction region sequence which is the hotspot for recombination). Sequences are analysed using reference-based mapping. However, long amplicons may be difficult to generate in samples with very low concentrations as in naturally-contaminated shellfish. Besides, as for the metabarcoding approach with shorter segments, PCR bias may arise. If we are able to generate sufficient NoV reads and coverage, long amplicon sequencing by Oxford Nanopore technologies (ONT) may be suitable to answer the CO 1, CO2, CO3 and CO 4.

An overview of the methods applied in the present project is shown in the table below (Table 7).

### Table 7: Overview of the three NGS methods proposed and their expected potential before the study

| | Metabarcoding | VirCapSeq Metagenomics | Long Amplicon Sequencing by ONT |
|---|---|---|---|
| NoV specific PCR step | Yes | No | Yes |
| Sequencing technology | Illumina MiSeq | Illumina MiSeq | Oxford Nanopore Technologies (ONT) |
| Sample preparation | Simple preparation, RT-PCR amplification, Suitable for low concentrated samples | No NoV specific primers used, sample concentration and purification needed | Simple preparation and RT-PCR amplification |
| Advantages | PCR sensitivity, Targets typing regions | Full genome possible, Detection of all vertebrate viruses, New strain discovery, Virome identification | PCR sensitivity, Longer fragment possible (1000 bp), Targets RdRp and VP1 typing regions together |
| Disadvantages | PCR selection, Only targeted genomes, Recombination cannot be assessed, Short genomic information (<500bp) | Low contamination difficult to detect, Concentration and enrichment steps needed | PCR selection, Only targeted genomes, Higher level of sequencing errors with the flow cells used in this study (R9), Difficulty to amplify long fragments |
| Estimated cost per sample* | €20 | €500** | €50 |
| Estimated investment costs (equipment)* | €200,000 | €200,000 | €5,000 |
| Time to result | 5 days | 8 days | 1 day |
| Suitable for surveillance*** | Yes | Possible, but limited by the cost and time to result | Yes |
| Suitable for outbreak sequencing*** | No – sequenced region may be too short to allow identification of strains and recombinants | Possible, but limited by cost and time to result | Yes |



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

| Suitable to identify transmission pathways*** | No – sequenced region may be too short to allow identification of strains, PCR induce biases | Yes | Maybe-but dependent on coverage. Sequence errors might complicate identification of transmission pathways |
|---|---|---|---|

* Estimated costs can be laboratory and/or country dependent; sequencing can be performed by private companies.
** Current cost estimate, may be lowered if VirCapSeq metagenomics of oyster samples is optimized.
*** Suitability for each application considering all characteristics listed above (advantages, disadvantages, cost, time).

The methods for each NGS approach are described in detail in the following sections, as well as the procedure used for the preparation of the test samples sets.

### 2.2.2. Application of the NGS methods to the different sample sets

The consortium has chosen three different NGS methods to answer the COs of this tender. These methods were applied for the sequencing of NoV in four different sample sets: the test sample sets 1 and 2 (TSS1 and TSS2), the sample set 3 (BLS) comprising 212 shellfish samples from the European BLS between November 2016 and October 2018, and the sample set 4 (OB), comprising human stool and shellfish samples related to sixteen confirmed historical shellfish-borne NoV outbreaks. This section describes how and why each NGS method is applied on each sample set (Figure 3).

First, the test sample sets were sequenced using the three different methods. Read coverage, length of contigs, genome coverage and the amount of NoV genotypes identified were considered to assess each method's specificity, sensitivity, and suitability for enabling different types of analyses (genotyping, NoV diversity, full genome sequencing, transmission pathways, recombination). This allows the comparison of the three different methods.

To maximise the outcome of this study, the consortium decided to apply the metabarcoding approach on selected samples from the BLS. Indeed, this method has already been tested on shellfish extracts where it allowed the identification of diverse NoV GII genotypes (Desdouits et al., 2020). This allowed to identify NoV genotypes present in shellfish throughout Europe at the time of the BLS, and estimate the diversity of NoV in such samples.

Yet, given the limitations of this approach (described in section 2.2.1), the consortium proposed to apply another NGS method, either VirCapSeq metagenomics or Long amplicon sequencing by Oxford Nanopore Technologies, to attempt analysis of transmission pathways and recombination events. The choice between VirCapSeq metagenomics or Long amplicon sequencing by Oxford Nanopore Technologies depended on the method comparison on TSS1 and TSS2, and especially on their respective sensitivity in relation to the low levels of NoV concentration in most BLS samples. For technical (very low target concentrations) and financial reasons, these expensive methods could not be applied on all BLS samples. Eventually, the VirCapSeq metagenomics approach was applied to sequence the outbreak related samples (OB sample set) as well as 20 samples from the BLS set, chosen based on the level of norovirus contamination.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

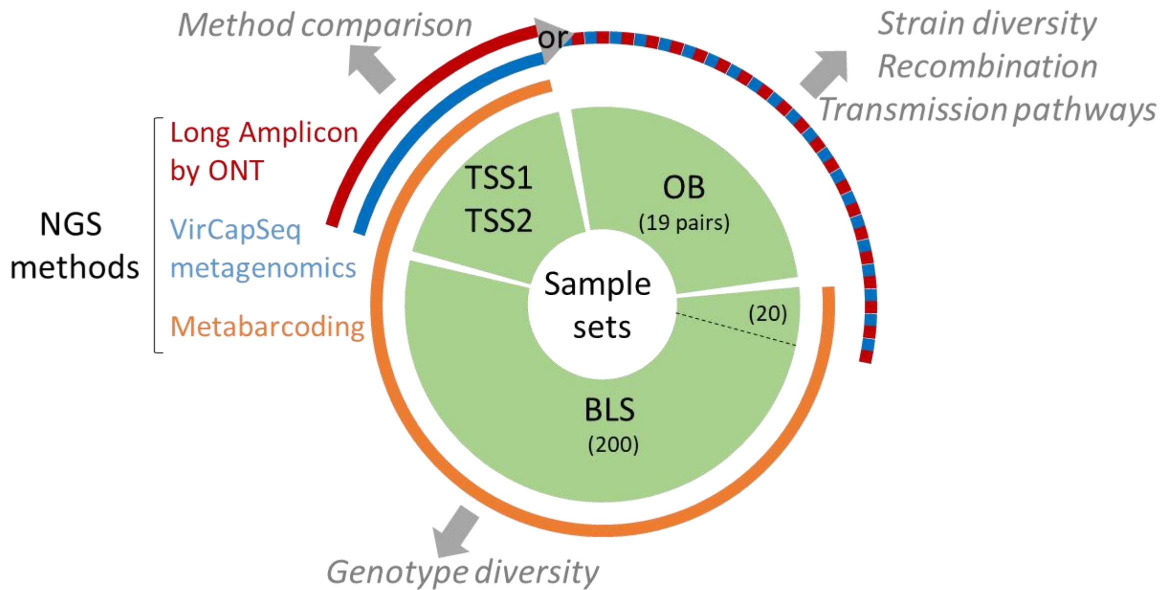



**FIGURE 3: IMPLEMENTATION OF THE NGS METHODS ON THE SAMPLE SETS.**

The consortium shared the work between partners according to Table 78, taking advantage of the specific expertise of each partner regarding the different methods and matrices analysed.

**TABLE 8: METHODS APPLIED TO THE DIFFERENT SAMPLE SETS BY THE FOUR PARTNERS OF THE CONSORTIUM**

| Sample set | Method | Ifremer | EMC | Cefas | DTU |
|---|---|---|---|---|---|
| **TSS1** | Metabarcoding | x (3)* | x (3) | | |
| | Metagenomics | x (3) | x (3) | | |
| | Long amplicon ONT | | x (3) | | |
| **TSS2** | Metabarcoding | x (6) | x (6) | x (6) | x (6) |
| | Metagenomics | x (6) | x (6) | | |
| | Long amplicon ONT | | x (6) | | |
| **BLS** | Metabarcoding | x (102) | | x (110) | |
| | Metagenomics | x (20) | | | |
| **OB - shellfish** | Metagenomics | x (20) | | | |
| **OB - stool** | Metagenomics | | x (10) | | |

*Number of samples analysed by each consortium partner

### 2.2.3. Preparation of Test sample sets 1 and 2

As described previously, two sets of test samples were prepared for validation of three NGS methods: the first one containing nucleic acids of different NoV strains extracted from NoV-positive stools (TSS1) and the second one constituted of digestive tissues (DT) of oysters contaminated with several NoV strains by bioaccumulation (TSS2). This section describes the methods used to prepare these samples.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

#### 2.2.3.1. Test sample set 1 (TSS1)

The selection criteria of NoV strains chosen for TSS1 and TSS2 are described in details in a chapter 2.1.2.1. The nucleic acids were extracted from selected NoV-positive stool samples from the consortium's collection, each containing one NoV strain of known genotype GI or GII. Firstly, stools samples were weighted and suspended in PBS at a final concentration of 10% (m/v), mixed by vortexing for 30 s and homogenised using a Fast-Prep (MP Biologicals) for 20s. Nucleic acids were extracted from 100 µL of stool suspension using NucliSens Magnetic Extraction Reagents (bioMérieux, France) according to manufacturer's instructions. The extracted nucleic acids were quantified pure, 10-fold and 100-fold diluted in molecular biology grade water, each in triplicates by one-step qRT-PCR according to ISO 15216-1 (ISO 15216-1, 2017), using Stratagene Mx3000Pro and Agilent AriaMx thermocyclers. The TSS1 was obtained by mixing the nucleic acids of each genotype in Tris-EDTA (TE) buffer to a final concentration of $10^3$ genome copies/µL for each initial strain (TSS1.1 sample). This equimolar mix was then serially diluted 10-fold in TE buffer to obtain the final concentrations of $10^2$ genome copies/µL (TSS1.2 sample) and 10 genome copies/µL (TSS1.3 sample). The final GI and GII NoV concentration of three TSS1 samples was quantified in triplicates by qRT-PCR according to ISO 15216-1 (ISO 15216-1, 2017), using Agilent Mx3000P and AriaMx thermocyclers, to verify that the target concentration was reached. In addition, NoV GI and GII concentrations were confirmed by one-step digital RT-PCR (dRT-PCR) as described in Desdouits et al. (2020), a method allowing the absolute quantification of target RNA independently from standards. Once the target concentration was validated, 8µL aliquots were prepared from all TSS1 samples and stored at -70°C or below before analysis by three NGS methods.

#### 2.2.3.2. Test sample set 2 (TSS2)

This test sample set consisted of DT from laboratory contaminated oysters by bioaccumulation, to test the performance of the used NGS approaches in oysters. Upon arrival in the laboratory, the oysters (*C. gigas*) were analysed according to ISO 15216-1 method (ISO 15216-1, 2017) to verify the absence of initial NoV contamination as well as the absence of inhibitors of enzymatic reactions in nucleic acids extracts from digestive tissues. The oysters were then kept at 4°C±2°C before being laboratory-contaminated by bioaccumulation to mimic the natural contamination process. Six norovirus strains (4 GII strains and 2 GI strains, listed in Table 2) were selected for bioaccumulation in oysters, as described above (chapter 2.1.2.2). Firstly, 10% suspensions of selected human stool samples were prepared in PBS, nucleic acids extracted and the concentration of NoV genome quantified, as described for the TSS1 (chapter 2.2.1.1). The cocktail for bioaccumulation was prepared by mixing 10% stool suspensions in PBS to obtain a final quantity of $10^7$ genome copies of each selected strain, and kept at 4°C ±2°C. An aquarium, placed in a cold room at a constant temperature of 12°C, was filled in with a sand-filtered sea water and left for temperature adjusting. The cocktail was added to the water and mixed to homogenize. Oysters (100 animals) were submersed and left for bioaccumulation during 24 hours with aeration, but no feeding.

After 24h, the oysters were taken out and kept at 4°C ±2°C. Ten animals awee randomly sampled, dissected and analysed following the ISO 15216-1 method to verify the bioaccumulation efficiency, using triplicates of pure nucleic acids extracted from digestive tissues. Once the NoV bioaccumulation was validated (at least $10^4$ genome copies/g DT), the remaining oysters were shucked, their digestive tissues (DT) dissected, pooled and homogenised using a FastPrep (MP Biological) and vortexing. In parallel, digestive tissues from another batch of NoV-free oysters were also dissected and homogenised. NoV-contaminated DT homogenate was serially diluted in NoV-negative DT to obtain four levels of NoV concentration: high (TSS2.1, no dilution), medium (TSS2.2, dilution 1:4), medium-low (TSS2.3, dilution 1:16) and very low (TSS2.4, dilution 1:32), close to the LOD of the quantification method). An additional sample was prepared DT used for the serial dilution (TSS2.5) as negative control since NoV were not detected in this batch. Finally, a background control was also included with non-contaminated oysters



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

from the same batch than the bioaccumulated ones (TSS2.0). Each diluted DT homogenate was split into 2g aliquots and stored at -70°C or below.

The quality of prepared samples and norovirus concentration wa verified on triplicates by one-step RT-qPCR according to ISO 15216-1. The three NGS methods were validated on TSS2 containing high (TSS2.1), medium (TSS2.2), medium-low (TSS2.3), very low (TSS2.4) NoV concentrations, background control (TSS2.0) and negative (TSS2.5) concentrations of NoV, to assess their LOD while taking into account the effect of the shellfish matrix. Thus, nucleic acids were extracted from TSS2 following different protocols optimized for each approach (see below).

### 2.2.4. Metabarcoding

The metabarcoding approach was applied on the TSS1, TSS2 and BLS oyster samples (Figure 3). This section describes the preparation of samples for metabarcoding, the generation of nucleic acid libraries, the sequencing and bioinformatics analysis leading to NoV genotyping. TSS1 was sequenced by Ifremer only, while TSS2 was sequenced in parallel by Ifremer, DTU and Cefas to test the repeatability of the method applied to shellfish samples, in different laboratories (Table 8). The 200 main list BLS samples were split between Ifremer (102 samples) and Cefas (98 samples) for metabarcoding sequencing (Table 8). The three laboratories followed the same method with minor differences restricted to final tagging and indexing of the amplicons.

#### 2.2.4.1. Nucleic acid extraction

All nucleic acids (NA) were extracted from samples using the NucliSens reagents (bioMérieux), except for the TSS1, which were already in the form of nucleic acids. Nucleic acids were extracted from BMS samples from TSS2 and BLS samples following the ISO 15216-1 method (ISO 15216-1, 2017). For this, 2g aliquots of DT were incubated in 2ml of proteinase K solution (100µg/ml) for 1 h at 37°C and then for 15 min. at 60°C. The sample was then centrifuged at 3000g for 5 min at room temperature, the total volume of supernatant was recorded and 0.5ml used for nucleic acid extraction (NucliSens, bioMérieux). For the BLS samples, the remaining original RNA generated during the BLS study, stored below -70°C, was used for the present study when possible. When the remaining RNA volume was not sufficient, the ISO 15216-1 method (2017) was used for extraction from remaining proteinase K supernatant, or from extra-DT samples, also stored at -70°C or below, in this order of priority, depending on availability of each material, for each sample.

#### 2.2.4.2. RT-PCR for metabarcoding

The RT-PCR method for the metabarcoding analysis consisted of 3 stages: reverse transcription (RT), first round PCR (referred below as PCR) and second round, semi-nested PCR (referred to as N-PCR).

For reverse transcription, a mix of random hexamers (Invitrogen, France) and nonamers (Sigma, France), both at 2.5 µM of final concentration were used as primers with the SuperScript II (Invitrogen, France) reverse transcriptase. The reaction was carried out in a total volume of 20µL, containing 7 µL of nucleic acids extract, according to the manufacturer's instructions, using a PTC-100 (MJ Research Ink), SimpliAmp (Applied Biosystems), Mastercycler Nexus X2 (Eppendorf) at Ifremer, Cefas and DTU. For samples for which several viruses and genes were targeted, duplicate reverse transcription reactions were necessary to generate enough complementary DNA (cDNA) for subsequent analysis. The efficiency of RT was controlled by quantitative PCR (qPCR) on the cDNA, using the same primers, probes and controls as described in ISO 15216-1 (2017), using the Taq polymerase instead of the RT-DNA polymerase enzyme mix and deleting the RT step from the thermal profile. The amplifications were performed using Stratagene Mx3000Pro, Mx3005Pro and Agilent AriaMx real-time thermocyclers.

Following reverse transcription, 10 µL of cDNA was used for each PCR. Then, to enhance specificity and sensitivity of the method, the N-PCR used 5 µL of the first round PCR product. For each amplification reaction mix, the Platinum Taq kit (Invitrogen, France) was used following the manufacturer's instruction, with a final volume of 50 µL, $MgCl_2$ (final concentration 1.5mM), and forward and reverse



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

primers at a final concentration of 0.2pM. Both amplification reactions (PCR and N-PCR) followed the same program, including 40 amplification cycles with an annealing temperature at 50°C. Conventional amplifications were run in the thermocyclers as specified above.

Each time a PCR mix was prepared, a negative control was added to the series of samples in the form of molecular grade water instead of nucleic acid extracts. In addition, for each N-PCR, the product of the first PCR on the negative control was also used as matrix. This generates, for each succession of PCR and N-PCR, two negative controls: one that was submitted to both rounds of PCR (2 PCR negative control), and one that was only generated during the semi-nested PCR (N-PCR negative control). These controls were verified together with PCR products by gel electrophoresis to ensure that there is no contamination during PCRs. The "2 PCR negative control" was added to the series of tagged amplicons to be sequenced. Finally, for the BLS samples (set 3), two "negative extract controls" were included in the library, in the form of nucleic acid extracted from a NoV-negative oyster DT sample submitted to exactly the same steps as the positive samples, at the same time. A control with water instead of nucleic acid matrix was also included.

Various sets of forward and reverse primers were used depending on the targeted NoV genogroup and genomic region (



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 9:). For each targeted genogroup (GI and GII), we have chosen or designed primers allowing either the separate amplification of RdRp and VP1 sequences, or the amplification of a longer segment of the genome comprising RdRp and VP1 junction. To allow the indexing of PCR products for multiplexed deep sequencing, the semi-nested PCR primers were tagged with an adapter sequence at 5’, specific of the sequencing platform used, and not indicated here. Following semi-nested amplification, N-PCR mixes were analysed by electrophoresis in 1% agarose gel to verify the presence of an amplicon with the expected size.

The sensitivity of two approaches (separate RdRp and VP1 amplification, or RdRp-VP1 segment) was tested on TSS 1 and 2 and according to the results, the separate RdRp and VP1 amplifications were applied on BLS samples as this approach was more sensitive. BLS samples were analysed for NoV GI, GII or both genogroups depending on the baseline survey results.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**TABLE 9: PRIMERS USED FOR THE METABARCODING APPROACH.**

| NoV genogroup | Target gene | PCR + direction | | Primer name | Primer sequence | Amplicon size | Reference |
|---|---|---|---|---|---|---|---|
| GI | VP1 | PCR | F | QNIF4 | CGCTGGATGCGNTTCCAT | 381 bp | (da Silva *et al.*, 2007) |
| | | N-PCR | F | GISKF | CTGCCCGAATTYGTAAATGA | 330 bp | (Kojima *et al.*, 2002) |
| | | PCR & N-PCR | R | GISKR | CCAACCCARCCATTRTACA | | (Kojima *et al.*, 2002) |
| | RdRp | PCR | F | NV4478 | AAATTGCCHATHAAAGTTGGNATG | 407 bp | *New design* |
| | | N-PCR | F | NV4562 | GATGCDGATTACACAGCHTGGG | 323 bp | (Yuen *et al.*, 2001) |
| | | PCR & N-PCR | R | P110 | ACDATYTCATCATCACCATA | | (Le Guyader *et al.*, 1996) |
| | RdRp-VP1 | PCR | F | Mon432 | TGGACICGYGGICCYAAYCA | 579 bp | (Beuret *et al.*, 2002) |
| | | N-PCR | F | GI_5196 | GIGARGCHICHCTNCAYGGNGA | 494 bp | *New design* |
| | | PCR & N-PCR | R | GISKR | CCAACCCARCCATTRTACA | | (Kojima *et al.*, 2002) |
| GII | VP1 | PCR | F | QNIF2D | ATGTTCAGRTGGATGAGRTTCTCWGA | 378 bp | (Loisy *et al.*, 2005) |
| | | N-PCR | F | GIISKF | CNTGGGAGGGCGATCGCAA | 344 bp | (Kojima *et al.*, 2002) |
| | | PCR & N-PCR | R | GIISKR | CCRCCNGCATRHCCRTTRTACAT | | (Kojima *et al.*, 2002) |
| | RdRp | PCR | F | NV4611 | CWGCAGCMCTDGAAATCATGG | 274 bp | (Yuen *et al.*, 2001) |
| | | N-PCR | F | NV4692 | GTGTGRTKGATGTGGGTGACTT | 193 bp | (Yuen *et al.*, 2001) |
| | | PCR & N-PCR | R | P110 | ACDATYTCATCATCACCATA | | (Le Guyader *et al.*, 1996) |
| | RdRp-VP1 | PCR | F | Mon431m | TGGACIAGRGGICCYAAYCA | 570 bp | (Beuret, *et al.*,2002) |
| | | N-PCR | F | NV4963 | GARHTNAARGAAGGTGGNATGGA | 426 bp | *New design* |
| | | PCR & N-PCR | R | GIISKR | CCRCCNGCATRHCCRTTRTACAT | | (Kojima *et al.*, 2002) |

Amplicon sizes are given for each forward primer (F) in combination with the reverse primer (R) used for the target sequence, which is the same during the first round PCR (PCR) and the second round semi-nested PCR (N-



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

PCR). Amplicon sizes for semi-nested products do not take the adapter sequences into account, as it may vary depending on the platform used for sequencing.

### 2.2.4.3. Retesting algorithm for BLS samples

The original plan for testing BLS samples devised before the start of the practical work was that the 200 main list samples selected according to 2.1.3.2 would be subject to testing using the metabarcoding RT-PCR as described in 2.2.4.2. In the case where a sample did not yield a detectable cDNA or N-PCR product of the correct size starting on RNA matrix (see 2.2.4.2), repeat testing was carried out. For this, RNA was re-extracted using stored PK supernatant or DT, following the ISO 15216-1 method (2017). The re-extracted RNA was analysed using the metabarcoding RT-PCR. If at this stage no cDNA was obtained (as determined by qPCR) or no N-PCR product was detected, the sample would be abandoned and a sample from the reserve list of the BLS sample set would be used instead. In this way the original target of submitting amplicons from 200 samples for sequencing would be preserved. The planned retesting algorithm is presented in Figure 4.

However, for reasons of time, at Cefas repeat testing of samples from the main list after initial negative results and testing of samples from the reserve list was carried out simultaneously instead of in sequence. This resulted in an excess of samples with positive amplicons available for sequencing from Cefas, (110 samples instead of 98, 212 samples instead of 200 across both labs, further details in 3.2.1). To maximise the value of testing it was decided to submit amplicons from all 212 available samples for sequencing, instead of limiting sequencing analysis to the original goal of 200 samples.

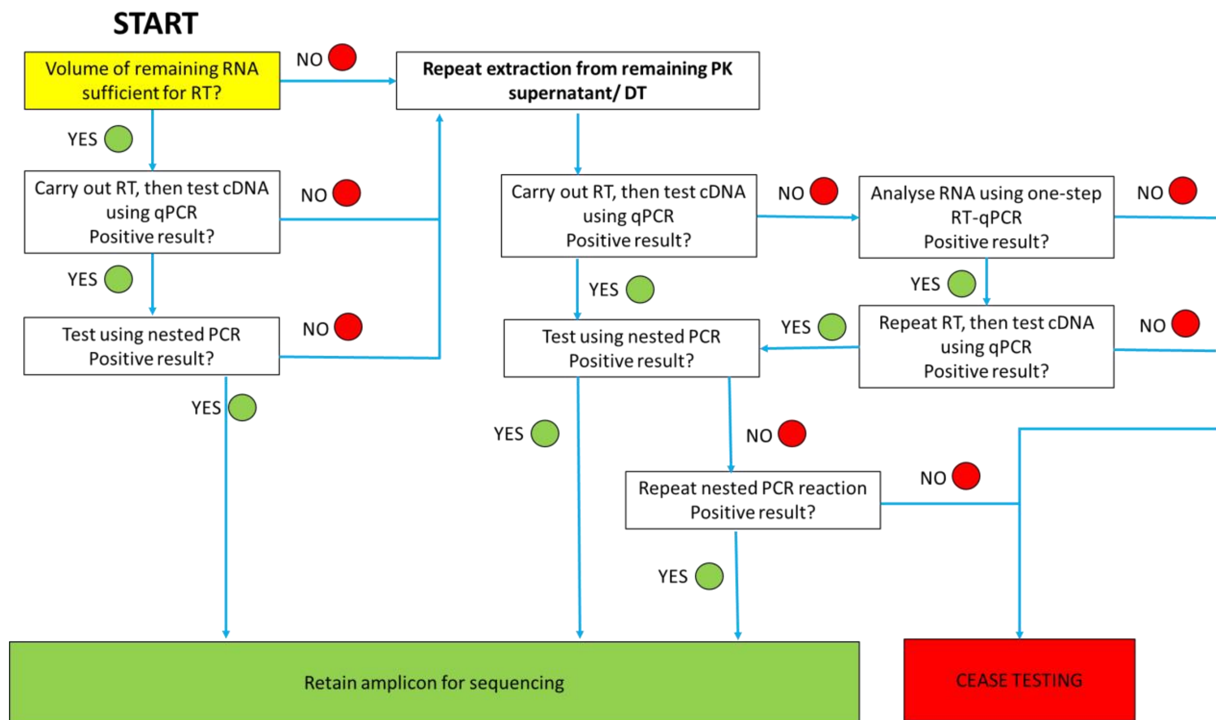



**FIGURE 4: RETESTING ALGORITHM FOR BLS SAMPLES.**

NOTE: BLS samples were only tested for those genogroups for which they tested positive during the BLS e.g. If a sample tested positive for GI but negative for GII during the BLS, in this project only qPCR and nested PCR assays for GI were applied. For samples that tested positive for both GI and GII during the BLS, qPCR and nested PCR for both GI and GII were applied in this project. The algorithm was followed separately for each of the assays required (for each genogroup and genome region amplified), however RNA volume checks took into account all required assays.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

#### 2.2.4.4. Sequencing and Bioinformatic analysis

Tagged PCR amplicons were indexed by an additional PCR, purified, quantified and mixed into a pool with equimolar quantities of each amplicon by the subcontractant selected to perfom the sequencing (Institut du Cerveau et la Moëlle, Paris, France). Sequencing was performed on the MiSeq platform (Illumina) to generate paired-end reads of 300nt (2x300).

Raw reads were demultiplexed directly by the Illumina MiSeq sequencer. Reads quality was verified using FastQC (V.0.11.8 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ Andrew, 2010) and MultiQC (V.1.9 https://multiqc.info/ Ewels et al., 2010) tools. For the GII-RdRp PCR reads, we only kept the 200bp first bases (as the expected amplicon size is 192pb) using Trim Galore (V.0.6.4 https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ Krueger, 2012). Reads wre merged, size-filtered and clustered by the Swarm algorithm using the FROGS suite (V.2.0.0 http://frogs.toulouse.inra.fr/ Escudié et al., 2018). FROGS pre-process parameters were a mismatch of 0.3, a swarm value of 2 and the denoising option. Chimera were removed from the generated clusters using FROGS Chimera tool. In addition, a manual detection of chimera was done based on cluster alignment and abundance. Non-chimeric clusters were filtered to discard those comprising less than 0.1% of reads using the FROGS Filter tool. Indeed, clusters with a very low abundance were numerous and they often represented chimeras of two different NoV genotypes and P-types that were not filtered out by the FROGS chimera removal filter. Thus, filtering out minority clusters ensured that most chimeras were removed while keeping most of the sequenced information.

Remaining clusters were identified using the Norovirus Typing Tool (V.2.0 https://www.rivm.nl/mpf/typingtool/norovirus/ Kroneman et al., 2011). When the Norovirus Typing Tool assigned an unexpected genogroup (for instance, GIII using a GI assay), or failed to assign a genotype, other tools were used to confirm or complete the assignation. BLASTn (V.2.6.0 https://blast.ncbi.nlm.nih.gov Camacho et al, 2009) was used on the NCBI nr/nt database, and the closest match was genotyped using the Norovirus Typing Tool to assign a genotype or P-type if possible, or to confirm the genogroup assignation.

To verify we did not miss important genotypes with FROGS, reads were aligned on NoV reference sequences from the Noronet database (downloaded the 26-03-2020), belonging to each known genotype, using Bowtie2 with default scoring options (sensitive -D 15 -R 2 -N 0 -L 22 -i S,1,1.15) (V.2.3.0 http://bowtie-bio.sourceforge.net/bowtie2/index.shtml Langmead and Salzberg, 2012) and the proportion of reads mapping to each genotype was used to quantify the genotype abundancy. The NoV genotype diversity and abundancy were then analysed using R (V.4.0.2 https://www.R-project.org/ R core Team, 2020) with a custom script (https://gitlab.ifremer.fr/lsem/r-script-moonstone).

All genotypes with less than 100 reads were removed to ensure reliability of the results. In addition, we used the negative control (water) of each run to establish the minimal number of NoV reads needed to consider a sample positive. Thus, water controls always appeared negative.

For TSS1 and TSS2, thresholds were different from zero for GII-RdRp (CEFAS, 200 reads; DTU 2000 reads), GI-VP1 (DTU, 10000 reads) and GI-RdRp (DTU, 500 reads).

For the BLS sample set, thresholds were set at 130 reads for both Ifremer and Cefas. In addition, clusters comprising less than 0,1% of total reads, but more than 130 reads in at least one sample, and that were assigned to a genotype or P-type not identified among the more abundant clusters, were retrieved and kept for analysis, since they could not be chimeras and added a valuable information on viral diversity. Finally, cluster sequences generated at Ifremer and Cefas were compared and aligned using AliView to identify identical sequences in both datasets and keep only unique sequences for the subsequent analyses.

#### 2.2.4.5. Controls and replicates.

The TSS1 and TSS 2, used for method validation, were tested in triplicates, starting at cDNA synthesis step for the TSS1 and three independent extractions for TSS2.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

For the BLS samples, one replicate was analysed using separate amplification of NoV GI and/or GII RdRp and VP1 amplicons from the same nucleic acid extract.

## 2.2.5. VirCapSeq metagenomics

The metagenomics approach using VirCapSeq-VERT capture was applied on: (i) the TSS 1, (ii) TSS 2, (iii) 20 samples selected among the BLS samples submitted to metabarcoding and (iv) matched stool and bivalve molluscan shellfish (BMS, including oysters and one mussel sample) samples linked to outbreaks (OB sample set) (Figure 3). Prior to nucleic acid extraction, samples must be prepared to increase the recovery of NoV genomes that are vastly outnumbered by genetic material of bacteria and host (human) or matrix (BMS). The following subsections describe this sample preparation for both BMS and human stool samples, the construction of the libraries using the VirCapSeq-VERT capture probes, the sequencing and bioinformatics analysis leading to identification of NoV genotypes.

### 2.2.5.1. Nucleic acid extraction from BMS samples

BMS samples from the TSS2, BLS and OB sample sets were stored at -70°C or below in the form of dissected, homogenised pools of DT. For each sample, 2g of DT were incubated with 2 mL of proteinase K for 15 min at 37°C and then 15 min at 60°C. After sonication (Bandelin UD 2200 with cup-horn adaptor) for 3x1 min at maximum power, the mixture was centrifugated for 5 min at 3,000g. Then the supernatant was mixed with 10mM sodium pyrophosphate (Bisseux et al., 2018) and incubated for 40 min at room temperature under gentle agitation. After 20 min of centrifugation at 8,000g and 4°C, the supernatant was recovered and 1.5mL of a PEG 24% (wt/vol)-sodium chloride (1.2 M) solution was added before incubating for 1 h at 4°C. After centrifugation for 20 min. at 11,000g, the pellet was resuspended in 1mL of glycine buffer (0.05 M) pH 9 and filtered through 5, 1.2 and 0.45 µM filters (Minisart NML 17594, NML17593, PES16533). Then the filtrate was treated with 10µL/mL OmniCleave™ Endonuclease (Lucigen Corporation) and 100µL $MgCl_2$ (100mM) for 1h at 37°C. Nucleic acids extraction was performed using the NucliSens Extraction Reagents (bioMérieux, France) according to the manufacturer's instructions to obtain 100µL of nucleic acids in elution buffer. As some inhibitors may persist and to remove endonucleases, an additional RNA purification was performed using the RNA Clean & Concentrator TM-5 kit (Zymo Research, Irvine USA) (Strubbia et al., 2019b). NoV RNA extraction efficiency was controlled by qRT-PCR using Stratagene Mx3000Pro, Mx3005Pro and Agilent AriaMx thermocyclers.

### 2.2.5.2. Nucleic acid extraction from stool samples

Stool samples from OB set were stored at -70°C or below. A clarified 10% (w/v) faecal suspension was prepared in phosphate-buffered saline, centrifuged and the supernatant was filtered through a 0.45-µm filter (Millipore). The viral-particle-containing filtrates was digested with a mixture of DNases and RNases (Strubbia et al. 2020). Viral RNA was extracted using the High Pure RNA isolation kit (Roche). The extraction efficiency of NoV RNA is controlled by RT-qPCR using the 7500 real time PCR system (AB Applied Biosystem).

### 2.2.5.3. Library preparation

Most steps of the library preparation were similar between Ifremer and EMC, but several details are specific of each laboratory due to their respective equipment. Both versions of the library preparation were applied on the same samples sets (TSS1 and TSS2) starting from RNA, by the two teams, which allowed to verify how these differences may affect the results, and validate the repeatability of the method in different laboratories with different equipments and constraints. In addition, the Ifremer version of the library preparation was applied to the 20 samples from the BLS sample set, and to the BMS samples from the OB sample set. The EMC version of the library preparation was applied to the stool samples from the OB sample set.

***At Ifremer***



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Random hexamers and the SuperScript II reverse-transcription kit (Invitrogen) were used to synthetize cDNA from 15µL of purified RNA. The second DNA strand was obtained using the reaction buffer from NEBNext Ultra RNA Library prep (New England Biolabs), followed by a physical fragmentation (Ultrasonicator M220, Covaris). Next steps of the library preparation (end repair, adaptor ligation and cleaning using AMPure beads) were performed using the KAPA Hyper Prep kit (Roche), according to the manufacturer's instructions. At the end of the process, a pre-capture PCR (15 cycles) was performed to reach the minimum recommended DNA concentration (5 ng/µL per library). VirCapSeq hybridization to enrich libraries in viral sequences were performed using the probes and protocol designed and published by Briese and colleagues (Briese et al., 2015) on pools of maximum 10 libraries, using the HyperCap Target Enrichment Kit. Finally, a post-capture PCR was performed to obtain a minimal recommended DNA concentration of 10 ng/µL. PCRs were done using a PTC-100 (MJ Research Ink) or a SimpliAmp (Applied Biosystems) thermocycler.

***At EMC***

Random hexamers and the SuperScript IV reverse-transcription kit (Invitrogen) were used to synthetize cDNA from 10 µL of purified RNA. The second DNA strand was obtained using Klenow Fragment (New England Biolabs). Next steps of the library preparation (end repair, adapter ligation and cleaning using AMPure beads) were performed using the KAPA Hyper Prep kit (Roche), according to the manufacturer's instructions, with the exception that the adapters were 1:10 diluted and the Post-Ligation Cleanup step was done twice. At the end of the process, a pre-capture PCR (24 cycles) was performed using a C1000 Thermal Cycler (Bio-Rad) to reach the minimum recommended DNA concentration (5 ng/µL per library). VirCapSeq hybridization to enrich libraries in viral sequences was performed using the probes and protocol designed and published by Briese and colleagues (Briese et al., 2015) on pools of maximum 6 libraries, using the HyperCap Target Enrichment Kit. Finally, a post-capture PCR was performed to obtain a minimal recommended DNA concentration of 10 ng/µL.

#### 2.2.5.4. Library quality check

During library preparation several quality checking were performed. The presence of NoV sequences in the cDNA was evaluated by qPCR (Le Guyader et al. 2009), using Stratagene Mx3000Pro and Agilent AriaMx thermocyclers. If Ct values were higher than the expected Ct based on the qRT-PCR results obtained post-RNA extraction, cDNA synthesis was repeated. After the two PCR (pre- and post-capture), the DNA concentration was quantified by Qubit (dS DNA Br Kit, Thermo Fisher Scientific). If the minimum quantity recommended for the next step (5ng/µl and 10ng/µl respectively) was not reached, the PCR was repeated with an adjusted number of cycles. Negative controls, which were expected to contain low amounts of DNA, were treated in the same way as positive samples, without adding more PCR cycles, even if they did not reach these minimal quantities. The size distribution of libraries before and after capture was verified using a Bioanalyser (Agilent), and should be between 150 and 500 bp with a maximum peak close to 320 bp. The distribution may be skewed by the presence of adaptor dimers (size < 127 bp), which can be removed by repeating the cleanup procedure with AMPure. The size of DNA fragments may also be too long in which case the library was prepared again with an increased time of fragmentation.

#### 2.2.5.5. Sequencing and Bioinformatic analysis

For TSS1 and TSS2, the sequencing and bioinformatics analysis were performed independently by Ifremer and EMC, on the library each team produced. This allowed to compare the repeatability of the method in different laboratories, from RNA extract to final analysis. In addition, data from TSS1, TSS2, BLS and OB sample sets generated by both laboratories were analysed by the bioinformatic pipeline set up by Ifremer, to account for possible differences due to sequencing or to the bioinformatics pipeline.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

### ***Sequencing and bioinformatics at Ifremer***

For TSS1 and TSS2, sequencing was performed on a NovaSeq 2X250bp 800 million reads to generate paired-end reads of 250 pb.

For BLS and OB, sequencing was performed on a NextSeq500 2X150bp 800 million reads to generate paired-end reads of 150 pb.

The bioinformatics pipeline used by Ifremer is presented in Figure 5 **Error! Reference source not found.**. Raw reads were quality-controlled with fastQC and filtered using Trim Galore with a quality threshold at 25 and remaining adapters from Illumina sequencing removed. A negative filtering was made by mapping the reads, using Bowtie 2 a reference sequence database containing the genomes *C. gigas* and closely related species, to remove sequences matching the shellfish matrix. Then, reads were assembled into contigs using SPAdes with the option --meta (metaSPAdes, V3.14.0, Nurk et al. 2017) and the following kmer sizes : -k 21,33,55. Assembled contigs were size-filtered (500 or 150 bp size threshold) and then identified with BLASTn on the ncbi nr/nt database (database uploaded 22/10/2021) keeping only the best hit and evalue at 0.00001. The taxonomy identifier provided by BLASTn was used to identify NoV contigs. These NoV sequences were genotyped using the Norovirus Typing Tool (V.2.0 https://www.rivm.nl/mpf/typingtool/norovirus/ Kroneman et al., 2011). For TSS1 and 2, a mapping of clean reads with Bowtie 2 with sensitive option (default) on this filtered assembly was also done to re-estimate the coverage and calculate the abundance of reads used to assemble the contigs. This abundance was then plotted using R.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

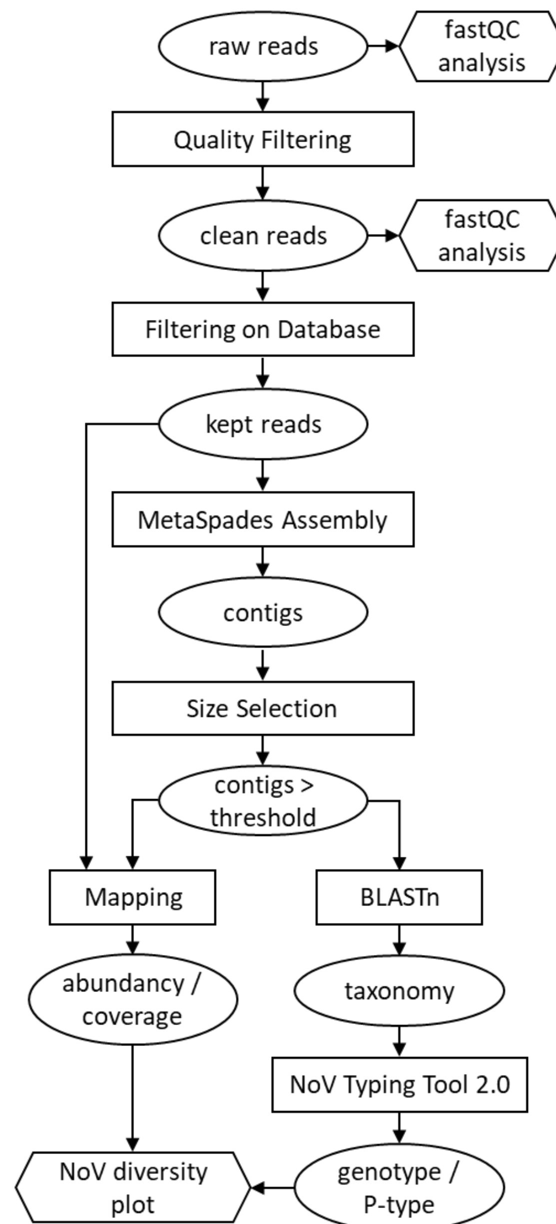



**FIGURE 5: METAGENOMIC PIPELINE AT IFREMER.**

Ellipse represent the data files, rectangles are processing steps, and hexahedras are outputs or verifications.

### *Sequencing and bioinformatics at EMC*

Sequencing was performed using Illumina MiSeq system (2x300bp) for TSS1 and 2, and for the stool samples from the OB set.

Sequenced reads were used for adapter removal and both ends of the reads were trimmed using a sliding window of 4bp and a quality score of 20 using Trimmomatic (V.0.6.4). *De novo* assembly was performed using SPAdes v3.11.1 and later by metaSPAdes (v3.13.1) both with standard settings. Both SPAdes and metaSPAdes were tested for comparison. The resulting contigs were used for BLASTn (v2.9.0), Boratyn et al., 2019) against a local database containing all downloaded NCBI Nov sequences (databases' date of download: 24-03-2020). Contigs with a minimal length of 500 bp and an identity score of ≥85% with NoV were selected for further analysis. Manual inspection was done through



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

alignment against a final BLASTn reference (first hit) using MAFFT (v7.407) and sequences were further analysed using mapping by Bowtie 2 (v2.3.4.3). Then, NoV sequences were genotyped using the Norovirus Typing Tool (V.2.0 https://www.rivm.nl/mpf/typingtool/norovirus/ Kroneman et al., 2011). All contigs were aligned using muscle (https://www.ebi.ac.uk/Tools/msa/muscle/) or MAFFT version 7 (https://mafft.cbrc.jp/alignment/server/) with reference strains to check the quality of the contigs (Chhabra P et al. 2019). In parallel samples were analysed through Genome detective https://www.genomedetective.com online versions 1.125-1.137. The bioinformatics pipeline used by EMC is presented in Figure 6.

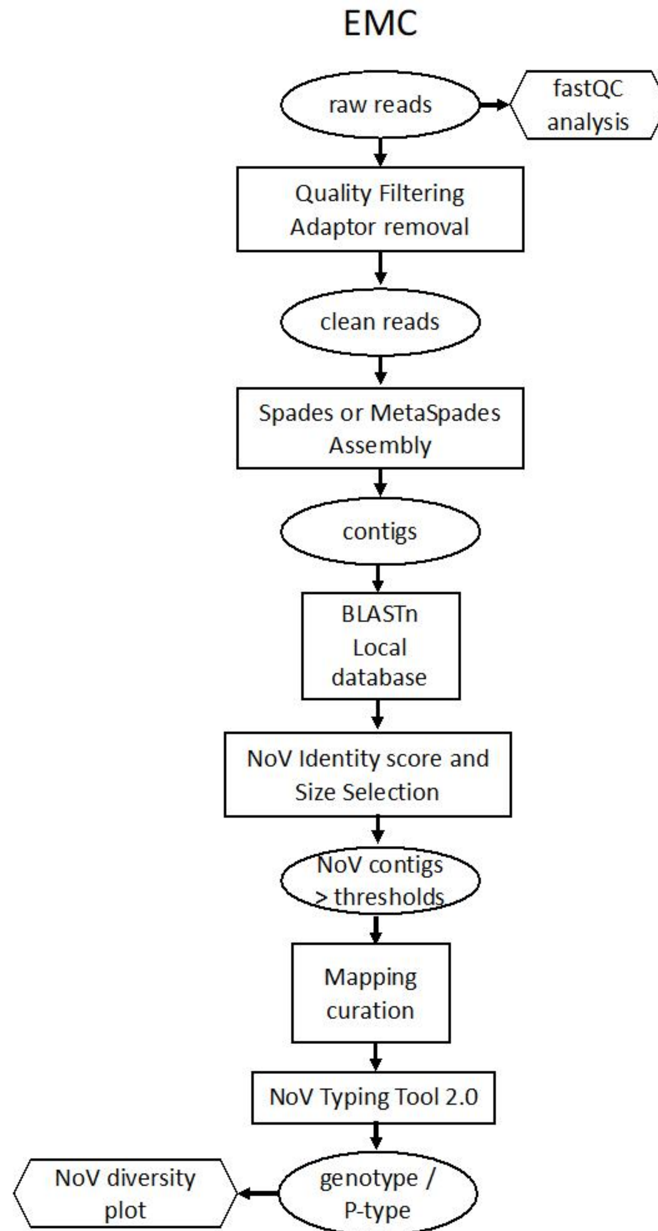



**FIGURE 6: METAGENOMIC PIPELINE AT EMC.**

Ellipse represent the data files, rectangles are processing steps, and hexahedra are outputs or verifications.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

#### 2.2.5.6. Controls and replicates

Among the TSS2 samples, we included one negative DT sample (TSS2.5) and background DT control (TSS2.0) (see section 3.1.1 on TSS preparation). In addition, during the preparation of the TSS2, BLS and OB samples libraries, a negative control was included using ultrapure water instead of sample nucleic acid for the cDNA synthesis and subsequent steps, and an extraction control was added in the form of a negative oyster DT treated in parallel with the samples. Both types of controls (control DT and ultrapure water) were submitted to exactly the same steps as the positive samples, and were added to a pool of libraries before VirCapSeq hybridization. The volume of negative sample library added to the pool was equal to the mean volume of positive samples libraries in the pool.

For TSS1 and TSS2, triplicates of libraries were prepared to allow method validation and comparison. The TSS1 being already in the form of nucleic acids, replicates were made starting at the cDNA synthesis step of library preparation (2.2.5.3). For TSS2, replicates were made starting at the extraction step (2.2.5.1).

### 2.2.6. Long amplicon sequencing by Oxford Nanopore Technology

#### 2.2.6.1. Nucleic extraction from BMS samples

For this method, nucleic acid extraction from BMS samples is the same as the procedure described in 2.2.4.1, and follows the ISO 15216-1 (ISO 15216-1, 2017).

#### 2.2.6.2. Sequencing and Bioinformatic analysis

RNA was reverse-transcribed into cDNA using random hexamers (Invitrogen, France) and SuperScript IV (Invitrogen, France) after which a norovirus specific PCR was performed using a C1000 Thermal Cycler (Bio-Rad). Various sets of forward and reverse primers were used depending on the targeted NoV genogroup and genomic region (



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 10). For each targeted genogroup (GI and GII), we have chosen or designed primers allowing the amplification of a longer segment of the genome comprising RdRp and VP1 sequences.

Bands of the GII typing PCR were cut out of the gel and purified with the MinElute Gel Extraction Kit (Qiagen). All the GI PCR samples were cleaned with AMPure beads. The next steps of the library preparation (end repair, Native barcode ligation, adapter ligation and cleaning using AMPure beads) were performed using the 1D Native barcoding genomic DNA protocol (EXP-NBD104, EXP-NBD114 and SQK-LSK109) (Oxford Nanopore Technologies), according to the manufacturer's instructions, with the exception that 100 ng sample was used for end-repair, 30 ng for native barcoding and 350 ng for adapter ligation.  At the end of the process, a R9.4 Flowcell was loaded and run on the GridION sequencing platform.

Sequence reads were demultiplexed using Porechop v0.2.4 (github.com/rrwick/Porechop) using a barcode identity threshold of 85% and the required_two_barcodes option. Next, samples were analysed through Genome detective https://www.genomedetective.com version 1.125-1.137. (Vilsker et al., 2019). Finally, contigs were typed through the Norovirus Typing Tool 2.0 (Kroneman et al., 2011). All contigs were aligned using muscle (https://www.ebi.ac.uk/Tools/msa/muscle/)  with reference strains to check the quality of the contigs (Chhabra et al., 2019).



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**TABLE 10: PRIMERS USED FOR THE LONG AMPLICON SEQUENCING WITH OXFORD NANOPORE TECHNOLOGY.**

| NoV genogroupe | Target sequence | PCR + direction | | Primer name | Primer sequence | Amplicon size | Reference |
|---|---|---|---|---|---|---|---|
| GI | RdRp-VP1 | PCR | F | A | AIYTTICCIGCIGWRAAIGCRTT | 1204 bp | van Beek et al.,2017 |
| | | | R | B | ATGAAYACAATNGARGAYGGNCC | | |
| | | N-PCR | F | D | GACTACAGCTTGGGAYTCNACNCAR | 1079 bp | van Beek et al.,2017 |
| | | | R | E | ACTCTCATATTCCCAACCCANCCRTTRTACAT | | |
| GII | RdRp-VP1 | PCR | F | A | AIYTTICCIGCIGWRAAIGCRTT | 1192 bp | van Beek et al., 2017 |
| | | | R | C | ATGAAYATGAAYGARGAYGGNCC | | |
| | | N-PCR | F | F | GACTACTCTCGGTGGGAYTCNACNCAR | 1072 bp | van Beek et al., 2017 |
| | | | R | G | ACCTCAAAACCACCTGCATANCCRTTRTACAT | | |

### 2.2.7 Sequence alignment and phylogenetic analysis of NoV sequences

All NoV sequences downloaded from NoroNet and NoV sequences obtained within this study were aligned using MAFFT (https://mafft.cbrc.jp/alignment/server/) with standard settings. All sequences were typed using the noronet typing tool, to ensure that only sequences from the same genotype or genogroup were aligned, and that all sequences were in the correct orientation. The typing tool provides this information on the orientation in the result file. Alternatively, the direction of nucleotide sequences can be adjusted in the MAFFT settings. If sequences are not aligned properly the gap penalty can be increased. Trees were inferred using IQ tree v1.6.12 (http://www.iqtree.org) with 1000 bootstraps. The nucleotide substitution model was determined using Modelfinder of IQtree. The trees were visualized using Figtree v1.4.3. (https://github.com/rambaut/figtree/releases). Only selected bootstraps >70 are shown.

# 3. Results

## 3.1. Method comparison on test sample sets

## 3.1.1. Preparation and characterization of TSS

### 3.1.1.1. Test sample set 1

The TSS1 was prepared as described in 2.2.3.1 with tenfold dilutions performed to obtain NoV concentrations ranging from 10 to 1000 gc/µL. The concentrations of GI and GII NoV were measured using qRT-PCR and dRT-PCR methods on pure RNA extracts. The mean concentrations obtained for both tests are shown in



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 11.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**TABLE 11: THE MEAN CONCENTRATIONS OF TSS1 SAMPLES**

| Sample code | Target concentration level per strain (gc/ µL) | Total norovirus GII concentration (x8) (gc/ µL) | | Total norovirus GI concentration (x4) (gc/ µL) | |
|---|---|---|---|---|---|
| | | qRT-PCR | dRT- PCR | qRT-qPCR | dRT- PCR |
| TSS 1.1 | $10^3$ | 4.97 x$10^3$ | 3.90 x$10^3$ | 2.32 x$10^3$ | 1.88 x$10^3$ |
| TSS 1.2 | $10^2$ | 8.52 x$10^2$ | 4.69 x$10^2$ | 5.17 x$10^2$ | 6.66 x$10^2$ |
| TSS 1.3 | $10^1$ | 9.50 x$10^1$ | 8.70 x$10^1$ | 6.36 x$10^1$ | 6.20 x$10^1$ |

gc /µL: genome copies per microlitre ; The total concentration in norovirus GII is expected to be 8 times the concentration per strain, and the total GI, 4 times the concentration per strain, since TSS1 contain 8 GII strains and 4 GI strains.

The target concentrations were achieved for each concentration level, with reasonable differences to be expected due to the accuracy of the quantification methods.

#### 3.1.1.2. Test sample set 2

#### *Oyster testing before bioaccumulation*

Oysters from a production zone classified B (CE N° 854/2004) in Brittany (France) were selected for the study. The sanitary quality of this zone, monitored by the French surveillance network (REMI) didn't reveal any alert of faecal contamination of the production area during 2 months preceding oyster sampling. The reproductively inactive, triploid (3n) oysters were chosen for the experiment to avoid PCR inhibitors usually present during the reproduction period. Moreover, the bioaccumulation experiments were programmed in autumn, when the oyster reproduction period is over and the ligand expression in oysters start to raise (Maalouf et al., 2010). At this period the risk of NoV circulation in environment is still relatively low because winter epidemics have not started in the human population yet. The oysters were submitted to standard purification procedure before arrival to the laboratory. After reception, the absence of NoV was verified on 50 oysters randomly chosen and analysed according to ISO 15216-1 method (2017). The test was then performed on five DT samples and each sample was analysed in triplicates of un-diluted RNA (Table 12).

**TABLE 12: OYSTER TESTING BEFORE BIOACCUMULATION**

| Code | Norovirus GI | | Norovirus GII | |
|---|---|---|---|---|
| | Ct | Concentration (gc/g DT) | Ct | Concentration (gc/g DT) |
| T0-1 | No Ct – 36.25 – No Ct | Detected, <LOQ | No Ct - No Ct - No Ct | Not detected |
| T0-2 | No Ct - No Ct - No Ct | Not detected | No Ct - No Ct - No Ct | Not detected |
| T0-3 | No Ct - No Ct - No Ct | Not detected | 35.96 - No Ct - No Ct | Detected, <LOQ |
| T0-4 | No Ct - No Ct - No Ct | Not detected | 36.08 - No Ct - No Ct | Detected, <LOQ |
| T0-5 | No Ct - No Ct - No Ct | Not detected | No Ct – 35.57 – 39.49 | Detected, <LOQ |

Ct: cycle threshold; three Ct values are shown as the analysis was done in triplicates of undiluted RNA; LOQ: limit of quantification of the qRT-PCR method; DT: digestive tissues; gc/ g DT: genome copies per gram of digestive tissues

Despite all the precautions implemented when choosing the oysters, the batch selected for the study was slightly contaminated by NoV, at very low concentrations. DT from these oysters before bio-accumulation was included as a background control in the TSS2 (TSS2.0). This resulted indeed in the detection of some NoV sequences using metabarcoding (see 3.1.2) and VirCapSeq metagenomics (see



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

3.1.3), but did not impact on the analysis of the other TSS since other negative controls (water) were presents and threshold were set accordingly.

***Homogeneity test on bioaccumulated oysters***

The oysters were contaminated as described in section 2.2.3.2 with a mix of six NoV strains, 4 strains belonging to GII and two strains to GI (as previously listed in Table 2). After bioaccumulation, the NoV concentrations as well as the homogeneity of contamination were evaluated on five DT samples by RT-qPCR according to ISO 15216-1 and using both ISO and VirCapSeq metagenomics RNA extraction methods, described in section 2.2.4.1 and 2.2.5.1, respectively. As presented in **Error! Reference source not found.** Table 13, the results confirm homogeneity of contamination. The NoV concentrations in DT are, as expected, close to $10^4$ gc/g of DT using the ISO 15216 method.

**TABLE 13: HOMOGENEITY TEST ON CONTAMINATED OYSTERS**

| Extraction method | Code | Norovirus GI | | Norovirus GII | |
|---|---|---|---|---|---|
| | | Mean Ct value | Concentration (gc / g DT) | Mean Ct value | Concentration (gc / g DT) |
| ISO 16216-1 | PK1 | 28.93 ±0.8 | 3.47 x$10^4$ | 27.48 ±0.3 | 8.32 x$10^4$ |
| | PK2 | 27.67 ±0.4 | 6.80 x$10^4$ | 26.22 ±0.3 | 1.58 x$10^5$ |
| | PK3 | 27.36 ±0.2 | 1.02 x$10^4$ | 26.35 ±0.3 | 1.74 x$10^5$ |
| | PK4 | 27.89 ±0.7 | 6.73 x$10^4$ | 26.88 ±0.1 | 1.17 x$10^5$ |
| | PK5 | 28.21 ±0.6 | 5.70 x$10^4$ | 26.27 ±0.0 | 1.86 x$10^5$ |
| | mean | NA | 6.57 x$10^4$ | NA | 1.44 x$10^5$ |
| VirCapSeq metagenomics | VCS1 | 26.10 ±0.3 | 1.53 x$10^5$ | 24.39 ±1.4 | 5.18 x$10^5$ |
| | VCS2 | 25.66 ±0.2 | 2.00 x$10^5$ | 24.21 ±0.5 | 5.58 x$10^5$ |
| | VCS3 | 26.83 ±1.0 | 8.05 x$10^4$ | 24.19 ±0.6 | 5.28 x$10^5$ |
| | VCS4 | 26.19 ±0.5 | 1.29 x$10^5$ | 23.98 ±0.2 | 6.15 x$10^5$ |
| | VCS5 | 25.29 ±0.8 | 2.51 x$10^5$ | 23.82 ±0.2 | 7.09 x$10^5$ |
| | mean | NA | 1.63 x$10^5$ | NA | 5.86 x$10^5$ |

Ct: cycle threshold; DT: digestive tissues; gc / g DT: genome copies per gram of digestive tissues; NA: not applicable

Both extraction approaches give similar results. For the ISO method the mean GI NoV concentration is 6.57x$10^4$ gc/g DT and 1.44x$10^5$ gc/g DT for GII while using VirCapSeq metagenomics extraction method the calculated values are 1.63x$10^5$ gc/g DT and 5.86x$10^5$ gc/g DT for norovirus GI and GII, respectively (Figure 7). The higher efficiency of the VirCapSeq metagenomics was observed before and could be explained by the removal of remaining inhibitors that may slightly impact on NoV RNA extraction or quantification in the ISO method.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

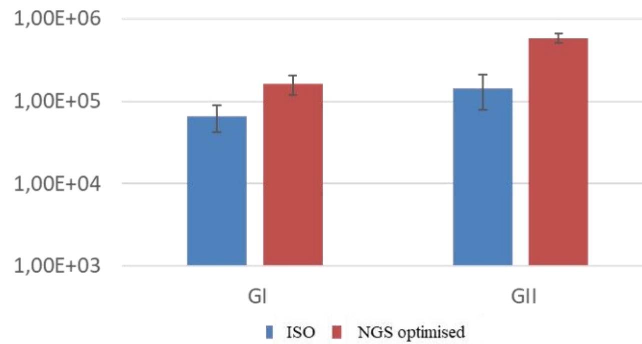



**FIGURE 7: MEAN NOROVIRUS CONCENTRATIONS CALCULATED FOR TWO EXTRACTION METHODS ON BIOACCUMULATED OYSTERS FOR TSS2 PREPARATION**

### *Preparation and quantification of TSS2 samples*

The TSS2 samples were prepared by diluting the homogenised DT of bioaccumulated oysters in homogenised negative DT. Eight dilution levels were prepared according to scheme presented on **Error! Reference source not found.**Figure 8.

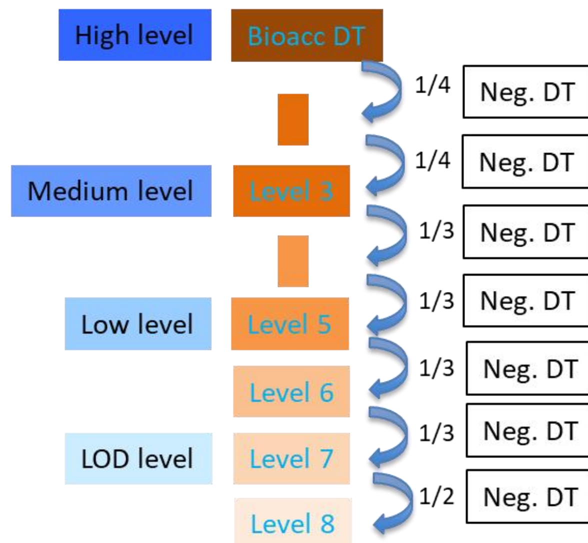



**FIGURE 8: SERIAL DILUTIONS OF CONTAMINATED DIGESTIVE TISSUES (DT)**

The contaminated DT were diluted in negative DT in cascade; the dilution factor is indicated with the arrows. Level 2 and level 4 were prepared in small quantities to serve entirely for the dilution cascade and are depicted as smaller boxes.

The NoV concentrations were evaluated for each dilution level (except levels 2 and 4 which were only prepared in small quantities as intermediate steps) using one-step qRT-PCR, and for both ISO and VirCapSeq metagenomics RNA extraction methods, as described in sections 2.2.4.1 and 2.2.5.1. Based on results of quantification, the pure DT was selected as the high contamination level (TSS2.1), level 3 as a medium concentrated (TSS2.2), level 5 as a low level (TSS2.3) and level 7 as a very low concentration level near the limit of detection of the ISO method (TSS2.4), as shown in Table 14.

Thus, nucleic acids are extracted from TSS2 following different protocols optimized for each approach (see below).



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**TABLE 14: QUANTIFICATION OF DIGESTIVE TISSUES DILUTIONS AND SELECTION OF TSS2 SAMPLES FOR NGS**

| Extraction method | Code | Norovirus GI | | Norovirus GII | |
|---|---|---|---|---|---|
| | | Ct | Concentration (gc/g DT) | Ct | Concentration (gc/g DT) |
| ISO | Pure | 28.22-28.24-28.19 | 3.46 x10$^4$ | 26.18-26.38-26.08 | 1.74 x10$^5$ |
| | L3 | 31.18-31.62-31.68 | 3.20 x10$^3$ | 29.37-31.38-28.99 | 1.52 x10$^4$ |
| | L5 | 35.3-35.83-35.44 | 1.71 x10$^2$ | 32.20-32.40-32.29 | 3.18 x10$^3$ |
| | L6 | 36.81-35.70-37.31 | Detected,<LOQ | 33.80-34.55-34.59 | 8.43 x10$^2$ |
| | L7 | No Ct-37.60-No Ct | Detected,<LOQ | 35.44-37.35-36.88 | Detected,<LOQ |
| | L8 | No Ct-No Ct-No Ct | Not detected | 36.21-36.78-36.47 | Detected,<LOQ |
| VirCapSeq metagenomics | Pure | 25.23-24.83-25.56 | 2.74 x10$^5$ | 23.68-23.70-23.70 | 7.32 x10$^5$ |
| | L3 | 28.83-29.07-29.63 | 1.63 x10$^4$ | 27.44-27.67-27.72 | 5.26 x10$^4$ |
| | L5 | 32.67-32.83-33.42 | 9.61 x10$^2$ | 30.82-31.90-30.88 | 4.17 x10$^3$ |
| | L6 | 34.32-34.73-34.24 | 3.25 x10$^2$ | 32.44-33.41-32.53 | 1.35 x10$^3$ |
| | L7 | 35.42-35.79-34.92 | 1.67 x10$^2$ | 33.34-34.59-34.65 | 5.24 x10$^2$ |
| | L8 | 37.61-42.32-36.82 | Detected,<LOQ | 35.36-35.99-35.29 | Detected,<LOQ |

Ct : cycle threshold ; three Ct values are shown as the analysis was done in triplicates of undiluted RNA; LOQ : limit of quantification of the qRT-PCR method ; DT : digestive tissues ; gc/ g DT : genome copies per gram of digestive tissues

Aliquots of 2g of DT were prepared for each of these dilution levels. Extractions and NGS analysis were organized in series containing one aliquot of each TSS2.1, TSS2.2, TSS2.3 and TSS2.4 samples as well as TSS2.0 (DT before bioaccumulation) and TSS2.5 (negative DT used for dilutions) controls.

### 3.1.2. Metabarcoding

#### 3.1.2.1. Metabarcoding on TSS1

The main objective of metabarcoding on TSS1, as described previously, was to evaluate the replicability of this approach and its ability to identify all NoV strains present in one sample, according to primer sets used for amplification (separate amplification of RdRp and VP1 versus long fragment RdRp-VP1). TSS1 metabarcoding was performed by Ifremer.

The results of the preparation and validation of TSS1 libraries are presented in the Appendix A.

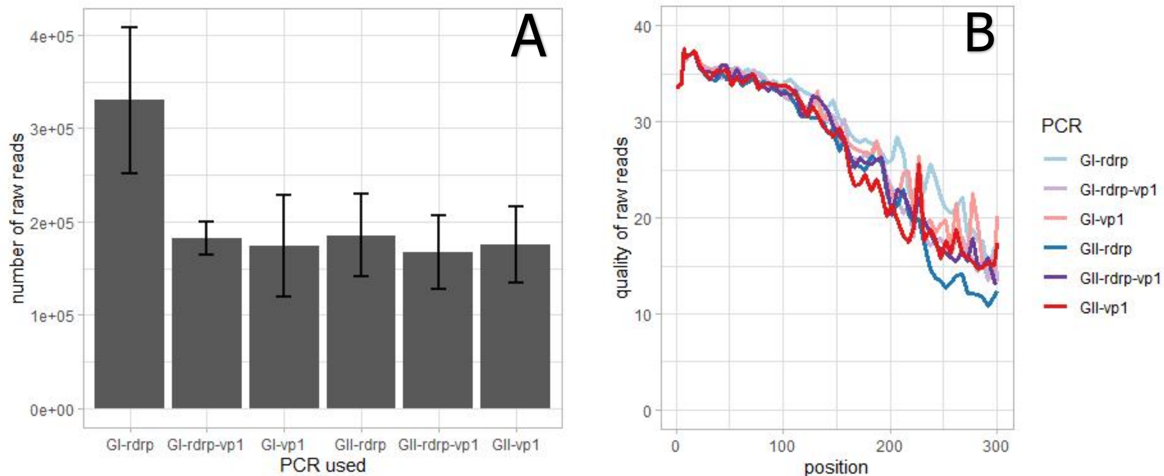
***Sequencing quality***

Following Illumina sequencing of TSS1 libraries by Ifremer, between 115K and 485K raw reads per sample and replicate were obtained, with a higher number of reads from the GI-RdRp PCR (Figure 9 A). For each initial TSS1 sample, the three replicates were labelled A, B and C throughout the analysis, and were further referred to as "samples". It is important to notice that two of these samples present a low number of reads (TSS1.3-B-GII-RdRp and TSS1.2-B-GII-RdRp). Base quality is decreasing at the end of



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

reads: quality is low (Qscore < 20, 1% error chance) after the 200 bp landmark (Figure 9 B), a profile that is often observed using 2x300 Illumina sequencing.




#### FIGURE 9: METABARCODING READS QUALITY AND COUNT FOR TSS1 SAMPLES

Number of raw reads in each group of sequencing (A), bars represent standard-deviations. Quality of raw reads among positions in the read (B), quality is the mean Q score given by the platform.

### *Diversity and relative abundancy of NoV GI genotypes identified using three sets of primers*

Four NoV GI strains were expected to be identified in the TSS1 (Table 2). Results are presented in Figure 10. Using the three primers set, targeting RdRp, VP1, and RdRp-VP1 junction, the GI.1[P1] strain was never detected. This could be due to a technical error during TSS1 preparation resulting in the absence or lower concentration of this strain in the mix. Indeed, since the three sets of primers were affected, a metabarcoding bias excluding GI.1[P1] sequencing is unlikely. In addition, the other methods (VirCapSeq metagenomics and Long amplicon sequencing by ONT) also failed to recover GI.1[P1] sequences in TSS1 (see section 3.1.3), suggesting that this strain was not added in TSS1 at the expected concentrations. Finally, the same metabarcoding approach was able to identify GI.1[P1] in TSS2, confirming that this was not due to the sequencing method, but to an error in TSS1 preparation.

An identification threshold was set to exclude clusters with very low numbers of reads, that were also present in the negative control (ultrapure water). These reads are likely not due to the cross-contamination of the negative sample, because there was no amplicon observed in the gel and very low DNA concentration during library preparation. Rather, they may result from mis-sequencing or swapping of Illumina indexes, and are observed in all runs. This underlines the need to include a negative sample in each sequencing run, and set thresholds on these controls for each run.

Using the VP1-primer set (Figure 10 A), we identified the three other capsid genotypes in all samples but one (TSS1.3-B), with a skewed diversity showing a higher relative abundancy for GI.4 than for GI.3 and GI.2. Using the RrRp-primer set (Figure 10 B), only GI.P2 and GI.P4 were detected, this time with a much higher relative abundancy for GI.P2. Finally, the primer set targeting the RdRp-VP1 junction allowed the typing of both VP1 and RdRp (dual typing) and the detection of GI.4[P4] in all samples, and GI.3[P3] in a minority of samples with a very low relative abundancy (Figure 10 C). The GI.2 capsid was also detected but not its corresponding polymerase sequence (GI.P2).

In conclusion for GI metabarcoding, the VP1-primer set allowed to detect all strains present in the sample (except GI.1[P1] which is actually absent), while the other sets detected two (RdRp) or one (RdRp-VP1) strain in most samples. Replicates were highly similar for each sample, with differences in



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

the detection of minority strains only in some of the less concentrated samples for VP1 (TSS1.3-B) and RdRp-VP1 (TSS1.2-C and TSS1.3-A/B).



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

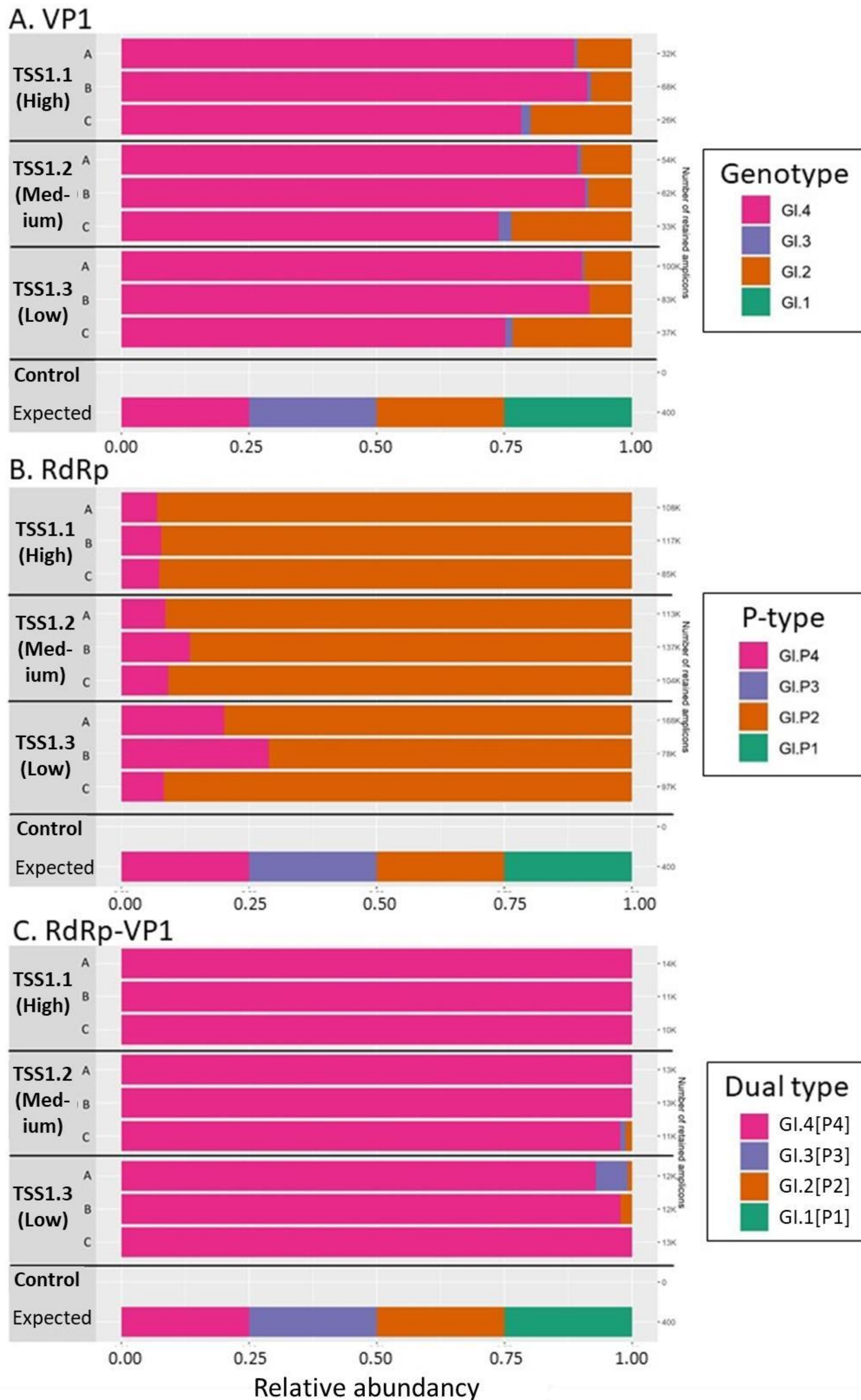



**FIGURE 10: RELATIVE ABUNDANCY OF CLUSTERS IDENTIFIED AS GI GENOTYPES IN TSS1 USING THREE PRIMER SETS.**



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

For each sample (TSS1.1 – high, TSS1.2 – medium – TSS1.3 – low), three replicates (A, B, C) were sequenced using primer sets targeting NoV GI VP1 (A.), RdRp (B.) and the junction spanning RpRp and VP1 (C.). Reads were clustered and assigned to a genotype or P-type, or both for the RdRp-VP1 junction (dual typing) and the relative abundancy of each genotype was compared to the expected one (bottom line). The few NoV sequences identified in the negative control consisting of ultrapure water were excluded by setting a minimal number of reads to be considered positive.

### *Diversity and relative abundancy of NoV GII genotypes identified using three sets of primers*

Eight NoV GII strains were expected to be identified in the TSS1, combining 7 capsid genotypes and 6 RdRp P-types (Table 2). Results are presented in Figure 11.

Identification threshold were set on the negative control sample (ultrapure water) to exclude clusters with very low numbers of reads. Of note, for the RdRp primer set, the second series of replicates (series B) showed a lower number of reads, ending up below the identification threshold for the three TSS1 samples. As these replicates were processed together, we suspect a technical issue during library preparation (PCR or N-PCR step). Other replicates displayed high number of reads and were highly similar among each sample.

Using the VP1-primer set (Figure 11, A), we identified the 7 NoV GII capsid genotypes in all samples but the TSS1.3-A replicate for which NoV GII.3 (purple) is missing. Relative abundancies differed moderately from the expected ones, with a tendency to over-representation of GII.17 and under-representation of GII.4.

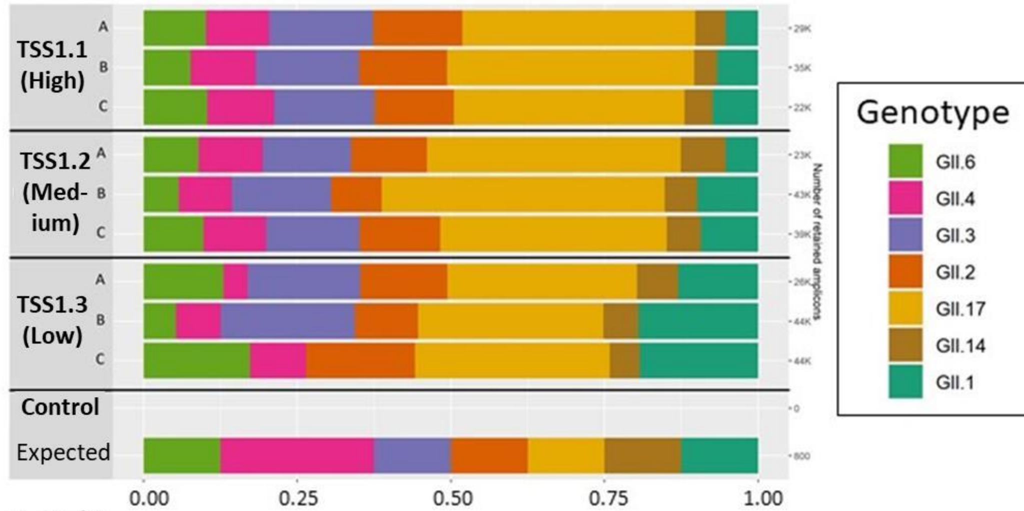
Using the RdRp-primer set (Figure 11, B), we identified 4 NoV GII polymerase P-types, with 2 P-types (GII.P7 and GII.P16) lacking from all samples. These genotypes were also not detected among the clean reads using direct mapping on references (not shown), so their absence in the FROGS analysis results was not due to the bioinformatic pipeline. This rather suggests that the primers used were less likely to catch these P-types in a mix of different strains. Among the P-types actually detected, the relative abundancies were also different from the expected ones, with over-representation of GII.P31 while the corresponding genotype, GII.4, is under-represented using VP1-primer set.

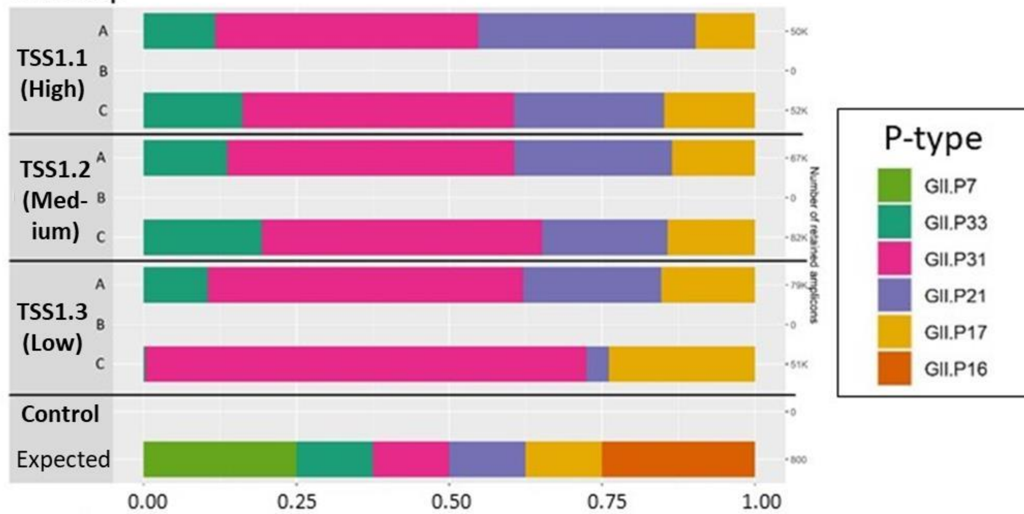
When targeting the RdRp-VP1 junction (Figure 11, C), we identified three of the expected RdRp-VP1 combinations: GII.4[P16] (blue), GII.2[P16] (orange), and GII.4[P31] (pink). The GII.17 capsid genotype was also identified, but unexpectedly combined to P31 and P16 instead of P17 (grey). This shows that chimeras can be formed during PCR or Illumina sequencing. During the bioinformatics analysis of metabarcoding results, chimeras were removed using a dedicated FROGS tool and by manual curation of the final clusters. Yet, both steps rely on the identification of parental strains to detect a chimera. Here, the GII.17[P17] parental strain was not identified, hence GII.17[P16] and GII.17[P31] could only be identified as chimera because the actual strains present in the sample were known.

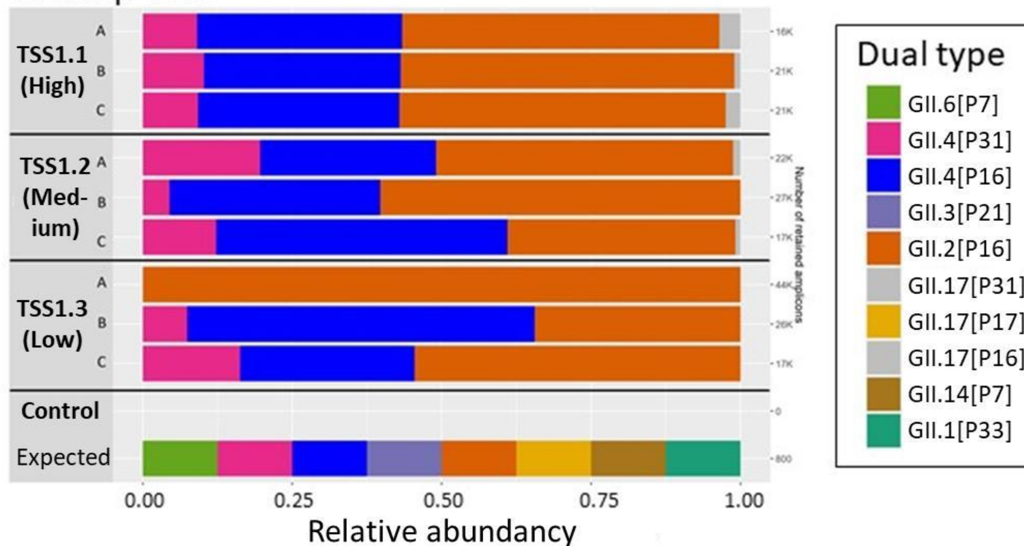
In conclusion for GII metabarcoding of TSS1, the VP1-primer set allowed to detect all strains present in the mix, while the other sets detect four (RdRp) or three (RdRp-VP1) strains. Between replicates, results were mainly similar except for some of the less concentrated samples, with differences in the detection of minority strains for VP1 (TSS1.3-B) and RdRp-VP1 (TSS1.2-C and TSS1.3-A/B).



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.






The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**FIGURE 11: RELATIVE ABUNDANCY OF CLUSTERS IDENTIFIED AS GII GENOTYPES IN TSS1 USING THREE PRIMER SETS.**

For each sample (TSS1.1 – high, TSS1.2 – medium – TSS1.3 – low), three replicates (A, B, C) were sequenced using primer sets targeting NoV GI VP1 (A.), RdRp (B.) and the junction spanning RpRp and VP1 (C.). Reads were clustered and assigned to a genotype or P-type, or both (dual typing) and the relative abundancy of each genotype was compared to the expected one.

### *Concluding remarks for metabarcoding on TSS1*

In conclusion for GI and GII metabarcoding on TSS1, we observed that all primer sets presented selection biases when applied on a mix of NoV strains. They yielded different relative abundancies for the different strains of each genogroup, which shows that relative abundancies do not reflect the actual proportion of each strain in the mix, but likely result from matrix-primers selective interactions. For both genogroups, the VP1-primer set allowed to detect all strains expected in the sample (except GI.1[P1] which was not detected with other methods either and may actually be missing in TSS1), while the other sets missed some strains. The primer sets targeting the RdRp-VP1 junction, although promising, exhibited several important limitations: less strains were identified than using the other primer sets both for GI and GII, and some partial sequences or chimera were detected, impairing the identification of recombinants in natural samples when the composition is unknown.

Regarding replicability, the method was robust, as exemplified by the high similarity between replicates for each sample. Differences in the number of identified strains were observed between some replicates in the less concentrated samples for GI VP1 (TSS1.3-B) and RdRp-VP1 (TSS1.2-C and TSS1.3-A/B), and GII VP1 (TSS1.3-C) and RdRp-VP1 (TSS1.3-A). There was no significant effect of the NoV concentration in this sample set except for this tendency for altered diversity in the lowest level (TSS1.3). This suggests that the method may lack repeatability on samples with a low NoV concentration.

Yet, TSS1 was not designed to test the method sensitivity and the concentration of NoV in these samples is higher than the one observed in contaminated BMS. To test the sensitivity of the method, and its repeatability when NoV concentration is very low, we used the test sample set 2 (TSS2) that present lower NoV concentrations and takes into account the possible BMS matrix effect.

## 3.1.2.2. TSS2 Metabarcoding results

The main objective of metabarcoding on TSS2 samples is to evaluate the sensitivity and repeatability of this approach when applied to shellfish samples. In addition, the entire process, from RNA extraction to Illumina sequencing was repeated three times starting at RNA extraction step from TSS2 by two laboratories (Ifremer and Cefas) to test the reproducibility of the method. DTU laboratory also performed metabarcoding analysis on TSS2 RNA extracted by Ifremer. The extraction results, libraries preparation and validation are presented in Appendix B.

### *Sequencing quality*

All three laboratories obtained a variable amount of raw reads (Figure 12, light grey) depending on the primer set and sample. Following the first step of the metabarcoding analysis (i.e. FROGS pre-process), a variable proportion of reads were kept for subsequent analysis, based on the presence of primer sequences and overlap between R1 and R2, and were referred to as clean reads (Figure 12, dark grey). At Ifremer, this proportion varied from 50% (NoV GII-RdRp-VP1) to 85% (NoV GI- RdRp). At DTU and Cefas, clean reads comprised between 70 to 95% of the raw reads. This difference may be due to different indexing strategies and to the overall quality of raw reads, which is lower for Ifremer (most observed Q score = 32 to 34) than for Cefas and DTU (most observed Q score = 34 to 37) after excluding controls and low-quality amplifications.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

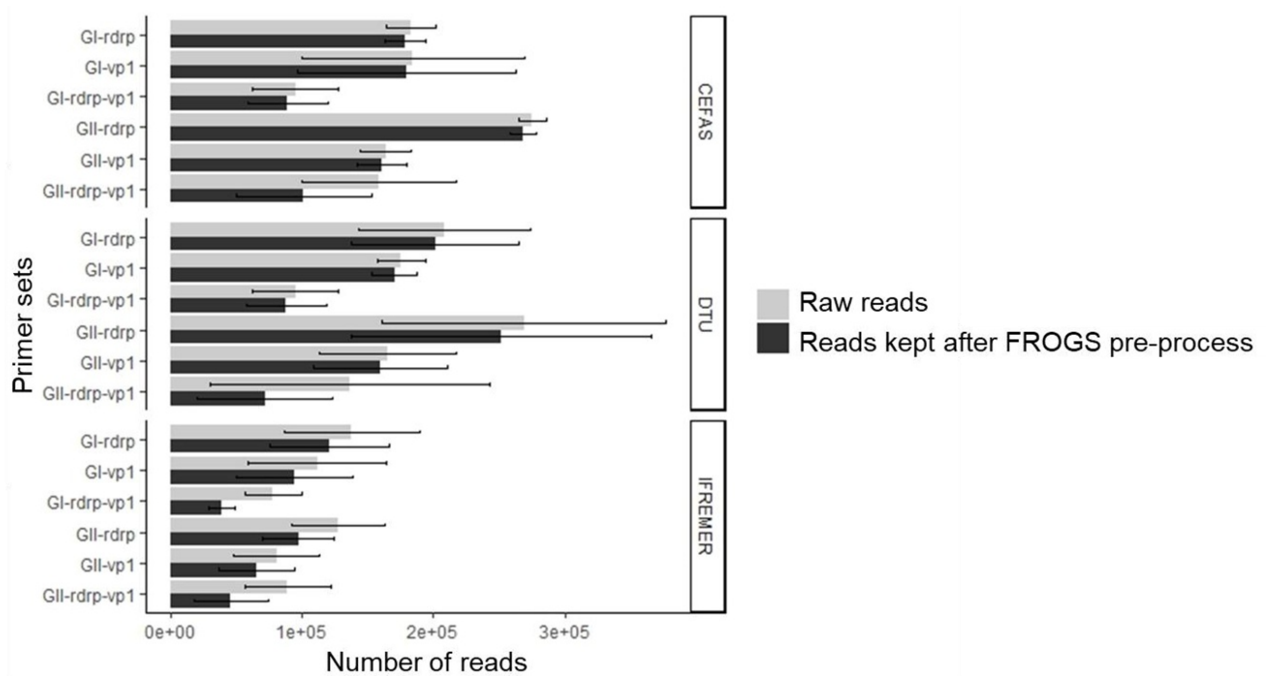



**FIGURE 12: MEAN NUMBER OF RAW AND READS KEPT AFTER FROGS PREPROCESS (CLEAN READS) OBTAINED FOR EACH SET OF PRIMERS BY THE THREE LABORATORIES**

Mean number of raw reads (light grey) obtained for each primer set on positive samples (TSS2.1, 2, 3 and 4), bars represent standard-deviations. Mean number of clean reads (dark grey), i.e. reads remaining for analysis after the first step of analysis, the FROGS pre-process, which includes selecting reads starting with the primer sequences and merging of R1 and R2 reads.

### *Diversity and relative abundancy of NoV GI genotypes identified using three sets of primers in the three laboratories*

Two NoV GI strains were expected in TSS2, GI.4[P4] and GI.1[P1]. Metabarcoding results using the three primer sets are presented in Figure 13, as the mean and standard deviation of the genotypes proportion in the three replicates, for each sample and laboratory. Identification thresholds were set on the negative controls (ultrapure water) to exclude NoV GII clusters with very low numbers of reads and only consider reliable results. Doing so, the few reads corresponding to NoV sequenced with the control oyster tissues (TSS2.0 – background control, and TSS2.5 – negative oysters used for dilution) were often filtered out and no NoV was identified in most of these samples, except in TSS2.5 with GI.4 and GI.P1 identified by DTU and GI.P4 by Cefas.

Importantly, GI.1[P1] was identified in all samples which confirms that the absence of this genotype in results from TSS1 is not due to impaired amplification by the primer sets, but rather to its absence or very low concentration in the TSS1.

Using the VP1-primer set (Figure 13 A), results were very similar between the three laboratories, with identification of the two expected capsid genotypes in samples TSS2.1, TSS2.2, TSS2.3. For TSS2.4, the sample with lowest NoV concentration levels, results were more variable between replicates with sometimes only one genotype identified. As observed with TSS1, relative abundancies differed from the expected ones and GI.4 is over-represented.

Using the RdRp-primer set (Figure 13 B), similar results were obtained by the three laboratories, with identification of the two P-types, in all replicates for TSS2.1, 2.2 and 2.3. Again, results were more variable for TSS2.4, as observed with the large error bars for Cefas and DTU, and the sole identification of GI.P4 by Ifremer. As for TSS1, the GI.P1 was over-represented, with relative abundancies varying between the three laboratories from roughly 60% (Ifremer) to 75% (DTU) and 95% (Cefas). In TSS2.5,



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Cefas identified the genotype GI.P4 and DTU, GI.P1, with a high number of reads equivalent to a positive sample (195 and 274 K reads, respectively), which is in agreement with the detection of a visible amplicon during electrophoresis at Cefas (Appendix B). The negative oysters used for DT dilution, that constituted the TSS2.5 sample, were never found positive for NoV using the ISO method, but giving the high sensitivity of the metabarcoding assay, very low contaminating NoV strains might have been detected here. Alternatively, although all recommended precautions were taken during library preparation, a contamination of this control cannot be excluded.

Using the RdRp-VP1 primer set (Figure 13 C), results were also similar between Cefas and DTU laboratories for TSS2.1, 2.2 and 2.3, with identification of GI.4P4 and GI.1P1 albeit with under-representation of the latter, while Ifremer detected only GI.P4 in most replicates. In TSS2.4, the three laboratories detected mainly GI.4[P4]. In addition, chimeric sequences combining the different genotypes and P-types were identified using this primer set. Here, these chimeras were kept during analysis because they represented a high proportion of the clusters, their relative abundance being higher than that of the parental GI.1[P1] strain, and the recombination point was located at the junction between RdRp and VP1 genes, the natural recombination hot-spot of noroviruses. Such chimeras would not be distinguishable from natural recombinants if identified in a natural sample where the contaminating strains are unknown.

In conclusion for GI metabarcoding, the three primer sets allowed to identify the two strains present in the mix, but the RdRp-VP1 primer set also identified chimeric recombinants. VP1 and RdRp primer sets yielded reproducible results among replicates and laboratories. An effect of the NoV concentration levels was observed with often less replicability of the results for the sample with very low levels of contamination (TSS2.4), or less diversity, as observed also for TSS2.3 with the RdRp-VP1 primer set.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

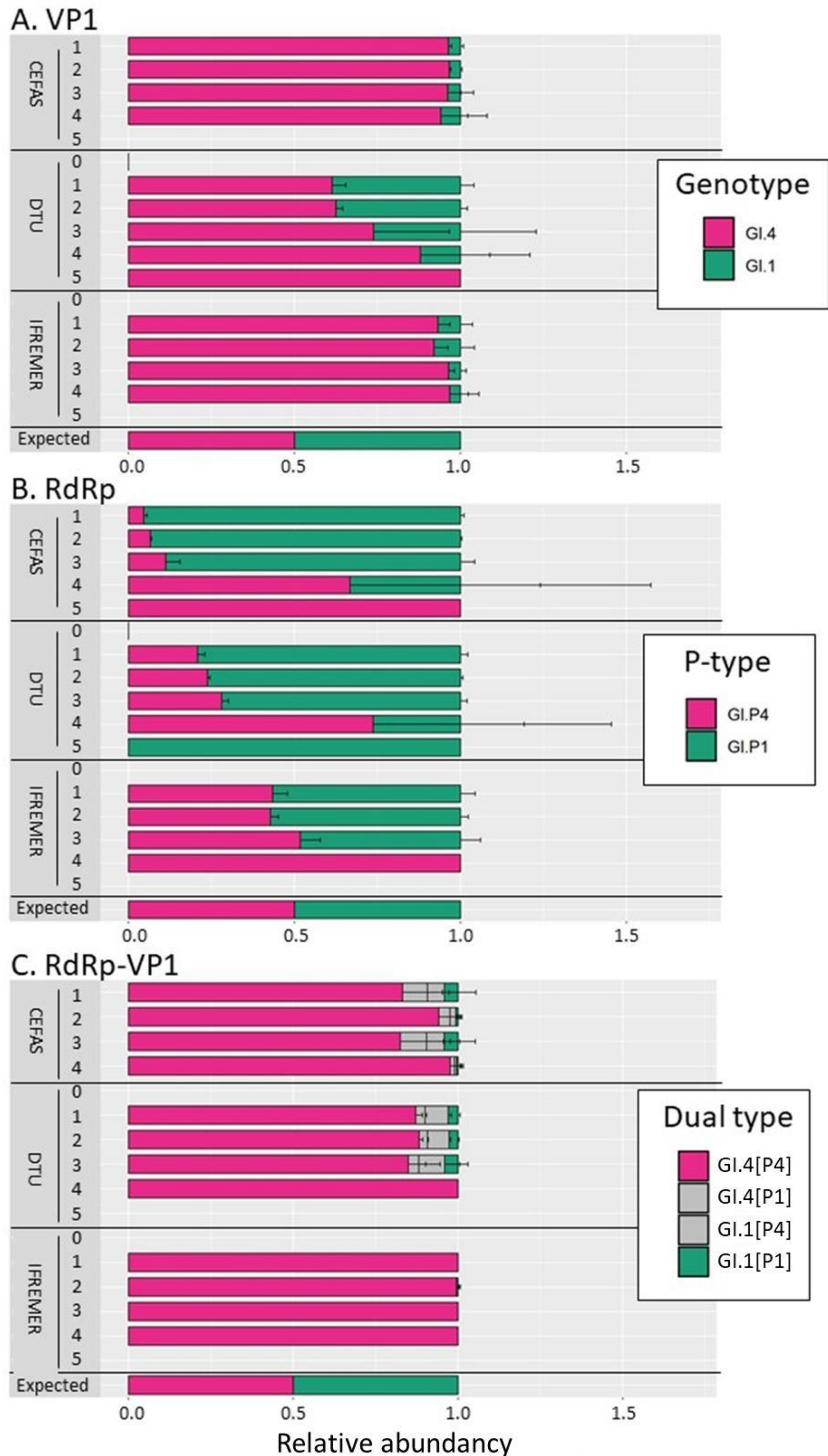



**FIGURE 13: RELATIVE ABUNDANCY OF NOV GI GENOTYPES IDENTIFIED IN TSS2 USING THREE PRIMER SETS.**



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

For each sample (TSS2.0 to TSS2.5), three replicates were sequenced using primer sets targeting NoV GI VP1 (A.), RdRp (B.) and the junction spanning RpRp and VP1 (C.) in three laboratories (Cefas, DTU, Ifremer), except TSS2.0 that was not sequenced by Cefas, and TSS2.5 that was sequenced by Cefas only with primers sets for VP1 and RdRp, not RdRp-VP1. Reads were clustered and assigned to a genotype or P-type, or both (dual typing) and the relative abundancy of each genotype was compared to the expected one. Mean relative abundancies across the three replicates are indicated for each sample and laboratory, with standard deviations as error bars.

### *Diversity and relative abundancy of NoV GII genotypes identified in TSS2 using three sets of primers in the three laboratories*

Four norovirus GII strains were expected to be identified in the TSS2. Metabarcoding results using the three primer sets are presented in Figure 14, as the mean and standard deviation of the genotypes proportion in the three replicates, for each sample and laboratory. Identification threshold were set on the negative control (water) to exclude NoV GII clusters with very low numbers of reads and only consider reliable results. Consequently, the few reads corresponding to NoV sequenced with the control oyster tissues (TSS2.0 – background control, and TSS2.5 – negative oysters used for dilution) were filtered out and no NoV was identified in these samples, except GII.4 and GII.3 in TSS2.0 by Cefas, albeit with a very low number of reads (2K) barely above the identification threshold. This could be due to the low NoV contamination of the oyster batch that was detected during TSS2 preparation.

Using the VP1-primer set, the three laboratories identified the four expected capsid genotypes in samples TSS2.1, TSS2.2, TSS2.3, sometimes with one genotype lacking in one replicate. For TSS2.4, which corresponds to the lowest NoV concentrations, results were more variable between laboratories and replicates, as observed with larger error bars for the three laboratories. Similar to TSS1 results, relative abundancies were skewed and GII.17 was over-represented.

Using the RdRp-primer set, very similar results were obtained by the three laboratories, with identification of three P-types out of the four expected ones, in all replicates for TSS2.1, 2.2 and 2.3. Again, results were less reproducible for TSS2.4, showing larger standard deviations. As observed with TSS1, the GII.P16 P-type was not identified using this primer set, regardless of the sample or laboratory.

Using the RdRp-VP1 primer set, results differed more between laboratories. Three strains were identified by Ifremer and DTU, which used the same RNA extracts: GII.4[P31] (pink), GII.2[P16] (orange) and GII.17[P17] (yellow). Cefas, which performed RNA extraction from other TSS2 aliquots, mostly detected GII.4[P31], and sometimes GII.4[P16] and GII.17[P17]. Thus, a laboratory effect at the RNA extraction step was likely involved in these discrepancies. GII.3[P21] was never detected. Replicability was good for the three laboratories for TSS2.1 and 2.2, with nearly identical results between replicates, but replicates were differing for TSS2.3 and 2.4, which showed larger standard deviations. Finally, as observed for TSS1, chimeric or partial sequences were identified using this primer set.

In conclusion for GII metabarcoding, the VP1-primer set allowed to identify all strains present in the TSS2 samples, while the other sets identify three strains in most samples. VP1 and RdRp primer sets yielded reproducible results among replicates and laboratories. An effect of the NoV concentration levels was observed with consistently less replicability of the results for the sample with very low levels of contamination (TSS2.4) for most primer sets and laboratories, and also TSS2.3 for the RdRp-VP1 primer set.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

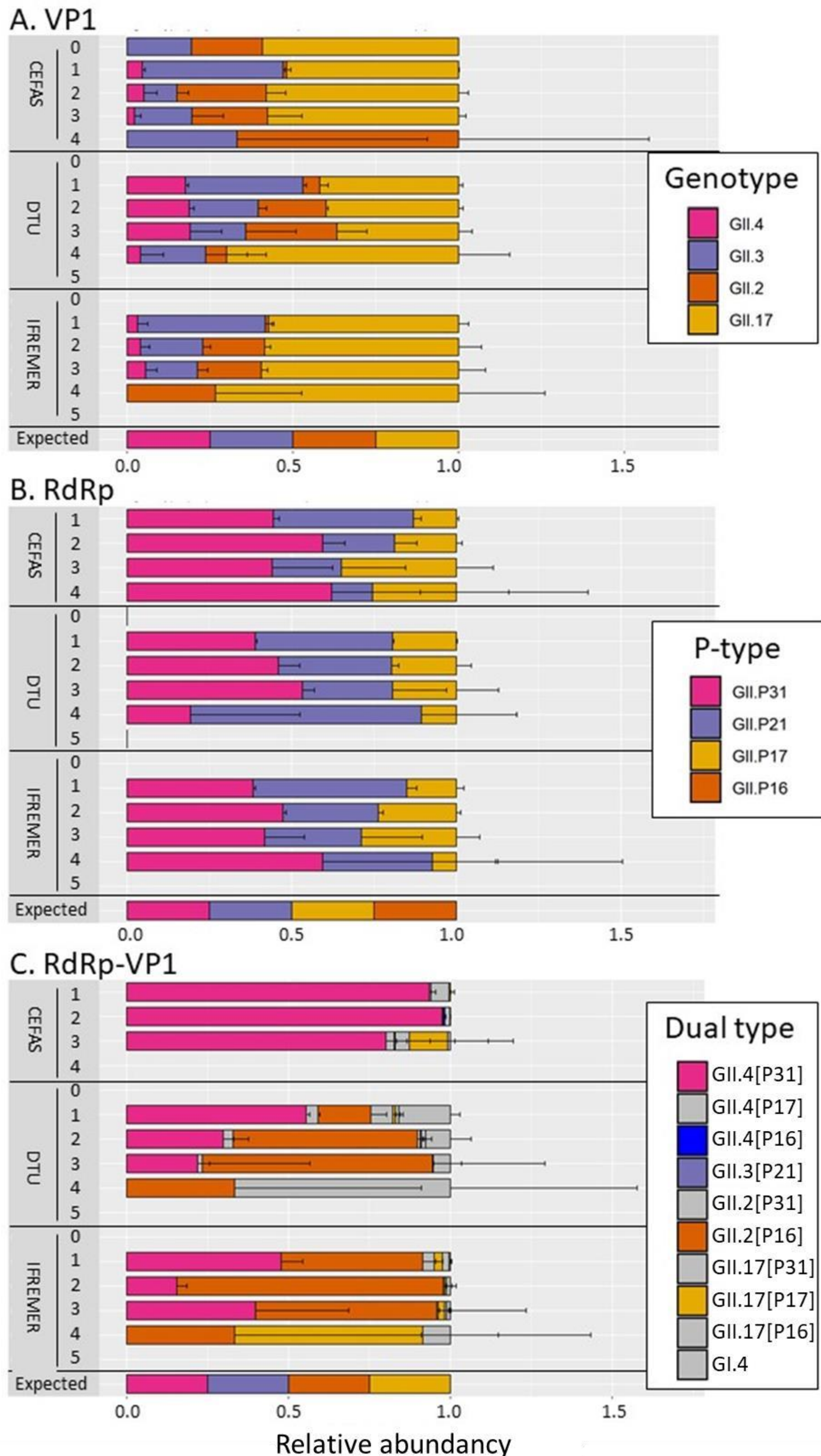



**FIGURE 14: RELATIVE ABUNDANCE OF NOV GII GENOTYPES IDENTIFIED IN TSS2 USING THREE PRIMER SETS.**



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

For each sample (TSS2.0 to TSS2.5), three replicates were sequenced using primer sets targeting NoV GII VP1 (A.), RdRp (B.) and the junction spanning RpRp and VP1 (C.) in three laboratories (Cefas, DTU, Ifremer), except TSS2.5 that was not sequenced by Cefas for all targets and TSS2.0 that was only sequenced by Cefas using primers for VP1. Reads were clustered and assigned to a genotype or P-type, or both (dual typing) and the relative abundancy of each type was compared to the expected ones. Mean relative abundancies across the three replicates are indicated for each sample and laboratory, with standard deviations as error bars.

### *Concluding remarks for metabarcoding on TSS2*

In conclusion for GI and GII metabarcoding on TSS2, we confirmed that all primer sets presented selection biases, also when applied on shellfish samples. This important result emphasizes that relative abundancies are only indicative and cannot be used to draw quantitative conclusions.

For both genogroups, the VP1-primer set allowed the detection of all expected strains, and the RdRp-primer sets also detected most of the strains, which confirms the results observed using TSS1. The primer sets targeting the RdRp-VP1 junction allowed the identification of most strains also but with detection of chimera in GI and GII. Importantly, these chimeras represented a higher proportion of reads than the parental strains. Since the recombination point was located at the junction between RdRp and VP1 genes, a natural recombination hotspot for NoV, they could only be distinguished here because the actual diversity of strains was known. In natural samples, these clusters would appear as new recombinants, a false positive result representing an important limitation for the use of this primer set.

Regarding technical replicability, the method was robust on samples up to concentrations of roughly $1.10^2$ and $1.10^3$ gc/g of DT for GI and GII, respectively, which corresponds to TSS2.3 samples, for RdRp and VP1. The RdRp-VP1 primer set was more affected by the dilution, with variability already at the level of TSS2.3. On TSS2.4 samples, whose concentration are below the LOQ of the qRT-PCR (ISO), replicates were more variable for most primer sets and laboratories. This is expected, as very low concentrations of NoV genomes result in sampling bias when taking the few µl required for the assay. Interestingly, here, combining results from the three replicates often allowed to observe more diversity. Thus, performing replicates can inform on the robustness of the result, but also lower the sampling bias when working on samples with very low concentrations. However, this increases both processing time and costs.

We conclude that the metabarcoding approach is suited to identify at least a subset of the existing NoV GI and GII diversity, even in samples below the LOQ of the qRT-PCR (124 and 389 gc/g of DT for GI and GII, respectively).

Regarding reproducibility between the different laboratories, the metabarcoding method showed identical results in terms of strain identification between the three laboratories for VP1 and RdRp primer sets for GI and GII, while the RdRp-VP1 primer set yielded less reproducible, but compatible, results.

### 3.1.2.3. Concluding remarks for metabarcoding approach on TSS samples

Taken together, the results obtained on TSS1 by Ifremer and on TSS2 by three consortium partners allow us to raise the following conclusions:

- The VP1 primer sets reflected the actual diversity of NoV GI and GII, while the RdRp primer sets allowed detection of a subset of this diversity. These results were obtained with 8 strains for GII and 4 strains for GI, selected among the most prevalent NoV strains, but should be confirmed on other and more rare NoV strains.
- The RdRp-VP1 primer sets displayed several important limitations: they resulted in the identification of chimeric strains that were similar to natural recombinants, and thus cannot allow a reliable identification of recombinants in samples with unknown composition; they were less sensitive than the two other primer sets, probably because of the longer amplicon, which is more difficult to synthetize when the matrix is rare; their results were affected by the extraction step and/or by the laboratory in which the method was applied. Therefore, this



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

approach was not selected for the sequencing of BLS samples, and separate VP1 and RdRp amplification and sequencing were carried out for this sample set (see below).

- Identification thresholds must be set for each run, since NoV reads are always present in the negative controls, at low levels, probably due to the swapping or mis-sequencing of Illumina indexes.
- Using VP1 and RdRp primer sets, we reached a limit of identification that was below the LOQ of the qRT-PCR, and similar to the LOD of this very sensitive method. Importantly, our results show that strains identified in samples with very low concentrations (below $1.10^2$ and $1.10^3$ gc/g of DT for GI and GII, respectively) may represent a subset of the actual diversity.
- The metabarcoding approach allowed to identify NoV GI and GII genotypes and P-types in shellfish samples, but not to quantify the relative abundancy of each strain, and could not exclude the presence of unidentified strains (false negative). Conversely, there are no false positive in our results using VP1 and RdRp primer sets, i.e. all detected strains were the expected ones.

### 3.1.3. VirCapSeq Metagenomics

Both TSS1 and TSS2 were submitted to VirCapSeq metagenomics as described in the methodologies section 2.2.5. The results on preparation and validation of the respective libraries are shown in the Appendix C. As for metabarcoding, the entire test was performed in triplicates to investigate reproducibility. To assess the repeatability of the method, all samples were sequenced by Ifremer and EMC independently. The data generated by both laboratories were analysed in parallel using the same pipeline by Ifremer to allow comparisons and investigate the relative abundancies of the detected strains for TSS1 and TSS2 (3.1.3.1 and 3.1.3.3, respectively). In addition, other pipelines were used by EMC on their data and compared to Ifremer's results to investigate the impact of the bioinformatic pipeline on the detection of NoV strains by VirCapSeq metagenomics, for TSS1 and TSS2 (3.1.3.2 and 3.1.3.4, respectively).

#### 3.1.3.1. Comparative analysis of TSS1 VirCapSeq metagenomics from Ifremer and EMC using Ifremer's pipeline

***Sequencing quality***

The TSS1 libraries were sequenced together with the TSS2 libraries by two laboratories, Ifremer and EMC. At Ifremer, a NovaSeq platform was used and between 39M and 139M raw reads per TSS1 sample were obtained, except for TSS1.1A (0.8M) and TSS1.3B (0.5M), which yielded a low amount of reads (**Error! Reference source not found.**). At EMC, a MiSeq platform was used, and between 1.3 and 4.3 M raw reads were obtained. In both runs, the negative controls comprised very few reads, as expected. A small fraction of reads were removed by the quality trimming step of Ifremer's pipeline (Figure 15, red portion of bars) and most reads were kept for subsequent analysis (Figure 15, blue portion of bars), for both datasets.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

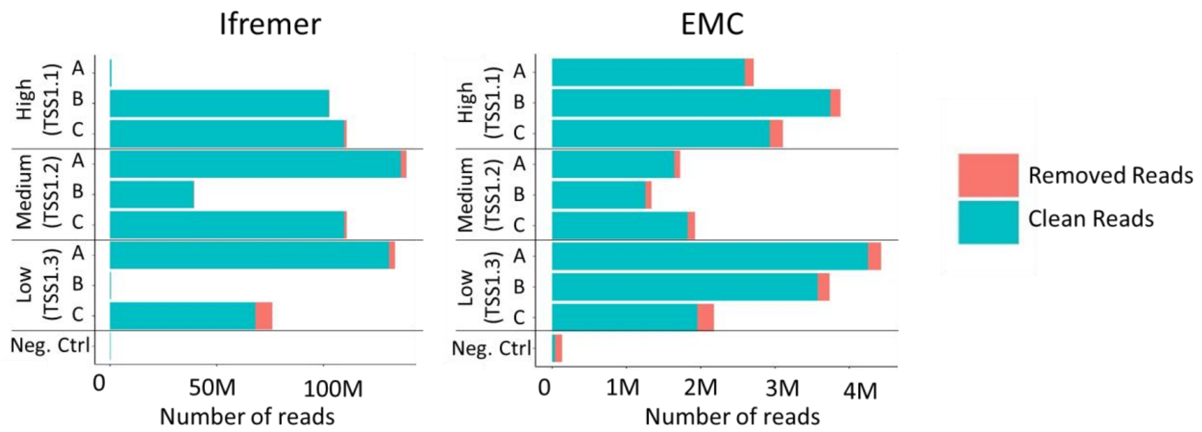



**FIGURE 15: NUMBER OF READS REMOVED FROM RAW READS AND CLEAN READS KEPT FOR ANALYSIS FOR TSS1**

A, B, C: replicates of TSS1 samples. Neg. Ctrl.: negative control (water). M: Millions of reads.

### *NoV GI and GII diversity in TSS1 sequenced by Ifremer and EMC and analysed using Ifremer's pipeline*

Both datasets were analysed as with Ifremer's pipeline as described in 2.2.5. This pipeline generates contigs, which are assemblies of overlapping reads. Among the contigs longer than 500bp, those corresponding to NoV sequences were identified and typed using the Norovirus Typing Tool 2.0. A fraction of the contigs could be both genotyped and P-typed (dual typing), while for the other, only the genotype, the P-type or the genogroup could be identified (



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 11). One contig with a RdRp-VP1 combination that was absent from the strains selected, was identified as a false recombinant where the RdRp sequence was assembled to the wrong VP1 sequence by the pipeline (



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 11). All other contigs with either P-type or genotype assignation (or both) corresponded to one of the initial strains.

**TABLE 15: NUMBER, MAXIMUM SIZE AND GENOTYPING OF NOV SEQUENCES (CONTIGS) OBTAINED FOR EACH SAMPLE OF THE TSS1 BY BOTH LABORATORIES**

| Dataset | Sample | Number of NoV contigs >500 bp | Max. contig size | Number of contigs with dual typing | Number of false recombinants |
|---|---|---|---|---|---|
| **Ifremer** | TSS1.1 | 44 | 5059 | 20 | 0 |
| | TSS1.2 | 32 | 4704 | 10 | 0 |
| | TSS1.3 | 22 | 6426 | 9 | 0 |
| | NC | 3 | 4059 | 2 | 0 |
| **EMC** | TSS1.1 | 80 | 7248 | 25 | 0 |
| | TSS1.2 | 16 | 3189 | 6 | 1 (GII.1[P21]) |
| | TSS1.3 | 34 | 1564 | 3 | 0 |
| | NC | 1 | 528 | 0 | 0 |

NC – negative control

The genotype and P-type assignations of these NoV contigs, and their relative reads abundancy in each sample and replicate are presented in Figure 16**Error! Reference source not found.** with the same colour code as for the metabarcoding analysis for the expected NoV strains. Contigs falling outside the typing regions, for which only the genogroup could be determined, are in grey.

In Ifremer's dataset, for each sample and replicate, between 625 (TSS1.3-B) and 779K (TSS1.2-C) reads were assembled in contigs assigned to NoV GI or NoV GII. In EMC's dataset, the number of reads in NoV contigs was lower, ranging from 60 (TSS1.2-C) to 87K (TSS1.3-B). This difference of roughly one order of magnitude between the two laboratories was expected due to the different sequencing platforms and overall depth of sequencing of the respective runs.

In both datasets, the negative control (N.C.), corresponding to molecular grade water that followed the same treatment as samples, showed from 0 to 834 reads in contigs assigned to NoV. This is likely due to mis-indexing of reads, but a contamination during library preparation cannot be excluded. It sets a threshold below which results cannot be considered reliable. In Ifremer's dataset, the TSS1.3-B NoV contigs comprised even less reads (625) than the negative control, therefore results should not be considered for this sample.

In Ifremer's dataset, for all samples with high and medium concentrations (TSS1.1 and 1.2), we identified all expected NoV GI and GII genotypes except NoV GI.1, which is missing from all samples, as observed using metabarcoding (3.1.2.1). In two replicates only, TSS1.1-A and TSS1.2-C, GII.17 or GI.3 were not detected. In the sample with low concentrations (TSS1.3), most GII contigs were not identified at the genotype level (grey). In EMC's dataset, all expected strains except GI.1 were identified in TSS1.1 (high concentration of NoV), but the diversity was much lower in samples with medium and low concentrations (TSS1.2 and 1.3). In addition, long contigs representing more than half of the NoV genome (4Kb to 7Kb) were identified as GI.4[P4] in almost all TSS1 samples and replicates. This strain may have been over-represented during the preparation of the samples, and/or may be preferentially selected by the VirCapSeq probes.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

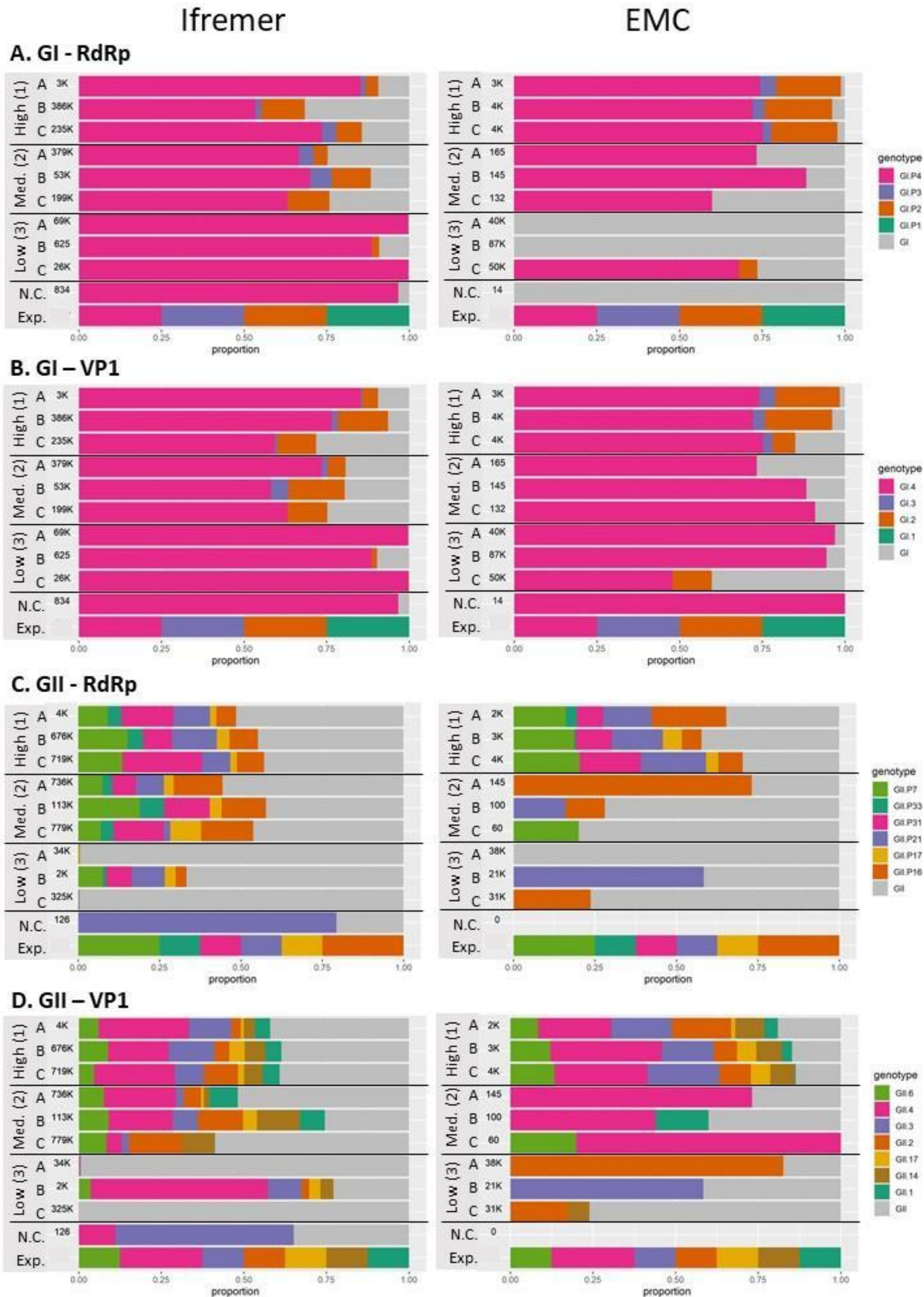



**FIGURE 16: PROPORTION OF READS ASSIGNED TO IDENTIFIED GI OR GII P-TYPES AND GENOTYPES USING IFREMER'S PIPELINE IN IFREMER (LEFT) AND EMC (RIGHT) TSS1 DATA.**



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Among contigs longer than 500 nt and comprising at least 10 reads, those falling in NoV GI – RdRp (A), GI – VP1 (B), GII – RdRp (C) or GII – VP1 (D) regions are assigned to a P-type or genotype and the proportion of their corresponding reads is depicted in colours. Contig not assigned to a P-type or genotype, but only to a genogroup, are in grey. The three TSS1 samples with high (1, TSS1.1), medium (Med., 2, TSS1.2) and low (3, TSS1.3) NoV contamination levels were sequenced in triplicates (A, B, C) together with a negative control (N. C., water). The total number of reads considered per genogroup for each sample and replicate is indicated before the bars (K: kilo). Exp: expected proportion.

#### 3.1.3.2. Comparison of datasets and bioinformatics pipelines for the detection of NoV strains in TSS1 by VirCapSeq metagenomics

The VirCapSeq metagenomics data on TSS1 were analysed by Ifremer using an in-house bioinformatics pipeline as detailed above, and by EMC using two in-house pipelines differing by the de-novo assembly tool (ie. metaSPAdes or SPAdes) and in parallel with the online pipeline Genome Detective, as described in 2.2.5.5. The NoV sequences assembled in these samples using the 4 different pipelines, and identified using the NoV typing tool are presented in



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 11

Genotype GI.4[P4] is detected by all pipelines in almost all samples. GI.3[P3] is found with Ifremer's pipeline and EMC's pipelines based on metaSPAdes and SPAdes, but not Genome Detective. GII.6[P7] and GII.1[P33] are only detected with Ifremer's pipeline and EMC'S pipeline with SPAdes and GII.4[P16] is only detected with Ifremer's pipeline and Genome Detective.

Together, these parallel analyses confirm that the highest NoV diversity is observed in the samples with the highest level of contamination, ie the TSS1.1, as detailed in 3.1.3.1 with Ifremer's pipeline. Importantly, the different replicates of the same sample showed similar results when analysed by the same pipeline.

These results also highlight that different pipelines are not equally efficient in assembling and identifying NoV sequences. The pipeline optimized by Ifremer for environmental samples with low levels of contamination tended to retrieve more diversity, but it also identified NoV sequences in the negative control (see 3.1.3.1). The two other in-house pipelines used by EMC identified less strains, but sometimes caught a strain not retrieved by the other ones. Finally, the open, on-line Genome Detective allowed to identify the most frequently assembled strains, ie GI.4[P4] which was probably present in higher concentrations than expected, and GII.4[P16], but missed others (GII.6[P7], GI.3[P3]). This tool may thus be useful for laboratories with few bioinformatics infrastructures to rely on, but it may underestimate the diversity of NoV strains in a sample.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**TABLE 16: DETECTION OF THE 10 EXPECTED STRAINS IN TSS1 IN BOTH DATASETS USING 4 DIFFERENT PIPELINES.**

| Sample | Rep. | GI.1 [P1] | GI.2 [P2] | GI.3 [P3] | GI.4 [P4] | GII.2 [P16] | GII.17 [P17] | GII.3 [P21] | GII.6 [P7] | GII.4 [P31] | GII.4 [P16] | GII.1 [P33] | GII.14 [P7] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TSS1.1 | A | | | | | | | | | | | | |
| TSS1.1 | B | | | | | | | | | | | | |
| TSS1.1 | C | | | | | | | | | | | | |
| TSS1.2 | A | | | | | | | | | | | | |
| TSS1.2 | B | | | | | | | | | | | | |
| TSS1.2 | C | | | | | | | | | | | | |
| TSS1.3 | A | | | | | | | | | | | | |
| TSS1.3 | B | | | | | | | | | | | | |
| TSS1.3 | C | | | | | | | | | | | | |
| NC | - | | | | | | | | | | | | |

Table legend :

| Colour | Pipeline |
|---|---|
| (blue) | Ifremer pipeline on Ifremer data |
| (light purple) | Ifremer pipeline on EMC data |
| (orange) | metaSPAdes-EMC pipeline on EMC data |
| (green) | SPAdes-EMC pipeline on EMC data |
| (pink) | Genome Detective on EMC data |



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

### *Concluding remarks for VirCapSeq on TSS1*

Results from the two laboratories were compatible but different due to the different sequencing platforms and sequencing depths achieved. They showed a marked effect of virus concentration: the diversity of NoV GI and GII was observed using the most concentrated samples, but only a subset of strains was observed in the most diluted ones. The sequencing depth had an impact on this effect as well: a deeper sequencing as performed by Ifremer resulted in reliable estimation of NoV diversity in the samples with high and medium NoV concentration levels while only the highest concentrated sample was well described following less deep sequencing at EMC. It should be mentioned here that such deep sequencing increases costs, and complicates the bioinformatics analysis. Finally, the choice of the bioinformatic pipeline applied on these data had an impact on the diversity of norovirus observed as well. Here, Ifremer's pipeline appeared to detect more NoV strains than the other pipelines, in EMC's dataset.

To further investigate the impact of NoV concentration on the sequencing results in a complex sample matrix like oyster, the TSS2 were sequenced together with TSS1 by the two laboratories.

## 3.1.3.3. Comparative analysis of TSS2 VirCapSeq metagenomics from Ifremer and EMC using Ifremer's pipeline

### *Sequencing quality*

In Ifremer's dataset, between 70M and 130M of raw reads were obtained for each A or C replicate and for the negative DT controls TSS2.0 and TSS2.5 (**Error! Reference source not found.**), but a lower number of reads was observed in the 4 TSS2 replicates from series B (TSS2.1B, TSS2.2B, TSS2.3B and TSS2.4B). In EMC's dataset, between 1M and 4.8M raw reads were obtained per TSS2 replicate. Here, due to the presence of genetic material from the matrix (oyster, bacteria), control DT TSS2.0 and TSS2.5 yielded the same amount of reads as other samples. The quality was good for both runs and a small proportion of reads (0-5%) were trimmed in the first steps of Ifremer's pipeline (Figure 17, red portion of bars), leaving most reads for subsequent analysis (Figure 17, blue portion of bars).

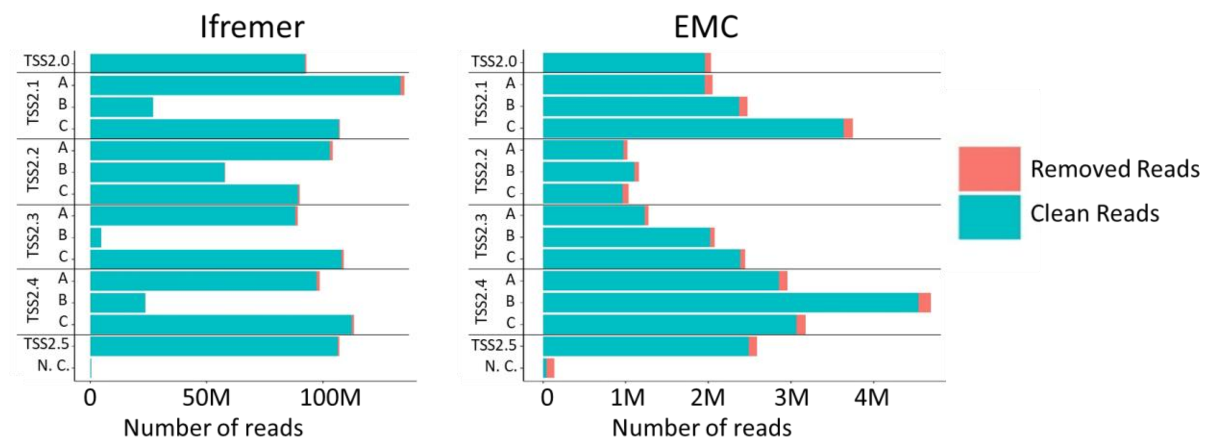



**FIGURE 17: NUMBER OF READS REMOVED FROM RAW READS (RED) AND CLEAN READS KEPT FOR ANALYSIS (BLUE) FOR TSS2**

A, B, C: replicates. N. C.: negative control (water). M: Millions.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

### NoV GI and GII diversity in TSS2 sequenced by Ifremer and EMC and analysed using Ifremer's pipeline

The sequences were analyzed using Ifremer's pipeline described in 2.2.5. A fraction of the NoV contigs generated by the pipeline could be both genotyped and P-typed (dual typing), while for the other, only the genotype, the P-type or the genogroup could be identified (Table 17). One contig with a RdRp-VP1 combination that was absent from the strains selected, was identified as a false recombinant where the RdRp sequence was assembled to the wrong VP1 sequence by the pipeline (



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 11). All other contigs with either P-type or genotype assignation (or both) corresponded to one of the initial strains. Importantly, several large contigs (0.5Kb to 5,5Kb) containing both RdRp and VP1 sequences allowed to identify the strains with both genes, with the expected combination (no generation of chimeras), except for the GII.4[P31] strain.

**TABLE 17: NUMBER, MAXIMUM SIZE AND GENOTYPING OF NOV SEQUENCES (CONTIGS) OBTAINED FOR EACH SAMPLE OF THE TSS1 BY BOTH LABORATORIES**

| Dataset | Sample | Number of NoV contigs >500 bp | Max. contig size | Number of contigs with dual typing | Number of false recombinants |
|---|---|---|---|---|---|
| **Ifremer** | TSS2.0 | 4 | 7193 | 1 | 0 |
| | TSS2.1 | 16 | 7497 | 5 | 0 |
| | TSS2.2 | 14 | 6896 | 3 | 0 |
| | TSS2.3 | 9 | 4759 | 3 | 0 |
| | TSS2.4 | 12 | 2983 | 3 | 0 |
| | TSS.2.5 | 5 | 7333 | 1 | 0 |
| **EMC** | TSS2.0 | 0 | - | 0 | 0 |
| | TSS2.1 | 43 | 7783 | 8 | 0 |
| | TSS2.2 | 7 | 5838 | 4 | 0 |
| | TSS2.3 | 8 | 6270 | 5 | 0 |
| | TSS2.4 | 19 | 2797 | 2 | 0 |
| | TSS2.5 | 0 | - | 0 | 0 |

In Ifremer's dataset, all DT samples yielded a high number of reads assigned to NoV GI (from 4K to 9M), above the number of reads in GI contigs from the negative control (834). For NoV GII, the number of reads kept after analysis are higher than those detected in the negative control (126) in most samples, where it ranged from 9K to 1M and are considered reliable. But a low number of NoV reads were detected in TSS2.3B (38) and TSS2.4B sample replicates (190). This agrees with the lower number of total raw reads for these samples, which should not be considered further.

In EMC's dataset, there were 14 NoV GI reads in contigs passing the thresholds in the water control, and no GII reads. All the other samples yielded more NoV reads (from 536 to 954K for GI, and from 10 to 176K for GII), representing reliable results.

For GI, the two expected strains (GI.4[P4] and GI.1[P1]) were identified in both datasets (Figure 18, A, B), but the first one was found in vast majority in most samples, including the negative DT controls (albeit with a low number of reads), while the latter is only identified in some replicates, and in the most concentrated samples. As for TSS1, this GI.4[P4] strain appears to have been favored during sample or library preparation.

For GII, the four expected strains (GII.4[P31], GII.2[P16], GII.3[P21], GII.17[P17]) were identified (Figure 18 C, D) at the genotype and P-type level, but not in all samples and replicates. The diversity is highest and results more reproducible between replicates in Ifremer's TSS2.1 and TSS2.2 and in EMC's TSS2.1. In the samples with lower NoV concentration levels, we observe some variability between the replicates; more strains are missing and a higher proportion of contigs were only assigned at the genogroup level because they fell outside the typing region.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

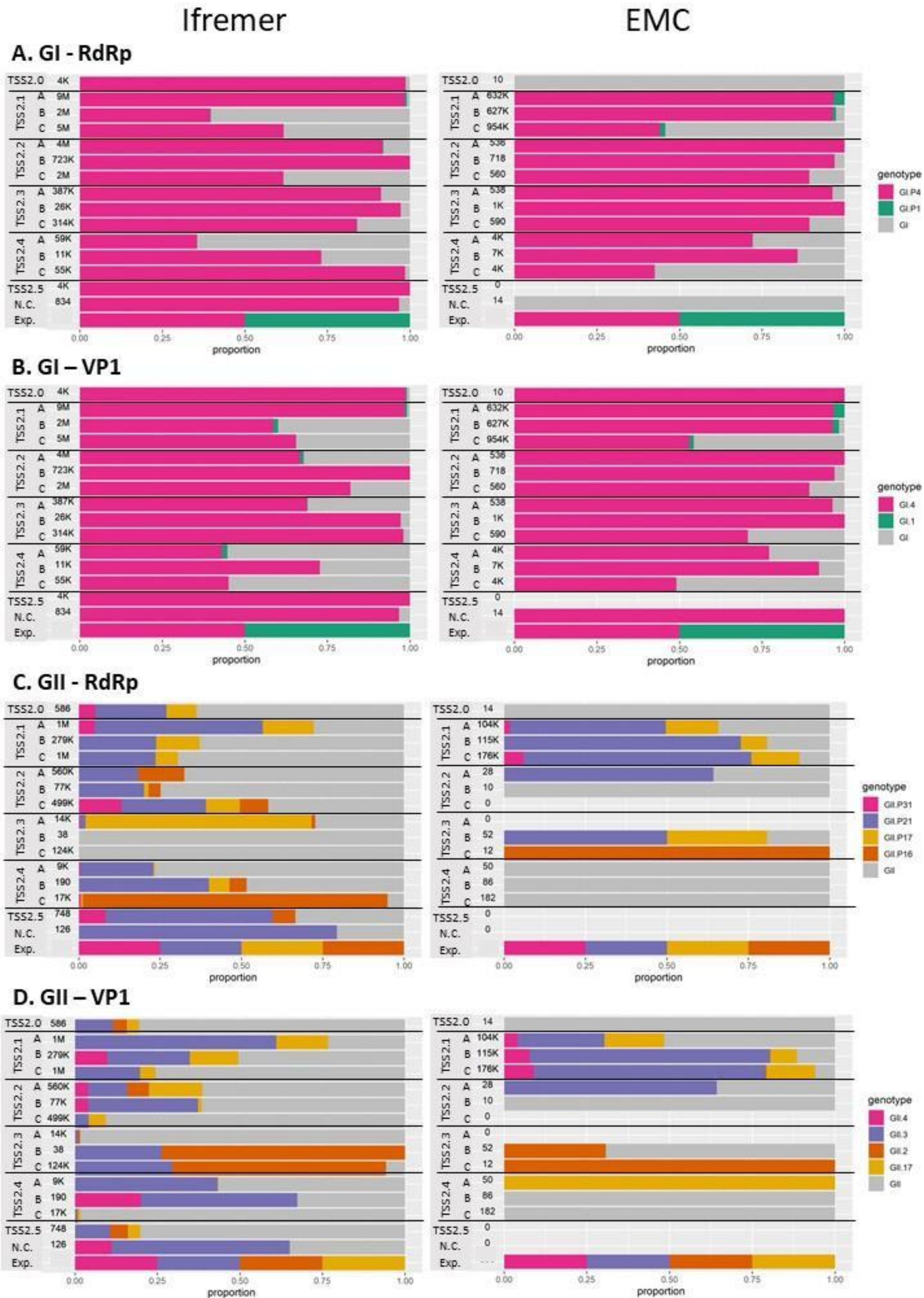



**FIGURE 18: PROPORTION OF READS ASSIGNED TO IDENTIFIED GI OR GII P-TYPES AND GENOTYPES USING IFREMER'S PIPELINE IN IFREMER (LEFT) AND EMC (RIGHT) TSS2 DATA.**



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Among contigs longer than 500 nt and comprising at least 10 reads, those falling in NoV GI – RdRp (A), GI – VP1 (B), GII – RdRp (C) or GII – VP1 (D) regions are assigned to a P-type or genotype and the proportion of their corresponding reads is depicted in colours. Contigs not assigned to a P-type or genotype, but only to a genogroup, are in grey. The four TSS2 samples with high (TSS2.1), medium (TSS2.2), medium-low (TSS2.3) and very-low (TSS2.4) NoV contamination levels were sequenced in triplicates (A, B, C) together with a negative control (N. C., water) and two control DT samples: the background control (TSS2.0) and the oysters used for dilution (TSS2.5). The total number of reads considered per genogroup for each sample and replicate is indicated before the bars (K: kilo, M: mega). Exp: expected proportion.

#### 3.1.3.4. Comparison of datasets and bioinformatics pipelines for the detection of NoV strains in TSS2 by VirCapSeq metagenomics

The VirCapSeq metagenomics data on TSS2 were analysed by Ifremer using an in-house bioinformatics pipeline, and by EMC using two in-house pipelines differing by the de-novo assembly tool (ie. metaSPAdes or SPAdes) and in parallel with the online pipeline Genome Detective, as described in 2.2.5.5. The NoV sequences assembled in these samples using the 4 different pipelines, and identified using the NoV typing tool are presented in



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 11.

For genotypes GI.4[P4], GI.1[P1], GII.17[P17] and GII.3[P21], both the polymerase and capsid genes are found in all datasets with all bioinformatics pipelines. GII.4[P31] is only found complete with SPAdes in one sample. GII.2[P16] is only found complete with Genome detective in three samples.

Like for TSS1, the highest NoV diversity is observed in the samples with the highest level of contamination, ie the TSS2.1, especially for the data generated by EMC, as detailed in 3.1.3.3 with Ifremer's pipeline. The impact of a deeper sequencing as performed by Ifremer is also clearly visible since these data allowed to identify more NoV strains. Importantly, the different replicates of the same sample showed similar results when analysed by the same pipeline, but some differences could be observed when one strain was detected in one replicate but not the others, highlighting the need for replicates to increase the viral diversity.

Here, the different pipelines detected the same strains and their efficiency in assembling and identifying NoV sequences were similar on EMC's dataset. The two in-house pipelines used by EMC and based on metaSPAdes and SPAdes tended to identify less strains except in samples with high levels of contamination, with only GI.4[P4] being assembled in TSS2.2, 2.3 and 2.4. Finally, both Ifremer's pipeline and the open, on-line Genome Detective identified several strains, including GI.4[P4] but also GII.2[P16] or GII.3[P21], in most samples except those with low or very low contamination levels (TSS2.3, TSS2.4).



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

### TABLE 18: DETECTION OF THE 10 EXPECTED STRAINS IN TSS2 IN BOTH DATASETS USING 4 DIFFERENT PIPELINES.

| Sample | Rep. | GI.1 [P1] | GI.4 [P4] | GII.2 [P16] | GII.17 [P17] | GII.3 [P21] | GII.4 [P31] |
|---|---|---|---|---|---|---|---|
| TSS2.1 | A | | | | | | |
| TSS2.1 | B | | | | | | |
| TSS2.1 | C | | | | | | |
| TSS2.2 | A | | | | | | |
| TSS2.2 | B | | | | | | |
| TSS2.2 | C | | | | | | |
| TSS2.3 | A | | | | | | |
| TSS2.3 | B | | | | | | |
| TSS2.3 | C | | | | | | |
| TSS2.4 | A | | | | | | |
| TSS2.4 | B | | | | | | |
| TSS2.4 | C | | | | | | |
| TSS2.5 | C | | | | | | |
| TSS2.0 | - | | | | | | |
| NC | - | | | | | | |



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table legend :

| | |
|---|---|
| (blue) | Ifremer pipeline on Ifremer data |
| (light purple) | Ifremer pipeline on EMC data |
| (orange) | metaSPAdes-EMC pipeline on EMC data |
| (green) | SPAdes-EMC pipeline on EMC data |
| (pink) | Genome Detective on EMC data |

***Concluding remarks for VirCapSeq metagenomics on TSS2 based on results of two consortium partners***

As for TSS1, results from the two laboratories show the impact of NoV dilution combined with sequencing depth. The expected diversity of NoV GI and GII was observed using the most concentrated samples, but only a subset of strains was observed in the most diluted ones, with deeper sequencing by Ifremer resulting in reliable sequencing of TSS2.1 and 2.2. whereas only TSS2.1 yielded contigs with high amounts of reads and diverse NoV GI and GII genotypes in all replicates for EMC. On the TSS2, the impact of the bioinformatics pipeline was not as important as on TSS1. Ifremer's pipeline and the online tool GenomeDetective tended to retrieve more diversity and showed similar results. For automation reasons, Ifremer's pipeline was selected for the analysis of VirCapSeq metagenomics on the BLS and OB samples (3.3, 3.4).

Importantly, several large contigs (0.5Kb to 5,5Kb) containing both RdRp and VP1 sequences allowed to identify the strains with both genes, with the expected combination (no generation of chimeras) expect for one false recombinant. However, even in the samples with the highest concentration level (TSS2.1), there were contigs comprising only RdRp or VP1. In a natural sample, this would impair the identification of recombinant strains.

### 3.1.4. Long amplicon sequencing by Oxford Nanopore Technology

The results of the preparation and validation of libraries are presented in the Appendix D.

The TSS1 and TSS2 libraries were quantified and sequenced using long amplicon sequencing with Oxford Nanopore Technology. Between 26K and 960K raw reads per sample were obtained.

Samples were analysed through Genome Detective https://www.genomedetective.com.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 19 shows the results of the Genome Detective analysis. All samples where we observed a specific band after PCR yielded norovirus sequences after GridION sequencing (Figure D 1 in the Appendix D). For TSS1, all GII specific PCRs were negative, for the GI specific PCRs only the MON432-G1SKR gave bands and GI.4[P4] was detected. For TSS2, both the GI and GII Noronet typing primers and MON432-G1SKR resulted in bands. For GI only GI.4[P4] was detected and no other GI genotypes. It was only possible to detect multiple GII genotypes when using the Noronet typing primers in combination with TSS2.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

### Table 19: Genome Detective results of TSS1 and TSS2 samples sequenced with long amplicon sequencing by Oxford Nanopore Technology.

All genotypes present in the samples are shown in the table and all genotypes detected through long amplicon sequencing by Oxford Nanopore Technology are shown in green.

| Noronet typing primers | | | | | | | |
|---|---|---|---|---|---|---|---|
| Sample | Concentration | GI.P1 GI.1 | GI.P4 GI.4 | GII.P31 GII.4 | GII.P16 GII.2 | GII.P17 GII.17 | GII.P21 GII.3 |
| TSS1.1 | 1000 (gc/µl) | | | | | | |
| TSS1.1 | 1000 (gc/µl) | | | | | | |
| TSS1.1 | 1000 (gc/µl) | | | | | | |
| TSS1.2 | 100 (gc/µl) | | | | | | |
| TSS1.2 | 100 (gc/µl) | | | | | | |
| TSS1.2 | 100 (gc/µl) | | | | | | |
| TSS1.3 | 10 (gc/µl) | | | | | | |
| TSS1.3 | 10 (gc/µl) | | | | | | |
| TSS1.3 | 10 (gc/µl) | | | | | | |
| Negative control | | | | | | | |
| TSS2.1.1 | High | | | | | | |
| TSS2.2.1 | Medium | | | | | | |
| TSS2.3.1 | Low | | | | | | |
| TSS2.1.2 | High | | | | | | |
| TSS2.2.2 | Medium | | | | | | |
| TSS2.3.2 | Low | | | | | | |
| TSS2.1.3 | High | | | | | | |
| TSS2.2.3 | Medium | | | | | | |
| TSS2.3.3 | Low | | | | | | |
| Negative control | | | | | | | |

| MON432-G1SKR (GI) MON431-G2SKR (GII) | | | | | | | |
|---|---|---|---|---|---|---|---|
| Sample | Concentration | GI.P1 GI.1 | GI.P4 GI.4 | GII.P31 GII.4 | GII.P16 GII.2 | GII.P17 GII.17 | GII.P21 GII.3 |
| TSS1.1 | 1000 (gc/µl) | | | | | | |
| TSS1.1 | 1000 (gc/µl) | | | | | | |
| TSS1.1 | 1000 (gc/µl) | | | | | | |
| TSS1.2 | 100 (gc/µl) | | | | | | |
| TSS1.2 | 100 (gc/µl) | | | | | | |
| TSS1.2 | 100 (gc/µl) | | | | | | |
| TSS1.3 | 10 (gc/µl) | | | | | | |
| TSS1.3 | 10 (gc/µl) | | | | | | |
| TSS1.3 | 10 (gc/µl) | | | | | | |
| Negative control | | | | | | | |
| TSS2.1.1 | High | | | | | | |
| TSS2.2.1 | Medium | | | | | | |
| TSS2.3.1 | Low | | | | | | |
| TSS2.1.2 | High | | | | | | |
| TSS2.2.2 | Medium | | | | | | |
| TSS2.3.2 | Low | | | | | | |



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

| TSS2.1.3 | High | | | | | | |
|---|---|---|---|---|---|---|---|
| TSS2.2.3 | Medium | | | | | | |
| TSS2.3.3 | Low | | | | | | |
| Negative control | | | | | | | |

Gc µL: genome copies per microliter

### *Concluding remarks for long amplicon sequencing using Oxford Nanopore Technology on TSS1 and TSS2*

If a band was obtained through PCR (Figure D 1 in Appendix D) we were able to retrieve NoV sequences. However, using this method we were unable to obtain a band for all samples and if we had a band we were unable to detect all NoV genotypes included in the sample.

## 3.2. Metabarcoding on BLS samples

Two hundred samples were chosen (as described in 2.1.3) to constitute the main list of samples to be sequenced by Ifremer (102 samples) and Cefas (98 samples).

### 3.2.1. Generation of amplicons from samples in the main and reserve lists

Two hundred samples from the main list were chosen for initial analysis as described in 2.1.3; subsets of these samples were analysed at Ifremer (102 samples) and Cefas (98 samples). Among the 102 samples in the main list analysed by Ifremer, 94 yielded at least one amplicon following application of the four N-PCR assays (GI-VP1, GI-RdRp, GII-VP1 and GII-RdRp) including retesting of any samples that were negative on initial testing. Therefore, 8 samples from the reserve list were tested; these all yielded at least one amplicon, allowing 102 samples to be submitted for sequencing as planned. Among the 98 samples in the main list analysed at Cefas, 79 yielded at least one amplicon upon initial testing. For reasons of time, retesting of the 19 samples from the main list and testing of 19 samples from the reserve list (to guarantee amplicons from 98 samples in total in the event that all main list sample retesting produced negative results) was carried out simultaneously instead of in sequence as originally planned and described in the retesting algorithm (2.2.4.3). After this retesting 93 main list samples and 17 reserve list samples provided at least one amplicon. Although the total number of samples with amplicons (110) was as a result greater than the original goal of 98 BLS samples submitted for sequencing from Cefas, it was decided to submit all amplicons for sequencing rather than limiting the number of reserve list samples to artificially preserve the original target number. In total, across both laboratories, 212 samples (102 from Ifremer and 110 from Cefas) yielded at least one amplicon and were submitted for sequencing (an increase compared with the original target of 200 samples). After replacement of main list samples with reserve list samples, two countries provided >50 samples for sequencing in total (maximum 58) compared with the maximum criteria of 50 for inclusion in the main list as described in 2.1.3.2, and 5 sampling sites provided >5 samples for sequencing in total (maximum 8) compared with the maximum criteria of 5 for inclusion in the main list as described in 2.1.3.2. Due to negative results during the metabarcoding N-PCR, three countries contributed <5 samples for sequencing (2, 3 and 4 samples respectively). The potential for these selection criteria for the main list to be exceeded/not met as a result of replacement of negative main list samples with reserve list samples was foreseen and was a result of the need to prioritise samples with high levels of virus during the retesting phase in order to ensure at least 200 samples were submitted for sequencing. The total amplicon numbers obtained for sequencing are presented in **Error! Reference source not found.**. The library preparation is described in Appendix E.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**TABLE 20. THE BLS AMPLICONS OBTAINED FOR METABARCODING SEQUENCING**

| N-PCR assay | Main list samples tested using N-PCR | Main list samples yielding amplicons | Reserve list samples yielding amplicons | Total samples yielding amplicons |
|---|---|---|---|---|
| **GI RdRp** | 166 | 144 | 16 | 160 |
| **GI VP1** | 166 | 158 | 18 | 176 |
| **GII RdRp** | 191 | 155 | 16 | 171 |
| **GII VP1** | 191 | 162 | 16 | 178 |
| **TOTAL** | **200** | **187** | **25** | **212** |

Following Illumina sequencing and bioinformatics analysis, at least one NoV sequence was obtained for 201 out of the 212 samples submitted (94.8%). Ten samples from the main list and one sample from the reserve list did not yield any NoV sequence. Two of these samples were initially positive for both GI and GII, one was initially positive for GI only, and seven were initially positive for GII only. For the four N-PCR assays applied, higher proportions of samples yielded no sequence for the GII-specific assays; for GI RdRp and GI VP1 155 out of 160 samples (96.9%) and 167 out of 176 samples (94.9%) respectively yielded at least one NoV sequence, for GII RdRp and GII VP1 136 out of 171 samples (79.5%) and 142 out of 178 samples (79.8%) respectively yielded at least one NoV sequence. This higher rate of failure for GII assays may be due in part to a higher loss of reads during bioinformatics analysis for GII VP1 and GII RdRp amplicons from Ifremer.

### 3.2.2. Molecular characterisation and genetic diversity of NoV strains in BLS samples set

#### 3.2.2.1. Overview of the diversity of NoV strains and genotypes identified in BLS

The analysis that was performed on the BLS samples yielded a total of 242 unique sequences ("clusters") across the four different N-PCR assays, of which 24 clusters were detected in samples submitted by both Cefas and Ifremer. The majority of clusters (129/242; 53.3%) were detected in more than one sample, and the most commonly found cluster was found in 110 different samples from 7 different countries, more than half of the total samples submitted for testing (110/212; 51.9%). For each assay, each detected cluster was assigned to a NoV genogroup and genotype/RdRp type using the Norovirus Typing Tool 2.0, plus additional tools if necessary as described in 2.2.4.4 (**Error! Reference source not found.**, **Error! Reference source not found.**, **Error! Reference source not found.**).



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**TABLE 21: NUMBER OF NOV CLUSTERS OBTAINED PER ASSAY PER NOV GENOGROUP.**

| Assay | Genogroup | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | GI | GII | GIII | GIV | GV | GVIII | GNA1 | nd* | Total |
| GI RdRp | 51(5)** | 1*** | 3 | - | - | - | - | 1**** | 56 |
| GI VP1 | 57(7)** | - | - | - | - | - | 2 | - | 59 |
| GII RdRp | - | 49(7)** | - | 1 | 1 | - | - | - | 51 |
| GII VP1 | - | 74(5)** | - | 1 | - | 2 | - | - | 77 |

* nd; NoV genogroup not determined.
** numbers of clusters found simultaneously in both Ifremer and Cefas datasets shown in parentheses.
*** a single GII cluster was found in one sample using the GI RdRp assay, this was identical in the region of overlap to a cluster found in the same sample using the GII RdRp assay; it is not considered a unique sequence and not considered in subsequent analyses
**** one cluster obtained using the GI RdRp assay was identified as NoV on the basis of phylogenetic analysis, but could not be definitively identified to genogroup level (closest matches GVII and GNA1).

For most clusters, the genogroup determined was the one targeted by the assay in question, e.g. 51/56 clusters (91.1%) detected using the GI RdRp assay were characterised as GI, however a small number of norovirus clusters from heterologous genogroups were identified. The GI VP1 assay yielded two clusters that were assigned to the tentative new genogroup, GNA1 (Chhabra et al., 2019), identified on the basis of a single whole genome sequence isolated from a Harbour Porpoise (de Graaf et al., 2016). Using GI RdRp, we retrieved one GII.P17 cluster (identical in the overlapping region to a cluster identified in the same sample using the GII RdRp assay and therefore not considered a unique sequence in subsequent analyses), three GIII (bovine NoV) clusters and one cluster (found in a single sample) assigned as GVII by the Norovirus Typing Tool 2.0 but most closely related by BLASTn to GNA1; although conclusively identified as NoV using phylogenetic analysis, definitive identification of this obtained cluster to genogroup was not possible. It may represent a rare NoV genogroup from a previously unsampled host. Using GII RdRp, one GIV (feline NoV) and one GV (rodent NoV) cluster were identified in addition to GII RdRp types, and with GII VP1, one GIV (canine NoV) and two GVIII (human NoV) clusters were identified. Identification of sequences from these less well-known NoV genogroups demonstrates the wide range of NoV circulating in the marine environment, but likely only represents a small sample of the true diversity, as the primers used in the N-PCR assays are designed specifically to amplify GI and GII sequences, and detection of other genogroups is only possible as a result of non-intentional cross-reactivity due to primer binding to non-target genome sequences.

In addition to the range of genogroups detected, considerable diversity within genogroups was identified, with 11 GI RdRp types, 7 GI genotypes (determined by VP1 sequence), 8 GII RdRp types and 11 GII genotypes found across all samples. Diversity within the types was also noted, with up to 15 different clusters identified for a single RdRp type (GII.P31; **Error! Reference source not found.**) and 19 clusters for a single genotype (GI.4; **Error! Reference source not found.**).



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

### TABLE 22: NOV GENOGROUP AND RDRP TYPE ASSIGNATION OF RDRP CLUSTERS IDENTIFIED

| Genogroup | RdRp type | Number of clusters |
|---|---|---|
| **GI** | GI.P1 | 1 |
| | GI.P2 | 14 |
| | GI.P3 | 2 |
| | GI.P4 | 9 |
| | GI.P5 | 1 |
| | GI.P6 | 2 |
| | GI.P7 | 7 |
| | GI.P9 | 1 |
| | GI.P11 | 4 |
| | GI.P13 | 9 |
| | nd* | 1 |
| **GII** | GII.P4 | 10 |
| | GII.P7 | 2 |
| | GII.P8 | 1 |
| | GII.P12 | 2 |
| | GII.P16 | 2 |
| | GII.P17 | 4 |
| | GII.P21 | 13 |
| | GII.P31 | 15 |
| **GIII** | nd | 3 |
| **GIV** | nd | 1 |
| **GV** | nd | 1 |
| **nd**** | nd | 1 |
| **TOTAL** | **23** | **106** |

nd: not determined.
* one GI RdRp cluster could not be definitively identified to RdRp type
** one cluster obtained using the GI RdRp assay was identified as NoV on the basis of phylogenetic analysis, but could not be definitively identified to genogroup level (closest matches GVII and GNA1).



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**TABLE 23: NOV GENOGROUP AND GENOTYPE ASSIGNATION OF VP1 CLUSTERS IDENTIFIED.**

| Genogroup | Genotype | Number of clusters |
|---|---|---|
| **GI** | GI.1 | 5 |
| | GI.2 | 6 |
| | GI.3 | 15 |
| | GI.4 | 19 |
| | GI.5 | 2 |
| | GI.6 | 5 |
| | GI.7 | 5 |
| **GII** | GII.1 | 1 |
| | GII.2 | 5 |
| | GII.3 | 17 |
| | GII.4 | 18 |
| | GII.6 | 16 |
| | GII.7 | 4 |
| | GII.8 | 1 |
| | GII.12 | 1 |
| | GII.13 | 5 |
| | GII.17 | 5 |
| | GII.18 | 1 |
| **GIV** | nd | 1 |
| **GVIII** | GVIII.1 | 2 |
| **GNA1** | nd | 2 |
| **TOTAL** | **21** | **136** |

nd: not determined.

### 3.2.2.2. Identification of multiple NoV strains per BLS sample.

Combining all targets, most samples allowed the detection of multiple NoV sequences in the same sample, with up to 18 different NoV sequences in one sample initially positive for both GI and GII (Figure 19, A). For each assay (GI RdRp, GI VP1, GII RdRp and GII VP1), most samples yielded between 0 and 2 different clusters, however up to 8 different clusters per assay were identified in some samples (Figure 19, B, C, D, E). Loss of reads during GII analysis at Ifremer may explain the higher number of samples without clusters for the two GII-specific assays, compared with the GI-specific assays.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

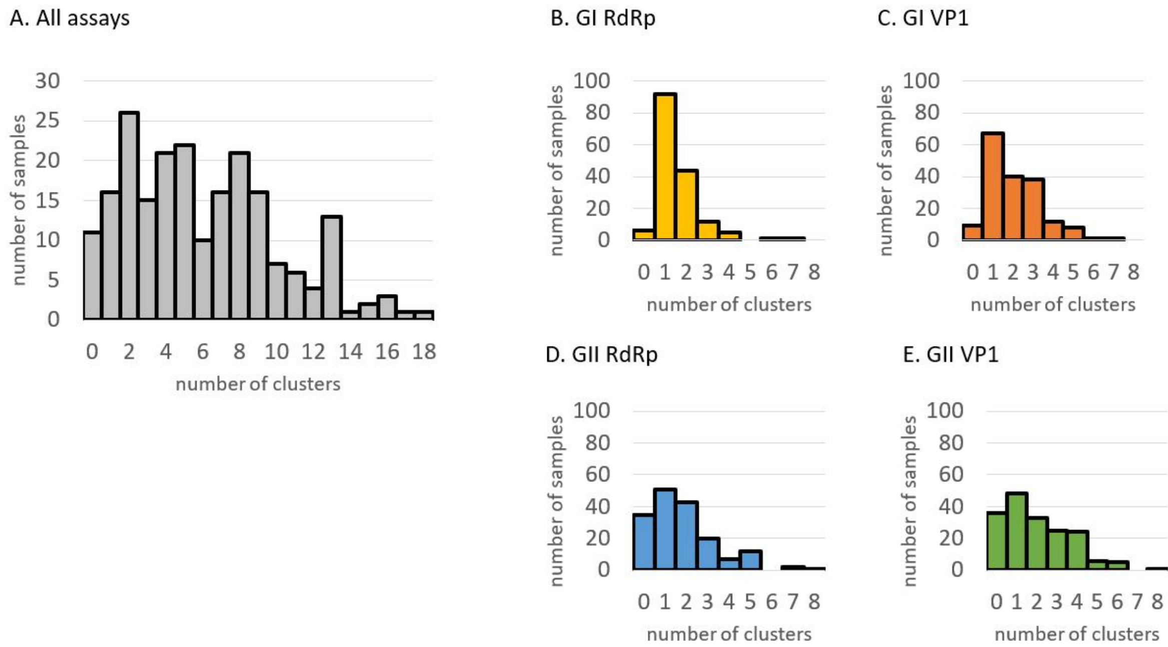



**FIGURE 19: DISTRIBUTION OF NOV CLUSTER NUMBERS IN BLS SAMPLES.**

A. shows a histogram of the number of samples (including all samples submitted for sequencing) containing a given number of clusters taking into account all 4 assays. B-E show equivalent histograms for each assay individually. In these cases, samples that were submitted for sequencing but only for different assays are not included in the counts; samples shown as containing zero clusters are those where sequencing was attempted for the relevant assay but not successful.

Following genogroup and genotype/RdRp type assignation, these clusters represented up to 5 different genotypes/RdRp types per sample and per assay, with up to 7 RdRp types and 9 genotypes found in a single sample when including sequencing of amplicons from both GI and GII assays (Figure 20).

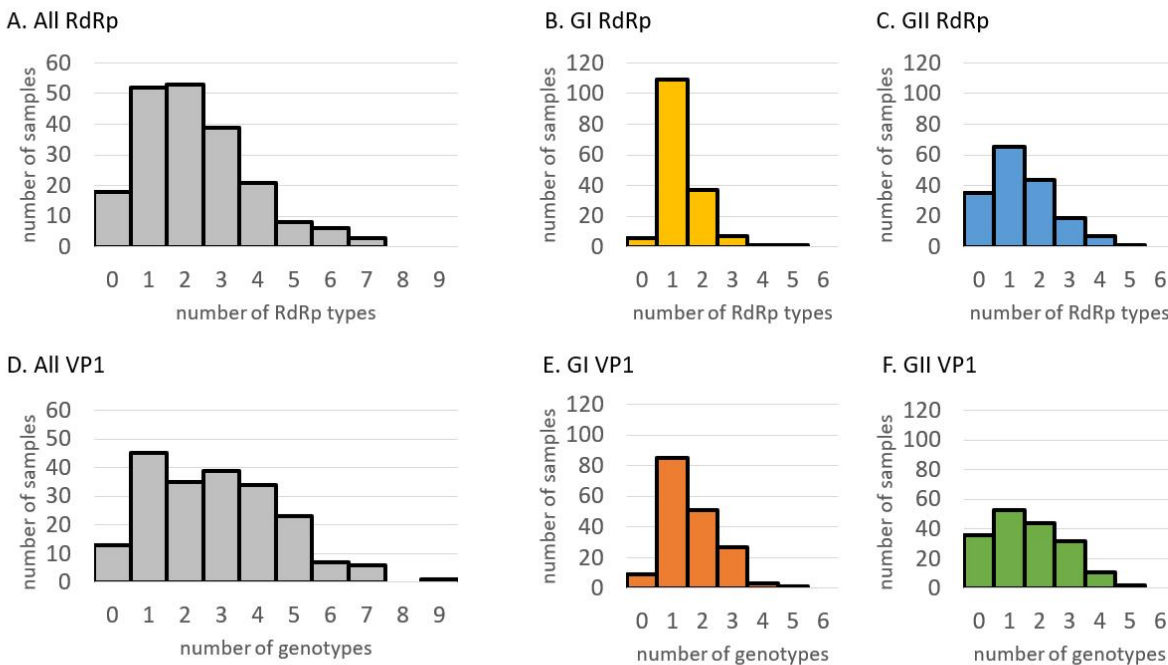



**FIGURE 20: DISTRIBUTION OF NOV GENOTYPE AND P-TYPE NUMBERS IN BLS SAMPLES.**



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

A. shows a histogram of the number of samples (including all samples where sequencing of RdRp amplicons was carried out) containing a given number of RdRp types taking into account both GI and GII assays. B-C show equivalent histograms for each assay individually. D. shows a histogram of the number of samples (including all samples where sequencing of VP1 amplicons was carried out) containing a given number of genotypes taking into account both GI and GII assays. E-F show equivalent histograms for each assay individually. In all cases, samples that were submitted for sequencing but only for different assays are not included in the counts; samples shown as containing zero clusters are those where sequencing was attempted for the relevant assay(s) but not successful.

### 3.2.2.3. Comparison of sequences obtained by Cefas and Ifremer

For technical and logistical reasons, the BLS metabarcoding analysis was split between two laboratories, Cefas and Ifremer. The raw sequencing data that were generated were analysed separately (using identical parameters). They could not be analysed together since the tagging and indexing strategies, the run quality and resulting sequencing depth were different. Nevertheless, the most commonly identified clusters, (those which were identified in the most different samples) were often the same in the two datasets. For instance, for the GI VP1 assay, the most commonly found cluster obtained in samples analysed by Cefas (a genotype GI.4 cluster, present in 66 samples) was identical to the most commonly found cluster obtained in samples analysed by Ifremer (present in 44 samples). Across all 4 assays, 15 out of the 20 most commonly found clusters were found in samples analysed at both Cefas and Ifremer. This suggests that the analysis is robust and metabarcoding carried out in different laboratories using similar approaches and bioinformatics pipelines can simultaneously identify widely circulating dominant strains.

### 3.2.2.4. Comparison of sequences obtained in different European countries

During the BLS, oyster samples were collected from 13 different European countries (all EU/EFTA Member States at the start of the survey). Samples from 12 of those countries were selected for analysis in this project (one country did not provide any positive results during the BLS so all its samples were excluded). Of these 12 countries, 6 provided 10 or more samples for sequencing; the following analysis is limited to these 6 countries to avoid biases due to small sample numbers. For each of these countries, the proportions of samples (amongst those samples that yielded at least one NoV sequence) containing each RdRP type and genotype identified is shown for each of the four assays in Figures 21, 22, 23 and 24.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

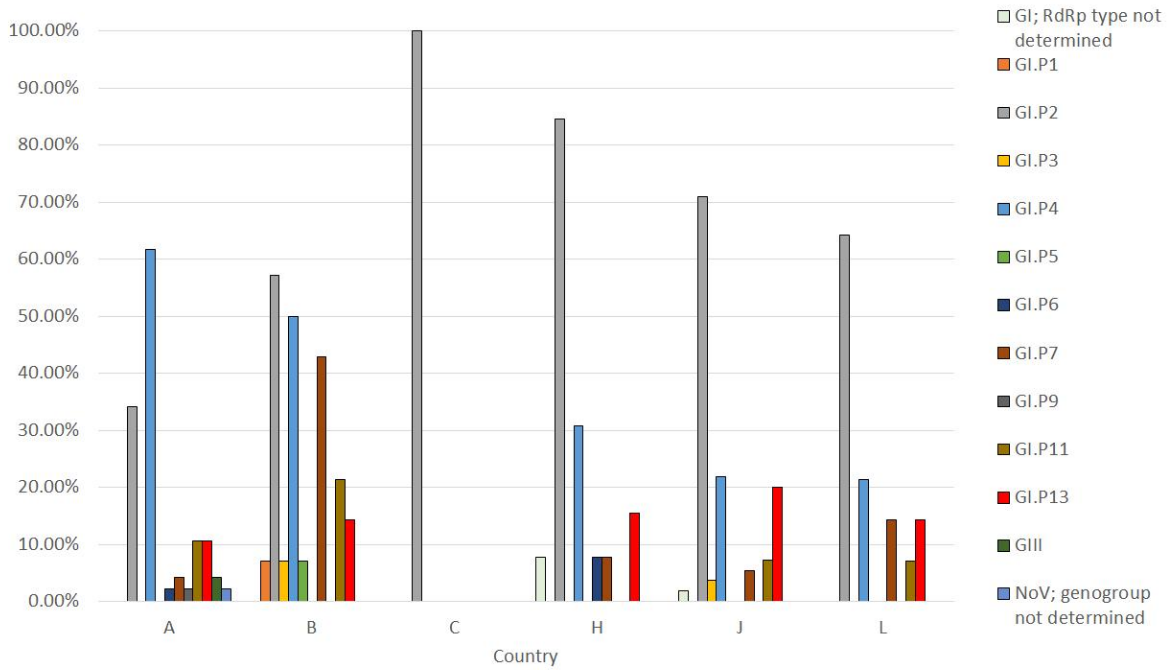



**FIGURE 21: PROPORTION OF SEQUENCED SAMPLES FROM DIFFERENT COUNTRIES CONTAINING DIFFERENT GI RDRP TYPES.**

Countries are identified using anonymised codes (A,B,C,H,J,L). Only samples providing at least one NoV sequence using amplicons from the GI RdRp assay are considered in calculation of the proportions for each country.

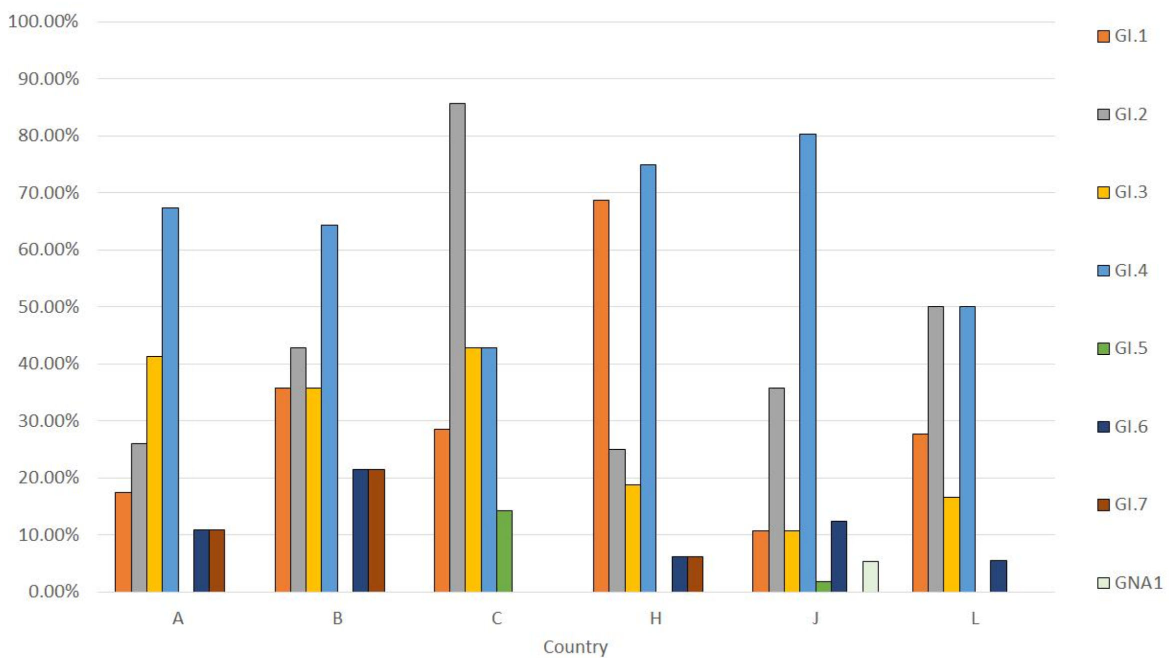



**FIGURE 22: PROPORTION OF SEQUENCED SAMPLES FROM DIFFERENT COUNTRIES CONTAINING DIFFERENT GI GENOTYPES.**

Countries are identified using anonymised codes (A,B,C,H,J,L). Only samples providing at least one NoV sequence using amplicons from the GI VP1 assay are considered in calculation of the proportions for each country.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

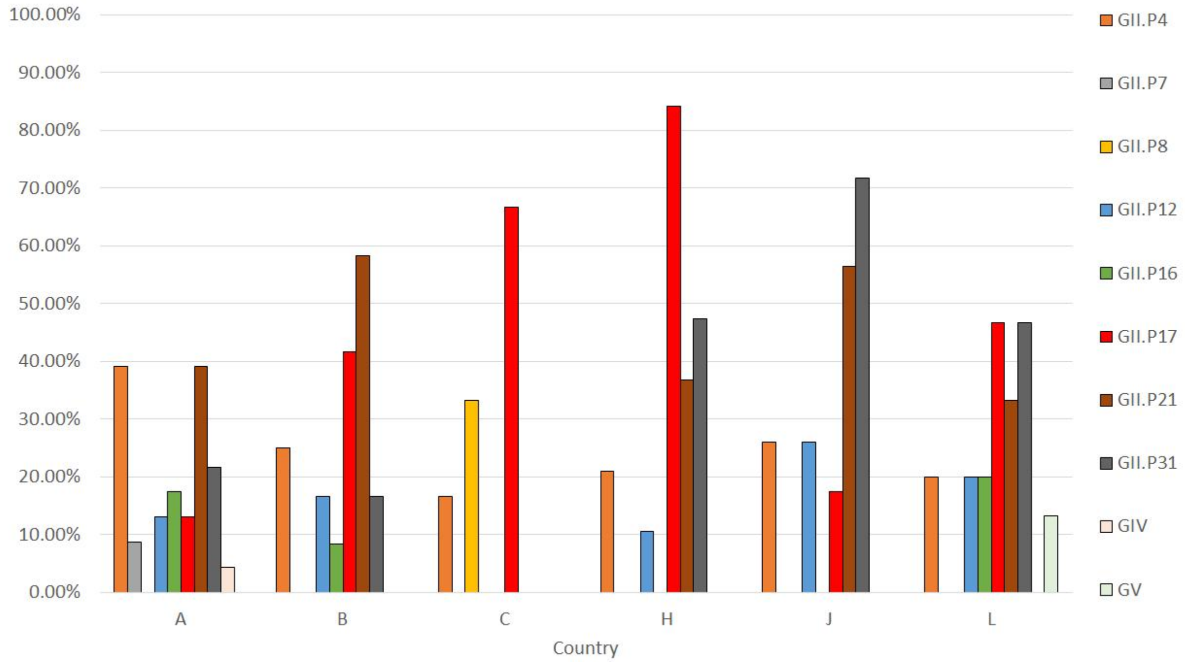



**FIGURE 23: PROPORTION OF SEQUENCED SAMPLES FROM DIFFERENT COUNTRIES CONTAINING DIFFERENT GII RDRP TYPES.**

Countries are identified using anonymised codes (A,B,C,H,J,L). Only samples providing at least one NoV sequence using amplicons from the GII RdRp assay are considered in calculation of the proportions for each country.

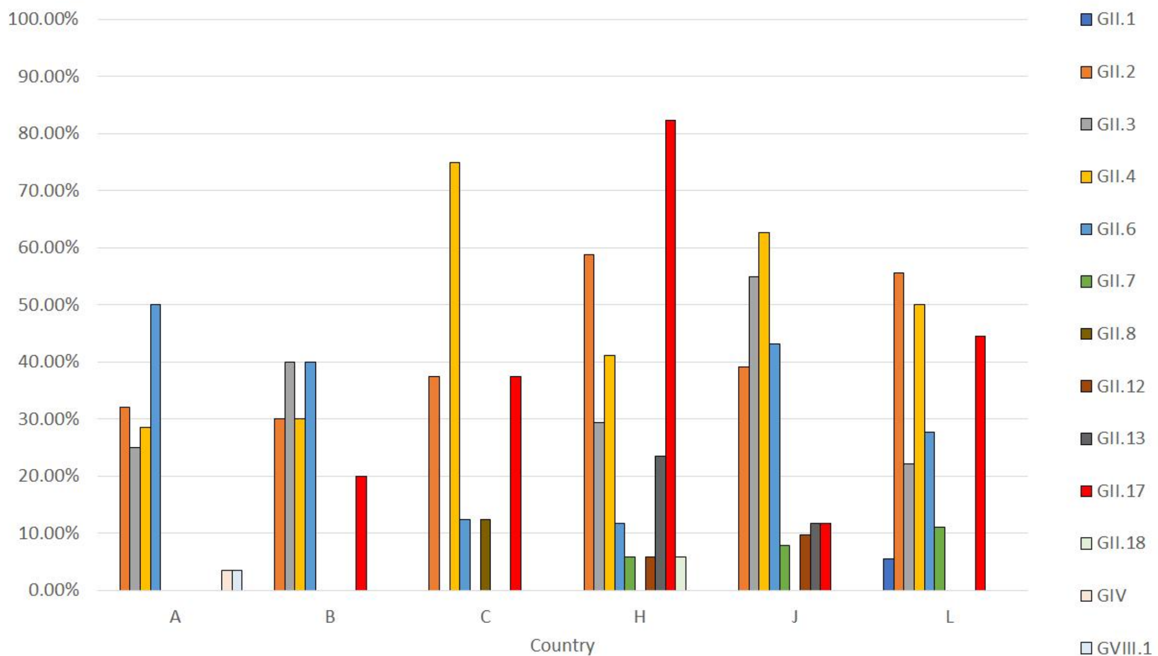



**FIGURE 24: PROPORTION OF SEQUENCED SAMPLES FROM DIFFERENT COUNTRIES CONTAINING DIFFERENT GII GENOTYPES.**

Countries are identified using anonymised codes (A,B,C,H,J,L). Only samples providing at least one NoV sequence using amplicons from the GII VP1 assay are considered in calculation of the proportions for each country.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

For GI RdRp (Figure 21), the proportion of samples containing GI.P2 ranges from 34.0% (Country A) up to 84.6% and 100% (Countries H and C, respectively). GI.P4 was not present in samples from Country C, and found in only 21.4% and 21.8% of samples from Countries J an L respectively, but was found in 61.7% of samples from Country A. GI.P7 again was absent from samples from Country C, and found in only 4.3% of samples from Country A, but was present in 42.9% of samples from Country B.

For GI VP1 (Figure 22) GI.4 was the most commonly found genotype, present in at least 42.9% of samples from all 6 countries examined. Other common genotypes included GI.1, GI.2 and GI.3 however these were all variable in their prevalence in different countries. The proportion of samples from Country J containing GI.1 was 10.7% compared with 68.8% from Country H. GI.2 was found in 25.0% of samples from Country H, but 85.7% of samples from Country C, while GI.3 was found in 10.7% of samples from Country J but 42.9% of samples from Country C.

For GII RdRp (Figure 23) commonly found RdRp types included GII.P4, GII.P17, GII.P21 and GII.P31. Proportions of samples containing GII.P17 ranged from 13.0% from Country A to 84.2% from Country H. GII.P21 was not found in samples from Country C, but otherwise found in between 33.3% (Country C) and 58.3% of samples (Country B). GII.P31 was found in 16.7% of samples from Country B but 47.4% of samples from Country J.

For GII VP1 (Figure 24) several genotypes were commonly found in multiple countries. Prevalence of GII.2 was variable, ranging between 30.0% (Country B) to 58.8% (Country H). GII.3 was present in most samples (54.9%) from Country J, but not found in samples from Country C, while GII.4 was present in 75.0% of samples from Country C but only 28.6% of samples from Country A. Genotype GII.6 was commonest in samples from Country A (50.0%) and least common in samples from Country H (11.8%) whereas GII.17 was absent in samples from Country A but present in the large majority (82.4%) of samples from country H.

The BLS was designed to analyse NoV prevalence and levels in oysters at a Europe-wide scale; no analysis of prevalence and levels in different countries was included in the final report on the European BLS. For these reasons it is necessary to maintain anonymity in terms of contributing countries in this report. It is therefore not possible to carry out some potentially informative geographical level analyses (comparison of sequences from oyster samples with sequences from clinical samples in the same country; comparison of sequences from oyster samples using broad geographical descriptors e.g. Northern vs. Southern Europe or from samples collected from different sea areas). Nevertheless Figures 19-22 highlight differences in patterns of RdRp type and genotype prevalence between different contributing countries. Unsurprisingly, rare types may only be present in samples from a minority of countries, however, proportions of commoner types vary widely between different countries.

#### 3.2.2.5. Comparison of sequences obtained in different time periods

The sampling element of the Baseline survey was carried out over a two-year period (November 2016 – October 2018). For the purposes of the analysis presented below, this was subdivided into 4 separate shorter 6-month periods, approximately equivalent to two consecutive winters (November 2016 – April 2017 and November 2017 – April 2018) and two consecutive summers (May – October 2017 and May – October 2018).

For each of these four periods, the proportions of samples (amongst those samples that yielded at least one NoV sequence) containing each RdRP type and genotype identified is shown for each of the four assays in Figures 25, 26, 27 and 28.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

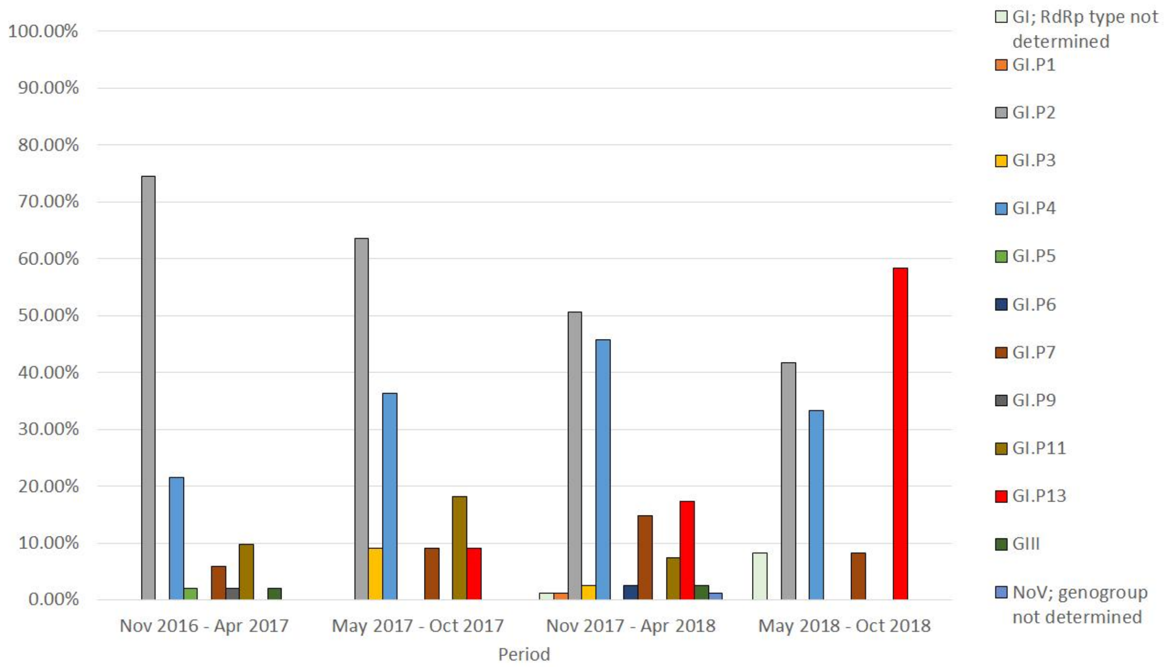



**Figure 25: Proportion of sequenced samples from different periods containing different GI RdRp types.**

Only samples providing at least one NoV sequence using amplicons from the GI RdRp assay are considered in calculation of the proportions for each period.

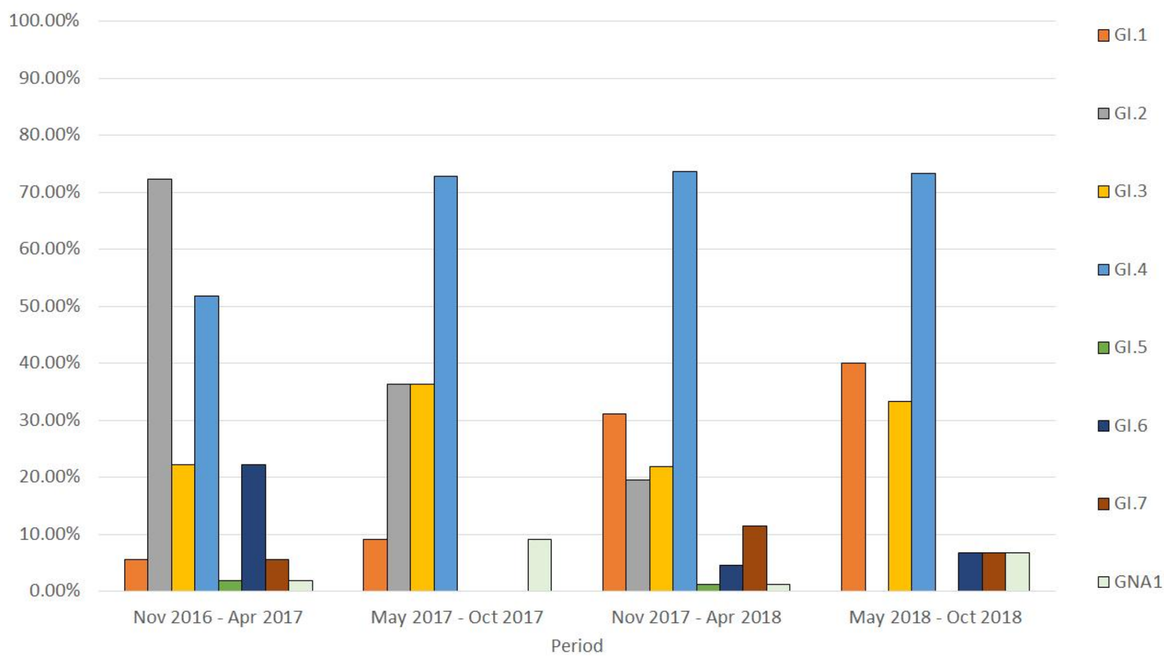



**Figure 26: Proportion of sequenced samples from different periods containing different GI genotypes.**

Only samples providing at least one NoV sequence using amplicons from the GI VP1 assay are considered in calculation of the proportions for each period.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

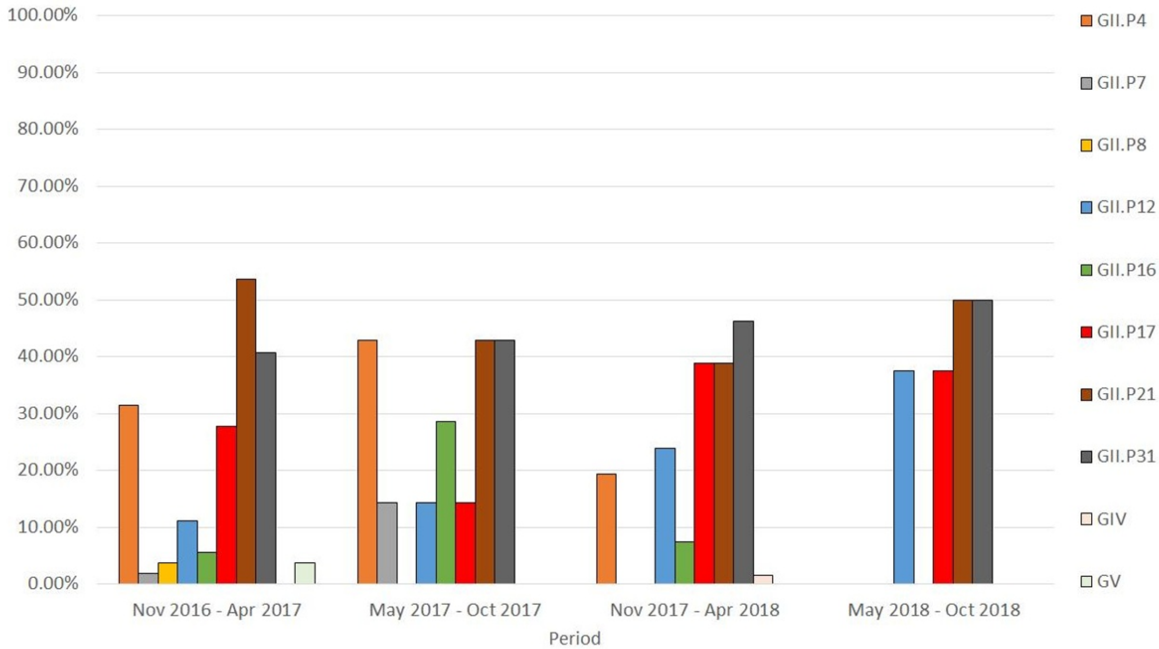



**FIGURE 27: PROPORTION OF SEQUENCED SAMPLES FROM DIFFERENT PERIODS CONTAINING DIFFERENT GII RDRP TYPES.**

Only samples providing at least one NoV sequence using amplicons from the GII RdRp assay are considered in calculation of the proportions for each period.

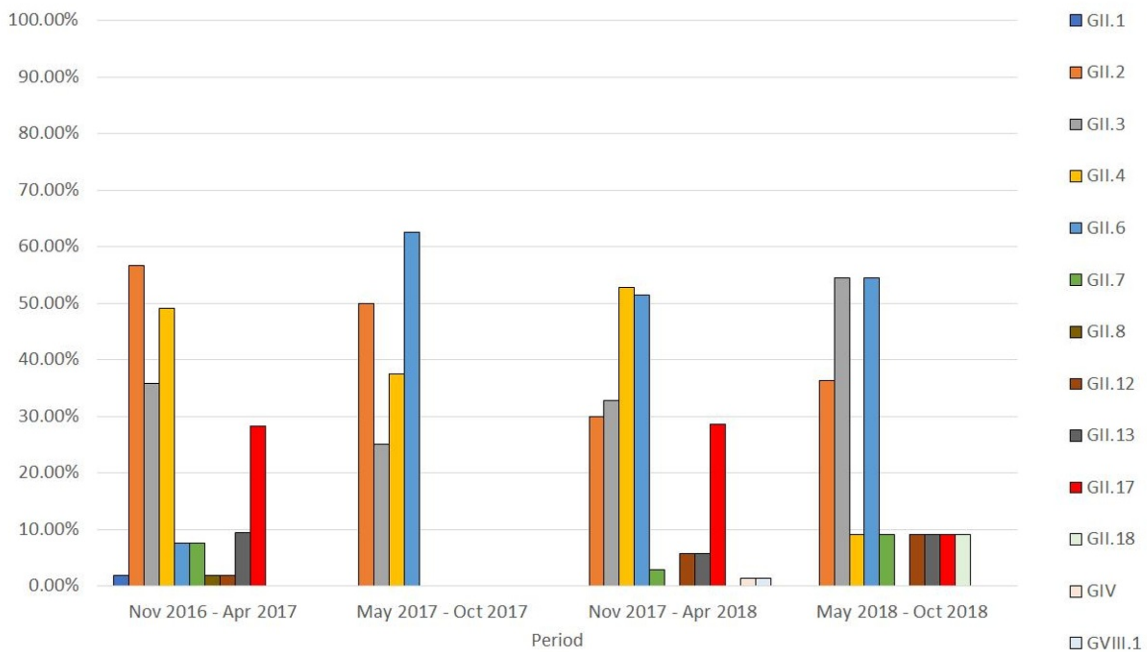



**FIGURE 28: PROPORTION OF SEQUENCED SAMPLES FROM DIFFERENT PERIODS CONTAINING DIFFERENT GII GENOTYPES.**

Only samples providing at least one NoV sequence using amplicons from the GII VP1 assay are considered in calculation of the proportions for each period.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

A number of trends in prevalence of different RdRp types and genotypes can be seen in the data. For GI RdRp, the proportion of samples containing GI.P2 decreased steadily across the lifetime of the BLS from 74.5% in the first period (November 2016 – April 2017) to 41.7% in the final period (May 2018 – October 2018). At the same time GI.P13, which was not present in any sample collected during the first period, increased to a prevalence of 58.3% in the final period. Amongst other GI RdRp types, GI.P4 was common across the BLS with prevalences ranging from 21.6 - 45.7% in different periods.

For GII VP1, GI.4 was present in the majority of samples (51.9 – 73.6%) in all periods of the BLS, however GI.1 became successively more common (5.6% in the first period, 40% in the last) and GI.2 became successively less common (72.2% in the first period, absent from samples taken in the final period).

Amongst GII RdRp types, GII.P4 was common in the first year of the BLS (31.5% and 42.9% in the first two periods) but not found in any samples in the final six-month period, while GII.P12 increased from 11.1% in the first period to 37.5% in the final period. GII.P17, GII.P21 and GII.P31 were present in moderate to high but variable proportions of samples throughout the period (ranging from 14.3 – 38.8%, 38.8 – 53.7% and 40.7 – 50.0% respectively).

Of the GII genotypes detected, GII.2 and GII.3 were always present at moderate to high prevalences (ranging from 30.0 – 56.6% and 25.0 – 54.5% respectively) but with no clear temporal trend. GII.4 was frequently present during the first 3 periods (ranging from 37.5 – 52.9%) but showed signs of a possible decrease and was only present in 9.1% of samples taken during the final period. By contrast, GII.6 showed an opposite trend, being present in only 7.6% of samples taken in the first period, but the majority of samples (51.4 – 62.5%) in the periods thereafter. GII.17 showed an unusual pattern, being detected in a moderate proportion of samples taken during the two "winter" periods (28.3% of samples taken November 2016 - April 2017, 28.6% of samples taken November 2017- April 2018), but being rarely detected in samples taken during the two "summer" periods (not detected in samples taken May – October 2017, detected in 9.1% of samples taken May – October 2018).

### 3.2.3. Diversity of genotypes and P-types detected in BLS samples and human outbreak samples submitted to NoroNet

The diversity of genotypes and P-type detected in BLS was compared to the diversity detected in human samples and submitted to NoroNet. All European norovirus sequences with sample dates between November 2016 to October 2018 were downloaded from Noronet. Although there are more than 40 norovirus genotypes and P-types (Chhabra et al 2019), not all were reported to Noronet during the time period of the baseline study. All Norovirus genotypes and P-types detected in the oysters reported in Noronet are shown in Figure 29 and 30. For these analyses only assigned GI and GII genotypes and P-types were included.

In the 212 BLS samples 7 GI genotypes, 11 GI P-types, 11 GII genotypes, 8 GII P-types were detected, in comparison in Noronet in 1853 submitted outbreak samples 7 GI genotypes, 11 GI P-types, 11 GII genotypes, 12 GII P-types were detected. Norovirus types that were reported in Noronet but not detected in BLS were; GII.14 (10), GII.16 (2), GI.P10 (1), GI.P14 (1), GII.P2 (6), GII.P22 (1), GII.P30 (2) and GII.P33 (5). Norovirus types that were detected in BLS but not reported to Noronet were; GII.18 (1), GII.8 (1) and GI.P9 (1). The fact that more genotypes were reported to Noronet is not surprising as many more samples were analysed and most were only detected in low quantities.

There was also variation in the prevalence of genotypes and P-types between human and BLS samples (Figure 29 and Figure 30). Striking differences were observed for GI.4 and GI.P2 that were present in much higher numbers in the BLS samples compared to Noronet and *vice versa* GII.P7 and GII.P16 that were present in much higher numbers in Noronet compared to the BLS samples. GII.4 was most often detected in human samples, consisting of 63% of all GII genotypes and less often (23%) in BLS samples.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

In human samples GII (88.5%) is more often detected than GI (11.5%), this difference in prevalence is less pronounced for BLS samples with 42% (GI) and 58% (GII).

To compare the Noronet data with the BLS data, the Noronet data was also subdivided into 4 separate 6-month periods, approximately equivalent to two consecutive winters (November 2016 – April 2017 and November 2017 – April 2018) and two consecutive summers (May – October 2017 and May – October 2018) (Figure 29 and Figure 30). Both for the BLS data and the Noronet data GII.P4 became less prevalent in the 2018 summer period, while GI.1 became more prevalent in 2018 compared to 2017. Although the prevalence of individual genotypes varied between these time periods, there was no clear correlation between this variation in BLS and Noronet data.

The correlation between norovirus genotypes detected in oysters and those in the human population have also been described by others. A Japanese study reported a time lag between circulation of norovirus genotypes in the human population and the detection of those in oysters (Pu et al., 2018). Another Japanese study (Imamura et al., 2016) reported a wide variety of norovirus genotypes in Japanese oysters collected in 2013-2014 and analysed using amplicon based NGS. Norovirus GI.4 and GII.4 were detected most often, in 25% of all GI genotypes and 50% of all GII genotypes detected in shellfish. No P-types were reported in this study.

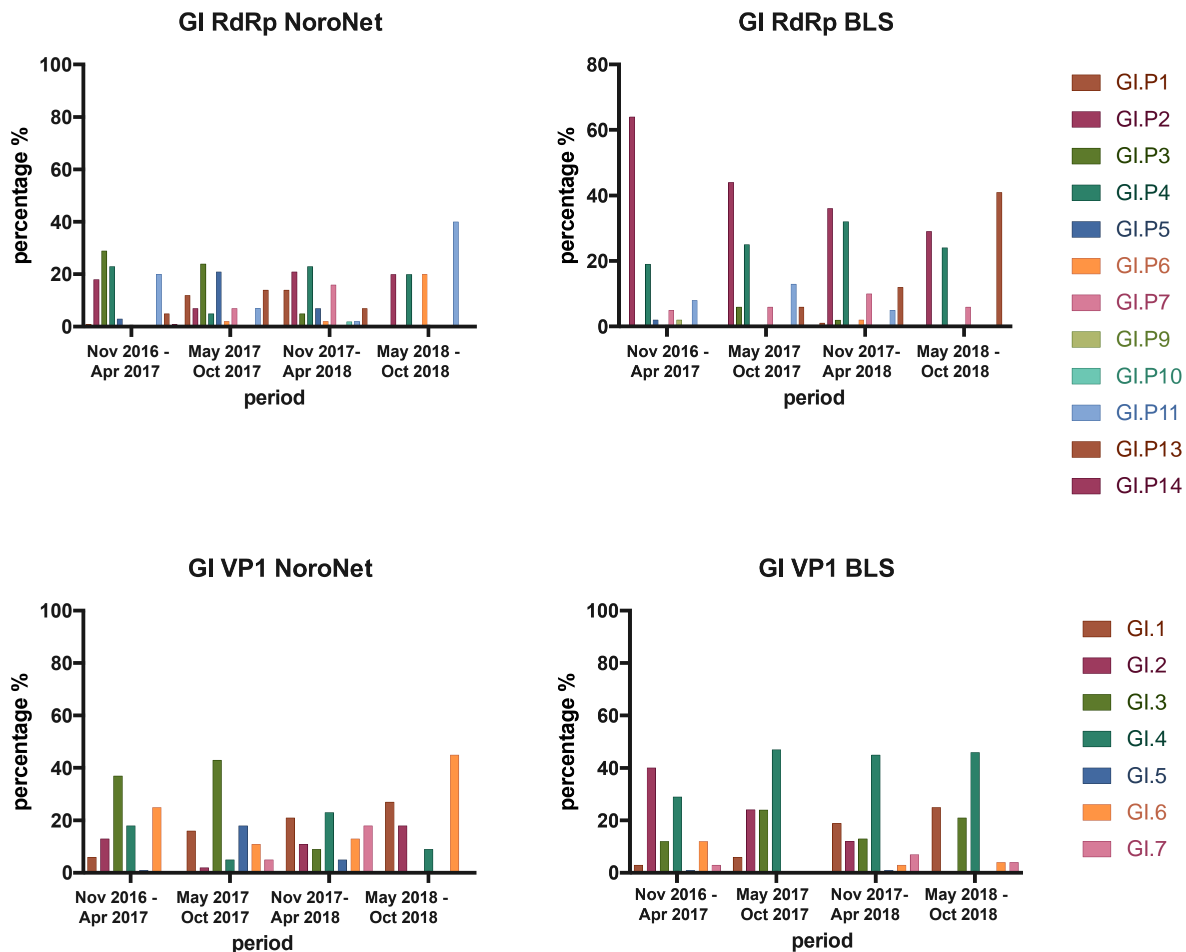



**FIGURE 29: PROPORTION OF GII P-TYPE AND GENOTYPE DIVERSITY IN BLS SAMPLES VS NORONET SURVEILLANCE SAMPLES ISOLATED BETWEEN 2016 AND 2018**



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

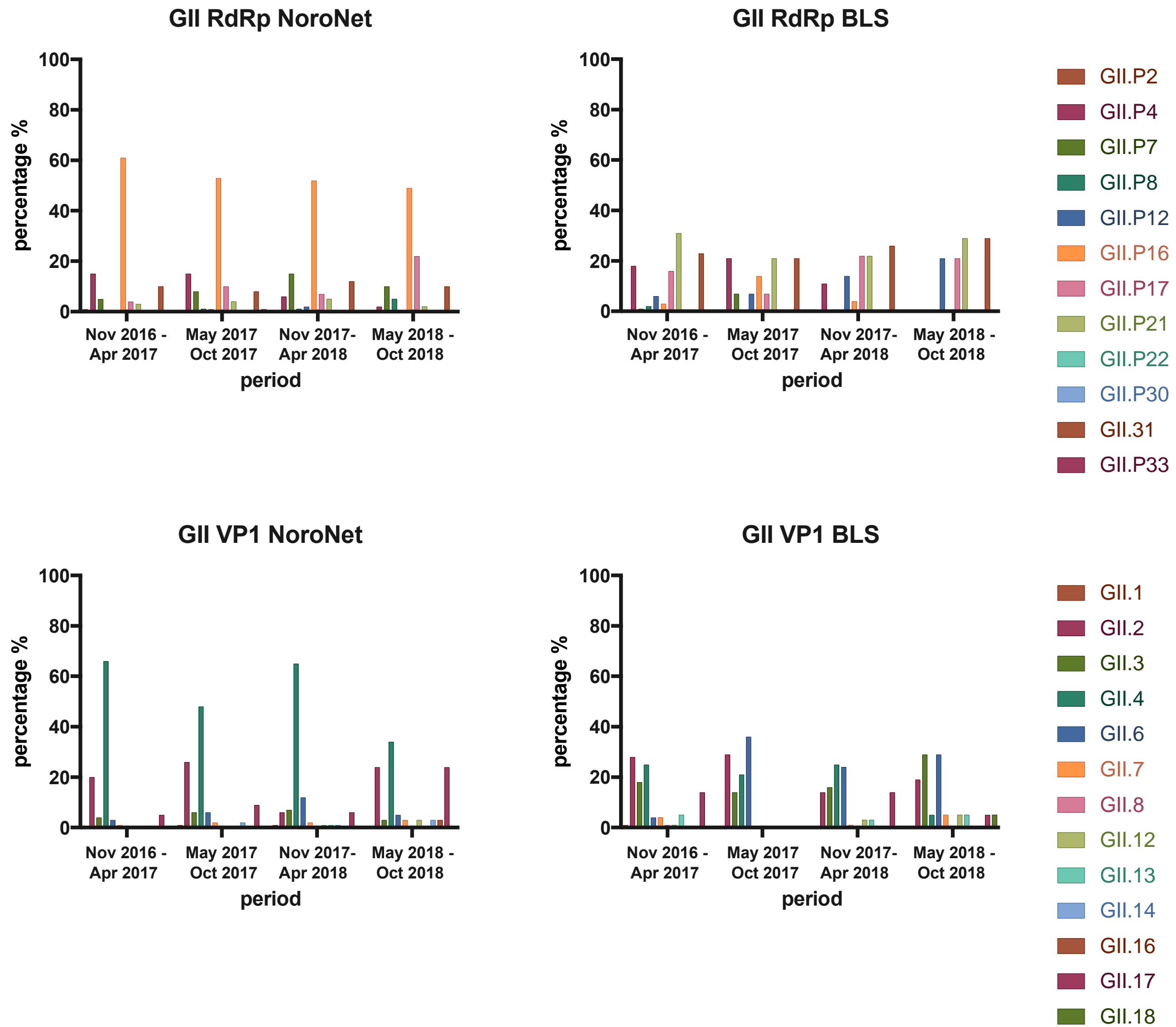



**FIGURE 30: PROPORTION OF GII P-TYPE AND GENOTYPE DIVERSITY IN BLS SAMPLES VS NORONET SURVEILLANCE SAMPLES ISOLATED BETWEEN 2016 AND 2018.**

All BLS samples subjected to metabarcoding analysis at CEFAS and IFREMER were included in the analysis. If a sample contained more clusters belonging to one genotype or P-type it was counted as a single positive. All genotyped sequences submitted to Noronet and with a sample date between November 2016 to October 2018 were downloaded. All norovirus genotypes and P-types that have been reported to Noronet between 2015 and 2019 were included in the legend.

Norovirus sequences are classified into genogroups and genotypes based on sequence diversity. However, there is also genetic diversity with a single genotype or P-type as the genomic nucleotide similarity between viruses from the same genotype can be as low as 69%. Visualisation of this diversity and the genetic relation between strains is needed to investigate transmission chains and linked outbreaks. Phylogenetic trees can be used to visualize this genetic relation, or the number of nucleotide mutations, between sequences. The closer sequences cluster in a phylogenetic tree the more similar they are.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

To investigate if the genetic diversity within genotype in human samples is similar to the genetic diversity observed for the BLS samples, maximum likelihood phylogenetic trees were inferred.

As an example, we selected two genotypes that were present in both human and BLS samples. Sixteen percent of all Noronet samples and 20,4% of all BLS GII samples contained GII.2 while 4.2% of all Noronet samples and 17,3% BLS samples contained GII.3 (Figure 31).

For the GII.2 sequences, there is limited diversity with many identical sequences. The BLS sequences (red) cluster throughout the tree, and there are several Noronet sequences that are identical to the BLS GII.2 VP1 cluster 11 sequence. There are no GII.2 BLS sequences from IFREMER or CEFAS that are identical to each other.

For the GII.3 sequences, there are several clusters and there is more genetic diversity compared to GII.2. Again, the GII.3 BLS sequences (red) cluster throughout the trees showing high diversity within genotype similar to that observed for human samples. There are several GII.3 BLS sequences that are identical to Noronet GII.3 sequences.

This shows that a diversity of GII.2 and GII.3 sequences can be obtained from BLS and that phylogenetic trees can be inferred with these. However, due to the short length of the sequences used for this analysis, the resolution of these trees was too low to provide conclusive evidence on transmission chains.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

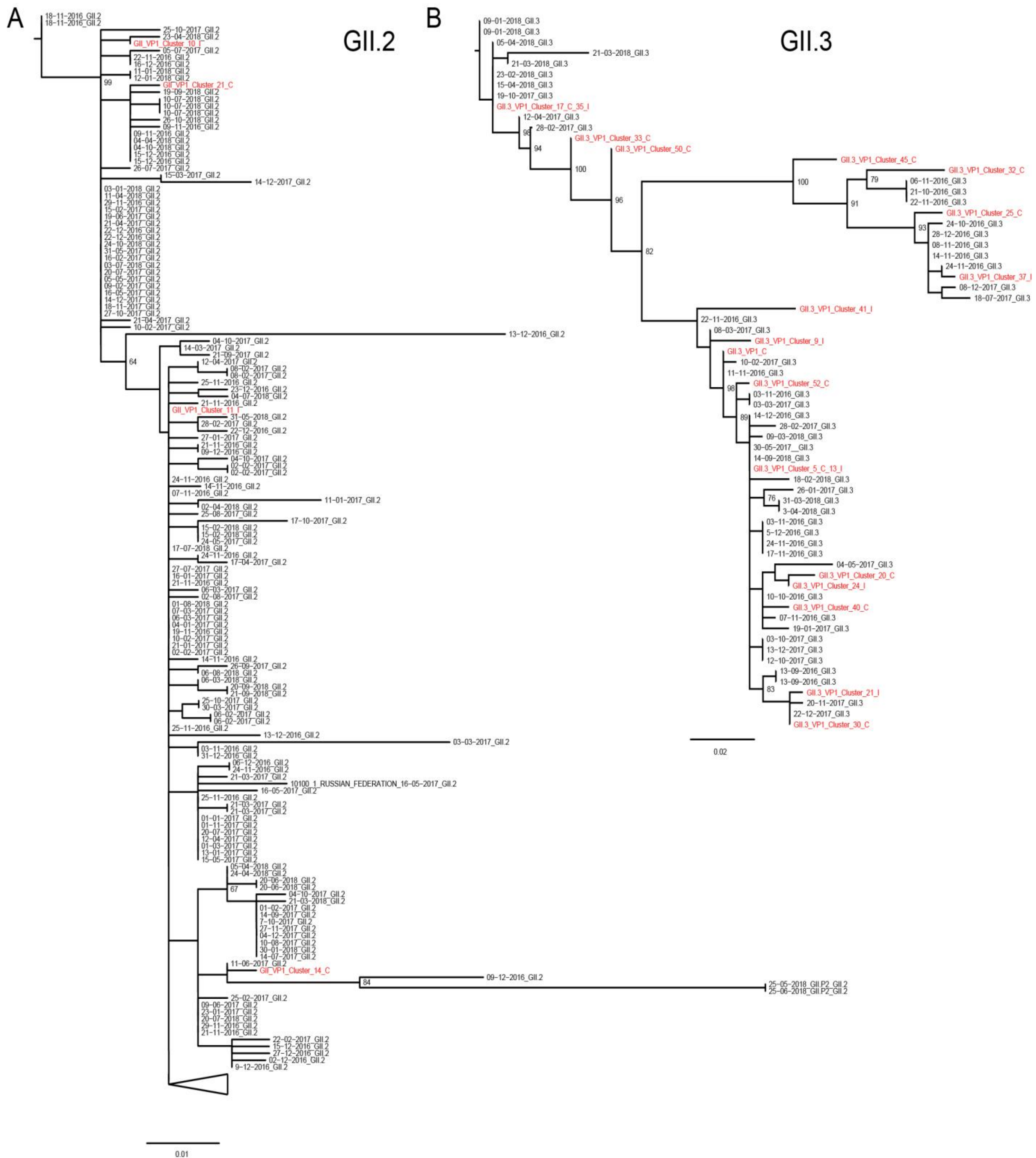



**FIGURE 31: MAXIMUM LIKELIHOOD TREE INFERRED WITH (A) GII.2 (246 NT) (B) GII.3 (307 NT)**

Sequences isolated between 2017 and 2018 from norovirus outbreak samples (Noronet) and sequences detected in BLS (red). Selected bootstrap values above 60 are given.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

### 3.2.4. Concluding remarks for metabarcoding on BLS samples

The metabarcoding approach applied to the BLS samples allowed the identification of norovirus sequences in 177 out of 200 samples (88.5%) from the main list, plus an additional 24 samples from the reserve list. A total of 106 unique RdRp sequences (representing 23 different RdRp types) and 136 unique VP1 sequences (representing 21 different norovirus genotypes) were identified across all samples, demonstrating the efficacy of this approach for describing NoV diversity in naturally contaminated oyster samples. The robustness of the method was further demonstrated by the fact that the majority of the most frequently identified sequence clusters were found independently by both laboratories carrying out the analysis.

The very high diversity of norovirus sequences identified across the BLS samples was notable; previous studies on norovirus diversity in shellfish in various countries and regions of the world have also identified a wide variety of genotypes across different samples (Imamura et al., 2017; Rajko-Nenow et al., 2013; Shin et al., 2013) but the scope and breadth of this study in terms of the number of samples and the number of different countries involved is greater than any previous systematic study and unsurprisingly the total diversity reported here is higher than previously reported (e.g. 21 norovirus genotypes reported here compared with ≤13 in the cited studies). This study is also unusual in discriminating between different sequences within RdRp types and genotypes, revealing the considerable number of similar strains in circulation.

In contrast to the situation with clinical norovirus samples as reported to Noronet, where GII sequences predominate to a significant degree, GI sequences were found almost as frequently as GII sequences in BLS samples, with a similarly high diversity. GI noroviruses are frequently reported in comparable numbers of shellfish samples to GII, including in the European BLS itself (EFSA, 2019). The higher prevalence of GI sequences in shellfish samples implicated in outbreaks was previously reported and linked to the selective transmission of GI NoV strains via oysters through specific binding to carbo-hydrate ligands (Le Guyader et al 2012, Yu et al. 2014). In addition, Noronet data by its nature is heavily biased towards strains that cause symptomatic illness whereas the BLS metabarcoding approach is agnostic in this respect. This factor may also contribute to the high abundance of GI sequences in the BLS dataset. Overall, at the genotype level, diversity in the BLS samples was rather similar to the sequences in the Noronet database, and sequence matches between the two sets were also observed. A small number of genotypes or RdRp types were found in the BLS samples but not in Noronet, and vice versa, however in all cases these types were found in a very small number of samples total. The commoner types in each sample set were always also found in the other set respectively, however there were some types that were relatively more abundant in one set or the other, for example GI.4 and GI.P2 were relatively more abundant in BLS samples, while GI.P7 and particularly GII.P16 were relatively more abundant in Noronet sequences. These differences may be explained by a combination of factors. Methodological biases in the N-PCR assays may artificially inflate or deflate proportions of particular genotypes or RdRp types, for example, analysis of TSS1 (3.1.2.1) revealed that the GII RdRp assay used here underestimates GII.P16, as appeared to be the case when comparing the BLS results with Noronet. On the other hand, the bias towards strains that cause symptomatic illness in the Noronet dataset, and the possible selective binding of certain strains in oyster tissues, described above as possible causes of the high relative abundance of GI sequences in the BLS dataset compared with Noronet, may also have contributed to higher relative abundances of certain sequence types within GI and GII also.

Comparing the temporal trends in abundance of genotypes and RdRp types in the two datasets it is clear that many of the most significant trends noticed were present in both shellfish and clinical samples, for example increases in GI.1 and GII.6 and decreases in GII.P4 and GII.4 over the lifetime of the BLS (November 2016 to October 2018). This is not surprising; the successive emergence and decline of different strains in the human population is well known, and is expected to be reflected in the strain



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

abundance in shellfish (with some delays caused by the transport of viruses through the aquatic environment and the retention of viruses in shellfish tissues), as ultimately the overwhelming majority of GI and GII norovirus strains contaminating shellfish originate from humans. In some cases temporal trends were only apparent in one database. For example, GI.2 become successively less common in BLS samples, but no particular trend with this genotype was noted in Noronet; it is possible that the apparent decrease in this case was an artefact.

More than half of the clusters identified were found in multiple samples, often collected from different countries, and the most commonly found cluster (a GI.4 VP1 sequence) was found in 110 different samples from 7 different countries. However, despite the ubiquity of certain sequences, overall patterns of norovirus diversity varied quite considerably between the different countries providing samples in the BLS. Comparison with strain diversity in clinical samples from the same countries was not possible due to anonymity requirements however the variability noted is consistent with previous studies showing considerably different profiles of norovirus genotypes in different countries within Europe and across the world (Van Beek et al., 2018), and indicates that, despite the global transmission of norovirus, epidemiology of particular strains can vary considerably at a regional and national level.

Diversity within individual samples was frequently high, with the majority of samples containing multiple clusters, genotypes and RdRp types. Within each of the four N-PCR assays diversity within individual samples ranged up to 8 clusters and 5 types. This heterogeneity of norovirus strains within samples has been reported previously, and as a result Sanger sequencing of norovirus from shellfish samples has normally required cloning of amplicons to avoid frequent occurrence of mixed and irresolvable sequences (Verhoef et al, 2010, Rajko-Nenow et al., 2013). The use of methods based on Next Generation sequencing technology, such as the metabarcoding described here, therefore offers the additional benefit over Sanger sequencing approaches of not requiring additional, often complex, cloning steps.

Interestingly, despite the primers targeting GI and GII noroviruses, a few sequences assigned as GIII, GIV, GV, GVIII and GNA1 were identified, hinting at the high diversity of both human and non-human norovirus types circulating in the marine environment. Shellfish, being exposed both to human and animal contaminations, have been suspected as a possible reservoir that may favor zoonotic events or emergence of recombinant viruses (Zakhour et al 2010, Rincé et al 2018).

## 3.3. VirCapSeq Metagenomics on BLS samples

According to results obtained on TSS samples, where long amplicon sequencing by ONT proved less efficient than the VirCapSeq-metagenomics to identify NoV strains in samples with low contamination levels, the latter method was chosen for the analysis of a subsection of 20 BLS samples.

The BLS samples for metagenomics were selected according to criteria described in chapter 2.1.3.5. The samples were tested in duplicates, starting at extraction step. In total, twenty-three samples were submitted to the VirCapSeq-optimized extraction procedure. Three samples gave negative or very weak positive signals (Ct>40) by qRT-PCR targeting NoV GI and GII. Thus, nucleic acids from 20 samples were used for library preparation (see the results in Appendix F) and Illumina sequencing.

### 3.3.1. Sequencing quality

Between 4.9M and 14.7M raw reads were obtained for the BLS samples by merging data from the three replicates. Fewer reads were obtained for the control DT (3.4M) and the negative water control (0.2M) as expected (Figure 32). A small fraction of reads was removed by the quality trimming step of Ifremer's pipeline (Figure 32, red portion of bars) and most reads were kept for subsequent analysis (Figure 32, blue portion of bars).



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

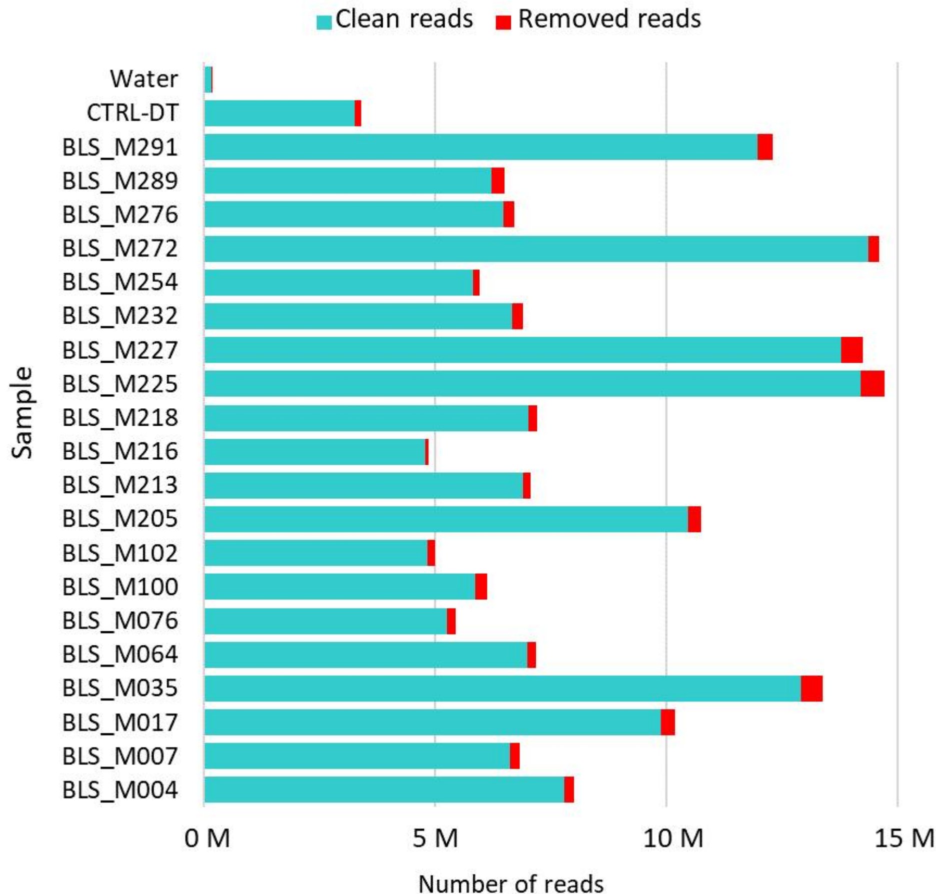



**FIGURE 32: NUMBER OF READS REMOVED FROM RAW READS AND CLEAN READS KEPT FOR ANALYSIS FOR BLS BY VIRCAPSEQ METAGENOMICS.**

CTRL-DT: control DT without NoV contamination; M: Millions of reads.

### 3.3.2. Norovirus detection and diversity by VirCapSeq metagenomics

Among the 20 sequenced BLS samples and their three replicates, none yielded NoV contigs longer than 500 bp, the initial size filter that was used when analysing TSS (see 3.1.3). To increase the chances to assemble NoV genomes from the data, raw reads from the three replicates were merged. The size threshold was also lowered to 150 bp.

In the control water sample, a NoV contig of 266 bp was found, with an abundancy of 10 reads. This 10 reads abundancy was thus set as an abundancy threshold under which results were considered unreliable. Eventually, 5 samples yielded from one to six NoV contigs mapped by more than 12 reads, with sizes ranging from 214 bp to 374 bp (**Error! Reference source not found.**).

Using the NoV Typing Tool 2.0, these contigs were assigned to GI, GII or GIII genogroups, and some could also be genotyped or P-typed when falling into the adequate regions of the genome (



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 24). Three BLS samples, M225, M232 and M272, yielded a unique contig identified as GII, GIII or GI.P2, respectively. One BLS sample, M216, harboured three different NoV sequences, one GI.P4, and two GI without genotype/P-type assignation. Finally, the BLS sample M213 yielded six different NoV contigs, two GI not genotyped, one GI.7[P7], one GII not genotyped, one GII.17 and one GII.6., suggesting contamination with raw sewage water or repeated contamination event.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**Table 24: Characteristics and identification of NoV contigs obtained after VirCapSeq metagenomics of BLS samples.**

| Sample name | Contig length | Reads | Position* | NoV Typing tool¤ | BLASTn# |
|---|---|---|---|---|---|
| **M213** | 249 | 36 | 2913-3162 | GII | *GII.P7* |
| | 265 | 18 | 5214-5479 | GI.7[P7] | *GI.7[P7]* |
| | 226 | 18 | 3975-4201 | GI | *GI.P7* |
| | 304 | 16 | 5712-6016 | GII.6 | *GII.6* |
| | 221 | 14 | 5358-5579 | GII.17 | *GII.17* |
| | 248 | 12 | 2002-2250 | GI | *GI.1* |
| **M216** | 279 | 54 | 1832-2111 | GI | *GI.P4*[a] |
| | 214 | 28 | 4996-5210 | GI.P4 | *GI.P4*[a] |
| | 374 | 20 | 3983-4357 | GI | *GI.P4*[a] |
| **M225** | 269 | 42 | 4408-4677 | GI.P2 | *GI.P2* |
| **M232** | 231 | 16 | 1489-1720 | GIII | *GIII* |
| **M272** | 241 | 14 | 3618-3859 | GII | *GII.P21* |

* contig position on the corresponding NoV reference genome as determined by the NoV Typing tool 2.0; ¤ assignation to a NoV genogroup, genotype or P-type by the NoV Typing tool 2.0; # NoV genotype of the best BLASTn hit as assigned by the NoV Typing tool ; [a] the same best hit was found by BLASTn for the three contigs of sample M216

To classify further the sequences falling outside the typing regions used by the Typing tool 2.0, we used BLASTn (https://blast.ncbi.nlm.nih.gov) to identify a best hit among the nucleotide collection database. The genotype or P-type of this best hit is indicated in Table 21. Percent identities between the contigs and their BLASTn best hit were higher than 96%, except for the contig obtained from sample M232 and assigned as GIII, for which it was only 84%. BLASTn best hit genotype / P-type assignation was identical to that of the contig when available. Interestingly, in sample M216, the three contigs shared the same best hit, a GI.P4 strain, suggesting that they might be different parts of the genome of the same strain, other parts missing for a complete assembly. However, this approach is only a proxy for genotype / P-type assignation and its result should be considered cautiously.

In summary, the results obtained by VirCapSeq metagenomics on BLS allowed to identify at least one NoV strain at the genogroup level in 5 of the 20 samples, and at the genotype / P-type level in 3 of them.

### 3.3.3. Comparison of metabarcoding and VirCapSeq metagenomics on BLS samples

The BLS samples submitted to VirCapSeq metagenomics were also analysed by metabarcoding (see section 3.2), which allows to compare the results of both methods on actual environmental samples (**Error! Reference source not found.**).



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**TABLE 25: COMPARISON OF GENOTYPES / P-TYPES IDENTIFIED BY METABARCODING AND VIRCAPSEQ METAGENOMICS IN BLS SAMPLES**

| Sample | Metabarcoding | Metagenomics* | Match ? |
|---|---|---|---|
| **M213** | GI.P2 (2) ; GI.1 (1) ; GII.P17 (1) ; GII.P31 (1) ; GII.3 (1) ; GII.6 (1) ; GII.17 (1) | GI.7[P7] ; GI.7 ; GII.6 ; GII.17 ; *GII.P7* ; GI.*1* | Partial |
| **M216** | GI.P2 (1) ; GI.P4 (1) ; GI.P7 (1); GI.4 (4) ; GI.6 (1) ; GII.P17 (2) ; GII.P21 (1) ; GII.P31 (1) ; GII.2 (1) ; GII.7 (1) | GI.P4 | Partial |
| **M225** | GI.P2 (1) ; GI.2 (2) ; GI.4 (1) ; GII.P4 (1) ; GII.P21 (3) ; GII.2 (1) ; GII.3 (2) ; GII.4 (1) | GI.P2 | Partial |
| **M232** | GI.P2 (1) ; GI.2 (1) ; GI.4 (1) ; GI.6 (1) ; GII.P4 (1) ; GII.2 (1) ; GII.3 (1) ; GII.7 (1) | GIII | No |
| **M272** | GII.P21 (1) ; GII.P31 (1) ; GII.3 (1) | GII.*P21* | Partial |

* genotype / P-type identified using the NoV Typing tool 2.0 when possible, or an estimation using BLASTn (in italic).

In three samples (M216, M232 and M272), one contig obtained by metagenomics had a P-type matching that of cluster(s) retrieved by metabarcoding in the same sample, GI.P4, GI.P2 and possibly GII.P21, respectively. In one sample, M213, the assignation of 4 contigs were compatible with the results from metabarcoding, while two contigs assigned to GI.7 and GI.7[P7] did not have a counterpart GI.7 or GI.P7 cluster among the metabarcoding results. In the last sample, M232, none of the strains identified by metabarcoding were observed by metagenomics, and the GIII strains sequenced by metagenomics does not appear in the metabarcoding results.

Interestingly, one contig from sample M225, assigned as GI.P2, overlaps with the position of the GI RdRp metabarcoding amplicon. Aligning the sequence of this contig (M225_NODE_6640) with that of the corresponding cluster (GI_RdRp_Cluster_1_C_1_I), we verified that both sequences were identical on the 93 shared nucleotide positions.

In conclusion, out of these 5 samples, results were discordant for one sample while 4 samples had partially matching results between the two approaches. Importantly, the diversity observed by VirCapSeq metagenomics in the BLS samples was a subset of the one observed by metabarcoding, suggesting that the sensitivity of the VirCapSeq metagenomics is lower than that of the metabarcoding in such samples.

### 3.3.4. Concluding remarks for VirCapSeq metagenomics on BLS samples

Together, these results show how the VirCapSeq metagenomics method can perform on actual environmental samples. Out of the 20 samples that were sequenced, only 5 yielded a NoV contig. Besides, these contigs were very short and often fell outside the typing regions, impairing confident assignation to a NoV genotype or P-type. For most samples, the contigs represented a subset of the NoV diversity observed using the metabarcoding approach. In one instance though, a NoV GIII contig was identify by this technique and not by metabarcoding, which could due to the use of primers targeting only GI and GII for this method.

The interest of the VirCapSeq metagenomics lies in its potential to yield long sequences allowing to identify strains in addition to genotypes or P-types. However, here the sequences were too short for such an application, on the same range than the metabarcoding amplicons (200-300nt), and much shorter than those obtained on the TSS1 and TSS2. This could be due to the longer storage of the BLS samples between their collection (from November 2016 to October 2018) and their processing for this study (2021) which may have altered the viral RNA.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

In conclusion, although promising, this method still requires optimisation and should be used in priority on samples with high levels of contamination or collected recently, when the identification of a strain is required and information on genotype / P-type not sufficient.

## 3.4. VirCapSeq Metagenomics on OB samples

### 3.4.1. VirCapSeq on OB BMS samples

Thirty bivalve molluscan shellfish (BMS) samples related to 16 outbreaks (OB) that occurred in France or Denmark were available for the project, including 29 *C. gigas* and one *M. edulis* sample. The digestive tissues were analysed in triplicates, starting at extraction step using the VirCapSeq-optimized protocol. For nine OB-BMS samples, no positive signal were obtained by qRT-PCR after nucleic acid extraction for any of replicates and these samples were excluded from the study. Twenty-one remaining samples were submitted to reverse transcription. A positive signal for NoV qPCR was obtained for 20 of the OB-BMS samples. The Appendix F present the results of library preparation.

#### *Sequencing quality*

Between 1.6M and 16.2M raw reads were obtained per OB-BMS sample when merging data from the two replicates. Fewer reads were obtained for the control DT (3.4M) and the negative water control (0.2M) (Figure 33). A small fraction of reads was removed by the quality trimming step of Ifremer's pipeline (Figure 33, red portion of bars) and most reads were kept for subsequent analysis (Figure 33, blue portion of bars).



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

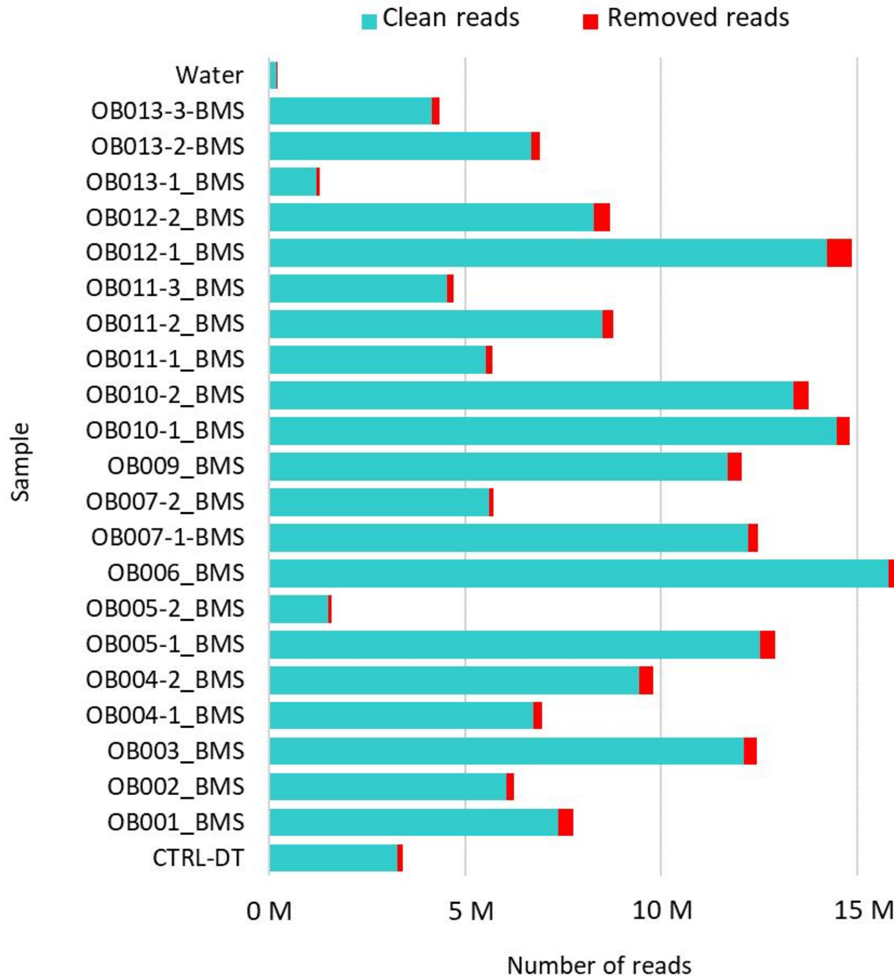



**FIGURE 33: NUMBER OF READS REMOVED FROM RAW READS (RED) AND CLEAN READS KEPT FOR ANALYSIS (BLUE) FOR BLS BY VIRCAPSEQ METAGENOMICS.**

CTRL-DT: control DT without NoV contamination; M: Millions of reads.

### *Identification and diversity of NoV sequences*

As for the analysis of BLS samples, that were sequenced in the same run, the raw reads from the two replicates were merged, the size threshold on contigs was lowered to 150 bp, and an abundancy threshold was set at >10 reads. Eventually, 3 samples yielded a NoV contig mapped by more than 12 reads, with sizes ranging from 266 bp to 275 bp (**Error! Reference source not found.**). Using the NoV Typing Tool 2.0, these contigs were assigned to GI or GII genogroups, but they could not be genotyped or P-typed since they fell outside the typing regions (**Error! Reference source not found.**). A tentative identification was done using that of the best BLASTn hit, but this should be considered with caution considering that the genomic regions that were sequenced could be more divergent or more conserved than the usual typing regions.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**TABLE 26: CHARACTERISTICS AND IDENTIFICATION OF NOV CONTIGS OBTAINED THROUGH VIRCAPSEQ METAGENOMICS OF OB-BMS SAMPLES.**

| Sample name | Contig length | Reads | Position* | NoV Typing tool¤ | BLASTn# |
|---|---|---|---|---|---|
| **OB011-2-BMS** | 275 | 22 | 1475-1750 | GI | *GI.P4* |
| **OB002-BMS** | 273 | 22 | 1400-1673 | GI | *GI.P3* |
| **OB004-1-BMS** | 266 | 12 | 3684-3950 | GII | *GII.P7* |

* contig position on the corresponding NoV reference genome as determined by the NoV Typing tool 2.0; ¤ assignation to a NoV genogroup, genotype or P-type by the NoV Typing tool 2.0; # NoV genotype of the best BLASTn hit as assigned by the NoV Typing tool

### 3.4.2. VirCapSeq metagenomics on human stool samples

Ten stool samples linked to the outbreaks that occurred in France were finally available for this study and sequenced at EMC as described in 2.2.

#### *Sequencing quality*

Between 59,812 and 3,713,038 raw reads were obtained per OB-stool sample. A fraction of reads was removed by the quality trimming step of EMCs pipeline (Figure 34, red portion of bars) and most reads were kept for subsequent analysis (Figure 34, blue portion of bars).

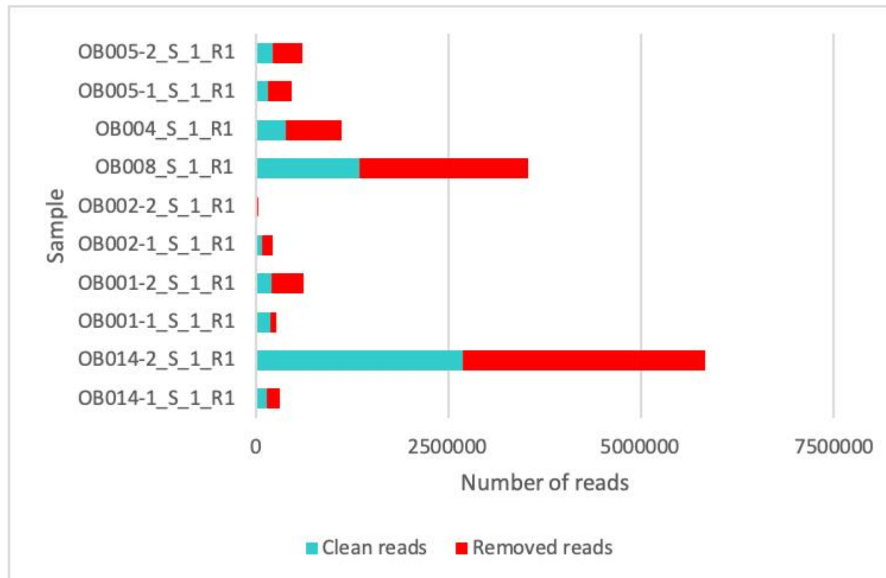



**FIGURE 34: NUMBER OF READS REMOVED FROM RAW READS (RED) AND CLEAN READS KEPT FOR ANALYSIS (BLUE) FOR OB-STOOL BY VIRCAPSEQ METAGENOMICS.**

#### *Identification and diversity of NoV sequences*

In the ten stool samples, two different genogroups were detected (GI and GII), 6 genotypes and 7 P-types (Figure 35 and **Error! Reference source not found.**). Detected genotype / P-type combinations were GI.5[P5], GI.1[P1], GII.17[P17], GII.1[P33] and GII.17[P25]. Interestingly, the GII.17[P25] recombinant has not been reported previously (Tohma et al., 2021), only two GII.P25 sequences have been reported to Noronet and none during 2016, 2017 and 2018. Importantly, five out of ten samples contained multiple NoV genotypes or P-types.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

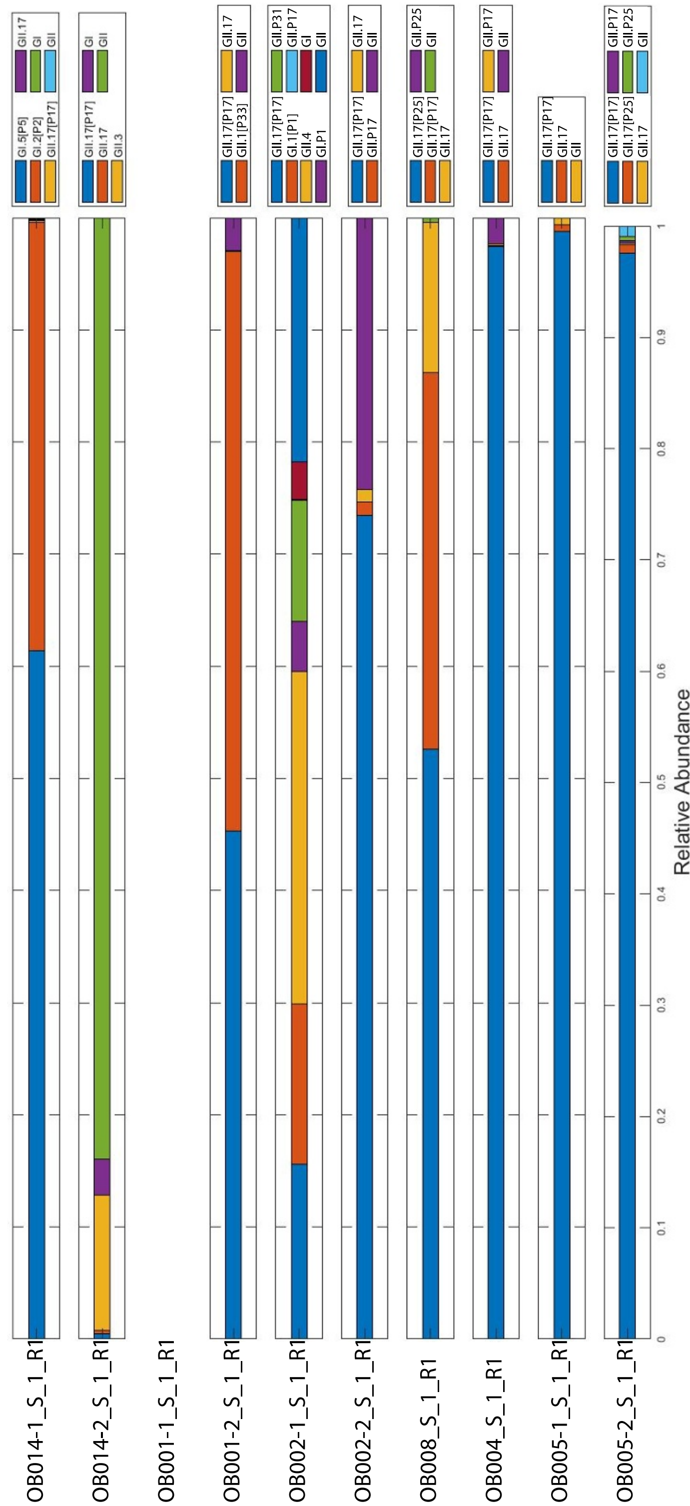



**Figure 35: Proportion of reads assigned to identified Genotypes or P-types using EMC pipeline in Ifremer (left) for OB-stool**



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**TABLE 27: CHARACTERISTICS AND IDENTIFICATION OF NOV CONTIGS OBTAINED THROUGH VIRCAPSEQ METAGENOMICS OF OB-STOOL SAMPLES.**

| Outbreak code | Sample code | Genogroups | Genotype | P-type |
|---|---|---|---|---|
| **OB-14** | OB014-1_S_1_R1 | GI, GII | GI.2, GI.5, GII.17 | GI.P2, GI.P5, GII.P17 |
| | OB014-2_S_1_R1 | GI, GII | GII.3, GII.17 | GII.P17 |
| **OB-001** | OB001-1_S_1_R1 | | | |
| | OB001-2_S_1_R1 | GII | GII.1, GII.17 | GII.P33, GII.P17 |
| **OB-002** | OB002-1_S_1_R1 | GI, GII | GI.1, GII.4, GII.17 | GI.P1, GII.P17, GII.P31 |
| | OB002-2_S_1_R1 | GII | GII.17 | GII.P17 |
| **OB-008** | OB008_S_1_R1 | GII | GII.17 | GII.P17, GII.P25 |
| **OB-004** | OB004_S_1_R1 | GII | GII.17 | GII.P17 |
| **OB-005** | OB005-1_S_1_R1 | GII | GII.17 | GII.P17 |
| | OB005-2_S_1_R1 | GII | GII.17 | GII.P17, GII.P25 |

### 3.4.3. Matching results between BMS and stool samples from the OB sample set

Metagenomic sequencing of matched OB BMS and stool samples was performed to investigate if the VirCapSeq metagenomic sequencing can be used to link NoV outbreaks in humans with NoV contaminated BMS. Unfortunately, not all human stool samples were available for sequencing and not all BMS samples resulted in NoV sequences. As a result, there is only one outbreak (OB-002) where there are sequences from two stool samples (OB002-1 and OB002-2) and one the BMS sample (OB002-BMS).

Analysis of these two stool samples resulted in contigs that could be typed as GI, GII, GII.17[P17], GII.P31, GII.4, GI.1[P1] and GII.P25. The BMS sample contained contigs that could be typed as GI, GII, GII.1 and GII.17, including NoV contig mapped by less than 12 reads (



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 28).

All GI and GI.1[P1] contigs were aligned, but the BMS GI contig (273 nt) differed with 87 of 273 nucleotides from the stool GI.1[P1] contig. BLASTn analyses showed that the BMS GI contig had 100% identity with a GI.3[P3] sequence (accession number MH218726.1).

All GII, GII.17[P17], GII.P31, GII.17 and GII.4 contigs were aligned. The BMS GII.1 sequence did not align with any of the GII contigs obtained from stool. The three other BMS GII and GII.17 contigs aligned with GII.17[P17] sequences from OB002-1 and OB002-2. The first contig (265 nt) was identical to OB002-1 and OB002-2, the second contig (208 nt) differed with 1 nucleotide from OB002-1 and OB002-2, and the last contig (239 nt) differed with 2 nucleotides from OB002-1 and OB002-2.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**TABLE 28: COMPARISON OF NOV CONTIGS OBTAINED THROUGH VIRCAPSEQ METAGENOMICS OF OB-STOOL AND OB-BMS SAMPLES.**

| Outbreak code | Sample code | Genogroups | Genotype | P-type |
|---|---|---|---|---|
| **OB-002** | OB002-1 | GI, GII | GI.1, GII.4, GII.17 | GI.P1, GII.P17, GII.P31 |
| | OB002-2 | GII | GII.17 | GII.P17 |
| **OB-002 BMS** | OB002_BMS | GI, GII | GII.1, GII.17 | GI.P3#, GII.P17# |

#NoV genotype of the best BLASTn hit as assigned by the NoV Typing tool

In conclusion, metagenomics can be used to match sequences from BMS samples with those from stool samples. However, since GII.17[P17] is highly prevalent and the sequenced regions of the BMS samples are relatively short it was for this case not possible to exclude another source for the NoV outbreak.

# 4. Discussion and conclusions

The different objectives of this project were designed to answer the EFSA specific objectives (ESO) proposed in the call. Rather than viral isolates, the selected samples consisted in nucleic acid extracted from BMS tissues or human stools, containing a mixture of host, bacterial and viral sequences. In this report, the terms of reference of the call were adapted and the consortium decided to focus on NoV contaminated oyster samples, as a proof of concept that may be extended to other food and shared with the different European Reference laboratories in charge of viral detection in food. BMS are known to be sensitive to viral contamination, especially by NoVs. As a consequence, BMS, and especially oysters, are frequently implicated in outbreaks (EFSA, 2012). Up until now, the determination of the genotype(s) of NoV implicated in human outbreaks or detected in contaminated foods relied on the amplification, cloning and classical (dideoxy) consensus sequencing of selected segments of their genome. This approach, recommended to describe the diversity of NoV strains in shellfish samples, is specific but labor- and time-consuming and may miss the low abundance strains, or strains for which the primers are mismatched (Symes et al., 2007, Le Guyader et al., 2008).

NGS methods have been developed to better characterize viral diversity in different samples including sewage or food (Nieuwenhuijse et al., 2020, Desdouits et al. 2020b). These methods are promising, but are still not benchmarked for routine application on food analysis (Desdouits et al. 2020b). Indeed, up to now, no real comparison has been carried out on approaches such as metabarcoding or metagenomic for virus characterization in naturally contaminated food. Methods, including bio-informatic analysis, for different important human pathogens such as bacteria, parasites or viruses have been developed through research projects such as H2020 COMPARE project. For bacterial or eukaryotic pathogens, conserved regions of the genome such as 16S or 18S genes or the capacity to grow the selected strains offer opportunities to increase the sensitivity of sequencing methods. This is not possible for human enteric viruses that present a large genomic diversity and most of them cannot be amplified in cell culture, as exemplified for NoVs.

Therefore, the key challenge for this project was to explore different NGS approaches that offer a sensitivity 'compatible' with NoV analysis in food. Indeed, contaminated foods such as oysters may harbor very low concentrations of viral genomes (close to the limit of detection of molecular techniques) and a large diversity of sequences when contaminated by human sewage.

We will first discuss the 2nd consortium objective, CO2, that pertains to methods comparison and benchmarking. Then, we will discuss the other COs and the corresponding results.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

***Consortium Objective 2 (CO2): To assess and compare the ability of the three NGS methods to characterize genetic diversity of NoVs.***

For this project, the consortium compared three different methods for NoV sequencing, ie metabarcoding based on short RdRp and VP1 sequences, metagenomics with enrichment of vertebrate viral sequences using the VirCapSeq-VERT probes, and a long amplicon sequencing using Oxford Nanopore technology. These methods were compared and assessed on artificial samples (TSS1 & TSS2) to evaluate their potential, limitations and bias. Thus, most conclusions on methods comparison arise from results obtained on the TSS. In addition, 20 BLS samples were also submitted to both metabarcoding and VirCapSeq metagenomics, which also allows for method comparisons on natural samples. To characterize a NoV strain, due to frequent genomics recombination events at the junction between the RdRp and VP1 genes, sequences from both genes need to be characterized (Chhabra et al., 2019). For the metabarcoding method, we selected several previously published primer sets and applied them on different sample sets. This method for typing NoV in oysters was applied previously for NoV GI and/or GII in oysters in Japan (Pu et al 2018) and in France (Desdouits et al, 2020), but was never systematically evaluated on both GI and GII with RdRp and VP1 genes.

All primer sets tested here presented selection biases when applied on a mix of NoV strains (TSS1). They yielded different relative abundancies for the different strains of each genogroup, which shows that relative abundancies do not reflect the actual proportion of each strain in the mix, but likely result from matrix-primers selective interactions. Regarding replicability, the method was robust, as exemplified by the high similarity between replicates for each sample. In the samples with lower NoV concentration the number of identified strains were sometimes lower, but the method still yielded NoV sequences. This suggests that the method may lack replicability on samples with a low NoV concentration, but is still very specific for NoV. As the RdRp-VP1 junction is important to identify recombinant NoV, amplification of this region was assayed. However, although promising, this approach exhibited several important limitations as for both GI and GII fewer genotypes and P-types were identified compared to the other primer sets. In addition, some partial sequences or chimera were detected, impairing the identification of recombinants in natural samples when the composition is unknown.

For GI and GII metabarcoding, when applied on laboratory contaminated oyster samples (TSS2), the same observation regarding primer selection biases held true. This important result emphasizes that relative abundancies are only indicative and cannot be used to draw quantitative conclusions. For both genogroups, the VP1-primer set allowed detection of all expected strains, the RdRp-primer sets also detected most of the strains, while primer sets targeting the RdRp-VP1 junction allowed the identification of most strains, but also resulted in the detection of GI and GII chimera. Importantly, these chimeras represented a higher proportion of reads than the parental strains. Since the recombination point was located at the junction between RdRp and VP1 genes, a natural recombination hotspot for NoV, they can only be distinguished here because the actual diversity of strains was known. In natural samples, these clusters would appear as new recombinants, a false positive result representing an important limitation for the use of this primer set.

Regarding technical replicability, the metabarcoding method is robust for samples up to concentrations close to the limit of quantification (LOQ) of the methods applied for shellfish analysis (ISO 15216). When NoV genome concentrations were below this LOQ, replicates were more variable for most primer sets and laboratories. This is expected, as very low concentrations of NoV genomes result in sampling bias when taking the few microliters required for the assay. Interestingly, here, combining results from the three replicates often allows to obtain more diversity. Thus, performing replicates can inform on the robustness of the results and, at the time, lower the sampling bias when working on samples with very low concentrations. Regarding repeatability between the different laboratories, the metabarcoding



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

method showed identical results in terms of strain identification between the three laboratories for VP1 and RdRp primer sets for GI and GII.

Due to the high genetic diversity of NoV, primer design needed for the metabarcoding approach is challenging and some strains may not be well recognized, as suggested with our metabarcoding results on TSS1. One advantage of the metagenomic approach is that this amplification step with specific primers is not needed, allowing the detection of new sequences somewhat divergent and not being amplified by commonly used primer sets. The drawback of this lack of specific amplification is the sensitivity threshold, as it may be complicated to detect sequences present at very low concentrations. This may be complemented by purification steps during the sample preparation (as for example elimination of oyster tissue, bacteria removal by filtration), DNA removal, or during the bio-informatic analysis (negative filtering of reads on oyster genomes). During the COMPARE project, different comparisons were performed to improve the sensitivity for the detection of short RNA viral genomes such as the one of norovirus: poly(A)-capture, non-ribosomal random primers, random primers that do not match the genome matrix (as oyster genome). None of these approaches were found fully efficient for oyster samples (Strubbia et al 2019a). Another strategy to enhance the recovery of human viruses in complex matrices is capture-based metagenomics, that includes an enrichment step in viral sequences using probed targeting all known vertebrate viruses during the library preparation (Briese et al 2015). This technique was found to efficiently obtain complete genomes of different human enteric viruses in sewage samples and related bioaccumulated oysters (Strubbia 2019). Such an approach was also applied to naturally contaminated clams collected in the Sanaga river (Cameroon) and allowed the detection of a large diversity of RNA viruses including human or animal norovirus sequences (Bonny et al 2021). Here, the method was promising on TSS samples but performed less well on naturally contaminated samples from the BLS and OB sets, presumably due to longer storage and loss of NoV genome integrity. Thus, further optimization is needed to increase the length of NoV sequences retrieved from such samples. In addition, due to the discontinuation of the VirCapSeq-VERT probes synthesis technology by the manufacturer (Roche), the design of other probe sets is needed. This could be the basis of a new version of the method, possibly with higher sensitivity by focusing on the viruses of interest for food investigation.

For microbiologists, analysis of the millions of reads obtained after NGS constitutes a new area of expertise. Some user-friendly tools were shared rapidly, such as the FROGS pipeline (Escudié et al. 2018) that was used for the metabarcoding method. However, even with this tool, the pre-process parameters need to be adapted and the automatic analysis does not exclude a careful manual curation to exclude chimera sequences. Negative controls are required together with quality rules such as the number of reads needed to be considered as a detected sequence to ensure reliability of the results. Hopefully for NoV, the Noronet database exists and constitutes an excellent support for NoV typing (Kroneman et al 2011).

The bioinformatic analysis of reads obtained using the metagenomic method was more challenging. In this project several pipelines, sharing similarities but not identical, were used and provided comparable results. However, many other parameters interfered such as the quality and number of reads, the diversity of sequence that prevent a clear conclusion. As for the metabarcoding method, the use of sequence databases such as Genbank or Noronet and the Noronet typing tool enabled the identification and typing of NoVs sequences detected in samples. The development of an automated online system for virus identification in NGS data such as Genome Detective is a promising tool that will allow providing laboratories with an easy to use website to identify the diversity of viruses present in the sample, in a relatively short time (Vilsker et al. 2019). As for in-house pipelines, quality controls and a critical reading of the results to detect possible chimeric sequences will be important to consider.

Thus, our final conclusions regarding method comparisons are the following:

- The three methods allowed to sequence NoV in most tested samples, however their performance in retrieving the expected NoV diversity differed. The most sensitive method was



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

the short metabarcoding. VirCapSeq metagenomics presented a lower sensitivity but this may be corrected partly by the sequencing depth and the bioinformatics pipeline. The long amplicon sequencing by ONT was the least sensitive, presumably due to amplification of a long amplicon, and was judged not compatible yet with BMS analysis.

- Metabarcoding was reproducible between replicate sample analysis, and repeatable between laboratories, but the results should be interpreted as qualitative and not quantitative due to PCR bias and differential amplification of strains. VirCapSeq metagenomics was also reproducible between sample replicates, and yielded similar results in different laboratories but with a high impact of the sequencing depth, and to a lesser extent, of the bioinformatics pipeline.

- Due to the amplification steps of metabarcoding, and the high sensitivity of NGS, the methods are prone to contamination and false positive results. The inclusion of negative controls and quality criteria are necessary to set thresholds to confidently interpret the results.

- Library preparation for metabarcoding and long amplicon sequencing by ONT are compatible with the ISO15216 extraction. They rely on PCR amplification, which may be moderately time consuming but easy to perform, similar to the current gold standard for NoV genotyping without the tedious cloning step. The VirCapSeq metagenomics requires a dedicated extraction protocol, more sophisticated and time-consuming.

- Metabarcoding results can be analyzed with user-friendly, mostly automatized pipelines, except for one step (NoV chimera identification) that still required careful manual curation. Metagenomics pipelines require bioinformatics know-how and the adaptation of parameters to the dataset. In addition, the automatized, online tool Genome Detective was suitable for analysis of VirCapSeq metagenomics and long amplicon sequencing by ONT.

- Metabarcoding and VirCapSeq metagenomics used an Illumina sequencing platform, while the long amplicon sequencing uses ONT, that may be easily implemented in laboratories since its sequencers have an affordable entry price. Further developments of the latter method would thus be a great interest for the scientific and regulatory community analysing contaminated foods.

- Short metabarcoding targeting separate RdRp and VP1 sequences was suitable to apply on natural samples, while amplification of the RdRp-VP1 junction led to additional biases and identification of false recombinants (chimera) and is thus not recommended.

- As expected, metabarcoding allowed the identification of multiple NoV strains in the same sample, at the genotype or P-type level. VirCapSeq metagenomics on stool samples yielded long sequences (even full genomes) suitable for strain identification. On shellfish TSS, this method yielded partial genomes, some long enough but others too short to unambiguously identify strains. Sequences were even shorter when applied on real environmental and food samples (BLS and OB BMS) and sometimes fell outside the typing region, impairing robust genotype / P-type identification using the Norovirus Typing Tool 2.0.

***Consortium Objective 1 (CO1): To perform molecular characterization by (whole) genome sequencing of a selection of NoV positive oyster samples using three currently available protocols for NGS analysis.***



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

The consortium partners proposed to share their knowledge and experiences in terms of shellfish microbiology, NGS, bioinformatics, and molecular epidemiology to characterize NoVs sequences in oyster samples collected during the European Baseline Survey of norovirus in oysters. Based on the conclusions of the Baseline Survey study (EFSA J. 2019), that presented the prevalence and concentrations of NoV in oyster samples, the complete molecular characterization of NoV strains present in these samples was a real challenge. Within this project different NGS approaches were thus assessed on test samples to maximize the outcome of the study.

- The high sensitivity of the metabarcoding approach allowed the characterization of NoV sequences in 212 oyster digestive tissue (DT) samples collected during the EFSA baseline study (BLS samples).
- A high diversity of NoV sequences was observed in BLS oyster samples by metabarcoding and, to a lesser extent, by VirCapSeq metagenomics, some being contaminated by several strains of NoV.
- Interestingly, NoV sequences belonging to genogroups known to infect animals (bovine, sea mammals) were also detected, demonstrating the efficacy of these approaches to analyse NoV diversity. This observation suggests that the design and use of primers targeting NoV outside the GI and GII genogroup should be considered for a scientific project in an one-health approach.
- Diverse NoV sequences were also retrieved from outbreak (OB)-related stool samples, with full or half genomes, including evidence of co-infecting NoV strains, and to a lesser extent from OB bivalve molluscan shellfish samples (OB BMS) samples, using the VirCapSeq metagenomics approach.

***Consortium Objective 3 (CO3): To assess how NGS outputs from the different methods can be used to determine the epidemiological relationship between NoV strains from different sources including the possible different genotype mixture within the same matrix and the potential viral recombination events.***

The metabarcoding approach applied on BLS samples allowed the detection of 242 NoV sequence clusters, out of the 212 samples analyzed, demonstrating the efficacy of this approach to investigate NoV diversity in such samples. More than half of these clusters were found in several samples collected from different countries. Interestingly, despite the primers targeting GI and GII, a few sequences assigned as GIII, GVII, GIV, GV and GVIII were identified. Shellfish being exposed both to human and animal contaminations have been suspected as a possible reservoir, that may favor zoonotic events or emergence of recombinant viruses (Zakhour et al 2010, Rincé et al 2018). Yet, such zoonotic events were never reported for NoV as of today.

The BLS were collected in Europe, with no analysis of prevalence and levels in different countries. The country anonymity requirements that needed to be followed, prevented any potentially informative geographical level analyses. Some differences were observed and, unsurprisingly, rare types were only present in samples from a minority of countries, and proportions of commoner types varied widely between different countries.

We compared the diversity of NoV genotypes and P-types identified in BLS samples to the one reported in the Noronet, the NoV database shared by European countries (van Beek et al 2018), at the same period than the BLS study. The variability noted is consistent with previous studies showing considerably different profiles of norovirus genotypes in different countries within Europe and across the world (Van Beek et al., 2018), and indicates that, despite the global transmission of norovirus, epidemiology of particular strains can vary considerably at a regional and national level.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Comparison of NoV strain diversity in clinical samples and BSL samples from the same countries was not possible due to country anonymity requirements, and that not all European countries analysed and reported NoV in clinical samples with the same approach. However the inter-country variability of NoV in BLS samples and NoroNet data indicated that, despite the global transmission of norovirus, the epidemiology of particular strains can vary considerably at a regional and national level. Comparing the temporal trends in abundance of genotypes and RdRp types in the two datasets it is clear that many of the most significant trends noticed were present both in shellfish and clinical samples. While Noronet collects a large set of NoV sequence data related to human outbreaks, many countries investigate outbreaks only based on epidemiological data on human samples without food analysis. NoV being mostly transmitted from human to human, the Noronet database also reports mainly community-acquired NoV, and foodborne or waterborne outbreaks are expected to be in minority. It is also known that strains associated to more severe symptoms are more likely to be investigated and reported in databases (Koopmans, 2013). Conversely, when being sick with moderate symptoms, some shellfish consumers do not seek a medical doctor (Lowther et al 2010, Koopmans, 2013) which hinders the reporting and sequencing of the implicated virus. Thus, NoV sequences from BMS, or from human cases following BMS consumption, are scarce in databases. Meanwhile, BLS samples sequenced here were collected in production areas or dispatch centers, and were not associated to human outbreaks. Therefore, the two datasets were expected to present some differences. For instance, the higher prevalence of GI sequences in BMS samples implicated in outbreaks was previously reported and linked to the selective transmission of NoV strains *via* oysters through specific binding to carbo-hydrate ligands (Le Guyader et al 2012, Yu et al. 2014). The high number of sequences obtained in this study will contribute to enrich the Noronet database.
Yet, our results mostly show that the same strains were detected in BLS and reported in Noronet at the time of the BLS study. Phylogenetic investigation of selected NoV GII genotypes also confirmed that the diversity of strains observed in BLS is a subset of the one of Noronet, and is not limited to one specific cluster of strains. Indeed, finding a correlation between sequences deposited in Noronet and BLS is expected as viruses detected in BLS mostly originate from human sewage, and therefore result from viruses circulating in the European population at the same time. This may imply that the genotypes that were more frequently detected in BLS than in Noronet were either excreted in higher concentration by the human population, more resistant to the depuration process or the coastal environmental conditions, or underreported in Noronet due to less severe symptoms. Conversely, those less observed in BLS could be genotypes or P-types that are less amplified by the primers we selected, as may be the case for GII.P16. Studying the diversity of NoV in sewage, that should represent all strains present in the human population, would allow to bypass the bias linked to Noronet reporting and investigate directly the selection operated during shellfish contamination. In addition, few sequences retrieved from the BLS corresponded to NoV of animal origin, such as GIII (bovine) but also GNA1, that was previously identified in harbor porpoise. Thus, apart from the vast majority of GI and GII HuNoV originating from human sewage, some animal sources appear to contribute occasionally to the diversity of NoV observed in shellfish, which raises the question of designing dedicated primers sets to target other genogroups than GI and GII. NoV zoonotic events have not yet been evidenced but need to be further investigated (Zakhour et al. 2010; Villabruna et al. 2021).

While most samples were analyzed using the metabarcoding method, the VirCapSeq metagenomics was also applied to a subset of the BLS samples and to some samples related to confirmed shellfish-borne outbreaks. When applied to a selection of 20 samples from the baseline survey, only five samples yielded NoV contigs. Besides, these contigs were very short and often fell outside the typing regions, impairing confident assignation to a NoV genotype or P-type. A possible explanation for this somewhat disappointing result is the long storage of the samples (BMS digestive tissues) that were also submitted to a new extraction and purification method. Partial degradation of viral particles under freezing condition or during shipping from one country to another probably occurred. The main interest of this VircapSeq metagenomic approach, consisting in providing long contigs, was clearly not observed. However, three samples yielded more than one NoV sequence using this method, and in one instance, a NoV GIII contig was identify by this technique and not by metabarcoding. This confirms the interest



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

of this method to detect a broad diversity of viral sequences. Further optimizations are needed to improve the sequencing sensitivity. It could also be interesting to improve shellfish DT storage and to build a shellfish tissue biobank to be able to retrospectively detect human pathogens, that we may be not able to identify now in the future, or serve as reference material after the emergence of novel human pathogens.

Thus, our final conclusions regarding how NGS tools can be used to study epidemiological relationship between NoV strains in oyster samples and humans are the following:

- At least two methods proposed here will offer the capacity to investigate the genotype mixture in BMS samples: the metabarcoding and the metagenomic methods.
- For now, the metabarcoding seems to be the more reliable method, presenting a high sensitivity and reproducibility and being easy to use even for the sequence analysis. However, the drawback of this method is that only the genotype diversity is accessible and that it does not allow to identify novel recombinants or emerging NoV strains due to the separate amplification of RdRp and VP1, and the specificity of the primer sets. Such new strains are more likely to be observed first in clinical samples.
- The metagenomics method showed promising results on fresh bio-accumulated test samples but despite an enrichment in viral sequences during the library preparation, it did not perform well on oyster DT from the BLS prevalence study and the OB set. This may be due to a decrease of NoV concentration and loss of genome integrity in the tissues during their storage at -20°C. However, through its potential to characterize all types of viral sequences, including virus discovery, this method still constitutes an important opportunity for food safety in the coming years and should be further optimized.
- Sequences obtained in the oysters encompassed most of the genotypes and P-types known to circulate in the European population at the time of the BLS study, but they did not reflect exactly the diversity observed in clinical samples, based on Noronet database. Some variations in sequence distribution were observed, that may be linked to differences in stability or bio-accumulation in shellfish (as for GI.1 compared to GII.4), but also to technical difficulties in amplifying certain strains (such as the GII.P16 using our GII-RdRp primers).

***Consortium Objective 4 (CO4): To perform retrospective analysis of samples from confirmed food-borne outbreaks in order to investigate the feasibility of NGS as tool for identification of sources and transmission pathways.***

To formally identify the contamination source, the gold standard is to demonstrate identical sequences in the food and in the stool of sick consumers. This is difficult to achieve when analyzing shellfish borne outbreaks, as BMS from the same batch as the consumed meal are rarely available, and clinical specimens (patient stool) are also rarely collected. In this study, from the matched OB BMS and stool samples, not all OB stool samples were finally available for sequencing and not all OB BMS samples resulted in NoV sequences. Metagenomic sequencing of stool samples resulted in (almost) complete genomes, and multiple genotypes could be detected within several samples, indicative of co-infections. Metagenomic sequencing of OB BMS samples only resulted in short contigs, but multiple genogroups could be detected within one sample. Three of the OB BMS NoV contigs could be matched to NoV genomes sequenced in the corresponding human sample, but they were too short to exclude other sources of infection for the investigated outbreaks. Also, for the other BMS samples, metagenomic sequencing only resulted in short sequences (max 374 nucleotides). These short sequences can in some cases be used to dismiss potential transmission chains, but especially if the detected genotype is highly prevalent, they do not carry enough information to definitely confirm transmission chains between contaminated BMS and NoV outbreak in humans.

Regarding the availability of outbreak-related shellfish batches, one option is to collect BMS from the same area providing that no new contaminant event occurred. For this study we collected as much



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

samples as possible from our archival collection but these were kept for many years at -20° and degradation may have occurred. We verified this by comparing the NoV concentrations after the new extraction for the metagenomic analysis and observed at least a one-log decrease compared to the initial concentration (not shown). Method impact can be excluded as the purification steps added for the metagenomic approach increase the concentrations compared to the ISO15216 protocol, as seen previously (Strubbia et al. 2019b) and here with TSS2.

On stool samples, this technique allowed to obtain several complete NoV genomes or very large contigs covering half the genome, highlighting the impact of viral load and virus integrity on the outcome of this method. In addition, VirCapSeq metagenomics showed the co-occurrence of several NoV strains in most of the stool samples, an evidence of co-infection that may be overlooked when deep sequencing is not performed. Such co-infection with multiple viral strains has been reported before in foodborne or waterborne outbreaks (Huang et al. 2013; Lysen et al 2009), and especially when shellfish were involved (Kageyama et al 2020; Le Guyader et al 2008; Wang et al 2015)

In this work, only one outbreak investigation provided sequences both from stool samples and from the implicated oyster sample. If the same genotypes were detected in the different types of samples, the short length of the sequences in the oyster sample prevented to reach a definitive conclusion, but rather suggested a possible link. When oysters are contaminated by several strains, it is somewhat complicated to match exactly with the clinical samples (Le Mennec 2017, Le Guyader et al 2008, 2014, Rasmussen et al. 2016). When multiple viruses are present this is even more difficult (Wang et al. 2015, Le Guyader et al. 2008). Such examples highlight the need to improve this method to generate longer sequenced and to be able to more precisely investigate shellfish-borne outbreaks and more generally foodborne outbreaks.

Thus, our final conclusions regarding how retrospective analysis of samples from confirmed food-borne outbreaks in order to investigate the feasibility of NGS as tool for identification of sources and transmission pathways are the following:

- To perform retrospective analysis, oyster DT need to be kept in conditions that ensure preservation of the viral particles. After 4 years at -20°C some degradation in viral particles occurred and prevented a good retrospective analysis of the viral diversity. Laboratories should consider to perform sequencing analysis rapidly after sampling, or preserve samples at -80°C.
- BMS implicated in outbreaks presented lower concentrations that the samples collected from the production areas (EFSA J. 2019), and thus even more sensitive methods need to be applied.
- The metabarcoding method can be applied to nucleic acids used for the official analysis (ISO 15216 method) while the metagenomic method need additional extraction with extended purifications. When dealing with BMS outbreaks this mean that additional portions of shellfish digestive tissues need to be available, which is not always possible.
- In the absence of clinical samples from consumers, and if the BMS samples collected for the analysis are not a leftover of the meal, the metabarcoding method should provide important information in terms of strain diversity, and the type of NoV. When combined with the quantification approach this will be sufficient to demonstrate the BMS contamination and probable implication in the clinical cases.
- When clinical samples are available as well as BMS from the batch suspected to be implicated in the outbreak, both the metagenomic and metabarcoding methods should be applied as soon as possible. Only long contigs obtained from both type of samples (clinical and food) will determine the BMS implication or not.
- It is important to continue the data collection from foodborne outbreaks, in relation with the Noronet database.

Overall, this study compared for the first time the advantages and the limitations of three NGS methods to investigate norovirus diversity in a specific environment/food matrix, ie. shellfish samples, in particular oysters. Our work provides a large unique collection of norovirus sequences detected in



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

oysters. We have shown that next generation sequencing methods are ready to substitute current methods despite some limitations. Metabarcoding is highly sensitive and allows to better capture the diversity of noroviruses present in BMS. Other methods such as ONT and metagenomics show promise, but require further optimization with regards to sensitivity. Importantly, these methods would allow for the identification of transmission chains and recombination events. Currently metagenomic sequencing is sensitive enough to investigate the human outbreak samples but not BMS samples. As the field of capture-based metagenomic sequencing is fast developing, we anticipate that these techniques will soon be suitable for BMS, although perhaps not to equal sensitivity as metabarcoding. Within this study we have used NGS to show that there is a wide diversity of norovirus genotypes in BMS which reflects the diversity observed for human samples. Observations from the data collected and presented in this report raise new scientific questions such as the impact of preferential detection of some genotypes in shellfish compared to those clinically reported or the impact of animal norovirus sequences. The optimized protocols and results presented in this study will aid future BMS research and risk assessment. Based on results presented here, it is clear that too many European shellfish samples are exposed to sewage pollution as evidenced by the multiple sequences detected in shellfish samples, highlighting that we still need to enhance coastal water quality. Providing efficient tools to describe norovirus diversity, as presented in this report, will help to enhance shellfish quality and thus to increase consumer safety.

# 5. Additional supporting information

Annex A – Supplementary Excel file containing the TSS master spreadsheet. This file contains the detailed information on TSS1 and TSS2 composition as well as their results of metabarcoding and metagenomics analysis.

Metabarcoding data include the genotypes and P-types identified, number of reads corresponding to each cluster identified in each sample or replicate and their sequences in fasta format.

Metagenomics data include the sequences of contigs corresponding to different strains present in these samples identified through their genotype and P-type. The additional contigs, either recombinants between two strains or those that could match two strains because they share the same genotype or P-type are also shown.

The raw sequences obtained for each sample or replicate in this project can be accessed at the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home) using the accession numbers indicated in the file.

Annex B – Supplementary Excel file containing the BLS master spreadsheet. This Excel file contains detailed information on metabarcoding and metagenomics results from BLS oyster samples. The metabarcoding data consist of:

- list of samples selected for this project and their metadata (i.e. sampling period),
- initial RT-qPCR results for norovirus testing and metabarcoding N-PCR results,
- sequencing results from this project including raw sequences.

The metagenomics data include:

- list of samples selected for this project and their metadata (i.e. sampling period),
- initial RT-qPCR results for norovirus testing, and cDNA qPCR results before library preparation,
- metagenomics sequencing results from this project (i.e. raw and clean reads, contigs assigned),
- norovirus contigs sequences and identification using Norovirus Typing Tool and BLASTn.

The raw sequences obtained for each sample in this project can be accessed at the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home) using the accession numbers indicated in the file.

Annex C - Supplementary Excel file containing the OB master spreadsheet. This Excel file contains detailed information on shellfish and human stool samples from shellfish-related outbreaks and their metagenomics sequencing results:

- sample metadata (sampling date, shellfish species and link to consumed batch),
- initial outbreak testing results and qPCR results on cDNA,



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

- metagenomics sequencing results (i.e. raw reads, clean reads and contigs assigned to norovirus using the bioinformatics pipeline, for all samples that were sequenced),
- norovirus contigs sequences and identification using Norovirus Typing Tool and BLASTn.

The raw sequences obtained for each sample in this project can be accessed at the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home) using the accession numbers indicated in the file.

Annex D - Supplementary PDF file containing the description of Noronet data. This PDF file describes the number of European NoV sequences submitted to NoroNet per GI and GII genotype and P-type between November 2016 and October 2018.

Annex A, B, C and D can be found in the online version of this output under the 'Supporting information' section at:

https://efsa.onlinelibrary.wiley.com/doi/10.2903/sp.efsa.2022.EN-7348#support-information-section

## References

Andrew S,2010. FastQC A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA, 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology 19: 455-477.

Beuret C, Kohler D, Baumgartner A, Luthi TM, 2002. Norwalk-like virus sequences in mineral waters: one year monitoring of three brands. Applied and Environmental Microbiology 68(4): 1925-1931.

Bisseux M, Colombet J, Mirand A,Roque-Afonso AM, Abravanel F, Izopet J, Archimbaud C, Peigue-Lafeuille H, Debroas D, Bailly JL, Henquell C, 2018. Monitoring human enteric viruses in wastewater and relevance to infections encountered in the clinical setting: a one-year experiment in central France, 2014 to 2015. Euro Surveillance Feb; 23(7): 17-00237.

Bonny P, Schaeffer J, Besnard A, Desdouits M, Essia Ngang JJ, Le Guyader FS, 2021. Human and Animal RNA Virus Diversity Detected by Metagenomics in Cameroonian Clams. Frontiers in Microbiology 12: 770385.

Boratyn GM, Thierry-Mieg J, Thierry-Mieg D, Busby B, Madden TL, 2019. Magic-BLAST, an accurate RNA-seq aligner for long and short reads. BMC Bioinformatics. 20(1):405.

Briese T, Kapoor A, Mishra N, Kumar A, Jabado OJ, Lipkin WI, 2015. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. mBio. 6: e01491-15.

Buchfink B, Xie C, Huson DH, 2015. Fast and sensitive protein alignment using diamond. Nature Methods 12: 59-60.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL, 2009. BLAST+: Architecture and applications. BMC Bioinformatics, 10(1), 421.

Chhabra P, de Graaf M, Parra G, Chi-Wai Chan M, Green K, Martella V, Wang Q, White P, Katayama K, Vennema H, Koopmans M, Vinje J, 2019. Updated classification of norovirus genogroups and genotypes. Journal of General Virology 100(10): 1393-1406.

Chen S, Zhou Y, Chen Y, Gu J., 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor." Bioinformatics 34: i884-i890.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

da Silva A, Le Saux J-C, Parnaudeau S, Pommepuy M, Elimelech M, Le Guyader FS, 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: 7891-7897.

de Graaf M, van Beek J, Koopmans MPG, 2016. Human norovirus transmission and evolution in a changing world. Nature Reviews Microbiology 14: 421-433.

Desdouits M, Wacrenier C, Ollivier J, Schaeffer J, Le Guyader FS, 2020. A targeted metagenomics approach to study the diversity of norovirus GII in shellfish implicated in outbreaks. Viruses 12(9): 978.

Desdouits M, de Graaf M, Strubbia S, Oude Munnink BB, Kroneman A, Le Guyader FS, Koopmans M, 2020. Novel opportunities for NGS-based one health surveillance of foodborne viruses. One Health Outlook 2, 14.

EFSA, 2011. Scientifc opinion on an update on the present knowledge on the occurence and control of foodborne viruses. EFSA Journal, 9: 1-96.

EFSA, 2012. Scientific opinion on norovirus (NoV) in oysters: methods, limits and control options. EFSA Journal 10: 39.

EFSA, 2016. Technical specifications for a European baseline survey of norovirus in oysters. EFSA Journal 14: 4414.

EFSA, 2019. Analysis of the European baseline survey of norovirus in oysters. EFSA Journal 17: 5762.

Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, maman S, Hernandez-Raquet G, Combes S, pascal G, 2018. FROGS: Find, Rapidly, OTUs with Galaxy solution. Bioinformatics. 34: 1287-1294.

Ettayebi K, Crawford SE, Murakami K, Broughman JR, Karandikar U, Tenge VR, Neill FH, Blutt SE, Zeng X-L, Qu L, Kou B, Opekum AR, Burrin D, Graham DY, Ramani S, Atmar RL, Estes MK, 2016. Replication of human noroviruses in stem cell-derived human enteroids. Science 353: 1387-1393.

Ewels P, Magnusson M, Lundin S, Käller M, 2016. MultiQC : Summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32(19), 3047-3048.

Huang J, Xu X, Weng Q, Hong H, Guo Z, He S, Niu J, 2013 Serial Foodborne Norovirus Outbreaks Associated with Multiple Genotypes. PLoS One, 8(5): e63327.

Imamura S, Haruna M, Goshima T, Kanezashi H, Okada, Akimoto K, 2016. Application of next-generation sequencing to investigation of norovirus diversity in shellfish collected from two coastal sites in Japan from 2013 to 2014. Japanese Journal of Veterinary Research 64(2): 113-22.

Imamura S, Kanezashi H, Goshima T, Haruna M, Okada T, Inagaki N, Uema M, Noda M, Akimoto K, 2017. Next-generation sequencing analysis of the diversity of human noroviruses in japanese oysters. Foodborne Pathogens and Disease 14(8):465-471.

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

Kageyama T, Shinohara M, Uchida K, Fukushi S, Hoshino B, Kojima S, Takai R, Oka T, Takeda N, Katayama K, 2020 Coexistence of multiple genotypes,, including newly identified genotypes, in outbreaks of gastroenteritis due to norovirus in Japan. Journal of Clinical Microbiology 42(7): 2988-95.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Kojima S, Kageyama T, Fukushi S,Hoshino FB, Shinohara M, Uchida K, Natori K, Takeda N, Katayama K, 2002. Genogroup -specific PCR primers for detection of Norwalk-like viruses. Journal of Virological Methods 100: 107-114.

Koopmans M, 2013. Surveillence strategy for early detection of unusual infectious disease events. Current Opinion in Virology 3(2): 185-91.

Kroneman A, Vennema H, Deforche K, Avoort HV, Penaranda S, Oberste MS, Vinje J, Koopmans M, 2011. An automated genotyping tool for enteroviruses and noroviruses. Journal of Clinical Virology 51(2):121-5.

Kroneman A, Vega E, Vennema H, Vinjé J, White PA, Hansman G, Green K, Martella V, Katayama K, Koopmans M, 2013. Proposal for a unified norovirus nomenclature and genotyping. Archives of Virology 158: 2059-2068.

Krueger F, 2012. Trimgalore A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

Langmead B,  Salzberg SL, 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4), 357-359.

Le Guyader F, Estes MK, Hardy ME, Neill FH, Green J, Brown DW, Atmar RL, 1996. Evaluation of a degenerate primer for the PCR detection of human caliciviruses. Archives of Virology 141: 2225-2235.

Le Guyader FS, Le Saux J-C, Ambert-Balay K, Krol J, Serais O, Parnaudeau S, Giraudon H, Delmas G, Pommepuy M, Pothier P, Atmar RL, 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.

Le Guyader FS., Atmar R., Le Pendu J. 2012 Transmission of viruses through shellfish: when specific ligands come into play. Current Opinion in Virology 2:103-10.

Le Mennec C, Parnaudeau S, Rumebe M, Le Saux JC, Piquet JC, Le Guyader FS,, 2017. Follow-up of norovirus contamination in an oyster production area linked to repeated outbreaks. Food and Environmental Virology 9: 54-61.

Loisy F, Atmar RL, Guillon P,Le Cann P, Pommepuy M, Le Guyader FS, 2005. Real-time RT-PCR for norovirus screening in shellfish. Journal of Virological Methods 123: 1-7.

Lowther J, Avant JM, Gizynski K, Rangdale RE, Lees DN, 2010. Comparison between quantitative real-time reverse transcription PCR results for norovirus in oysters and self-reported gastroenteric illness in restaurant customers. Journal of Food Protection 73(2): 305-11.

Lowther JA, Gustar NE, Hartnell RE, Lees DN, 2012. Comparison of norovirus RNA levels in outbreak-related oysters with background environmental levels. Journal of Food Protection 75: 389-393.

Lowther J., Bosch A., Butot S., Ollivier J., Mäde D., Rutjes S., Hardouin G., Lombard B., In't Veld P., Leclerq A, 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.

Lun JH, Hewitt J, Sitabkhan A, Eden JS, Tuipulotu DE, Netzler NE, Morell L, Merif J, Jones R, Huang B, Warrilow D, Ressler KA, Ferson MJ, Dwyer DE, Kok J, Rawlinson WD, Deere D, Crosbie ND, White PA, 2018. Emerging recombinant noroviruses identified by clinical and waste water screening. Emerging Microbes and Infections 7: 50.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Lysén M, Thorhagen M, Brytting M, Hjertqvist M, Andersson Y, Hedlund KO, 2009. Genetic diversity among food-borne and waterborne norovirus strains causing outbreaks in Sweden. Journal of Clinical Microbiology 47(8): 2411-8.

Maalouf H, Zakhour M, Le Pendu J, Le Saux JC, Atmar RL, Le Guyader FS, 2010 Distribution in tissue and seasonal variations of norovirus genogroup I and II ligands in oysters. Applied and Environmental Microbiology 76: 5621-5630.

Maalouf H, Schaeffer J, Parnaudeau S, Le Pendu J, Atmar RL, Crawford SE, Le Guyader FS, 2011. Strain-dependent norovirus bioaccumulation in oysters. Applied and Environmental Microbiology 77: 3189-3196.

McLeod C, Polo D, Le Saux J-C, Le Guyader FS, 2017. Depuration and relaying: a review on potential removal of norovirus from oysters. Comprehensive Reviews in Food Science and Food Safety 16: 692-706.

Nieuwenhuijse DF, Aude Munnink B, Phan My VT, Global Sewage Surveillance project consortium; Munk P, Venkatakrishnan S, Aerstrup FM, Cotten M, Koopmans MPG, 2020. Setting a baseline for global urban virome surveillance in sewage. Scientific Reports 10: 13748.

Nooij S, Schmitz D, Vennema H, Kroneman A, Koopmans MPG, 2018. Overview of virus metagenomic classification methdos and their biological applications. Frontiers in Microbiology 9: 749.

Nurk, Sergey, Dmitry Meleshko, Anton Korobeynikov, et Pavel A. Pevzner. MetaSPAdes: A New Versatile Metagenomic Assembler. Genome Research 27, nº 5 (5 janvier 2017): 824-34.

Oshiki, M, Miura T, Kazama S, Segawa T, Ishii S, Hatamoto M, Yamaguchi T, Kubota K, Iguchi A, Tagawa T, Okubo T, Uemura S, Harade H, Kobayashi N, Araki N, Sano D, 2018. Microfluidic PCR amplification and MiSeq amplicon sequencing techniques for high-throughput detection and genotyping of human pathogenic RNA viruses in human feces, sewage, and oysters. Frontiers in Microbiology 9: 830.

Pu J, Miura T, KazamaS, Konta Y, Azraini ND, Itp E, Ito H, Omura T, Watanabe T, 2018. Weekly variations in norovirus genogroup II genotypes in Japanese oysters. International Journal of Food Microbiology 284: 48-55.

Rajko-Nenow P., Waters A., Keaveney S., Flannery J., Tuite G., Coughlan S., O'Flaherty V., Doré W. (2013) Norovirus genotypes present in oysters and in effluent from a wastewater treatment plant during the seasonal peak of infections in Ireland in 2010. Appl Environ Microbiol 79(8):2578-87.

Rasmussen LD, Schultz AC, Uhrbrand K, et al. 2016. Molecular evidence of oysters as vehicle of norovirus GII.P17-GII.17. Emerging Infectious Diseases 22: 2024-2025.

Rincé A, Balière C, Hervio-Heat D, Cozien J, Lozach S, Parnaudeau S, Le Guyader FS, Le Hello S, Giard JC, Sauvageot N, Benachour A, Strubbia S, Gourmelon M, 2018. Occurrence of Bacterial Pathogens and Human Noroviruses in Shellfish-Harvesting Areas and Their Catchments in France. Frontiers in Microbiology 9: 2443.R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Shin SB, Oh EG, Yu H, Son KT, Lee HJ, Park JY, Kim JH, 2013. Genetic diversity of noroviruses detected in oysters in Jinhae Bay, Korean Food Science and Biotechnology 22:1–8.

Strubbia S, Phan MVT, Schaeffer J, Koopmans M, Cotton M, Le Guyader FS, 2019a. Characterization of norovirus and other human enteric viruses in sewage and stool samples through next-generation sequencing. Food and Environmental Virology 11(4): 400-409.




The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Strubbia S, Schaeffer J, Oude Munnink BB,Besnard A, Phan MVT, Nieuwenhuijse DF, de Graaf M, Schapendonk CME, Wacrenier C, Cotton M, Koopmans MPG, Le Guyader FS, 2019b. Metavirome sequencing to evaluate norovirus diversity in sewage and related bioaccumulated oysters. Frontiers in Microbiology 10, 2394.

Symes SJ, Gunesekere IC, Marshall JA, Wright PJ, 2007. Norovirus mixed infection in an oyster-associated outbreak: an opportunity for recombination. Archives of Virology 152: 1075-1086.

Tohma K, Lepore CJ, Degiuseppe JI, Khamrin P, Saito M, Mayta H, Nwaba AUA, Ford-Siltz LA, Green KY, Galeano ME, Zimic M, Stupka JA, Gilman RH, Maneekam N, Ushijima H, Parra GI,2021. Genome-wide analyses of human noroviruses provide insights on evolutionary dynamics and evidence of coexisting viral populations evolving under recombination constraints. PLoS Pathogens 17(7): e1009744.

van Beek J, van der Eijk A, Fraaij P, Caliskan K, Cransberg K, Dalinghaus M, Hoek R, Metselaar H, Roodnat J, Vennema H, Koopmans M, 2017. Chronic norovirus infection among solid organ recipients in a tiertiary care hospital, the Netherlands, 2006-2014. Clinical Microbiology and Infection 23(4)265.e9 - 265.e13.

van Beek J, de Graaf M, Al-Hello H, Allen DJ, Ambert-Balay K, Bottldorn N, Brytting M, Buesa J, Cabrerizo M, Chan M, Cloak F, Di Bartolo I, Guix S, Hewitt J, Iritani N, Jin M, Johne R, Lederer I, Mans J, Martella V, Maunula L, McAlister G, Niendorf S, Niesters HG, Podkolzin AT, Poljsak-Prijateji M, Dam Rasmussen L, Reuter G, Tuite G, Kroneman A, Vennema H, Koopmans MPG, NoroNet, 2018. Molecular surveillance of norovirus, 2005-16: an epidemiological analysis of data collected from the NoroNet network. Lancet Infectious Diseases 18(5): 545-553.

Verhoef L, Vennema H, van Pelt W, Lees D, Boshuizen H, Henshilwood K, Koopmans M, Food-Borne Viruses in Europe Network. 2010. Use of norovirus genotype profiles to differentiate origins of foodborne outbreaks. Emerging Infectious Diseases 16(4): 617-24.

Villabruna N, Schapendonk CME, Aron GL, Koopmans MPG, de Graaf M, 2021. Human noroviruses attach to intestinal tissues of a broad range of animal species. Journal of Virology 95(3): e01492-20

Vilsker M, Moosa Y, Nooij S, Fonseca V, Ghysens Y, Dumon K, Pauwels R, Alcantara LC, Eynden EV, Vandamme A-M, Deforche K, de Oliveira T, 2019. Genome Detective: an automated system for virus identification from high-throughput sequencing data. Bioinformatics 35: 871-873.

Watson SJ, Welkers MRA, Depledge DP, Coulter E, Breuer JM, de Jong MD, Kellam P, 2013. Viral population analysis and minority-variant detection using short read next-generation sequencing. Philosophical Transactions of the Royal Society B. 368.

Wang Y, Zhang J, Shen Z, 2015 The impact of calicivirus mixed infection in an oyster-associated outbreak during a food festival. Journal of Clinical Microbiology 73: 55-63.

Wylie KM, Wylie T, Buller R, Herter B, Cannella MT, Storch GA, 2018. Detection of viruses in clinical samples by use of metagenomic sequencing and targeted sequence capture. Journal of Clinical Microbiology 56: e01123-01118.

Yu Y, Cai H, Hu L, Lei R, Pan Y, Yan S, Wang Y, 2015 Molecular Epidemiology of oyster-related human Noroviruses and their global enetic diversity and temporal-geographical distribution from 1983 to 2014. Applied and Environmental Microbiology 81(21):7615-24.

Yuen LK, Catton MG, Cox BJ, Wright P, Marshall JA, 2001. Heminested multiplex reverse-transcription-PCR for detection and idfferenciation of norwalk-like virus genogroups 1 and 2 in fecal samples. Journal of Clinical Microbiology 39: 2690-2694.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Zakhour M, Maalouf H, Di Bartolo I, Haugarreau L, Le Guyader FS, Rouoven-Clouet N, Le Saux J-C, Ruggeri FM, Pommepuy M, Le Pendu J, 2010 Bovine norovirus: carbohydrate ligand, environmental contamination, and potential cross-species transmission via oysters. Applied and Environmental Microbiology 76(19):6404-11.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

## Abbreviations

| | |
|---|---|
| bp | Base pair |
| BLS | Baseline Survey of norovirus in oysters |
| BMS | Bivalve Molluscan Shellfish |
| cDNA | Complementary DNA |
| Ct | Cycle threshold |
| CO | Consortium Objective |
| DT | Digestive tissues |
| EFSA | European Food Safety Authority |
| ESO | EFSA Specific Objective |
| FROGS | Find Rapidly OTU with Galaxy Solution |
| GI | Norovirus Genogroup I |
| GII | Norovirus Genogroup II |
| LOD | Limit of detection |
| LOQ | Limit of quantification |
| NGS | Next Generation Sequencing |
| N-PCR | Nested Polymerase chain reaction |
| NoV | Norovirus |
| nt | Nucleotide |
| OB | Outbreak-related samples |
| PBS | Phosphate-buffered saline |
| PCR | Polymerase chain reaction |
| qPCR | quantitative (real-time) Polymerase chain reaction |
| RdRp | RNA-dependent RNA polymerase |
| RT-PCR | Reverse transcription – polymerase chain reaction |
| RT-dPCR | Reverse transcription – digital PCR |
| RT-qPCR | Reverse transcription (RT)- quantitative (real-time) PCR |
| TE | Tris-EDTA buffer, 10mM |
| TSS1 | Test sample set 1 |
| TSS2 | Test sample set 2 |
| VirCapSeq VERT | The VirCapSeq-VERT Capture Panel which covers the genomes of 207 viral taxa known to infect vertebrates (including humans) and enables detection of viral sequences in complex sample types |
| VP1 | Viral protein 1 |



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

## Appendix A - Validation of TSS1 libraries for the metabarcoding

The cDNA were obtained for all the replicates as confirmed by qPCR (data not shown). The libraries obtained following N-PCR were controlled by electrophoresis in 1% agarose gel (Figure A1). The sizes of amplicons with tags sequences, used for this study were 349 bp (VP1), 342 bp (RdRp), 514 bp (RdRp-VP1) for the GI and 364 bp (VP1), 212 bp (RdRp) and 557 bp (RdRp-VP1) for the GII, respectively. No amplicon was detected in negative controls (water). Three replicate amplicons were generated for each sample, and are labelled R1 (corresponds to extraction series A), R2 (series B) and R3 (series C) on the gels.

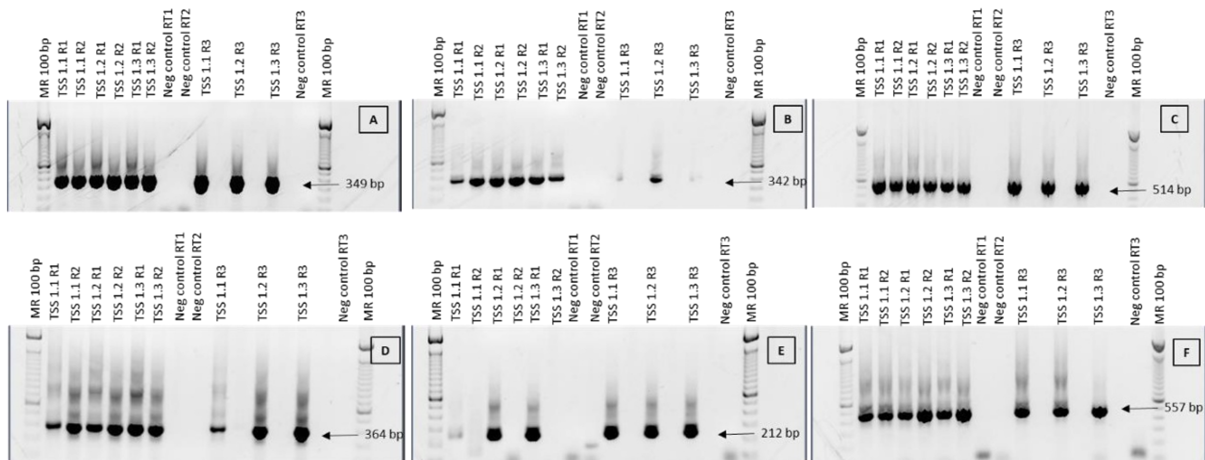



**FIGURE A 1: ELECTROPHORESIS OF TSS1 LIBRARIES IN 1% AGAROSE GELS**

NoV GI: VP1 (A), RdRp (B) and RdRp-VP1 (C); NoV GII: VP1 (D), RdRp (E) and RdRp-VP1 (F)



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

# Appendix B - Preparation and validation of the TSS2 libraries for the metabarcoding

## B.1 RNA extraction and quantification at Ifremer

The test is performed in extraction series including each high to limit of detection contamination level (TSS2.1 to 2.4), negative control (TSS2.5) and background control (TSS2.0) samples. Results of quantification by qRT-PCR for the three series of extractions are presented in **Error! Reference source not found.**

**Table B 1: Quantification of RNA extracts from TSS2 samples.**

| Sample | Norovirus GI concentration (gc/g DT) | | | Norovirus GII concentration (gc/g DT) | | |
|---|---|---|---|---|---|---|
| | Replicate A | Replicate B | Replicate C | Replicate A | Replicate B | Replicate C |
| TSS2.1 | 3.46 x$10^{4}$ | 5.62 x$10^{4}$ | 3.14 x$10^{4}$ | 1.74 x$10^{5}$ | 1.68 x$10^{5}$ | 1.29 x$10^{5}$ |
| TSS2.2 | 3.20 x$10^{3}$ | 5.66 x$10^{3}$ | 5.28 x$10^{3}$ | 1.52 x$10^{4}$ | 1.97 x$10^{4}$ | 1.91 x$10^{4}$ |
| TSS2.3 | 1.71 x$10^{2}$ | 7.18 x$10^{2}$ | 8.34 x$10^{2}$ | 3.18 x$10^{3}$ | 1.50 x$10^{3}$ | 2.83 x$10^{3}$ |
| TSS2.4 | Detected,<LOQ | Detected,<LOQ | Not detected | Detected,<LOQ | Detected,<LOQ | Detected,<LOQ |
| TSS2.5 | Detected,<LOQ | Not detected | Not detected | Not detected | Not detected | Not detected |
| TSS2.0 | Not detected | Not detected | Not detected | Detected,<LOQ | Not detected | Detected,<LOQ |

## B.2 Preparation and validation of TSS2 libraries at Ifremer

The viral cDNA is obtained from all samples and replicates, except for the negative controls, as confirmed by NoV-specific qPCR (data not shown). The libraries obtained after N-PCR are controlled by agarose gel electrophoresis. The presence of a specific band of correct size was confirmed for all positive samples. The negative controls (TSS2.0, TSS2.5 and water) remained negative (Figure B 1). Three replicate amplicons were generated for each sample, and are labelled R1 (corresponds to extraction series A), R2 (series B) and R3 (series C) on the gels.

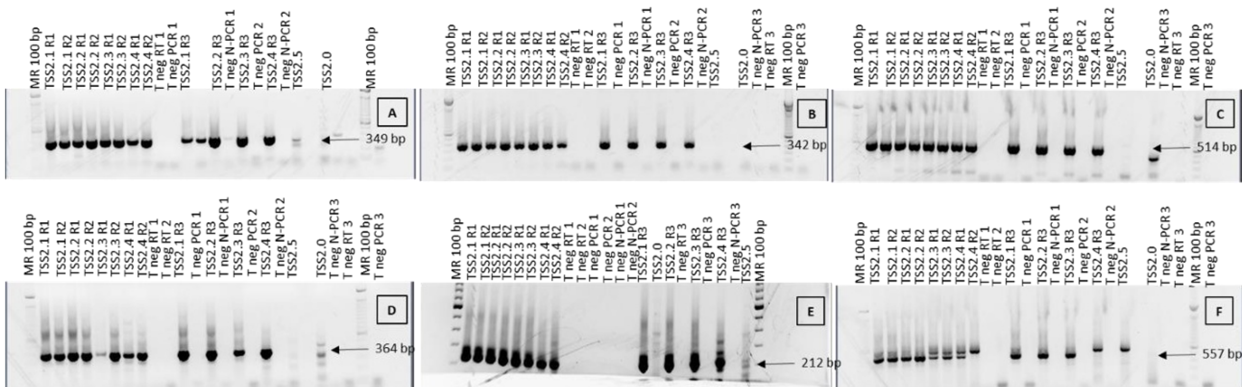



**Figure B 1: Electrophoresis of TSS2 libraries in 1% agarose gels at Ifremer**

NoV GI: VP1 (A), RdRp (B) and RdRp-VP1 (C); NoV GII: VP1 (D), RdRp (E) and RdRp-VP1 (F); R1-3 – replicates of TSS2 samples



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

## B.3 Preparation and validation of TSS2 libraries at Cefas

Following RNA extraction of all TSS2 samples, the presence of viral cDNA was confirmed by qPCR for all samples except for TSS2.1 R1 and TSS2.4 R1 which gave negative results for NoV GII only (results not shown). In the case of TSS2.4 R1 this is not unexpected due to the low contamination level in the sample, however the negative result for TSS2.1 R1 is difficult to explain due to the high levels of norovirus in the sample; this sample gave a Ct value of ~27.0 for NoV GI indicating high levels and a successful extraction, and the other TSS2.1 replicates both gave positive results for NoV GII (with Ct values 28-29 indicating high levels).

Due to time constraints it was decided to proceed directly to library preparation using N-PCR without repeating the qPCR analysis.

After N-PCR all libraries were subjected to agarose gel electrophoresis. The presence of a specific band of correct size was confirmed for all positive samples with the exception of all 3 replicates for TSS2.4 (very low contamination level) when amplified using the NoV GII RdRp-VP1 assay (long amplicons); in these cases a non-specific band of the wrong size was obtained. The negative controls (TSS2.0, TSS2.5 and water) remained negative with the exception of TSS2.5 for the NoV GI RdRP assay which gave a band of the expected size. TSS2.1 R1 gave positive results as expected for all 3 NoV GII assays, indicating that the negative qPCR result for this sample was an artefact, possibly caused by a pipetting error (Figure B 2).

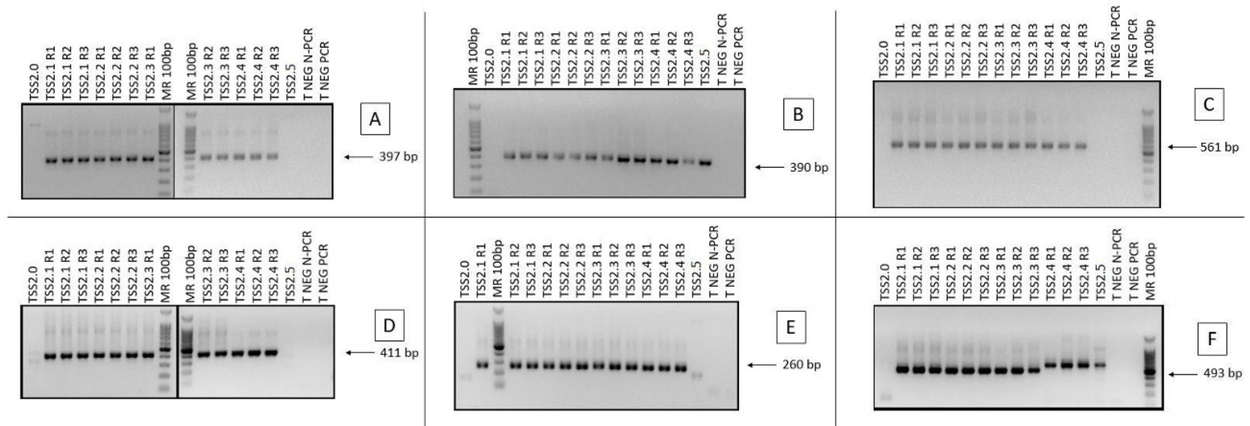



**FIGURE B 2: ELECTROPHORESIS OF TSS2 LIBRARIES IN 1% AGAROSE GELS AT CEFAS**

NoV GI: VP1 (A), RdRp (B) and RdRp-VP1 (C); NoV GII: VP1 (D), RdRp (E) and RdRp-VP1 (F); R1-3 – replicates of TSS2 samples.

NOTE: amplicon sizes are different compared with those in Figure A2 as the two laboratories use N-PCR primers targeting identical regions of the NoV genome but with different Illumina sequencing tags.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

# Appendix C - Preparation and validation of the TSS libraries for the VirCapSeq metagenomics approach

## C.1 VirCapSeq metagenomics on TSS1 samples

### C.1.1 Preparation and validation of TSS1 libraries at Ifremer

TSS1 samples consisting of RNA extract were submitted directly to reverse transcription and cDNAs were controlled by qPCR, as described previously. The cDNA was used for the library constructions, as described in section 2.2.5.3. The libraries, before pooling, were quantified by Qubit (Invitrogen, France), showing concentrations ranging from 10 to 100 ng/µL. Their profile and mean size were controlled by the use of 2100 Bioanalyzer (Agilent, France) (Figure C 1).

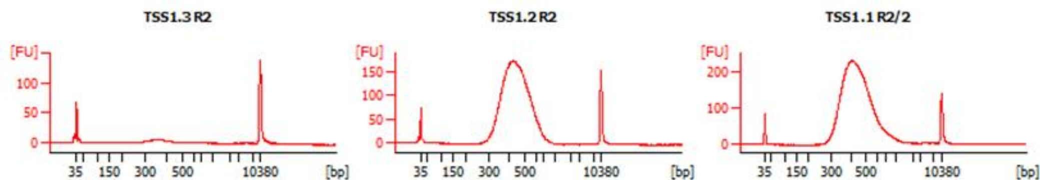



**FIGURE C 1: BIOANALYZER PROFILE OF TSS1 SAMPLES, EXAMPLE FOR REPLICATE 2.**

Nine libraries of TSS1 are pooled together before VirCapSeq capture step. Post capture qualification using 2100 BioAnalyzer (Agilent, France) performed in duplicate shows that the mean library length is 437 pb. Non-specific, secondary peak is observed, containing fragments of 2,000bp (Figure C 2). The pool concentration, measured by Qubit, is 27.6 ng/µL.

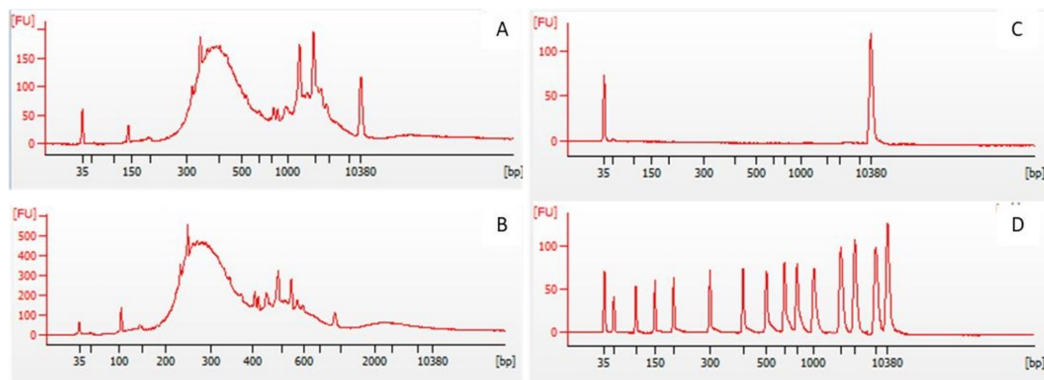



**FIGURE C 2: BIOANALYZER PROFILE OF LIBRARY POOL I**

A and B – library pool replicates 1 and 2; C – negative control; D- size marker

### C.1.2 Preparation and validation of TSS1 libraries at EMC

The libraries are prepared as described in 2.2.5.3.

The libraries, before pooling, are quantified by Qubit (Invitrogen, France), showing concentrations ranged from 25 to 40 ng/µL. Their profile and mean size are controlled by the use of 2100 Bioanalyzer (Agilent, France) (Figure C 3).



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

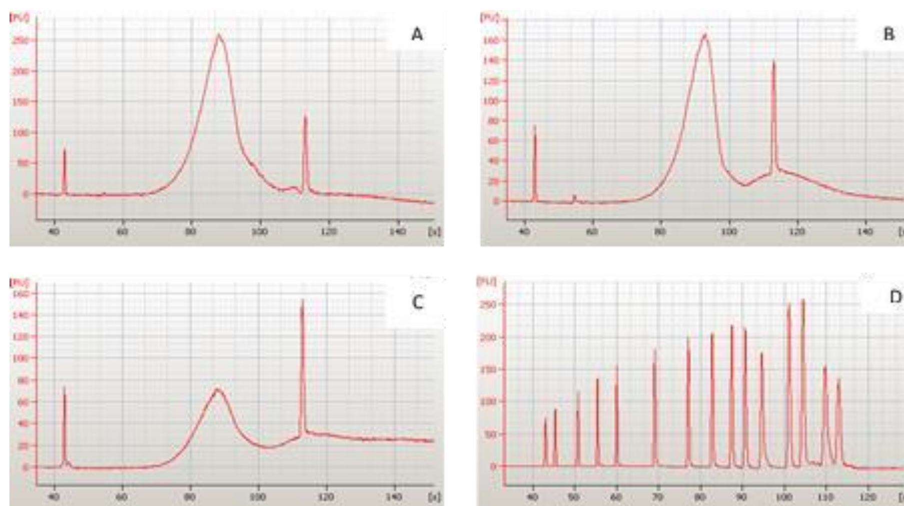


**FIGURE C 3: BIOANALYZER PROFILE OF TSS1 SAMPLES**

A - TSS1.1, B - TSS1.2, C - TSS1.3, D - size marker.

Nine libraries of TSS1 together with 15 libraries of TSS2 are mixed together to form four pools before VirCapSeq capture step. Post capture qualification using 2100 BioAnalyzer (Agilent, France) shows that the mean library length is 661 bp for pool I, 655 bp for pool II, 587 bp for pool III and 613 bp for pool IV. Their profile and mean size are controlled by the use of 2100 Bioanalyzer (Figure C 3). Pool III is likely diluted too much, resulting in a low profile. The pool concentrations, measured by Qubit (Invitrogen, France), are 36 ng/µL, 37.4 ng/µL, and 25.8 ng/µL for the pool I, II, and III respectively.

## C.2 VirCapSeq metagenomics on TSS2 samples

### C.2.1 Preparation and validation of TSS2 libraries at Ifremer

As for metabarcoding, three extraction series are performed, including each contamination level (TTS 2.1 to 2.4), DT of negative control (TSS 2.5) and DT for the background control (TSS 2.0). The extracted RNA was quantified by one-step qRT-PCR and showed less efficient RNA extraction in series C (Table C 1). As no process control virus was added to the samples, the extraction efficiency was not calculated.

**TABLE C 1: QUANTIFICATION OF RNA EXTRACTS FROM TSS2 SAMPLES.**

| Sample | Norovirus GI concentration (gc/g DT) | | | Norovirus GII concentration (gc/g DT) | | |
|---|---|---|---|---|---|---|
| | Replicate A | Replicate B | Replicate C | Replicate A | Replicate B | Replicate C |
| TSS2.1 | 3.09 x10$^5$ | 3.37 x10$^5$ | 2.27 x10$^3$ | 8.44 x10$^5$ | 1.02 x10$^6$ | 3.21 x10$^4$ |
| TSS2.2 | 1.54 x10$^4$ | 2.23 x10$^4$ | 7.03 x10$^3$ | 5.05 x10$^4$ | 5.71 x10$^4$ | 5.79 x10$^3$ |
| TSS2.3 | 8.58 x10$^2$ | 99.76 x10$^2$ | 9.49 x10$^2$ | 3678 x10$^3$ | 2.46 x10$^3$ | 1.83 x10$^3$ |
| TSS2.4 | 1.45 x10$^2$ | Detected,<LOQ | 1.25 x10$^2$ | 4.50 x10$^2$ | Detected,<LOQ | Not detected |
| TSS2.5 | Not detected | Not detected | Not detected | Not detected | Not detected | Not detected |
| TSS2.0 | Not detected | Not detected | Not detected | Not detected | Not detected | Not detected |

The cDNA obtained by reverse transcription with random hexamers, are controlled by qPCR (data no shown) and submitted to library preparation. The DNA concentration in libraries, quantified using Qubit (Invitrogen, France) range between 4 and 89 ng/µL. Their mean size and profile are controlled using



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

2100 Bioanalyzer (Agilent, France). Figure C 4 shows the profiles obtained for the replicates 2 of TSS2 samples.

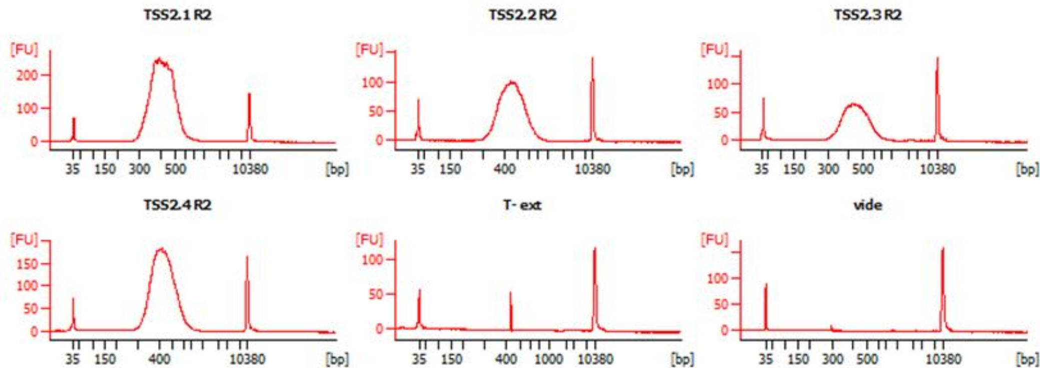



**FIGURE C 4: BIOANALYZER PROFILES OF TSS2 SAMPLES OF REPLICATES 2**

As all quality tests are valid, the libraries are mixed together to form two pools, one containing 8 libraries (pool II) and second with 7 libraries (pool III) before VirCapSeq capture step. Post capture qualification using 2100 BioAnalyzer performed in duplicate for each pool show that the mean library length was 430 bp for pool II and 436 bp for the pool III. The bioanalyzer profiles are presented on Figure C 5 and Figure C 6. The DNA concentration, measured by Qubit (Invitrogen, France), is 18.2 ng/µL and 18.5 ng/µL for the pool II and III, respectively.

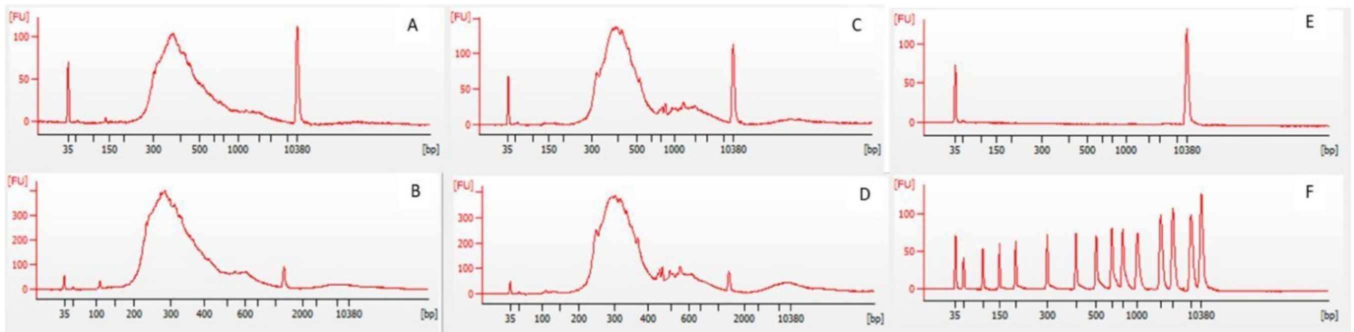



**FIGURE C 5: BIOANALYZER PROFILE OF LIBRARY POOL II AND III**

A and B – library pool II replicates 1 and 2; C and D - library pool III replicates 1 and 2; E – negative control; F – size marker.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

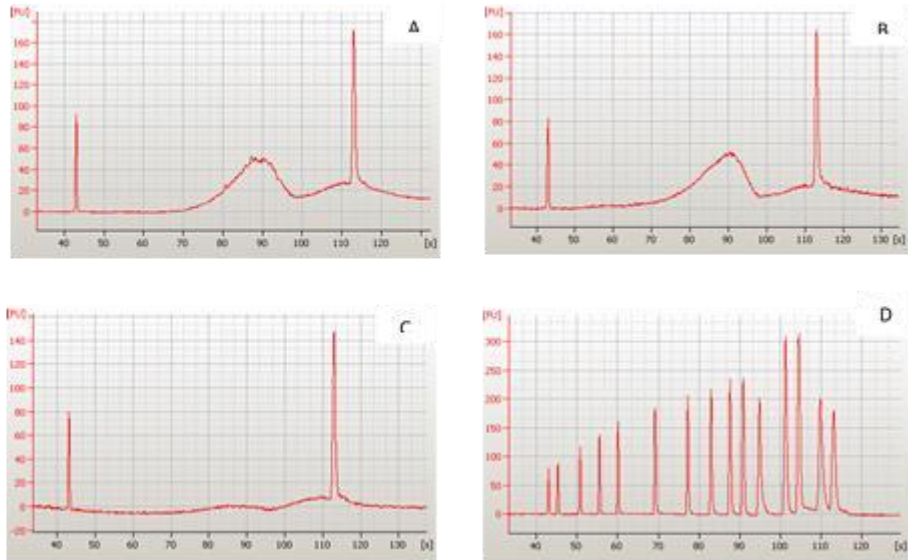



**FIGURE C 6: BIOANALYZER PROFILE OF LIBRARY POOLS**

A, B and C – library pool containing TSS1 samples; D- size marker

## C.2.2 Preparation and validation of TSS2 libraries at EMC

cDNA's obtained by reverse transcription with random hexamers, were submitted to library preparation. The DNA concentration in libraries, quantified using Qubit (Invitrogen, France) ranged between 30 and 77 ng/µL. Their mean size and profile were controlled using 2100 Bioanalyzer (Agilent, France) (Figure C 7).

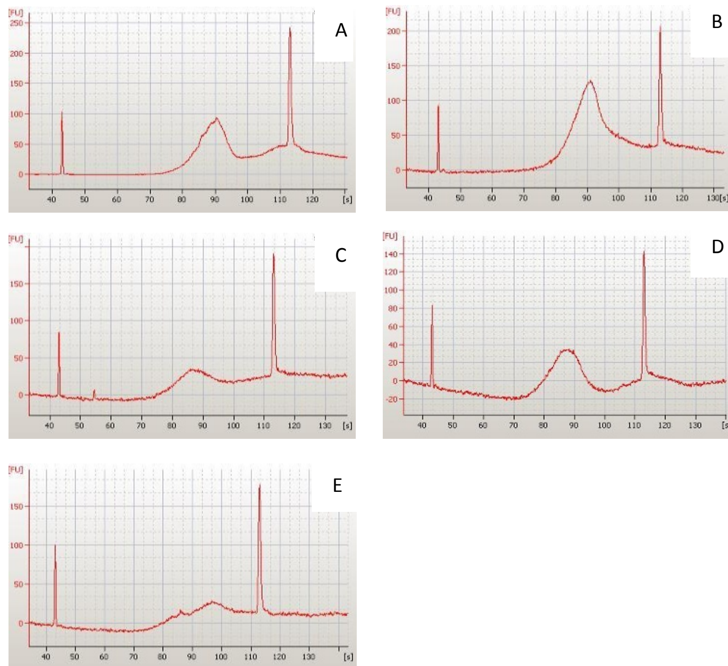



**FIGURE C 7: BIOANALYZER PROFILE OF TSS2 SAMPLES**

A - TSS2.1 extraction A, B - TSS2.2 extraction B, C - TSS2.3 extraction C, D - TSS2.4 extraction C, E - TSS2.0 extraction C

As all quality tests were valid, the libraries were mixed together with TSS1 libraries to form four pools containing 6 libraries before VirCapSeq capture step. Post capture qualification using 2100 BioAnalyzer



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

(Agilent, France) showed that the mean library length was 661 bp for pool I, 655 bp for pool II, 587 bp for pool III and 613 bp for pool IV. The bioanalyzer profiles are presented in Figure C 8. The pool concentration, measured by Qubit (Invitrogen, France), was 36 ng/µL, 37.4 ng/µL, 25.8 and 31.8 ng/µL for the pool I, II, III and IV, respectively.

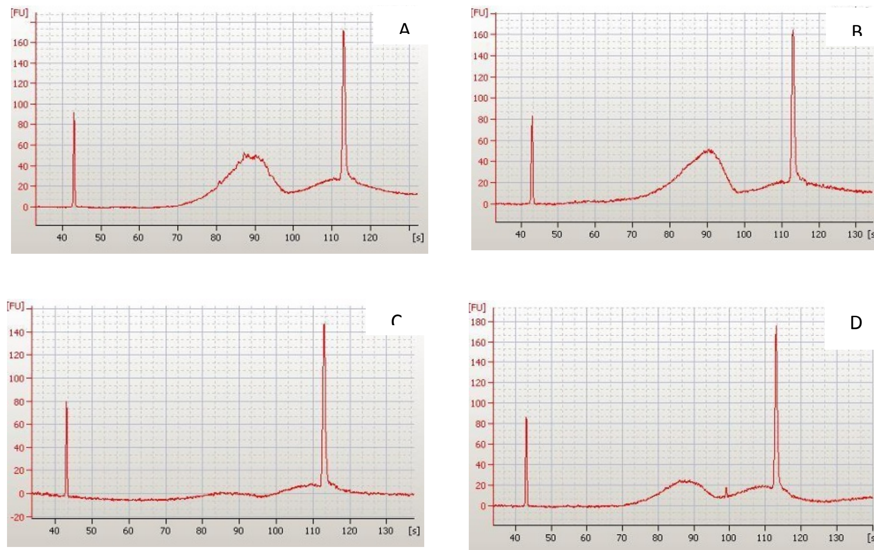



**FIGURE C 8: BIOANALYZER PROFILE OF LIBRARY POOL II AND III**

A - library pool I; B - Library pool II; C - Library pool III and D - library pool IV



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

# Appendix D - Library preparation for long amplicon sequencing using Oxford Nanopore Technology on TSS samples

## D.1 cDNA PCR and long amplicon PCR

Reactions to generate cDNA were performed for all reactions. PCR performed using primers from



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 9 and



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 10 were controlled by electrophoresis in 1% agarose gel (Figure D 1). For GI RdRp-VP1 target sequences we used primer combinations MON432-G1SKR (



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 9) and the primers A, B, D and E (



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 10 Noronet typing primers), PCR and N-PCR bands are shown in Figure D 1 C and A, respectively. For GII RdRp-VP1 target sequences we used primer combinations MON431-G2SKR (



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 9) and the primers F, G, H and I (



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 10 Noronet typing primers), PCR and N-PCR bands are shown in Figure D 1 D and B, respectively.

The presence of a specific band of correct size was confirmed for 22 samples (17 NoV GI and 5 NoV GII). The negative controls (TSS2.0, TSS2.5 and water) remained negative (Figure D 1).

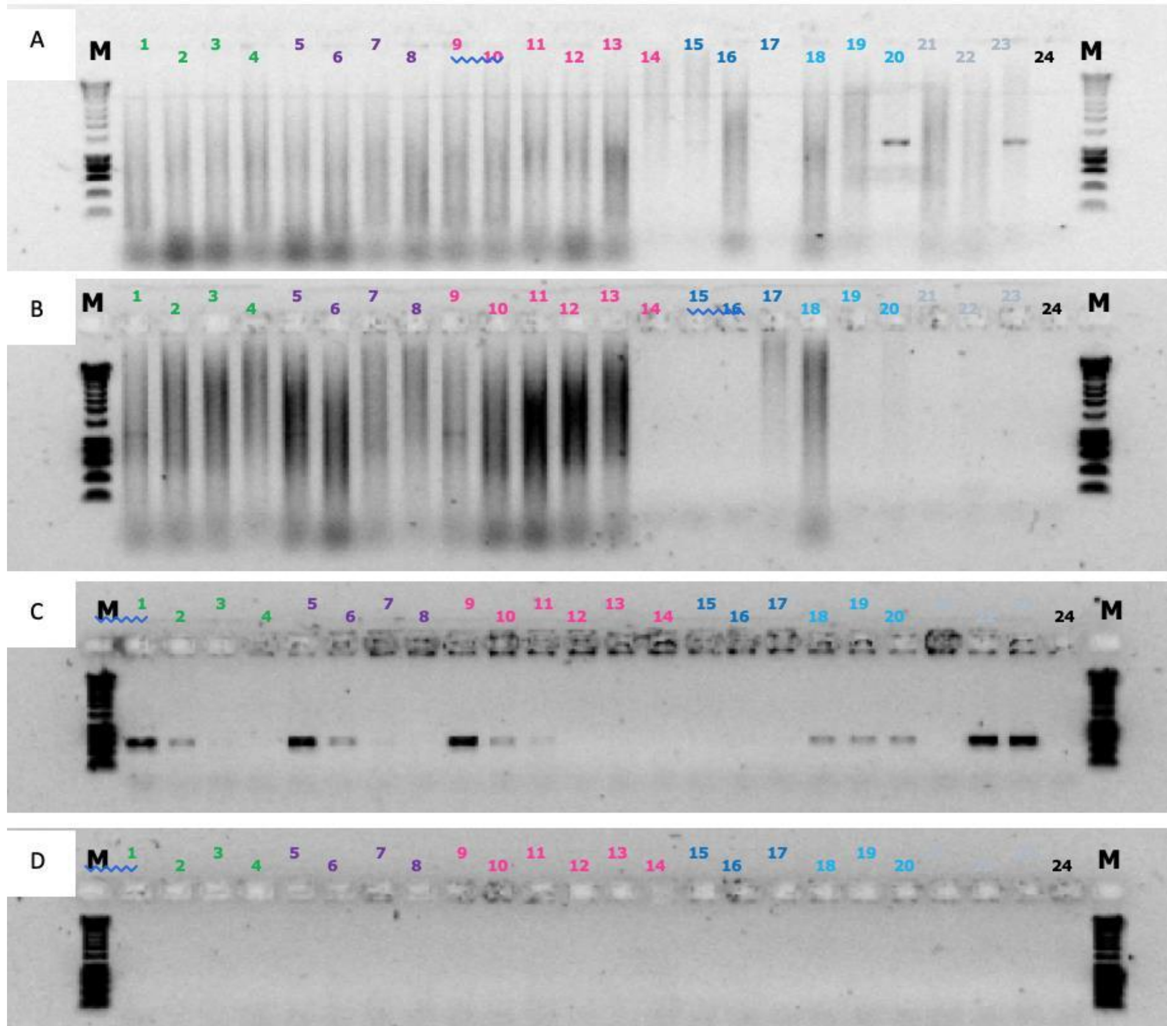



**FIGURE D 1: ELECTROPHORESIS OF TSS1 AND TSS2 LIBRARIES USING LONG AMPLICON PCR FOR OXFORD NANOPORE TECHNOLOGY SEQUENCING IN 1% AGAROSE GELS.**

A – GI Noronet typing primers, B – GII Noronet typing primers, C – GI MON432-G1SKR, D – GII MON431-G2SKR



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

## Appendix E - Preparation of BLS samples for metabarcoding

The reverse transcription efficiency and cDNA synthesis was confirmed by qPCR (data not shown). The libraries obtained following N-PCR were controlled by electrophoresis in 1% agarose gel. The sizes of amplicons with tags sequences, used for this study were 349 bp (VP1), 342 bp (RdRp), 514 bp (RdRp-VP1) for the GI and 364 bp (VP1), 212 bp (RdRp) and 557 bp (RdRp-VP1) for the GII. No amplicon was detected in negative controls (water). One example of electrophoresis of the libraries are showed on Figure E 1.

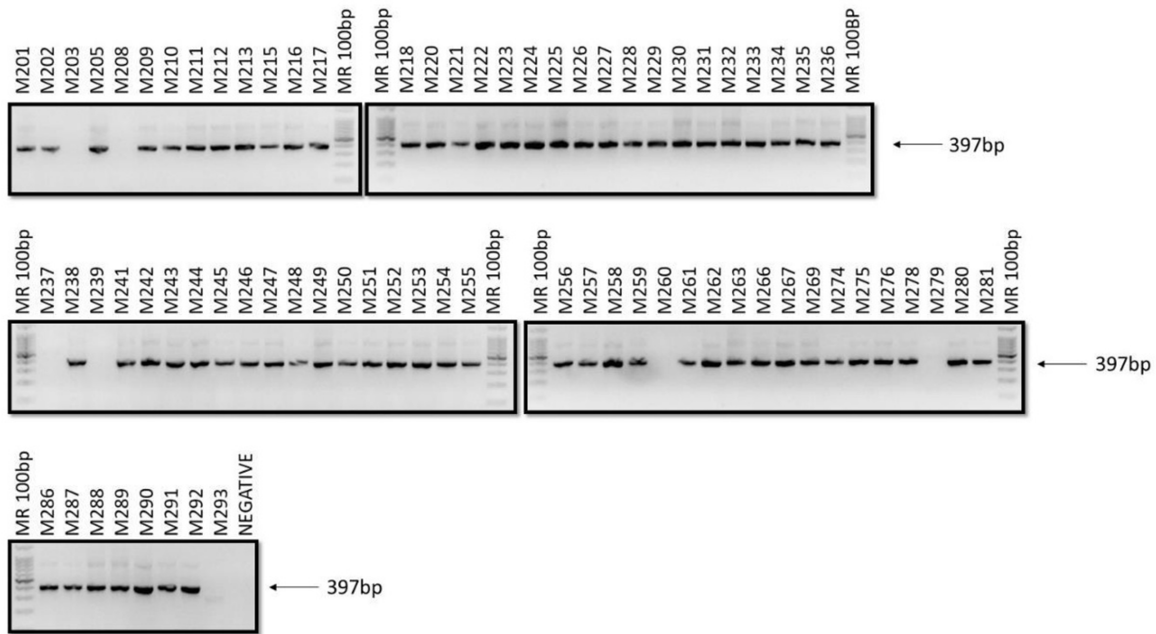



**FIGURE E 1: ELECTROPHORESIS OF BLS GI NOV VP1 LIBRARIES IN 1% AGAROSE GELS AT CEFAS**



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

## Appendix F – BLS samples and OB shellfish samples library preparation and validation for VirCapSeq metagenomics

cDNAs obtained by reverse transcription with random hexamers, were submitted to library preparation.

The DNA concentration in the libraries, quantified using Qubit (Invitrogen, France) ranged between 12 and 87 ng/µL for the BLS samples and between 7 and 78 ng/µL for the OB shellfish samples. Their mean size and profile were controlled using 2100 Bioanalyzer (Agilent, France) (data not shown).

As all quality tests are valid, the libraries are mixed together to form the pools, each pool containing 9 or 10 libraries before VirCapSeq capture step. The pools were done together with BLS and OB shellfish libraries. Nine pools were obtained in total.

Post capture qualification using 2100 BioAnalyzer performed for each pool show that the mean library length was 350 bp. The bioanalyzer profiles are presented on Figure F 1. The DNA concentration, measured by Qubit (Invitrogen, France), ranged from 41,8 ng/µL to 67,6 ng/µL.

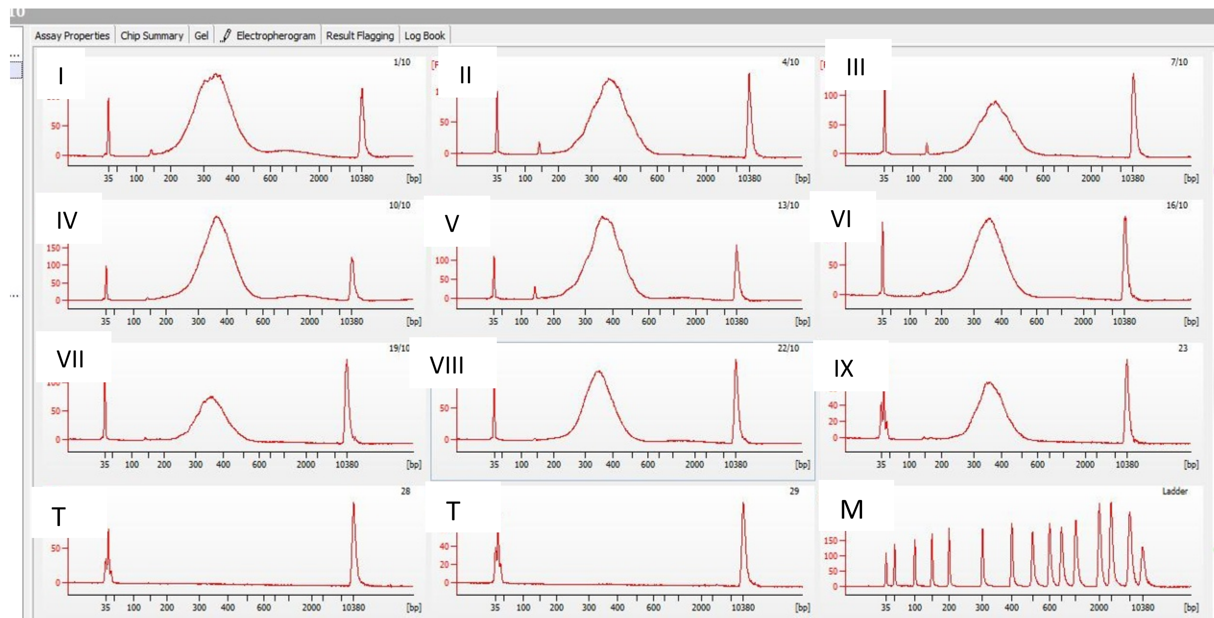



**FIGURE F 1: BIOANALYZER PROFILE OF LIBRARY POOLS**

I to IX – library pools, T – negative control, M – size marker



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.