

MOLECULAR EPIDEMIOLOGICAL STUDY ON  
INFECTIOUS PANCREATIC NECROSIS VIRUS  
ISOLATES FROM AQUAFARMS IN SCOTLAND  
OVER THREE DECADES

**Kristina Ulrich**

---

**UNIVERSITY of  
STIRLING**



---

Thesis submitted for the degree of Doctor of Philosophy  
Faculty of Natural Sciences  
Institute of Aquaculture

February 2018

---

*to Olga and Vladimir Ulrich*

## DECLARATION

---

I, Kristina Ulrich, submit this thesis for the degree of Doctor of Philosophy. I certify that the thesis has been written entirely by myself and embodies the results of my own research. I have acknowledged the nature and extent of work carried out in collaboration with others included in the thesis.

## ACKNOWLEDGEMENTS

---

First and special thanks to my parents, my foster-parents, my family and friends (new and old, near and far; including friends within the *Canis lupus familiaris* species) for the biggest support one can only get. The patience, love, constant encouragement and support in any possible way they gave me is above and beyond what I was expecting. The way they handled me and my emotional status during the PhD with all its ups and down is beyond my understanding. I am thankful for everything, especially for making this all possible.

Thanks to my supervisors Prof. Manfred Weidmann and Dr. Michaël Bekaert who supported and always believed in me throughout the PhD using the phrase “Please don’t panic!” quite a few times. They made it possible for me to go on conferences, gave me opportunities to build networks and present my data. I learned so much, so I feel like I can now call myself a real scientist.

Thanks to the University of Stirling, MASTS and IMPACT for the financial support.

Thanks to Dr. Meik Dilcher, Dr. Emma Thomson, Dr. Stefanie Wehner, Nicholas Di Paola, Dr. Christos Palaiokostas, Dr. John Taggart, Katherine Fiona Muir and the Virology lab and Dr. Kerry Bartie for their professional support.

Thanks to Dr. Benjamin Lopez Jimena (University of Stirling, Institute of Aquaculture, UK), Dr. Iveta Matejusova (Marine Scotland – Marine Laboratory Aberdeen, UK) and Dr. Harun Albayrak (Ondokuz Mayıs University, Department of Virology, Turkey) for providing samples as well to Dr. Tharangani Herath for sharing her IPNV RT-qPCR protocol with me.

### Introduction

RNA viruses are economically important pathogens of fish, and among these viruses, *infectious pancreatic necrosis virus* (IPNV) is of particular concern for the aquaculture industry, especially for farmed rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*). This non-enveloped aquatic virus, which was first isolated in the UK in 1971, belongs to the family of *Birnaviridae* and has a bi-segmented dsRNA genome of about 6kb. IPNV is classified in 6 genogroups with correspondence to 10 known serotypes and an additional proposed genogroup of marine aquabirnaviruses (MABV).

IPNV causes high mortality in fry and a reduced mortality in adult fish, respectively. Fish, which survive, can become carriers and this can lead to a clinical outbreak by releasing infective material into water or by vertical transmission via oocytes, milt and seminal fluids.

### Methods

This project aimed at determining the phylogeny and genomic changes of IPNV in Scotland by whole genome sequence analysis of IPNV isolates (diagnostic TCID<sub>50</sub> supernatants) spanning 3 decades since 1982, using next generation sequencing technology. Viral RNA of IPNV culture supernatant (CHSE-214 and TO cell culture) was processed for next generation sequencing on an Illumina MiSeq platform. Library preparation was performed using the Nextera XT DNA Library Kit, prior to sequencing according to the manufacturer's MiSeq Reagent Kit v3 (150cycles) protocol. To optimize whole genome next generation sequencing for IPNV, we compared two RNA processing protocols, the Glasgow (GLAP) and the Goettingen protocol (GOEP) with focus on missing terminal nucleotides after a *de novo* genome assembly. Sequences were used to determine the phylogeny and selection pressure on the genome as well as a possible virus-host adaptation.

### Results

The results showed that both protocols were able to give full length genomes as well as genomes with missing terminal nucleotides.

The phylogenetic analysis of 57 sequenced IPNV isolates shows that 78.95 % of the isolates group within genogroup V, which includes serogroup Sp and 5.26 % within genogroup I which

includes serogroup Ja. Segment A of 15.79 % of the isolate grouped within genogroup III, which includes serotype Ca1 and Te but only 7.02 % of the segment B isolates grouped in the genogroup III. The remaining 8.77 % of segment B groups within genogroup II, containing the Ab serotype.

Previous research has shown that residue substitutions at positions 217 and 221 in the major capsid protein VP2 have an impact on the virulence of the virus, leading to different virulence types: virulent (T217, A221), low virulence (P217, A221), avirulent (T217, T221) and persistent (P217, T221). Whole genome sequence results show that 58.93 % of the sequenced isolates belong to the persistent, 32.14 % to the low virulent type, only one isolate was of a virulent type and 7.15 % had not virulence assigned amino acid compositions in positions 217 and 221. The selection pressure analysis showed that especially VP2 is experiencing selection pressure in the variable region. In the VP1 protein we see two sites under positive selection pressure within specific motifs. VP5 showed positive selected sites mostly within the truncated region of the protein. Other proteins showed no particular interesting sites of selection.

The codon adaptation analysis showed highest adaptation index for VP2. Besides VP5, which had an CAI index below one, therefore showing negative adaptation, other IPNV proteins had an CAI of barely above the value of 1.

The dinucleotide abundance, focussing on CpG, showed that CpG is underrepresented in segment A and B.

## **Discussion**

Phylogenetic analysis of the sequenced IPNV strains shows separate clustering of different genogroups. Genetic reassortment is observed in segment B showing a grouping within genogroup III and II although the segment A of these isolates was grouping exclusively within III.

We found that over 50 % of the isolates belong to the persistent and over 30 % to the low virulent type, assuming that due to not sterilising vaccination these types were selected in the vaccinated population.

The results from the CAI calculations indicate an adaptation of IPNV to its host. Together with the findings that CpG is underrepresented in IPNV it suggests that this leads to an immune escape. Especially since the selection pressure analysis showed positive selection in VP2

within the virulence determination sites of the protein, indicating that IPNV “tries” to downregulate immune recognition.

The prevalence of mostly persistent type of isolates indicates together with the assumption of adaptation and immune escape that IPNV is evolving with the host in order to ensure survival.

# Table of Contents

<b>TABLE OF CONTENTS</b>	<b>8</b>
<b>I LIST OF FIGURES</b>	<b>10</b>
<b>II LIST OF TABLES</b>	<b>12</b>
<b>1 INTRODUCTION</b>	<b>15</b>
<b>1.1 INFECTIOUS PANCREATIC NECROSIS VIRUS</b>	<b>15</b>
1.1.1 THE DISEASE	15
1.1.2 VIRUS CHARACTERISTICS	15
1.1.3 PATHOGENICITY	26
1.1.4 IPNV DIAGNOSTICS	26
1.1.5 EPIDEMIOLOGY	27
1.1.6 IPNV VACCINE	28
<b>1.2 NEXT GENERATION SEQUENCING (NGS)</b>	<b>30</b>
1.2.1 OVERVIEW	30
1.2.2 IPNV RNA PROCESSING	30
1.2.3 MISEQ	31
1.2.4 MINION SEQUENCING TECHNOLOGY	34
<b>1.3 SEQUENCE ANALYSIS</b>	<b>38</b>
1.3.1 OVERVIEW	38
1.3.2 CALCULATING PHYLOGENIES USING BEAST	38
1.3.3 CALCULATING PHYLOGENIES USING RAXML	39
1.3.4 CAI	39
1.3.5 SELECTION ANALYSIS USING HYPHY	40
1.3.6 DINUCLEOTIDE COMPOSITION	40
<b>OBJECTIVES</b>	<b>42</b>
<b>2 METHODS</b>	<b>43</b>
<b>2.1 WHOLE GENOME SEQUENCING OF DIAGNOSTIC TCID 50 ISOLATES</b>	<b>43</b>
2.1.1 BULKING UP DIAGNOSTIC TCID50 IPNV ISOLATES IN CELL CULTURE	43
2.1.2 IPNV RNA PROCESSING AND SEQUENCING	48
<b>2.2 IMPROVING THE GLASGOW PROTOCOL</b>	<b>64</b>
2.2.1 IMPROVING THE DS-CDNA SYNTHESIS	64
2.2.2 IMPROVING RNA PROCEDURE	67
<b>2.3 PERSISTENCY EXPERIMENT</b>	<b>69</b>
2.3.1 INTRODUCTION	69
2.3.2 METHODS	69
<b>2.4 MINION SEQUENCING</b>	<b>72</b>
2.4.1 INTRODUCTION	72
2.4.2 METHODOLOGY	72
2.4.3 IMPROVING SIZE SELECTION FOR MINION SEQUENCING	78
<b>2.5 BIOINFORMATICS ANALYSIS</b>	<b>79</b>
2.5.1 PROCESSING THE RAW DATA FROM MISEQ (ILLUMINA)	79
2.5.2 PROCESSING RAW DATA FROM MINION (OXFORD NANOPORE TECHNOLOGIES)	80
<b>2.6 PHYLOGENETIC ANALYSIS</b>	<b>82</b>
2.6.1 GENERAL HANDLING OF SEQUENCES	82
2.6.2 MEGA	82
2.6.3 BEAST PACKAGE	82



2.6.4	RAXML	85
2.6.5	CAI CALCULATION	85
2.6.6	HYPHY PACKAGE	85
2.6.7	DINUCLEOTIDE COMPOSITION	86
2.6.8	SPLITSTREE	86
2.6.9	STATISTICS AND GRAPHS	87
<b>3</b>	<b>RESULTS</b>	<b>88</b>
<b>3.1</b>	<b>WGS PROTOCOL DEVELOPMENT AND IMPROVEMENT</b>	<b>88</b>
3.1.1	IMPROVEMENTS TO THE DOUBLE STRAND CDNA SYNTHESIS STEP OF THE GLASGOW PROTOCOL	92
3.1.2	IMPROVEMENTS TO RNA PROCEDURES OF THE GLASGOW PROTOCOL	94
<b>3.2</b>	<b>SEQUENCING RESULTS</b>	<b>97</b>
3.2.1	ILLUMINA MISEQ SEQUENCING	97
3.2.2	MINION	102
<b>3.3</b>	<b>SEQUENCING ANALYSIS</b>	<b>106</b>
3.3.1	PHYLOGENETIC ANALYSIS	106
3.3.2	ANALYSIS OF AMINO ACID SEQUENCE FEATURES OF IPNV PROTEINS	109
3.3.3	CAI ANALYSIS	118
3.3.4	SELECTION ANALYSIS	123
3.3.5	DINUCLEOTIDE COMPOSITION	126
<b>4</b>	<b>DISCUSSION</b>	<b>128</b>
<b>4.1</b>	<b>WHOLE GENOME SEQUENCING PROTOCOLS</b>	<b>128</b>
4.1.1	COMPARING THE GOETTINGEN AND THE GLASGOW PROTOCOL	128
4.1.2	MINION SEQUENCING	131
<b>4.2</b>	<b>PHYLOGENETIC ANALYSIS INTERPRETATION</b>	<b>133</b>
<b>4.3</b>	<b>CODON ADAPTATION OF IPNV TO SALMONIDS</b>	<b>139</b>
<b>4.4</b>	<b>SELECTION ANALYSIS INTERPRETATION</b>	<b>144</b>
<b>4.5</b>	<b>CONCLUSION</b>	<b>146</b>
<b>4.6</b>	<b>LIMITATIONS AND OUTLOOK</b>	<b>147</b>
<b>III</b>	<b>APPENDIX</b>	<b>148</b>
<b>III.1</b>	<b>FIGURES AND TABLES</b>	<b>148</b>
III.1.1	FIGURES	148
III.1.2	TABLES	157
III.1.3	DATA PROCESSING – ADDITIONAL APPENDIX	169
<b>III.2</b>	<b>ABBREVIATIONS AND ACRONYMS</b>	<b>170</b>
<b>III.3</b>	<b>MATERIAL</b>	<b>172</b>
III.3.1	PRIMER, PROBES, ADAPTORS	172
III.3.2	CHEMICALS	174
III.3.3	KITS	176
III.3.4	CONSUMABLES	178
III.3.5	DEVICES	179
<b>IV</b>	<b>REFERENCES</b>	<b>180</b>

**V APPENDIX – SCRIPTS**

**PROVIDED SEPARATELY**

## I List of Figures

Figure 1. IPNV in FHM cells; short arrow indicates virus-specific tubules; long arrow shows crystalline arrays of virions. Scale represents 200nm (Crane and Hyatt 2011). .....	16
Figure 2. Segment A and B of IPNV; VP1 is bound to each 5'-end of the bi-segmented dsRNA genome, which encodes the proteins VP2, VP3, VP4 and VP5 in segment A and VP1 in segment B (Dobos 1995a). .....	17
Figure 3. IPNV and IBDV VP2; (A) Ribbon diagram of IBDV-VP2 with nuanced domains B, S and P. (B) Ribbon diagram of IPNV-VP2 with nuanced domains B, S and P. (C) Alignment of IPNV (blue) and IBDV (red) VP2 subunits. Blue and red stars highlight residues decisive for virulence of IPNV and IBDV, respectively. The line indicates the molecular 3-fold axis. Boxes highlight regions structural divergence and stars show the position of the amino acid residues associated with virulence in IPNV (blue; 217 and 221) and in IBDV (red; 253, 279, 284) (Coulibaly et al. 2010, Coulibaly et al. 2005). .....	20
Figure 4. Model for self-interaction and VP1 and dsRNA binding sites of the IPNV-VP3 protein (Pedersen, Skjesol, and Jørgensen 2007). .....	21
Figure 5. Segment A encoded polyprotein with VP4 cleavage sites (Munro and Midtlyng 2011). .....	21
Figure 6. IPNV replication model proposed by Villanueva et al., showing different steps in the infective cycle, considering intermediates possibly such as defective interfering particles (Villanueva et al. 2004). .....	25
Figure 7. Template with different termini sites required for Illumina sequencing (Illumina, adapted). .....	31
Figure 8. Template amplification technology of Illumina NGS devices (Goodwin, McPherson, and McCombie 2016). .....	32
Figure 9. Overview of the MinION sequencing technology (Oxford Nanopore Technologies, adapted). .....	35
Figure 10. Template for MinION sequencing containing a leader and a hairpin structure (Oxford Nanopore Technologies, adapted). .....	36
Figure 11 Overview of the main steps of GOEP and the GLAP. ....	48
Figure 12. Coverage (number of reads/reference bases) plots of one of the three randomly selected samples processed with the GOEP and the GLAP. Plots for the other samples see Figure 28 in the Appendix. Plots were generated using BOWTIE and SAMTOOLS (2.5.1.2). .....	91
Figure 13. Coverage (number of reads/reference bases) plots of protocol variants with added GP32. Plots for the other variants see Figure 29 in the Appendix section. Plots were generated using BOWTIE and SAMTOOLS (2.5.1.2). .....	93
Figure 14. Coverage (number of reads/reference bases) plots of two randomly selected samples processed with different incubation conditions for RNA prior ds-cDNA synthesis. Plots were generated using BOWTIE and SAMTOOLS (2.5.1.2). .....	95
Figure 15. Forward and reverse reads recovered from IPNV samples processed with the GLAP or the GOEP. ....	100
Figure 16. Mean coverage (number of reads/reference bases) of segment A and B for 57 processed IPNV samples with the GLAP or GOEP. ....	101
Figure 17. Missing terminal nucleotides of both segments of IPNV samples processed with the GLAP or the GOEP after de novo assembly. ....	102
Figure 18. Comparing the coverage (number of reads/reference bases) plots of one randomly selected sample (upper: segment A, lower: segment B) sequenced on the MinION (blue,	

left coordinates) and on the MiSeq (green, right coordinates). Plots were generated using BOWTIE and SAMTOOLS (2.5.1.2). ..... 104

Figure 19. Bioanalyzer results of one randomly selected betanodavirus sample used for improvement of fragment length for MinION sequencing; upper shows the fragment range with 1x beads to sample ratio, lower shows the fragment range with 0.6x beads to sample ratio. .... 105

Figure 20. Phylogenetic classification of 57 IPNV isolates based on segment A sequences using reference sequences (Table 32). The unrooted tree was calculated using BEAST with 10,000,000 chains per cycle with every 1000<sup>th</sup> chain being recorded. Posterior probability values are shown at corresponding nodes in percent. The scale represents 0.02 substitutions per nucleotide site. .... 107

Figure 21. Phylogenetic classification of 57 IPNV isolates based on segment B sequences using reference sequences (Table 33). The unrooted tree was calculated using BEAST with 10,000,000 chains per cycle with every 1000<sup>th</sup> chain being recorded. Posterior probability values are shown at corresponding nodes in percent. The scale represents 0.06 substitutions per nucleotide site. The serotype clusters separately but is still defined as genogroup III (see Table 2). .... 108

Figure 22. Non-homologous amino acid alignment based Neighbour Network of the VP2 protein for genogroup V IPNV samples. The scale bar represents the split support for the edges. Sample 53 (red) is the only virulent type (TA). Green samples represent persistent type (PT), purple low virulent (PA) and blue not assigned types (AT and PS). .... 112

Figure 23. Non-Homologous amino acid alignment based Neighbour Network of the VP3 protein for genogroup V IPNV samples. The scale bar represents the split support for the edges. .... 115

Figure 24. Selected region of the amino acid alignment of the VP5 protein for 57 sequenced IPNV samples showing deletions and years when they occurred. Stars indicate truncated proteins. Deletions highlighted in red. Dots represent homologous amino acids. Amino acid variations are highlighted by respective amino acid code letters. .... 118

Figure 25. CAI analysis of each IPNV protein to show signs of codon usage adaptation towards the host genome of Atlantic salmon (*Salmo salar*) and rainbow trout (*Onchorhynchus mykiss*). Values above 1.00 show a sign of codon usage adaptation. Stars indicate a significant relation. .... 120

Figure 26. CAI analysis of each IPNV to show signs of codon usage adaptation towards the host genome of Atlantic Atlantic salmon (*Salmo salar*) and rainbow trout (*Onchorhynchus mykiss*) between low virulent and persistent type of samples. Values above 1.00 show a sign of codon usage adaptation. Stars indicate a significant relation; significance level  $p \leq 0.05$ . .... 121

Figure 28. Distribution of variants of 56 sequenced IPNV strains in each decade. .... 136

Figure 29. Coverage plots of the other two randomly selected samples processed with the Goettingen and the Glasgow protocol. .... 148

Figure 30. Coverage plots of the other protocol variants. .... 149

Figure 31. Coverage plots of 57 processed IPNV samples. The coverage plots were created with R, using the NanoOK tool as a guidance with changes in the R script for this purpose. .... 156

## II List of Tables

Table 1. IPNV-protein overview. ....	18
Table 2. Classification of IPNV based on seroneutralization and genotyping proposed by Hill and Way, 1995 (serotype) and Blake et al., 2001 (genogroup). Hosts were listed according to the publications (Hill and Way 1995, Blake et al. 2001). ....	23
Table 3. Overview of methods for IPNV diagnosis (Munro and Midthlyng 2011).....	27
Table 4. Comparing Illumina MiSeq and Oxford Nanopore Technologies MinION sequencing systems (Goodwin, McPherson, and McCombie 2016). ....	36
Table 5. Ligation reaction-set up.....	52
Table 6. Synthesis reaction, part I/ II- set up.....	52
Table 7. Synthesis reaction, part II/ II- set up.....	53
Table 8. Synthesis reaction- cycling conditions.....	53
Table 9. Amplification with EvaGreen- set up.....	53
Table 10. Amplification with EvaGreen- cycling conditions in the LC480II. ....	54
Table 11. Amplification without EvaGreen- set up. ....	54
Table 12. Amplification without EvaGreen- cycling conditions in a thermocycler.....	54
Table 13. cDNA synthesis- master mix. ....	56
Table 14. cDNA synthesis- cycling conditions.....	56
Table 15. cDNA second strand synthesis reaction- master mix. ....	57
Table 16. cDNA second strand synthesis- cycling conditions. ....	57
Table 17. IPNV samples used to compare the Goettingen and the Glasgow protocol. ....	58
Table 18. IPNV RT-qPCR- set up. ....	59
Table 19. IPNV RT-qPCR- cycling conditions.....	59
Table 20. Correlation between Agarose gel %-age and DNA size distribution (Ref.: <a href="http://www.promega.co.uk/resources/pubhub/enotes/what-percentage-agarose-is-needed-to-sufficiently-resolve-my-dna-sample/?activeTab=0">http://www.promega.co.uk/resources/pubhub/enotes/what-percentage-agarose-is-needed-to-sufficiently-resolve-my-dna-sample/?activeTab=0</a> ).....	60
Table 21. PCR amplification cycling conditions. ....	62
Table 22. Changes in the GLAP compared to standard GLAP.....	64
Table 23. Primer for IPNV segment A and B.....	65
Table 24. End-prep set-up . All reagents are included in the NEBNext Ultra II End-Prep kit.....	72
Table 25. Adaptor ligation set-up. All reagents are included in the NEB Blunt/TA kit or the SQK-NSK007 (v.R9) kit. ....	73
Table 26. PCR mix set-up of ligated sample. All reagents are included in the LongAmp Taq 2x kit or the SQK-NSK007 (v.R9) kit.....	73
Table 27. PCR cycling conditions. ....	74
Table 28. End-prep set-up after the PCR. All reagents are included in the NEBNext End-Repair/dA-tailing kit. ....	75
Table 29. Ligation set-up of the hairpin adaptor. All reagents are included in the NEB Blunt/TA Ligase kit or the SQK-NSK007 (v.R9) kit.....	75
Table 30. Preparation of the priming mix. All reagents are included in the SQK-NSK007 (v.R9) kit.....	77
Table 31. Preparation of the library mix to load on the flow cell. All reagents are included in the SQK-NSK007 (v.R9) kit.....	77
Table 32. Reference strains used for phylogenetic analysis of segment A; NA means not assigned.....	82
Table 33. Reference strains used for phylogenetic analysis of segment B; NA means not assigned.....	84

Table 34. Control values of selected libraries in Section I: randomly selected samples processed with the Goettingen (GOEP) and the Glasgow (GLAP) sequencing protocol; Section II: Improvement variants for GLAP; Section III: temperature variants tested for GLAP; Section IV: range of control values of 57 sequenced IPNV samples. Conc.=concentration. ....	90
Table 35. Mean coverage (number of reads/reference bases) of segment A and B for the samples processed with the GLAP and the GOEP. ....	91
Table 36. Comparison of missing nucleotides and the overall mapping rate in samples processed with the GOEP and the GLAP after de novo assembly. ....	91
Table 37. Mean coverage (number of reads/reference bases) of segment A and B for the sample processed with different changes in the GLAP. ....	92
Table 38. Comparison of missing nucleotides and the overall mapping rate of one randomly selected sample and its improvement variants after de novo assembly. ....	93
Table 39. Mean coverage (number of reads/reference bases) of segment A and B for two samples processed with RNA incubation changes in the GLAP. ....	95
Table 40. Comparison of missing nucleotides and the overall mapping of two randomly selected samples processed with different incubation conditions for RNA prior ds- cDNA synthesis after de novo assembly. ....	96
Table 41. Sequenced IPNV samples and their sampling information and the protocol used; NA=not assigned. ....	97
Table 42. Control values of one randomly selected sample processed for MinION 2D sequencing. ....	103
Table 43. Bioanalyzer results of the library of one randomly selected sample processed for MinION 2D and MiSeq sequencing. ....	103
Table 44. Mean coverage (number of reads/reference bases) of segment A and B for one samples sequenced on the MiSeq and the MinION. ....	104
Table 45. Missing nucleotides and the overall mapping rate of one randomly selected sample sequenced on the MinION and the MiSeq in comparison after de novo assembly. ....	104
Table 46. Bioanalyzer results of two randomly selected samples used for improvement of fragment length for MinION sequencing. ....	105
Table 47. Genogroup distribution of the sequenced IPNV isolates. ....	109
Table 48. VP2 residue characteristics pattern of sequenced IPNV isolates compared to the virulent reference strain excluding IoA032. Amino acid positions identified in previous studies as important in terms of viral virulence are listed, with a focus on the positions 217 and 221 (underlined). ....	110
Table 49. Distribution of variants of 56 sequenced IPNV strains; NA=not assigned. ....	111
Table 50. Amino acid composition of five randomly selected samples grown in two different cell lines to prove persistency characteristics. All Isolates $\geq 2$ passages. ....	113
Table 51. Amino acid features of specific IPNV RdRp (VP1) sites and motifs. S (Serine), A (Alanine), D (Aspartic acid), N (Asparagine), I (Isoleucine), V (Valine), H (Histidine), Y (tyrosine), G (Glycine), E (Glutamic acid), K (Lysine), L (Leucine), T (Threonine). ....	117
Table 52. Amino acid features of catalytic sites of the IPNV protease (VP4). S (Serine), V (Valine), K (Lysine), F (Phenylalanine). ....	117
Table 53. Significance levels of CAI threshold over time of three decades by comparing two decades with each other for both hosts <i>Salmo salar</i> and <i>Onchorhynchus mykiss</i> ; significance level $p \leq 0.05$ . ....	122
Table 54. Significant differences in CAI values between hosts against a sample variant persistent (P) type and low virulent (LV) type; significance level $p \leq 0.05$ . ....	123

Table 55. Significant differences in CAI values low virulent (LV) and persistent (P) type of samples against host species; significance level $p \leq 0.05$ . .....	123
Table 56. Results of a positive selection analysis of all IPNV proteins for all 57 sequenced samples using SLAC on HYPHY. ....	124
Table 57 Results of site-specific positive and diversifying selection analysis of all IPNV proteins for all 57 sequenced samples using MEME on HYPHY, applying the MG94xREV codon model. P describes the p-value ( $p \leq 0.1$ ) for positive selection at one site.....	125
Table 58. Dinucleotide composition of IPNV segment A and B calculated with EMBOSS compseq with focus on CG (bold italic). Results CpGO/E $\leq 1$ : underrepresented dinucleotide; CpGO/E $\geq 1$ : overrepresented dinucleotide. ....	126
Table 59 Distribution of variants of 56 sequenced IPNV strains with indicated flips leading to virulent strains.....	135
Table 60. Control values of sequenced IPNV samples; NA=not assigned. ....	157
Table 61. Number of forward and reverse decontaminated reads of the randomly selected IPNV samples processed with the Goettingen or the Glasgow protocol for whole genome sequencing using NGS on the MiSeq.....	159
Table 62. Mean coverage (number of reads/reference bases) of segment A and B for 57 processed IPNV samples. ....	161
Table 63. Missing nucleotides and the overall mapping rate of processed IPNV samples after de novo assembly.....	163
Table 64. Sequenced IPNV isolates and their genogroups according to segment A and segment B phylogeny; NA=not assigned.....	165
Table 65. Detailed VP2 residue characteristics of sequenced IPNV isolates compared to the virulent reference strain excluding IoA032; NA=not assigned. Amino acid positions identified in previous studies as important in terms of viral virulence are listed, with a focus on the positions 217 and 221 (underlined).....	167

# 1 Introduction

## 1.1 Infectious pancreatic necrosis virus

### 1.1.1 The disease

Infectious pancreatic necrosis virus (IPNV) is a wide-spread pathogen of aquatic animals (Munro and Midtlyng 2011) that has been isolated worldwide. The regions with the highest IPNV reports are the ones with intensive aquaculture. The virus and the disease patterns in Atlantic salmon (*Salmo salar*) as the host organism are described in the following. As this work is about the molecular evolution of IPNV I will mainly focus on these aspects, while only briefly describing disease patterns, diagnostic and preventive measurements.

IPNV causes a high mortality of up to 100 % in salmonid fry in freshwater, and a reduced mortality in adult fish. Fish, which survive, can shed the virus for up to 6 years as IPNV can persist in B-cells, neutrophils and head kidney derived salmon leucocytes (Rønneseth, Pettersen, and Wergeland 2012) (Ahne and Thomsen 1986). Horizontal transmission by IPNV virions shed into the water by persistently infected fish or vertical IPNV transmission via oocytes, milt and seminal fluids can lead to new infections in naïve fish.

Infectious pancreatic necrosis (IPN) mostly occurs in young animals and can be symptomatic or asymptomatic. The symptom can differ and not every infected fish develops all described symptoms. Infected young Atlantic salmon lose appetite, are lethargic and show an abnormal swimming pattern. Furthermore, they are darker in appearance, present internal, as well as external lesions, abdominal swelling, pale gills, haemorrhages and necrosis of pancreatic cells (Wood, Snieszko, and Yasutake 1955) (Wolf 1988) (Smail et al. 1992) (McKnight and Roberts 1976, Roberts and McKnight 1976).

The severity of IPNV depends on factors such as age, feeding, stress and temperature. Smolts are most susceptible 4-12 weeks after transfer into sea water (Wolf, Dunbar, and Pyle 1961) (Wolf, Fish, and Research 1966) (Smail et al. 1992) (Roberts and Pearson 2005) (Skjelstad 2003) and IPN develops after an incubation period of 10-14 days, depending on the temperature (Ramstad et al. 2007) (Okamoto et al. 1993, Okamoto et al. 1987) (Wetten et al. 2007).

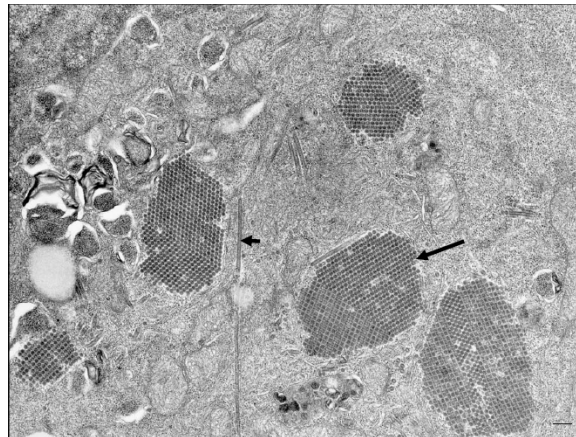
### 1.1.2 Virus characteristics

IPNV was first isolated in the US during an outbreak in rainbow trout (*Onchorhynchus mykiss*) fingerlings in 1955 (Wood, Snieszko, and Yasutake 1955). This isolate was given the reference

number VR299. It appears that the first description of the disease was already published earlier by M'Gonigle in Canada in the 1940s. In Europe IPNV was first detected in 1964 in southern France (Besse and de Kinkelin 1965). From 1968 on IPNV was isolated also in other European countries (Jorgensen and Bregnballe 1969). The first isolation in the UK was in 1971 from rainbow trout on a farm in Scotland (Ball et al. 1971).

#### 1.1.2.1 IPNV virion

IPNV, as seen in Figure 1, is a member of the genus *Aquabirnavirus* in the family *Birnaviridae*, has a non-enveloped, single-shelled capsid of a T=13 icosahedral symmetry composed of 260 trimeric subunits of VP2 peptides (Lightner and Post 1969) (Hulo et al. 2010). It is 60 nm in diameter (Cohen, Poinard, and Scherrer 1973) and has a molecular weight of  $55 \times 10^6$  Da (Dobos 1995a, b).



**Figure 1. IPNV in FHM cells; short arrow indicates virus-specific tubules; long arrow shows crystalline arrays of virions. Scale represents 200nm (Crane and Hyatt 2011).**

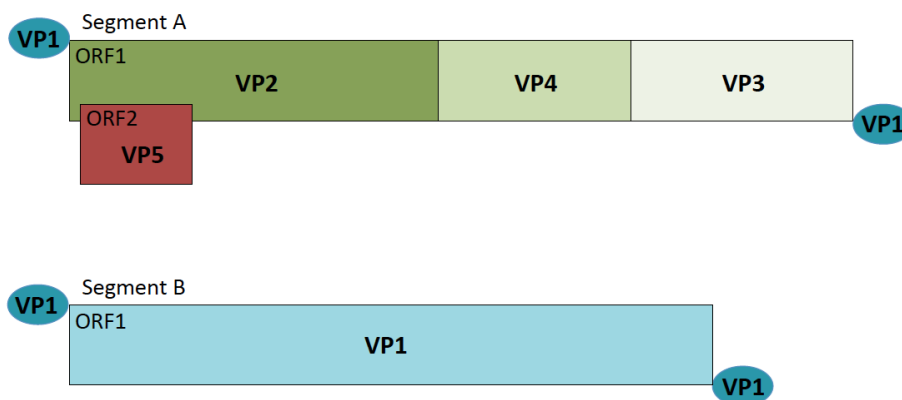
The stability and infectivity of the virus depends on water temperature, salinity and the presence of microbial flora (Mortensen, Nilsen, and Hjeltnes 1998) (Ahne 1982) (ToReNzo 1983). Studies have shown that IPNV is stable at salinity ranges from 0 to 40 % and at temperatures from -80 °C to +40 °C, with a reduction of infectivity above 20 °C. Freezing and thawing was shown to decrease the virus titre (Mortensen, Nilsen, and Hjeltnes 1998) (Barja et al. 1983), with differences between the IPNV strains (McMichael, Fryer, and Pilcher 1975). Taken together, greatest IPNV stability is achieved in filter-sterilised or autoclaved estuarine water at 15 °C (Barja et al. 1983) (ToReNzo 1983).



### 1.1.2.2 *IPNV genome*

The IPNV genome is a bi-segmented, double-stranded, RNase resistant RNA genome of 5,881kb (Jasper strain) (Duncan et al. 1987) which is composed of a 3097 bp long segment A and a 2784 bp long segment B (NC\_001915.1; NC\_001916.1). Each segment is linked to a Serine residue at the 5'-end of the genome-linked protein (VPg/VP1) and has no 3'-poly-A tail (Dobos 1995a, b).

Segment A contains two overlapping open reading frames (ORFs) that are read through leaky scanning, where the 40 S ribosome subunit skips the first ORF and continues scanning until it gets to the second overlapping ORF. The larger first ORF encodes a 106 kDa polyprotein composed of 5'-pVP2-VP4-VP3-3'. The second (small) overlapping ORF encodes the Arginine-rich non-structural 15 kDa VP5 protein (Dobos 1995a, b). Segment B is monocistronic and contains only one ORF that encodes the RNA-dependent RNA polymerase (RdRp/VP1), in form of a 94 kDa VP1/VPg protein (Duncan and Dobos 1986, Duncan et al. 1987) (Dobos 1995a, b).



**Figure 2. Segment A and B of IPNV; VP1 is bound to each 5'-end of the bi-segmented dsRNA genome, which encodes the proteins VP2, VP3, VP4 and VP5 in segment A and VP1 in segment B (Dobos 1995a).**

### 1.1.2.3 *IPNV proteins*

IPNV encodes five proteins, four located on segment A and one located on segment B. The proteins VP2, VP3 and VP4 derive from one polyprotein as shown in Figure 2 and Figure 5. This polyprotein generates first the 62 kDa protein pVP2, which is further cleaved during the virus maturation to the 54 kDa VP2 protein and the propeptides p1-3, that are associated with the virion, and the 31 kDa VP3. This is achieved via post-translational modification by the 29 kDa Serine-Lysine protease VP4 (Figure 5 and Table 1) (Huang et al. 1986) (Duncan et al. 1987)

(Nagy et al. 1987) (Manning and Leong 1990);(Petit et al. 2000) (Galloux, Chevalier, Henry, Huet, Da Costa, et al. 2004).

**Table 1. IPNV-protein overview.**

Protein	Segment	Size (kDa)	Function
VP1/ VPg	B	94	RdRp/ genome-linked VP1
Polyprotein	A	106	precursor of VP2, VP3 and VP4
pVP2	A	63	pre-VP2
VP2	A	54	major capsid protein
VP3	A	31	minor capsid protein
VP4	A	29	Serine-Lysine protease
VP5	A	15-17	anti-apoptosis/ non-structural protein

#### 1.1.2.3.1 VP1/ RdRp

VP1 is encoded on segment B and is a RNA-dependent RNA polymerase (RdRp, VP1) with a GTP-binding region (G motif) and 6 conserved motifs I – VI (I=F; II=C; III=A; IV=B; V=D; VI=E; GTP binding motif=G), which performs RNA transcription and genome replication. The polymerase sequence motifs are II – III – IV with ADN being the catalytic residue in motif II (Graham et al. 2011). It is larger compared to the polymerases of other RNA viruses. *In vitro* experiments showed that this polymerase acts as a primer and uses a semi-conservative (displacement) mechanism for RNA replication. Furthermore, it is active in the virion without any processing such as proteolysis. The optimal enzymatic activity was shown to be at 30 °C, pH of 8.0 and in the presence of 6 mM Mg<sup>2+</sup>-ions, although the pH and the Mg<sup>2+</sup> concentrations were not critical for the polymerase activity (Mertens, Jamieson, and Dobos 1982). Further studies showed, that the IPNV polymerase also has a terminal (deoxy)nucleotide transferase, RNA-dependent DNA polymerase (reverse transcriptase) and template-independent self-guanylation activity (Graham et al. 2011).

It was also shown that the IPNV RdRp exists in two forms within the virion: as VPg, which binds to the 5'-end of each genome segment by a N-terminal Serine-5'-guanosinemonophosphate phosphodiester bond (Calvert et al. 1991) which occurs only in the *Birnaviridae* dsRNA family, and as a free polypeptide VP1 that functions as a RdRp. Comparison of RdRps of other dsRNA viruses showed that the RdRp of IPNV does not have the

Glycine-Aspartic acid-Aspartic acid (GDD) motif, which is typical for this enzyme family but a structurally similar Alanine-Aspartic acid-Asparagine (ADN) motif (Dobos 1995a, b, Graham et al. 2011, Gorbalenya et al. 2002).

Additionally, VP1 is the only protein in the virion that, as VP1pG, auto-catalyses in a template-independent manner the addition of a GMP to a Serine residue (within the N-terminal 27 residues) so it can act as a primer for the replication during the viral ssRNA(+) synthesis (Dobos 1993) (Xu, Si, and Dobos 2004) (Graham et al. 2011). The catalytic binding site is conserved within the *Birnaviridae*, which makes it a general feature of the polymerases of this virus family. Protein structure studies revealed that the N-terminus (Serine 2) of VP1 interacts with its active site and is ligated to the nascent RNA daughter strand during RNA replication. This finding provides a model for the mechanism of viral genome-polymerase association that is observed *in vivo*. It is suggested that the N-terminus of the polymerase domain plays a role in recruiting dNTPs, since it was observed close to the (d)NTP entry tunnel (Graham et al. 2011). Since we know that VP1 has a reverse transcriptase activity it can utilise dNTPs beside NTPs and form RNA/DNA hybrids.

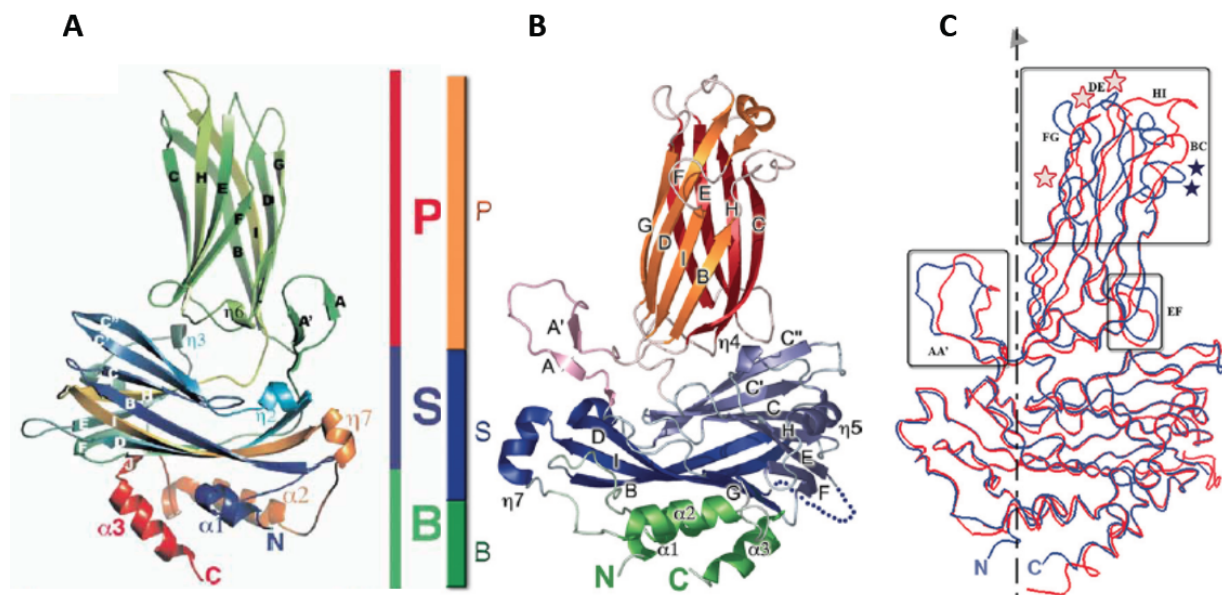
Genetic comparison experiments of the IPNV Jasper/genogroup I and Sp/genogroup V RdRp revealed that there is an 88 % homology between each other, but only 41 % with the Birnavirus Infectious Bursal Disease Virus (IBDV) RdRp. Nevertheless, homologous regions exist within the central part of the proteins (Duncan et al. 1991).

#### 1.1.2.3.2 VP2

When Santi et. al. (2004) compared the sequences of different IPNV Sp/ genogroup V isolates, they found that the major capsid protein VP2 is more variable in the coding region than VP1, VP3 and VP4 sequences. VP2 contains a variable domain in the central region between the residues 183-335, which is an important antigenic site (Blake et al. 2001) and two hypervariable region between the residues 239-257 and 271-284. Further experiments revealed that a Threonine (Thr) instead of Proline (Pro) at position 217 and Alanine (Ala) at position 221 are characteristic for virulent IPNV Sp isolates (Santi, Vakharia, and Evensen 2004b). Interestingly, both residues are located within the hypervariable region. The observation of the Ala/Thr change at position 221 was always made after 2 passages in CHSE-214 cell culture, which probably shows an attenuation of the virus in cell culture (Santi, Vakharia, and Evensen 2004b), it was however not shown in RTG-2 cell line even after several

passages (Song et al. 2005). However, it is assumed that Alanine at position 221 reduces binding affinity to cell receptors, resulting in more effective virus release and therefore higher replication (Mutoloki et al. 2016).

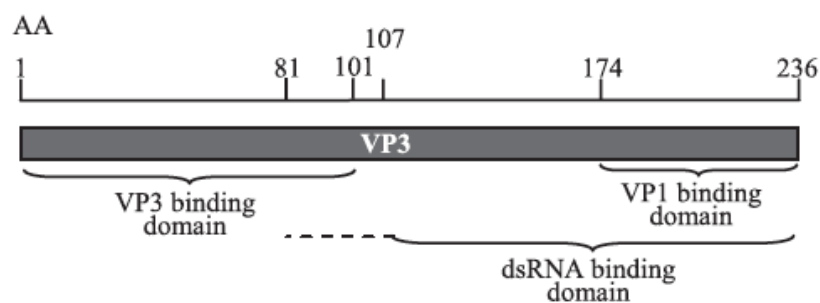
As previously mentioned, the capsid of the virion is made of 260 VP2 subunits. Each subunit is folded into three domains, base (B), shell (S) and projection (P) (see Figure 3). Domain P harbours loops, which are the most exposed part of the VP2 trimer and contain the antigenic sites and virulence factors.



**Figure 3.** IPNV and IBDV VP2; (A) Ribbon diagram of IBDV-VP2 with nuanced domains B, S and P. (B) Ribbon diagram of IPNV-VP2 with nuanced domains B, S and P. (C) Alignment of IPNV (blue) and IBDV (red) VP2 subunits. Blue and red stars highlight residues decisive for virulence of IPNV and IBDV, respectively. The line indicates the molecular 3-fold axis. Boxes highlight regions structural divergence and stars show the position of the amino acid residues associated with virulence in IPNV (blue; 217 and 221) and in IBDV (red; 253, 279, 284) (Coulibaly et al. 2010, Coulibaly et al. 2005).

#### 1.1.2.3.3 VP3

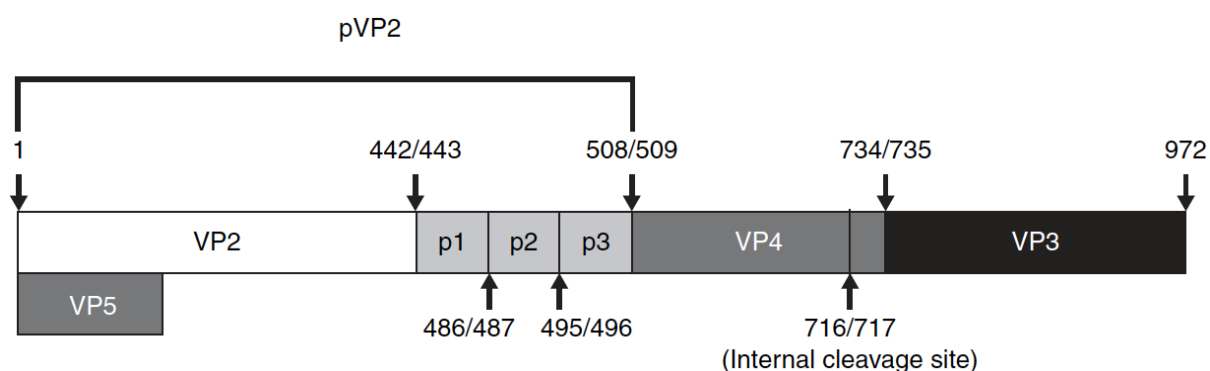
The multifunctional VP3 protein, which is found in the inner layer of the virus capsid, is a self-binding protein. It also interacts with VP1 and the dsRNA genome segments and seem to activate the RdRp (Pedersen, Skjesol, and Jørgensen 2007) (see Figure 4). The VP1-VP3 interaction does not require the presence of dsRNA. Interestingly, the binding between VP1 and VP3 occurs before the mature virion formation. This interaction could play a role in the assembly process of IPNV (Pedersen, Skjesol, and Jørgensen 2007). Furthermore, VP3 inhibits IFN $\alpha$  induction in salmon cells by targeting the RIG-I/MDA5 pathway to block immune response during infection (Lauksund et al. 2015).



**Figure 4. Model for self-interaction and VP1 and dsRNA binding sites of the IPNV-VP3 protein (Pedersen, Skjesol, and Jørgensen 2007).**

#### 1.1.2.3.4 VP4

The VP4 protein is a Serine633-Lysine674 protease that is essential for the polypeptide processing. It cleaves the (Ser/Thr)-X-Ala ↓ (Ser/Ala)-Gly motif of the polyprotein at its own N-terminal site at amino acid position 508-509 of the pVP2-VP4 junction and at its C-terminal site at position 734-735 of the VP4-VP3 junction, as shown in Figure 5 (Petit et al. 2000). As VP3, VP4 also has an immune response blocking activity by inhibiting the IFN $\alpha$ 1 promotor in salmon cells (Lauksund et al. 2015).



**Figure 5. Segment A encoded polyprotein with VP4 cleavage sites (Munro and Midtlyng 2011).**

#### 1.1.2.3.5 VP5

VP5 is a non-structural Arginine-rich anti-apoptosis protein and is expressed early in the replication cycle. It shares homologous domains with Bcl-2 and shuts off the host's apoptosis system by down regulation of Mcl-1, which induces cells to undergo apoptotic cell death. This way IPNV enhances cell survival for the benefit of viral proliferation (Hong, Gong, and Wu 2002). Besides that, together with VP4, VP5 inhibits the IFN expression facilitated by the Mx promoter and has an antagonistic effect (Skjesol et al. 2009). The small ORF, which encodes

the protein, shows a high level of variation within the genome of different IPNV Sp serotype/genogroup V isolates with an additional in-frame stop codon at nucleotide position 472, that creates a 12 kDa version of the VP5 protein and at nucleotide position 496 that results in an additional alternative (truncated) form of VP5. It was shown that IPNV with a truncated VP5 is virulent but this feature could not be confirmed in other challenge experiments (Santi, Sandtrø, et al. 2005, Santi, Song, et al. 2005, Song et al. 2006, Song et al. 2005), where fish challenged with IPNV virions containing a truncated VP5 showed only moderate mortality. The fact that some isolates did not even encode the VP5 protein but still induced mortality, supports the findings that absence of VP5-expression does not influence the viral growth and virulence (Heppell et al. 1995) (Weber et al. 2001, Santi, Song, et al. 2005).

#### 1.1.2.4 IPNV classification

Wolf and Quimby found in 1971 (Wolf and Quimby 1971) that antibodies against the IPNV isolate VR299 only partly deactivated European isolates. Through further experiments Hill and Way suggested a classification of aquatic Birnaviruses in serotypes A and B and divided serogroup A in A1-A9, due to results in reciprocal cross-neutralisation assays (Hill and Way 1995). IPNV belonging to serotype A are found worldwide, whereas serotype B, which was first described in the mussel *Tellina tenuis* by Underwood et al. in 1977 (Underwood et al. 1977), seems to be endemic primarily in the North Sea close to Denmark.

Although studies, which were done with polyclonal antibodies, confirm the 9 different A serotypes, experiments with monoclonal antibodies (MAb) in an ELISA assay, targeting the major neutralisation epitopes within VP2 (Frost et al. 1995) showed that they could not distinguish between them. Until now, there are no suitable MAb-assays available to determine the serotype of an IPNV isolate (Christie, Ness, and Djupvik 1990, Frost et al. 1995, Melby and Christie 1994).

**Table 2. Classification of IPNV based on seroneutralization and genotyping proposed by Hill and Way, 1995 (serotype) and Blake et al., 2001 (genogroup). Hosts were listed according to the publications (Hill and Way 1995, Blake et al. 2001).**

Serotype Nomenclature	Serotype Name	Genogroup Nomenclature	Host	Origin Country
A1	West Buxton (WB)	I	Trout	USA
A1	Buhl	I	Trout	USA
A1	Reno	I	Trout	USA
A1	VR299	I	Trout	USA
A1	Jasper (Ja-Dobos)	I	Trout	Canada
A1	Dry Mills (DM)	I	Trout	USA
A2	Sparajub (Sp)	V	Trout	Denmark
A2	N1	V	Atlantic salmon ( <i>Salmo salar</i> )	Norway
A3	Abild (Ab)	II	Trout	Denmark
A4	Hecht (He)	VI	Pike	Germany
A5	Tellina (Te)	III	Tellina	UK
A6	Canada1 (C1)	III	Trout	Canada
A7	Canada2 (C2)	IV	Trout	Canada
A8	Canada3 (C3)	IV	Arctic char ( <i>Salvelinus alpinus</i> )	Canada
A9	Jasper (Ja)	I	Trout	Canada
B1	Tellina virus (TV-1)	III	Tellina	UK
MABV (Marine Aquabirnavirus) (Zhang and Suzuki 2004, Nishizawa, Kinoshita, and Yoshimizu 2005)	Yellowtail virus (Y-6, YT-01A)	VII	Yellowtail	Japan

On the other hand, sequencing the genome of IPNV can be used to genogroup IPNV. Based on the large ORF in segment A and the VP2 region, IPNV is now classified in 6 genogroups with correspondence to the known 10 serotypes as shown in Table 2 with an additional genogroup for Marine *Aquabirnavirus* as proposed by Zhang and Suzuki in 2004 (Zhang and Suzuki 2004) (Hill and Way 1995) (Blake et al. 2001) (Cutrin et al. 2004).

Nevertheless, since Romero-Brey reported a natural reassortment between two different strains in 2009, due to coinfection of two different strains in one fish, where segment A from one strain and segment B from another strain rearranged together in a new virion. Therefore, both segment A and B should be sequenced for better genomic characterisation of each IPNV isolate (Romero-Brey et al. 2009). Further research showed that reassortment in IPNV is not random and that some segments replicate more efficiently than others in a cell-dependent manner and that it can as well influence the virulence (Lago et al. 2017).

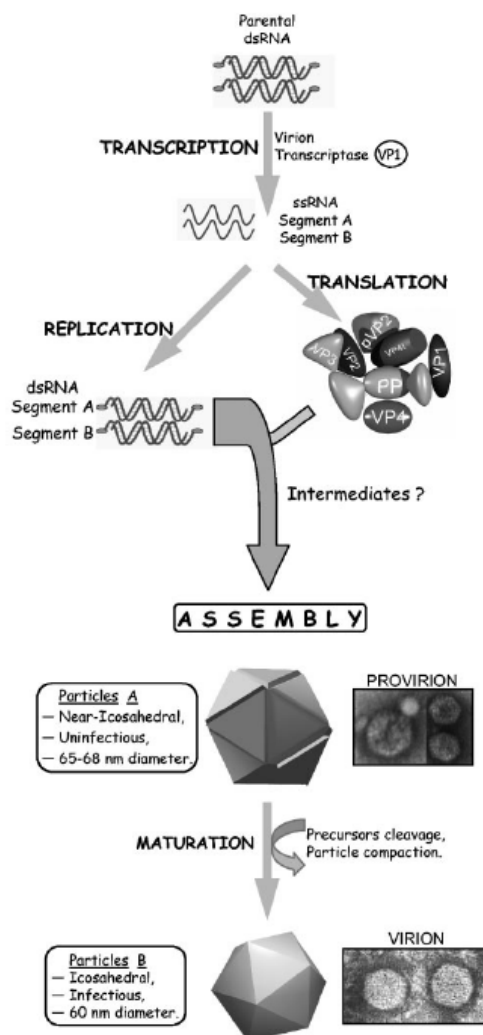
#### 1.1.2.5 IPNV replication model

IPNV replicates in the cytoplasm of a host cell. Since eukaryotic cells do not produce dsRNA, they recognise the viral dsRNA as a pathogen associated molecular pattern (PAMP) and inactivate it, which leads to development of immune escape mechanism due to mutation pressure initiated by the immune system of the host. This will be discussed in detail further in the thesis, especially in the discussion.

One replication cycle, as first shown in details in CHSE-214 cells, takes approximately 16-20 h depending on the temperature, although the RdRp is active even at temperatures above 30 °C (Mertens, Jamieson, and Dobos 1982). In some cells, researchers observed a persistent infection with a very low viral replication without any cytopathic effect (CPE) probably due to the overproduction of defective interfering particles (MacDonald and Kennedy 1979) (Hedrick and Fryer 1981). IPNV enters the host through gills, skin and gut penetrating the cell by binding VP2 to polypeptides on the cell surface (Dobos 1995a, b) (Kuznar et al. 1995) to be adsorbed into the cell through endocytosis (Couve, Kiss, and Kuznar 1992). For CHSE-214 specifically it was shown that IPNV enters via macropinocytosis (Levicán et al. 2017). It is assumed that the virion is not being uncoated during the entry process (Cohen 1975, Patton and Spencer 2000) and that during the replication cycle, the genome of dsRNA viruses remains within the capsid, to protect the viral genome from Rig-I like receptors (RLRs) of the innate immune response, in the presence of the active RdRp inside the virion. At 2-4 h post



infection a first transcription intermediate can be detected as well as the viral mRNA and viral polypeptides after 4-6 hours. The maximum amount of viral mRNA, which does not have a 5' poly A tail, is found after 8-10 hours and decreases after 14 h (Somogyi and Dobos 1980) (Dobos 1995a, b). Viral proteins are produced in approximately the same proportions between 3 and 14 h after infection (Dobos et al. 1977, Dobos 1977). The viral particles are formed from non-infectious premature provirions that contain pVP2. The maturation process, which involves the cleavage of pVP2 to VP2 by the VP4 protease, creates then mature infectious virions (Villanueva et al. 2004).



**Figure 6. IPNV replication model proposed by Villanueva et al., showing different steps in the infective cycle, considering intermediates possibly such as defective interfering particles (Villanueva et al. 2004).**

Further *in vitro* experiments showed, that transcription, which is primed via VP1 and is proceed as an asymmetric, semi-conservative strand-displacement mechanism, results in

synthesis of only (+) strands of the genome segments that are attached to the template via base pairing (Dobos 1995a, b). Produced viral mRNA functions as template for translation to produce viral proteins and replication of the viral genome. Villanueva and colleagues proposed a model of IPNV replication based upon the research done so far, which shows the different steps of the replication (Villanueva et al. 2004).

### 1.1.3 Pathogenicity

Cell culture is a diagnostic method, that provides proof of the presence of infectious intact IPNV at the time of sampling. Usually, the pathogen achieves high titres in infected tissues and there is no known stage of the disease at which the virus is not detectable in cell culture.

An infection with IPNV results in a clear visible cytopathic effect (CPE) in cells. Many cell types are susceptible to IPNV infection. Several cell lines are used for IPNV detection and isolation. For best results the CHSE-214 cell line should be used (Lorenzen, Carstensen, and Olesen 1999). It was further shown that IPNV does attenuate (undergoes mutations which lead to a weaker virus type) in CHSE-214 cells but not in RTG-2 cell culture (Song et al. 2005) even after several passages (see 1.1.2.3.2). Another difference between cell lines is the cellular immune response to IPNV. RTG-2 cells produce interferon, whereas CHSE-214 do not, so IPNV in CHSE-214 results in higher virus titres. Therefore, CHSE-214 is mostly used for IPNV growth (MacDonald and Kennedy 1979, Dobos 1995a). IPNV grows in cell culture at 10-25 °C, dependent on the cell line used. CHSE-214 cells are cultured in the dark at 21°C and 5 % CO<sub>2</sub>. Another cell line used for IPNV infection is the TO cell line derived from head kidney leucocytes of the Atlantic salmon cultured at 20 °C with 5 % CO<sub>2</sub> (Wergeland and Jakobsen 2001).

### 1.1.4 IPNV diagnostics

Standard diagnosis of IPN disease is carried out by isolating IPNV on cell culture followed by antibody-based detection. However, this is time (up to three or four weeks) and laboratory-space consuming and needs specially trained technicians with cell culture experience. Over the years immunological and molecular techniques have been developed that allow a faster, more sensitive and even quantitative detection of IPNV (Table 3).

The diagnostic criteria for an aquatic pathogen are defined by the World Organization for Animal Health and American Fisheries Society (OIE 2000). Sampling for IPNV diagnostics is

different for each method and the size of the fish but most methods require the sampling of the liver, kidney and spleen.

**Table 3. Overview of methods for IPNV diagnosis (Munro and Midtlyng 2011).**

<b>Technique</b>	<b>Concept</b>
Cell culture	Viral isolation
Neutralisation Assay	
FAT	
ELISA	
Immunohistochemistry	Antibody binding assays
Immunodot blot	
Flow cytometry	
Coagglutination test	
RT-PCR	
real-time PCR	Amplification and detection
RT-LAMP	of viral nucleic acid
In situ hybridisation	

### 1.1.5 Epidemiology

First disease outbreaks caused by IPNV on Atlantic salmon farms in Europe were reported in Norway in the 1980s (Christie et al. 1988) and later, in the 1990s, in Scotland (Smail et al. 1992). It is assumed that IPNV infects a wide range of up to 100 hosts, including fish, water birds and crustaceans (Munro and Midtlyng 2011) and has been detected worldwide. It is the most important disease in its impact on farmed salmon production in the European Union and in Norway (Ariel and Olesen 2002) (Murray, Busby, and Bruno 2003). There are no publications available that provide statistics about losses in aquaculture due to IPN, although the Fish Health Report for Norway reports a reduction on IPNV positive tested salmon farming sites from 223 in 2009 to 27 in 2016 (Hjeltnes et al. 2017). A study done by the Marine Laboratory in Aberdeen, UK, estimates that Scottish farms lost up to 2 M GBP (Murray, Busby, and Bruno 2003) and the loss in Ireland was up to € 31 M after an IPNV outbreak in 2006 according to the Marine Institute in Galway, Ireland (Geoghegan, Ó Cinneide, and Ruane 2007). Furthermore, all waters in Shetland are thought to be IPNV-contaminated. As the

disease can stay for up to 3 months in post-transfer fish and can be shed via oocytes, milt and seminal fluid, this led to the expansion of IPNV down the coast of Scotland by 2001 and reaching Ireland in 2002 possibly through trading and transmission of the virus into the wild fish population. To prevent huge losses farms can use specific post-transfer diets and transfer systems which are less stressful for the stocked fish. There are also IPN resistant fish strains and vaccines available (Roberts and Pearson 2005).

Today we know that certain IPNV strains are less virulent than others. In 2004 Santi et al., 2004, described that certain amino acid positions in the VP2 gene determine the pathogenicity of different IPNV strains (Santi, Vakharia, and Evensen 2004b).

Today we differ between virulent (T217 A221), persistent (P217 T221), avirulent (T217 T221) and low virulent (P217 A221) types with less than 10 % mortality. The virulent strains have a high morbidity and mortality (Santi, Vakharia, and Evensen 2004b) (Song et al. 2005). Persistent strains induce a high morbidity and in general no mortality but may induce some mortality later in the disease (Mutoloki, Munang'andu, and Evensen 2013) (Julin, Mennen, and Sommer 2013). This pattern was not shown in a big study of Scottish IPNV isolates where some of the PT isolates induced high mortalities (Bain, Gregory, and Raynard 2008) as mentioned previously. Isolates with the residues T217 T221 have only a mortality of 15 %, compared to T217 A221 with >60 % mortality (Song et al. 2005).

#### 1.1.6 IPNV vaccine

Vaccines are an important measure to prevent disease, especially when there is no treatment available, as it is mostly the case for viral diseases. There are different types of vaccines, such as non-replicating (inactivated whole virus or subunit vaccines), replicating and DNA vaccines. Non-replicating vaccines contain killed pathogens or parts of pathogens that are extracted or synthesized, for example, a specific site of a protein. They are widely used as they cannot cause the disease after vaccination. However, they require higher amount of the antigen of interest, they have to be administered with adjuvants to increase the immune response, which is a B-cell response that results in antibody production against the administered antigen (humoral immunity) and they result in a shorter protection compared to replicating vaccines. Replicating vaccines on the other hand simulate a natural infection and therefore are capable of inducing a humoral and a cellular immune response. The administered virus replicates in host cells and is therefore presented on the cell surface as well. This results in a much longer

protection than given by non-replicating vaccines. DNA vaccines contain a genetic sequence of the antigen of interest that is embedded in a genetic delivery system like a plasmid and is expressed in the host cell to stimulate humoral and cellular immune response (Flint et al. 2015, Gudding, Lillehaug, and Evensen 2014).

First vaccine trials were launched in the 1970s with inactivated or attenuated IPNV and failed to protect against the virus (Dorson 1988). Then Manning and Leong showed that the VP2 coding region has an immunogenic effect able to induce immunity (Manning and Leong 1990). After that several recombinant vaccines were launched, such as in 1995 in Norway. IPN vaccines available today include either the main immunogenic protein VP2 and other components of the virus (Rimstad 2014) or are inactivated vaccines (Gomez-Casado, Estepa, and Coll 2011). However, there is an issue as fish are most susceptible to IPNV before they reach their immunocompetent age, so the administration of a vaccine would result in a full protection (Rodriguez Saint-Jean, Borrego, and Perez-Prieto 2003).

Quite often vaccine trial results as well as details on the exact antigen used were not published. The protection shown in the trials were said to be more than 80 % but this was not the case in the field and it didn't stop IPNV outbreaks. DNA vaccines were developed as well but they could not be administered on a bigger scale due to regulatory issues.

Vaccines are administered via injection, submersion bath or per os to salmonids in the seawater stage (post-smolts). The best results were achieved by vaccination via injection although that is difficult due to the size of the vulnerable fish and tissue damage caused.

IPN outbreaks in Norway continued up to 2010 and dropped from 2011 onward, when commercial IPN resistant salmon breed became available (Olsen and Hellberg 2012). It is possible to assume a similar scenario for the UK as IPN resistant salmon with identified quantitative trait loci (QTL) for IPN resistance (Houston, Bishop, et al. 2009, Houston et al. 2008) also were available from 2010s on (Gheyas et al. 2010). The IPNV-resistant QTL has been shown to block the entry of the T217, A221 variant (virulent type) in infected cells and in infected fish (Moen et al. 2015, Houston, Haley, et al. 2009).

IPN resistant fish seem to be more promising than vaccines as vaccines show a big variation in terms of protection depending on the fish's immune-genetic background and the IPNV strain involved (Ramstad and Midtlyng 2008). Initial reports from northern Norway indicate that in recent years IPNV has evolved to overcome this resistance (O. Evensen, personal communication).

## 1.2 Next generation sequencing (NGS)

### 1.2.1 Overview

Next generation sequencing describes high throughput sequencing technology. With its higher throughput, the ability to process multiple whole genomes from different organisms at the same time and lower costs than Sanger sequencing it provided a next level tool for basic research. Furthermore, it is used in clinical diagnostics, for organism identification and as the technology gets smaller the devices can be even used in remote locations (Wetterstrand 2013) (Gargis et al. 2012).

The first one of its kind was available in the 2000s and since then this type of technology was called NGS. By now there are different types of NGS technologies, differing in read length and sequencing biochemistry.

I will briefly describe the technologies I used in this study to sequence whole IPNV genomes: The Illumina MiSeq and the Oxford Nanopore Technologies MinION system, as well as the two RNA processing protocols, Goettingen (GOEP) and Glasgow (GLAP) protocol.

### 1.2.2 IPNV RNA processing

In this study, IPNV RNA was sequenced with two different protocols. The Goettingen (GOEP) protocol, that was initially developed for 454 Pyrosequencing of RNA viruses at Goettingen University, Germany (Dilcher et al. 2012). This protocol was transferred to whole genome sequencing (WGS) using next generation Illumina sequencing technology (NGS). During the study it was assessed in comparison to a simplified protocol, the Glasgow protocol (GLAP), from a team at MRC Virology Glasgow University, UK (Thomson et al. 2016) that developed this protocol in order to sequence HCV directly from whole blood samples.

Compared to GOEP, GLAP requires a quantitative RT-qPCR prior sample processing as GLAP has no template-amplifying step (PCR) prior to library preparation. The RT-qPCR is necessary as the Illumina MiSeq system has a detection limit of  $1 \times 10^4$  genome copies/mL ( $1 \times 10^1$  copies/ $\mu$ l) (Frey et al. 2014). When sequencing viral genomes, it is of advantage to use a minimum of amplification steps to avoid introduction of mutations by the polymerase. Using the GOEP includes performing two amplification steps: one during the ds-cDNA synthesis and another during the library preparation. Using GLAP results only in one amplification during the library preparation step itself. Besides that, the GLAP has a shorter DNase step of 5 minutes at

37 °C compared to the GOEP of 30 minutes at 37 °C. Assuming RNase activity through contamination, the short DNase step ensures a maximum preservation of viral RNA.

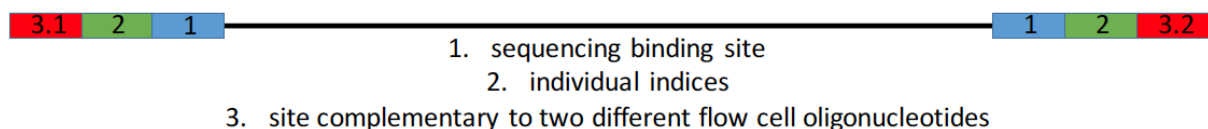
In summary, both protocols convert viral dsRNA into ds-cDNA in their specific manner. The ds-cDNA is further used for library preparation using the Illumina Nextera XT DNA Library Preparation Kit.

### 1.2.3 MiSeq

#### 1.2.3.1 MiSeq sequencing technology

The Illumina MiSeq system is a short-read Sequencing by Synthesis (SBS) system, which sequences by using cyclic reversible termination. In this case the nucleotides that terminate the sequencing reaction have labelled protected 3'-OH groups, terminating elongation and allowing for further fluorescence detection.

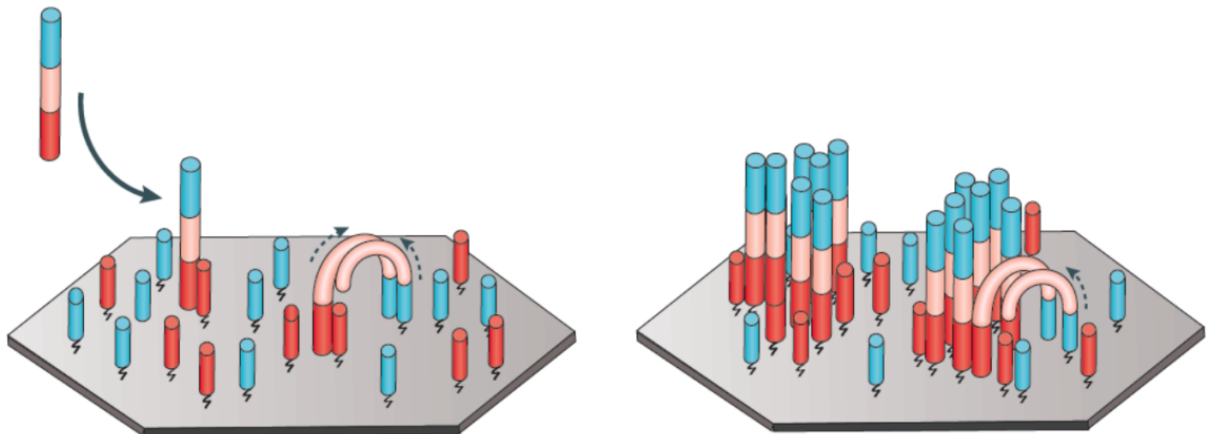
There is a variety of library preparation systems depending on the input sample. The Illumina Nextera XT DNA system uses transposons for fragmentation of DNA into pieces of approximately 300 bp and for adding adapters in the same step (Figure 6). This is followed by a PCR with a low number of cycles which adds indices and allows unique labelling of up to 384 samples. Ampure XP magnetic bead purification then selects for a certain amplicate size by adjusting the sample to beads ratio. This is important, as library size affects the choice of the sequencing biochemistry that is used. Normalisation of libraries is achieved by adjusting the molarity of each sample.



**Figure 7. Template with different termini sites required for Illumina sequencing (Illumina, adapted).**

Single stranded library fragments generated by denaturation are pooled and loaded into the flow cell where they hybridise to oligonucleotides immobilised on the flow cell surface complementary to the adapters of the fragments. A polymerase now initiates amplification of the first strand (forward strand) and generates many copies of the initial template (Figure 7) which is flushed away after the amplification. The amplified strands fold over to hybridise with a second complementary immobilised oligonucleotides on the flow cell surface to initiate

bridge amplification, generating clusters of fragments. After the bridge amplification the reverse strands are washed off.



**Figure 8. Template amplification technology of Illumina NGS devices (Goodwin, McPherson, and McCombie 2016).**

The sequencing of the forward strand begins, which is defined as read 1. During this step a mix of terminating 3'-end blocked and labelled dNTPs is added. When such a terminating dNTP binds, the reaction is flushed and an image is taken, followed by the removal of the blocking group and the fluorophore and a new cycle begins. As only one terminating dNTP is added per cycle, it assures that repeating nucleotides of a sequence (i.e. TTTT) are still accurately sequenced. The number of cycles determines the length of the read. Therefore, the sequence accuracy decreases with increasing read length, which makes this technology only accurate for short reads. The detection system is optical (fluorescence) so the base call is defined by the emission wavelength and the signal density. After the forward strand was sequenced, the index read is generated by sequencing the index. After that a polymerase uses the template to generate a reverse strand, which now can be sequenced. This is defined as read 2. This paired-end sequencing gives a high coverage per site, which makes the data more accurate.

After sequencing the Illumina software separates the samples through their individual indices. The raw reads with similar base calls are clustered together and can be used for genome assembly. As forward and reversed reads are paired, which is helpful when alignments are not clear, this system makes it possible to create accurate genome alignments based on a reference sequence and is perfectly capable of *de novo* assembly (Metzker 2010, Goodwin,



McPherson, and McCombie 2016, Liu et al. 2012, Glenn 2011, Ambardar et al. 2016, Rodríguez-Ezpeleta, Hackenberg, and Aransay 2011).

Data processing is divided into several steps, which involve several processing tools. As the MiSeq demultiplexes the data and clips adaptors there is no necessity for a tool, although it sometimes happens that due to mutations in the adaptors and indices MiSeq fails to clip properly. This is where the tools PRINSEQ (Schmieder and Edwards 2011b) and TRIMMOMATIC (Bolger, Lohse, and Usadel 2014) can be used. They identify not removed adaptors and perform additional quality trimming or eventually full removal of bad quality reads. Prior to these steps FASTQC (Andrews 2010) is used to check the quality of the reads and assess the quality of the sequencing output. DECONSEQ is used to eliminate contaminating reads not belonging to the genome of interest (Schmieder and Edwards 2011a). The reads are then ready for a *de novo* or a reference-based assembly. In this study SPADES (Nurk et al. 2013) was used for *de novo* and TANOTI (TANOTI: a rapid BLAST-guided read mapper for small, divergent genomes (*manuscript communicated*)) and STAMPY (Lunter and Goodson 2011) were used for reference-based assembly.

The Illumina MiSeq system creates very low non-random errors (0.1 %), which are hard to exclude even through high coverage at a site. The error rate of one read is very low, but as the error is non-random errors will persist at any one site even with very high coverages (Metzker 2010, Goodwin, McPherson, and McCombie 2016, Liu et al. 2012, Glenn 2011, Ambardar et al. 2016, Rodríguez-Ezpeleta, Hackenberg, and Aransay 2011).

Sequencing features of the Illumina MiSeq system are further listed in Table 4.

#### 1.2.3.2 SPADES for *de novo* genome assembly

SPADES is a *de novo* genome assembler, which uses different sizes of k-mers to analyse the reads. A k-mer is a nucleotide sequence with all possible sequences of a certain defined length. Reads are chopped into the k-mer size and assembled (Nurk et al. 2013). The k-mer length that yields the best assembly outcome is determined. Subsequently the evaluated k-mer length is then selected to puzzle the reads into a full genome. The k-mer analysis of SPADES is automated in terms of k-mers-modulation, while ABYSS, another commonly used *de novo* assembling software, requires manual setting of k-mers (Simpson et al. 2009).

After assembling a sequence, it is then fully aligned to the reference genome to check for any problems. If the sequence is fine, no changes are needed. If the assembly is unsatisfactory

(gaps, low coverage), the pipeline-protocol has to be optimised i.e. improved checks for contaminations, trim adaptors/indices, decrease coverage of the reads. Decontamination is very important and crucial as cleaner reads give better assembly results. The termini of the assembled viral genome have eventually to be optimised manually by assembling first/last 300-400 bp using the reference-based approach.

#### 1.2.3.3 TANOTI for reference-based genome assembly

TANOTI assembles short reads of highly variable sequencing data by mapping the reads to a reference genome (TANOTI: a rapid BLAST-guided read mapper for small, divergent genomes (*manuscript communicated*)) and was developed to detect resistance mutations in HCV (Hepatitis C Virus). In contrast to SPADES it is a reference-based genome assembly method created specifically for short viral genomes with high variability. The results are mapped as read-mountains or “stacks” that further need to be calculated into a consensus sequence. The consensus sequence is then aligned to the reference genome to see if the assembly was successful. If it is not the case, especially when the genome varies, it has to be edited manually.

Furthermore, using a reference-based genome assembly does not show genome rearrangements and other changes so important genomic information can get lost. Therefore, it is better to use a *de novo* approach as it takes any change in the genome into account.

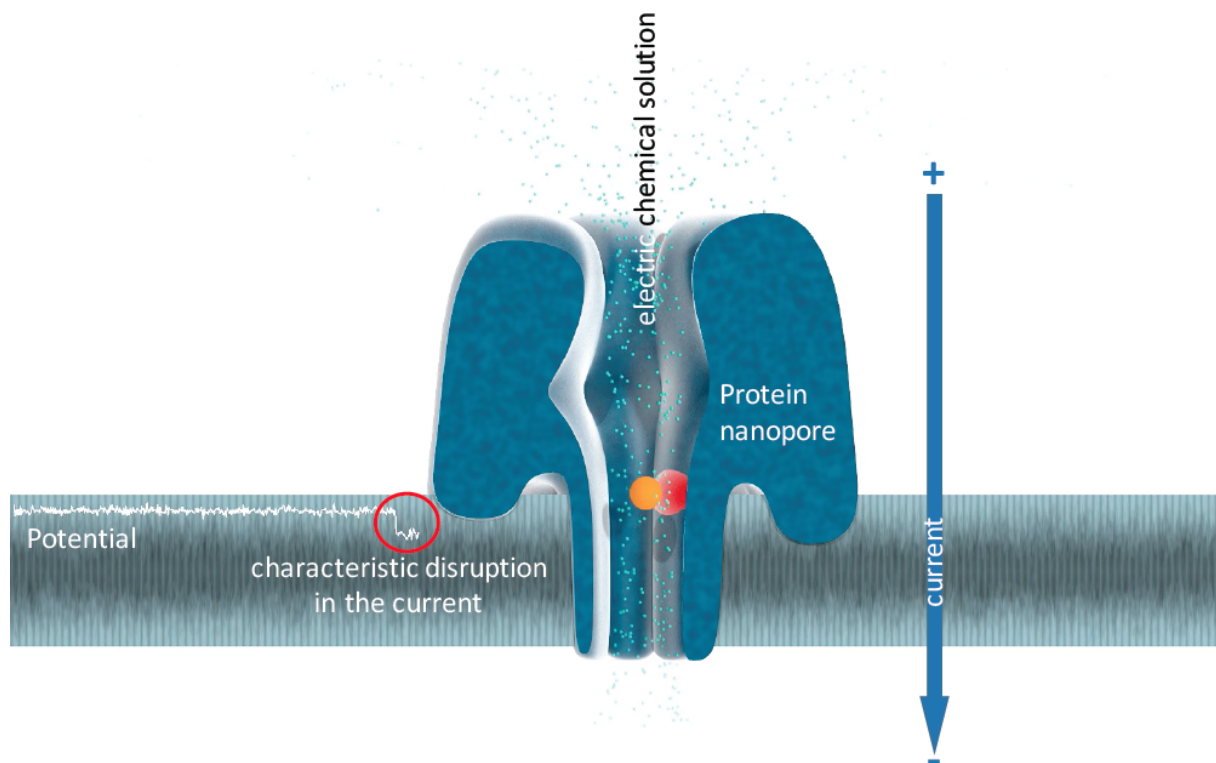
#### 1.2.3.4 STAMPY for reference-based genome assembly

STAMPY is a reference-based genome assembler that was specifically developed for short Illumina generated reads and is very user friendly. Its distinctive feature is that it can map reads to reference genome faster and more sensitive than other software. Even if reads contain variations and mutations such as insertions or deletions making it a great tool with the ability for construction of the whole genome, including termini, but considering sequence variations (Lunter and Goodson 2011).

#### 1.2.4 MinION sequencing technology

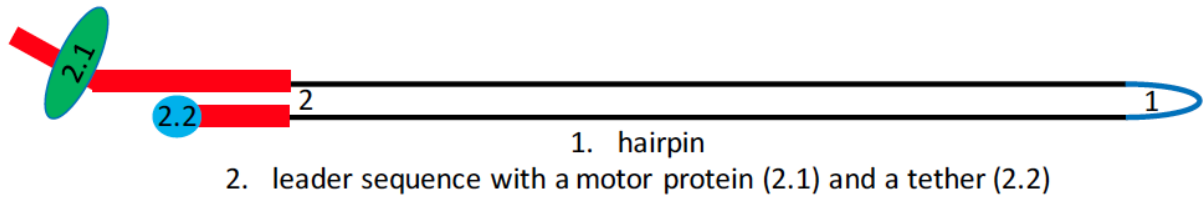
Oxford Nanopore Technologies created a long read single-molecule sequencing set up which uses a membrane made of a synthetic polymer. The membrane has a very high electrical resistance and protein nanopores are inserted into it. By applying an electrochemical solution,

a potential across the membrane is created. This results in an ionic current through the pores. When a molecule (RNA, DNA, proteins) enters the pore, it creates characteristic changes in the current. These current changes are also specific for a chain-group of nucleotides, such as k-mers (Figure 8). Briefly, each possible k-mer can be identified as an individual signal, which results in more than 1000 different signals rather than 4.



**Figure 9. Overview of the MinION sequencing technology (Oxford Nanopore Technologies, adapted).**

During the library preparation the DNA is physically sheared into a range of 6 to 20 kb by forcing the nucleic acid through an orifice by centrifugation. First the fragment ends are repaired adding an A-tail. This is followed by the addition of a leader sequence with a motor protein and a tether, as well as a hairpin structure (Figure 9). The purified library is loaded into the flow cell, the leader sequence interacts with the pore and the motor protein attaches to the pore. The attachment causes structural changes in the motor protein and it starts to unzip the leader sequence of the double stranded template, which then passes through the pore nucleotide per nucleotide until the first strand is sequenced and reaches the hairpin structure, which allows a subsequent sequencing of the second strand without any interruption.



**Figure 10. Template for MinION sequencing containing a leader and a hairpin structure (Oxford Nanopore Technologies, adapted).**

This technology is quite interesting in terms of template termini. As the template can be sequenced fully it allows the detection of terminal nucleotides.

The MinION flowcell has 512 channels, so 512 molecules can be sequenced/identified at the same time, which are connected to 4 wells, providing data from one well at a time (Lu, Giordano, and Ning 2016). 5-mer or 6-mer define base calling for each nucleotide (ATCG) which gives  $4^5$  or  $4^6$  possible combinations. The MinION has a throughput of ca. 70 bp per second and can process templates of up to 100 kb.

Compared to the low but non-random error rate of the Illumina MiSeq the MinION system creates high levels of random errors of 10 %, but as error sites are distributed across the genome in a random manner, with increasing coverage per site the overall error rate is much reduced in comparison to Illumina reads.

**Table 4. Comparing Illumina MiSeq and Oxford Nanopore Technologies MinION sequencing systems (Goodwin, McPherson, and McCombie 2016).**

Sequencing Technology	Read length	Throughput	Generated reads	Costs per Gb
MiSeq	2x 75 paired end	3.3-3.8 Gb	44-50 M	250 USD
MinION	> 8 kb	up to 1.5 Gb	> 100000	750 USD

The MinION can generate 10-20 Gb of DNA data with read lengths of up to 100 kb. The output files that are produced in the sequencing run are FAST5 files and have to be processed in METRICHOR (Oxford Nanopore Technologies), a real time base-calling software for MinION sequencing. Data that pass this step are stored in a new FAST5 file. NANOOK (Leggett et al. 2015) uses this file system to create a run report with basic information about the run such as coverage plots, lengths of sequenced fragments, pore occupation etc. The FAST5 file can be

converted into FastA or FastQ files using PORETOOLS (Loman and Quinlan 2014) for further workflow and analysis such as quality control, adaptor trimming and reference-based genome assembly using BWA (Li 2013). PORECHOP (Wick et al. 2017) can be used to solve concatemer issues.

### 1.3 Sequence Analysis

#### 1.3.1 Overview

There are many possibilities to analyse a sequenced genome. A variety of algorithms, scripts and softwares can calculate a value of interest to describe a particular feature of a site in a genome, or the genome itself and genome evolution. These analysis techniques can contribute to understanding the character, the evolution and spread of virus genomes. Here I will briefly introduce the models used in this study to analyse IPNV genomes.

#### 1.3.2 Calculating Phylogenies using BEAST

To look at the phylogeny of sequences bioinformaticians and mathematicians have developed several tools, that differ in models and therefore in computing time. This work only uses the methodology for the purpose of phylogenetic analysis so I will only briefly explain the subject matter. It was shown that tools which are based on Bayesian Markov Chain Monte Carlo analysis perform stable and accurate analysis of molecular sequences taking less time to perform the analysis compared to heuristic optimization based on maximum likelihood (ML) (Huelsenbeck and Ronquist 2001, Huelsenbeck et al. 2001). MCMC does not require marginal likelihood calculation (Peter G. Foster, 2009). By using random sampling MCMC generates equitable samples out from a distribution of samples, therefore it can be used to analyse posterior distributions in Bayesian inference. It handles complex models and parameters (Peter G. Foster, 2009). MCMC perturbs stochastically the current tree and proposes a new tree. The new tree is accepted or rejected. The new tree is then used for further perturbation (Huelsenbeck et al., 2001). The BEAST (Bayesian Evolutionary Analysis by Sampling Trees) package offers such an approach to analyse sequences to create evolutionary models (Drummond et al. 2012). Comparing to ML the Bayesian approach describes the branch stability not via bootstrap values but rather through posterior probabilities, which is a combination of a phylogeny's prior probability and tree likelihood, calculated using MCMC (Huelsenbeck and Ronquist 2001, Huelsenbeck et al. 2001). The tree with the highest probability is considered as the best estimated phylogeny. Comparing to the ML approach, where the tree with the highest likelihood is considered as the best (Rannala and Yang 1996).

### 1.3.3 Calculating phylogenies using RAXML

To perform phylogenetic analysis using a maximum likelihood algorithm, RAXML (Randomised accelerated maximum likelihood) provides a great graphic user interface. Its particular benefit is the fast-maximum likelihood calculating performance from a nucleotide substitution model, which describes the process of genetic variation, returning trees with high likelihood scores with the opportunity for bootstrap analysis.

Nucleotide substitution models calculate evolutionary distances (divergence) between sequences that originated from a common ancestor by measuring the number of changes between sequences. The calculated genetic distance number is inverse proportional to the, genetic difference of the respective sequences.

A bootstrap analysis shows the robustness of a tree by providing a statistical error, even if the sampling distribution is unknown. This is done by generating new alignments from the original data but with replacements in the chosen columns. This new set of sequences is then used to generate trees. By calculating the proportion of each clade among the replicates it gives a statistical confidence to support the calculated tree. (Silvestro and Michalak 2012, Lemey 2009, Stamatakis 2014, 2006, Stamatakis, Ludwig, and Meier 2004, Kaehler et al. 2014).

### 1.3.4 CAI

The codon adaptation index measures synonymous codon usage biases given a genome reference. CAI can describe synonymous changes of virus genomes towards a host genome over time showing how the codon preference of the virus adapts or evolves towards codon usage of the host. Through selection pressure and drifts a virus can conserve mutations and this way optimise replication efficiency and increase the transmission rate. For example, if a virus codes for tRNA<sup>CUC</sup> (Leucine) but the hosts translational system offers more tRNA<sup>CUU</sup> for Leucine instead, there is a selective pressure on the virus that could drive the evolution towards a mutation for coding tRNA<sup>CUU</sup> in order to optimise the virus' replication and translation. In some persistent viruses we even see a codon usage pattern that reduces protein expression in order to avoid activation of the immune system. Calculating the CAI of a virus to its host can help to gain insight in viral evolution and virulence of a virus as viruses optimise their survival by not killing the host immediately but rather try to be shed as long as possible and for long distances.

The CAI or normalized CAI (nCAI) for each coding protein is calculated as followed:  $nCAI = rCAI/eCAI$ . This can be performed on an online platform <http://genomes.urv.es/CAIcal> and <http://genomes.urv.es/CAIcal/E-CAI>. The tool CAICAL is used to calculate raw CAI (rCAI). Expected CAI (eCAI) is generated by creating multiple random sequences similar to those that are used for calculation. The output is understood as greater or lower than 1. If the values are higher than 1 it indicates that the virus is adapting its codon usage to the one of the host (Sharp, Tuohy, and Mosurski 1986, Puigbo, Bravo, and Garcia-Vallve 2008b, a) (Di Paola, de Melo Freire, and de Andrade Zanotto 2018).

### 1.3.5 Selection analysis using HYPHY

Selection analysis gives the opportunity to see which sites in a genome or a protein coding region are under positive selection pressure. There are a lot of different established models. For this study I decided to use the HYPHY package (Pond, Frost, and Muse 2005), with variety of models which work based on an optimised and extended version of the codon based MG94 model which describes protein coding sequence evolution by modelling codon substitution using a Markov process (Muse and Gaut 1994, Goldman and Yang 1994).

One possibility to calculate selection is by using SLAC (Single Likelihood Ancestor Counting) (Kosakovsky Pond and Frost 2005, Pond and Frost 2005). This method is able to model variations in synonymous and non-synonymous rates under the MG94xREV codon model.

Another model is MEME (Mixed Effects Model of Evolution) (Murrell et al. 2012) which also works under the MG94xREV codon model. MEME hypothesise that sites have been under episodic positive or diversifying selection under a proportion of branches.

### 1.3.6 Dinucleotide composition

Dinucleotides occur in 16 variants, AA, AC, AG, AT, CA, CC, CG, CT, GA, GC, GG, GT, TA, TC, TG, TT. Especially the motif CG or CpG is of big interest as it is a methylation cite in vertebrates and besides, suppressed throughout the DNA (Karlin, Doerfler, and Cardon 1994), due to cytosine methylation. According to a study by Cheng et al. there is no under-representation of CpG observed for dsRNA viruses infecting vertebrate hosts as their mean  $CpG_{O/E}$  of 0.88 defines as not under-represented in that study (Cheng et al. 2013), There are several explanations for under-representation of CpG motifs beside methylation, especially not for RNA viruses without DNA intermediates during replication (Cheng et al. 2013). One possible



explanation could be the higher energy of CpG in DNA, although the CpG energy in RNA is not specifically higher than for other dinucleotides (Breslauer et al. 1986, Serra and Turner 1995). Another idea is immune escape mechanism of a virus. This is quite possible as non-methylated CpG have an immune stimulating effect (Goldberg, Urnovitz, and Stricker 2000, Chen et al. 2001, Cheng et al. 2013), as it was shown for the Toll-like receptor 9, which initiated an innate immune response after binding of CpG motifs (Tyagi, Kumar, and Singh 2017). On the other hand, it was shown that systems that recognise dsRNA, like the Toll-like receptor 3, the RIG-I-like RNA helicase and the dsRNA binding protein, do not have a CpG preference, indicating that there might be no pressure on the CpG content from the host. Either way, the representation of CpG plays an important role in terms of viral evolution with the host, as the mean  $CpG_{O/E}$  in vertebrates is 0.47 (Cheng et al. 2013).

To estimate the dinucleotide composition, a ratio between observed CpG and estimated CpG is calculated ( $CpG_{O/E}$ ). A CpG abundance is considered low when  $CpG_{O/E} \leq 1$  and high when  $CpG_{O/E} \geq 1$  (Cheng et al. 2013, Tyagi, Kumar, and Singh 2017). There are different possibilities to achieve that. In this study EMBOSS COMPSEQ was used on an online platform [<http://emboss.bioinformatics.nl/cgi-bin/emboss/compseq>] to perform the calculation (Rice, Longden, and Bleasby 2000).

## Objectives

This project aimed at studying the molecular epidemiology and virus to host adaptation of IPNV in Scotland by whole genome sequence analysis of IPNV isolates spanning 3 decades from 1982 to 2014, as well as at optimising a protocol avoiding specific terminal primers in order to have a WGS protocol that can be used for any possible virus.

Specific objectives were:

- To select and culture IPNV from diagnostic TCID<sub>50</sub> supernatants.
- To process viral RNA of IPNV culture supernatants for next generation sequencing.
- To optimise whole genome next generation sequencing for IPNV by comparing two RNA processing protocols, the Goettingen (GOEP) (Dilcher et al. 2012) and the Glasgow (GLAP) (Thomson et al. 2016) protocol, as well as two sequencing platforms, Illumina MiSeq and Oxford Nanopore Technologies MinION.
- To optimise IPNV genome assembly
- To assess IPNV genome evolution using several bioinformatics assessment tools.
- To investigate a possible virus-to-host adaptation.
- To see if IPNV is under a selective pressure and what could possibly drive it.
- To detect possible reassortant isolates.

## 2 Methods

### 2.1 Whole genome sequencing of diagnostic TCID 50 Isolates

#### 2.1.1 Bulking up diagnostic TCID50 IPNV isolates in cell culture

##### 2.1.1.1 Maintenance of cell lines used for IPNV growth

###### 2.1.1.1.1 General description

For this work, Chinook salmon embryonic cells (CHSE-214) (Fryer, Yusha, and Pilcher 1965), which derived from a Chinook salmon (*Oncorhynchus tshawytscha*) embryo and TO cells, which derived from head kidney leucocytes of the Atlantic salmon (*Salmo salar*) (Wergeland and Jakobsen 2001), were used. The CHSE-214 and TO cell lines were grown at 22 °C and ca. 4 % CO<sub>2</sub> in Eagle's Minimum Essential Medium (EMEM; Gibco) with 10 % Foetal Bovine Serum (FBS; Gibco), 1x non-essential amino acids (NEAA; Gibco) and 2µM L-glutamine (Gibco). In order to maintain a healthy and even monolayer cells had to be passaged regularly.

###### 2.1.1.1.2 Splitting CHSE-214 or TO cells with a 1:3 split ratio

CHSE-214 or TO cells were examined visually every day under a microscope at 4x optical magnification. When the cell monolayer showed 100 % density and the cells appeared healthy, old media was decanted and the cell monolayer was washed twice with 5mL PBS (Gibco). Then 1 mL trypsin-EDTA (0.05 %; Gibco) was added to detach the cells from the flask surface. After ca. 1min some of the trypsin-EDTA was poured off, while the cells were still 100 % attached to the flask surface, leaving approximately 0.5 mL in the flask. The enzymatic reaction of the trypsin was observed under the microscope to determine the point of cell detachment. When the cells rounded up and no cell-to-cell contact was visible, the flask was shaken slightly to detach the cells from the surface. The cells were then re-suspended in EMEM+10 % FBS to stop the enzymatic activity of trypsin and aspirated by pipetting up and down several times to create a single cell suspension. The content was equally transferred into one to three new tissue culture flasks (TC; Sarstedt) of the same size as the parent flask, containing 5 mL EMEM+10 % FBS. The content of each flask was mixed by pipetting up and down several times and the cells were incubated at 22 °C and 4 % CO<sub>2</sub> with a 1/4<sup>th</sup> opened lid for gas exchange.

#### 2.1.1.1.3 Sub-culturing CHSE-214 and TO cells of a 25 cm<sup>2</sup> TC into a 75 cm<sup>2</sup> TC with a 1:3 split ratio

Cells of a 25 cm<sup>2</sup> TC were processed in the same manner as described in 2.1.1.1.2. The cells were then re-suspended in EMEM+10 % FBS and aspirated by pipetting up and down several times to create a single cell suspension. The cell suspension was transferred into a 75 cm<sup>2</sup> TC containing 17 mL EMEM+10 % FBS. The content was mixed by pipetting up and down several times and the cells were incubated at 22°C and 4% CO<sub>2</sub> with a 1/4<sup>th</sup> opened lid for gas exchange.

#### 2.1.1.1.4 Sub-culturing CHSE-214 and TO cells of a 75 cm<sup>2</sup> TC into a 175 cm<sup>2</sup> TC with a 1:3 split ratio

Cells of a 75 cm<sup>2</sup> TC were processed in the same manner as described in 2.1.1.1.2. The cells were then re-suspended in EMEM+10 % FBS (adjusted to a total volume of 9 mL) and aspirated by pipetting up and down several times to create a single cell suspension. Seven mL of the cell suspension was transferred into a 175 cm<sup>2</sup> TC containing 36 mL EMEM+10 % FBS and 2x 1 mL was transferred into two 25 cm<sup>2</sup> TCs containing 5 mL EMEM+10 % FBS each. The content of each flask was mixed by pipetting up and down several times and the cells were incubated at 22 °C and 4 % CO<sub>2</sub> with a 1/4<sup>th</sup> opened lid for gas exchange.

#### 2.1.1.1.5 Seeding a 24 well plate (WP) of CHSE-214 and TO cells for inoculation via adsorption

The cell monolayer of a tissue culture flask containing CHSE-214 or TO cells that are not older than 7 days, with a cell density of ca. 95 % and are in the replicative mode were examined visually under the microscope to ensure the cells fulfilled the requirements. Old media was decanted off and the cell monolayer was washed twice with 5 mL PBS. Then 1 mL trypsin-EDTA (0.05 %) was added to detach the cells from the flask surface. After ca. 1 min some of the trypsin-EDTA was poured off, while the cells were still 100 % attached to the flask surface, leaving approximately 0.5-1 mL in the flask. The enzymatic reaction of the trypsin was observed under the microscope to determine the point of cell detachment. When cells rounded up and no cell-to-cell contact was visible, the flask was shaken slightly to detach the cells from the surface. The cells were then re-suspended in EMEM+10 % FBS and aspirated by pipetting up and down several times to create a single cell suspension. Adjusting the resuspension media to 3 or 6 mL was performed if further subculturing with a 1:3 split ratio

was required. The cell suspension was aspirated by pipetting up and down. Cells were counted and seeded at a density of  $1.5 \times 10^5$  cells per well (see 2.1.1.1.5.1). The plate was incubated at 22 °C and 4 % CO<sub>2</sub>. The next day the plate was examined visually to ensure that each well had approximately the same cell density and the cells looked healthy. If the requirements were fulfilled, the well plate was used for inoculation with IPNV samples.

#### 2.1.1.1.5.1 Determination of cell numbers in a haemocytometer

A mix of 0.1 mL cell suspension in 0.1 mL Trypan Blue (0.5 %; Sigma Aldrich) was prepared in a polystyrene container (dilution factor: 2). A cover glass was placed on the haemocytometer so that Newton's rings were visible. One of the two counting chambers was filled with the cell-Trypan Blue mix and the haemocytometer was placed under the microscope. The focus was set on 10x optical magnification without orange filter. It was first ensured that the cell density was approximately the same in each of the 9 squares of the haemocytometer. Count of living cells (not stained blue) in 3 of the 9 squares at least was performed and the average cell amount was calculated. The depth of the haemocytometer is 0.1 mm. The volume of 1 mm<sup>2</sup> is  $1 \times 1 \times 0.1 = 0.1 \text{ mm}^3 = 0.0001 \text{ cm}^3 = 10^{-4} \text{ cm}^3$ . When  $1 \text{ cm}^3 = 1 \text{ mL}$  the cell concentration/mL is: average cell count per 1 mm<sup>2</sup> x dilution factor (2) x 10<sup>4</sup>. The total cell amount was calculated by multiplying the cell amount/mL with the volume of the cell suspension. The cell number per well needed for CHSE-214 and TO cells was  $1.5 \times 10^5$  cells. For 24 wells + 2 wells extra (26 wells) a cell number of  $1.5 \times 10^5 \times 26 = 3.9 \times 10^6$  cells were required. The following calculation was performed to estimate the volume of the cell suspension containing  $3.9 \times 10^6$  cells:  $(3.9 \times 10^6 / \text{total cell number}) \times \text{total volume of cell suspension}$ . This volume was adjusted with EMEM+10 % FBS up to 26 mL and 1 mL of the dilution was pipetted into each well.

#### 2.1.1.1.6 Seeding a 24WP of CHSE-214 and TO cells for simultaneous inoculation

Cells were processed in the same manner as in 2.1.1.1.5. Cells were counted and seeded at a density of  $1.75 \times 10^5$  cells per well (see 2.1.1.1.5.1). The 24WP (Sarstedt) was incubated at 22 °C and 4 % CO<sub>2</sub> for maximum 30 min so cells do not attach to the well bottom and used for IPNV inoculation the same day.

### 2.1.1.2 Growing IPNV in CHSE-214 and TO cells

#### 2.1.1.2.1 General description

CHSE-214 and TO cells were inoculated with diagnostic TCID<sub>50</sub> IPNV isolates via adsorption and via simultaneous inoculation and processed the same way. HBSS+2 % FBS was used as a diluent and as a negative control. Wells with a certain isolate in either CHSE-214 or TO cells that showed a CPE were collected and passaged into a 25 cm<sup>2</sup> TC containing the corresponding cell line. When a full CPE was visible in the 25 cm<sup>2</sup> TC the content was harvested and passaged into a 175 cm<sup>2</sup> TC with the corresponding cell line to bulk up the viral particles. Bulking up viral particles in a 175 cm<sup>2</sup> TC was only required for the Goettingen protocol. Furthermore, 4x 0.5 mL of the harvested supernatant of a 25 cm<sup>2</sup> TC were stored at -70 °C as viable IPNV stocks.

Negative wells were passaged up to 2 times into a new 24WP containing the corresponding cell line. The chosen passage method was a freeze (-70 °C)/thaw passage via adsorption according to the OIE guidelines for IPNV diagnostics.

EMEM+10 % FBS was used as growth media for inoculation after the adsorption period. To prevent bacterial contamination a 1x kanamycin/penicillin/streptomycin antibiotics mix was added into EMEM+10 % FBS. One flask of the same size was treated exactly the same, but without inoculation, was used as an internal control.

In order to perform the Goettingen protocol for whole genome sequencing, supernatant of a 175 cm<sup>2</sup> TC was used. For the Glasgow protocol the supernatant of a 25 cm<sup>2</sup> TC was used.

#### 2.1.1.2.2 Inoculation via adsorption

IPNV isolates of interest were defrosted and kept on ice. A 1:10 and 1:50 dilutions was prepared. Media of a 24WP prepared on the previous day was pipetted off and 150 µL of the neat sample, the 1:10 and the 1:50 dilutions were pipetted in a column in following order: neat, 1:10, 1:50, negative control (NC). The plate was sealed with plastic paraffin film and incubated at 15 °C and 1 % CO<sub>2</sub> for 3 h. The plate was rocked every hour to ensure that the cells were equally covered with the inoculum. Then 1.5 mL of EMEM-10 % FBS media were added to each well and the plate was incubated at 15 °C and 1 % CO<sub>2</sub>. Every day the plate was examined for a cytopathic effect (CPE).

#### 2.1.1.2.3 Simultaneous inoculation

IPNV isolates of interest were defrosted and kept on ice. A 1:10 and 1:50 dilutions was prepared. Into each well of a freshly prepared 24WP 100  $\mu\text{L}$  of the neat sample, the 1:10 and the 1:50 dilutions were pipetted in a column in following order: neat, 1:10, 1:50, NC. The plate was incubated at 15 °C and 1 %  $\text{CO}_2$ . Every day the plate was examined for a CPE.

#### 2.1.1.2.4 Passage of positive well(s) into a 25 cm<sup>2</sup> TC

When a full CPE was visible in a well, or in several wells of the same isolate, the supernatant of the specific well was collected or multiple wells were pooled into a centrifuge tube and centrifuged at 2500x  $g$  and 4 °C for 15 min. The supernatant was decanted in a polystyrene container. A dilution of  $10^{-2}$  up to  $10^{-4}$  (depending on the virulence of the isolate) was prepared on ice. Old media of a 25 cm<sup>2</sup> TC was decanted off and 1mL of the diluted supernatant was added into the flask. For a NC, another 25 cm<sup>2</sup> TC was incubated with 1mL of the diluent. The flasks were incubated for 1 h at 15 °C and 1 %  $\text{CO}_2$ . After incubation, the inoculum was discarded and 5mL EMEM+10 % FBS media were added. The flasks were incubated at 15 °C and 1 %  $\text{CO}_2$  and examined every day for a CPE.

When a full CPE was visible in the 25 cm<sup>2</sup> TC, the supernatant was collected into a 15mL centrifuge tube and centrifuged at 2500x  $g$  and 4 °C for 15 min. The supernatant was further filtered through a 0,2  $\mu\text{m}$  sterile filter (Sartorius). The supernatant was stored at -70 °C or it was used immediately for inoculation of a 175 cm<sup>2</sup> TC for the Goettingen protocol or used immediately for the Glasgow protocol.

#### 2.1.1.2.5 Passage of positive 25 cm<sup>2</sup> TC into a 175 cm<sup>2</sup> TC

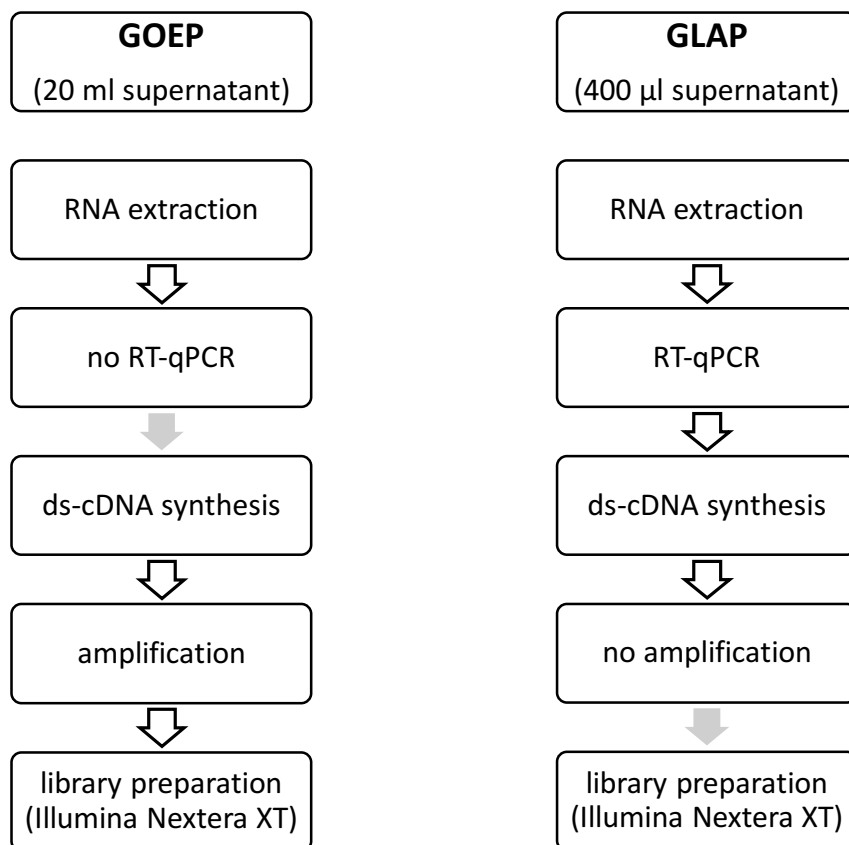
The supernatant from a 25 cm<sup>2</sup> TC was used for inoculation of a 175 cm<sup>2</sup> TC. A dilution of  $10^{-3}$  up to  $10^{-5}$  (depending on the virulence of the isolate) was prepared on ice as followed: 100  $\mu\text{L}$  of the harvested supernatant + 900  $\mu\text{L}$  of the diluent = 1 mL of a  $10^{-1}$  dilution, diluting it down to  $10^{-4}$ ; 1 mL of the  $10^{-4}$  dilution + 9 mL of the diluent = 10 mL of a  $10^{-5}$  dilution. Old media of a 175 cm<sup>2</sup> TC was decanted off and 7 mL of the dilution was added into the flask. For a NC, a 25 cm<sup>2</sup> TC, that was prepared from the same parent flask as the 175 cm<sup>2</sup> TC, was incubated with 1mL of the diluent. The flasks were incubated for 1h at 15 °C and 1 %  $\text{CO}_2$ . After incubation, the inoculum was discarded and 30 mL of EMEM+10 % FBS were added to the 175 cm<sup>2</sup> TC

(5mL to the NC). The flasks were incubated at 15 °C and 1 % CO<sub>2</sub> and examined every day for a CPE.

#### 2.1.1.2.6 Harvest IPNV supernatant of a 175 cm<sup>2</sup> TC for RNA extraction

The culture medium of infected cells was harvested when the cells showed 80-90 % CPE. Medium of a 175 cm<sup>2</sup> TC was transferred into a 50 mL centrifuge tube (Sarstedt) and clarified by centrifugation for 10 min at 2000 rpm (700x g). The new supernatant was centrifuged again for 5min at 4000 rpm (2800x g). After that, the supernatant was filtered into a new centrifuge tube through a 0.2 µm sterile filter. The purified supernatant was stored at -70 °C before it was processed further in the Goettingen protocol.

#### 2.1.2 IPNV RNA processing and sequencing



**Figure 11** Overview of the main steps of GOEP and the GLAP.

Cell culture supernatants containing IPN virus were used for RNA extraction, ds-cDNA conversion and library preparation (Figure 10). First 27 IPNV samples were processed with the



Goettingen (GOEP) protocol and 30 IPNV samples were processed with the Glasgow protocol (GLAP).

#### 2.1.2.1 RNA extraction and processing

##### 2.1.2.1.1 Goettingen protocol

This protocol was developed by Dr. Meik Dilcher, Department of Virology, University Goettingen, Germany and was used for whole genome sequencing of diagnostic TCID50 IPNV isolates. The initial protocol (GOEP) was designed for use of Titanium Shotgun Rapid Library Preparation and Pyrosequencing system by Roche. In this study, for both, the GOEP and GLAP, the Nextera XT DNA Library Preparation Kit (Illumina) was used instead.

##### 2.1.2.1.1.1 Solutions and Buffers

###### 2.1.2.1.1.1.1 1 M Tris, pH6,5 (10 0mL)

Tris (12.11 g; Sigma Aldrich) was dissolved in 80mL distilled water (Sigma Aldrich). The pH was formulated with HCl (VWR Chemicals). Then the solution was adjusted to 100mL with distilled water, autoclaved and stored at room temperature.

###### 2.1.2.1.1.1.2 0.5 M EDTA, pH8,0 (100 mL)

EDTA (18.62 g; Sigma Aldrich) was dissolved in 80mL distilled water. The pH was formulated with 10 N NaOH (Sigma Aldrich). Then the solution was adjusted to 100 mL with distilled water, autoclaved and stored at room temperature.

###### 2.1.2.1.1.1.3 5 M NaCl (500 mL)

NaCl (146.1 g; Sigma Aldrich) was dissolved in 500 mL distilled water, autoclaved and stored at room temperature.

###### 2.1.2.1.1.1.4 10x NTE (200 mL)

Twenty mL of 1 M Tris (pH6.5), 4 mL of 0.5 M EDTA (pH8.0) and 40 mL of 5 M NaCl were mixed. The volume was adjusted to 200 mL with distilled water and the pH was checked (pH ca. 7.0). The solution was then autoclaved and stored at room temperature.

#### 2.1.2.1.1.1.5 30 % PEG in NTE (500 mL)

PEG8000 (150 g; Sigma Aldrich) were dissolved in 50 mL 10x NTE. The volume was adjusted to 500 mL with distilled water. The solution was autoclaved and stored at room temperature.

#### 2.1.2.1.1.2 Extraction and processing of IPNV RNA

##### 2.1.2.1.1.2.1 IPNV RNA extraction

Supernatant from 2.1.1.2.6 was defrosted. To 20 mL supernatant 1.48 mL sterile 5 M NaCl and 10.80 mL sterile 30 % (w/v) PEG8000 in NTE (10 mM Tris, pH6.5; 1 mM EDTA; 100 mM NaCl) were added. The mix was incubated for 30min on a rocker at 4 °C and then centrifuged for 1.5 h at 3877x *g* and 4 °C. The pellet was re-suspended in 250 µL PBS (Gibco). To exclude contaminating nucleic acids an RNase-digestion of intact virus particles was performed by adding 12.5 µL of RNase A (20 mg/mL; final concentration: 1 mg/mL; PeqLab) to the sample. The reaction was incubated at 37 °C for 30 min. After that the volume was adjusted to 500 µL with PBS. The sample was further processed by adding 1.5 mL Trifast (PeqLab) per sample and incubated at room temperature for 5 min. The sample was then split into two 2 mL tubes. After that 200 µL chloroform (Sigma Aldrich) per 1 mL/sample were added followed by incubation for 10 min at room temperature. Next the samples were centrifuged for 5 min at 12.000x *g* and room temperature. The clear supernatant was transferred to a new tube and 2 µL glycogen (35 ng/µL; PeqLab) was added. Then 500 µL 100 % isopropanol (Sigma Aldrich) was added. The samples were incubated at -20 °C over night. The samples were centrifuged on the next day for 15 min at 12.000x *g* and 4 °C and washed twice with 1 mL cold 75 % ethanol (Sigma Aldrich). The pellets were dried for 20-30 min at room temperature in the open tube and re-suspended in 25 µL molecular grade water each. The eluted RNA was incubated for 5 min at 57 °C in a heating block and the duplicates were combined in one tube. The RNA was stored at -70 °C or processed further immediately. The RNA concentration was measured spectrophotometrically via NanoDrop 2.1.2.2.1.1.

##### 2.1.2.1.1.2.2 DNase treatment

To exclude contamination DNA 0.1x volume of 10x Turbo DNase buffer and 1 µL Turbo DNA (Life Technologies) were added and incubated for 5min at 37 °C in a heating block. To inactivate the DNase 0.1x volume of well re-suspended DNase Inactivation Reagent (Life

Technologies) were added. The sample was incubated for 5min at room temperature and centrifuged for 1.5 min at 10000x g. The clear supernatant was transferred to a new tube.

#### 2.1.2.1.1.2.3 Pellet Paint Precipitation

To further purify viral RNA a Pellet-Paint-Precipitation was performed. Pellet Paint® NF Co-Precipitant (PPP; Merck Millipore) and 3 M Na-Acetate (pH5,2; part of the PPP kit, Merck Millipore) were equilibrated to room temperature. Two µL of PPP, 0.1x volume of 3 M Na-Acetate and 2x volume 100 % ethanol were added. The mixture was incubated for 2 min at room temperature and centrifuged for 5 min at maximum speed. Afterwards the pellet was washed twice with 1 mL 70 % ethanol and one time with 1 mL 100 % ethanol. The pellet was then dried for 20-30 min at room temperature in the open tube, re-suspended in 25 µL molecular grade water and incubated for 5 min at 57 °C in the heating block. The RNA concentration was measured spectrophotometrically via NanoDrop. To estimate the exact RNA concentration an additional reading was performed at  $\lambda=600$  nm to exclude the PPP. The RNA concentration was then calculated as followed:  $\text{RNA concentration (ng/}\mu\text{L)} = (\text{reading}_{A260} - (\text{reading}_{A600} \times 0.84 / \text{PPP constant})) \times 40 / \text{RNA factor}$ . Additionally, 2x 1 µL of RNA was used for an in-house IPNV RT-qPCR, developed by Dr. Tharangani Herath, Institute of Aquaculture, University of Stirling, Stirling, UK. The RNA was then stored at -70 °C or processed further immediately.

#### 2.1.2.1.1.2.4 Ligation of a 3' iSP9 adapter

For a better termini coverage, a 3'-end iSP9 adaptor (MW=10902.14 g/mol; Synthesis: 1.00D = 30.3 µg or 2.7 nmol; HPLC-grade) was ligated to the viral RNA according to the modified protocol of (Maan et al. 2007). The reaction was set up as shown in Table 5 using components from the T4 Ligase 1 Kit (NEB):

**Table 5. Ligation reaction-set up.**

<b>Reagent</b>	<b>per sample</b>
RNA (500–1000 ng)	max. 12 $\mu$ L
3' iSP9 adaptor (500-1000 ng)	2.0 $\mu$ L
10x T4 RNA Ligase 1 Buffer	2.0 $\mu$ L
10 mM ATP	2.0 $\mu$ L
T4 RNA Ligase 1 (10000 Units/mL)	1.0 $\mu$ L
Protector RNase-Inhibitor (40 Units/ $\mu$ L; Roche)	1.0 $\mu$ L
Molecular grade water	adjust to 20 $\mu$ L
<b><math>\Sigma</math> Volume:</b>	<b>20 <math>\mu</math>L</b>

The reaction was incubated over night at 4 °C and purified the next day with the Qiagen RNeasy-MinElute Cleanup Kit according to the manufacturer's protocol. Since the adaptor sequence could not be found in the sequenced samples, it was decided to leave this step out in order to avoid RNA losses during this procedure.

#### *2.1.2.1.1.2.5 Transplex Whole Transcriptome Amplification*

The Transplex Whole Transcriptome Amplification Kit (WTA2) by Sigma-Aldrich (modified protocol) was used to generate fragmented ds-cDNA of the viral RNA.

The first part synthesis reaction was prepared as listed in Table 6:

**Table 6. Synthesis reaction, part I/ II- set up.**

<b>Reagent</b>	<b>NC</b>	<b>Per sample</b>
60 ng RNA	$\emptyset$	max. 8.46 $\mu$ L
Library Synthesis Solution	1.5 $\mu$ L	1.5 $\mu$ L
Nuclease-free water	8.46 $\mu$ L	adjust to 9.96 $\mu$ L
<b><math>\Sigma</math> Volume:</b>	<b>9.96 <math>\mu</math>L</b>	<b>9.96 <math>\mu</math>L</b>

The mix was incubated for 5 min at 95 °C, then directly transferred on ice and incubated on ice for 2 min. Following components were then added to 9.96 µL of cooled-primed RNA (Table 7) and cycled as followed (Table 8):

**Table 7. Synthesis reaction, part II/ II- set up.**

Library Synthesis Buffer	1.5 µL
Molecular grade water	2.34 µL
Library Synthesis Enzyme	1.2 µL
<b>Σ Volume:</b>	<b>15 µL</b>

**Table 8. Synthesis reaction-cycling conditions.**

18 °C	11 min
25 °C	10 min
37 °C	30 min
42 °C	10 min
70 °C	20 min
12 °C	Hold

Samples were stored at -20 °C or processed with the Amplification reaction step. This step includes a nucleotide stain, EvaGreen (Jena Bioscience), which helps to estimate the amplification plateau, so amplification of “unspecific” genomes or primer-dimer formations can be avoided. The reaction was set up in a 96 well plate (WP) and cycled as followed (Table 9, Table 10) in the Light Cycler 480 (Roche):

**Table 9. Amplification with EvaGreen- set up.**

Molecular grade water	59.7 µL
Amplification Mix	7.5 µL
WTA2 dNTPs	1.5 µL
Amplification Enzyme	0.75 µL
EvaGreen (50x)	0.75 µL
Library Synthesis Reaction	4.8 µL

**Table 10. Amplification with EvaGreen- cycling conditions in the LC480II.**

Denaturation	94 °C	2 min	
Amplification	94 °C	30 sec	25
	70 °C	5 min	cycles
Cooling	40 °C	5 min	

When the amplification curves fulfilled the requirements and it was clear when the amplification reaches its plateau, samples without EvaGreen were prepared in duplicates and cycled as followed in a thermal cycler (Table 11, Table 12):

**Table 11. Amplification without EvaGreen- set up.**

Molecular grade water	60.45 µL
Amplification Mix	7.5 µL
WTA2 dNTPs	1.5 µL
Amplification Enzyme	0.75 µL
Library Synthesis Reaction	4.8 µL

**Table 12. Amplification without EvaGreen- cycling conditions in a thermocycler.**

Denaturation	94°C	2min	
Amplification	94°C	30sec	X cycle(s)
	70°C	5min	+ 2-3 cycles
Cooling	40°C	5min	

The samples were stored at -20 °C or purified directly via the QiaQuick PCR-Purification-Kit. The DNA concentration was measured via NanoDrop. The values were used to dilute the sample to 50 ng/µL in 100 µL (in molecular grade water) for a following Ampure Bead purification to exclude DNA fragments smaller than 300 bp. The diluted sample was mixed with 70 µL of well re-suspended beads (0.9x) and incubated for 10 min on a mixer at room temperature. The sample was placed in a tube magnet (Life Technologies) to discard the supernatant. After three washing steps (30 sec each) with 500 µL freshly prepared 70 % ethanol while keeping the sample in the magnet, the pellet was dried in the magnet for at

least 3 min at room temperature. The DNA was eluted in 18  $\mu\text{L}$  molecular grade water and transferred into a new tube. The DNA concentration was measured via NanoDrop. Additionally, 1  $\mu\text{L}$  of the sample, diluted to 25 ng/ $\mu\text{L}$  with molecular grade water, based on the NanoDrop results, was used for a fragment analysis on a Bioanalyzer (Agilent Technologies) (see 2.1.2.2.3). The DNA was used immediately for library preparation via Illumina Nextera XT DNA Library Preparation Kit or stored at  $-20\text{ }^{\circ}\text{C}$ .

#### 2.1.2.1.2 Glasgow protocol

This protocol was developed by Dr. Emma Thomson, MRC-University of Glasgow, Centre for Virus Research, Glasgow, UK and was used for whole genome sequencing of diagnostic TCID50 IPNV isolates. This protocol requires less input material, has one PCR step less than the Goettingen protocol, is cheaper and faster. The initial protocol used the NEB Next<sup>®</sup> Ultra<sup>™</sup> Directional RNA Library Prep Kit for Illumina (New England Biolabs) for library preparation and the KAPA SYBR<sup>®</sup> FAST qPCR Kit (Kapa Biosystems) for library normalisation. In this study, the Nextera XT DNA Library Preparation Kit (Illumina) was used instead.

##### 2.1.2.1.2.1 IPNV RNA extraction

IPNV RNA was extracted via Agencourt RNAdvance Blood Kit according to Part C – Agencourt RNAdvance Blood 2mL Tube Protocol (Beckman Coulter) with following changes:

Filtered cell culture supernatant from 2.1.1.2.4 (400  $\mu\text{L}$ ) was treated with a Protease K (50 mg/mL)-lysis buffer mix for 15 min at  $55\text{ }^{\circ}\text{C}$ . After an incubation of 2 min at room temperature 410  $\mu\text{L}$  of Bind 1/isopropanol solution, containing 10  $\mu\text{L}$  Bind 1 and 400  $\mu\text{L}$  100 % isopropanol, were added. The mix was incubated for 5 min at room temperature before the tubes were placed into a magnetic rack. The clear supernatant was discarded. The tubes were removed from the magnetic rack and washed with 400  $\mu\text{L}$  of the wash buffer. The tubes were placed into the magnet to remove the wash buffer. The second wash was performed by adding 800  $\mu\text{L}$  of 70 % ethanol into the tubes, which remained in the magnet. After 1 min the ethanol was removed and the beads were air-dried. The tubes were removed from the magnet and 100  $\mu\text{L}$  of the DNase solution, containing 85  $\mu\text{L}$  water, 10  $\mu\text{L}$  10x DNase buffer and 5  $\mu\text{L}$  DNase I (Life Technologies), was added to each tube prior to incubating at  $37\text{ }^{\circ}\text{C}$  for 5 min. Then 200  $\mu\text{L}$  of the Bind 2 buffer was added and the mix was incubated at room temperature before the tubes were placed in the magnet and the clear supernatant was discarded. The beads were

washed twice by adding 800  $\mu\text{L}$  of 70 % ethanol, incubating for 1 min in the magnetic rack and discarding the ethanol. The beads were air-dried in the magnet for 10 min. The RNA was eluted by re-suspending the beads in 20 $\mu\text{L}$  molecular grade water and incubating the mix for 2 min at room temperature. The tubes were placed into the magnet and the clear RNA solution was transferred into a fresh tube. An aliquot was prepared for concentration analysis and in-house IPNV RT-qPCR and the stock was stored at  $-70\text{ }^{\circ}\text{C}$  until further processing.

#### 2.1.2.1.2.2 ds-cDNA synthesis

##### 2.1.2.1.2.2.1 Synthesis of the first cDNA strand

The RNA was converted into cDNA by mixing 11  $\mu\text{L}$  of RNA (maximum 1  $\mu\text{g}$  total RNA) with 10  $\mu\text{L}$  of the following master mix using components from Life Technologies (Table 13):

**Table 13. cDNA synthesis- master mix.**

5x Reverse Transcriptase Buffer	4 $\mu\text{L}$
Super Script III RT	2 $\mu\text{L}$
RNaseOUT Recombinant Ribonuclease Inhibitor	1 $\mu\text{L}$
Random Hexamers (50 $\mu\text{M}$ )	1 $\mu\text{L}$
dNTP's (2.5mM)	1 $\mu\text{L}$
DTT	1 $\mu\text{L}$
<b><math>\Sigma</math> Volume</b>	<b>10 <math>\mu\text{L}</math></b>

The cDNA reaction was incubated in a thermocycler as followed (Table 14):

**Table 14. cDNA synthesis-cycling conditions.**

25 $^{\circ}\text{C}$	10 min
55 $^{\circ}\text{C}$	60 min
70 $^{\circ}\text{C}$	15 min
4 $^{\circ}\text{C}$	hold



#### 2.1.2.1.2.2.2 Synthesis of the second cDNA strand

Using the NEB Next mRNA Second Strand Synthesis Module Kit the ds-cDNA was prepared by adding 20  $\mu\text{L}$  of the previously produced single stranded cDNA (ss-cDNA) to 60  $\mu\text{L}$  of the following master mix (Table 15, Table 16) using components from the NEBNext<sup>®</sup> mRNA Second Strand Synthesis Module (NEB):

**Table 15. cDNA second strand synthesis reaction- master mix.**

10x Second Strand Synthesis Reaction Buffer	8 $\mu\text{L}$
Second Strand Synthesis Enzyme	4 $\mu\text{L}$
Molecular Grade Water	48 $\mu\text{L}$
<b><math>\Sigma</math> Volume</b>	<b>60 <math>\mu\text{L}</math></b>

The ds-cDNA reaction was incubated in a thermocycler as followed:

**Table 16. cDNA second strand synthesis- cycling conditions.**

25 °C	2 min 30 sec
4 °C	hold

#### 2.1.2.1.2.2.3 Clean-up of the ds-cDNA

After this step, the product was either stored at -20 °C or processed directly. The product was cleaned up by adding 80  $\mu\text{L}$  of Ampure XP beads (Beckman Coulter) per sample (1x) and incubation for 5 min at room temperature. The sample was placed into a magnet rack so the clear supernatant could be discarded. The magnetic beads pellet was washed twice with 200  $\mu\text{L}$  of freshly prepared 70 % ethanol, while keeping the tube in the magnet and dried in the magnet for 5 min at room temperature to remove all ethanol traces. The pellet was eluted in 25  $\mu\text{L}$  molecular grade water and transferred into a new Eppendorf<sup>®</sup> DNA LoBind microcentrifuge 1.5 mL tube (Eppendorf). The DNA was stored at -20 °C or used for concentration measurement and library preparation via Illumina Nextera XT DNA Library Preparation Kit.

### 2.1.2.1.3 Comparison of the Glasgow and the Goettingen protocols regarding whole genome sequencing

For the purpose of saving time and funding the Glasgow protocol, which is cheaper and faster, and the Goettingen protocols were compared using samples IoA024 and IoA030 in terms of virus sequence completeness. The samples were bulked up as listed in Table 17:

**Table 17. IPNV samples used to compare the Goettingen and the Glasgow protocol.**

<b>Sample</b>	<b>Cell line</b>	<b>Flask size</b>	<b>Passage in cell culture</b>	<b>Protocol</b>
IoA024	TO	25 cm <sup>2</sup>	P3	Glasgow
IoA024	TO	175 cm <sup>2</sup>	P4	Goettingen
IoA030	TO	25 cm <sup>2</sup>	P1	Glasgow
IoA030	TO	175 cm <sup>2</sup>	P2	Goettingen

The samples were processed with either the Glasgow or the Goettingen protocol. The ds-cDNA was prepared as in 2.1.2.1.1 or 2.1.2.1.2 and libraries were prepared. The library was prepared and run as in 2.1.2.3.

### 2.1.2.2 Nucleic acid analysis

#### 2.1.2.2.1 RNA analysis

##### 2.1.2.2.1.1 Measure RNA concentration via Nano Drop

The sample panel of the Nano Drop (Thermo Fisher Scientific) was cleaned with molecular grade water before the first use. The instrument was adjusted with 2 µL molecular grade water. The setting was changed to RNA. The measurement was carried out with 2 µL of the sample.

##### 2.1.2.2.1.2 IPNV RT-qPCR

In order to estimate the threshold of concentration of IPNV genome equivalents higher than 10<sup>4</sup> copies/mL, which are needed for a successful whole genome sequencing, an in-house RT-qPCR for IPNV was used. It was developed by Dr. Tharangani Herath in the Institute of Aquaculture, Stirling, UK and includes a cloned IPNV standard (Ct=16; 10<sup>8</sup> copies/µl). The working dilution (Ct=19-20; 10<sup>7</sup> copies/µl) was diluted with molecular grade water in 1:10 steps down to 10<sup>1</sup> copies/µl. Samples, all standards and the NC (molecular grade water) ran in

duplicate in a Light Cycler 480 (Roche). The master mix was set up as followed according to the LightCycler 480 RNA Master Hydrolysis Probes Kit by Roche (Table 18):

**Table 18. IPNV RT-qPCR- set up.**

1. 2.7x Buffer mix	7.5 $\mu$ L (1x)
2. Activator (MnCl <sub>2</sub> ) 50 mM	1.3 $\mu$ L (3.25 mM)
Forward primer 10 $\mu$ M (10 pmol)	1.0 $\mu$ L (500 nM)
Reverse primer 10 $\mu$ M (10 pmol)	1.0 $\mu$ L (500 nM)
Probe 10 $\mu$ M (10 pmol)	0.4 $\mu$ L (200 nM)
3. PCR grade water	7.8 $\mu$ L
<b><math>\Sigma</math> Volume</b>	<b>19 <math>\mu</math>L</b>

The master mix was added to 1  $\mu$ L standard, NC or sample RNA in a 96WP, mixed, pelleted at approximately 2000 rpm and cycled in the Light Cycler 480II (Roche) as followed (Table 19):

**Table 19. IPNV RT-qPCR- cycling conditions.**

Reverse transcription	63 °C	3 min	
Activation	95 °C	30 sec	
Amplification	95 °C	5 sec	45
	60 °C	15 sec	cycles
Cooling	40 °C	40 sec	

For data analysis the Abs Quant/Fit Points algorithm (Roche) was used to estimate the Ct value and the copy number.

#### 2.1.2.2.2 DNA analysis

##### 2.1.2.2.2.1 Measure DNA concentration via Nano Drop

The sample panel was cleaned with molecular grade water before the first use. The instrument was adjusted with 2  $\mu$ L molecular grade water. The setting was changed to DNA. The measurement was carried out with 2  $\mu$ L of the sample.

#### 2.1.2.2.2.2 Gel electrophoresis

Gel electrophoresis was used as a PCR control. A 1 % Agarose gel (Bioline), which was considered as appropriate for the size of the PCR product (Table 20), containing 0.5 µg/mL (final concentration) ethidium bromide (Thermo Fisher Scientific) was loaded with 1 µL sample, which was prior diluted in 5 µL 6x Loading Dye (Thermo Fisher Scientific). A DNA Ladder (Thermo Fisher Scientific) was also loaded as a size control according to manufacturer's protocol. The gel was run for 40 min at 75 V.

**Table 20. Correlation between Agarose gel %-age and DNA size distribution (Ref.: <http://www.promega.co.uk/resources/pubhub/enotes/what-percentage-agarose-is-needed-to-sufficiently-resolve-my-dna-sample/?activeTab=0>).**

<b>% Gel and DNA-size distribution (Promega)</b>	<b>Agarose</b>	<b>0.5x TAE</b>
1 % Gel (500–10000 bp)	0.3 g	30 mL
1.5 % Gel (200-3000 bp)	0.45 g	30 mL
2 % Gel (50-2000 bp)	0.7 g	35 mL

#### 2.1.2.2.2.3 Bioanalyzer analysis

To verify the size distribution and quality of fragmented DNA after the fragmentation and purification step in the Goettingen protocol or after the PCR-purification step of the Nextera XT DNA Library Preparation Kit the Agilent High Sensitivity DNA Kit was used according to manufacturer's protocol (Agilent Technologies).

The gel matrix was prepared by adding 15 µL of the dye into the gel. The mix was vortexed and loaded on a filter membrane to centrifuge at 2250x g for 10 minutes.

The DNA chip was prepared in the priming station by loading 9 µL of the gel-dye mix into the distribution well. Further 9 µL of the gel-dye mix were added to other gel wells. After that 5 µL DNA marker were added in every sample well as well as the DNA ladder well. One µL of the ladder was loaded into the ladder well. After pipetting 1 µL of sample into sample wells, the chip was vortexed at 2500 rpm for 1 min and ran in the Bioanalyzer. Results were processed using the corresponding software.

#### 2.1.2.2.2.4 Qubit analysis

A Qubit dsDNA HS Assay Kit (Life Technologies) was used for an exact concentration measurement of ds-cDNA on a Qubit 2.0 Fluorometer (Life Technologies). The DNA

concentration was first measured via NanoDrop and the results were used to check if the DNA concentration was within the accurate working ratio from 10 pg/μL to 100 ng/μL of the used assay. The Qubit working solution and the standards were prepared according to the manufacturer's protocol. The samples were prepared by mixing 1 μL of the DNA with 199 μL of the Qubit working solution. The reading in ng/mL was calculated with the following formula to get the exact concentration: concentration of the sample in ng/μL = (concentration in ng/mL x 200/ X μL of sample input)/1000 mL. The concentration provided by the Qubit analysis was used to dilute samples more precisely for the Nextera XT DNA Library preparation, since the exact DNA input is very crucial to the outcome of the library preparation.

#### 2.1.2.3 Library preparation

For whole genome sequencing of IPNV samples processed with Goettingen or Glasgow protocol NGS libraries were prepared according to the Nextera XT DNA Library Preparation Kit (Illumina).

##### 2.1.2.3.1 Tagmentation

For the tagmentation 2.5 μL DNA (0.5 ng of total DNA) were mixed with 5 μL TD (Tagment DNA Buffer) and 2.5 μL AMT (Amplicon Tagment Mix) in a 0.2 mL PCR tube. The mix was centrifuged at 280x g and 20 °C for 1 min and incubated for 5 min at 55 °C in a thermocycler. The mix was cooled down to 10 °C and the tagmentation reaction was neutralized by pipetting 2.5 μL of NT (Neutralize Tagment Buffer) into the sample mix. The mix was centrifuged at 280x g and 20 °C for 1 min and incubated for 5 min at room temperature.

##### 2.1.2.3.2 PCR amplification

The PCR mix was set up by pipetting 7.5 μL NPM (Nextera PCR Mastermix) and 2.5 μL of index primer 1 and 2 of the Nextera XT Index Kit (for 24 libraries) to the sample mix from the tagmentation step. The mix was centrifuged at 280x g and 20 °C for 1 min and cycled as followed (Table 21):

**Table 21. PCR amplification cycling conditions.**

72 °C	3 min	
95 °C	30 sec	
95 °C	10 sec	
55 °C	30 sec	12 cycles
72 °C	30 sec	
72 °C	5 min	
10 °C	hold	

The PRC reaction was processed immediately or stored at 4-8 °C for up to 2 days.

#### 2.1.2.3.3 PCR Clean-Up

The PCR tubes were pelleted at approximately 2000 rpm and the samples were transferred into a 1.5 mL tube. Forty-five µL of AMPure XP beads were added to each sample (1.8x) to select for 300-500 bp long sample fragments and incubated for 5 min at room temperature. The tubes were placed into a magnetic rack and incubated for minimum 2 min until the supernatant had cleared. The supernatant was discarded. The samples (still in the magnetic rack) were washed twice with freshly prepared 80 % ethanol for 30 sec. The samples were left on the magnetic rack for 15 min to air-dry the beads. The tubes were removed from the magnetic rack to elute the libraries in 26 µL RSB (Resuspension Buffer). The elution mix was incubated for 2 min at room temperature and the tubes were placed back on the magnetic rack to separate the beads from the eluted libraries. After the supernatant has cleared, 23 µL of the supernatant were transferred into a new 1.5 mL tube. The samples were processed immediately or stored at -15-25 °C for up to a week.

#### 2.1.2.3.4 Library Normalisation

To normalise the quantity of each library 45.8 µL/sample of LNA1 and 8.3 µL/sample of LNB1 were mixed together and 45 µL of the LNA1/LNB1 mix were added to each sample and incubated for 30 min on a mixer at 99 rpm. The samples were placed into a magnetic rack and incubated until the supernatant has cleared. The supernatant was removed and the samples were washed with 45 µL LNW1 by incubating the tubes at 99 rpm for 5 min on a mixer. Then the tubes were placed back in the magnetic rack. The clear supernatant was removed and the

samples were washed as described one more time. The tubes were removed from the magnetic rack and 30  $\mu$ L of freshly prepared 0.1 N NaOH were added to each sample. The elution mix was incubated for 5 min at 99 rpm on a mixer. During the incubation step 30  $\mu$ L of LNS1 were pipetted into a fresh 1.5 mL tube. The samples were placed in the magnetic rack after the incubation and 30  $\mu$ L of the clear supernatant were transferred to the 1.5 mL tube, which contains 30  $\mu$ L LNS1. The eluted library was pelleted at approximately 2000 rpm and processed immediately or stored at -15-25  $^{\circ}$ C.

#### 2.1.2.3.5 Library Pooling and loading

Each library was mixed and pelleted at approximately 2000 rpm. Then 5  $\mu$ L of each library were pooled into a 1.5 mL tube. Another 1.5 mL tube was used to mix 570  $\mu$ L of HT1 and 12  $\mu$ L of PhiX Control v3. Eighteen  $\mu$ L of the pooled library were added to this mix. Right before loading the mix into the MiSeq reagent cartridge of the MiSeq Reagent Kit v3 (150 cycles) the mix was pelleted at approximately 2000 rpm and incubated at 96  $^{\circ}$ C for 2 min in a heat block. Then the library mix was cooled for 5 min in an ice-ethanol mix and loaded into the MiSeq reagent cartridge, which was previously thawed at room temperature. The rest of the pooled library mix was stored at -15-25  $^{\circ}$ C for up to a week. The MiSeq reagent cartridge was placed into the previously maintained MiSeq sequencer (Illumina) and a paired end sequencing run was performed according to the manufacturer's protocol.







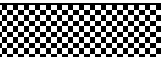



## 2.2 Improving the Glasgow protocol

### 2.2.1 Improving the ds-cDNA synthesis

#### 2.2.1.1 Introduction

In order to improve the coverage of the 3'-end of IPNV segment A and B some changes were introduced to the Glasgow protocol using the sample IoA039:

**Table 22. Changes in the GLAP compared to standard GLAP.**

<b>IoA039 Sample variation</b>	<b>GP32 protein</b>	<b>3' primer</b>	<b>RNase treatment</b>	<b>Elute RNA in dNTPs and random hexamers</b>	<b>Incubate RNA at 75 °C for 5 min</b>	<b>Normal Glasgow protocol</b>
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						

GP32 is a ssDNA binding protein and it was used to improve the yield and efficiency of RT during RT-PCR (Shamoo et al. 1995). In this experiment 1 µg/25µL reaction volume of GP32 (Roche) were used to improve amplification. For that, GB32 was included in the ss-cDNA step (Table 22 1) or the ds-cDNA step (Table 22 2) or in both steps (Table 22 3).

To improve termini coverage a primer was created that is as long as the random primers/oligos from the respective kits, within 30-100 bp of the 3'-end. The primer was used additionally to the random primers/oligos from the respective kit in the first cDNA strand



synthesis (Table 22 4) or second cDNA strand synthesis (Table 22 5) or in both reactions (Table 22 6).

The 5'-3' sequence of the primer is complementary to the 3'-end of the 3'-5' sequence and will initiate a transcription to the end. The primers for segment A and B were created by creating a contig sequence of multiple IPNV sequences available (Table 23).

**Table 23. Primer for IPNV segment A and B.**

	<b>5' – 3' sequence</b>	<b>Tm</b>	<b>Position</b>
IPNV A	CGACGACCCCG	45 °C	2933-2943
IPNV B	GCTCCR(A or G)CGCC	46 °C	2694-2703

In this optimisation an RNase treatment after the first strand synthesis was performed for 20 min at 37 °C (Table 22 7).

Another optimisation step was to elute the extracted IPNV RNA in a mix of water, random hexamers and dNTPs (Table 22 8).

Further the IPNV RNA was incubated at 75 °C for 5 min prior to the first strand cDNA synthesis instead of 65 °C as in the original protocol (Table 22 9) as higher incubation temperature allows a better unfolding of potential complex secondary structures at the genomic ends.

### 2.2.1.2 Setting up the experiment

#### 2.2.1.2.1 GP32

IPNV RNA was extracted from 400 µL cell supernatant as in Glasgow protocol step 2.1.2.1.2.1 (RNAAdvance Blood Kit) and eluted in 20 µL water. The RNA was stored at -70 °C if necessary or processed immediately. IPNV RNA concentration was measured using the NanoDrop and the in-house IPNV RT-qPCR.

Three different reactions were performed by adding GP32 (final concentration: 0.28 mg/mL) only in the ss-cDNA synthesis, only in the ds-cDNA synthesis and in both steps. The ds-cDNA synthesis was performed according to the Glasgow protocol.

The reactions were cleaned up as described in step 2.1.2.1.2.2.3 and stored at -20 °C until next procedure. Ds-cDNA was analysed on a High-Sensitivity DNA Bioanalyzer Chip (concentration, fragment size, molarity) and compared with others. Sequencing libraries were prepared

according to the Illumina Nextera XT DNA Library Preparation Kit and sequenced according to the MiSeq Reagent Kit (v3; 150 cycles).

#### 2.2.1.2.2 Specific 3' - Primer

IPNV RNA was extracted from 400  $\mu$ L cell supernatant as in Glasgow protocol step 2.1.2.1.2.1 (RNAAdvance Blood Kit) and eluted in 20  $\mu$ L water. The RNA was stored at -70 °C if necessary or processed immediately. IPNV RNA concentration was measured using the NanoDrop and the in-house IPNV RT-qPCR.

Three different reactions were performed by adding specific 3'-primer only in the ss-cDNA synthesis, only in the ds-cDNA synthesis and in both steps. The additional primer had a final concentration in the master mix of 2.38  $\mu$ M (same as the random hexamers in the first cDNA strand synthesis step). The ds-cDNA synthesis was performed according to the Glasgow protocol.

The reactions were cleaned up as described in step 2.1.2.1.2.2.3 and stored at -20 °C until next procedure. Ds-cDNA was analysed on a High-Sensitivity DNA Bioanalyzer Chip (concentration, fragment size, molarity) and compared with others. Sequencing libraries were prepared according to the Illumina Nextera XT DNA Library Preparation Kit and sequenced according to the MiSeq Reagent Kit (v3; 150 cycles).

#### 2.2.1.2.3 changes in RNA incubation

IPNV RNA was extracted from 400  $\mu$ L cell supernatant as in Glasgow protocol step 2.1.2.1.2.1 (RNAAdvance Blood Kit) and eluted in 20  $\mu$ L water. The RNA was stored at -70 °C if necessary or processed immediately. IPNV RNA concentration was measured using the NanoDrop and the in-house IPNV RT-qPCR. Following the Glasgow protocol as usual, the RNA was incubated at 75 °C for 5 min instead of 65 °C.

The reactions were cleaned up as described in step 2.1.2.1.2.2.3 and stored at -20 °C until next procedure. Ds-cDNA was analysed on a High-Sensitivity DNA Bioanalyzer Chip (concentration, fragment size, molarity) and compared with others. Sequencing libraries were prepared according to the Illumina Nextera XT DNA Library Preparation Kit and sequenced according to the MiSeq Reagent Kit (v3; 150 cycles).

#### 2.2.1.2.4 Eluting RNA in a 'premix'

IPNV RNA was extracted from 400  $\mu$ L cell supernatant as in Glasgow protocol step 2.1.2.1.2.1 (RNAAdvance Blood Kit) and eluted in 13  $\mu$ L Premix (11  $\mu$ L water + 1  $\mu$ L dNTPs + 1  $\mu$ L random hexamers) The RNA was stored at -70 °C if necessary or processed immediately according to the Glasgow protocol.

The reactions were cleaned up as described in step 2.1.2.1.2.2.3 and stored at -20 °C until next procedure. Ds-cDNA was analysed on a High-Sensitivity DNA Bioanalyzer Chip (concentration, fragment size, molarity) and compared with others. Sequencing libraries were prepared according to the Illumina Nextera XT DNA Library Preparation Kit and sequenced according to the MiSeq Reagent Kit (v3; 150 cycles).

#### 2.2.1.2.5 RNase treatment

IPNV RNA was extracted from 400  $\mu$ L cell supernatant as in Glasgow protocol step 2.1.2.1.2.1 (RNAAdvance Blood Kit) and eluted in 20  $\mu$ L water. The RNA was stored at -70 °C if necessary or processed immediately. IPNV RNA concentration was measured using the NanoDrop and the in-house IPNV RT-qPCR. After the ss-cDNA synthesis step the RNA was removed incubating the mix with *E. coli* RNaseH (Life Technologies) for 20 min at 37 °C. The rest was performed according to the standard Glasgow protocol.

The reactions were cleaned up as described in step 2.1.2.1.2.2.3 and stored at -20°C until next procedure. Ds-cDNA was checked on a High-Sensitivity DNA Bioanalyzer Chip and compared with others. Sequencing libraries were prepared according to the Illumina Nextera XT DNA Library Preparation Kit and sequenced according to the MiSeq Reagent Kit (v3; 150 cycles).

#### 2.2.2 Improving RNA procedure

In order to further improve the termini coverage if the IPNV genome samples IoA035 and IoA111 were randomly selected to perform changes in the RNA incubation step of the Glasgow protocol.

IPNV RNA was extracted from 400  $\mu$ L cell supernatant as in Glasgow protocol step 2.1.2.1.2.1 (RNAAdvance Blood Kit) and eluted in 20  $\mu$ L water. The RNA was stored at -70 °C if necessary or processed immediately. IPNV RNA concentration was measured using the NanoDrop and the in-house IPNV RT-qPCR.

Prior ss-cDNA synthesis step RNA of each sample was incubated 5min at 95 °C and at 65 °C as in the standard Glasgow protocol. Further steps were performed according to the protocol. The reactions were cleaned up as described in step 2.1.2.1.2.2.3 and stored at -20 °C until next procedure. Ds-cDNA was analysed on a High-Sensitivity DNA Bioanalyzer Chip (concentration, fragment size, molarity) and compared with others. Sequencing libraries were prepared according to the Illumina Nextera XT DNA Library Preparation Kit and sequenced according to the MiSeq Reagent Kit (v3; 150 cycles).

## 2.3 Persistency experiment

### 2.3.1 Introduction

The whole genome sequencing analysis of several IPNV samples revealed a large amount of persistent strains, which have the amino acid Proline at position 217 and a Threonine at position 221 in the VP2 protein.

It is known that cell culture adaptations lead to attenuation of the virus to replicate more efficiently in the cell line (Santi, Vakharia, and Evensen 2004b, Song et al. 2005). In order to estimate whether the persistence pattern observed in the analysed IPNV strains was natural or due to cell culture adaptation, IPNV isolates IoA068, 039, 090, 076 and 035 known to have the persistence pattern were passaged at least 3 times in CHSE-214 and RTG-2 cells (Officer 1964), since it was shown that even after 10 passages in RTG-2 cells, there are no changes in the amino acid (AA) sequence of the VP2 protein which makes RTG-2 cells a useful cell line for virulence/persistence analysis of IPNV (Song et al. 2005).

### 2.3.2 Methods

#### 2.3.2.1 RTG-2 cell line maintenance

Cells were processed in the same manner as described in 2.1.1.1 but with several changes:

- L-15 (Leibovitz's L-15 medium; Gibco)+10 % FBS media instead
- split ratio was 1:2
- cells were incubated at 22 °C and no CO<sub>2</sub> with a fully closed lid.

#### 2.3.2.2 Seeding density estimation of RTG-2 cells for a 24 well plate (WP)

In order to have an optimal cell density in each well of a 24WP, several seeding densities were tested. For that cells were processed as in 2.1.1.1.5 but with L-15+10 % FBS. A cell calculation was performed for several densities:  $5 \times 10^4$ ,  $7.5 \times 10^4$ ,  $1.0 \times 10^5$ ,  $1.5 \times 10^5$ ,  $1.75 \times 10^5$ , and  $2.0 \times 10^5$  cells/well. One mL of each calculated cell dilution was pipette in 4 wells of each of the 6 rows. The plate was incubated at 22 °C and no CO<sub>2</sub>. The next day the plate was examined visually to ensure that each well of the six different seeding densities had approximately the same cell density and the cells looked healthy. The cell density with the best suitable monolayer was marked and the plate was incubated over night at 15 °C and no CO<sub>2</sub> which are the incubation requirements for IPNV inoculation. The next day the plate was again examined visually, especially the marked row with the best suitable seeding density. If the seeding

density of interest fulfilled the requirements, it was used for initial inoculation with IPNV in a 24WP.

#### 2.3.2.3 Seeding a 24WP of RTG-2 cells for inoculation via adsorption

Cells were processed as described in 2.3.2.1 and 2.1.1.1.5. The cell number per well that was needed for RTG-2 cells is  $1.0 \times 10^5$  cells/well. Cell dilution was pipetted in each well. The plate was incubated at 22 °C and no CO<sub>2</sub>. The next day the plate was examined visually to ensure that each well had approximately the same cell density and the cells looked healthy. If the requirements were fulfilled, the well plate was used for inoculation with IPNV samples.

#### 2.3.2.4 Infecting a RTG-2 24WP with IPNV

Isolates (TCID<sub>50</sub>) IaA068, 039, 090, 076 and 035 that were shown to belong to the persistent strain type were randomly selected for inoculation via adsorption of RTG-2 cells in a 24WP.

IPNV isolates of interest were defrosted and kept on ice. 1:10 and 1:50 dilutions were prepared using HBSS+2 % FBS media as diluent. Media of a 24WP that was prepared on the previous day was pipetted off and 150 µL of the neat sample, the 1:10 and the 1:50 dilutions were pipetted in a column in following order: neat, 1:10, 1:50, negative control (NC= HBSS+2 % FBS). The plate was sealed with plastic paraffin film and incubated at 15 °C w/o CO<sub>2</sub> for 3 h. The plate was rocked every hour to ensure that the cells were equally covered with the inoculum. Then 1.5 mL of L15+10 % FBS media were added to each well and the plate was incubated at 15 °C w/o CO<sub>2</sub>. Every day the plate was examined for a cytopathic effect (CPE).

#### 2.3.2.5 Passage of positive well(s) into a 25 cm<sup>2</sup> TC

When a full CPE was visible in a well, or in several wells of the same isolate, the supernatant of the specific well was collected or multiple wells were pooled into a centrifuge tube and centrifuged at 2500x g and 4 °C for 15 min. The supernatant was decanted in a polystyrene container and filtered through a 0.2 µm filter into a new polystyrene container. A dilution of 10<sup>-2</sup> up to 10<sup>-4</sup> (depending on the virulence of the isolate) was prepared on ice by using HBSS+2 % FBS media. Old media of a 25 cm<sup>2</sup> TC was decanted off and 1 mL of the diluted supernatant was added into the flask. For a NC, another 25 cm<sup>2</sup> TC was incubated with 1mL of the diluent (HBSS+2 % FBS). The flasks were incubated for 1h at 15 °C w/o CO<sub>2</sub>. After incubation, the inoculum was discarded and 5mL L15+10 % FBS media were added to the

sample and the NC flask. The flasks were incubated at 15 °C w/o CO<sub>2</sub> and examined every day for a CPE.

*2.3.2.6 Harvest IPNV supernatant of a 25 cm<sup>2</sup> TC for RNA extraction*

The culture medium of infected cells was harvested when the cells showed 80-90 % CPE as previously described in 2.1.1.2.4 (second paragraph).

*2.3.2.7 Processing IPNV RNA, ds-cDNA synthesis and sequencing*

The RNA was processed via the Glasgow protocol and sequenced according to the Illumina Nextera XT DNA Library Preparation Kit (see 2.1.2.1.2 and 2.1.2.3).

## 2.4 MinION sequencing

### 2.4.1 Introduction

In order to improve termini coverage of the IPNV samples beside changes in the ds-cDNA protocol, it was decided to try to sequence on the MinION device as it has a different NGS technology, which could yield in better termini coverage as it was previously seen for whole genome sequencing of Influenza A (H3N2) virus by (Wang et al. 2015).

### 2.4.2 Methodology

#### 2.4.2.1 DNA input

The ds-cDNA of one randomly selected IPNV sample IoA039 was used as for the MiSeq sequencing. The MinION 2D cDNA R9 version protocol using the SQK-NSK007 R9 version (v.R9; Oxford Nanopore Technologies) sequencing Kit required 1 µg cDNA as input. As viral RNA recovered from cell culture supernatant could not reach such yields, the maximum volume of ds-cDNA was used as input.

#### 2.4.2.2 End-prep

Forty-five µL of ds-cDNA were mixed with 15µL of the following master mix and pelleted at approximately 2000 rpm (Table 24):

**Table 24. End-prep set-up . All reagents are included in the NEBNext Ultra II End-Prep kit.**

<b>Reagent</b>	<b>Volume</b>
Ultra II End-prep buffer	7 µL
Ultra II End-prep enzyme mix	3 µL
Water	5 µL

The mix was incubated at 20 °C for 5 min and 65 °C for 5 min in a thermal cycler and cleaned up using 60 µL of AMPure XP beads (1x). After adding the beads to the reaction, it was incubated on a mixer at 70 rpm on the UU setting and pelleted at approximately 2000 rpm before placed into a magnetic rack. The beads were washed while remaining in the magnet twice with 200 µL 70 % ethanol. To pipette off any residual ethanol the tube was removed from the magnet to spin the content down. The beads were briefly dried and the end-prepped



DNA was eluted in 16  $\mu\text{L}$  water outside the magnet. The mix was incubated for 2 min at room temperature. The tube was placed back into the magnet and the clear supernatant with the DNA was transferred into a fresh tube.

#### 2.4.2.3 Ligation of PCR adapters and amplification

The mix was set up as followed (Table 25):

**Table 25. Adaptor ligation set-up. All reagents are included in the NEB Blunt/TA kit or the SQK-NSK007 (v.R9) kit.**

Reagent	Volume
dA-tailed DNA	15 $\mu\text{L}$
PCR adapters	5 $\mu\text{L}$
Blunt/TA Ligase Master Mix	20 $\mu\text{L}$

The reaction was mixed and incubated for 15 min at room temperature and cleaned up using 28  $\mu\text{L}$  of AMPure XP beads. After adding the beads to the reaction, it was incubated on a mixer at 70 rpm on the UU setting and pelleted at approximately 2000 rpm before placed into a magnetic rack. The beads were washed while remaining in the magnet twice with 200  $\mu\text{L}$  70 % ethanol. To pipette off any residual ethanol the tube was removed from the magnet to spin the content down. The beads were briefly dried and the end-prepped DNA was eluted in 25  $\mu\text{L}$  water outside the magnet. The mix was incubated for 2 min at room temperature. The tube was placed back into the magnet and the clear supernatant with the DNA was transferred into a fresh 0.2 mL PCR tube for the PCR, which was set up as followed (Table 26):

**Table 26. PCR mix set-up of ligated sample. All reagents are included in the LongAmp Taq 2x kit or the SQK-NSK007 (v.R9) kit.**

Reagent	Volume
Adapter-ligated library	25 $\mu\text{L}$
PRM	2 $\mu\text{L}$
LongAmp Taq 2x Master mix	50 $\mu\text{L}$
Water	23 $\mu\text{L}$

The reaction was mixed and cycled as followed (Table 27):

**Table 27. PCR cycling conditions.**

Step	Temperature	Time	Cycle
Initial denaturation	95 °C	3 min	1
Denaturation	95 °C	15 sec	
Annealing	62 °C	15 sec	18 cycles
Extension	72 °C	3 min	
Final extension	65 °C	10 min	1
Hold	4 °C	hold	

The reaction was then cleaned up using 1.8x reaction volume in a 1.5 mL centrifuge tube. The beads were mixed with the PCR reaction and incubated for 5 minutes at room temperature. The mix was placed in a tube magnet rack and incubated for 2 minutes until the supernatant was clear, which was then discarded. The beads were washed twice with 200 µL of 70 % ethanol without disturbing the beads. The beads were air dried and re-suspended in 46 µL molecular grade water. The elution reaction was incubated 2 minutes at room temperature outside the magnetic rack. After this the tube was placed back in the magnetic rack and the clear supernatant was transferred into a new centrifuge tube. One µL of the clean DNA was used for concentration measurement in the Qubit fluorometer (as described in 2.1.2.2.4).

#### 2.4.2.4 End-prep of post-PCR DNA

The post-PCR DNA was transferred into a fresh 0.2 mL PCR tube containing the following reaction (Table 28):

**Table 28. End-prep set-up after the PCR. All reagents are included in the NEBNext End-Repair/dA-tailing kit.**

<b>Reagent</b>	<b>Volume</b>
DNA	45 $\mu$ L
Ultra II End-prep reaction buffer	7 $\mu$ L
Ultra II End-prep enzyme mix	3 $\mu$ L
Water	5 $\mu$ L
<b><math>\Sigma</math> Volume</b>	<b>60 <math>\mu</math>L</b>

The mix was incubated at 20 °C for 5 min and 65 °C for 5 min in a thermal cycler and cleaned up using 60  $\mu$ L of AMPure XP beads. After adding the beads to the reaction, it was incubated on a mixer at 70 rpm on the UU setting and pelleted at approximately 2000 rpm before placed into a magnetic rack. The beads were washed while remaining in the magnet twice with 200  $\mu$ L 70 % ethanol. To pipette off any residual ethanol the tube was removed from the magnet to spin the content down. The beads were briefly dried and the end-prepped DNA was eluted in 31  $\mu$ L water outside the magnet. The mix was incubated for 2 min at room temperature. The tube was placed back into the magnet and the clear supernatant with the DNA was transferred into a fresh tube. One  $\mu$ L of the clean DNA was used for concentration measurement in the Qubit.

#### 2.4.2.5 Adapter ligation

The DNA was diluted to 500 ng in 22.5  $\mu$ L for the reaction, which was set up as followed (Table 29):

**Table 29. Ligation set-up of the hairpin adapter. All reagents are included in the NEB Blunt/TA Ligase kit or the SQK-NSK007 (v.R9) kit.**

<b>Reagent</b>	<b>Volume</b>
700 ng of DNA	30 $\mu$ L
Water	88 $\mu$ L
Adapter Mix	10 $\mu$ L
HP adapter	2 $\mu$ L
Blunt/TA Ligase master mix	50 $\mu$ L
<b><math>\Sigma</math> Volume</b>	<b>100<math>\mu</math>L</b>

The reaction was mixed, pelleted at approximately 2000rpm and incubated at room temperature for 10 minutes. After that 1  $\mu$ L of the HP Tether was added to the mix, mixed, pelleted at approximately 2000 rpm and incubated at room temperature for 10 minutes.

#### 2.4.2.6 Preparation of MyOne C1 beads

The beads (Life Technologies) were mixed and 50  $\mu$ L of the beads were transferred into a new tube. The tube was placed into a magnetic rack and the clear supernatant was discarded. Then 100  $\mu$ L of the bead binding buffer was added to the beads. The beads were mixed and placed into the magnet. The clear supernatant was discarded and the wash was repeated. In the end, the beads were re-suspended in 100  $\mu$ L of the bead binding buffer.

#### 2.4.2.7 Library purification and elution

One hundred  $\mu$ L of previously prepared MyOne C1 beads were added to the tethered sample, mixed and incubated on a mixer for 5 minutes at 50 rpm. The tube was then placed into a magnetic rack and the clear supernatant was discarded. The beads were washed twice by re-suspending them in 150  $\mu$ L Bead Binding Buffer outside the magnetic rack, placing the tube back into the rack and taking the clear supernatant off. To get rid of any residual liquid, the tube was centrifuged, placed back in the magnet and any possible liquid was discarded. The beads were then re-suspended in 25  $\mu$ L of Elution buffer, mixed and incubated at 37 °C for 10 minutes. The tube was then placed in a magnet and the clear supernatant was transferred into a new tube. One  $\mu$ L was used for quantification in a Qubit fluorometer (as described in 2.1.2.2.2.4).

#### 2.4.2.8 Priming the Flow Cell and loading the library

Before priming the flow cell, a QC had to be performed. For that the flow cell FLO-MIN104 R9 version was placed into the MinION sequencer and the QC option NC\_Platform\_QC\_FLO\_MIN104 was selected within the MinKNOW software.

To prime the flow cell, the sample port was opened and the flow cell was checked for bubbles. To remove any bubbles a small amount of liquid was aspirated and discarded. The flow cell priming mix was prepared as followed (Table 30):

**Table 30. Preparation of the priming mix. All reagents are included in the SQK-NSK007 (v.R9) kit.**

Reagent	Volume
RBF1	500 $\mu$ L
Water	500 $\mu$ L
Total	1000 $\mu$ L

First 500  $\mu$ L of the mix were loaded into the flow cell through the sample port. After an incubation of 5 minutes procedure was repeated.

The library loading mix was set up as followed (Table 31):

**Table 31. Preparation of the library mix to load on the flow cell. All reagents are included in the SQK-NSK007 (v.R9) kit.**

Reagent	Volume
RBF1	75 $\mu$ L
Water	63 $\mu$ L
Adapted and tethered library	12 $\mu$ L
Total	150 $\mu$ L

The library mix was mixed and pelleted at approximately 2000 rpm before it was loaded into the flow cell drop by drop. After loading the sample port was closed and the set up NC\_48H\_Seq\_Run\_FLO\_MIN\_104 was run.

#### 2.4.2.9 Washing the flow cell for immediate reuse

After 24 h the run was stopped and the flow cell was washed using the Wash Kit (Oxford Nanopore Technologies) to be ready for loading the second half of the sample. The sample port cover was opened and 150  $\mu$ L of the Solution A were added through the sample port. After an incubation of 10 minutes 150  $\mu$ L of Solution B were added through the sample port and the library was added as described in previous steps. No platform QC and following priming were performed when running another library immediately.

#### 2.4.2.10 Washing and storage of the flow cell

After 24 h the run was stopped and the flow cell was washed to be ready for loading the second half of the sample. The sample port cover was opened and 150 µL of the Solution A were added through the sample port. After an incubation of 10 minutes 500 µL of Storage buffer were added through the sample port. The sample port was closed and all buffer was removed from the waste compartments through the waste port. The flow cell was stored at 4 °C for up to 72 h. A platform QC and following priming were performed before running another library as described before.

#### 2.4.3 Improving size selection for MinION sequencing

As the MinION is capable of sequencing very large fragments, it was of interest to see if a change in the sample to AmpureXP beads ratio from 1x to 0.6x can select for fragments longer than 500 bp. This was done by testing it on one randomly selected betanodavirus (provided by Dr. Benjamin Lopez-Jimena) ds-cDNA sample that was processed according to the Glasgow protocol, using 1x AmpureXP beads to sample ratio in the last clean up step.

Twenty µL of betanodavirus ds-cDNA were mixed with 12 µL AmpureXP beads and processed as in the post ds-cDNA synthesis clean up step in the Glasgow protocol 0.

## 2.5 Bioinformatics analysis

### 2.5.1 Processing the raw data from MiSeq (Illumina)

#### 2.5.1.1 Overview

A successfully assembled genome results in 2 big segments (both full) and some little ones. If the assembly resulted in several semi-big segments, they have to be put together manually using the reference genome for orientation. The assembled sequence is then aligned to the reference genome to check for any problems. If the sequence is fine, no changes are needed. If the sequence turned out bad, the pipeline-protocol has to be optimised: increased check for contaminations, trim adaptors/indices, decrease coverage of the reads. The decontamination is very important and crucial as cleaner reads give better assembly results. The ends have to be optimised manually by assembling first/last 300-400 bp using the reference-based approach.

#### 2.5.1.2 Workflow for Illumina MiSeq raw data processing

The workflow for processing the raw data through to the assembled genome in a FastA file was developed by Dr. Stefanie Wehner, containing multiple steps of data processing. The detailed workflow is given in the separate additional appendix under V.1.

The MiSeq generates per cycle base call (bcl) files as primary files. The files are combined and converted into FASTQ by BCL2FASTQ (Illumina, Inc.) by the MiSeq. This software is able to perform the file conversion and demultiplexing at the same time using the sample sheet information providing index – sample information. During this step one error in the index can be detected and corrected.

The quality of the raw reads was then checked using FASTQC. The report outcome was to ignore as this tool was not directly used for filtering but it provides a first simple overview of the data, gives information about the sequence quality and shows overrepresented sequences such as contaminations or adapter sequences.

After that the reads were trimmed and clipped using PRINSEQ and TRIMMOMATIC to remove low quality reads and eventually not removed adapter sequences.

To remove reads belonging to the host organism (fish), contamination (human, bacteria) and the Illumina PhiX control genome, DECONSEQ was used. In this step it was crucial to watch out for possible “hidden” virus sequences such as those that are embedded into a cell line etc. which should then also be removed.

In case of short genomes, a faster and better assembly can be achieved by down sampling the reads with KHMER (Crusoe et al. 2015). In this step redundant short reads are eliminated, while the individual reads are not modified.

After that the reads were used for a *de novo* assembly via SPADES. In parallel, reference-based assemblies were created using TANOTI and STAMPY with NC\_001915.1 (segment A) and NC\_001916.1 (segment B) as reference sequences. The resulting sequences were compared with the reference sequence using BLAST (Altschul et al. 1990) to find the most complete genome. *De novo* assembled sequences were preferred in this approach as it takes into account potential insertions, deletions and rearrangement of the genome. The reference-based assembly was used to eventually cover the *de novo* sequences at the genome termini.

BOWTIE2 (Langmead and Salzberg 2012) and SAMTOOLS (Li et al. 2009) were used to map the MiSeq reads to the assembled genome to get the number of mapped reads for the coverage plots. They show how many reads map per base of the genome. Mean coverage was calculated as well and refers to the mean coverage of all mean coverages. So the mean coverage of each assembled segment was calculated. Based on these values an overall mean coverage was calculated. Whereas for minimum/maximum mean coverage, for every assembled segment the minimal/maximal coverage was taken and then again a mean was calculated (based on the minimum/maximum values).

For single protein analysis and CAI Dr. Michaël Bekaert aligned the sequences to the reference genome using CLUSTAL OMEGA (Goujon et al. 2010, Sievers et al. 2011), transferred the genome annotation from the reference sequence using RATT (Otto et al. 2011) and extracted the sequences for each aligned protein in a FastA file. The detailed workflow is given in the separate additional appendix under V.3.

#### 2.5.2 Processing raw data from MinION (Oxford Nanopore Technologies)

The workflow for processing the raw data up until the assembled genome in a FastA file was done by Dr. Stefanie Wehner. The detailed workflow is given in the separate additional appendix under V.2.

The reads were first processed by the corresponding software METRICHOR. NANOOK was used to create a run report with basic information about the run. To convert MinION filing system FAST5 into FastA PORETOOLS was used. PORECHOP was used additionally to eventually trim adaptors and indices as there were sometimes issues observed such as



adaptor bleed over and concatemer formations. The reads were aligned with BWA to generate a reference-based genome assembly. The coverage plots for the MinION runs were generated using a script in R, which based on the NANOOK script.

## 2.6 Phylogenetic analysis

### 2.6.1 General handling of sequences

CLUSTAL Omega was used on an online platform <https://www.ebi.ac.uk/Tools/msa/clustalo/> (Goujon et al. 2010, Sievers et al. 2011) to align IPNV sequences.

For conversion of sequence file formats, <http://sequenceconversion.bugaco.com/converter/biology/sequences/index.html> was used.

### 2.6.2 MEGA

MEGA v6.0.6 (Molecular Evolutionary Genetics Analysis) (Tamura et al. 2013) was used for work with alignment files and for eventual manual editing. It was used to create amino acid alignments overviews of the IPNV sequences to visually estimate differences.

### 2.6.3 BEAST package

To calculate phylogenetic properties, sequences of segment A and segment B of 57 IPNV isolates were put together in one FastA file with additional reference genomes (Table 32, Table 33). BEAUTI v1.8.3 was then used to generate an xml file with 10,000,000 chains per cycle with every 1000<sup>th</sup> chain being recorded, using the HKY substitution model, a strict clock and a coalescent tree according to (Kingman 1982a, Kingman 1982b). The xml file was loaded in BEAST v1.8.3 to perform the calculation of the tree (Drummond et al. 2012). The trees were viewed and edited using FIGTREE v1.4.0 (Rambaut 2012).

**Table 32. Reference strains used for phylogenetic analysis of segment A; NA means not assigned.**

Strain	GenBank	Geographical Origin	Reference
Ca3	AF342734	Canada	(Blake et al. 2001)
Connecticut1	JF440810	USA	(Glenney et al. 2012)
Ca2	AF342733	Canada	(Blake et al. 2001)
TABV2013	KP268663.1	Australia	(Mohr et al. 2015)
TABV1998	NC_028252.1	Australia	(Mohr et al. 2015)
1146	AJ489222.1	Spain	(Cutrin et al. 2004)
Sp	KF279643	Iran	(Dadar et al. 2013)
Gre	NA	Turkey	Unpublished

SDF4	NA	Turkey	Unpublished
1054	NA	Turkey	Unpublished
Sp	AF342728	Denmark	(Blake et al. 2001)
Ireland	HQ457197.1	Ireland	(Mutoloki and Evensen 2011)
DPL	AY026485.1	Thailand	(Blake et al. 2001)
Fr21	AY026483.1	France	(Blake et al. 2001)
OV2	AY026484.1	England, UK	Blake et al., 2001 (Blake et al. 2001)
N1	D00701	Norway	(Havarstein et al. 1990)
Chile	HQ457169	Chile	(Mutoloki and Evensen 2011)
Norway	HQ457199	Norway	(Mutoloki and Evensen 2011)
Sp vir.	AY374435.1	Norway	(Santi, Vakharia, and Evensen 2004a)
Ab	AF342729	Denmark	(Blake et al. 2001)
2284	AJ489223	Spain	(Cutrin et al. 2004)
Ab	L40580.1	Denmark	(Heppell et al. 1993)
CVHB1	AY026489.1	Taiwan	(Blake et al. 2001)
S	AY026487	Taiwan	
V	AY026486.1	Japan	
PV	AY026488.1	Taiwan	
Ca1	AF342732	Canada	
ASV	AY026490.1	Canada	
Te2	AF342731	England, UK	
AY98	AY283785.1	Japan	(Zhang and Suzuki 2004)
YTAV_Y6	AY283781	Japan	
GC1	AY064396.1	Korea	unpublished
Ja	AF342735.1	Canada	(Blake et al. 2001)
Reno	AY026345	USA	
Ja2310	AJ489225	Spain	(Cutrin et al. 2004)
HL1	D26526.1	Korea	(Chung et al. 1994)
Ja	NC_001915.1	Canada	(Duncan et al. 1987)
VR299	AF343572	USA	(Blake et al. 2001)
93	AY026346.1	USA	
11	AY026347.1	USA	
Mexico	JX174178	Mexico	unpublished

114	AY026348	USA	(Blake et al. 2001)
91	AF343570	USA	
Buhl	AF343573	USA	
DryMills	AF343571.1	USA	
WB	AF342727	USA	
He	AF342730	Germany	
VTAB	NC_030242.1	Australia	(Mohr et al. 2015)

**Table 33. Reference strains used for phylogenetic analysis of segment B; NA means not assigned.**

Strain	GenBank	Geographical Origin	Reference
Mexico	EU665685.1	Mexico	unpublished
HL1	D26527.1	Korea	(Lee, Chung, and Lee 1994)
2310	AJ489241	Spain	(Cutrin et al. 2004)
Ja Duncan	NC_001916	Canada	(Duncan et al. 1991)
WB	AF078669.1	USA	Yao et al., 1998 (Yao and Vakharia 1998)
YTAV-06	AB281674	Japan	(Hirayama et al. 2007)
POBV	EU161286.1	China	(Zhao et al. 2008)
YTAV-NC1	AY129666	Korea	(Zhang and Suzuki 2003)
Y-6	AY129662	Japan	(Zhang and Suzuki 2003)
AY-98	AY123970	Japan	unpublished
VTAB	NC_030244.1	Australia	(Mohr et al. 2015)
He	JF734351.1	Spain	(Bandin et al. 2014)
Ca2	JF734354	Canada	
Ca3	JF734355	Canada	
Ca1	JF734353	Canada	
Ab2	JF734350.1	NA	
Ab1	AM114033.1	NA	(Cutrin et al. 2004)
2284	AJ489239.1	Spain	(Cutrin et al. 2004)
Sp1	AJ622823	France	(Galloux, Chevalier, Henry, Huet, Costa, et al. 2004)
IRIPNV	KC900161.1	Iran	(Dadar et al. 2013)
Gre	NA	Turkey	Unpublished
SDF4	NA	Turkey	Unpublished

1054	NA	Turkey	Unpublished
Sp3	AM889221	Denmark	(Dixon et al. 2008)
Sp2	M58757.1	NA	(Duncan et al. 1991)
Sp122	AY354524.1	Norway	(Shivappa et al. 2004)
Chile	KU609618	Chile	(Jorquera et al. 2016)
Sp116	AY354523	Norway	(Shivappa et al. 2004)
Sp103	AY354522.1	Norway	
TABV2013	KP268678.1	Australia	(Mohr et al. 2015)
TABV1998	NC_028253.1	Australia	
Te	AJ920336	UK	(Nobiron et al. 2008)

#### 2.6.4 RAXML

To calculate phylogenetic properties, sequences of segment A and segment B of 57 IPNV isolates were put together in one FastA file with additional reference genomes. RAXML v1.5 beta was then used to generate a maximum likelihood tree using the GTRGAMMA nucleotide substitution model and supporting bootstrap calculation of 1000 replicates of random generated trees (Silvestro and Michalak 2012). The trees were viewed and edited using FigTree v1.4.0 (Rambaut 2012).

#### 2.6.5 CAI calculation

CAI was calculated by Nicholas Di Paola according to his publication using the webserver at <http://genomes.urv.es/CAIcal> and <http://genomes.urv.es/CAIcal/E-CAI> (Puigbo, Bravo, and Garcia-Vallve 2008b, a) using the protein alignment FastA files created by Dr. Michaël Bekaert. The script and the used *Salmo salar* and *Onchorhynchus mykiss* genomes used in the calculation are listed in the separate additional appendix under V.4.

#### 2.6.6 HYPHY package

To perform positive selection calculation on HYPHY, FastA files of each IPNV protein (all 57 sequenced isolates aligned) with a corresponding tree file (created using BEAST package; newick file format) were used.

SLAC on HYPHY was performed using the following set up:

1. Type of selection: Positive Selection → Quick Selection Detection
2. Genetic code: Universal Code

3. Model: 012345
4. dN/dS option: estimate dN/dS only
5. Ancestor counting option: Single Ancestor counting
6. SLAC option: Full Tree
7. Ancestral state reconstruction and counting: Averaged
8. Test statistics: Approximate
9. Significance level: 0.1
10. Output option: Chart
11. dN/dS count: Skip

MEME on HYPHY was performed using the following set up:

1. Type of selection: Positive Selection → Quick Selection Detection
2. Genetic code: Universal Code
3. Model: 012345
4. dN/dS option: estimate dN/dS only
5. Ancestor counting option: MEME
6. Significance level: 0.1

#### 2.6.7 Dinucleotide composition

To estimate the abundance of dinucleotides in IPNV segment A and B sequences, EMBOSS COMPSEQ was used online (<http://emboss.bioinformatics.nl/cgi-bin/emboss/compseq>). The expected frequencies were calculated based on the single base frequency of input sequence and not by assumption, that every dinucleotide has equal frequency. A CpG abundance is considered low when  $CpG_{O/E} \leq 1$  and high when  $CpG_{O/E} \geq 1$ .

#### 2.6.8 SPLITSTREE

SPLITSTREE v4.14.6 (Huson and Bryant 2005) was used to generate Neighbour Networks of the IPNV VP2 and VP3 proteins using nucleotide acid alignments provided by Dr. Michaël Bekaert. The sequence was converted in amino acid sequence using MEGA v6.0.6 and homologue amino acids were removed with UGENE v1.25.0 (Okonechnikov et al. 2012) prior loading in SPLITSTREE.

### 2.6.9 Statistics and graphs

The graphic and statistical analysis was performed using the Mann-Whitney test in GRAPHPAD PRISM v.6.

Graphical finish of trees, networks and figures was done using Adobe Illustrator CS5.

### 3 Results

To analyse genomic changes in IPNV over the last three decades IPNV isolates from the diagnostic collection of the IoA were subjected to whole genome sequencing (WGS) using next generation sequencing (NGS) technology on the Illumina MiSeq platform.

#### 3.1 WGS protocol development and improvement

Initially a protocol for whole genome sequencing (WGS) using next generation sequencing technology (NGS) of RNA viruses developed at Goettingen University (GOEP) was used (Dilcher et al. 2012). This protocol was assessed in comparison to a simplified protocol from a team at MRC Virology Glasgow University (GLAP) (Thomson et al. 2016). Altogether, 27 IPNV isolates were sequenced with the GOEP and 30 with the GLAP.

Both protocols have a specific type of RNA extraction and ds-cDNA synthesis prior to the library preparation according to the Illumina protocol 2.1.2.3.

To determine if the GLAP was efficient for IPNV WGS, three samples (IoA024, IoA030 and IoA107) were randomly selected and processed with both the GOEP and the GLAP for comparison of yield and the amount of terminal nucleotides determined by de novo assembly. All RNA samples had a mean concentration of IPNV genome equivalents of higher than  $10^4$  copies/mL which is the minimum requirement for Illumina NGS. There was no significant difference estimated using the Mann-Whitney test in the ds-cDNA concentration, as measured on Qubit (see 2.1.2.2.2.4), between the two protocols ( $p=0.1000$ ) as seen in Figure 34, Section I. The concentrations ( $p=0.1000$ ), the molarity ( $p=0.1000$ ) and the average fragment size ( $p=0.7000$ ) of the library, estimated by Bioanalyzer (see 2.1.2.2.2.3), did not differ significantly between the protocols (see Table 34 Section I). The GOEP resulted in an overall higher mean coverage throughout the whole genome. The first third of the genome in most cases had a higher coverage (see Figure 12).

The mean coverage was higher in samples IoA024 and IoA030 using the GLAP but lower in sample IoA107. However, there was no significant difference between the two protocols used ( $p>0.9999$  segment A;  $p=0.7000$  segment B) (Table 35).

An analysis of the terminal nucleotides for each segment and the overall mapping rate before manual editing yielded no significant difference between the two protocols ( $p=0.2000$  missing terminal nucleotides;  $p>0.9999$ ). Furthermore, both protocols did not yield sequences



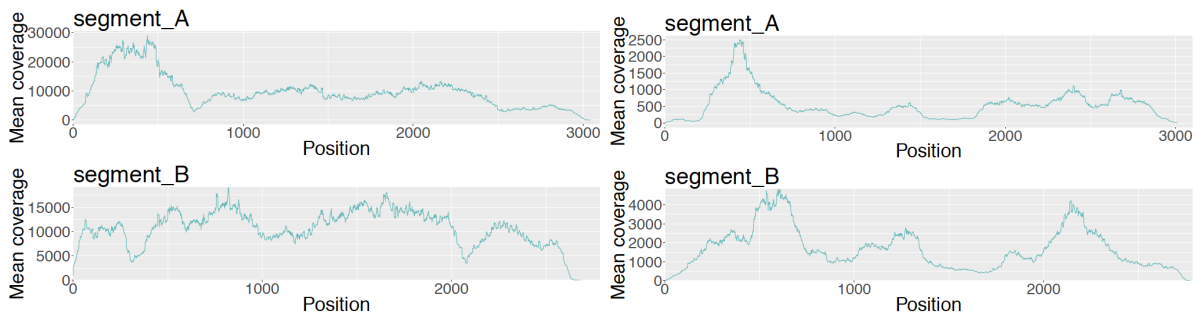
sufficient for a full genome assembly with a *de novo* approach as seen in the total number of missing terminal nucleotides (Table 36).

**Table 34. Control values of selected libraries in Section I: randomly selected samples processed with the Goettingen (GOEP) and the Glasgow (GLAP) sequencing protocol; Section II: Improvement variants for GLAP; Section III: temperature variants tested for GLAP; Section IV: range of control values of 57 sequenced IPNV samples. Conc.=concentration.**

	RNA		ds-cDNA		Library			
	Sequence ID and variant	RNA conc. NanoDrop	Mean conc. qPCR	ds-cDNA conc. NanoDrop (ng/μL)	ds-cDNA conc. Qubit (ng/μL)	Average fragment size (bp)	Conc. (pg/μL)	Mol. (pmol/l)
Section I	IoA024 GOEP	558.97	3.93E8	85.56	34.20	268	1649.56	9887.4
	IoA024 GLAP	8.67	8.69E10	8.07	0.57	249	983.33	6301.6
	IoA030 GOEP	330.83	5.31E8	76.69	35.00	277	2733.47	15937.0
	IoA030 GLAP	17.39	2.47E11	10.75	1.298	253	926.00	5887.6
	IoA107 GOEP	620.15	5.24E11	24.25	8.68	279	1472.53	8492.7
	IoA107 GLAP	19.81	1.94E9	5.26	0.369	368	24.64	108.7
Section II	Initial sample IoA039	21.20	4.77E7					
	ss-cDNA-GP32			19.46	2.14	281	2090.31	12160.0
	ds-cDNA-GP32			12.11	1.34	280	3306.55	19334.5
	ss/ds-cDNA-GP32			20.66	1.58	232	565.39	3947.7
	ss-cDNA-3'Primer			17.85	1.41	274	3975.92	23764.9
	ds-cDNA-3'Primer			12.63	1.51	281	3755.30	21969.5
	ss/ds-cDNA-3'Primer			18.06	1.17	279	2632.13	15502.0
	RNaseH treatment			17.19	1.72	276	3025.48	18038.7
	dNTPs and rand.hex.			11.26	1.85	282	2671.15	15559.0
	RNAincub.75°C/5min			18.63	1.25	285	1690.73	9724.8
GLAP unchanged			19.44	1.50	276	2438.04	14458.2	
Section III	IoA035 GLAP	26.82	1.01E7	12.39	3.06	332	9983.41	53481.7
	IoA035 GLAP 95°C	26.82	1.01E7	13.25	1.81	437	6687.40	30316.7
	IoA111 GLAP	49.43	2.62E8	9.03	0.49	246	1746.83	12334.0
	IoA111 GLAP 95°C	49.43	2.62E8	7.29	0.50	278	4159.70	31451.3
Section IV	GLAP lowest value	3.07	2.99E1	5.64	0.18	154	51.26	544.2
	GLAP highest value	135.77	8.48E8	17.85	2.80	473	8208.93	48193.5
	GOEP lowest value	21	1.56E1	1.79	0.72	260	813.00	5785.1
	GOEP highest value	1845.73	5.24E11	172.28	89.80	457	5085.71	31114.9

**Table 35. Mean coverage (number of reads/reference bases) of segment A and B for the samples processed with the GLAP and the GOEP.**

Sequence ID	Mean coverage segment A	Mean coverage segment B
IoA024 GOEP	534.7	1665.3
IoA024 GLAP	9784.3	10731.3
IoA030 GOEP	15120.6	41682.0
IoA030 GLAP	20066.3	23085.9
IoA107 GOEP	6222.1	19385.3
IoA107 GLAP	126.4	167.6



IoA024 GLAP

IoA024 GOEP

**Figure 12. Coverage (number of reads/reference bases) plots of one of the three randomly selected samples processed with the GOEP and the GLAP. Plots for the other samples see Figure 28 in the Appendix. Plots were generated using BOWTIE and SAMTOOLS (2.5.1.2).**

**Table 36. Comparison of missing nucleotides and the overall mapping rate in samples processed with the GOEP and the GLAP after de novo assembly.**

Sequence ID	segment A (bp)		segment B (bp)		Total (bp)	Mapping rate (%)
	5'	3'	5'	3'		
IoA024 GOEP	54	31	1	0	86	2.20
IoA024 GLAP	30	26	89	2	147	39.5
IoA030 GOEP	0	19	8	39	66	69.2
IoA030 GLAP	36	2	7	32	77	50.7
IoA107 GOEP	1	17	9	1	28	41.7
IoA107 GLAP	36	157	36	76	305	4.8

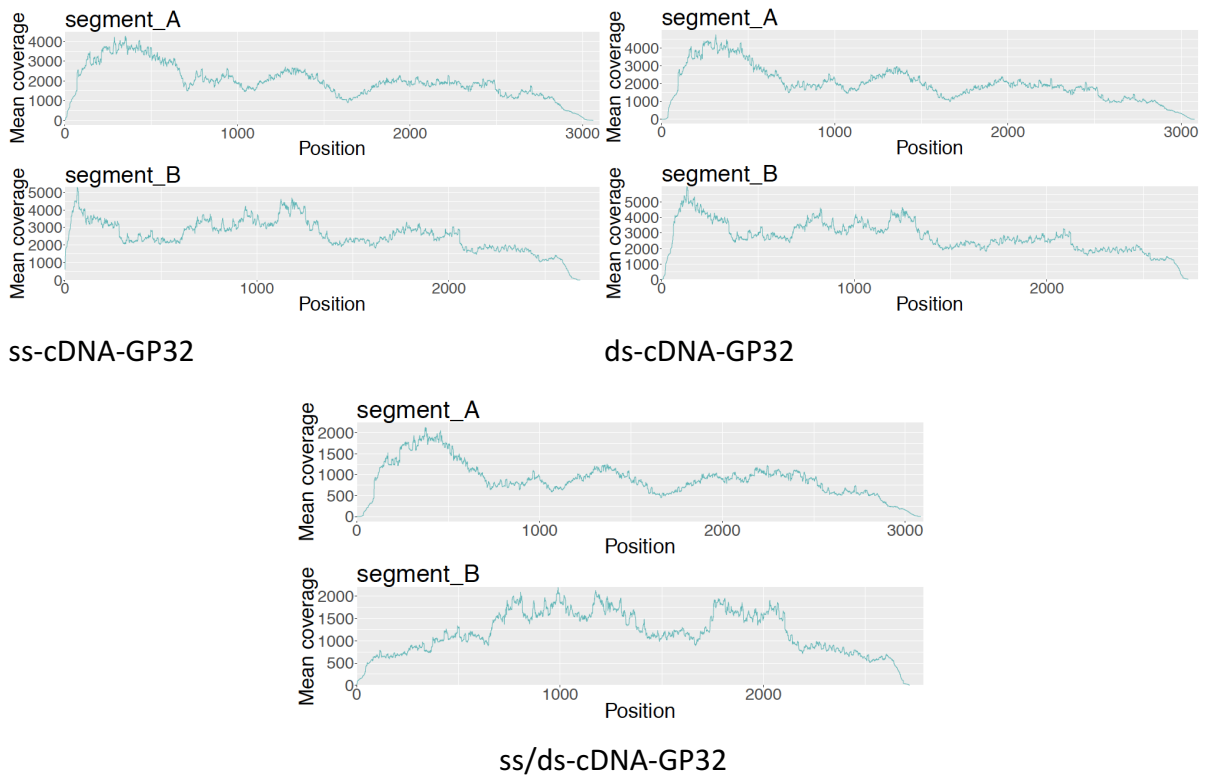
### 3.1.1 Improvements to the double strand cDNA synthesis step of the Glasgow protocol

In order to improve the termini coverage in the GLAP, an IPNV sample (IoA039) was randomly chosen and several changes to the ds-cDNA synthesis step were introduced to the GLAP (see 2.2.1):

- the single strand binding protein GP32 was added in the first, or the second or in both steps of the ds-cDNA synthesis.
- a 3' primer was added in the first, or the second or in both steps of the ds-cDNA synthesis.
- ss-cDNA was treated with RNase H prior the second strand synthesis.
- random hexamers and dNTPs were added to the RNA prior the RNA heat incubation step.
- the RNA was incubated at 75 °C for 5 minutes.

**Table 37. Mean coverage (number of reads/reference bases) of segment A and B for the sample processed with different changes in the GLAP.**

Sequence ID	Mean coverage segment A	Mean coverage segment B
ss-cDNA-GP32	1966.5	2553.8
ds-cDNA-GP32	1968.6	2759.1
ss/ds-cDNA-GP32	890.2	1178.1
ss-cDNA-3'Primer	1550.1	2255.1
ds-cDNA-3'Primer	1462.8	2315.0
ss/ds-cDNA-3'Primer	1691.9	2405.9
RNaseH treatment	1017.7	1461.5
dNTPs and rand.hex.	2109.5	3296.6
RNAincub.75 °C/5 min	754.2	1169.2
GLAP unchanged	1238.6	1732.7



**Figure 13.** Coverage (number of reads/reference bases) plots of protocol variants with added GP32. Plots for the other variants see Figure 29 in the Appendix section. Plots were generated using BOWTIE and SAMTOOLS (2.5.1.2).

**Table 38.** Comparison of missing nucleotides and the overall mapping rate of one randomly selected sample and its improvement variants after de novo assembly.

Sequence ID variant	segment A (bp)		segment B (bp)		Total (bp)	Mapping rate (%)
	5'	3'	5'	3'		
ss-cDNA-GP32	31	2	89	2	124	4.12
ds-cDNA-GP32	5	12	23	20	60	4.31
ss/ds-cDNA-GP32	6	4	41	19	70	2.77
ss-cDNA-3'Primer	79	6	42	2	129	2.76
ds-cDNA-3'Primer	22	3	89	978	1092	2.20
ss/ds-cDNA- 3'Primer	8	32	23	21	84	3.39
RNaseH treatment	79	18	42	19	158	2.27
dNTPs and rand.hex.	37	2	24	993	1056	3.86

RNAincub.75 °C/5 min	36	6	42	15	99	2.27
GLAP unchanged	8	28	42	19	97	3.23

The sequencing results of these variant protocols were compared to the results with the unchanged GLAP in terms of the genome coverage after *de novo* assembly.

Significant improvements between variants in regard to the ds-cDNA concentration and the Bioanalyzer results were not observed (see Table 34 Section II).

RNaseH treatment, RNA incubation at 75 °C/5 min, and ss/ds-cDNA-GP32 resulted in a lower mean coverage in both segments A and B compared to the unchanged Glasgow protocol (Table 37).

The variations in the protocol had no significant influence on the coverage plots and did not improve read yields for full termini coverage regarding missing nucleotides after a *de novo* genome assembly (Figure 13, Table 38,  $p=0.061772$ ).

### 3.1.2 Improvements to RNA procedures of the Glasgow protocol

Finally, it was investigated if specific changes in the RNA incubation before ds-cDNA synthesis have an impact on full genome *de novo* assembly. For this 2 samples were randomly chosen and the RNA was processed at a standard incubation temperature of 65°C and in comparison at 95 °C, as it is known that RNA unfolds at very high temperatures (See Table 34 Section III) (Li, Viereg, and Tinoco Jr 2008, Tinoco Jr 2004).

Sample IoA111 showed higher ds-cDNA concentration by Qubit measurement, larger average fragment size and higher concentration and molarity estimated by the Bioanalyzer after the 95°C incubation. However, this was not observed for sample IoA035.

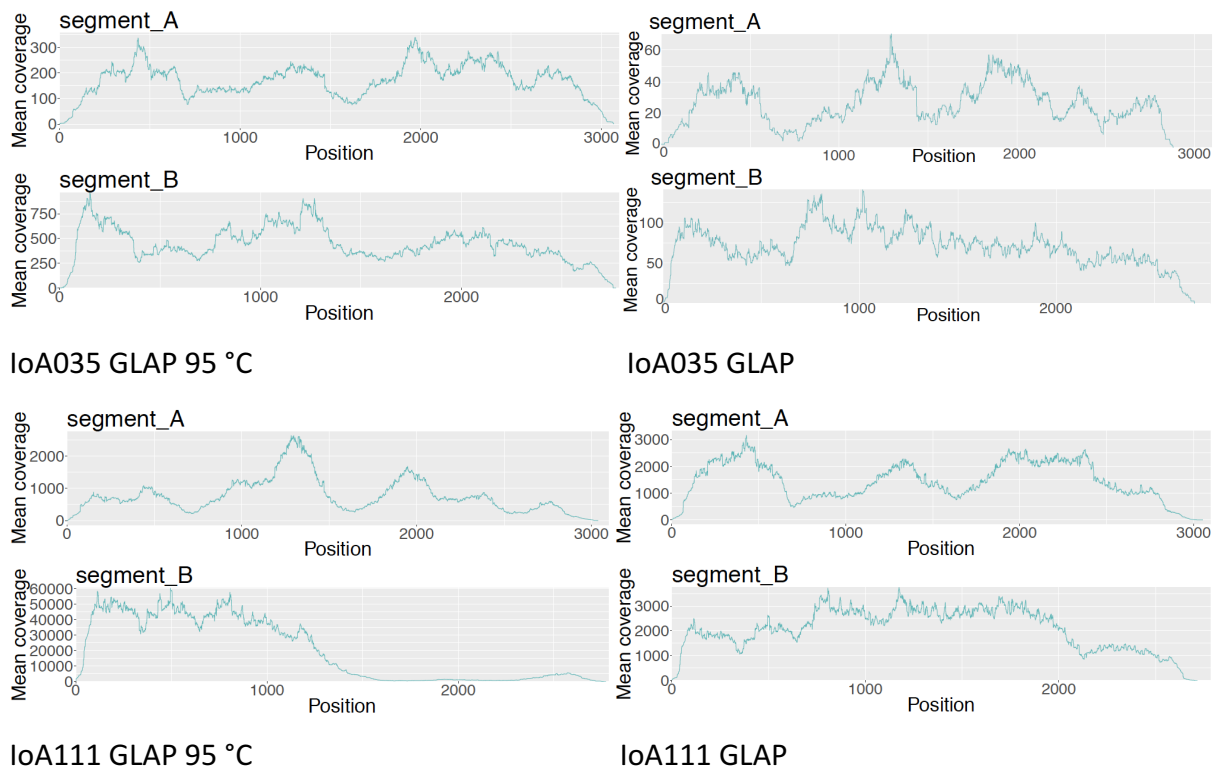
The obtained coverage plots show that samples incubated with 95 °C have a higher mean coverage for segment A and B, except for segment A of IoA111 95 °C RNA where the unchanged GLAP results in a higher mean coverage (Figure 14). The coverage (Table 39) was not significantly better for both segments between the two variants ( $p=0.6667$  segment A;  $p=0.3333$  segment B).

The total number of missing nucleotides was not significant between samples whose RNA was incubated at 95 °C prior the ds-cDNA synthesis step and samples processed with the unchanged Glasgow protocol ( $p= 0.6667$ ). The overall mapping rate showed no significant

difference ( $p=0.3333$ ). Full genomes could not be assembled using *de novo* without editing despite the changes in the incubation step of the protocol (Table 40).

**Table 39. Mean coverage (number of reads/reference bases) of segment A and B for two samples processed with RNA incubation changes in the GLAP.**

Sequence ID	Mean coverage segment A	Mean coverage segment B
IoA035 GLAP 95 °C	33761.8	64944.6
IoA035 GLAP	27.3	71.4
IoA111 GLAP 95 °C	767.0	19879.1
IoA111 GLAP	1452.9	2064.0



**Figure 14. Coverage (number of reads/reference bases) plots of two randomly selected samples processed with different incubation conditions for RNA prior ds-cDNA synthesis. Plots were generated using BOWTIE and SAMTOOLS (2.5.1.2).**

**Table 40. Comparison of missing nucleotides and the overall mapping of two randomly selected samples processed with different incubation conditions for RNA prior ds- cDNA synthesis after de novo assembly.**

Sample variant	segment A (bp)		segment B (bp)		Total (bp)	Mapping rate (%)
	5'	3'	5'	3'		
IoA035 GLAP 95 °C	31	11	41	4	87	93.46
IoA035 GLAP	59	156	44	29	288	0.12
IoA111 GLAP	40	7	41	20	108	8.54
IoA111 GLAP 95 °C	26	31	41	15	113	43.22



## 3.2 Sequencing Results

### 3.2.1 Illumina MiSeq sequencing

57 IPNV samples were randomly selected from the diagnostic collection and subjected to whole genome sequencing (WGS) using next generation technology (NGS) on the Illumina MiSeq platform. The selected isolates were collected from 1982 – 2014. Thirty-seven were isolated from *S. salar* and 5 from *O. mykiss* (Table 41).

IPNV samples were grown in the corresponding cell line and harvested in order to perform RNA extraction, which was then used for ds-cDNA synthesis, according to the GOEP (see 2.1.2.1.1) and later to the GLAP (see 2.1.2.1.2). Library preparation was performed according to the Illumina protocol in the same way for all samples (see 2.1.2.3).

**Table 41. Sequenced IPNV samples and their sampling information and the protocol used; NA=not assigned.**

Sequence ID	Collection Year	Cell line	Protocol	Host	Country of origin	Region
IOA_IPNV006	1982	CHSE-214	Glasgow	<i>Salmo salar</i>	UK	Scotland, Argyll
IOA_IPNV060	1983	CHSE-214	Goettingen	NA	NA	NA
IOA_IPNV007	1984	CHSE-214	Glasgow	<i>Oncorhynchus mykiss</i>	UK	Scotland, Borders
IOA_IPNV008	1984	CHSE-214	Glasgow	<i>Oncorhynchus mykiss</i>	UK	Scotland, Borders
IOA_IPNV011	1986	CHSE-214	Glasgow	<i>Salmo salar</i>	UK	Scotland, Highlands
IOA_IPNV061	1986	CHSE-214	Goettingen	<i>Scophthalmus maximus</i>	NA	NA
IOA_IPNV019	1987	CHSE-214	Glasgow	<i>Oncorhynchus mykiss</i>	UK	Scotland, Highlands
IOA_IPNV021	1987	CHSE-214	Glasgow	<i>Oncorhynchus mykiss</i>	UK	Scotland, Strathclyde
IOA_IPNV062	1987	CHSE-214	Goettingen	<i>Salmo salar</i>	NA	NA
IOA_IPNV063	1987	CHSE-214	Goettingen	<i>Salmo salar</i>	NA	NA
IOA_IPNV064	1987	CHSE-214	Goettingen	NA	UK	Scotland, Highlands
IOA_IPNV065	1987	CHSE-214	Goettingen	NA	UK	Scotland, Highlands

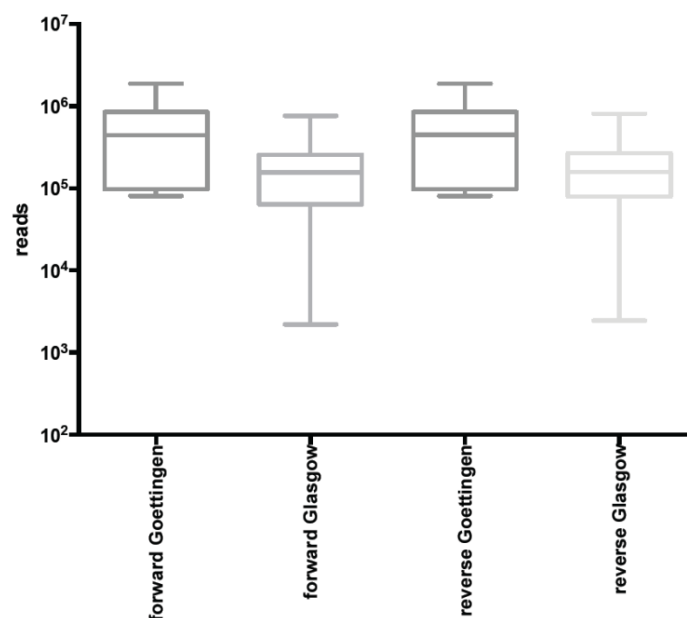
IOA_IPNV066	1988	CHSE-214	Goettingen	<i>Salmo salar</i>	UK	Scotland, Highlands
IOA_IPNV027	1989	CHSE-214	Glasgow	<i>Oncorhynchus mykiss</i>	UK	Scotland, Borders
IOA_IPNV032	1989	CHSE-214	Glasgow	<i>Salmo salar</i>	UK	Scotland, Shetland
IOA_IPNV068	1989	CHSE-214	Goettingen	<i>Salmo salar</i>	UK	Scotland, Highlands
IOA_IPNV026	1990	CHSE-214	Glasgow	<i>Salmo salar</i>	UK	Scotland, Shetland
IOA_IPNV069	1992	CHSE-214	Goettingen	<i>Salmo salar</i>	NA	NA
IOA_IPNV070	1992	CHSE-214	Goettingen	NA	Norway	Norway
IOA_IPNV071	1992	TO	Goettingen	NA	UK	Scotland, Argyll
IOA_IPNV072	1992	CHSE-214	Goettingen	NA	UK	Scotland, Shetlands
IOA_IPNV073	1993	CHSE-214	Goettingen	<i>Salmo salar</i>	NA	NA
IOA_IPNV075	1993	TO	Goettingen	<i>Salmo salar</i>	Norway	Norway
IOA_IPNV076	1994	TO	Goettingen	<i>Salmo salar</i>	NA	NA
IOA_IPNV081	1994	CHSE-214	Glasgow	<i>Salmo salar</i>	NA	NA
IOA_IPNV082	1996	CHSE-214	Goettingen	<i>Salmo salar</i>	NA	NA
IOA_IPNV083	1997	CHSE-214	Goettingen	<i>Salmo salar</i>	NA	NA
IOA_IPNV084	1997	CHSE-214	Glasgow	<i>Salmo salar</i>	NA	NA
IOA_IPNV087	1997	CHSE-214	Glasgow	<i>Salmo salar</i>	NA	NA
IOA_IPNV088	1997	CHSE-214	Glasgow	<i>Salmo salar</i>	UK	Scotland, Shetlands
IOA_IPNV090	1998	CHSE-214	Goettingen	<i>Salmo salar</i>	NA	NA
IOA_IPNV107	1999	CHSE-214	Goettingen	<i>Salmo salar</i>	UK	Scotland, Shetlands
IOA_IPNV111	1999	CHSE-214	Glasgow	<i>Salmo salar</i>	UK	Scotland, Shetlands
IOA_IPNV001	2000	TO	Glasgow	<i>Salmo salar</i>	NA	NA
IOA_IPNV003	2002	TO	Glasgow	<i>Salmo salar</i>	NA	NA
IOA_IPNV009	2003	CHSE-214	Goettingen	<i>Salmo salar</i>	UK	Scotland, Argyll
IOA_IPNV010	2003	TO	Glasgow	<i>Salmo salar</i>	NA	NA
IOA_IPNV013	2004	TO	Glasgow	<i>Salmo salar</i>	UK	Scotland, Argyll

IOA_IPNV014	2004	CHSE-214	Goettingen	NA	NA	NA
IOA_IPNV018	2004	TO	Glasgow	NA	NA	NA
IOA_IPNV024	2005	TO	Goettingen	<i>Salmo salar</i>	NA	NA
IOA_IPNV030	2006	TO	Goettingen	<i>Salmo salar</i>	NA	NA
IOA_IPNV031	2006	TO	Glasgow	<i>Salmo trutta</i>	UK	Scotland, Shetlands
IOA_IPNV035	2007	CHSE-214	Glasgow	<i>Salmo trutta</i>	UK	Scotland, Shetlands
IOA_IPNV039	2008	CHSE-214	Glasgow	<i>Salmo salar</i>	NA	NA
IOA_IPNV051	2010	CHSE-214	Glasgow	<i>Salmo salar</i>	UK	Scotland, Highlands
IOA_IPNV052	2011	CHSE-214	Glasgow	<i>Salmo salar</i>	UK	Scotland, Highlands
IOA_IPNV053	2011	CHSE-214	Glasgow	<i>Salmo salar</i>	UK	Scotland, Strathclyde
IOA_IPNV054	2011	CHSE-214	Glasgow	<i>Salmo salar</i>	UK	Scotland, Shetland
IOA_IPNV055	2012	CHSE-214	Glasgow	<i>Salmo salar</i>	UK	Scotland, Highlands
IOA_IPNV056	2012	CHSE-214	Glasgow	<i>Salmo salar</i>	UK	Scotland, Orkney
IOA_IPNV057	2013	CHSE-214	Glasgow	<i>Salmo salar</i>	UK	Scotland, Shetland
IOA_IPNV047	2014	CHSE-214	Goettingen	NA	NA	NA
IOA_IPNV048	2014	CHSE-214	Goettingen	NA	NA	NA
IOA_IPNV049	2014	CHSE-214	Goettingen	NA	NA	NA
IOA_IPNV050	2014	CHSE-214	Goettingen	NA	NA	NA
IOA_IPNV058	2014	CHSE-214	Glasgow	<i>Salmo salar</i>	UK	Scotland, Outer Hebrides

During sample preparation, measurements were performed on each sample to control the amount of RNA, ds-cDNA and of the library. As a quality measure for samples that were processed with the GLAP the quantity of target RNA molecules/mL was determined using RT-qPCR (see 2.1.2.2.1.2). All samples had a mean concentration over  $10^4$  copies of IPNV genome equivalents/mL, which was considered as a minimum threshold for successful virus genome sequencing using NGS.

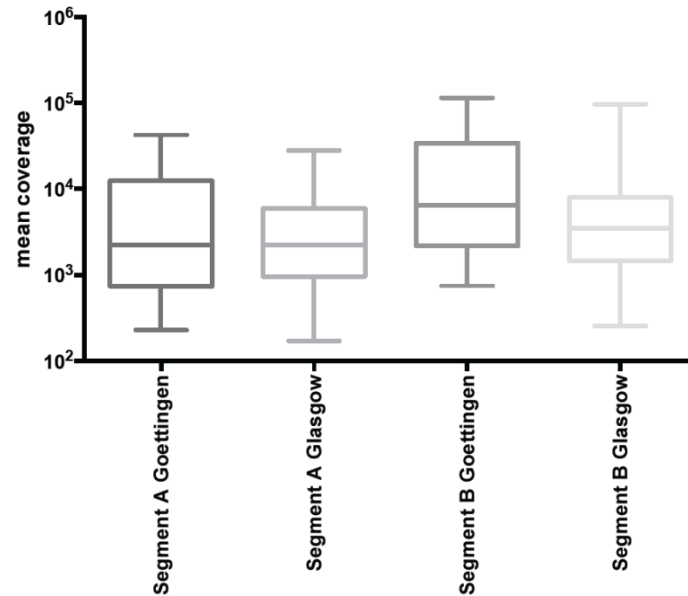
The library fragment sizes ranged from 154 bp to 473 bp as determined with capillary electrophoresis. We observed a significant difference ( $p < 0.0001$ ) in the initial RNA concentration measured on the NanoDrop (see 2.1.2.2.1.1) as well as in the ds-cDNA concentration measured on the Qubit ( $p < 0.0001$ ) between the GOEP and the GLAP with higher concentrations in samples processed with the GOEP (See Table 34 Section IV and Table 60).

Before sequence reads were used for the *de novo* genome assembly, they were checked for quality and were decontaminated using a specific tool that deletes non-viral reads (see Methods 2.5.1). A significant difference ( $p = 0.0032$  forward reads;  $p = 0.0034$  reverse reads) between the RNA to ds-cDNA protocols was noted, with the GOEP resulting in higher number of reads remaining after read preparation steps that is later used for genome assembly (Figure 14, Table 61).



**Figure 15. Forward and reverse reads recovered from IPNV samples processed with the GLAP or the GOEP.**

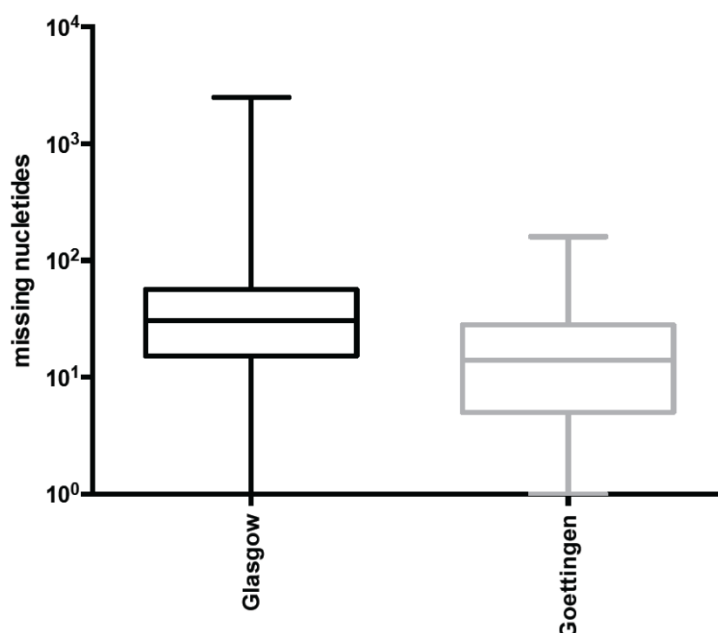
Sequencing reads were then used to estimate the genome coverage based on a reference genome for segment A and B (Figure 15). There was no difference in the mean coverage between the two protocols ( $p = 0.5203$  segment A;  $p = 0.0695$  segment B) using the Mann-Whitney test.



**Figure 16. Mean coverage (number of reads/reference bases) of segment A and B for 57 processed IPNV samples with the GLAP or GOEP.**

The mean maximum coverage for segment A was 18004.51 and for segment B 38575.39. The mean minimum coverage for segment A was 0.88, and for segment B 5.26 (Table 62). Some samples shared a similar coverage pattern with a high coverage towards the 5'-end and a lower coverage towards the 3'-end of the sequence, as well as the position of the peaks. Some dips were observed in several IPNV samples in the second thirds of the genome (Figure 30).

There are no visible clear differences between the reads obtained from the GLAP and the GOEP. Coverage at the termini of both segments in all cases was very low, independent of the protocol used. For segment A missing nucleotides ranged from 1-92 bp at the 5'-end and 1-157 bp at the 3'-end, and for segment B from 1-26 bp at the 5'-end to 1-37 bp at the 3'-end. Segment A showed a higher number of missing nucleotides than segment B (Table 63).



**Figure 17. Missing terminal nucleotides of both segments of IPNV samples processed with the GLAP or the GOEP after *de novo* assembly.**

An important part of post *de novo* assembly is to quantify missing terminal nucleotides. This allows to assess, how well a protocol or procedure is working for the determination of the sequence of linear virus genomes. After assembling the IPNV genomes with the *de novo* approach, nucleotides at the termini were missing in most samples (Figure 16, Table 63) with a significant difference between the protocols ( $p=0.0143$ ). The overall mapping rate was significantly different between the two protocols used ( $p=0.0210$ ) resulting in higher mapping rates for the GOEP.

Whole genome *de novo* assemblies with no missing nucleotides at the termini and therefore without manual editing were achieved only in a few samples with both, the GOEP and the GLAP, as seen in samples loA60, 011, 019, 021, 062, 066, 071 and 081.

### 3.2.2 MinION

In order to further improve the termini coverage, MinION sequencing technology (Oxford Nanopore Technology, UK) was tested, which sequences the DNA strand by threading it through a pore protein embedded in a membrane to identify individual nucleotides by the unique change in the electrical conductivity as the DNA molecule passes through the nanopore protein (see 1.2.4).

One IPNV sample (IoA039) was randomly selected and sequenced on the MiSeq, using the GLAP for RNA extraction and ds-cDNA synthesis and the Illumina protocol for library preparation, and compared it to the sequences obtained from MinION sequencing, using the protocol according to Oxford Nanopore sequencing (see 2.4, Table 42).

MinION sequencing resulted in a consistent coverage throughout the genome.

The average fragment size of the library for MinION sequencing was higher but the concentration and molarity was lower than in the library for MiSeq sequencing (Table 43).

The coverage plots of a sample sequenced on the MiSeq and the MinION seen in Figure 18 show an up to 40 times higher mean coverage recovered from the MiSeq (Table 44). Reads recovered from both methods cover the full genome only with manual editing after the *de novo* assembly, although the mapping rates are higher on the MiSeq (Table 45).

There is no significant difference in missing terminal nucleotides between the genomes recovered from sequencing on the MiSeq and the MinION ( $p=0.3143$ ).

**Table 42. Control values of one randomly selected sample processed for MinION 2D sequencing.**

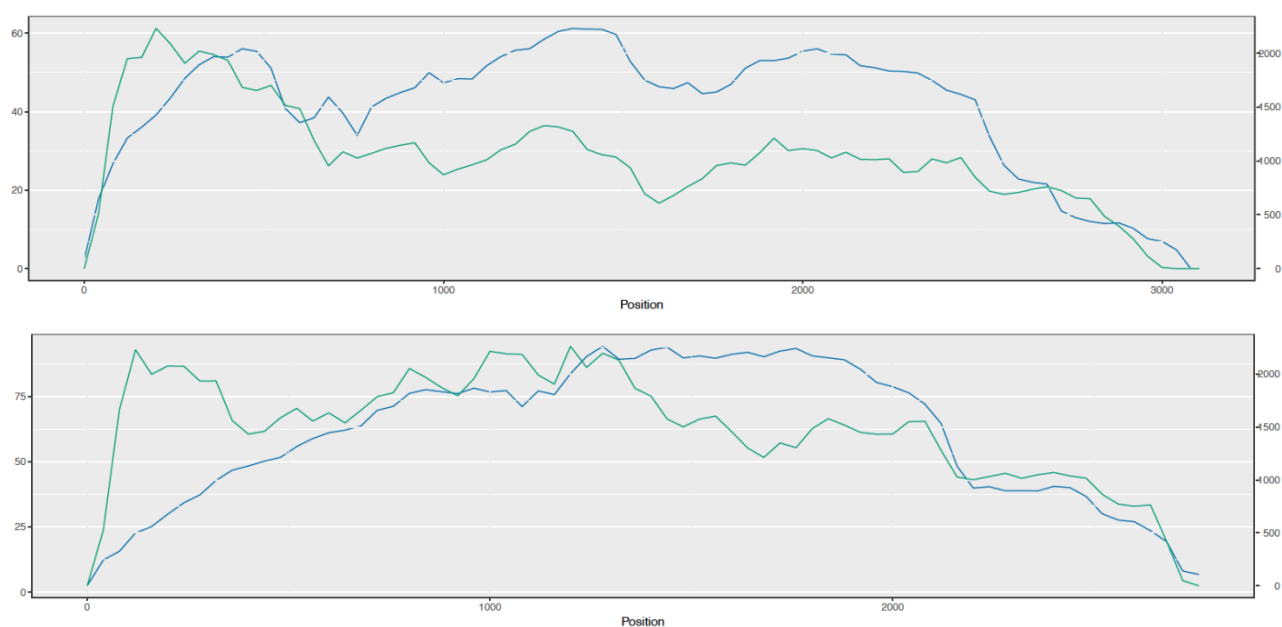
Sequence ID	initial conc. of ds-cDNA Qubit (ng/μL)	cDNA amplification		Library preparation	
		post Ligation conc. (ng/μL)	post PCR amplification conc. (ng/μL)	post EndPrep conc. (ng/μL)	post Elution conc. (ng/μL)
IoA039	1.50	0.52	2.26	4.16	0.95

**Table 43. Bioanalyzer results of the library of one randomly selected sample processed for MinION 2D and MiSeq sequencing.**

Sequence ID	Library preparation		
	Average fragment size (bp)	Conc. (pg/μL)	Mol. (pmol/l)
IoA039 MinION	1110	1501.18	3199.5
IoA039 MiSeq	274	3975.92	23764.9

**Table 44. Mean coverage (number of reads/reference bases) of segment A and B for one samples sequenced on the MiSeq and the MinION.**

Sequence ID	Mean coverage segment A	Mean coverage segment B
IoA039 MinION	41.7	61.1
IoA039 MiSeq	1534.4	2228.7



**Figure 18. Comparing the coverage (number of reads/reference bases) plots of one randomly selected sample (upper: segment A, lower: segment B) sequenced on the MinION (blue, left coordinates) and on the MiSeq (green, right coordinates). Plots were generated using BOWTIE and SAMTOOLS (2.5.1.2).**

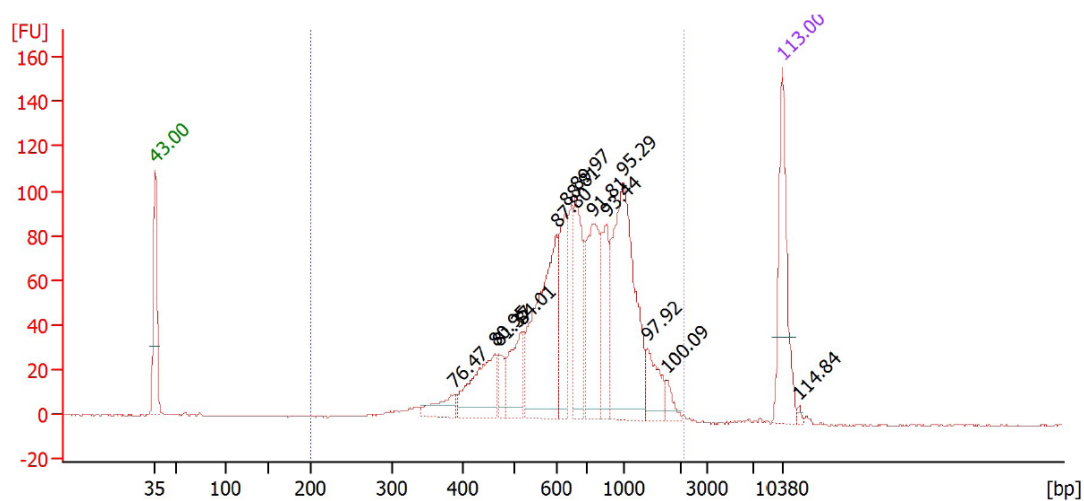
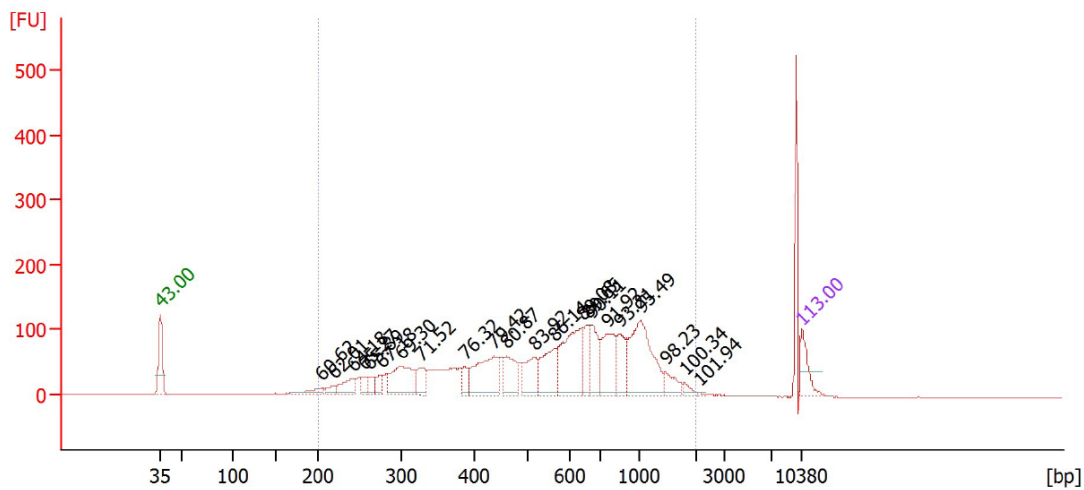
**Table 45. Missing nucleotides and the overall mapping rate of one randomly selected sample sequenced on the MinION and the MiSeq in comparison after de novo assembly.**

Sequence ID	Segment A (bp)		Segment B (bp)		Total (bp)	Mapping rate (%)
	5'	3'	5'	3'		
IoA039 MinION	10	12	0	14	36	0.85
IoA039 MiSeq	1	5	2	2	10	2.78

### 3.2.2.1.1 Improvement of MinION sequencing

In order to check if the fragment length for MinION sequencing can be in general increased, it was decided to change the sample to AmpureXP beads ratio by testing it on one randomly selected betanodavirus ds-cDNA. For this the beads to sample ratio was changed from 1x to 0.6x (see 2.4.3). Using this approach, the fragments size eluted yielded a much more pronounced peak at 600-1000 bp as seen in Figure 18.





**Figure 19.** Bioanalyzer results of one randomly selected betanodavirus sample used for improvement of fragment length for MinION sequencing; upper shows the fragment range with 1x beads to sample ratio, lower shows the fragment range with 0.6x beads to sample ratio.

**Table 46.** Bioanalyzer results of two randomly selected samples used for improvement of fragment length for MinION sequencing.

Sequence ID		Average fragment size (bp)	Conc. (pg/μL)	Mol. (pmol/l)
Noda	before size selection	642	3163.08	10387.5
Pt/08/Sba	after size selection	770	983.43	2325.2

Table 46 shows that after the change to 0.6x beads to sample ratio, the average fragment size increased although the concentration and molarity decreased.

### 3.3 Sequencing Analysis

#### 3.3.1 Phylogenetic analysis

In order to obtain an insight into the phylogenetic relationship of the 57 IPNV isolates a tree was calculated using the BEAST (MCMC method) software package (2.6.3) and RAXML (2.6.4; results not shown as trees showed the same result) in reference to published sequences available from GenBank (Table 32 and Table 33).

Fifty seven IPNV sequences determined from isolates in Scotland spread across genogroups I (Jaster), III (Tellina/Canada1) and V (Sparajub) in case of segment A and across genogroups I, II (Abild), III and V in case of segment B (Figure 20 and Figure 21, Table 64).

Samples IoA 060, 061 and 081 group in genogroup I both in segment A and B. Segment A of samples IoA 006, 011, 019, 063, 064, 065, 066, 068 and 069 cluster in genogroup III (Figure 20) whereas for segment B only samples IoA 006, 011, 019 and 069 belong to genogroup III (Figure 21) while IoA063, 064, 065, 066 and 068 cluster in genogroup II. All other samples cluster within genogroup V for both segments.

In segment B we observe that genogroup III seems to be split up as the Tellina reference isolate is grouping separately and not as in segment A within the genogroup III. This might be due to serotyping/genotyping issue. In serotyping the neutralizing antibodies to specific epitopes, with in IPNV are within the VP2 protein, which is coded in segment A. Sequencing and genotyping the segment B therefore might result in a different grouping of the previously serotyped and defined (by segment A) isolates.

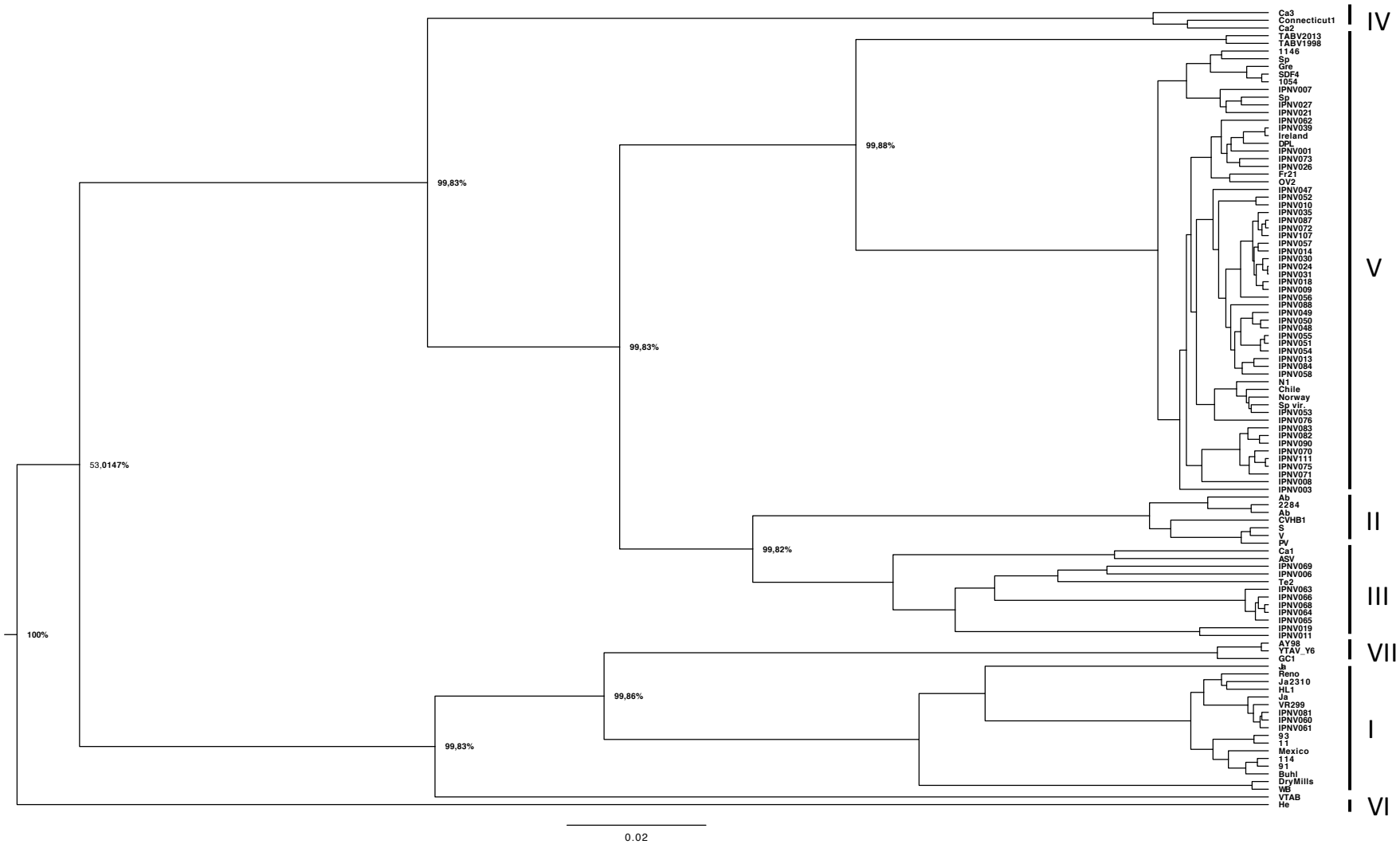
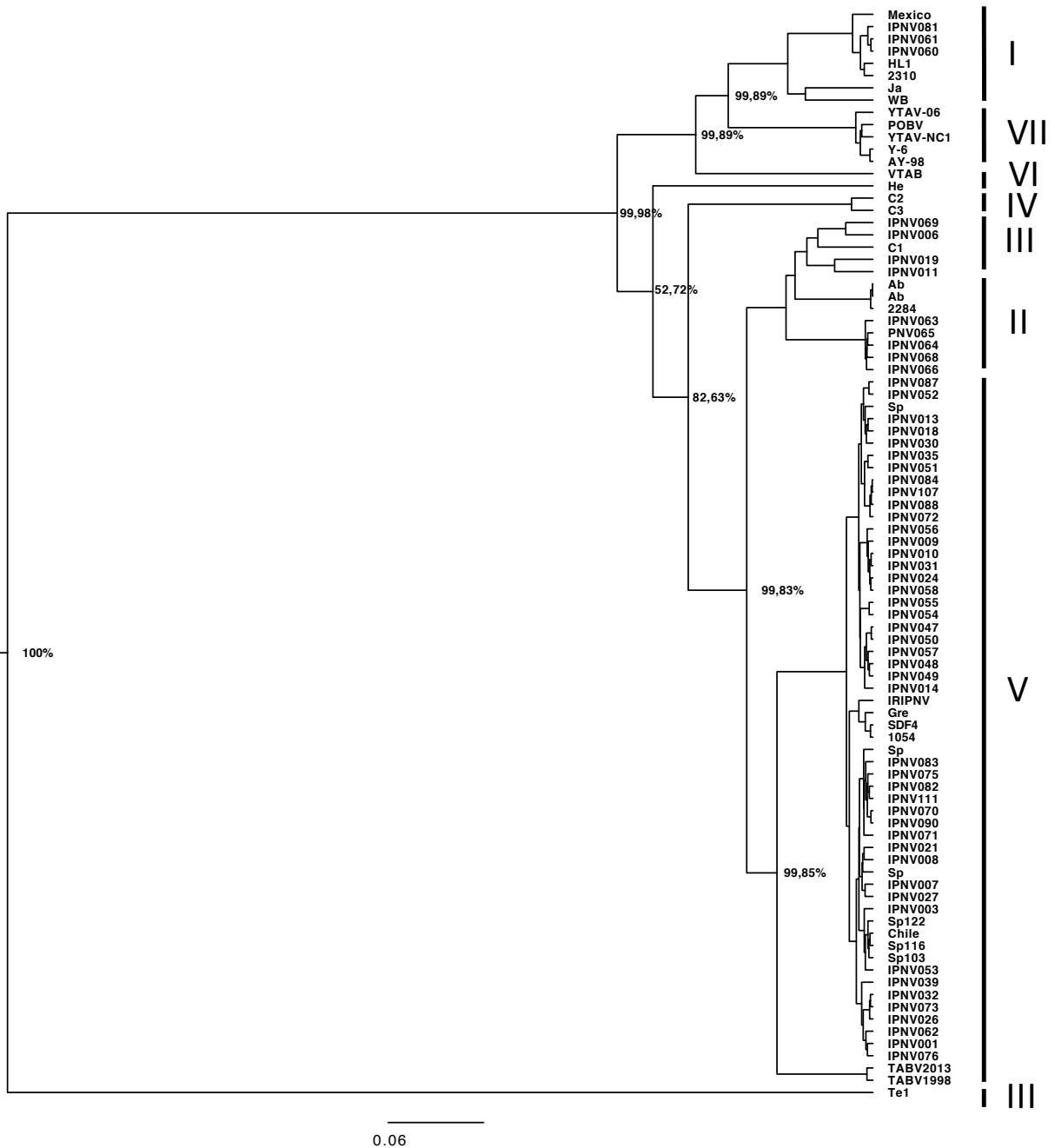


Figure 20. Phylogenetic classification of 57 IPNV isolates based on segment A sequences using reference sequences (Table 32). The unrooted tree was calculated using BEAST with 10,000,000 chains per cycle with every 1000<sup>th</sup> chain being recorded. Posterior probability values are shown at corresponding nodes in percent. The scale represents 0.02 substitutions per nucleotide site.



**Figure 21. Phylogenetic classification of 57 IPNV isolates based on segment B sequences using reference sequences (Table 33). The unrooted tree was calculated using BEAST with 10,000,000 chains per cycle with every 1000<sup>th</sup> chain being recorded. Posterior probability values are shown at corresponding nodes in percent. The scale represents 0.06 substitutions per nucleotide site. Te serotype clusters separately but is still defined as genogroup III (see Table 2).**

**Table 47. Genogroup distribution of the sequenced IPNV isolates.**

<b>Genogroup</b>				
<b>V</b>	<b>I</b>		<b>II</b>	<b>III</b>
45 (78.95 %)	3 (5.26 %)	segment A	0	9 (15.79 %)
		segment B	5 (8.77 %)	4 (7.02 %)

An analysis of the distribution of genogroups revealed that 78.95 % of the 57 IPNV sequences belong to genogroup V and only 5.26 % to genogroup I. Of the remaining isolates 15.79 % of segment A sequences cluster within genogroup III and 8.77 % of segment B sequences cluster within genogroup II and 7.02 % within the genogroup III (Table 47)

### 3.3.2 Analysis of amino acid sequence features of IPNV proteins

When comparing the amino acid composition of the 57 sequenced IPNV isolates (IoA32 excluded for segment A and its encoded proteins) several genogroup-specific differences in all IPNV proteins genes were observed. Some amino acids are specific for two of the three genogroups throughout the genomes. Within the genogroup II/III we observe amino acid differences between samples that cluster within genogroup III in both, segment A and B and samples that cluster within genogroup III in segment B. In the following the comparative analysis of the individual protein sequences highlights the main observed features.

#### 3.3.2.1 Analysis of features of VP2

Previous work on IPNV has indicated the importance of amino acid sites 217, 221, 247, 252, 281, 282 and 319 in VP2 for the pathogenicity of IPNV, with a focus on the positions 217 and 221. To analyse and compare the VP2 protein sequence of 56 IPNV isolates (excluding IoA032, as it was showing gaps in the region of interest) the segment A protein sequences of all 56 isolates were aligned in reference to the virulent Sp strain NVI-001 documented by Santi et al. in 2004.

All genogroup III strains share the same amino acid residue pattern (Table 48).

Two genogroup I isolates, IoA060 from 1983 and IoA061 (ex *Scophthalmus maximus*) from 1986 share the same residue pattern whereas IoA081 from 1994 (ex *Salmo salar*) has different amino acids in positions 217, 247 and 282.

Within the genogroup V strains a range of amino acid variations in all decades and hosts can be seen.

**Table 48. VP2 residue characteristics pattern of sequenced IPNV isolates compared to the virulent reference strain excluding IoA032. Amino acid positions identified in previous studies as important in terms of viral virulence are listed, with a focus on the positions 217 and 221 (underlined).**

Cluster	Sequence ID	Collection Year	Genogroup	Amino acid position						
				<u>217</u>	<u>221</u>	247	252	281	282	319
1	IOA_IPNV 06,11,19, 63, 64, 65, 66, 68, 69	1982, 1986, 1987, 1988, 1989, 1992	III	P	T	P	A	T	A	A
2	IOA_IPNV 7, 8, 21, 62, 26, 70, 71, 73, 75, 76, 81, 83, 84, 107, 1, 3, 39	1984, 1987, 1990, 1992, 1993, 1994, 1997, 1999, 2000, 2002, 2008	I or V	P	T	A	N	T	N	A
3	IOA_IPNV 27	1989	V	P	T	A	N	T	N	V
4	IOA_IPNV 24, 31, 57, 47, 48, 50	2005, 2006, 2013, 2014	V	P	T	A	V	T	N	A
5	IOA_IPNV 72	1992	V	P	A	A	T	T	N	A
6	IOA_IPNV 87, 88, 90, 111	1997, 1998, 1999	V	P	A	A	I	T	N	A
7	IOA_IPNV 9, 10, 13, 18, 30, 35, 51, 52, 54, 55, 56, 49, 58	2003, 2004, 2006, 2007, 2010, 2011, 2012, 2014	V	P	A	A	V	T	N	A
8	IOA_IPNV 60, 61, 82	1983, 1986, 1996	I or V	A	T	E	N	T	A	A
9	IOA_IPNV 14	2004	V	P	S	A	V	T	N	A
10	IOA_IPNV 53	2011	V	T	A	A	V	T	N	A
	Santi et al., 2004	2000	V	T	A	T	V	T	N	A

Apart from the configuration for virulence in VP2 (T217, A221), Santi also described low virulent (P217, A221), avirulent (T217, T221) or persistent (P217, T221) configurations respectively.

Of the 56 IPNV isolates analysed here, more than half belong to the persistent type, followed by low virulent strains. Only one sequence shows a virulent amino acid characteristic, as defined by previous research. No avirulent isolates were found. One isolate shows an P217, S221 amino acid characteristic and 3 isolates show a A217, T221 amino acid characteristic.

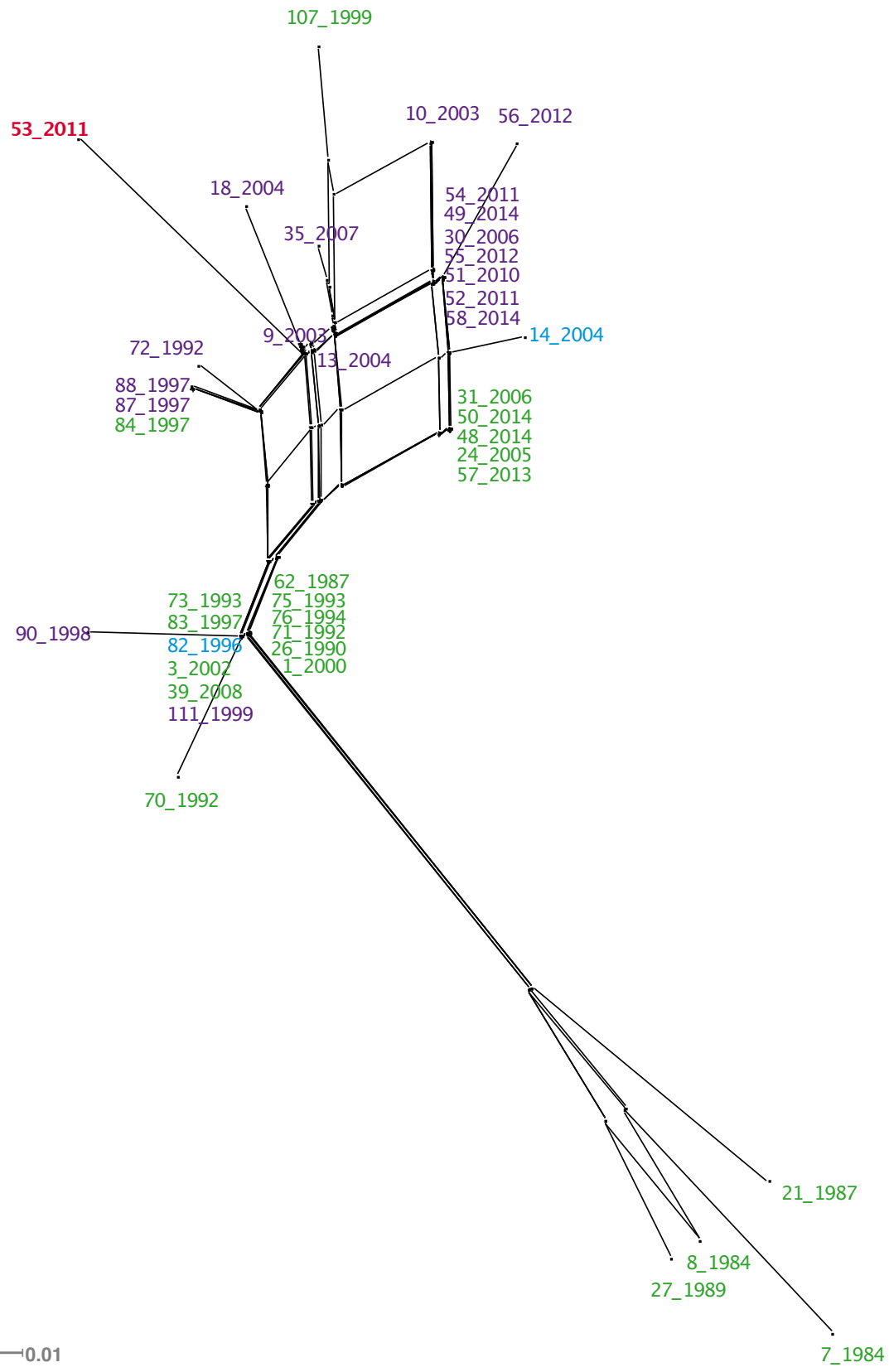
One isolate could not be characterised due to non-assigned nucleotides in the alignment. The virulence characteristic distribution shows that persistent type is dominating in this study (see Table 49 and Table 65).

The persistent types were from samples from all three decades and ex *S. salar* as well as ex *O. mykiss*. Low virulent types were exclusively ex *S. salar* starting in the 1990's.

**Table 49. Distribution of variants of 56 sequenced IPNV strains; NA=not assigned.**

	Amino acid position		N	%
	217	221		
Virulent	T	A	1	1.79
Low virulent	P	A	18	32.14
Avirulent	T	T	0	0
Persistent	P	T	33	58.93
NA	P	S	1	1.79
NA	A	T	3	5.36

To gain a deeper insight into the VP2 protein, which is the major capsid protein that is recognised by the immune system and contains the sites that determine the virulence of the virus, a Neighbour Network was calculated using only non-homologues amino acids from genogroup V IPNV isolates. The results show that the main cluster is composed of persistent (PT) and low virulent type (PA) isolates. The main cluster is divided in sub clusters. Four samples are outside the main cluster (Figure 22). These four samples are all of the persistent type. Sample 53, the only virulent (TA) isolate, as defined by a specific amino acid residue characteristics obtained from previous research, found in this study is located away from the main cluster near samples IoA9 and 13 which have the amino acid composition PA (see Table 48).



**Figure 22. Non-homologous amino acid alignment based Neighbour Network of the VP2 protein for genogroup V IPNV samples. The scale bar represents the split support for the edges. Sample 53 (red) is the only virulent type (TA). Green samples represent persistent type (PT), purple low virulent (PA) and blue not assigned types (AT and PS).**



### 3.3.2.1.1 Persistency experiment

Previous studies have shown that IPNV attenuates in CHSE-214 cell culture after several passages, which is shown in amino acid changes at VP2 position 217 and 221 e.g. from virulent (T217, A221) to avirulent (T217, T221).

All original diagnostic cultures had been performed on CHSE-214 cells (1-2 passages). To rule out, for the additional pre-sequencing CHSE-214 cell culture (maximum 3 passages), that the VP2 residues observed in the IPNV sequences determined here, were not biased due to cell culture attenuation, 5 randomly selected IPNV isolates were additionally cultured for 3 passages on rainbow trout gonad (RTG-2) cells which are known not to introduce cell culture attenuation mutations in these specific VP2 amino acid positions.

The analysis of the whole genome sequences obtained from the selected isolates cultured on RTG-2 cells showed no differences in amino acid positions 217 and 221 in the VP2 gene whether grown in CHSE-214 or RTG-2 cells, and independent of the number of passages (Table 50).

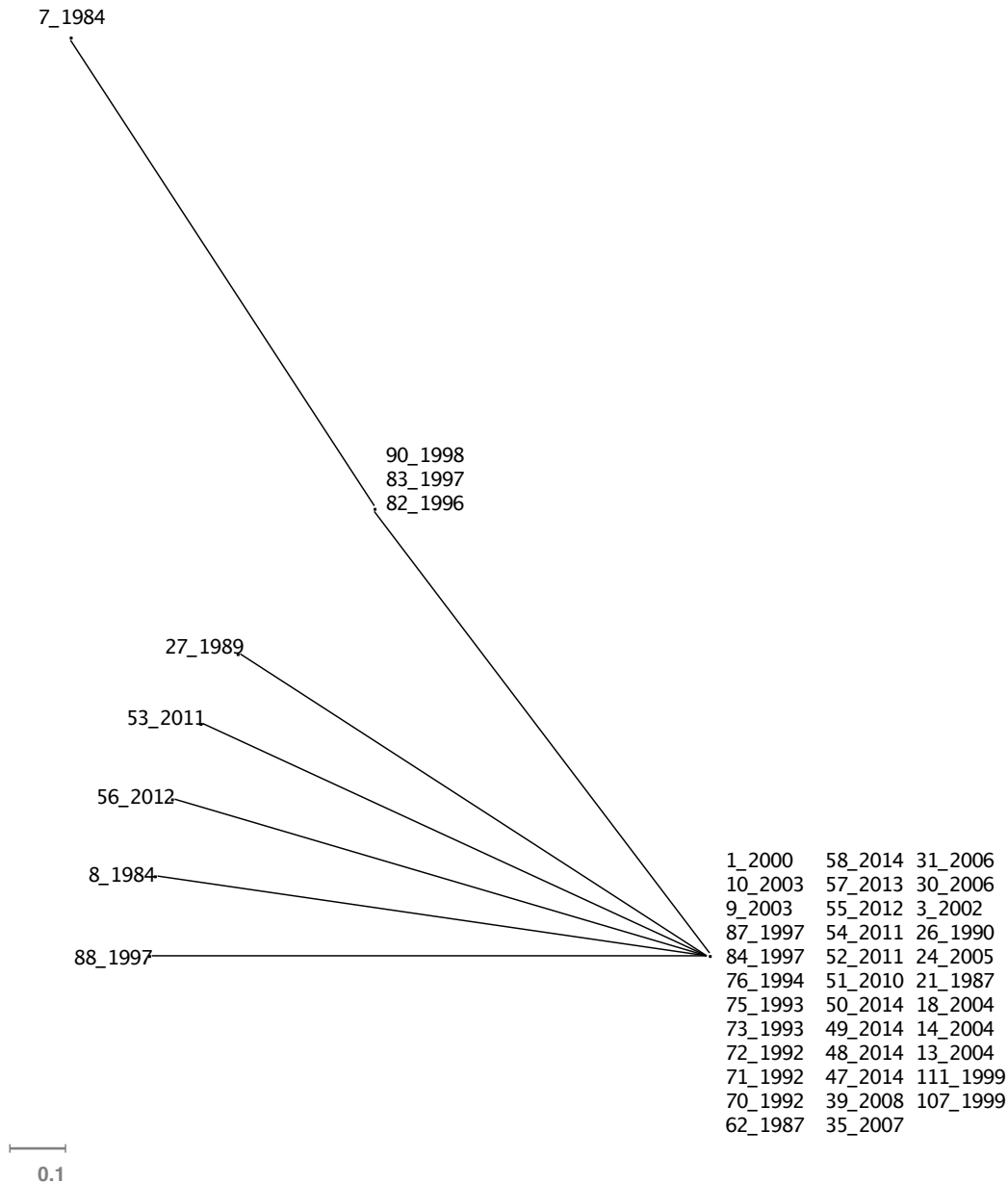
**Table 50. Amino acid composition of five randomly selected samples grown in two different cell lines to prove persistency characteristics. All Isolates  $\geq 2$  passages.**

Sequence ID	Amino acid position	
	217	221
IoA068CHSE-214	P	T
IoA068RTG2	P	T
IoA090CHSE-214	P	T
IoA090RTG2	P	T
IoA076CHSE-214	P	T
IoA076RTG2	P	T
IoA035CHSE-214	P	T
IoA035RTG2	P	T
IoA039CHSE-214	P	T
IoA039RTG2	P	T

### 3.3.2.2 Analysis of features of other IPNV proteins: VP1, VP3, VP4, VP5

VP1 has a characteristic N-terminal Serine at the 5'-end of the RNA-RdRp complex which is confirmed in the sequenced IPNV samples (Table 51). Furthermore, the alignment shows in the GTP- binding motif (G) site 242-252 that samples belonging to the genogroup I have an Isoleucine I243 whereas all other genogroups have a Valine V243. In the RdRp specific motifs I-VI (A-F) amino acid changes are observed in the motifs I (F), IV (B), V (D) and VI (E) that differ between genogroups. It is known that the RdRp lacks the typical GDD motif and has an ADN motif instead, which is confirmed for all 57 IPNV sequences.

The Neighbour Network analysis of non-homologue amino acids of the VP3 protein (genogroup V only) shows one main cluster. Nine samples are outside the main cluster with sample IoA7 (1984) being further away and deriving from the isolates IoA82, 83 and 90, that were all sampled in the 1990s (Figure 23).



**Figure 23. Non-Homologous amino acid alignment based Neighbour Network of the VP3 protein for genogroup V IPNV samples. The scale bar represents the split support for the edges.**

The alignment of VP4 (Table 52) shows the main conserved catalytic residues Serine S179 and Lysine K220 as expected, as well as the catalytic residue F89 in samples belonging to genogroup I or V89 in samples belonging to genogroups II, III and V. Other catalytic sites V91, H93, L184, A221, A223, H225, L229, L231, I232, G233 and D239 do not differ between the genogroups and are therefore not explicitly shown.

In VP5 four sequences all belonging to genogroup V show a truncated protein. One sample is of the persistent type from the 1980's and the other three are of the low virulent type from the 2000's (Figure 24).

**Table 51. Amino acid features of specific IPNV RdRp (VP1) sites and motifs. S (Serine), A (Alanine), D (Aspartic acid), N (Asparagine), I (Isoleucine), V (Valine), H (Histidine), Y (tyrosine), G (Glycine), E (Glutamic acid), K (Lysine), L (Leucine), T (Threonine).**

	<b>Terminal AA</b>	<b>GTP binding site (G) (242-252)</b>	<b>RdRp specific ADN motif (387-389)</b>	<b>RdRp specific site I (F) (315-324)</b>	<b>RdRp specific site IV (B) (462-493)</b>	<b>RdRp specific site V (D) (502-520)</b>	<b>RdRp specific site VI (E) (553-568)</b>
AA site	2	243		316	467 486 492	503 509	566
I	S	I	ADN	K	H I V	G E	I
II/III	S	V	ADN	K	Y L H	E T	V
V	S	V	ADN	L	Y I H	E A	V

**Table 52. Amino acid features of catalytic sites of the IPNV protease (VP4). S (Serine), V (Valine), K (Lysine), F (Phenylalanine).**

	<b>Main catalytic site</b>		<b>Further catalytic site</b>
AA site	179	220	89
I	S	K	F
II/III	S	K	V
V	S	K	V



was performed, assuming both hosts have the same tRNA usage, as the both belong to the Salmonidae family (see 2.6.5 according to Nicholas Di Paola's approach).

In the first analysis a CAI was calculated for each IPNV protein in terms of codon usage adaptation towards the salmon and the rainbow trout genome (see Figure 25). Proteins VP2, VP3 and VP4 have a CAI above 1.00 which indicates adaptation for both, salmon and rainbow trout genomes, although the adaptation in VP2 and VP4 seems to be higher than in VP3. Furthermore, VP2 appears to have a significantly higher adaptation towards the salmon than the rainbow trout genome ( $p=4.867e-05$ ). In VP5 we observe a completely random codon usage for both hosts as the CAI is below 1.00 indicating the virus does not adapt to the codon choice of any of the hosts, although there is a significant difference between the two hosts ( $p=1.544e-08$ ). The polymerase shows only a minor codon usage adaptation.

Analysing CAI over time (see Table 53) a significant increase of codon usage adaptation is evident for both hosts in VP2 and VP4, with the highest increase between the 1980's and 1990's. (Figure 26). In VP3 a significant decrease in codon adaptation over time for both hosts can be observed. VP5 demonstrates a random codon usage constantly over time, but with significant fluctuations between the decades for both hosts. The polymerase shows significant fluctuations over time (Table 53) but does not show signs of overall codon adaptation (Figure 25).

Finally, CAI values were calculated for all proteins of IPNV isolates presenting a persistent and low virulent phenotype in VP2 (see Table 54 and Table 55).

In VP2 we see significant differences between salmon and rainbow trout for low virulent as well as persistent isolates. Significant differences are also obvious between salmon and rainbow trout within the persistent type of samples. In VP3 we see a significant difference between persistent and low virulent samples towards the rainbow trout genome. VP4 shows a significant difference in the persistent type of samples between the two hosts. RdRp shows a significant difference between low virulent and persistent types for each host. VP5 shows significant differences in CAI between salmon and rainbow trout in low virulent and persistent samples. Further there is a significant difference between low virulent and persistent samples for salmon (**Fehler! Verweisquelle konnte nicht gefunden werden.**).

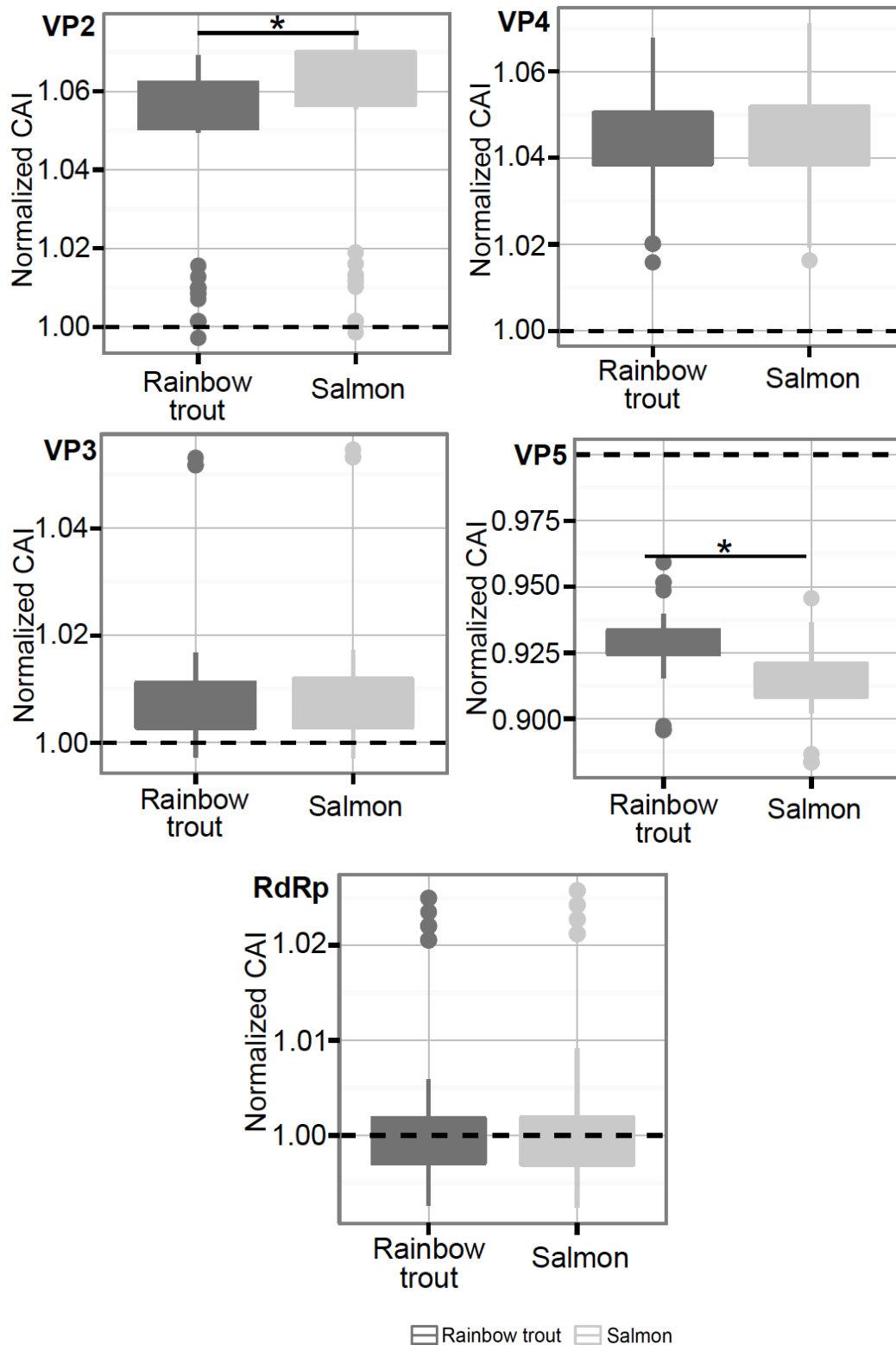
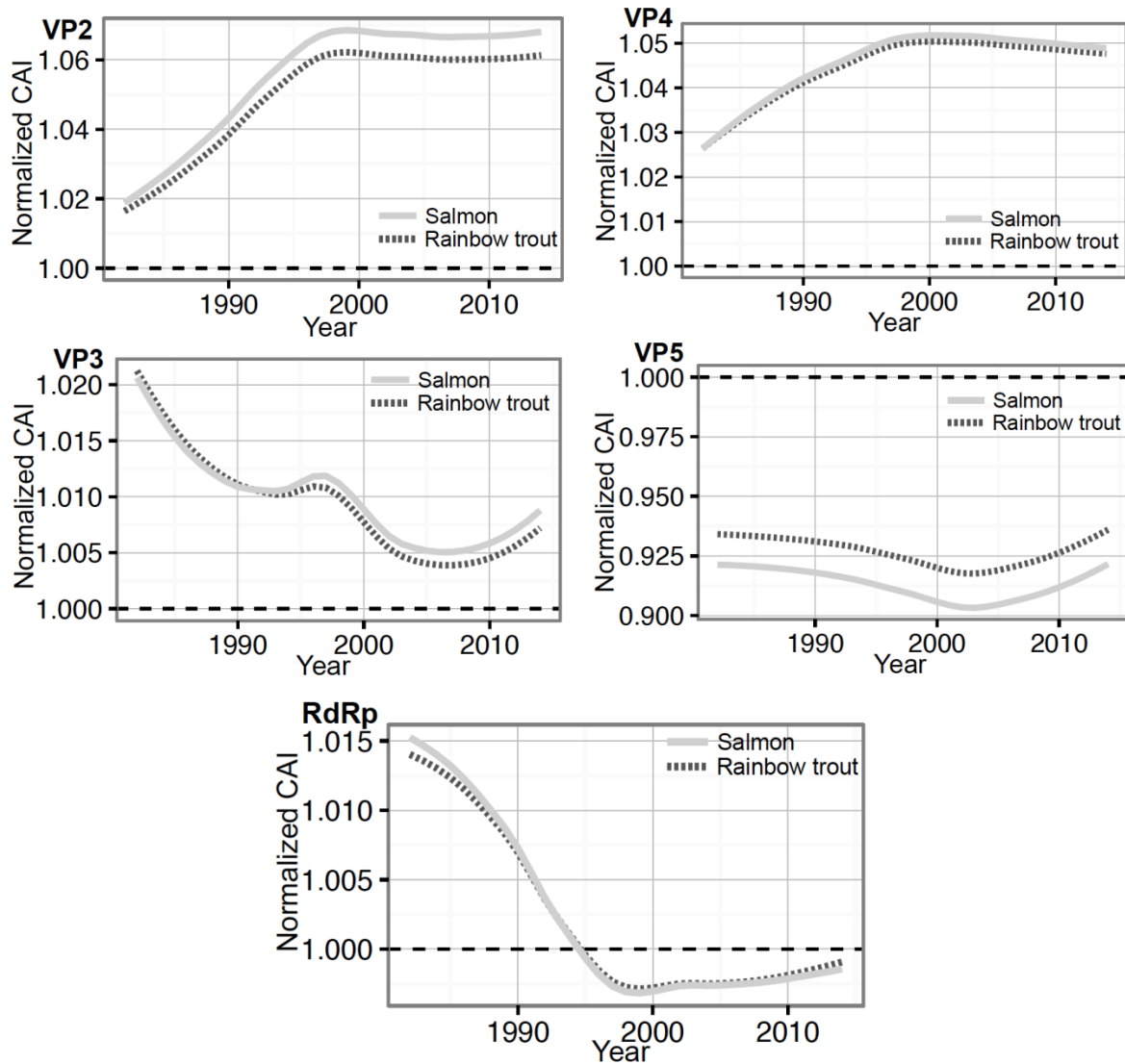


Figure 25. CAI analysis of each IPNV protein to show signs of codon usage adaptation towards the host genome of Atlantic salmon (*Salmo salar*) and rainbow trout (*Onchorhynchus mykiss*). Values above 1.00 show a sign of codon usage adaptation. Stars indicate a significant relation.





**Figure 26.** CAI analysis of each IPNV to show signs of codon usage adaptation towards the host genome of Atlantic Atlantic salmon (*Salmo salar*) and rainbow trout (*Onchorhynchus mykiss*) between low virulent and persistent type of samples. Values above 1.00 show a sign of codon usage adaptation. Stars indicate a significant relation; significance level  $p \leq 0.05$ .

**Table 53. Significance levels of CAI threshold over time of three decades by comparing two decades with each other for both hosts *Salmo salar* and *Oncorhynchus mykiss*; significance level  $p \leq 0.05$ .**

<b>Protein</b>	<b>Host</b>	<b>Comparison</b>	<b>p- value</b>
VP1	<i>Oncorhynchus mykiss</i>	1980s - 1990s	2.165e-05
	<i>Salmo salar</i>	1980s - 1990s	2.165e-05
	<i>Oncorhynchus mykiss</i>	2000s - 2010s	0.003261
	<i>Salmo salar</i>	2000s - 2010s	0.003261
VP2	<i>Oncorhynchus mykiss</i>	1980s - 1990s	2.165e-05
	<i>Salmo salar</i>	1980s - 1990s	2.165e-05
VP3	<i>Oncorhynchus mykiss</i>	1980s - 1990s	2.165e-05
	<i>Salmo salar</i>	1980s - 1990s	0.0006495
	<i>Oncorhynchus mykiss</i>	1990s - 2000s	1.083e-05
	<i>Salmo salar</i>	1990s - 2000s	1.083e-05
	<i>Oncorhynchus mykiss</i>	2000s - 2010s	0.03717
	<i>Salmo salar</i>	2000s - 2010s	0.01985
VP4	<i>Oncorhynchus mykiss</i>	1980s - 1990s	2.165e-05
	<i>Salmo salar</i>	1980s - 1990s	2.165e-05
	<i>Oncorhynchus mykiss</i>	2000s - 2010s	0.003261
	<i>Salmo salar</i>	2000s - 2010s	0.003261
VP5	<i>Oncorhynchus mykiss</i>	1980s - 1990s	2.165e-05
	<i>Salmo salar</i>	1980s - 1990s	2.165e-05
	<i>Oncorhynchus mykiss</i>	1990s - 2000s	0.003886
	<i>Salmo salar</i>	1990s - 2000s	0.002879
	<i>Oncorhynchus mykiss</i>	2000s - 2010s	0.003261
	<i>Salmo salar</i>	2000s - 2010s	0.003261

**Table 54. Significant differences in CAI values between hosts against a sample variant persistent (P) type and low virulent (LV) type; significance level  $p \leq 0.05$ .**

Protein	Host	Comparison	p- value
VP1	<i>Oncorhynchus mykiss</i>	P - LV	0.00037
	<i>Salmo salar</i>	P - LV	0.0004383
VP2	<i>Oncorhynchus mykiss</i>	P - LV	0.014
	<i>Salmo salar</i>	P - LV	0.01627
VP3	<i>Oncorhynchus mykiss</i>	P - LV	0.048

**Table 55. Significant differences in CAI values low virulent (LV) and persistent (P) type of samples against host species; significance level  $p \leq 0.05$ .**

Protein	Virulence type	Comparison	p- value
VP2	LV	<i>Oncorhynchus mykiss</i> - <i>Salmo salar</i>	8.178e-06
	P	<i>Oncorhynchus mykiss</i> - <i>Salmo salar</i>	0.01589
VP5	LV	<i>Oncorhynchus mykiss</i> - <i>Salmo salar</i>	0.0008824
	P	<i>Oncorhynchus mykiss</i> - <i>Salmo salar</i>	7.267e-06

#### 3.3.4 Selection analysis

To identify specific sites in any of the IPNV proteins under a positive selection pressure, a positive selection analysis using the HYPHY package was performed. Initially, the SLAC (Single Likelihood Ancestor Counting) model, with a significance level of  $p \leq 0.1$ , which is able to model variations in synonymous and non-synonymous rates under the MG94xREV codon model, was applied (see 2.6.6).

The results in Table 56 show that only one site, site 217, in the VP2 protein is under significant positive selection. A total of 6 substitutions at this site, with 0 of them being synonymous and 6 non-synonymous was calculated. The determination of site 217 as being under positive selection pressure is supported by the calculated p-value of 0.087.

**Table 56. Results of a positive selection analysis of all IPNV proteins for all 57 sequenced samples using SLAC on HYPHY.**

<b>Protein</b>	<b>Codon</b>	<b>p-value</b>
VP1	FOUND NO POSITIVE SELECTED SITES	
VP2	217	0.087791
VP3	FOUND NO POSITIVE SELECTED SITES	
VP4	FOUND NO POSITIVE SELECTED SITES	
VP5	FOUND NO POSITIVE SELECTED SITES	

Another, more detailed way to find selection sites is to apply MEME (Mixed Effects Model of Evolution) within the HYPHY package under the MG94xREV codon model. MEME hypothesises that sites have been under episodic positive or diversifying selection under a proportion of branches. A p-value of  $p \leq 0.1$  defines positive selection at one site. Table 57 shows that all IPNV proteins have sites of positive selection. In VP1 these sites are distributed across the genome with codon 316 located in the 1<sup>st</sup> and 467 in the 4<sup>th</sup> RdRp specific motif respectively.

VP2 shows sites of selection within the variable domain. Sites 217, 248 and 315 are under positive selection which represent regions determining the virulent character of the virus (see 1.1.2.3.2). VP3 shows sites of selection within the VP3 and the RNA binding domain. Sites in VP4 are not within the active region of the protease. Sites 2-68 in VP5 are all within the truncated region of the protein.

**Table 57 Results of site-specific positive and diversifying selection analysis of all IPNV proteins for all 57 sequenced samples using MEME on HYPHY, applying the MG94xREV codon model. P describes the p-value ( $p \leq 0.1$ ) for positive selection at one site.**

VP1		VP2		VP3		VP4		VP5	
Codon	p-value	Codon	p-value	Codon	p-value	Codon	p-value	Codon	p-value
50	0.01	195	0.02	66	0.10	37	0.09	2	0.01
81	0.00	208	0.02	145	0.10	46	0.01	13	0.19
82	0.00	217	0.00			131	0.04	17	0.03
140	0.04	241	0.00			212	0.09	36	0.09
226	0.05	242	0.03			218	0.05	53	0.07
232	0.04	248	0.00					68	0.09
316	0.00	315	0.09					98	0.08
358	0.07	333	0.07					106	0.04
368	0.05							111	0.08
467	0.01								
687	0.07								
699	0.07								
737	0.09								

### 3.3.5 Dinucleotide composition

In order to get more insight into possible selection on nucleotide level of IPNV sequences obtained from whole genome sequencing, a dinucleotide abundance calculation, focusing on the CG dinucleotide, was performed for IPNV segment A and B using the EMBOSS COMPSEQ online platform. As previously mentioned in 1.3.6, the motif CG or CpG is of big interest as it is a methylation site in vertebrates and besides, suppressed throughout the DNA (Karlin, Doerfler, and Cardon 1994), due to cytosine methylation and other reasons such as higher energy of CpG in DNA and immune escape mechanism of a virus.

**Table 58. Dinucleotide composition of IPNV segment A and B calculated with EMBOSS compseq with focus on CG (bold italic). Results CpGO/E ≤1: underrepresented dinucleotide; CpGO/E ≥1: overrepresented dinucleotide.**

Dinucleotide	Segment A (CpG <sub>O/E</sub> )	Segment B (CpG <sub>O/E</sub> )
AA	0.91	0.98
AC	1.05	1.03
AG	1.10	1.01
AT	0.91	0.96
CA	1.15	1.09
CC	0.99	0.98
<b><i>CG</i></b>	<b><i>0.72</i></b>	<b><i>0.74</i></b>
CT	1.19	1.26
GA	1.20	1.15
GC	0.76	0.81
GG	1.07	1.12
GT	0.98	0.89
TA	0.56	0.66
TC	<b>1.33</b>	<b>1.27</b>
TG	1.22	1.26
TT	0.83	0.78

Results with  $CpG_{O/E} \leq 1$  mean that the dinucleotide is underrepresented in the sequence,  $CpG_{O/E} \geq 1$  means the dinucleotide is overrepresented in the sequence.

The results show that for both segments, the CG dinucleotide is underrepresented (see Table 58, bold italic) compared to previously estimated mean  $CpG_{O/E}=0.88$  for vertebrate infecting dsRNA viruses, by Cheng et al. (Cheng et al. 2013). The most abundant dinucleotide is TC (bold) and the less abundant is TA (italic).

## 4 Discussion

To observe possible changes in the genome of IPNV isolates from the last three decades of IPNV, the whole genome of 57 IPNV TCID50 isolates was sequenced and analysed using next generation sequencing technology on the Illumina MiSeq platform. Two whole genome sequencing protocols, the Goettingen (GOEP) and the Glasgow protocol (GLAP), were compared in terms of terminal nucleotide coverage. Furthermore, the whole genome of one IPNV isolate was recovered from whole genome sequencing on the Oxford Nanopore Technologies MinION sequencer.

### 4.1 Whole genome sequencing protocols

#### 4.1.1 Comparing the Goettingen and the Glasgow protocol

A sequencing protocol for virus RNA had been developed in the group of Dr. Weidmann, University of Goettingen, Germany using the 454 Pyrosequencing approach (Dilcher et al. 2012). For the purpose of this project the virus RNA protocol was transferred to whole genome sequencing (WGS) on the Illumina MiSeq next generation sequence (NGS). The protocol is complex and costly so upon contacting the group of Dr. Emma Thompson at MRC, University of Glasgow, UK, who had developed a WGS protocol for sequencing Hepatitis C Virus (Thomson et al.) directly from whole blood samples (Thomson et al. 2016), protocols were compared and it was decided to test if the Glasgow protocol could simplify and speed up IPNV whole genome sequencing in comparison to the Goettingen protocol. Special attention was on obtaining the full genome with *de novo* assembly and to obtain terminal sequences without the use of specific primers, to make it possible to use the protocol for WGS of potentially any viral genome in a sample.

For this several IPNV isolates were chosen. IPNV RNA was processed with both, GOEP (2.1.2.1.1) and GLAP (2.1.2.1.2) and sequenced in the same way, according to the Illumina Nextera XT DNA Library Preparation Kit (2.1.2.3) as seen in the overview (Figure 11).

The results showed no significant difference in terms of missing nucleotides at the termini of the IPNV genome. Both protocols were not able to yield a full genome after a *de novo* assembly. Manual editing of missing nucleotides was achieved using a reference-based approach. In general reference-based assembly works better than *de novo* at sites of low genome coverage, as in the case of termini regions of the IPNV genome segments. The fact that the GLAP had a significantly lower coverage and reads than the GOEP did not lower its



assembly quality, as *de novo* assembly often works better with lower read numbers for short sequences typically produced by the MiSeq. As the genomes recovered through manual editing from GLAP sequencing were as good as those recovered from GOEP sequencing, it was decided to switch to the GLAP, as it was additionally cheaper and faster. Compared to GOEP it did not have a ds-cDNA amplification step, which potentially can introduce mutations into the viral genome sequencing amplicates.

Altogether, 27 IPNV isolates were sequenced using the GOEP and 30 isolates were sequenced using the GLAP. Comparison of the sequencing results indicated that the GOEP determined significantly more terminal nucleotides (0-60nt missing) than the GLAP results (0-2495nt missing) (Table 63). Interestingly, both protocols were able to yield full length genomes (4 samples each). However, the data demonstrated that the GLAP is suitable for whole genome sequencing of IPNV and was subsequently used based on qualities such as lower costs, less time consumption and fewer amplification steps.

#### 4.1.1.1 *Improving GLAP*

To improve the termini coverage of the GLAP, several changes were introduced to the protocol.

RNA from one IPNV sample was processed in different ways during the ds-cDNA synthesis step (2.2) and sequenced as described in 2.1.2.3.

The bacteriophage T4 single strand DNA binding (SSB) GP32 protein was used to protect ss-cDNA during the ds-cDNA synthesis, as it is known that GP32 stabilises ssDNA intermediates and protects them during the reaction from degradation (Shereda et al. 2008, Shamoo et al. 1995, Bittner, Burke, and Alberts 1979).

Termini specific primers were used in ds-cDNA synthesis to increase the overall number of determined nucleotides at the genome termini.

The use of RNase H after ss-cDNA synthesis was not part of the original GLAP but was considered as it is recommended in the SuperScript III Reverse Transcriptase original protocol. It was assumed to obtain an overall better sequencing results by using the RT enzyme according to the original protocol.

The use of a template premix containing the RNA, the dNTPs and the random hexamers prior to the RNA incubation step was a variation that was suggested in personal correspondence with Dr. Emma Thomson.

Finally increasing the RNA incubation temperature, up to 75 °C and above, was introduced to allow a better unfolding of potential complex secondary structures at the genomic ends, as RNA secondary structures unfold completely at different temperatures depending on the structure complexity (Li, Viereg, and Tinoco Jr 2008, Tinoco Jr 2004).

The results showed that none of the protocol variations resulted in a full IPNV genome after a *de novo* assembly and termini were still missing nucleotides in the range of 60-1092 nucleotides (Table 38 and Table 40). Even improvements in the RNA incubation step prior to the ds-cDNA synthesis by increasing the RNA incubation temperature did only improve termini coverage slightly (87 nt missing compared to 288 nt at 65 °C RNA incubation temperature). Manual editing using a reference-based approach was still needed.

This could mean that the successful unfolding of the secondary structure of the IPNV genome termini might include other factors such as salts in the mix, that are known to play a role in resolving secondary RNA structures (Li, Viereg, and Tinoco Jr 2008, Tinoco Jr 2004).

Usage of RNase H after ss-cDNA synthesis did not improve sequencing results. This could be due to the influence of the RNaseH in the mix that was used for the following ds-cDNA synthesis step as according to the protocol for RNaseH a clean-up step had not to be performed. It might be an option to clean the ss-cDNA synthesis reaction up before using it for the ds-cDNA synthesis step, although this could decrease the template concentration.

Using a 3'-end specific primer in both steps of ds-cDNA synthesis had only a minimal effect on increasing the number of terminal nucleotides determined. It might be considered to work on primer concentration optimisation or optimise the primer sequence especially for IPNV sequencing. However, in general is not of interest to use a specific primer as the goal is to use the WGS protocol for possibly all viruses. Besides, reference-based assembly is still mostly used for viral genome assembly which means that sites with rather poor coverage can still be assembled fully without any gaps by providing reference sequences if they are available.

Additionally, the GOEP includes a step of ligating a 3'-iSP9 adapter (Maan et al. 2007) to the dsRNA but this adapter sequence was never found in the obtained data, indicating that the ligation did not work for IPNV. This suggested that RNA termini maybe are not properly accessible due to the secondary structure of the terminal RNAs of the IPNV genome.

The number of missing terminal nucleotides was smaller when additionally using GP32 in the ds-cDNA synthesis step (60 missing nucleotides) or during both steps (70 missing

nucleotides) but not in the ss-cDNA step (124 missing nucleotides), indicating that it has only a small impact during the ds-cDNA synthesis step (see Table 38). This could be due to primary incomplete unfolding of the template RNA to begin with. The concentration of the GP32 protein in the mix was possibly higher or lower than needed for optimal process. Another possibility could be the need of further SSBs in order to unwind the RNA, stabilise the primary RNA structure thus, optimise the ds-cDNA synthesis in terms of recovering all nucleotides in the genome termini.

Taken together, comparing two WGS protocols and improving the GLAP in order to increase the termini coverage of the IPNV genome segments, showed that RNA secondary structure could be the cause of failing to achieve recovering of a full IPNV genome. The results made clear that changes in the GLAP that affect the secondary structure like the use of the GP32 protein and higher RNA incubating temperatures showed, even if not statistically significant, improvement resulting in fewer missing terminal nucleotides. WGS of betanodavirus samples using NGS (MiSeq) with the standard GLAP showed a higher terminal coverage in general than observed for IPNV (personal communication with Dr. Benjamin-Lopez Jimena). betanodavirus also is a bi-segmented virus (RNA1 and RNA2) and it was observed that RNA2 was recovered completely in almost every sample, whereas RNA1 was missing just a few terminal nucleotides, thus supporting the idea that RNA secondary structure is the major cause for not recovering the terminal sequences.

#### 4.1.2 MinION sequencing

A publication on whole genome sequencing of Influenza virus on the MinION platform (Wang et al. 2015) gave the idea to try to sequence IPNV on the MinION to see if this would recover a full IPNV genome after a reference-based assembly, without any missing nucleotides in the terminal genome regions. The data from the study done by Wang et al. showed that the whole genome of an influenza virus was >99 % identical to the genome obtained by Sanger sequencing, which proved the accuracy of the MinION device as such. Although the group used termini specific primer to amplify the viral template (PCR sequencing approach), the sequencing technology of the MinION itself is very different. As a ssDNA strand passes through a pore and tetranucleotides are transduced into a signal, this might potentially result in better covered terminal regions.

RNA from one IPNV sample was processed with the GLAP and sequenced on the MiSeq (2.1.2.3) and on the MinION (2.4). Comparing sequence yields showed no improvement, as nucleotides at the IPNV genome termini were missing with both sequencing approaches. The mean coverage of the MinION-obtained sequence was more than 30 times lower than for the MiSeq-obtained sequence (Table 44). The number of missing terminal nucleotides was only 10 for the MiSeq yield compared to MinION yield with 36 (Table 45), although the genome assembly for the MinION data was done using a reference genome in the first place. The MinION technology allows the sequencing of very long fragments. It was therefore of interest to see if a change in the magnetic bead to sample ratio in the post ds-cDNA synthesis purification step could increase the fragment size. To prove that, one randomly selected betanodavirus sample was processed as describe in 3.2.2.1.1 with a change of the beads to sample ratio from 1x to 0.6x. The library preparation protocol for the MinION could be improved successfully in terms of recovering longer fragments shifting the average fragment size from 642 bp to 770 bp. This shows that the MinION sequencing protocol can be adapted to individual sequencing needs.

MinION sequencing was also performed with betanodavirus samples, although with a slightly changed library preparation protocol and yielded full recovery of RNA2 whereas RNA1 was still missing a few terminal nucleotides.

The use of two different NGS technologies for betanodavirus as well as for IPNV WGS indicates that not the ds-cDNA synthesis protocol, the library preparation or the sequencing technology but rather the RNA structure of the virus is the factor that determines the recovery of a full genome with fully recovered terminal sites.

Comparing the GLAP to other published WGS protocols for NGS reveals that usually termini specific primers for segment A and B are used in the cDNA synthesis step in order to obtain the full genome of Birnaviruses, as shown in a study of Lu et al. (Lu et al. 2015).

This study aimed at optimising a protocol avoiding specific terminal primers in order to have a WGS protocol that can be used for any possible virus. The GLAP was indeed, successfully used to sequence the whole genomes of more IPNV samples from Turkey, white spot syndrome virus (WSSV, dsDNA genome), piscine myocarditis virus (PMCV, dsRNA genome), and viruses from the *Betanodaviridae* family (ssRNA(+)) genome); data not shown as it is unpublished work by Dr. Benjamin Lopez Jimena and not described in the the introduction and methods section.

## 4.2 Phylogenetic analysis interpretation

To analyse the phylogeny of IPNV isolates from Scottish aquafarms from three decades, the whole genome of 57 IPNV isolates was sequenced using NGS. Phylogenetic trees were calculated using the BEAST package and double checked with phylogenetic trees calculated via RAXML (not shown). The results showed that over 70 % of the sequenced IPNV isolates in this study belong to genogroup V, which was also seen in studies in Finland between 1985-2015 (Holopainen, Eriksson-Kallio, and Gadd 2017), in Scotland between 1979-2004 (Bain, Gregory, and Raynard 2008), in Ireland between 1993-2013 (Ruane et al. 2015) and in Norway between 2002-2003 (Mutoloki and Evensen 2011). Only a few IPNV isolates belong to the genogroup I, II or III.

In five samples the results showed that segment A belonged to genogroup III but segment B to the genogroup II, indicating reassortment. The samples are dated from 1987 to 1989. Three samples (IoA63, 66, 68) were extracted from salmon and in case of IoA64 and 65 the host is not known.

Natural reassortment in wild IPNV isolates resulting in a combination of serotypes WB/Ab (I/II) (see Table 2 for IPNV classification) was observed by Romero-Brey et al., in 1999 off the Canadian east coast isolated from Atlantic cod (*Gadus morhua*), Greenland halibut (*Reinhardtius hippoglossoides*), Witch flounder, (*Glyptocephalus cynoglossus*), American plaice (*Hippoglossoides platessoides*), Deepwater redfish (*Sebastes mentella*), Onion-eye grenadier (*Macrourus berglax*), Blue antimora (*Antimora rostrata*) and Atlantic wolf fish (*Anarhichas lupus*) (Romero-Brey et al. 2009). More cases of natural reassortants were found between 2010 and 2011 in 29 different wild fish in Spain and represented WB/Ab (I/II) and WB/Ja (I/I) combination (Moreno et al. 2014). Cases of natural reassortment were also reported in IBDV, a Birnavirus infecting chicken (Wei et al. 2006). No indication for reassortants of the combination II/III in IPNV was found in literature.

Interactions between wild and farmed fish are found for many infectious diseases. IPNV is found in wild fish near farming sites but also up to 50km away, persistently for many years, which indicates that it is passed on from farmed to wild animals, through escaped animals. There is evidence for the opposite route as several pathogens were detected in wild fish near farming sites before there were any signs of the disease in the farm or even before a farm was built. Once IPNV is transmitted from wild to farmed animals it can be enriched in the dense farmed population and reach outbreak potential and can potentially spill over to

wild populations (pathogen spill back) (Wallace, McKay, and Murray 2017). The high host number allowing massive replication is a suitable ground for selection of isolates with specific properties, and reassortants, that could be deriving from cross-transmissions between farmed and wild fish.

The detected IPNV reassortants may have been generated in wild fish and then infected the farmed salmon and trout. However, it is also possible that the reassortant strains in this study derived from 2 types of IPNVs co-infecting farmed salmon and trout.

The IPNV reassortants found here could be a product of early aquaculture but they were obviously not able to persist. Still, reassortants could occur anytime due to the use of non-sterilising IPNV vaccines used on farming sites, which allows IPNV replication in vaccinated salmon and acquiring with different IPNV genogroups (co-infection) (Julin, Mennen, and Sommer 2013). Viruses with more than one genomic segment possibly use reassortment due to selection pressure to rearrange genomic properties in order to overcome elimination, which is perfectly documented in human influenza virus (Carrat and Flahault 2007, Shao et al. 2017).

Only one virulent Sp isolate (genogroup V) was found according to the criteria published by Santi et al (Santi, Vakharia, and Evensen 2004b). The two most represented variants were the persistent (58.93 %) and the low virulent (32.14 %) type.

Similar results were obtained in a study of 36 IPNV isolates from Scotland between 1999 and 2004 by Bain et al. (Bain, Gregory, and Raynard 2008) where 41.67 % of the isolates were of the persistent, 50.00 % the low virulent type and only three isolates showed a virulent amino acid composition. Surprisingly the low virulent P(217) A(221) isolates showed high mortality in field and challenge experiments (Bain, Gregory, and Raynard 2008). Furthermore, it is known that variants with the persistent P(217) T(221) motif cause high mortality in rainbow trout (Ahmadivand et al. 2018). Furthermore, it was reported by Ruane et al. that all IPNV isolates from Ireland (including clinical IPNV) from 1993 to 2013 were either P(217) T(221) or P(217) A(221) (Ruane et al. 2015).

Unfortunately, there is no information available if the isolates in this study caused an outbreak or on the mortality induced, so it is not possible to verify the pathogenic nature of the isolates. However, the key information that can be drawn from this study and Bain's observation is that pathogenic IPNV viruses are prevalent in vaccinated farmed salmon and trout.

Interestingly, only two amino acid flips are required to move from the persistent to the virulent type in VP2 (see Table 59). This could be of importance as there could be a potential for a high number of virulent strains if such flips occur due to stress in fish or selection mutations, as a high density of hosts favours such phenomena.

A study by Skjesol et al. in 2004 in Norway demonstrated that such flips can occur from pathogenic to low virulent in IPNV isolates from field outbreaks with T(217) A(221) changing to T(217) T(221) in CHSE-214 cell culture after 2-3 passages and to P(217) A(221) after 13 passages post infection (Skjesol et al. 2011). This shows that IPNV can accumulate mutations which result in decreased replication success, thus favouring fish survival but indirectly ensuring its own survival at low replication level.

The incidence of the virulence types described here suggests that the amino acid switches in IPNV can happen from low virulent/persistent strains to high virulent ones.

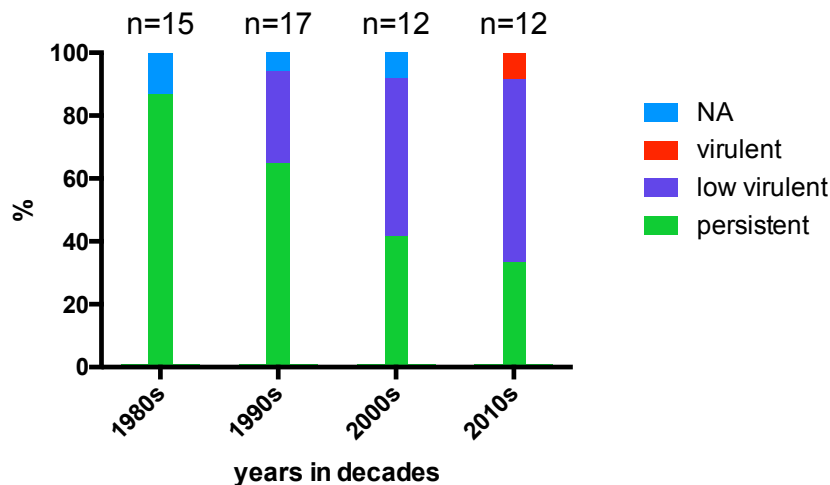
**Table 59 Distribution of variants of 56 sequenced IPNV strains with indicated flips leading to virulent strains.**

Amino acid site			
	213 (217)	217 (221)	
persistent	P	T	58.93 %
		↓	
low virulent	P	A	32.14 %
	↓		
virulent	T	A	1.79 %

It can be assumed that the high number of persistent and low virulent strains enables the virus to replicate on a certain level without harming the host and avoiding a strong immune response. The host can potentially shed the virus for a longer period and over bigger distances (see 4.3 for more). This was already suggested for shrimp viruses, where it was shown that changes in the codon usage allow a shift towards lower virus replication to avoid hosts immune response (Tyagi, Kumar, and Singh 2017).

Looking at the data presented in this study, we see that most isolates are of the persistent type and only persistent types prevail in the time before vaccine introduction in 1995 (see Table 65). After 1995 the data shows that there seemed to be a switch from persistent

P(219) T(221) to low virulent types P(217) A(221), with continuous increase of the low virulent type over the decades (see Figure 27). Before 1995 highly virulent strains were not detected, after 1995 one isolate from 2011 shows the high virulent amino acid composition T(217) A(221).



*Figure 27. Distribution of variants of 56 sequenced IPNV strains in each decade.*

This could mean that the amino acid switches proposed in **Fehler! Verweisquelle konnte nicht gefunden werden.** are potentially happening in farmed salmon. The first switch from persistent to low virulent types occurred after the introduction of the IPNV vaccine. The next change could have been triggered by the introduction of the QTL fish in 2010 which are resistant to IPNV. They specifically block the entry of the T217, A221 variant in infected fish. This possibly introduced an enormous selection pressure in terms of the binding affinity of IPNV, as it was shown that the QTL region contains genes coding for epithelial cadherin (cdh1) gene, which is involved in virus internalization into the cell (Moen et al. 2015), and may have yielded the only high virulent isolate from 2011. To confirm this, it would be necessary to study more recent isolates as this assumption is based on a small dataset. Recently, persistent P217, T221 types have been isolated from groups of fish carrying the IPN-QTL in northern Norway (Oystein Evensen, personal communication). As we know from previous research, Alanine at position 221 results in lower binding affinity and therefore better virus release and spreading (Mutoloki et al. 2016). So a mutation from P217, T221 (persistent) to P217, A221 (low virulent) would enable IPNV to replicate better in host cells, selecting for low virulent types. To understand whether, in spite of this, T217, A221 IPNV



types are currently evolving in QTL fish in Scotland, it would be necessary to study more recent isolates.

However, this study and previous research might show that IPNV is evolving towards mutations, such as Alanine at position 221, to reduce binding affinity.

Another way to explain the high prevalence of persistent strains is that the IPNV vaccines commercially available from 1995 must all be failed vaccines. The efficiency of IPNV vaccines was assessed mainly in terms of their ability to suppress disease symptoms and improve fish survival. In general vaccines can fail for different reasons: (i) don't induce optimal immune response, (ii) an already existing infection, (iii) not all strains are included in the vaccine, (iv) not 100 % efficiency against the agent, (v) or vaccine-vaccine interaction with antigenic interaction in multivalent vaccines (Heininger et al. 2012).

Julin et al. showed that IPNV isolates extracted from vaccinated fish from salmon farms in Norway were almost all of the high virulent type T(217) A(221) and when these isolates were used in challenge experiments in naïve fish they induced mortality of up to 56 %, even isolates with the low virulent P(217) A(221) motif induced a mortality of more than 30 %, showing that vaccinated fish harbour virulent strains. One sample showed attenuation in the passage experiment changing from high virulent T(217) A(221) to avirulent T(217) T(221) and when challenged with the attenuated type, it showed highly reduced mortality (Julin, Mennen, and Sommer 2013). But in general, the study showed that field isolates from vaccinated fish induced very high mortalities in unvaccinated populations. In one case the IPN-vaccine had apparently no effect at all.

All of these observations indicate failure of the IPN-vaccines in the sense that they are non-sterilising vaccines which do not inhibit IPNV replication.

It also shows that IPNV is very capable of changing the motifs that determine its virulence after passages of the virus, or due to stress and hosts immune response. Julin et al. showed that both, infections with the high virulent or low virulent IPNV strains in Atlantic salmon induced mortalities after sea water transfer, considered as stressful for fish due to smoltification (Julin et al. 2015). Similar was shown by Mutoloki et al. when P(217) T(221) isolates caused subclinical infections after sea water transfer and T(217) A(221) isolates clinical infections (Mutoloki et al. 2016). Finally, Gadan et al. showed that a reversion of a non-virulent T(217) T(221) IPNV strain to a high virulent T(217) A(221) IPNV strain can be induced by stress in salmon (Gadan et al. 2013).

Therefore, IPNV clearly shows the potential to change its virulence pattern in both directions: due to stress to a high virulent type or due to attenuation to a low virulent type, making **Fehler! Verweisquelle konnte nicht gefunden werden.** indicating the emergence of more and more high virulent strains a possible scenario.

Humoral and cellular response induced by vaccines are usually directed against viral epitopes, in this case against the IPNV VP2 epitopes (1.1.6). RNA viruses show a high evolution rate, due to error prone RdRp, with substitution rates ranging from  $10^{-2}$  to  $10^{-5}$  substitutions (mutations)/site/year, whereas non-viral evolving rates are between  $1.2 \times 10^{-8}$  and  $1.2 \times 10^{-10}$  substitutions/site/year (Hanada, Suzuki, and Gojobori 2004, Silva et al. 2013). This type of mutation rate allows for the selection of variants resistant to control via vaccination, i.e. of variants escaping immune response of the host. IPNV as an RNA virus has therefore a high variability and mutation rate (Pereira and Amorim 2013, Domingo and Holland 1997) and is quite capable of creating variants to escape vaccines. Skjesol et al. observed that VP1, VP2, VP3 and VP5 can acquire altogether more than 14 sites per 1 kb and that VP2, VP3 and VP5 contain equal or more non-synonymous than synonymous mutations, especially in the hyper variable region of VP2, indicating selection towards low virulent variants (Skjesol et al. 2011).

Although usually high rates of non-synonymous mutations indicate better viral reproduction it seems that IPNV is actually developing towards co-existence with the host.

This indicates that it is possible that vaccination escape variants are observed in this study. It is very likely that the virulent isolates used for the development of the vaccines did not have the P(217) T(221) or P(217) A(221) motifs. The IPNV population was subsequently selected for these variants in order to overcome the vaccine control. Only recently research is concentrating on the PT motif for vaccine development, as it was shown to induce anti-IPNV antibody production and was able to protect against IPNV (Ahmadivand et al. 2018).

### 4.3 Codon adaptation of IPNV to salmonids

In order to see if IPNV is adapting to salmonid host codon usage (*S. salar* and *O. mykiss*), to improve IPNV replication in these hosts and to undermine immune response, a codon adaptation index (CAI) was calculated for each of the 5 IPNV proteins towards both hosts over time and for both, persistent and low virulent strains.

The results indicate that the major capsid protein VP2 and the protease VP4 show the highest CAIs for adapting their codon usage to both hosts, although the adaptation of VP2 in *S. salar* is significantly higher than in *O. mykiss*. This is not observed for VP4 (see Figure 25).

This difference could be explained by the large number of salmon (162817 tons produced in 2016) in comparison to rainbow trout (8096 tons produced in 2016) farmed in Scotland with an approximate 30 times higher salmon production per year (Munro and Wallace 2017). This allows for a much higher number of IPNV replication cycles in the salmon population and consequently more scope for codon adaptation. The fact that VP4 seems not to fully optimise codon usage could be explained with a study done by Skjesol, which showed that VP4 has the lowest mutation/kb rate with preferred synonymous mutations (Skjesol et al. 2011). The adaptation of the codon usage in VP2 and VP4 ensures an efficient translation of the capsid protein essential for packaging and the protease, which enables maturation. Interestingly, VP2 and VP4 both show an increase of CAI over time (see Figure 26) with the biggest increase in salmon farming between the 1980s and 1990s of 98.15 %, compared to 74.91 % production increase between 1990s and 2000s and 16.35 % production increase between 2000s and 2010s (according to Scottish Fish Farm Annual Production Surveys 1980, 1990, 2000 and 2010). Still, this has to be viewed critical, as the sample sizes varies between the decades in this study.

During this time period the IPNV vaccine was introduced and salmonid farming was growing yearly, thus putting a pressure on the IPNV genome and allowing for viral enrichment in the dense farmed population with a suitable ground for selecting isolates with specific properties.

The CAI for the RdRp (VP1) is very low (Figure 25) which is plausible, as the RdRp is strongly conserved within RNA viruses (Shwed et al. 2002) and rarely adapts because changes in the coding sequence could result in non-synonymous mutations, leading to fatal changes in the RdRp protein structure and a possible decrease or loss of the function of this enzyme. It

correlates with findings that the IPNV RdRp has (following VP4) the lowest mutation/kb rate with preferred synonymous mutations (Skjesol et al. 2011).

Previous research showed that VP3 has the highest mutation/kb rate of all IPNV proteins with preferred non-synonymous mutations (Skjesol et al. 2011).

In this study the minor capsid protein VP3 showed only minimal codon adaptation. One of the functions of the multifunctional VP3 is to block the hosts' immune response during infection by suppressing IFN $\alpha$  induction. A small amount of VP3 will still allow for a low level innate immune response, whereas a bigger amount of VP3 could result in a potentially defenceless host that can even succumb to disease too fast. This would be contra productive for virus survival, as it would be more efficient if the host could shed the virus for a longer period over longer distances. Therefore, a non-optimised VP3 as observed here might be an important toggle to keep innate immune response at bay for just the right low level of IPNV replication.

The CIA values for VP5 are lower than 1, meaning that codon adaptation is not happening.

This is quite interesting as VP5 shuts down hosts apoptosis system so the host does not initiate apoptosis of infected cells and together with VP4 inhibits IFN signalling (Skjesol et al. 2009, Hong, Gong, and Wu 2002). By this, IPNV enhances cell survival and undermines the immune response for the benefit of viral proliferation. VP5 is a homologue of the anti-apoptotic Bcl-2 family and it therefore has to retain the ability to mimic this conservative protein.

Previous research conducted by Do Yew et al. showed that in IBDV the VP5 ORF has a different base usage than the VP2 ORF indicating that through mutation the original VP5 ORF was overprinted to create a novel gene (de novo gene) (Do Yew et al. 2004) (Sabath, Wagner, and Karlin 2012). Similar results were shown for IPNV in a study by Pavesi et al., 2013, where VP2 showed a codon usage more similar to the viral genome than to VP5 (Pavesi, Magiorkinis, and Karlin 2013). De novo genes are usually more connected to viral pathogenicity than to structure and replication. My study of CAI for VP2 and VP5 confirms that VP5 shows no adaptation to the hosts codon usage, due to its unusual codon usage, deriving from an overprinting strategy and not being essential for replication. Looking at this, and the fact that some natural IPNV strains were found to lack VP5 expression while still replicating and inducing mortality, the protein appears to be not essential for virus survival

and therefore there is might be no need to adapt the codon usage for this protein to the hosts' codon usage to increase replication efficiency.

This could mean that VP5 might have evolved at farming sites due to the high replication level possible in densely farmed populations and the selection pressures described above, i.e. to maintain low level replication in vaccinated hosts.

Bahir et al. analysed CAI values of mammalian viruses and showed that viral structural genes show the highest similarity in codon usage to the host to optimise translational fitness.

Genes that encode proteins that make up the capsid but are not involved in hosts immune system interaction show higher CAI values than non-structural proteins such as the RdRp (Bahir et al. 2009).

In this study VP2, which is a capsid protein but is also interacting with the host's immune system, shows the highest CAI. This only partly correlates with Bahir's observation. As a capsid protein VP2 adapts to the host's codon usage in order to increase translational efficiency but it also contains epitopes that are recognised by the hosts immune system. The selection analysis clearly showed however that the variable domain within VP2 at position 183-335 that plays a role in immune system response is under selection pressure and prone to changes in order to bypass the immune system.

A study by Tello et al., 2013 (Tello, Vergara, and Spencer 2013), analysing CAI for ISAV (ssRNA(-) genome) proteins showed that the nucleoprotein NP, the matrix protein M1 and the non-structural protein 1 NS1 had the highest CAI values. The high CAI values of M1 and the nucleoprotein correlate with findings by Bahir, that structural genes show the highest adaptation to the host's codon usage. NS1 however showed higher CAI values than the matrix protein M1 and M2 for some analysed ISAV samples indicating that fish/aquatic viruses may not follow the general adaptation models that were suggested for mammalian viruses.

A possible explanation could be that the immune response of fish differs to that of mammals and therefore the selection pressure on viral proteins that interact with the fish hosts immune system is different than described for interaction with mammalian hosts.

A similar observation was made in a study of CAI in shrimp viruses (Tyagi, Kumar, and Singh 2017). Almost all viruses in that study had proteins that showed an adaptation to their host's codon usage, as showed by elevated CAI values. In case of the hepatopancreatic parvovirus and infectious hypodermal and hematopoietic necrosis virus only non-structural proteins

had higher CAI values showing again that aquatic viruses adapt to their host in a different way from that shown for mammalian viruses. Interestingly, the only analysed dsRNA shrimp infecting virus in that study showed no codon usage adaptation at all, which supports the possibility of different codon usage preference between vertebrates and invertebrates infecting viruses.

Cheng et al. looked at the difference in the CpG content of different viruses, which beside the CAI can be considered as another adaptation feature and compared it to that of their hosts. This revealed that invertebrate infecting dsRNA viruses had a higher CpG abundance than vertebrate infecting dsRNA viruses but only because the  $CpG_{O/E}$  of invertebrates themselves is 1.02 compared to 0.47 for vertebrates (Cheng et al. 2013). This suggests indeed a difference in host adaptation between vertebrates and invertebrates infecting dsRNA viruses, which could explain why the shrimp infecting dsRNA virus did not show elevated CAI values but IPNV does.

Furthermore, it is known from Epstein-Barr virus (EBV) that it de-optimises its codon bias when it is in the latent stage in order to replicate on a non-detectable level (Karlin, Blaisdell, and Schachtel 1990). A de-optimisation was also found in shrimp viruses, by which the virus does not compete with the host translational system, thus possibly supporting a low replication strategy to avoid immune defence (Tyagi, Kumar, and Singh 2017).

It is possible that IPNV uses a similar mechanism, as most proteins show hardly any host codon adaptation and VP5 suggests even a de-optimisation of its codon bias due to CAI values lower than 1. This correlates with the finding that most isolates belong to the persistent type. Basically, viral replication on a low level ensures host survival and therefore, survival of the virus.

CpGs are also pathogen associated pattern molecules (PAMPs) recognised by innate immunity receptors. Small RNA viruses show CpG suppression in their genome which could enhance their replication (Karlin, Doerfler, and Cardon 1994). This was observed in a dsRNA shrimp virus (Tyagi, Kumar, and Singh 2017) with the  $CpG_{O/E}$  value of 0.61.

The CpG abundance calculation done in this study shows that the CpG dinucleotide is under-represented in both IPNV segments with a  $CpG_{O/E}=0.72$  for segment A and a  $CpG_{O/E}=0.74$  for segment B (Table 58; CG value). These calculated values are lower than the previously estimated mean  $CpG_{O/E}=0.88$  for vertebrate infecting dsRNA viruses, by Cheng et al., which in that study meant a normal or an over-representation of CpG motifs. The findings here

correlate with previous studies which showed that small RNA viruses show a suppression of CpG in order to escape immune recognition, as Toll-like receptor 9 is known to recognise CpG motifs in dsRNA and to initiate innate immune response (Tyagi, Kumar, and Singh 2017). Non-methylated CpG have an immune stimulating effect (Goldberg, Urnovitz, and Stricker 2000, Chen et al. 2001, Cheng et al. 2013), as it was shown for the Toll-like receptor 9, which initiated an innate immune response after binding of CpG motifs (Tyagi, Kumar, and Singh 2017). The immune escape suggestion might have been proven, for fish as hosts, in a study done by Jorgensen et al. (Jørgensen et al. 2001), which showed that Atlantic salmon leucocytes produce antiviral cytokines after stimulation with unmethylated CpG.

Although the immune escape theory regarding the CpG underrepresentation is ambiguous, the suppression of CpG in IPNV may suggest an immune escape mechanism. IPNV has acquired this feature before the year 1982, as a separate calculated  $CpG_{O/E}$  for segment A and B for the oldest sample in our study from that year yielded 0.73 and 0.74, respectively (not shown in results).

#### 4.4 Selection analysis interpretation

In order to see if sites of the IPNV genomes are under selection pressure a positive selection analysis on each IPNV protein sequence alignment of the 57 sequenced IPNV samples was performed using two different algorithms.

The results from the SLAC calculation showed that only VP2 has a site (codon 217) which is under positive selection pressure. The MEME calculation on the other hand revealed possible sites under positive selection pressure in each IPNV protein.

The site 217 in VP2 is under positive selection pressure calculated with both approaches. This is one of the sites described by Santi et al. important for virulence determination. The fact that this site is under pressure correlates with previous findings that this site is variable and changes in the codons determine the virulence of the virus. Other sites in VP2 obtained with MEME show that 5 of them (217, 241, 242, 248 and 315) are within the variable domain or represent sites that determine the virulence character of the virus. This indicates that these sites are constantly under pressure to change in terms of virulence behaviour. It correlates with the fact that there is a tendency to changes between different types of strains and the possibility to create highly virulent strains with only two amino acid flips. As described above, one of the flips is from persistent to low virulent strains (**Fehler! Verweisquelle konnte nicht gefunden werden.**).

As this site is under pressure such a flip is very likely to happen. The variable region also contains the neutralising epitopes so it is possible that pressure coming from the immune system results in changes within this region in order to escape immune response resulting in changes to virulence beneficial to low level virus replication. Evidence for such flips to happen is seen in the network analysis of non-homologous amino acid sequences within the hypervariable region (HVR) of the genogroup V isolates of the VP2 protein (Figure 22). The virulent (TA) isolate 53 separates from the main cluster of persistent (PT) and low virulent (PA) strains by acquiring the second amino acid change. The network shows that the main cluster of persistent and low virulence isolates evolves in the constraints of a certain sequence space from which as shown by samples 72 or 14, variants with specific new amino acid patterns can emerge (Table 48). Because of the pressure of the hosts immune system, IPNV escapes through low virulent or persistent infections, and the main cluster of the network is composed of these types of isolates. But factors such as stress, not sterilising vaccines and viral enrichment in the densely farmed populations lead to the emergence of



isolates with changes in the genome that can lead to a change in the virulence properties, even towards full virulence as seen in isolate 53 from 2011 as clearly indicated in the network analysis.

VP1 (RdRp) sites calculated under positive selection pressure were mostly found outside conserved motifs I-VI. Only site 316 which is in the motif I and site 467 in the motif IV experience positive selection pressure. These conserved motifs are crucial for the RdRp function (Graham et al. 2011) and it is not surprising that there can be only very few or even no changes in the codons.

VP3 showed one site under positive selection pressure within the self-binding domain and one within the dsRNA binding domain. No sites were under selection pressure within the RdRp binding domain. This indicates that this region is quite conserved as it binds to the RdRp, which is a conserved protein itself, meaning that changes in this site can negatively affect the RdRp binding. The self-binding site and the RNA binding sites are possibly more flexible to changes as certain differences in the amino acid structure might not disturb the binding efficiency.

In VP4 no catalytic active sites are affected by positive selection. It is quite obvious that these sites are very crucial to protease functioning and cannot undergo changes.

In VP5 the first 6 sites under positive selection pressure are within the truncated region of the protein indicating high local variability. As stated above VP5 appears not being essential, some IPNVs have a truncated VP5, and some IPNVs lack VP5 but still are capable to induce mortality. The role of VP5 remains to be elucidated.

#### 4.5 Conclusion

This project made it possible to determine the phylogeny and genomic changes of IPNV in Scotland by sequencing and analysing the whole genome of diagnostic IPNV isolates spanning 3 decades since 1982, using next generation sequencing technology.

Two different protocols were compared, with GLAP being suitable for whole genome sequencing of viral genomes on the next generation sequencing Illumina MiSeq platform. During the protocol optimisation and sequencing of IPNV on the Oxford Nanopore Technologies MinION platform it could be shown that the secondary structure of IPNV RNA has to be the reason why there were always termini nucleotides missing after a *de novo* IPNV genome assembly.

The phylogenetic and genomic analyses of IPNV done in this study indicate that selection pressure by the hosts immune response and failed vaccines result in escape mechanisms of IPNV in farmed salmon and rainbow trout which are seen in:

- (i) changes within the hypervariable region in the VP2 protein, thus changing virulence beneficial to low level virus replication
- (ii) adapting the codon usage of the major capsid protein VP2 and the VP4 for a better maturation
- (iii) reassorting the dsRNA segments between different genogroups
- (iv) occurrence of a virulent (TA) strain (IoA53 from 2011)

These mechanisms are possible due to:

- (i) high mutation rate of IPNV
- (ii) high virus replication and enrichment in dense fish populations on farms and interactions with wild animals
- (iii) non-sterilising IPNV vaccines which enable IPNV replication in farmed vaccinated fish.

The high prevalence of persistent and low virulent strains shows that IPNV is evolving towards the host by lowering the replication levels so the fish host do not succumb to the disease too fast and the virus can be shed over a longer period of time over longer distances. Still, the emergence of virulent strains from the persistent and low virulent pool that is observed in this study is highly possible by flipping only 1 amino acid.

#### 4.6 Limitations and Outlook

The samples studied here have no clinical data available, so it was not possible to make a correlation between different virulence variants and sub or clinical infections.

As geographical data was sparse, phylogeography could not be done.

As sequencing of IPNV showed that RNA secondary structures are an issue in whole virus genome sequencing, it would be of interest to overcome this problem by testing other approaches such as higher incubations temperatures, alternative single strand binding proteins, secondary structure relaxing additives such as tetramethylammonium chloride, formamide, or salts already established in RT-PCR, to improve virus genome sequencing without using termini specific primer.

In order to verify the data described here a 2<sup>nd</sup> study looking at IPNV in trout aquaculture in Turkey has been initiated and it remains to be seen if the data obtained in this new study will verify the observations made here.

Fish challenge experiments should verify if isolates that have been characterised according to their specific amino acids in VP2 (persistent, low virulent or virulent) are causing subclinical or clinical infections. On the other hand, the not yet assigned amino acid combinations need to be characterized in fish experiments to elucidate their virulence character.

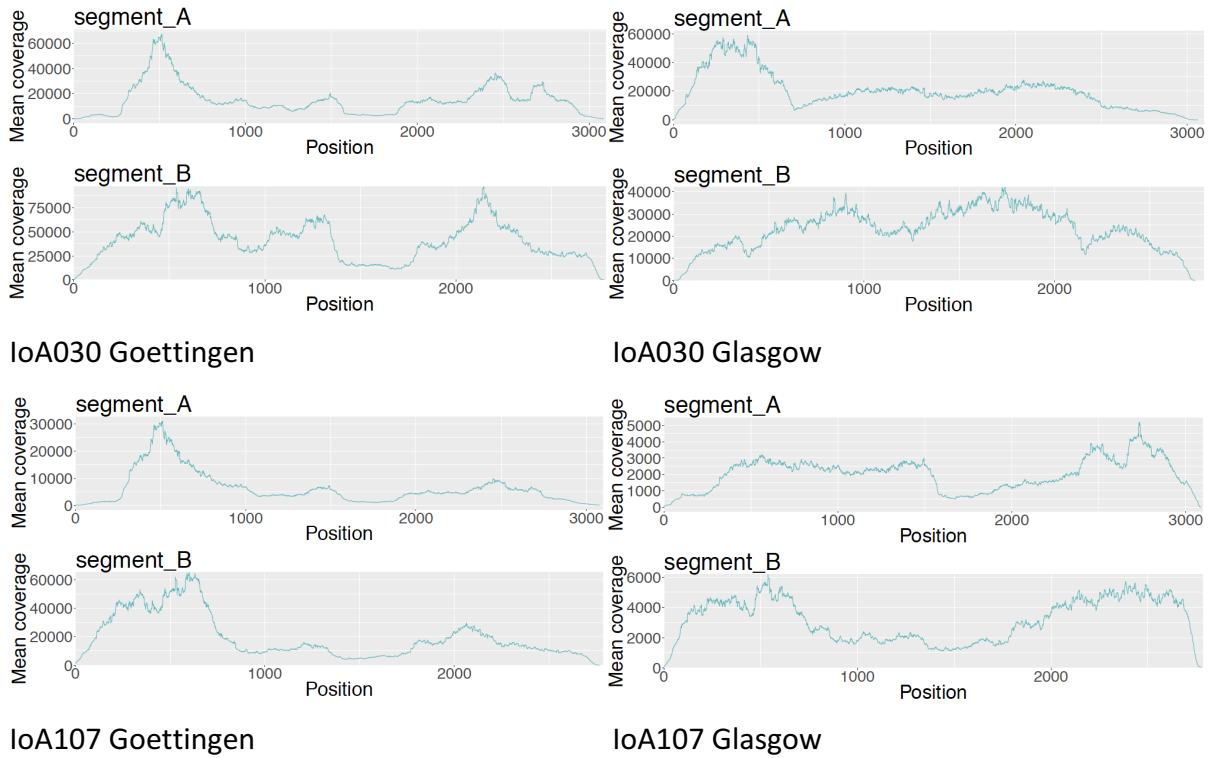
Since IPNV QTL fish are dominating salmon aquaculture, an effort should be made to attempt isolating and characterising more IPNV isolates evolving in these more selective conditions. Indeed, the industry should be convinced to contribute to a new surveillance effort so that the emergence of new IPNV and their characteristics are understood before they can cause any harm.

The data obtained in this study shows the need for a sterilising vaccine and furthermore, the industry should make an attempt to make vaccine data available for science to ease research to improve aquaculture.

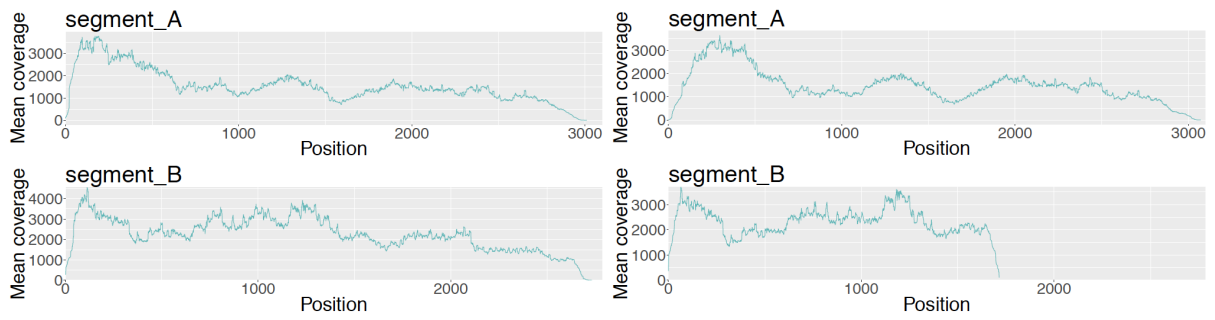
### III Appendix

#### III.1 Figures and Tables

##### III.1.1 Figures



**Figure 28.** Coverage plots of the other two randomly selected samples processed with the Goettingen and the Glasgow protocol.

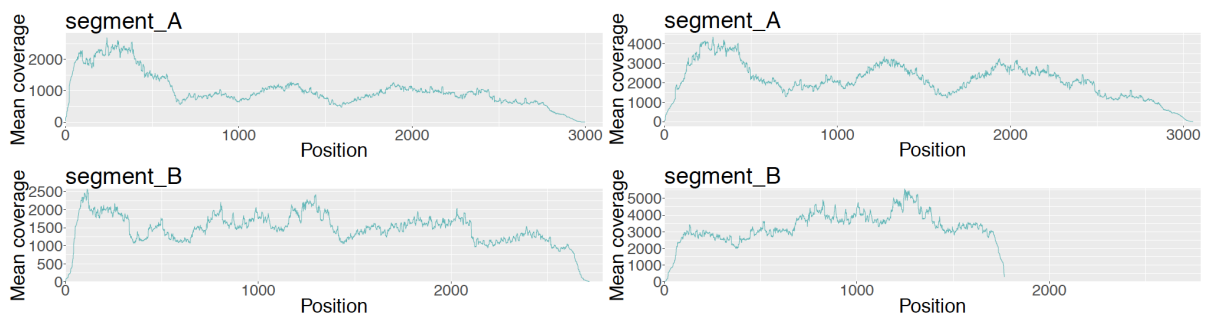


ss-cDNA-3'Primer

ds-cDNA-3'Primer



ss/ds-cDNA-3'Primer



RNaseH treatment

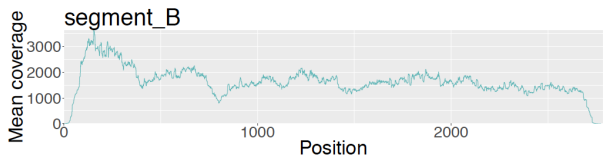
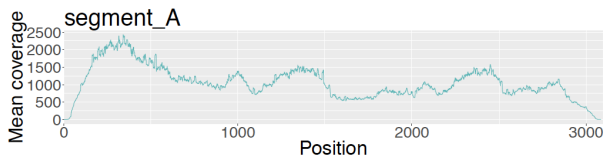
dNTPs and rand.hex



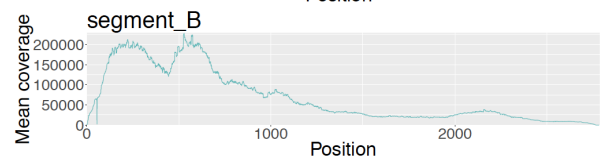
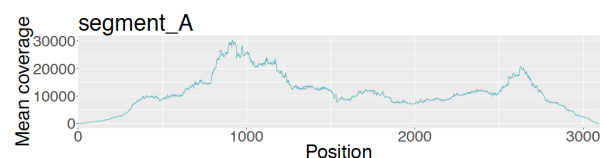
RNAincub.75C/5min

normal Glasgow prot.

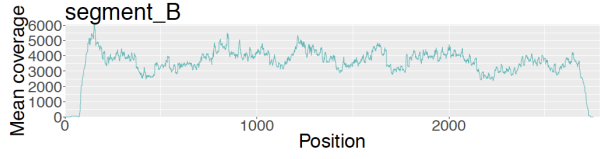
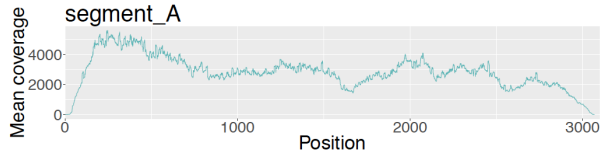
**Figure 29. Coverage plots of the other protocol variants.**



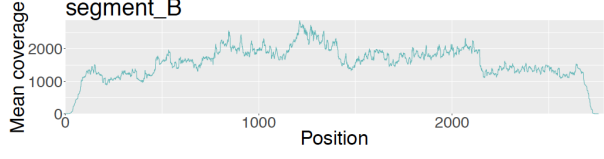
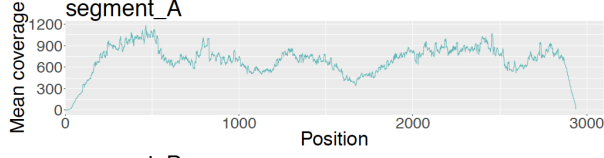
**IoA006 Glasgow**



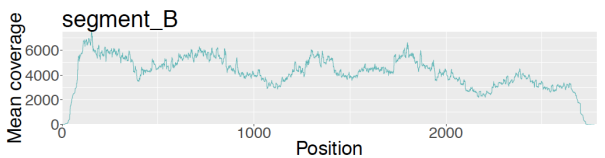
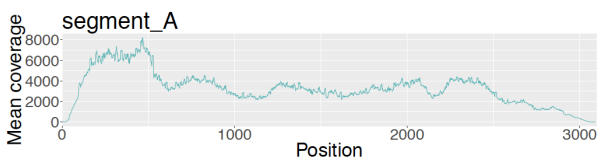
**IoA060 Goettingen**



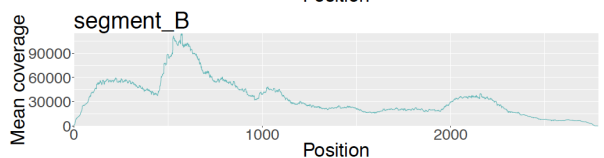
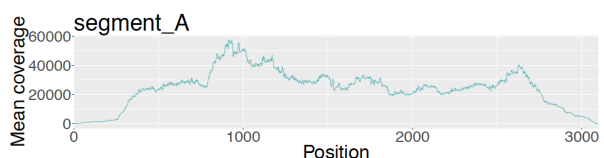
**IoA007 Glasgow**



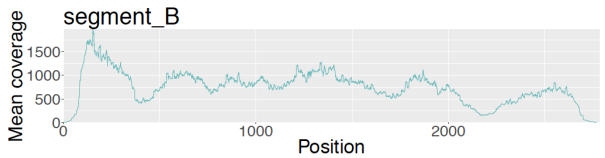
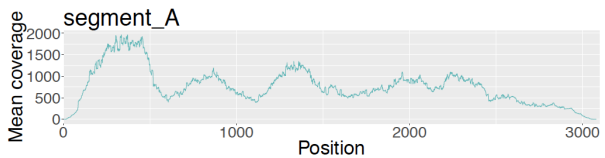
**IoA008 Glasgow**



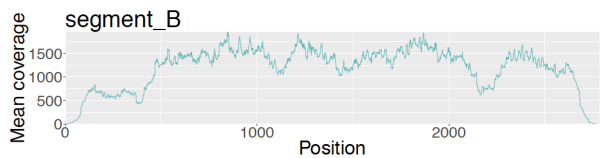
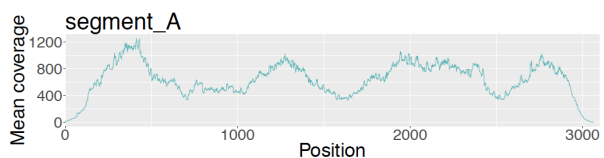
**IoA011 Glasgow**



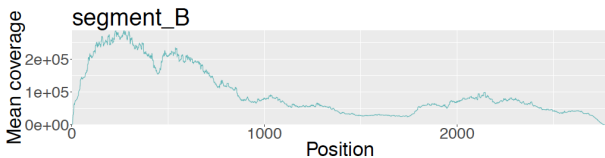
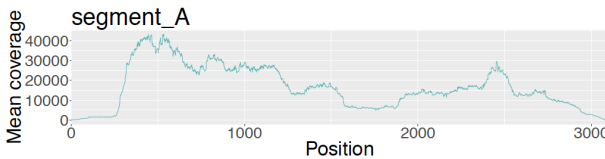
**IoA061 Goettingen**



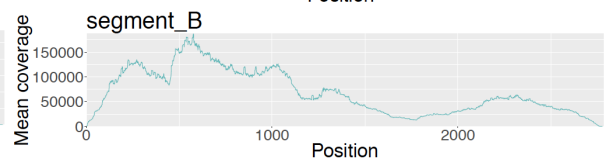
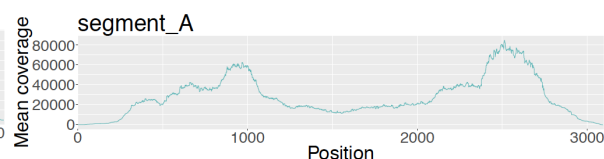
**IoA019 Glasgow**



**IoA021 Glasgow**

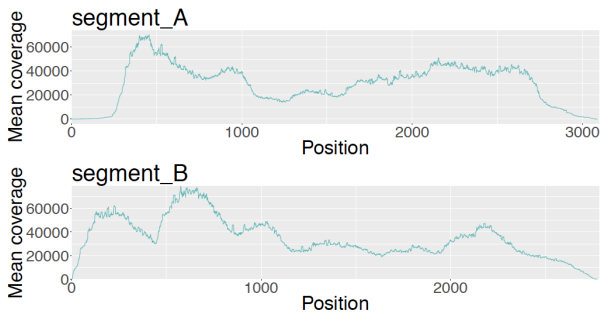


**IoA019 Glasgow**

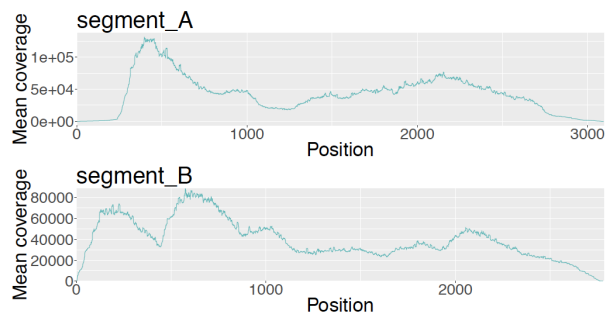


**IoA021 Glasgow**

IoA062 Goettingen



IoA063 Goettingen



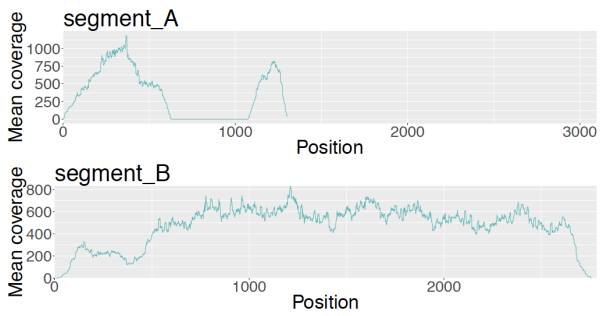
IoA064 Goettingen



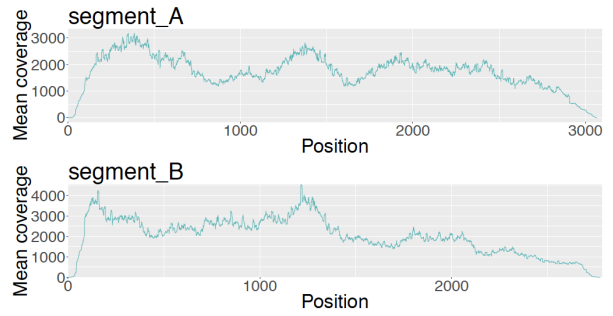
IoA065 Goettingen



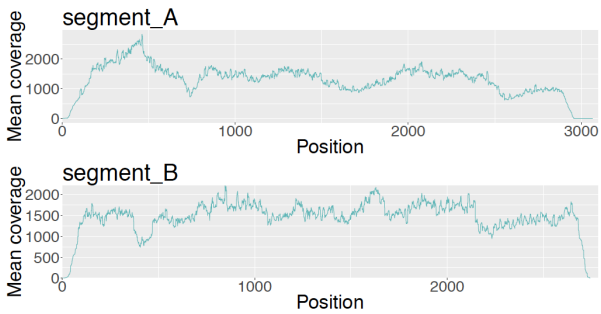
IoA066 Goettingen



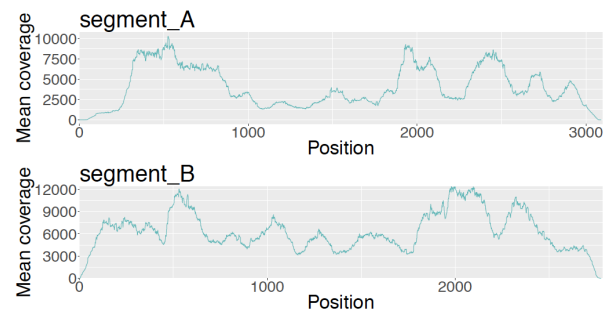
IoA027 Glasgow



IoA032 Glasgow

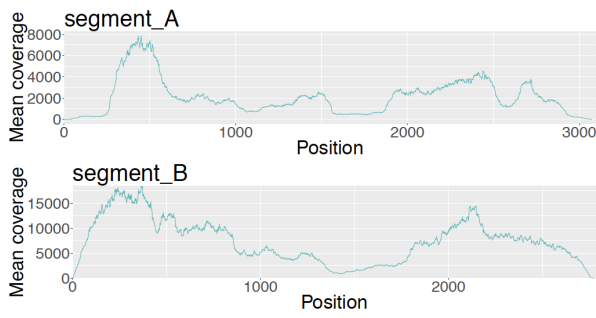


IoA068 Goettingen



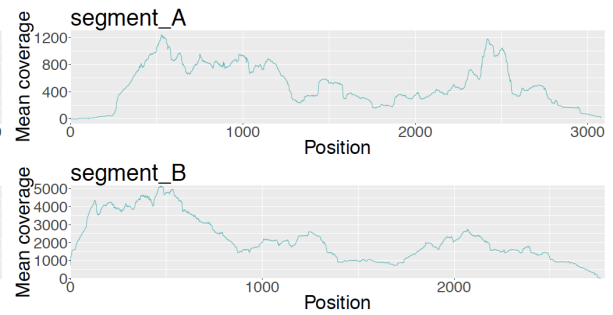
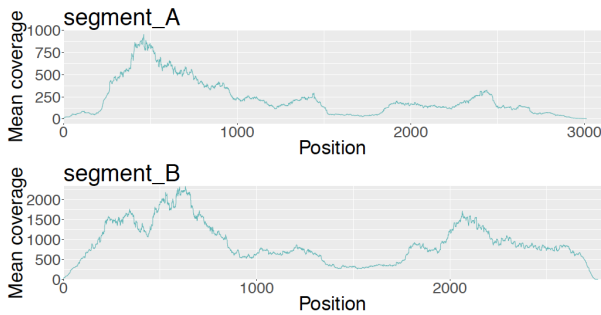
IoA026 Glasgow

IoA069 Goettingen



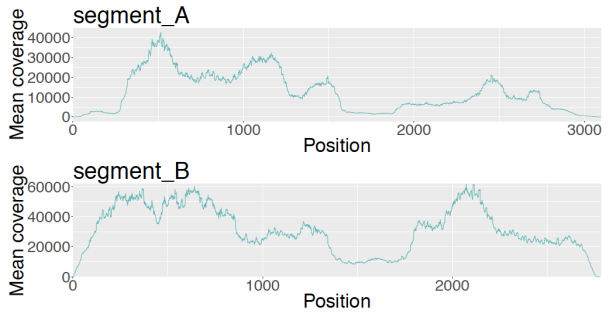
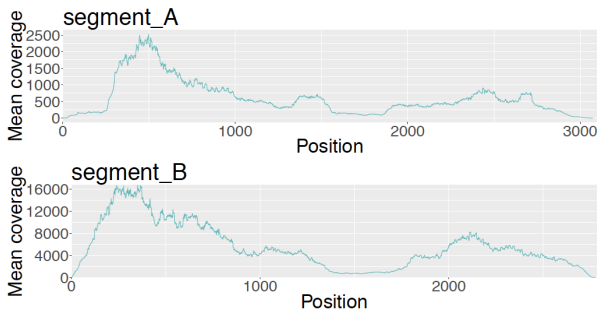
IoA070 Goettingen

IoA071 Goettingen



IoA072 Goettingen

IoA073 Goettingen



IoA075 Goettingen

IoA076 Goettingen



IoA081 Glasgow

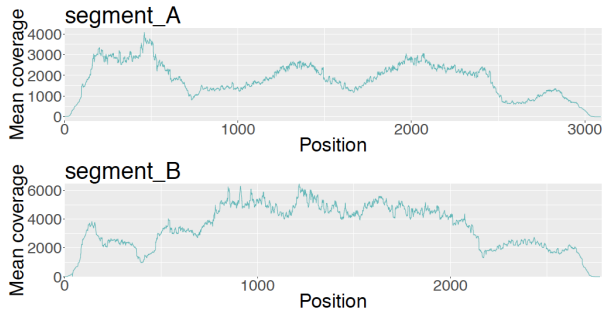
IoA082 Goettingen



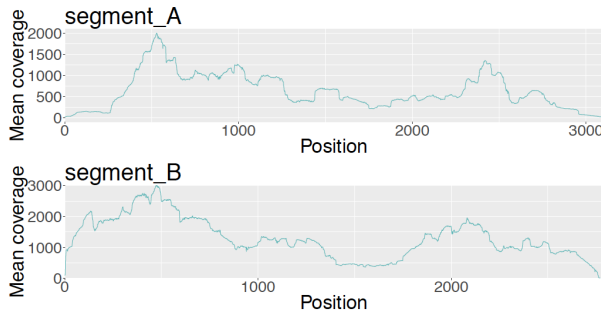
IoA083 Goettingen



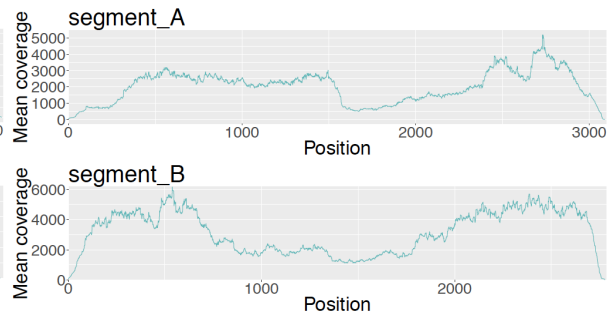
IoA084 Glasgow



IoA087 Glasgow



IoA088 Glasgow



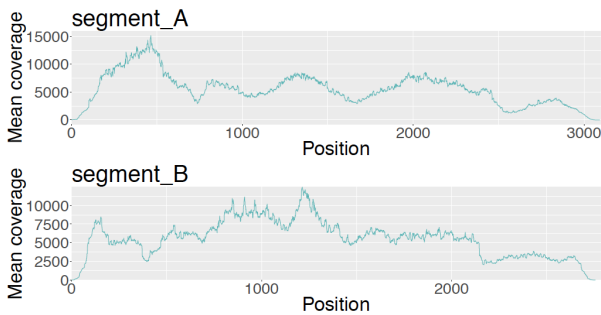
IoA090 Goettingen



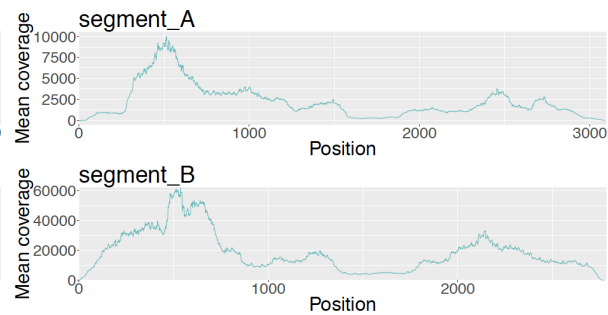
IoA107 Glasgow



IoA111 Goettingen



IoA001 Glasgow



IoA003 Glasgow

IoA009 Goettingen



### IoA039 Glasgow



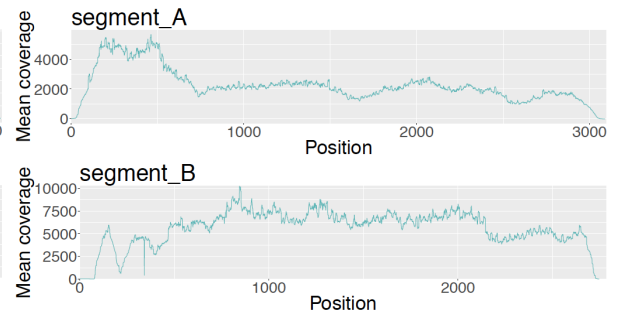
### IoA051 Glasgow



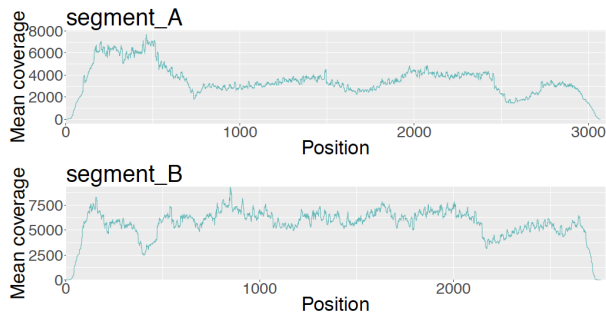
### IoA052 Glasgow



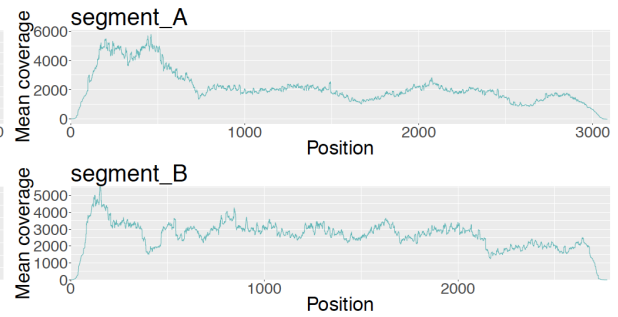
### IoA053 Glasgow



### IoA054 Glasgow



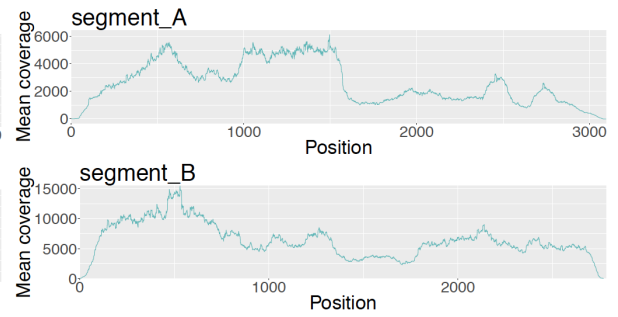
### IoA055 Glasgow



### IoA056 Glasgow

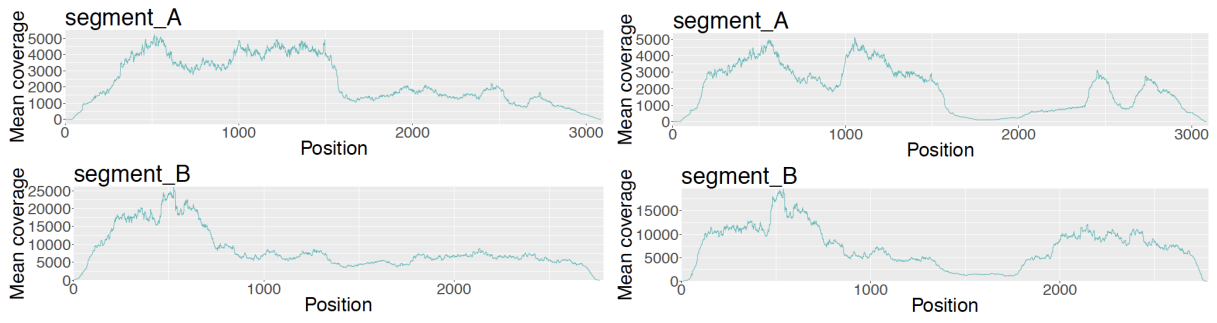


### IoA057 Glasgow



### IoA047 Goettingen

### IoA048 Goettingen



IoA049 Goettingen

IoA050 Goettingen



IoA058 Glasgow

**Figure 30. Coverage plots of 57 processed IPNV samples. The coverage plots were created with R, using the NanoOK tool as a guidance with changes in the R script for this purpose.**

### III.1.2 Tables

**Table 60. Control values of sequenced IPNV samples; NA=not assigned.**

Sequence ID	RNA			ds-cDNA		Library		
	RNA conc. (ng/μL) NanoDrop	Ct value qPCR	Mean conc. qPCR (genome equivalent; copies/mL)	dscDNA conc. (ng/μL) NanoDrop	dscDNA conc. (ng/μL) Qubit	Average fragment size (bp)	Conc. (pg/μL)	Mol. (pmol/l)
IOA_IPNV006	12.28	NA	NA	9.28	1.08	258	3037.91	18868.1
IOA_IPNV060	21.00	40.62	9.14E1	122.29	115.00	NA	NA	NA
IOA_IPNV007	7.14	20.37	1.72E7	10.73	0.63	221	370.37	2625.7
IOA_IPNV008	15.02	21.26	9.20E6	6.77	0.50	260	4297.47	26478.4
IOA_IPNV011	13.95	36.41	2.99E1	7.77	0.96	247	1941.34	12532.4
IOA_IPNV061	162.50	43.08	1.56E1	76.77	68.50	NA	NA	NA
IOA_IPNV019	21.85	37.43	2.44E2	7.62	0.36	203	1155.30	10354.1
IOA_IPNV021	135.77	23.66	3.16E6	8.93	0.58	193	678.74	6406.6
IOA_IPNV062	139.10	25.44	5.01E6	43.00	38.80	NA	NA	NA
IOA_IPNV063	60.80	41.63	4.41E1	67.66	53.20	NA	NA	NA
IOA_IPNV064	353.10	NA	NA	31.04	26.10	NA	NA	NA
IOA_IPNV065	239.70	NA	NA	28.67	25.40	NA	NA	NA
IOA_IPNV066	72.90	38.48	4.75E2	86.81	75.80	NA	NA	NA
IOA_IPNV027	46.74	24.80	1.45E6	8.40	0.34	184	1004.50	9854.8
IOA_IPNV032	57.73	24.09	2.70E6	9.05	0.80	178	605.03	5936.3
IOA_IPNV068	1845.73	NA	NA	32.84	9.82	260	813.00	4991.0
IOA_IPNV026	7.45	22.07	5.28E6	6.13	0.98	237	1168.22	7825.0
IOA_IPNV069	333.15	NA	NA	1.79	0.72	NA	NA	NA
IOA_IPNV070	173.49	36.69	5.57E3	102.13	47.10	NA	NA	NA
IOA_IPNV071	579.64	25.51	8.73E8	53.39	25.60	290	1817.05	10238.5
IOA_IPNV072	1615.24	24.88	7E9	38.75	11.42	269	962.46	5785.1
IOA_IPNV073	434.93	NA	NA	133.50	41.40	429	1672.72	6424.7
IOA_IPNV075	650.39	23.51	9.22E7	88.20	32.20	293	1783.87	9942.1
IOA_IPNV076	685.94	21.38	3.63E8	68.33	26.20	267	2574.63	15530.1
IOA_IPNV081	26.97	28.84	5.09E4	9.01	0.86	280	4878.27	29337.5
IOA_IPNV082	1057.00	28.84	3.52E7	37.43	14.62	272	1148.10	6791.0
IOA_IPNV083	1407.65	22.62	2.35E10	19.87	6.86	267	1297.05	7827.0
IOA_IPNV084	20.02	17.19	2.66E8	7.95	0.40	303	6218.10	35167.2

IOA_IPNV087	31.00	16.81	3.60E8	11.92	0.63	294	8208.93	48193.5
IOA_IPNV088	3.07	18.71	4.01E8	9.41	0.18	305	5580.07	31571.5
IOA_IPNV090	391.59	NA	NA	141.14	59.60	456	1811.74	6745.1
IOA_IPNV107	863.79	19.97	5.24E11	24.25	8.68	NA	NA	NA
IOA_IPNV111	73.39	17.25	2.62E8	9.03	0.49	246	1746.83	12334.0
IOA_IPNV001	4.51	20.61	2.17E7	6.79	0.57	248	1204.42	7764.1
IOA_IPNV003	2.08	18.66	4.19E8	9.39	0.21	309	6859.15	38430.9
IOA_IPNV009	330.83	18.32	3.10E9	102.73	44.00	265	5085.71	31114.9
IOA_IPNV010	24.79	17.02	3.02E8	12.42	0.85	271	6153.88	36620.1
IOA_IPNV013	11.44	15.99	6.49E8	7.91	0.86	253	2289.94	14425.3
IOA_IPNV014	465.51	NA	NA	172.28	89.80	457	1842.64	6500.5
IOA_IPNV018	29.01	16.88	3.61E8	11.92	1.25	250	1819.34	11612.7
IOA_IPNV024	781.58	21.25	3.93E8	85.56	34.20	268	1649.56	9887.4
IOA_IPNV030	558.97	22.16	5.31E8	76.79	35.00	277	2733.47	15937.8
IOA_IPNV031	11.73	15.63	8.48E8	5.64	0.64	250	2463.72	15731.5
IOA_IPNV035	2.07	28.29	3.47E5	11.72	0.52	473	6378.78	26009.0
IOA_IPNV039	21.20	18.48	4.77E7	17.85	1.41	274	3975.92	23764.9
IOA_IPNV051	11.03	19.89	2.35E7	9.07	1.31	273	1995.74	11783.0
IOA_IPNV052	14.09	16.12	3.16E8	10.47	2.08	246	1424.90	9282.8
IOA_IPNV053	9.48	23.38	2.13E6	9.76	1.45	154	51.26	544.2
IOA_IPNV054	20.85	15.70	4.22E8	11.56	2.14	249	1752.89	11315.4
IOA_IPNV055	16.85	17.32	1.39E8	10.45	2.80	253	1980.77	12639.1
IOA_IPNV056	9.65	15.54	4.72E8	9.16	0.95	245	3959.42	25968.4
IOA_IPNV057	6.27	17.04	1.68E8	6.88	0.90	239	2420.07	16058.7
IOA_IPNV047	458.44	25.35	5.33E6	89.52	80.30	NA	NA	NA
IOA_IPNV048	286.06	27.14	7.11E6	89.85	50.20	NA	NA	NA
IOA_IPNV049	180.96	31.04	3.82E5	86.56	58.40	NA	NA	NA
IOA_IPNV050	231.32	29.31	1.39E6	66.18	43.20	NA	NA	NA
IOA_IPNV058	36.53	18.96	8.01E7	10.05	1.20	192	294.75	2785.8

**Table 61. Number of forward and reverse decontaminated reads of the randomly selected IPNV samples processed with the Goettingen or the Glasgow protocol for whole genome sequencing using NGS on the MiSeq.**

Sequence ID	Collection Year	Raw reads	Raw reads	RNA to ds-cDNA protocol
		forward	reverse	
IOA_IPNV006	1982	157513	159043	Glasgow
IOA_IPNV060	1983	847250	847403	Goettingen
IOA_IPNV007	1984	156883	157428	Glasgow
IOA_IPNV008	1984	203001	206777	Glasgow
IOA_IPNV011	1986	262412	263952	Glasgow
IOA_IPNV061	1986	666022	666071	Goettingen
IOA_IPNV019	1987	56400	58897	Glasgow
IOA_IPNV021	1987	58105	59652	Glasgow
IOA_IPNV062	1987	1130819	1130831	Goettingen
IOA_IPNV063	1987	1030840	1031013	Goettingen
IOA_IPNV064	1987	716053	716233	Goettingen
IOA_IPNV065	1987	921373	921546	Goettingen
IOA_IPNV066	1988	645778	645936	Goettingen
IOA_IPNV027	1989	34563	35875	Glasgow
IOA_IPNV032	1989	51422	52702	Glasgow
IOA_IPNV068	1989	108577	155139	Goettingen
IOA_IPNV026	1990	116230	118318	Glasgow
IOA_IPNV069	1992	97497	97571	Goettingen
IOA_IPNV070	1992	84454	84490	Goettingen
IOA_IPNV071	1992	1869779	1871585	Goettingen
IOA_IPNV072	1992	233195	239646	Goettingen
IOA_IPNV073	1993	114205	114188	Goettingen
IOA_IPNV075	1993	341263	348589	Goettingen
IOA_IPNV076	1994	794345	796701	Goettingen
IOA_IPNV081	1994	255231	276768	Glasgow
IOA_IPNV082	1996	378916	378509	Goettingen
IOA_IPNV083	1997	345097	346524	Goettingen
IOA_IPNV084	1997	149023	156194	Glasgow
IOA_IPNV087	1997	51928	83714	Glasgow
IOA_IPNV088	1997	117249	138934	Glasgow
IOA_IPNV090	1998	87701	87705	Goettingen
IOA_IPNV107	1999	47114	48257	Glasgow
IOA_IPNV111	1999	943838	947100	Goettingen

IOA_IPNV001	2000	275019	279827	Glasgow
IOA_IPNV003	2002	568821	572761	Glasgow
IOA_IPNV009	2003	803601	810307	Goettingen
IOA_IPNV010	2003	102959	152299	Glasgow
IOA_IPNV013	2004	762381	806229	Glasgow
IOA_IPNV014	2004	87292	87281	Goettingen
IOA_IPNV018	2004	156887	183702	Glasgow
IOA_IPNV024	2005	441114	454357	Goettingen
IOA_IPNV030	2006	931946	933601	Goettingen
IOA_IPNV031	2006	405181	427456	Glasgow
IOA_IPNV035	2007	196167	202745	Glasgow
IOA_IPNV039	2008	548752	609277	Glasgow
IOA_IPNV051	2010	102370	104648	Glasgow
IOA_IPNV052	2011	177755	178843	Glasgow
IOA_IPNV053	2011	2200	2457	Glasgow
IOA_IPNV054	2011	284878	287228	Glasgow
IOA_IPNV055	2012	194832	196394	Glasgow
IOA_IPNV056	2012	185364	186123	Glasgow
IOA_IPNV057	2013	132886	134701	Glasgow
IOA_IPNV047	2014	449718	449735	Goettingen
IOA_IPNV048	2014	82315	82347	Goettingen
IOA_IPNV049	2014	94259	94302	Goettingen
IOA_IPNV050	2014	80821	80846	Goettingen
IOA_IPNV058	2014	65844	67027	Glasgow



**Table 62. Mean coverage (number of reads/reference bases) of segment A and B for 57 processed IPNV samples.**

Sequence ID	Collection Year	Protocol	Mean coverage segment A	Mean coverage segment B
IOA_IPNV006	1982	Glasgow	1057.42	1627.85
IOA_IPNV060	1983	Goettingen	11225.5	68372.8
IOA_IPNV007	1984	Glasgow	2838.87	3557.37
IOA_IPNV008	1984	Glasgow	704.312	1584.97
IOA_IPNV011	1986	Glasgow	24625.3	34155.3
IOA_IPNV061	1986	Goettingen	3260.57	4180.43
IOA_IPNV019	1987	Glasgow	28061.4	68500.2
IOA_IPNV021	1987	Glasgow	16605.2	97132.9
IOA_IPNV062	1987	Goettingen	42507.8	39525.6
IOA_IPNV063	1987	Goettingen	29486.7	34777.7
IOA_IPNV064	1987	Goettingen	763.912	744.757
IOA_IPNV065	1987	Goettingen	637.947	1198.1
IOA_IPNV066	1988	Goettingen	15903.1	39622.8
IOA_IPNV027	1989	Glasgow	252.93	255.255
IOA_IPNV032	1989	Glasgow	359.017	475.389
IOA_IPNV068	1989	Goettingen	1728.51	2025.94
IOA_IPNV026	1990	Glasgow	1294.14	1472.15
IOA_IPNV069	1992	Goettingen	4041.39	6344.43
IOA_IPNV070	1992	Goettingen	2101.18	7359.07
IOA_IPNV071	1992	Goettingen	29768.6	115328
IOA_IPNV072	1992	Goettingen	228.237	900.192
IOA_IPNV073	1993	Goettingen	495.345	2176.05
IOA_IPNV075	1993	Goettingen	607.765	5601.92
IOA_IPNV076	1994	Goettingen	12416.8	31406.9
IOA_IPNV081	1994	Glasgow	7382.67	8689.62
IOA_IPNV082	1996	Goettingen	1474.65	6441.11
IOA_IPNV083	1997	Goettingen	737.929	2650.62
IOA_IPNV084	1997	Glasgow	3585.32	4951.66
IOA_IPNV087	1997	Glasgow	1059.14	1422.46
IOA_IPNV088	1997	Glasgow	1800.65	3423.65
IOA_IPNV090	1998	Goettingen	643.839	1299.1
IOA_IPNV107	1999	Goettingen	2022.02	3144.74
IOA_IPNV111	1999	Glasgow	5941.4	56484.6

IOA_IPNV001	2000	Glasgow	5924.04	7736.91
IOA_IPNV003	2002	Glasgow	5503.64	5560.97
IOA_IPNV009	2003	Goettingen	2231.4	18619.2
IOA_IPNV010	2003	Glasgow	2225.12	2286.15
IOA_IPNV013	2004	Glasgow	421.222	1631.82
IOA_IPNV014	2004	Goettingen	15744.5	28408.6
IOA_IPNV018	2004	Glasgow	2518.9	5401.02
IOA_IPNV024	2005	Goettingen	524.901	1664.13
IOA_IPNV030	2006	Goettingen	15076.9	41065
IOA_IPNV031	2006	Glasgow	9645.72	12206.9
IOA_IPNV035	2007	Glasgow	170.414	441.67
IOA_IPNV039	2008	Glasgow	1534.39	2228.68
IOA_IPNV051	2010	Glasgow	442.107	816.657
IOA_IPNV052	2011	Glasgow	3136.61	5685.87
IOA_IPNV053	2011	Glasgow	172.811	321.778
IOA_IPNV054	2011	Glasgow	6184.14	9461.81
IOA_IPNV055	2012	Glasgow	2255.28	5646.46
IOA_IPNV056	2012	Glasgow	3508.96	5632.67
IOA_IPNV057	2013	Glasgow	2150.06	2615.85
IOA_IPNV047	2014	Goettingen	2593.66	6378.74
IOA_IPNV048	2014	Goettingen	1946.87	7189.44
IOA_IPNV049	2014	Goettingen	2359.31	8515.45
IOA_IPNV050	2014	Goettingen	6574.6	34302
IOA_IPNV058	2014	Glasgow	1045.52	1315.39

**Table 63. Missing nucleotides and the overall mapping rate of processed IPNV samples after de novo assembly.**

Sequence ID	Segment A		Segment B		Total (bp)	Mapping rate (%)	RNA to ds-cDNA protocol
	(bp)		(bp)				
	5'	3'	5'	3'			
IOA_IPNV006	1	8	1	2	12	7.44	Glasgow
IOA_IPNV060	0	0	0	0	0	87.85	Goettingen
IOA_IPNV007	1	26	1	20	48	31.86	Glasgow
IOA_IPNV008	0	157	4	19	180	4.87	Glasgow
IOA_IPNV011	0	0	0	0	0	88.31	Glasgow
IOA_IPNV061	1	2	1	2	6	15.84	Goettingen
IOA_IPNV019	0	0	0	0	0	86.79	Glasgow
IOA_IPNV021	0	0	0	0	0	90.82	Glasgow
IOA_IPNV062	0	0	0	0	0	82.71	Goettingen
IOA_IPNV063	0	5	0	0	5	84.20	Goettingen
IOA_IPNV064	7	2	1	4	14	4.59	Goettingen
IOA_IPNV065	32	0	2	5	39	7.88	Goettingen
IOA_IPNV066	0	0	0	0	0	86.01	Goettingen
IOA_IPNV027	24	2	26	15	67	1.67	Glasgow
IOA_IPNV032	92	2386	2	15	2495	2.15	Glasgow
IOA_IPNV068	0	26	0	3	29	2.71	Goettingen
IOA_IPNV026	1	105	1	32	139	7.35	Glasgow
IOA_IPNV069	0	2	0	0	2	87.66	Goettingen
IOA_IPNV070	7	9	0	0	16	89.77	Goettingen
IOA_IPNV071	0	0	0	0	0	88.02	Goettingen
IOA_IPNV072	52	31	8	1	92	2.22	Goettingen
IOA_IPNV073	0	18	10	0	28	23.02	Goettingen
IOA_IPNV075	0	10	0	0	10	72.88	Goettingen
IOA_IPNV076	17	5	0	0	22	9.02	Goettingen
IOA_IPNV081	0	0	0	0	0	54.22	Glasgow
IOA_IPNV082	1	18	0	0	19	27.33	Goettingen
IOA_IPNV083	0	26	0	2	28	13.45	Goettingen
IOA_IPNV084	0	13	0	0	13	5.38	Glasgow
IOA_IPNV087	1	2	1	28	32	24.71	Glasgow
IOA_IPNV088	1	36	0	19	56	4.18	Glasgow
IOA_IPNV090	2	156	0	2	160	9.92	Goettingen
IOA_IPNV107	0	8	0	8	16	22.21	Glasgow

IOA_IPNV111	0	7	0	0	7	68.80	Goettingen
IOA_IPNV001	0	6	0	20	26	30.72	Glasgow
IOA_IPNV003	0	3	1	19	23	17.88	Glasgow
IOA_IPNV009	0	9	0	0	9	16.04	Goettingen
IOA_IPNV010	1	115	2	19	137	6.43	Glasgow
IOA_IPNV013	0	9	0	8	17	23.09	Glasgow
IOA_IPNV014	0	23	1	2	26	42.91	Goettingen
IOA_IPNV018	0	26	1	2	29	15.00	Glasgow
IOA_IPNV024	0	26	0	0	26	2.20	Goettingen
IOA_IPNV030	0	10	0	0	10	69.31	Goettingen
IOA_IPNV031	1	22	1	8	32	34.34	Glasgow
IOA_IPNV035	28	0	4	4	36	1.06	Glasgow
IOA_IPNV039	1	5	2	2	10	2.78	Glasgow
IOA_IPNV051	1	2	17	2	22	3.42	Glasgow
IOA_IPNV052	1	26	1	30	58	30.84	Glasgow
IOA_IPNV053	43	157	23	35	258	4.83	Glasgow
IOA_IPNV054	0	4	1	37	42	39.49	Glasgow
IOA_IPNV055	1	5	1	29	36	18.84	Glasgow
IOA_IPNV056	1	26	1	15	43	30.42	Glasgow
IOA_IPNV057	1	11	2	2	16	14.62	Glasgow
IOA_IPNV047	0	0	4	0	4	90.12	Goettingen
IOA_IPNV048	0	9	0	0	9	90.55	Goettingen
IOA_IPNV049	0	11	10	0	21	90.21	Goettingen
IOA_IPNV050	2	10	0	30	42	89.80	Goettingen
IOA_IPNV058	1	1	2	15	19	12.09	Glasgow

**Table 64. Sequenced IPNV isolates and their genogroups according to segment A and segment B phylogeny; NA=not assigned.**

Sequence ID	Collection Year	Serotype	Genogroup (segment A)	Genogroup (segment B)
IOA_IPNV006	1982	NA	III	III
IOA_IPNV060	1983	NA	I	I
IOA_IPNV007	1984	NA	V	V
IOA_IPNV008	1984	NA	V	V
IOA_IPNV011	1986	NA	III	III
IOA_IPNV061	1986	Sp	I	I
IOA_IPNV019	1987	NA	III	III
IOA_IPNV021	1987	NA	V	V
IOA_IPNV062	1987	NA	V	V
IOA_IPNV063	1987	Ab	III	II
IOA_IPNV064	1987	Ab	III	II
IOA_IPNV065	1987	Ab	III	II
IOA_IPNV066	1988	Ab	III	II
IOA_IPNV027	1989	NA	V	V
IOA_IPNV032	1989	NA	V	V
IOA_IPNV068	1989	Ab	III	II
IOA_IPNV026	1990	NA	V	V
IOA_IPNV069	1992	Ab	III	III
IOA_IPNV070	1992	Sp	V	V
IOA_IPNV071	1992	Sp	V	V
IOA_IPNV072	1992	Sp	V	V
IOA_IPNV073	1993	Sp	V	V
IOA_IPNV075	1993	Sp	V	V
IOA_IPNV076	1994	Sp	V	V
IOA_IPNV081	1994	NA	I	I
IOA_IPNV082	1996	Sp	V	V
IOA_IPNV083	1997	Sp	V	V
IOA_IPNV084	1997	Sp	V	V
IOA_IPNV087	1997	Sp	V	V
IOA_IPNV088	1997	Sp	V	V
IOA_IPNV090	1998	Sp	V	V
IOA_IPNV107	1999	Sp	V	V
IOA_IPNV111	1999	Sp	V	V
IOA_IPNV001	2000	Sp	V	V

IOA_IPNV003	2002	Sp	V	V
IOA_IPNV009	2003	Sp	V	V
IOA_IPNV010	2003	Sp	V	V
IOA_IPNV013	2004	Sp	V	V
IOA_IPNV014	2004	Sp	V	V
IOA_IPNV018	2004	Sp	V	V
IOA_IPNV024	2005	Sp	V	V
IOA_IPNV030	2006	Sp	V	V
IOA_IPNV031	2006	Sp	V	V
IOA_IPNV035	2007	Sp	V	V
IOA_IPNV039	2008	Sp	V	V
IOA_IPNV051	2010	NA	V	V
IOA_IPNV052	2011	NA	V	V
IOA_IPNV053	2011	NA	V	V
IOA_IPNV054	2011	NA	V	V
IOA_IPNV055	2012	NA	V	V
IOA_IPNV056	2012	NA	V	V
IOA_IPNV057	2013	NA	V	V
IOA_IPNV047	2014	Sp	V	V
IOA_IPNV048	2014	Sp	V	V
IOA_IPNV049	2014	Sp	V	V
IOA_IPNV050	2014	Sp	V	V
IOA_IPNV058	2014	NA	V	V

**Table 65. Detailed VP2 residue characteristics of sequenced IPNV isolates compared to the virulent reference strain excluding IoA032; NA=not assigned. Amino acid positions identified in previous studies as important in terms of viral virulence are listed, with a focus on the positions 217 and 221 (underlined).**

Sequence ID	Collection Year	Genogroup by genotyping	Amino acid position							Variant
			<u>217</u>	<u>221</u>	<u>247</u>	<u>252</u>	<u>281</u>	<u>282</u>	<u>319</u>	
Santi et al., 2004	2000	V	T	A	T	V	T	N	A	Virulent
IOA_IPNV006	1982	III	P	T	P	A	T	A	A	Persistent
IOA_IPNV060	1983	I	A	T	E	N	T	A	A	NA
IOA_IPNV007	1984	V	P	T	A	N	T	N	A	Persistent
IOA_IPNV008	1984	V	P	T	A	N	T	N	A	Persistent
IOA_IPNV011	1986	III	P	T	P	A	T	A	A	Persistent
IOA_IPNV061	1986	I	A	T	E	N	T	A	A	NA
IOA_IPNV019	1987	III	P	T	P	A	T	A	A	Persistent
IOA_IPNV021	1987	V	P	T	A	N	T	N	A	Persistent
IOA_IPNV062	1987	V	P	T	A	N	T	N	A	Persistent
IOA_IPNV063	1987	III	P	T	P	A	T	A	A	Persistent
IOA_IPNV064	1987	III	P	T	P	A	T	A	A	Persistent
IOA_IPNV065	1987	III	P	T	P	A	T	A	A	Persistent
IOA_IPNV066	1988	III	P	T	P	A	T	A	A	Persistent
IOA_IPNV027	1989	V	P	T	A	N	T	N	V	Persistent
IOA_IPNV068	1989	III	P	T	P	A	T	A	A	Persistent
IOA_IPNV026	1990	V	P	T	A	N	T	N	A	Persistent
IOA_IPNV069	1992	III	P	T	P	A	T	A	A	Persistent
IOA_IPNV070	1992	V	P	T	A	N	T	N	A	Persistent
IOA_IPNV071	1992	V	P	T	A	N	T	N	A	Persistent
IOA_IPNV072	1992	V	P	A	A	T	T	N	A	Low virulent
IOA_IPNV073	1993	V	P	T	A	N	T	N	A	Persistent
IOA_IPNV075	1993	V	P	T	A	N	T	N	A	Persistent
IOA_IPNV076	1994	V	P	T	A	N	T	N	A	Persistent
IOA_IPNV081	1994	I	P	T	A	N	T	N	A	Persistent
IOA_IPNV082	1996	V	A	T	E	N	T	A	A	NA
IOA_IPNV083	1997	V	P	T	A	N	T	N	A	Persistent
IOA_IPNV084	1997	V	P	T	A	N	T	N	A	Persistent
IOA_IPNV087	1997	V	P	A	A	I	T	N	A	Low virulent
IOA_IPNV088	1997	V	P	A	A	I	T	N	A	Low virulent
IOA_IPNV090	1998	V	P	A	A	I	T	N	A	Low virulent

IOA_IPNV107	1999	V	P	T	A	N	T	N	A	Persistent
IOA_IPNV111	1999	V	P	A	A	I	T	N	A	Low virulent
IOA_IPNV001	2000	V	P	T	A	N	T	N	A	Persistent
IOA_IPNV003	2002	V	P	T	A	N	T	N	A	Persistent
IOA_IPNV009	2003	V	P	A	A	V	T	N	A	Low virulent
IOA_IPNV010	2003	V	P	A	A	V	T	N	A	Low virulent
IOA_IPNV013	2004	V	P	A	A	V	T	N	A	Low virulent
IOA_IPNV014	2004	V	P	S	A	V	T	N	A	NA
IOA_IPNV018	2004	V	P	A	A	V	T	N	A	Low virulent
IOA_IPNV024	2005	V	P	T	A	V	T	N	A	Persistent
IOA_IPNV030	2006	V	P	A	A	V	T	N	A	Low virulent
IOA_IPNV031	2006	V	P	T	A	V	T	N	A	Persistent
IOA_IPNV035	2007	V	P	A	A	V	T	N	A	Low virulent
IOA_IPNV039	2008	V	P	T	A	N	T	N	A	Persistent
IOA_IPNV051	2010	V	P	A	A	V	T	N	A	Low virulent
IOA_IPNV052	2011	V	P	A	A	V	T	N	A	Low virulent
IOA_IPNV053	2011	V	T	A	A	V	T	N	A	Virulent
IOA_IPNV054	2011	V	P	A	A	V	T	N	A	Low virulent
IOA_IPNV055	2012	V	P	A	A	V	T	N	A	Low virulent
IOA_IPNV056	2012	V	P	A	A	V	T	N	A	Low virulent
IOA_IPNV057	2013	V	P	T	A	V	T	N	A	Persistent
IOA_IPNV047	2014	V	P	T	A	V	T	N	A	Persistent
IOA_IPNV048	2014	V	P	T	A	V	T	N	A	Persistent
IOA_IPNV049	2014	V	P	A	A	V	T	N	A	Low virulent
IOA_IPNV050	2014	V	P	T	A	V	T	N	A	Persistent
IOA_IPNV058	2014	V	P	A	A	V	T	N	A	Low virulent



### III.1.3 Data processing – Additional Appendix

V.1.1 MiSeq reads process shows the detailed workflow, provided by Dr. Stefanie Wehner, of how the data, generated with the MiSeq, was processed in order to obtain whole IPNV genomes.

V.1.2 Workflow for coverage plots generation shows the detailed workflow, provided by Dr. Stefanie Wehner, of how coverage plots were generated for each sequenced IPNV sample on the MiSeq.

V.2 MinION data process shows the detailed workflow, provided by Dr. Stefanie Wehner, of how the data generated with the MinION was processed to obtain a full IPNV genome.

V.3 Genome annotation shows the detailed process of IPNV genome annotation for sequenced IPNV samples on the MiSeq. It was provided by Dr. Michaël Bekaert.

V.4 CAI shows the detailed workflow, provided by Nicholas Di Paola, of how the CAI for each IPNV protein, based on the single protein alignment from all sequenced IPNV isolates, was calculated, with corresponding statistical analysis and graphical output.

### III.2 Abbreviations and Acronyms

°C	Celsius degree
A	Adenine
bp	Base pair
C	Cytosine
CaCl <sub>2</sub>	Calcium Chloride
cDNA	Complementary DNA
CO <sub>2</sub>	Carbon dioxide
Cp	Crossing point
CPE	Cytopathic effect
CpG	5'—Cytosine—phosphate—Guanine—3'
Ct	Threshold Cycle
DNA	Deoxyribonucleic acid
dNTP	Desoxyribonucleosidtriphosphate
ds	Double stranded
EDTA	Ethylenediaminetetraacetic acid
EMEM	Eagle's Minimum Essential Medium
FBS	Fetal bovine serum
G	Guanine
h	Hour
HCl	Hydrochloric acid
HPLC	High-performance liquid chromatography
iSP9	Integrated C9 (phosphoramidite) spacer
kbp	Kilo base pair
kg	Kilo gram
L	Litre
M	Molar
mg	Mille gram
min	Minute
mL	Mille litre
mM	Mille molar

MnCl <sub>2</sub>	Manganese (II) chloride
mol	Mole
MW	Molecular weight
N	Normality
NaCl	Sodium chloride
ng	Nano gram
nM	Nano molar
OD	Optical density
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid
RT-qPCR	Real-time quantitative PCR
sec	Second
ss	Single stranded
T	Thymine
TAE	Tris-Acetic acid-EDTA-buffer
TC	Tissue culture flask
Tris	Tris (hydroxymethyl) aminomethane
U	Uracil
V	Volt
WP	Well plate
μL	Micro litre
μM	Micro molar
RNase H	Ribonuclease H

### III.3 Material

#### III.3.1 Primer, Probes, Adaptors

3' iSP9 adaptor (Oligonucleotide)	5'-P-GACCTCTGAGGATTCTAAACXTCCAGTTTAGAATCC-OH-3' X= C9-phosphoamiditespacer
qPCR forward primer	5'-GCCTACCCCCGTTCT
qPCR reverse primer	5'-CCCGTCACTGTTGTTGAGTTGA
qPCR probe	5'-6FAM-ACTCTCTACGAGGGAAACGCCGAC--BBQ
WTA2 adaptor	5'-GTGGTGTGTTGGGTGTGTTTGGG
S501	5'-TAGATCGC
S502	5'-CTCTCTAT
S503	5'-TATCCTCT
S504	5'-AGAGTAGA
S505	5'-GTAAGGAG
S506	5'-ACTGCATA
S507	5'-AAGGAGTA
S508	5'-CTAAGCCT
N701	5'-TAAGCGCA
N702	5'-CGTACTAG
N703	5'-AGGCAGAA
N704	5'-TCCTGAGC
N705	5'-GGA CTCT
N706	5'-TAGGCATG
N707	5'-CTCTCTAC
N708	5'-CAGAGAGG
N709	5'-GCTACGCT
N710	5'-CGAGGCTG
N711	5'-AAGAGGCA
N712	5'-GTAGAGGA

Nextera®	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
transposase	(a) Read 1 -->
sequences	
	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
	(d) Read 2 -->
Nextera® Index	5'-AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC
Kit - PCR primers	(c) i5 Index read -->
	5'-CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGGGCTCGG
	<-- i7 Index read (b)
IPNV A	5'-CGACGACCCCG
IPNV B	5'-GCTCCR(A or G)CGCC

### III.3.2 Chemicals

10 M NaOH BioUltra	Sigma Aldrich
Agarose Molecular Grade	Bioline
Chloroform, Chromasolv for HPLC	Sigma Aldrich
DNA decontamination reagent 250 mL	Sigma Aldrich
Dulbecco's Phosphate Buffered Solution (1x; -CaCl <sub>2</sub> ; -MgCl <sub>2</sub> ), 500 mL	Gibco
Ethanol for molecular biology	Sigma Aldrich
Ethidium bromide	Thermo Fisher Scientific
Ethylenediaminetetraacetic acid disodium salt dihydrate	Sigma Aldrich
FBS batch number: 41F6547K	Gibco
Hank's Balanced Salt Solution (1x; -CaCl <sub>2</sub> ; -MgCl <sub>2</sub> ), 500 mL	Gibco
HCl	VWR CHEMICALS
L-glutamine, 200 mM (100x), 100 mL	Gibco
Minimum Essential Medium (1x; +Eagle's Salt; +L-glutamine), 500 mL	Gibco
Minimum Essential Medium Non-Essential Amino Acids (100x), 100 mL	Gibco
MyOne C1 streptavidine beads	Life Technologies
NaCl	Sigma Aldrich
peqGold Trifast	PeqLab
Poly ethylene glycol, BioUltra, 8000, 1 kg	Sigma Aldrich
Presept (100x 2.5 g)	Division of Ethicon, Inc.
Propan-2-ol	Fisher Scientific
RNase A (20 mg/mL)	PeqLab
RNase away, 250 mL	Sigma Aldrich
TAE (50x), 1 L	Fisher Bio Reagents
Tris	Sigma Aldrich
Trypan Blue Solution, 0,4 %, 100 mL	Sigma Aldrich
Trypsin EDTA, 0,05 % (1x), 500 mL	Gibco
Virkon Desinfectant (5 kg)	SLS

---

---

Molecular grade water	Sigma- Aldrich
isopropanol	Sigma Aldrich

---

---

### III.3.3 Kits

Agencourt AMPure XP beads	Beckman Coulter
Agencourt RNAdvance® Blood Kit	Beckman Coulter
Agilent RNA 6000 Nano Kit	Agilent Technologies
DNA Gel Loading Dye	Thermo Scientific
dNTP Mix (2.5 mM)	Life Technologies
Flow cell wash Kit EXP-WSH002	Oxford Nanopore Technologies
GeneRuler 100 bp Plus DNA Ladder	Thermo Scientific
GeneRuler Express Marker	Thermo Scientific
Glycogen 35 mg/mL, 1 mL	PeqLab
High Sensitivity DNA Kit	Agilent Technologies
Light Cycler 480 RNA Master Hydrolysis Probes (Version 07)	Roche
MasterPure Complete DNA & RNA Purification Kit	Epicenter
MiSeq Reagent Kit v3 150 cycles	Illumina
Molecular grade water	Sigma Aldrich
Nanopore Sequencing Kit SQK-NSK007 R9 version	Oxford Nanopore Technologies
NEBNext mRNA Second Strand Synthesis Module	NEB
Nextera XT DNA Library Preparation Kit	Illumina
Nextera XT Index Kit (24 indices)	Illumina
Pellet Paint Precipitant	Merck Milipore
PhiX Control v3	Illumina
Protector Rnase Inhibitor (2000 U)	Roche
QIAquick PCR Purification-Kit	Qiagen
Qiagen RNeasy-MinElute Cleanup Kit	Qiagen
Qubit dsDNA HS assay	Life Technologies
Random Hexamers (50 µM)	Life Technologies
RNaseOUT Recombinant Ribonuclease Inhibitor (5000 U)	Life Technologies
SuperScript III Reverse Transcriptase (10000 U)	Life Technologies



T4 RNA Ligase 1	NEB
Turbo DNA-free Kit Ambion	Life Technologies
WTA2 Complete Whole Transcriptome Amplification Kit	Sigma Aldrich
T4 gene 32 protein (100 µg)	Roche
RNase H (2 U/µL)	Life Technologies

### III.3.4 Consumables

0.2 mL Thin Wall Tube Flat Cap	Alpha Laboratories
1.5 mL APEX Tough Microtube Natural	Alpha Laboratories
Biosphere Filter Tips, long, 10 µL	Sarstedt
DNA LoBind Microcentrifuge tubes 1.5 mL	Eppendorf
epDualfilter TIPS 0.1-10 µl	Eppendorf
Gilson pipette 1000 µL, 200 µL, 100 µL, 20 µL, 10 µL, 2 µL	Gilson
Lens cleaning tissue	GE Healthcare
Light Cycler 480 multiwell plate 96, white	Roche
Light Cycler 480 Sealing foil	Roche
Maxymum Recovery, Filter tips: 1000 µL, 200 µL, 100 µL, 20 µL	Axygen
Microcentrifuge tubes with flat cap, 2 mL	VWR
Minisart High-Flow single use syringe filter, 0.2 µm	Sartorius
Qubit Assay Tubes, 0.5 mL	Life Technologies
Research plus pipette 1000 µL, 100 µL, 10 µL, 2.5 µL	Eppendorf
Serological pipette: 10 mL, 25 mL	Sarstedt
Spectrum Disinfectant Wipes	Advanced Sterilized Products
Syringe, 20 mL	Therumo
Tissue culture flask, canted neck, non-pyrogenic, non cytotoxic: 25 cm <sup>2</sup> , 75 cm <sup>2</sup> , 175 cm <sup>2</sup>	Sarstedt
Tissue culture plate with 24 wells, flat bottom, with lid	Sarstedt
DynaMag-2 Magnet	Life Technologies
Qubit assay tubes	Life Technologies

### III.3.5 Devices

Agilent 2100 Bioanalyzer	Agilent Technologies
B.S.5726 Class II Flow Cabinet	Peteric Ltd
Easy Breeze Gel Dryer	Hoefer Scientific Instruments
Electrophoresis Apparatus GPS 200/400	LKB/ Pharmacia
Flow cabinet TC-48	Gelaire
Fluorescence microscope TH4-200	Olympus
Heating Block QBD1	Grant
HU6 Mini Horizontal Gel Unit	SCIE-PLAS
Inverted Research Microscope IMT-2	Olympus
MinION MKI	Oxford Nanopore Technologies
MiSeq	Illumin
MyFuge 12 Mini Centrifuge	Benchmark Scientific
NanoDrop ND-1000	Thermo Fisher Scientific
Pipetus, 0,1 mL - 200 mL	Hirschmann
Refrigerated Microcentrifuge 5418 R	Eppendorf
Refrigerated Centrifuge 4-16KS	Sigma
Research CO <sub>2</sub> Incubator GA2010	Leec
Research CO <sub>2</sub> Incubator GA3000	Leec
Rotator, Vortex and Blotter Intelli-Mixer RM-2S	ELMI
Thermocycler TGradient	Biometra
Tube shaker Lab dancer	IKA
WT16 Mini Rocking Platform	Biometra
Light Cycler 480 II	Roche
Qubit 2.0	Life Technologies

## IV References

- Ahmadvand, Sohrab, Mehdi Soltani, Mahdi Behdani, Øystein Evensen, Ehsan Alirahimi, Elahe Soltani, Reza Hassanzadeh, and Javad Ashrafi-Helan. 2018. "VP2 (PTA motif) encoding DNA vaccine confers protection against lethal challenge with infectious pancreatic necrosis virus (IPNV) in trout." *Molecular Immunology* 94:61-67.
- Ahne, W. 1982. "Vergleichende Untersuchungen über die Stabilität von vier fischpathogenen Viren (VHSV, PFR, SVCV, IPNV)." *Zoonoses and Public Health* 29 (6):457-476.
- Ahne, W., and I. Thomsen. 1986. "Infectious pancreatic necrosis: detection of virus and antibodies in rainbow trout IPNV-carrier (*Salmo gairdneri*)." *Zentralbl Veterinarmed B* 33 (7):552-4.
- Altschul, Stephen F, Warren Gish, Webb Miller, Eugene W Myers, and David J Lipman. 1990. "Basic local alignment search tool." *Journal of molecular biology* 215 (3):403-410.
- Ambaradar, S., R. Gupta, D. Trakroo, R. Lal, and J. Vakhlu. 2016. "High Throughput Sequencing: An Overview of Sequencing Chemistry." *Indian J Microbiol* 56 (4):394-404. doi: 10.1007/s12088-016-0606-4.
- Andrews, Simon. 2010. FastQC: a quality control tool for high throughput sequence data.
- Ariel, E, and NJ Olesen. 2002. "Finfish in aquaculture and their diseases-a retrospective view on the European community." *Bulletin-European Association of Fish Pathologists* 22 (2):72-85.
- Bahir, I., M. Fromer, Y. Prat, and M. Linial. 2009. "Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences." *Mol Syst Biol* 5:311. doi: 10.1038/msb.2009.71.
- Bain, N, A Gregory, and RS Raynard. 2008. "Genetic analysis of infectious pancreatic necrosis virus from Scotland." *Journal of fish diseases* 31 (1):37-47.
- Ball, H. J., A. L. Munro, A. Ellis, K. G. Elson, W. Hodgkiss, and I. S. McFarlane. 1971. "Infectious pancreatic necrosis in rainbow trout in Scotland." *Nature* 234 (5329):417-8.
- Bandin, I., S. Souto, J. M. Cutrin, C. Lopez-Vazquez, J. G. Oliveira, C. Esteve, E. Alcaide, and C. P. Dopazo. 2014. "Presence of viruses in wild eels *Anguilla anguilla* L, from the Albufera Lake (Spain)." *J Fish Dis* 37 (7):597-607. doi: 10.1111/jfd.1392.
- Barja, JL, AE Toranzo, ML Lemos, and FM Hetrick. 1983. "Influence of water temperature and salinity on the survival of IPN and IHN viruses [infectious pancreatic necrosis virus, infectious haematopoietic necrosis virus]." *Bulletin of the European Association of Fish Pathologists*.
- Besse, P., and P. de Kinkelin. 1965. "[On the existence in France of pancreatic necrosis in the rainbow trout (*Salmo gairdneri*). (Preliminary note)]." *Bull Acad Vet Fr* 38 (5):185-90.
- Bittner, M, RL Burke, and BM Alberts. 1979. "Purification of the T4 gene 32 protein free from detectable deoxyribonuclease activities." *Journal of Biological Chemistry* 254 (19):9565-9572.
- Blake, S, J-Y Ma, DA Caporale, S Jairath, and BL Nicholson. 2001. "Phylogenetic relationships of aquatic birnaviruses based on deduced amino acid sequences of genome segment A cDNA." *Diseases of Aquatic Organisms* 45 (2):89-102.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. "Trimmomatic: a flexible trimmer for Illumina sequence data." *Bioinformatics* 30 (15):2114-20. doi: 10.1093/bioinformatics/btu170.
- Breslauer, Kenneth J, Ronald Frank, Helmut Blöcker, and Luis A Marky. 1986. "Predicting DNA duplex stability from the base sequence." *Proceedings of the National Academy of Sciences* 83 (11):3746-3750.

- Calvert, Jay G, Eva Nagy, Maritza Soler, and Peter Dobos. 1991. "Characterization of the VPg-dsRNA linkage of infectious pancreatic necrosis virus." *Journal of general virology* 72 (10):2563-2567.
- Carrat, Fabrice, and Antoine Flahault. 2007. "Influenza vaccine: the challenge of antigenic drift." *Vaccine* 25 (39):6852-6862.
- Chen, Y, P Lenert, R Weeratna, M McCluskie, T Wu, HL Davis, and AM Krieg. 2001. "Identification of methylated CpG motifs as inhibitors of the immune stimulatory CpG motifs." *Gene therapy* 8 (13):1024.
- Cheng, Xiaofei, Nasar Virk, Wei Chen, Shuqin Ji, Shuxian Ji, Yuqiang Sun, and Xiaoyun Wu. 2013. "CpG usage in RNA viruses: data and hypotheses." *PLoS one* 8 (9):e74109.
- Christie, K. E., L. S. Havarstein, H. O. Djupvik, S. Ness, and C. Endresen. 1988. "Characterization of a new serotype of infectious pancreatic necrosis virus isolated from Atlantic salmon." *Arch Virol* 103 (3-4):167-77.
- Christie, KE, S Ness, and HO Djupvik. 1990. "Infectious pancreatic necrosis virus in Norway: partial serotyping by monoclonal antibodies." *Journal of Fish Diseases* 13 (4):323-327.
- Chung, HK, SH Lee, HH Lee, DS Lee, and YS Kim. 1994. "Nucleotide sequence analysis of the VP2-NS-VP3 genes of infectious pancreatic necrosis virus DRT strain." *Molecules and Cells (Korea Republic)*.
- Cohen, J, A Poinard, and R Scherrer. 1973. "Physico-chemical and morphological features of infectious pancreatic necrosis virus." *Journal of General Virology* 21 (3):485-498.
- Cohen, Jean. 1975. "Ribonucleic acid polymerase activity in purified infectious pancreatic necrosis virus of trout." *Biochemical and biophysical research communications* 62 (3):689-695.
- Coulibaly, Fasséli, Christophe Chevalier, Bernard Delmas, and Félix A Rey. 2010. "Crystal structure of an Aquabirnavirus particle: insights into antigenic diversity and virulence determinism." *Journal of virology* 84 (4):1792-1799.
- Coulibaly, Fasséli, Christophe Chevalier, Irina Gutsche, Joan Pous, Jorge Navaza, Stéphane Bressanelli, Bernard Delmas, and Félix A Rey. 2005. "The birnavirus crystal structure reveals structural relationships among icosahedral viruses." *Cell* 120 (6):761-772.
- Couve, E., J. Kiss, and J. Kuznar. 1992. "Infectious pancreatic necrosis virus internalization and endocytic organelles in CHSE-214 cells." *Cell Biol Int Rep* 16 (9):899-906.
- Crane, Mark, and Alex Hyatt. 2011. "Viruses of fish: an overview of significant pathogens." *Viruses* 3 (11):2025-2046.
- Crusoe, Michael R, Hussien F Alameldin, Sherine Awad, Elmar Boucher, Adam Caldwell, Reed Cartwright, Amanda Charbonneau, Bede Constantinides, Greg Edvenson, and Scott Fay. 2015. "The khmer software package: enabling efficient nucleotide sequence analysis." *F1000Research* 4.
- Cutrin, J. M., J. L. Barja, B. L. Nicholson, I. Bandin, S. Blake, and C. P. Dopazo. 2004. "Restriction Fragment Length Polymorphisms and Sequence Analysis: an Approach for Genotyping Infectious Pancreatic Necrosis Virus Reference Strains and Other Aquabirnaviruses Isolated from Northwestern Spain." *Applied and Environmental Microbiology* 70 (2):1059-1067. doi: 10.1128/aem.70.2.1059-1067.2004.
- Dadar, M., R. Peyghan, H. R. Memari, M. R. Shapouri, R. Hasanzadeh, L. M. Goudarzi, and V. N. Vakharia. 2013. "Sequence analysis of infectious pancreatic necrosis virus isolated from Iranian reared rainbow trout (*Oncorhynchus mykiss*) in 2012." *Virus Genes* 47 (3):574-8. doi: 10.1007/s11262-013-0981-4.

- Di Paola, Nicholas, Caio César de Melo Freire, and Paolo Marinho de Andrade Zanotto. 2018. "Does adaptation to vertebrate codon usage relate to flavivirus emergence potential?" *PloS one* 13 (1):e0191652.
- Dilcher, Meik, Lekbira Hasib, Marcus Lechner, Nicolas Wieseke, Martin Middendorf, Manja Marz, Andrea Koch, Martin Spiegel, Gerhard Dobler, and Frank T Hufert. 2012. "Genetic characterization of Tribeč virus and Kemerovo virus, two tick-transmitted human-pathogenic Orbiviruses." *Virology* 423 (1):68-76.
- Dixon, P. F., G. H. Ngoh, D. M. Stone, S. F. Chang, K. Way, and S. L. Kueh. 2008. "Proposal for a fourth aquabirnavirus serogroup." *Arch Virol* 153 (10):1937-41. doi: 10.1007/s00705-008-0192-9.
- Do Yew, Tan, Mohd Hair Bejo, Aini Ideris, Abdul Rahman Omar, and Goh Yong Meng. 2004. "Base usage and dinucleotide frequency of infectious bursal disease virus." *Virus Genes* 28 (1):41-53.
- Dobos, PETER. 1977. "Virus-specific protein synthesis in cells infected by infectious pancreatic necrosis virus." *Journal of virology* 21 (1):242-258.
- Dobos, Peter. 1993. "In vitro guanylylation of infectious pancreatic necrosis virus polypeptide VP1." *Virology* 193 (1):403-413.
- Dobos, Peter. 1995a. "The molecular biology of infectious pancreatic necrosis virus (IPNV)." *Annual Review of Fish Diseases* 5:25-54.
- Dobos, Peter. 1995b. "Protein-primed RNA synthesis in vitro by the virion-associated RNA polymerase of infectious pancreatic necrosis virus." *Virology* 208 (1):19-25.
- Dobos, PETER, ROSS Hallett, DT Kells, O Sorensen, and DAVID Rowe. 1977. "Biophysical studies of infectious pancreatic necrosis virus." *Journal of virology* 22 (1):150-159.
- Domingo, E., and J. J. Holland. 1997. "RNA VIRUS MUTATIONS AND FITNESS FOR SURVIVAL." *Annual Review of Microbiology* 51 (1):151-178. doi: 10.1146/annurev.micro.51.1.151.
- Dorson, M. 1988. "Vaccination against infectious pancreatic necrosis." *Fish Vaccination. London, Academic Press Limited*:162-171.
- Drummond, Alexei, M. A. Suchard, D. Xie, and A. Rambaut. 2012. *Bayesian phylogenetics with BEAUti and the BEAST 1.7*. Vol. 22.
- Duncan, R, and P Dobos. 1986. "The nucleotide sequence of infectious pancreatic necrosis virus (IPNV) dsRNA segment A reveals one large ORF encoding a precursor polyprotein." *Nucleic Acids Research* 14 (14):5934.
- Duncan, R, E Nagy, PETER J Krell, and PETER Dobos. 1987. "Synthesis of the infectious pancreatic necrosis virus polyprotein, detection of a virus-encoded protease, and fine structure mapping of genome segment A coding regions." *Journal of virology* 61 (12):3655-3664.
- Duncan, Roy, Carla L Mason, Eva Nagy, Jo-Ann Leong, and Peter Dobos. 1991. "Sequence analysis of infectious pancreatic necrosis virus genome segment B and its encoded VP1 protein: a putative RNA-dependent RNA polymerase lacking the Gly-Asp-Asp motif." *Virology* 181 (2):541-552.
- Flint, Jane, Vincent R. Racaniello, Glenn F. Rall, and Anna Marie Skalka. 2015. *Principles of Virology, Fourth Edition, Bundle*: American Society of Microbiology.
- Frey, Kenneth G, Jesus Enrique Herrera-Galeano, Cassie L Redden, Truong V Luu, Stephanie L Servetas, Alfred J Mateczun, Vishwesh P Mokashi, and Kimberly A Bishop-Lilly. 2014. "Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood." *BMC genomics* 15 (1):96.

- Frost, Petter, Leiv Sigve Håvarstein, Bjarte Lygren, Stefan Ståhl, Curt Endresen, and Karen Elina Christie. 1995. "Mapping of neutralization epitopes on infectious pancreatic necrosis viruses." *Journal of General Virology* 76 (5):1165-1172.
- Fryer, John L., Alexander Yusha, and K. S. Pilcher. 1965. "THE IN VITRO CULTIVATION OF TISSUE AND CELLS OF PACIFIC SALMON AND STEELHEAD TROUT\*." *Annals of the New York Academy of Sciences* 126 (1):566-586. doi: 10.1111/j.1749-6632.1965.tb14303.x.
- Gadan, K., A. Sandtro, I. S. Marjara, N. Santi, H. M. Munang'andu, and O. Evensen. 2013. "Stress-induced reversion to virulence of infectious pancreatic necrosis virus in naive fry of Atlantic salmon (*Salmo salar* L.)." *PLoS One* 8 (2):e54656. doi: 10.1371/journal.pone.0054656.
- Galloux, M., C. Chevalier, C. Henry, J. C. Huet, B. D. Costa, and B. Delmas. 2004. "Peptides resulting from the pVP2 C-terminal processing are present in infectious pancreatic necrosis virus particles." *J Gen Virol* 85 (Pt 8):2231-6. doi: 10.1099/vir.0.80012-0.
- Galloux, Marie, Christophe Chevalier, Celine Henry, Jean-Claude Huet, Bruno Da Costa, and Bernard Delmas. 2004. "Peptides resulting from the pVP2 C-terminal processing are present in infectious pancreatic necrosis virus particles." *Journal of general virology* 85 (8):2231-2236.
- Gargis, A. S., L. Kalman, M. W. Berry, D. P. Bick, D. P. Dimmock, T. Hambuch, F. Lu, E. Lyon, K. V. Voelkerding, B. A. Zehnbaauer, R. Agarwala, S. F. Bennett, B. Chen, E. L. Chin, J. G. Compton, S. Das, D. H. Farkas, M. J. Ferber, B. H. Funke, M. R. Furtado, L. M. Ganova-Raeva, U. Geigenmuller, S. J. Gunselman, M. R. Hegde, P. L. Johnson, A. Kasarskis, S. Kulkarni, T. Lenk, C. S. Liu, M. Manion, T. A. Manolio, E. R. Mardis, J. D. Merker, M. S. Rajeevan, M. G. Reese, H. L. Rehm, B. B. Simen, J. M. Yeakley, J. M. Zook, and I. M. Lubin. 2012. "Assuring the quality of next-generation sequencing in clinical laboratory practice." *Nat Biotechnol* 30 (11):1033-6. doi: 10.1038/nbt.2403.
- Geoghegan, F, M Ó Cinneide, and NM Ruane. 2007. "Infectious pancreatic necrosis virus and its impact on the Irish salmon aquaculture and wild fish sectors."
- Gheyas, AA, CS Haley, DR Guy, A Hamilton, AE Tinch, JC Mota-Velasco, and JA Woolliams. 2010. "Effect of a major QTL affecting IPN resistance on production traits in Atlantic salmon." *Animal genetics* 41 (6):666-668.
- Glenn, T. C. 2011. "Field guide to next-generation DNA sequencers." *Mol Ecol Resour* 11 (5):759-69. doi: 10.1111/j.1755-0998.2011.03024.x.
- Glenney, G. W., P. A. Barbash, J. A. Coll, and W. M. Quartz. 2012. "Isolation and molecular characterization of a novel infectious pancreatic necrosis virus strain in returning Atlantic salmon *Salmo salar* from the Connecticut River, USA." *J Aquat Anim Health* 24 (2):63-72.
- Goldberg, Billi, Howard B Urnovitz, and Raphael B Stricker. 2000. "Beyond danger: unmethylated CpG dinucleotides and the immunopathogenesis of disease." *Immunology letters* 73 (1):13-18.
- Goldman, N., and Z. Yang. 1994. "A codon-based model of nucleotide substitution for protein-coding DNA sequences." *Mol Biol Evol* 11 (5):725-36.
- Gomez-Casado, E., A. Estepa, and J. M. Coll. 2011. "A comparative review on European-farmed finfish RNA viruses and their vaccines." *Vaccine* 29 (15):2657-71. doi: 10.1016/j.vaccine.2011.01.097.
- Goodwin, S., J. D. McPherson, and W. R. McCombie. 2016. "Coming of age: ten years of next-generation sequencing technologies." *Nat Rev Genet* 17 (6):333-51. doi: 10.1038/nrg.2016.49.

- Gorbalenya, Alexander E, Fiona M Pringle, Jean-Louis Zeddari, Brian T Luke, Craig E Cameron, James Kalmakoff, Terry N Hanzlik, Karl HJ Gordon, and Vernon K Ward. 2002. "The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage." *Journal of molecular biology* 324 (1):47-62.
- Goujon, Mickael, Hamish McWilliam, Weizhong Li, Franck Valentin, Silvano Squizzato, Juri Paern, and Rodrigo Lopez. 2010. "A new bioinformatics analysis tools framework at EMBL–EBI." *Nucleic Acids Research* 38 (suppl\_2):W695-W699. doi: 10.1093/nar/gkq313.
- Graham, Stephen C, L Peter Sarin, Mohammad W Bahar, Reg A Myers, David I Stuart, Dennis H Bamford, and Jonathan M Grimes. 2011. "The N-terminus of the RNA polymerase from infectious pancreatic necrosis virus is the determinant of genome attachment." *PLoS pathogens* 7 (6):e1002085.
- Gudding, Roar, Atle Lillehaug, and Oystein Evensen. 2014. *Fish vaccination*: John Wiley & Sons.
- Hanada, Kousuke, Yoshiyuki Suzuki, and Takashi Gojobori. 2004. "A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes." *Molecular biology and evolution* 21 (6):1074-1080.
- Havarstein, L. S., K. H. Kalland, K. E. Christie, and C. Endresen. 1990. "Sequence of the large double-stranded RNA segment of the N1 strain of infectious pancreatic necrosis virus: a comparison with other Birnaviridae." *J Gen Virol* 71 ( Pt 2):299-308. doi: 10.1099/0022-1317-71-2-299.
- Hedrick, RP, and JL Fryer. 1981. "Persistent infection of three salmonid cell lines with infectious pancreatic necrosis virus (IPNV)." *Fish Pathology* 15 (3-4):163-172.
- Heininger, U, NS Bachtar, P Bahri, A Dana, A Dodoo, J Gidudu, and E Matos dos Santos. 2012. "The concept of vaccination failure." *Vaccine* 30 (7):1265-1268.
- Heppell, Joël, Laurent Berthiaume, François Corbin, Esther Tarrab, Jacqueline Lecomte, and Maximilien Arella. 1993. "Comparison of Amino Acid Sequences Deduced from a cDNA Fragment Obtained from Infectious Pancreatic Necrosis Virus (IPNV) Strains of Different Serotypes." *Virology* 195 (2):840-844. doi: <https://doi.org/10.1006/viro.1993.1441>.
- Heppell, Joël, Esther Tarrab, Laurent Berthiaume, Jacqueline Lecomte, and Maximilien Arella. 1995. "Characterization of the small open reading frame on genome segment A of infectious pancreatic necrosis virus." *Journal of general virology* 76 (8):2091-2096.
- Hill, BJ, and K Way. 1995. "Serological classification of infectious pancreatic necrosis (IPN) virus and other aquatic birnaviruses." *Annual Review of Fish Diseases* 5:55-77.
- Hirayama, Takeshi, Ichiro Nagano, Hajime Shinmoto, Ken-ichi Yagyu, and Syun-ichirou Oshima. 2007. "Isolation and characterization of virulent yellowtail ascites virus." *Microbiology and immunology* 51 (4):397-406.
- Hjeltnes, Br, C Walde, B Bang Jensen, and A Haukaas. 2017. "Fish Health Report 2016." *Nowegian Veterinary Institute, Oslo*.
- Holopainen, R., A. M. Eriksson-Kallio, and T. Gadd. 2017. "Molecular characterisation of infectious pancreatic necrosis viruses isolated from farmed fish in Finland." *Arch Virol* 162 (11):3459-3471. doi: 10.1007/s00705-017-3525-8.



- Hong, Jiann-Ruey, Hong-Yi Gong, and Jen-Leih Wu. 2002. "IPNV VP5, a novel anti-apoptosis gene of the Bcl-2 family, regulates Mcl-1 and viral protein expression." *Virology* 295 (2):217-229.
- Houston, R. D., C. S. Haley, A. Hamilton, D. R. Guy, J. C. Mota-Velasco, A. A. Gheyas, A. E. Tinch, J. B. Taggart, J. E. Bron, W. G. Starkey, B. J. McAndrew, D. W. Verner-Jeffreys, R. K. Paley, G. S. E. Rimmer, I. J. Tew, and S. C. Bishop. 2009. "The susceptibility of Atlantic salmon fry to freshwater infectious pancreatic necrosis is largely explained by a major QTL." *Heredity* 105:318. doi: 10.1038/hdy.2009.171.
- Houston, RD, SC Bishop, A Hamilton, DR Guy, AE Tinch, JB Taggart, A Derayat, BJ McAndrew, and CS Haley. 2009. "Detection of QTL affecting harvest traits in a commercial Atlantic salmon population." *Animal genetics* 40 (5):753-755.
- Houston, Ross D, Chris S Haley, Alastair Hamilton, Derrick R Guy, Alan E Tinch, John B Taggart, Brendan J McAndrew, and Stephen C Bishop. 2008. "Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*)." *Genetics* 178 (2):1109-1115.
- Huang, MT, DS Manning, M Warner, EB Stephens, and JC Leong. 1986. "A physical map of the viral genome for infectious pancreatic necrosis virus Sp: analysis of cell-free translation products derived from viral cDNA clones." *Journal of virology* 60 (3):1002-1011.
- Huelsenbeck, John P, and Fredrik Ronquist. 2001. "MRBAYES: Bayesian inference of phylogenetic trees." *Bioinformatics* 17 (8):754-755.
- Huelsenbeck, John P, Fredrik Ronquist, Rasmus Nielsen, and Jonathan P Bollback. 2001. "Bayesian inference of phylogeny and its impact on evolutionary biology." *science* 294 (5550):2310-2314.
- Hulo, Chantal, Edouard De Castro, Patrick Masson, Lydie Bougueleret, Amos Bairoch, Ioannis Xenarios, and Philippe Le Mercier. 2010. "ViralZone: a knowledge resource to understand virus diversity." *Nucleic acids research* 39 (suppl\_1):D576-D582.
- Huson, Daniel H, and David Bryant. 2005. "Application of phylogenetic networks in evolutionary studies." *Molecular biology and evolution* 23 (2):254-267.
- Jørgensen, Jorunn B., Audny Johansen, Børge Stenersen, and Ann-Inger Sommer. 2001. "CpG oligodeoxynucleotides and plasmid DNA stimulate Atlantic salmon (*Salmo salar* L.) leucocytes to produce supernatants with antiviral activity." *Developmental & Comparative Immunology* 25 (4):313-321. doi: [https://doi.org/10.1016/S0145-305X\(00\)00068-9](https://doi.org/10.1016/S0145-305X(00)00068-9).
- Jorgensen, PE, and F Bregnballe. 1969. "Infectious pancreatic necrosis in rainbow trout (*Salmo gairdneri*) in Denmark." *Nordisk Veterinaer Medicin* 21 (3):142-+.
- Jorquera, Esteban, Paz Morales, David Tapia, Pamela Torres, Yoanna Eissler, Juan C Espinoza, Pablo Conejeros, and Juan Kuznar. 2016. "Chilean IPNV isolates: Robustness analysis of PCR detection." *Electronic Journal of Biotechnology* 19 (2):28-32.
- Julin, K, L-H Johansen, A-I Sommer, and JB Jørgensen. 2015. "Persistent infections with infectious pancreatic necrosis virus (IPNV) of different virulence in Atlantic salmon, *Salmo salar* L." *Journal of fish diseases* 38 (11):1005-1019.
- Julin, K, S Mennen, and A-I Sommer. 2013. "Study of virulence in field isolates of infectious pancreatic necrosis virus obtained from the northern part of Norway." *Journal of fish diseases* 36 (2):89-102.
- Kaehler, Benjamin D, Von Bing Yap, Rongli Zhang, and Gavin A Huttley. 2014. "Genetic distance for a general non-stationary markov substitution process." *Systematic biology* 64 (2):281-293.

- Karlin, S, W Doerfler, and LR Cardon. 1994. "Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses?" *Journal of virology* 68 (5):2889-2897.
- Karlin, SAMUEL, B EDWIN Blaisdell, and GABRIEL A Schachtel. 1990. "Contrasts in codon usage of latent versus productive genes of Epstein-Barr virus: data and hypotheses." *Journal of virology* 64 (9):4264-4273.
- Kingman, John FC. 1982a. "On the genealogy of large populations." *Journal of Applied Probability* 19 (A):27-43.
- Kingman, John Frank Charles. 1982b. "The coalescent." *Stochastic processes and their applications* 13 (3):235-248.
- Kosakovsky Pond, Sergei L, and Simon DW Frost. 2005. "Not so different after all: a comparison of methods for detecting amino acid sites under selection." *Molecular biology and evolution* 22 (5):1208-1222.
- Kuznar, J, M Soler, GILDA Farias, and JC Espinoza. 1995. "Attachment and entry of infectious pancreatic necrosis virus (IPNV) into CHSE-214 cells." *Archives of virology* 140 (10):1833-1840.
- Lago, María, Isabel Bandín, José G. Oliveira, and Carlos P. Dopazo. 2017. "In vitro reassortment between Infectious Pancreatic Necrosis Virus (IPNV) strains: The mechanisms involved and its effect on virulence." *Virology* 501 (Supplement C):1-11. doi: <https://doi.org/10.1016/j.virol.2016.11.003>.
- Langmead, Ben, and Steven L. Salzberg. 2012. "Fast gapped-read alignment with Bowtie 2." *Nat Meth* 9 (4):357-359. doi: 10.1038/nmeth.1923  
<http://www.nature.com/nmeth/journal/v9/n4/abs/nmeth.1923.html - supplementary-information>.
- Lauksund, S., L. Greiner-Tollersrud, C. J. Chang, and B. Robertsen. 2015. "Infectious pancreatic necrosis virus proteins VP2, VP3, VP4 and VP5 antagonize IFN $\alpha$ 1 promoter activation while VP1 induces IFN $\alpha$ 1." *Virus Res* 196:113-21. doi: 10.1016/j.virusres.2014.11.018.
- Lee, Hyung-Hoan, Hye-Kyung Chung, and Seong-Hun Lee. 1994. "Nucleotide sequence analysis of the RNA-dependent RNA polymerase gene of infectious pancreatic necrosis virus DRT strain." *Journal of Microbiology and Biotechnology* 4 (4):264-269.
- Leggett, Richard M, Darren Heavens, Mario Caccamo, Matthew D Clark, and Robert P Davey. 2015. "NanoOK: multi-reference alignment analysis of nanopore sequencing data, quality and error profiles." *Bioinformatics* 32 (1):142-144.
- Lemey, P., Salemi, M., & Vandamme, A. (Eds.). 2009. *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. 2 ed. Cambridge: Cambridge University Press.
- Levican, J., C. Miranda-Cardenas, R. Soto-Rifo, F. Aguayo, A. Gaggero, and O. Leon. 2017. "Infectious pancreatic necrosis virus enters CHSE-214 cells via macropinocytosis." *Sci Rep* 7 (1):3068. doi: 10.1038/s41598-017-03036-w.
- Li, Heng. 2013. "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM." *arXiv preprint arXiv:1303.3997*.
- Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and Subgroup Genome Project Data Processing. 2009. "The Sequence Alignment/Map format and SAMtools." *Bioinformatics* 25 (16):2078-2079. doi: 10.1093/bioinformatics/btp352.
- Li, Pan TX, Jeffrey Viereggs, and Ignacio Tinoco Jr. 2008. "How RNA unfolds and refolds." *Annu. Rev. Biochem.* 77:77-100.

- Lightner, Donald, and George Post. 1969. "Morphological Characteristics of Infectious Pancreatic Necrosis Virus in Trout Pancreatic Tissue." *Journal of the Fisheries Board of Canada* 26 (8):2247-2250.
- Liu, L., Y. Li, S. Li, N. Hu, Y. He, R. Pong, D. Lin, L. Lu, and M. Law. 2012. "Comparison of next-generation sequencing systems." *J Biomed Biotechnol* 2012:251364. doi: 10.1155/2012/251364.
- Loman, N. J., and A. R. Quinlan. 2014. "Poretools: a toolkit for analyzing nanopore sequence data." *Bioinformatics* 30 (23):3399-401. doi: 10.1093/bioinformatics/btu555.
- Lorenzen, E., B. Carstensen, and N. J. Olesen. 1999. "Inter-laboratory comparison of cell lines for susceptibility to three viruses: VHSV, IHNV and IPNV." *Dis Aquat Organ* 37 (2):81-8. doi: 10.3354/dao037081.
- Lu, H., F. Giordano, and Z. Ning. 2016. "Oxford Nanopore MinION Sequencing and Genome Assembly." *Genomics Proteomics Bioinformatics* 14 (5):265-279. doi: 10.1016/j.gpb.2016.05.004.
- Lu, Zhen, Lizhou Zhang, Nian Wang, Yuming Chen, Li Gao, Yongqiang Wang, Honglei Gao, Yulong Gao, Kai Li, and Xiaole Qi. 2015. "Naturally occurring reassortant infectious bursal disease virus in northern China." *Virus research* 203:92-95.
- Lunter, Gerton, and Martin Goodson. 2011. "Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads." *Genome research* 21 (6):936-939.
- Maan, Sushila, Shujing Rao, Narender Singh Maan, Simon John Anthony, Houssam Attoui, Alan Richard Samuel, and Peter Paul Clement Mertens. 2007. "Rapid cDNA synthesis and sequencing techniques for the genetic study of bluetongue and other dsRNA viruses." *Journal of Virological Methods* 143 (2):132-139. doi: <https://doi.org/10.1016/j.jviromet.2007.02.016>.
- MacDonald, R. D., and J. C. Kennedy. 1979. "Infectious pancreatic necrosis virus persistently infects chinook salmon embryo cells independent of interferon." *Virology* 95 (1):260-4.
- Manning, D Scott, and JC Leong. 1990. "Expression in Escherichia coli of the large genomic segment of infectious pancreatic necrosis virus." *Virology* 179 (1):16-25.
- McKnight, I. J., and R. J. Roberts. 1976. "The pathology of infectious pancreatic necrosis. I. The sequential histopathology of the naturally occurring condition." *Br Vet J* 132 (1):76-85.
- McMichael, John, JL Fryer, and KS Pilcher. 1975. "An antigenic comparison of three strains of infectious pancreatic necrosis virus of salmonid fishes." *Aquaculture* 6 (3):203-210.
- Melby, HP, and KE Christie. 1994. "Antigenic analysis of reference strains and Norwegian field strains of aquatic birnaviruses by the use of six monoclonal antibodies produced against the infectious pancreatic necrosis virus N1 strain." *Journal of Fish Diseases* 17 (4):409-415.
- Mertens, Peter PC, Peter B Jamieson, and Peter Dobos. 1982. "In vitro RNA synthesis by infectious pancreatic necrosis virus-associated RNA polymerase." *Journal of General Virology* 59 (1):47-56.
- Metzker, M. L. 2010. "Sequencing technologies - the next generation." *Nat Rev Genet* 11 (1):31-46. doi: 10.1038/nrg2626.
- Moen, Thomas, Jacob Torgersen, Nina Santi, William S. Davidson, Matthew Baranski, Jørgen Ødegård, Sissel Kjølglum, Bente Velle, Matthew Kent, Krzysztof P. Lubieniecki, Eivind Isdal, and Sigbjørn Lien. 2015. "Epithelial Cadherin Determines Resistance to Infectious Pancreatic Necrosis Virus in Atlantic Salmon." *Genetics* 200 (4):1313-1326. doi: 10.1534/genetics.115.175406.

- Mohr, P. G., N. J. Moody, L. M. Williams, J. Hoad, and J. Crane M. St. 2015. "Molecular characterization of Tasmanian aquabirnaviruses from 1998 to 2013." *Dis Aquat Organ* 116 (1):1-9. doi: 10.3354/dao02903.
- Moreno, Patricia, José G Olveira, Alejandro Labella, Juan Manuel Cutrín, Jorge C Baro, Juan Jose Borrego, and Carlos P Dopazo. 2014. "Surveillance of viruses in wild fish populations in areas around the Gulf of Cadiz (South Atlantic Iberian Peninsula)." *Applied and environmental microbiology:AEM*. 02090-14.
- Mortensen, S. H., R. K. Nilsen, and B. Hjeltnes. 1998. "Stability of an infectious pancreatic necrosis virus (IPNV) isolate stored under different laboratory conditions." *Dis Aquat Organ* 33 (1):67-71. doi: 10.3354/dao033067.
- Munro, A. L., and I. S. Wallace. 2017. SCOTTISH FISH FARM PRODUCTION SURVEY 2016. Edinburgh, UK: Marine Scotland Science.
- Munro, E. S., and P. J. Midthlyng. 2011. *Infectious pancreatic necrosis and associated aquatic birnaviruses*. Edited by Woo P. T. K. and Bruno D. W. Vol. 3.
- Murray, Alexander G, Corina D Busby, and David W Bruno. 2003. "Infectious pancreatic necrosis virus in Scottish Atlantic salmon farms, 1996–2001." *Emerging Infectious Diseases* 9 (4):455.
- Murrell, B., J. O. Wertheim, S. Moola, T. Weighill, K. Scheffler, and S. L. Kosakovsky Pond. 2012. "Detecting individual sites subject to episodic diversifying selection." *PLoS Genet* 8 (7):e1002764. doi: 10.1371/journal.pgen.1002764.
- Muse, Spencer V, and Brandon S Gaut. 1994. "A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome." *Molecular biology and evolution* 11 (5):715-724.
- Mutoloki, S, HM Munang'andu, and Ø Evensen. 2013. "Clinical and subclinical forms of infectious pancreatic necrosis virus infections show specific viral genetic fingerprints that link differences in virulence to immunogenicity." *Fish & Shellfish Immunology* 34 (6):1667.
- Mutoloki, S., and O. Evensen. 2011. "Sequence similarities of the capsid gene of Chilean and European isolates of infectious pancreatic necrosis virus point towards a common origin." *J Gen Virol* 92 (Pt 7):1721-6. doi: 10.1099/vir.0.030270-0.
- Mutoloki, S., T. B. Jossund, G. Ritchie, H. M. Munang'andu, and O. Evensen. 2016. "Infectious Pancreatic Necrosis Virus Causing Clinical and Subclinical Infections in Atlantic Salmon Have Different Genetic Fingerprints." *Front Microbiol* 7:1393. doi: 10.3389/fmicb.2016.01393.
- Nagy, Eva, Roy Duncan, Peter Krell, and Peter Dobos. 1987. "Mapping of the large RNA genome segment of infectious pancreatic necrosis virus by hybrid arrested translation." *Virology* 158 (1):211-217.
- Nishizawa, T., S. Kinoshita, and M. Yoshimizu. 2005. "An approach for genogrouping of Japanese isolates of aquabirnaviruses in a new genogroup, VII, based on the VP2/NS junction region." *J Gen Virol* 86 (Pt 7):1973-8. doi: 10.1099/vir.0.80438-0.
- Nobiron, I., M. Galloux, C. Henry, C. Torhy, P. Boudinot, N. Lejal, B. Da Costa, and B. Delmas. 2008. "Genome and polypeptides characterization of Tellina virus 1 reveals a fifth genetic cluster in the Birnaviridae family." *Virology* 371 (2):350-61. doi: 10.1016/j.virol.2007.09.022.
- Nurk, S., A. Bankevich, D. Antipov, A. A. Gurevich, A. Korobeynikov, A. Lapidus, A. D. Prjibelski, A. Pyshkin, A. Sirotkin, Y. Sirotkin, R. Stepanauskas, S. R. Clingenpeel, T. Woyke, J. S. McLean, R. Lasken, G. Tesler, M. A. Alekseyev, and P. A. Pevzner. 2013.

- "Assembling single-cell genomes and mini-metagenomes from chimeric MDA products." *J Comput Biol* 20 (10):714-37. doi: 10.1089/cmb.2013.0084.
- Officer, Julius E. 1964. "Ability of a Fish Cell Line to Support the Growth of Mammalian Viruses." *Proceedings of the Society for Experimental Biology and Medicine* 116 (1):190-194. doi: 10.3181/00379727-116-29198.
- OIE. 2000. "Diagnostic Manual for Aquatic Diseases." In, edited by OIE Fish Diseases Commission. France: OFFICE INTERNATIONAL DES EPIZOOTIES, 2000.
- Okamoto, N., R. Yasutomi, H. Shibazaki, S. Hanzawa, and T. Sano. 1987. "The influence of immersing temperature for inoculation with IPNV and/or rearing temperature on mortality of rainbow trout fry postinfection." *Nippon Suisan Gakkaishi (Bulletin of the Japanese Society of Scientific Fisheries)* 53 (7):1125-1128.
- Okamoto, Nobuaki, Takuo Tayama, Makoto Kawanobe, Noboru Fujiki, Yoshihiro Yasuda, and Tokuo Sano. 1993. "Resistance of a rainbow trout strain to infectious pancreatic necrosis." *Aquaculture* 117 (1):71-76. doi: [https://doi.org/10.1016/0044-8486\(93\)90124-H](https://doi.org/10.1016/0044-8486(93)90124-H).
- Okonechnikov, Konstantin, Olga Golosova, Mikhail Fursov, and UGENE team. 2012. "Unipro UGENE: a unified bioinformatics toolkit." *Bioinformatics* 28 (8):1166-1167.
- Olsen, AB, and H Hellberg. 2012. "Fiskehelsesrapporten 2011." *Norwegian Veterinary Institute*.
- Otto, Thomas D., Gary P. Dillon, Wim S. Degraeve, and Matthew Berriman. 2011. "RATT: Rapid Annotation Transfer Tool." *Nucleic Acids Research* 39 (9):e57-e57. doi: 10.1093/nar/gkq1268.
- Patton, J. T., and E. Spencer. 2000. "Genome replication and packaging of segmented double-stranded RNA viruses." *Virology* 277 (2):217-25. doi: 10.1006/viro.2000.0645.
- Pavesi, Angelo, Gkikas Magiorkinis, and David G Karlin. 2013. "Viral proteins originated de novo by overprinting can be identified by codon usage: application to the "gene nursery" of Deltaretroviruses." *PLoS computational biology* 9 (8):e1003162.
- Pedersen, Torunn, Astrid Skjesol, and Jorunn B Jørgensen. 2007. "VP3, a structural protein of infectious pancreatic necrosis virus, interacts with RNA-dependent RNA polymerase VP1 and with double-stranded RNA." *Journal of virology* 81 (12):6652-6663.
- Pereira, F, and A Amorim. 2013. "Evolution: Viruses."
- Petit, Stéphanie, Nathalie Lejal, Jean-Claude Huet, and Bernard Delmas. 2000. "Active residues and viral substrate cleavage sites of the protease of the birnavirus infectious pancreatic necrosis virus." *Journal of virology* 74 (5):2057-2066.
- Pond, Sergei L Kosakovsky, and Simon DW Frost. 2005. "Datamonkey: rapid detection of selective pressure on individual sites of codon alignments." *Bioinformatics* 21 (10):2531-2533.
- Pond, Sergei L. Kosakovsky, Simon D. W. Frost, and Spencer V. Muse. 2005. "HyPhy: hypothesis testing using phylogenies." *Bioinformatics* 21 (5):676-679. doi: 10.1093/bioinformatics/bti079.
- Puigbo, P., I. G. Bravo, and S. Garcia-Vallve. 2008a. "CAIcal: a combined set of tools to assess codon usage adaptation." *Biol Direct* 3:38. doi: 10.1186/1745-6150-3-38.
- Puigbo, P., I. G. Bravo, and S. Garcia-Vallve. 2008b. "E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI)." *BMC Bioinformatics* 9:65. doi: 10.1186/1471-2105-9-65.
- Rambaut, A. 2012. FigTree v. 1.4.0.

- Ramstad, A, and PJ Midtlyng. 2008. "Strong genetic influence on IPN vaccination-and-challenge trials in Atlantic salmon, *Salmo salar* L." *Journal of fish diseases* 31 (8):567-578.
- Ramstad, A., A. B. Romstad, D. H. Knappskog, and P. J. Midtlyng. 2007. "Field validation of experimental challenge models for IPN vaccines." *J Fish Dis* 30 (12):723-31. doi: 10.1111/j.1365-2761.2007.00858.x.
- Rannala, Bruce, and Ziheng Yang. 1996. "Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference." *Journal of molecular evolution* 43 (3):304-311.
- Rice, Peter, Ian Longden, and Alan Bleasby. 2000. EMBOSS: the European molecular biology open software suite. Elsevier Current Trends.
- Rimstad, Espen. 2014. "Vaccination against Infectious Pancreatic Necrosis." *Fish Vaccination*:303-312.
- Roberts, R. J., and I. J. McKnight. 1976. "The pathology of infectious pancreatic necrosis. II. Stress-mediated recurrence." *Br Vet J* 132 (2):209-14.
- Roberts, R. J., and M. D. Pearson. 2005. "Infectious pancreatic necrosis in Atlantic salmon, *Salmo salar* L." *J Fish Dis* 28 (7):383-90. doi: 10.1111/j.1365-2761.2005.00642.x.
- Rodriguez Saint-Jean, S., J. J. Borrego, and S. I. Perez-Prieto. 2003. "Infectious pancreatic necrosis virus: biology, pathogenesis, and diagnostic methods." *Adv Virus Res* 62:113-65.
- Rodríguez-Ezpeleta, Naiara, Michael Hackenberg, and Ana M Aransay. 2011. *Bioinformatics for high throughput sequencing*: Springer Science & Business Media.
- Romero-Brey, I, I Bandín, JM Cutrín, VN Vakharia, and CP Dopazo. 2009. "Genetic analysis of aquabirnaviruses isolated from wild fish reveals occurrence of natural reassortment of infectious pancreatic necrosis virus." *Journal of fish diseases* 32 (7):585-595.
- Ronneseth, A., E. F. Pettersen, and H. I. Wergeland. 2012. "Flow cytometry assay for intracellular detection of Infectious Pancreatic Necrosis virus (IPNV) in Atlantic salmon (*Salmo salar* L.) leucocytes." *Fish Shellfish Immunol* 33 (6):1292-302. doi: 10.1016/j.fsi.2012.09.020.
- Ruane, Neil Martin, Stephen John McCleary, Lorraine Julie McCarthy, and Kathleen Henshilwood. 2015. "Phylogenetic analysis of infectious pancreatic necrosis virus in Ireland reveals the spread of a virulent genogroup 5 subtype previously associated with imports." *Archives of virology* 160 (3):817-824.
- Sabath, Niv, Andreas Wagner, and David Karlin. 2012. "Evolution of viral proteins originated de novo by overprinting." *Molecular biology and evolution* 29 (12):3767-3780.
- Santi, N., V. N. Vakharia, and O. Evensen. 2004a. "Identification of putative motifs involved in the virulence of infectious pancreatic necrosis virus." *Virology* 322 (1):31-40. doi: 10.1016/j.virol.2003.12.016.
- Santi, Nina, Ane Sandtrø, Hilde Sindre, Haichen Song, Jiann-Ruey Hong, Beate Thu, Jen-Leih Wu, Vikram N Vakharia, and Øystein Evensen. 2005. "Infectious pancreatic necrosis virus induces apoptosis in vitro and in vivo independent of VP5 expression." *Virology* 342 (1):13-25.
- Santi, Nina, Haichen Song, Vikram N Vakharia, and Øystein Evensen. 2005. "Infectious pancreatic necrosis virus VP5 is dispensable for virulence and persistence." *Journal of virology* 79 (14):9206-9216.
- Santi, Nina, Vikram N Vakharia, and Øystein Evensen. 2004b. "Identification of putative motifs involved in the virulence of infectious pancreatic necrosis virus." *Virology* 322 (1):31-40.

- Schmieder, Robert, and Robert Edwards. 2011a. "Fast identification and removal of sequence contamination from genomic and metagenomic datasets." *PloS one* 6 (3):e17288.
- Schmieder, Robert, and Robert Edwards. 2011b. "Quality control and preprocessing of metagenomic datasets." *Bioinformatics* 27 (6):863-864.
- Serra, Martin J, and Douglas H Turner. 1995. "[11] Predicting thermodynamic properties of RNA." *Methods in enzymology* 259:242-261.
- Shamoo, Yousif, Alan M. Friedman, Mark R. Parsons, William H. Konigsberg, and Thomas A. Steitz. 1995. "Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA." *Nature* 376 (6538):362-366.
- Shao, Wenhan, Xinxin Li, Mohsan Ullah Goraya, Song Wang, and Ji-Long Chen. 2017. "Evolution of Influenza A Virus by Mutation and Re-Assortment." *International journal of molecular sciences* 18 (8):1650.
- Sharp, P. M., T. M. Tuohy, and K. R. Mosurski. 1986. "Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes." *Nucleic Acids Res* 14 (13):5125-43.
- Shereda, Robert D, Alexander G Kozlov, Timothy M Lohman, Michael M Cox, and James L Keck. 2008. "SSB as an organizer/mobilizer of genome maintenance complexes." *Critical reviews in biochemistry and molecular biology* 43 (5):289-318.
- Shivappa, RB, H Song, K Yao, A Aas-Eng, Ø Evensen, and VN Vakharia. 2004. "Molecular characterization of Sp serotype strains of infectious pancreatic necrosis virus exhibiting differences in virulence." *Diseases of aquatic organisms* 61 (1-2):23-32.
- Shwed, Philip S, Peter Dobos, Lynne A Cameron, Vikram N Vakharia, and Roy Duncan. 2002. "Birnavirus VP1 proteins form a distinct subgroup of RNA-dependent RNA polymerases lacking a GDD motif." *Virology* 296 (2):241-250.
- Sievers, Fabian, Andreas Wilm, David Dineen, Toby J Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Söding, Julie D Thompson, and Desmond G Higgins. 2011. "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." *Molecular Systems Biology* 7 (1). doi: 10.1038/msb.2011.75.
- Silva, Fernanda MF, Pedro MP Vidigal, Luciana W Myrrha, Juliana LR Fietto, Abelardo Silva, and Márcia R Almeida. 2013. "Tracking the molecular epidemiology of Brazilian Infectious bursal disease virus (IBDV) isolates." *Infection, Genetics and Evolution* 13:18-26.
- Silvestro, Daniele, and Ingo Michalak. 2012. "raxmlGUI: a graphical front-end for RAxML." *Organisms Diversity & Evolution* 12 (4):335-337. doi: 10.1007/s13127-011-0056-0.
- Simpson, Jared T., Kim Wong, Shaun D. Jackman, Jacqueline E. Schein, Steven J. M. Jones, and İnanç Birol. 2009. "ABySS: A parallel assembler for short read sequence data." *Genome Research* 19 (6):1117-1123. doi: 10.1101/gr.089532.108.
- Skjelstad, Bård. 2003. *IPN in salmonids: a review*. [Trondheim]: VESO.
- Skjesol, Astrid, Toril Aamo, Marit Nøst Hegseth, Børre Robertsen, and Jorunn B Jørgensen. 2009. "The interplay between infectious pancreatic necrosis virus (IPNV) and the IFN system: IFN signaling is inhibited by IPNV infection." *Virus research* 143 (1):53-60.
- Skjesol, Astrid, Ingrid Skjæveland, Marianne Elnæs, Gerrit Timmerhaus, Børge N Fredriksen, Sven Martin Jørgensen, Aleksei Krasnov, and Jorunn B Jørgensen. 2011. "IPNV with high and low virulence: host immune responses and viral mutations during infection." *Virology journal* 8 (1):396.

- Smail, D. A., D. W. Bruno, G. Dear, L. A. McFarlane, and K. Ross. 1992. "Infectious pancreatic necrosis (IPN) virus Sp serotype in farmed Atlantic salmon, *Salmo salar* L., post-smolts associated with mortality and clinical disease." *Journal of Fish Diseases* 15 (1):77-83. doi: 10.1111/j.1365-2761.1992.tb00639.x.
- Somogyi, Paul, and Peter Dobos. 1980. "Virus-specific RNA synthesis in cells infected by infectious pancreatic necrosis virus." *Journal of virology* 33 (1):129-139.
- Song, Haichen, Jennifer L Baxter-Roshek, Jonathan D Dinman, and Vikram N Vakharia. 2006. "Efficient expression of the 15-kDa form of infectious pancreatic necrosis virus VP5 by suppression of a UGA codon." *Virus research* 122 (1):61-68.
- Song, Haichen, Nina Santi, Øystein Evensen, and Vikram N Vakharia. 2005. "Molecular determinants of infectious pancreatic necrosis virus virulence and cell culture adaptation." *Journal of virology* 79 (16):10289-10299.
- Stamatakis, Alexandros. 2006. "RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models." *Bioinformatics* 22 (21):2688-2690.
- Stamatakis, Alexandros. 2014. "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies." *Bioinformatics* 30 (9):1312-1313.
- Stamatakis, Alexandros, Thomas Ludwig, and Harald Meier. 2004. "RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees." *Bioinformatics* 21 (4):456-463.
- Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0." *Mol Biol Evol* 30 (12):2725-9. doi: 10.1093/molbev/mst197.
- Tello, Mario, Francisco Vergara, and Eugenio Spencer. 2013. "Genomic adaptation of the ISA virus to *Salmo salar* codon usage." *Virology journal* 10 (1):223.
- Thomson, E., C. L. Ip, A. Badhan, M. T. Christiansen, W. Adamson, M. A. Ansari, D. Bibby, J. Breuer, A. Brown, R. Bowden, J. Bryant, D. Bonsall, A. Da Silva Filipe, C. Hinds, E. Hudson, P. Klenerman, K. Lythgow, J. L. Mbisa, J. McLauchlan, R. Myers, P. Piazza, S. Roy, A. Trebes, V. B. Sreenu, J. Witteveldt, Stop-Hcv Consortium, E. Barnes, and P. Simmonds. 2016. "Comparison of Next-Generation Sequencing Technologies for Comprehensive Assessment of Full-Length Hepatitis C Viral Genomes." *J Clin Microbiol* 54 (10):2470-84. doi: 10.1128/JCM.00330-16.
- Tinoco Jr, Ignacio. 2004. "Force as a useful variable in reactions: unfolding RNA." *Annu. Rev. Biophys. Biomol. Struct.* 33:363-385.
- ToReNzo, AE. 1983. "Stability of infectious pancreatic necrosis virus (IPNV) in untreated, filtered and autoclaved estuarine water." *Bull. Eur. Ass. Fish pathot.* 4:5t\_5j.
- Tyagi, A., B. T. N. Kumar, and N. K. Singh. 2017. "Genome dynamics and evolution of codon usage patterns in shrimp viruses." *Arch Virol.* doi: 10.1007/s00705-017-3445-7.
- Underwood, BO, CJ Smale, F Brown, and BJ Hill. 1977. "Relationship of a virus from *Tellina tenuis* to infectious pancreatic necrosis virus." *Journal of General Virology* 36 (1):93-109.
- Villanueva, Rodrigo A, José L Galaz, Juan A Valdés, Matilde M Jashés, and Ana María Sandino. 2004. "Genome assembly and particle maturation of the birnavirus infectious pancreatic necrosis virus." *Journal of virology* 78 (24):13829-13838.
- Wallace, I. S., P. McKay, and A. G. Murray. 2017. "A historical review of the key bacterial and viral pathogens of Scottish wild fish." *J Fish Dis* 40 (12):1741-1756. doi: 10.1111/jfd.12654.



- Wang, J., N. E. Moore, Y. M. Deng, D. A. Eccles, and R. J. Hall. 2015. "MinION nanopore sequencing of an influenza genome." *Front Microbiol* 6:766. doi: 10.3389/fmicb.2015.00766.
- Weber, Siegfried, Dieter Fichtner, Thomas C Mettenleiter, and Egbert Mundt. 2001. "Expression of VP5 of infectious pancreatic necrosis virus strain VR299 is initiated at the second in-frame start codon." *Journal of General Virology* 82 (4):805-812.
- Wei, Yongwei, Jianrong Li, Jiangtao Zheng, Hong Xu, Long Li, and Lian Yu. 2006. "Genetic reassortment of infectious bursal disease virus in nature." *Biochemical and Biophysical Research Communications* 350 (2):277-287. doi: <https://doi.org/10.1016/j.bbrc.2006.09.040>.
- Wergeland, H. I., and R. A. Jakobsen. 2001. "A salmonid cell line (TO) for production of infectious salmon anaemia virus (ISAV)." *Dis Aquat Organ* 44 (3):183-90. doi: 10.3354/dao044183.
- Wetten, Marte, Torunn Aasmundstad, Sissel Kjølglum, and Arne Storset. 2007. "Genetic analysis of resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar* L.)." *Aquaculture* 272 (1):111-117. doi: <https://doi.org/10.1016/j.aquaculture.2007.08.046>.
- Wetterstrand, Kris A. 2013. DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP).
- Wick, Ryan R, Louise M Judd, Claire L Gorrie, and Kathryn E Holt. 2017. "Completing bacterial genome assemblies with multiplex MinION sequencing." *bioRxiv*:160614.
- Wolf, K. 1988. *Fish Viruses and Fish Virus Diseases*. Ithaca, NY: Cornell University Press. Book.
- Wolf, K., U.S. Fish, and Wildlife Service. Division of Fishery Research. 1966. *Infectious Pancreatic Necrosis (IPN) of Salmonid Fishes*: U.S. Department of the Interior, Fish and Wildlife Service, Bureau of Sport Fisheries and Wildlife.
- Wolf, Ken, C. E. Dunbar, and E. A. Pyle. 1961. "Infectious Pancreatic Necrosis of Trout: II. Experimental Infections with Brook Trout." *The Progressive Fish-Culturist* 23 (2):61-65. doi: 10.1577/1548-8659(1961)23[61:IPNOT]2.0.CO;2.
- Wolf, Ken, and MC Quimby. 1971. "Salmonid viruses: infectious pancreatic necrosis virus." *Archiv für die gesamte Virusforschung* 34 (2):144-156.
- Wood, E. M., S. F. Snieszko, and W. T. Yasutake. 1955. "Infectious pancreatic necrosis in brook trout." *AMA Arch Pathol* 60 (1):26-8.
- Xu, Hong-Tao, Wei-Duo Si, and Peter Dobos. 2004. "Mapping the site of guanylation on VP1, the protein primer for infectious pancreatic necrosis virus RNA synthesis." *Virology* 322 (1):199-210.
- Yao, Kun, and Vikram N Vakharia. 1998. "Generation of infectious pancreatic necrosis virus from cloned cDNA." *Journal of virology* 72 (11):8913-8920.
- Zhang, C. X., and S. Suzuki. 2003. "Comparison of the RNA polymerase genes of marine birnavirus strains and other birnaviruses." *Arch Virol* 148 (4):745-58. doi: 10.1007/s00705-002-0951-y.
- Zhang, CX, and S Suzuki. 2004. "Aquabirnaviruses isolated from marine organisms form a distinct genogroup from other aquabirnaviruses." *Journal of Fish Diseases* 27 (11):633-643.
- Zhao, Zhe, Fei Ke, Zhengqiu Li, Jianfang Gui, and Qiya Zhang. 2008. "Isolation, characterization and genome sequence of a birnavirus strain from flounder *Paralichthys olivaceus* in China." *Archives of virology* 153 (6):1143-1148.