SCIENTIFIC REPORTS
nature research

OPEN

A novel nudivirus infecting the invasive demon shrimp *Dikerogammarus haemobaphes* (Amphipoda)

Thomas W. Allain[1], Grant D. Stentiford[2,3], David Bass[2,3], Donald C. Behringer[1,4,5] & Jamie Bojko[6,7]✉

The *Nudiviridae* are a family of large double-stranded DNA viruses that infects the cells of the gut in invertebrates, including insects and crustaceans. The phylogenetic range of the family has recently been enhanced via the description of viruses infecting penaeid shrimp, crangonid shrimp, homarid lobsters and portunid crabs. Here we extend this by presenting the genome of another nudivirus infecting the amphipod *Dikerogammarus haemobaphes*. The virus, which infects cells of the host hepatopancreas, has a circular genome of 119,754 bp in length, and encodes a predicted 106 open reading frames. This novel virus encodes all the conserved nudiviral genes (sharing 57 gene homologues with other crustacean-infecting nudiviruses) but appears to lack the p6.9 gene. Phylogenetic analysis revealed that this virus branches before the other crustacean-infecting nudiviruses and shares low levels of gene/protein similarity to the *Gammanudivirus* genus. Comparison of gene synteny from known crustacean-infecting nudiviruses reveals conservation between *Homarus gammarus nudivirus* and *Penaeus monodon nudivirus*; however, three genomic rearrangements in this novel amphipod virus appear to break the gene synteny between this and the ones infecting lobsters and penaeid shrimp. We explore the evolutionary history and systematics of this novel virus, suggesting that it be included in the novel *Epsilonnudivirus* genus (*Nudiviridae*).

The family *Nudiviridae* comprises a group of non-occluded, double-stranded DNA (dsDNA) viruses infecting arthropods. The family includes two recognised genera, *Alphanudivirus* and *Betanudivirus*, with two other genera, *Gammanudivirus* and *Deltanudivirus*, recently proposed to contain aquatic-host-infecting viruses [1,2]. *Alphanudivirus* contains two species; *Gryllus bimaculatus nudivirus* and *Oryctes rhinoceros nudivirus*, while *Betanudivirus* contains a single species, *Heliothis zea nudivirus* [3]. In addition, several other insect-infecting nudiviruses have been morphologically and/or genomically characterized but remain to be formally recognised [4]. In recent years, closely related viruses have been identified infecting marine crustaceans, using genomics and ultrastructural data. The most complete descriptions include *Penaeus monodon nudivirus* (PmNV) infecting the farmed penaeid shrimp *Penaeus monodon*[1] and recently, *Homarus gammarus nudivirus* (HgNV) infecting juveniles of the European lobster, *Homarus gammarus*[2].

Other crustaceans are also infected with putative nudiviruses, often referred to in the published literature as "bacilliform viruses" in lieu of available genomic data[5–11]. In addition to descriptions of putative nudiviruses in these decapod crustaceans, observations of nudivirus-like infections of Amphipoda have also been reported; e.g. for *Dikerogammarus villosus* [12], *Dikerogammarus haemobaphes* [11], *Gammarus roeselii* [8], *Pontogammarus robustoides* [7] and *Gammarus varsoviensis* [7]. In all cases, putative nudivirus infection is observed within cells of the hepatopancreas, causing nuclear hypertrophy but no observable host immune response to infection, in

[1]School of Forest Resource and Conservation, University of Florida, Gainesville, FL 32611, USA. [2]International Centre of Excellence for Aquatic Animal Health, Centre for Environment, Fisheries and Aquatic Science (Cefas), Weymouth, Dorset DT4 8UB, UK. [3]Centre for Sustainable Aquaculture Futures, Biosciences, University of Exeter, Stocker Road, Exeter EX4 4PY, UK. [4]Fisheries and Aquatic Sciences, University of Florida, Gainesville, FL 32653, USA. [5]Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611, USA. [6]School of Health and Life Science, Teesside University, Middlesbrough TS1 3BA, UK. [7]National Horizons Centre of Excellence in Bioscience Industry, Teesside University, Darlington DL1 1HG, UK. ✉email: J.Bojko@tees.ac.uk

histological Section[8,11]. Histopathologically, nudiviral infection results in nuclear hypertrophy of the hepatopancreatocytes, caused primarily by a growing viroplasm; this is explored in full for the virus in *D. haemobaphes* in Bojko et al. [11]. Also in *D. haemobaphes*, infection prevalence of up to 77.7% has been correlated with altered behaviour in infected animals [11]. The behavioural change was associated with increased activity, which positively correlated with viral burden, potentially indicating some benefit to viral transmission through increased movement [11]. Many of the amphipod hosts in which putative nudiviruses have been reported are non-native or invasive species present outside of their native ranges. Infection with this viral family may therefore have potential for transboundary transmission when their hosts are present in their invasive range [7,10,11].

Genomic data is currently lacking for all the putative nudiviruses infecting amphipods. In this study, we provide full genome characterisation of a nudivirus infecting the amphipod *D. haemobaphes* collected from outside of its native range. We use these data to provisionally name the virus as *Dikerogammarus haemobaphes nudivirus* (DhNV) and place the virus within a newly suggested genus *Epsilonnudivirus* of the family *Nudiviridae*.

Results

Genome structure of DhNV. The circular genome of DhNV is 119,754 bp and contains 106 hypothetical ORFs, with 56 on the positive strand and 50 on the negative strand (Fig. 1). Fifty-nine of the ORFs met our comparative e-value threshold of < 0.001 and were directly comparable to other members of the *Nudiviridae*. Up to 17% and 37% of the ORFs aligned most closely with genes from HgNV and PmNV, respectively. Two ORFs, DhNV_008 and DhNV_002, scored above 50% similarity to protein sequences on BLASTp. DhNV_008 was 52.33% similar to *pif-2* from HgNV (QBB28614) with *per os* infectivity as the only identified protein domain. DhNV-002 was 50% similar to a hypothetical protein from PmNV (YP_009051845) where the cytoplasmic, non-cytoplasmic, tmhelix, and transmembrane domains were identified. Among the 47 ORFs that provided no similarity to other protein sequences within the threshold, InterProScan assessment identified 20 ORFs with functional domains. A protein signature match to the inhibitor of apoptosis repeat superfamily in DhNV_059 may indicate the presence of a homolog of the *Iap* nudivirus gene; however, BLASTp annotations did not yield any similarity results to the *Iap* gene found in other nudiviruses. The remaining 19 ORFs contained proteins with Zinc finger domains (DhNV_045), Tmhelix (DhNV_044), signal peptide (DhNV_084), p-loop containing nucleoside triphosphate hydrolase (DhNV_047), non-cytoplasmic domain (DhNV_025, 028, and 057), disorder predictions (DhNV_011, 020, 024, 072, 077, and 078), predicted to be cytoplasmic (DhNV_019 and 065), or had 'coil' feature(s) (DhNV_022, 026, 042, and 086).

Seven genes involved with DNA processing were identified: *DNA polymerase, helicase,* two copies of *helicase 2, integrase, DNA ligase,* and *fen-1*. Five genes involved with RNA transcription were also identified: *p47, lef-4, lef-5, lef-8,* and *lef-9*. Eight genes known to be involved in *per os* infectivity were found: *vp91* (*pif-8*), *pif-1, pif-2, pif-3, odv-e28* (*pif-4*), *odv-e56* (*pif-5*), *ac68* (*pif-6*), and *p74* (*pif-0*), along with the *11 k* gene. Nine genes can be grouped into packaging, assembly, and release: two copies of *ac92* (*p33*), two copies of *odv-e66,* two copies of *vlf-1*, *38 k, ac81, and 31 K* (*vp39*). The recently identified *p6.9* gene in crustacean-infecting nudivirus genomes, a baculovirus core gene associated with encapsulation of the viral genome, could not be found in DhNV despite similarity searches and checks for hypothetical SRSR repeat regions common to the p6.9 protein. A putative DUTPase, cg30-1, p-loop NTPase, guanosine monophosphate kinase, esterase, and *p51* were also identified. Definitively, 20 out of the 21 core baculovirus genes conserved among nudivirus were identified in the DhNV genome (Table 1).

Multiple genes were specific to DhNV; however, some were similar to other protein groups from various taxa. One ORF, DhNV_076, revealed 31.98% to 35.97% similarity to proteins from seven different organisms. The LOC108666550-like protein from HgNV is 32.35% similar (85% coverage) to DhNV_076 (Table 1). Uncharacterized proteins from *Hyalella azteca* (QBB28676, sim. 35.97% cov. 82%), an amphipod native to North America; *Penaeus vannamei* (XP_027235023, sim. 34.07% cov. 83%), the Whiteleg shrimp; *Crassostrea gigas* (XP_011433877, sim. 31.98% cov. 85%), the Pacific Oyster; *Armadillidium vulgare* (RXG54766, sim. 34.15% cov. 83%), the common pill-bug; and *Tigriopus californicus* (TRY76277, similarity 33.26% coverage 88%), a North American coastal copepod, all met the e-value threshold, as did the actin-binding IPP-like protein from *Brachionus plicatilis* (RMZ96256 similarity 33.65% coverage 87%). Only the undescribed PANTHER protein family, PTHR38566, was identified as a domain in DhNV_076.

Gene synteny among the *Epsilonnudivirus*, *Gammanudivirus* and *Deltanudivirus* genera. A comparison of gene synteny between DhNV and three other nudiviruses determined that DhNV had a different gene synteny to members of the *Gammanudivirus* and *Deltanudivirus* genera (Fig. 2). Comparison between ToNV and DhNV revealed a high level of genomic rearrangement, where the 32 genes that showed genetic similarity (e < 0.001) with those ORFs on the DhNV genome were located across the respective genomes, showing little conserved synteny (Fig. 2a). Comparison between DhNV and PmNV/HgNV revealed higher levels of gene synteny (Fig. 2b,c). A comparison using all three viruses identified 12 major regions of genetic novelty in the DhNV genome (Fig. 2d). This included 47 hypothetical ORFs that were unique to DhNV and showed little genetic/protein relatedness to other nudiviruses within the e-value threshold of < 0.001, one of which (DhNV_070) showed highest similarity to a gene from *Pyricularia oryzae*, a fungal plant pathogen, and another (DhNV_029) with highest similarity to *Sucra jujuba nucleopolyhedrovirus* (Table 1).

Using the protein similarity data, we determined that there were 11 crustacean-infecting nudivirus genes (DhNV_006, 024, 032, 034, 050, 062, 067, 080, 103, 104 and 106) that show conservation across the crustacean-infecting nudiviruses (Table 1) (i.e. present in PmNV, HgNV and DhNV) but appear absent from other nudiviruses that do not infect crustaceans. Using these genes in addition to the conserved baculovirus core genes across DhNV, HgNV and PmNV [1,2], 35 genes were comparable in a "gene-block" fashion relative to their genomic loci

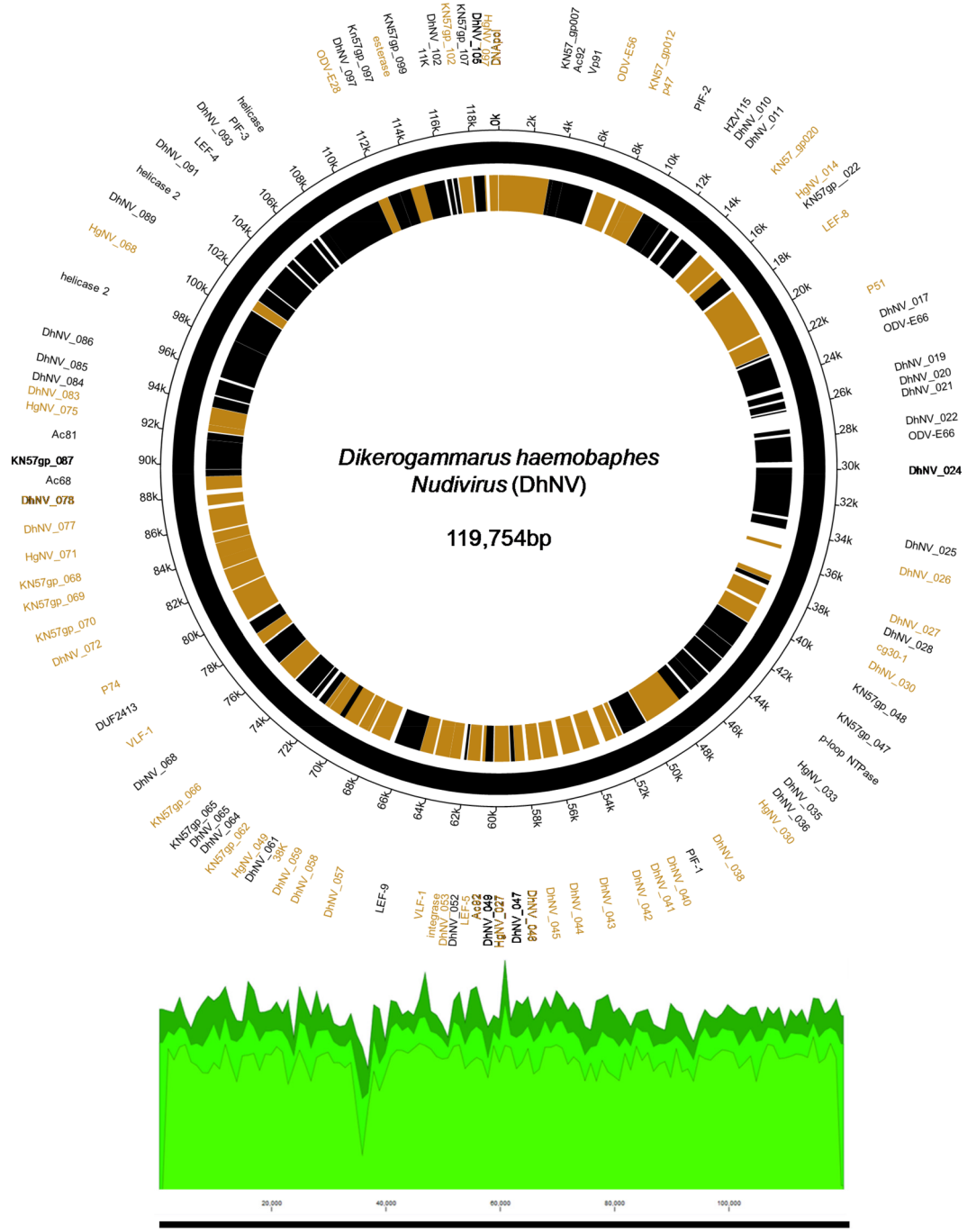

Figure 1. The '*Dikerogammarus haemobaphes nudivirus*' circular genome with open reading frames (ORF) graphically plotted across the 119,754 bp sequence. Locus tags for each similar virus are denoted in the place of DhNV genes, where possible. Gold regions and text indicate predicted genes on the positive strand and black regions and text indicate those on the negative strand. Below the genome, the green plot displays genome coverage across the genome using the MiSeq and HiSeq trimmed sequence reads. ORFs with significant similarity (e-value < 0.001) to other nudivirus genes are listed as the reference gene number on the corresponding viral genome. The circular plot was developed in Circa (www.omgenomics.com/circa/) and the coverage map from CLC genomics workbench v.12 (Qiagen).

(Fig. 2e). This revealed three major rearrangement events. Reordering of the DhNV_032, DhNV_034, and *pif-1* gene block, which is reversed in PmNV and HgNV (Fig. 2, 'X'). Reordering of the *vlf-1* and *p74* gene block, which is reversed in PmNV and HgNV (Fig. 2, 'Y') and a larger rearrangement of 6 genes (*vlf-1, ac68,* DhNV_080,

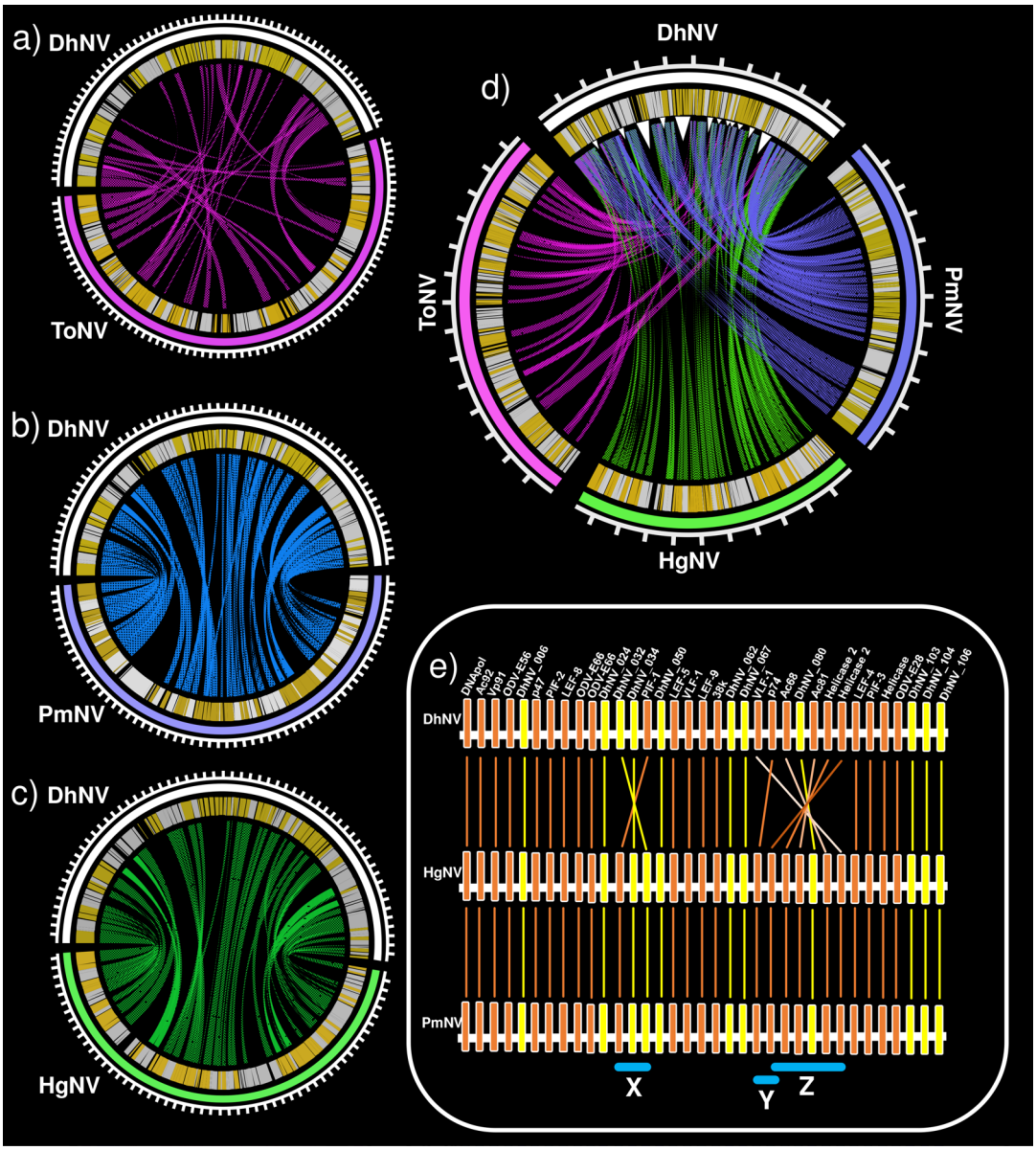

Figure 2. (**a**) Gene synteny among the *Epsilonnudivirus*, *Gammandivirus* and *Deltanuvirus* reveals a gene order among decapod-infecting nudiviruses, which is missing from peracarid-infecting nudiviruses. The genomes are annotated with positive strand (gold) and negative strand (silver) coding regions. In Fig. 2a, *Dikerogammarus haemobaphes nudivirus* gene synteny (white) is compared to Tipula oleracea nudivirus gene synteny (pink). Ribbons connecting the two genomes link up the homologous gene and its location on the viral genome. Scale ticks = 2 kb. The comparative plots were developed in Circa (www.omgenomics.com/circa/). (**b**) *Dikerogammarus haemobaphes nudivirus* gene synteny (white) is compared to *Penaeus monodon nudivirus* gene synteny (blue). Ribbons connecting the two genomes link up the homologous gene and its location on the viral genome. Scale ticks = 2 kb. The comparative plots were developed in Circa (www.omgenomics.com/circa/). (**c**) *Dikerogammarus haemobaphes nudivirus* gene synteny (white) is compared to *Hommarus gammarus nudivirus* gene synteny (green). Ribbons connecting the two genomes link up the homologous gene and its location on the viral genome. Scale ticks = 2 kb. The comparative plots were developed in Circa (www.omgenomics.com/circa/). (**d**) All four virus are compared together, indicating regions of novelty in the *Dikerogammarus haemoabphes nudivirus* genome. The white triangles on the DhNV genome highlight the areas of novel sequence information that do not correspond to genes on the other nudivirus genomes. Scale ticks = 10 kb. The comparative plots were developed in Circa (www.omgenomics.com/circa/). (**e**) The crustacean nuidviruses contain 24 nudivirus core genes (VLF-1, ODV-E66 and Helicase 2 are duplicated) (p6.9 is missing from *Dikerogammarus haemobaphes nudivirus*) (orange) and 11 other gene homologs conserved across the crustacean-infecting nudiviruses (yellow). These conserved homologues are based on similarity, synteny and functional identity. Comparison with DhNV (peracarid-infecting nudivirus) highlights three main areas of gene reorganization. 'X' corresponds to a rearrangement of the DhNV_032, DhNV_34 and *pif-1* genes, 'Y' corresponds to a rearrangement of the *vlf-1* and *p74* genes and finally, 'Z' corresponds to a rearrangement of six genes (*vlf-1*, *ac68*, DhNV_080, *ac81* and both *helicase 2* homologues). The *Gammanudivirus* members that infect decapods share the same gene synteny across these conserved motifs.

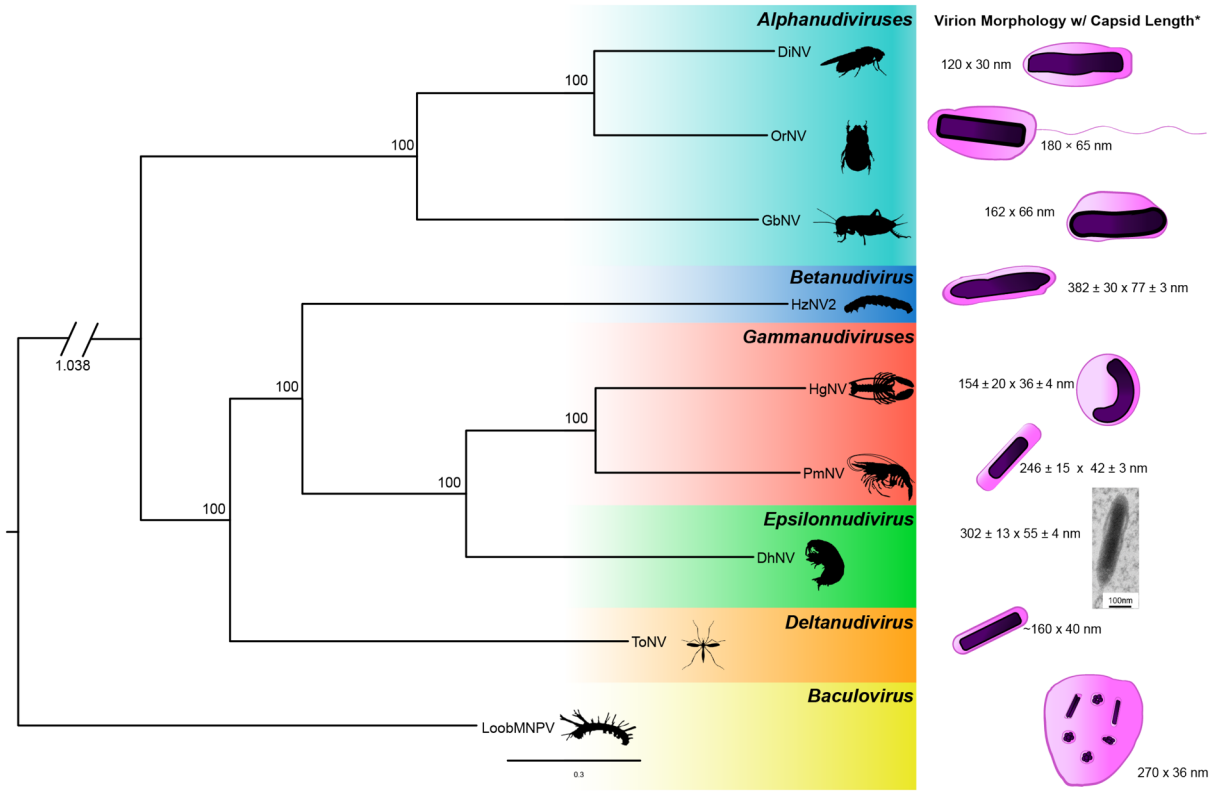

Figure 3. A concatenated phylogeny using 18 core nudivirus genes among 8 nudiviruses with LoobMNV as an outgroup. Node labels indicate bootstrap support in percent. Nudivirus genera are displayed in coloured groups. Illustrations of virion morphology are presented to the right of the tree and are based off electron micrographs of electron dense cores surrounded by a membrane. Capsid length approximations and averages are displayed from relevant publications cited in the methods. Genome accession numbers include: DiNV (NC_040699), OrNV (MN623374), GbNV (NC_009240), HzNV2 (NC_004156), HgNV (MK439999), PmNV (NC_024692), DhNV (MT488302), ToNV (NC_026242) and LoobMNPV (NC_043520). The tree was annotated in FigTree v.1.4.3. For additional information on DhNV virion morphology and pathology, please consult Bojko et al.[11].

ac81, and both copies of *helicase 2*), which is reversed in PmNV and HgNV and overlaps the 'Y' rearrangement event (Fig. 2, 'Z').

Morphological and phylogenetic comparison to other *Nudiviridae*. A concatenated maximum likelihood phylogenetic analysis of eight nudiviruses and one baculovirus (outgroup) using 18 core nudivirus genes (see Sect. 4) supported the positioning of DhNV outside of the two crustacean-infecting nudiviruses with bootstrap values of 100% (Fig. 3). Within this grouping, DhNV is an early branching member of the *Gammanudivirus* genus and may constitute a different genus altogether. The *Betanudivirus* genus branches outside of the *Gammanudivirus* cluster and the *Deltanudivirus* member ToNV is the earliest branching member of these three genera. The *Alphanudivirus* group represents the most phylogenetically distinct nudivirus genus represented on our diagram (Fig. 3).

Illustrations of virus morphology provide another dimension of comparison among the *Nudiviridae* species (Fig. 3). DhNV virions consist of a double membrane surrounding an electron-dense core measuring (n = 30, mean ± SD) 302 ± 13 nm in length and 55 ± 4 nm at its diameter [11]. The rod-shaped structure is maintained across all the nudiviruses. DhNV represents one of the larger nudiviruses discovered to date, second to HzNV2, which has a length of 382 ± 30 nm.

A second concatenated maximum likelihood phylogenetic analysis of putative *iap* and *pif-2* genes supported DhNV as an earlier branch of the crustacean-infecting nudiviruses. The addition of *Macrobrachium rosenbergii nudivirus* CN-SL2011 (MrNV) (NCBI:txid1217568), which only has the aforementioned genes available, branched in the *Gammanudivirus* genus. ToNV (*Deltanudivirus*) is the earliest branch of these genera, followed by HzNV2 (*Betanudivirus*), and the four crustacean-infecting nudiviruses. The *Alphanudiviruses* represent the most phylogenetically distinct lineage among nudiviruses in this tree (Fig. 4), following the same general phylogenetic theme as the details in Fig. 3.

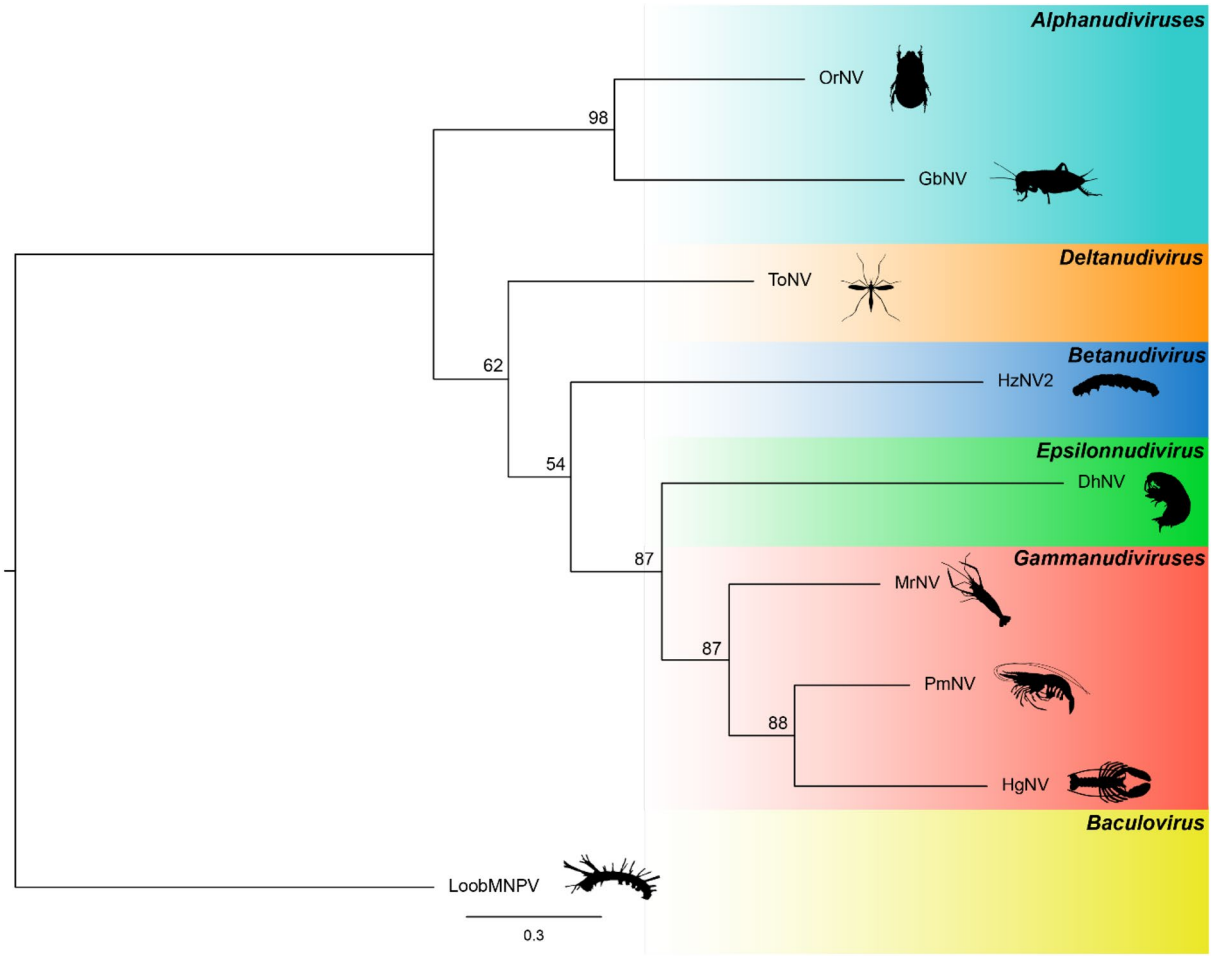

Figure 4. A concatenated phylogeny using *iap* and *pif-2* genes from 8 nudiviruses with LoobMNPV as an outgroup. Nodes are assigned bootstrap support values from 1,000 bootstrap replicates. Accession numbers include: OrNV (MN623374), GbNV (NC_009240), ToNV (NC_026242), HzNV2 (NC_004156), DhNV (MT488302), MrNV (JQ804994; JQ804993), PmNV (NC_024692), HgNV (MK439999) and LoobMNPV (NC_043520). The tree was annotated in FigTree v.1.4.3.

Discussion

We provide a full genome characterisation of DhNV, a novel member of the *Nudiviridae* infecting the freshwater amphipod host, *Dikerogammarus haemobaphes*. The genome size, ORFs and morphology of this virus correspond with related viruses from crustaceans and insects. The identification of this virus is discussed relative to its genetic and protein content, its gene synteny and the gene synteny of related viruses, and finally, its phylogenetic relatedness to other *Nudiviridae*. These combined data suggest a novel genus may be appropriate: *Epsilonnudivirus*.

A novel member of the *Nudiviridae* (*Epsilonnudivirus*) from an amphipod. Using a combination of the core genes conserved across the *Nudiviridae*, we show that DhNV is most related to the *Gammanudivirus* genus; however, with a low level of protein similarity at most loci (< 50%) it seems pertinent to explore the erection of a new genus. Our concatenated phylogenetic analysis of eight nudiviruses representing four genera is concordant with previously published trees [2]. DhNV appears to branch early from the three marine nudiviruses (Figs. 3, 4), suggesting an ancestral position to the HgNV, MrNV and PmNV.

The DhNV genome encoded all core nudivirus genes, apart from *p6.9*, a nucleotide-binding protein. These proteins function for DNA processing, RNA transcription, and *per os* infectivity [13]. The *p6.9* gene, which is responsible for the encapsulation of the viral genome, is characterized by a serine-arginine repeat region that could not be identified from the DhNV genome and was not present at the predicted locus where *p6.9* lies in other *Gammanudivirus* members: between *lef-5* (DhNV_051) and *vlf-1* (DhNV_055). In addition to the core genes (n = 24, including three repeat homologues), most genes show similarity to other *Nudiviridae* under a conservative e-value threshold (< 0.001), providing strong evidence that this virus belongs within the *Nudiviridae* (Table 1).

The primary source of protein similarity information for DhNV ORF's came from PmNV and HgNV, two genomically characterised viruses from the *Gammanudivirus* genus. *Gammanudivirus* members contain unique apoptosis inhibitor genes that lack a predicted RING domain [1] and appear twice in the HgNV genome. DhNV_059

ORF	Strand	Left End	Right End	Protein Length	Hit/Accession	Similarity (%)	Coverage (%)	e-value	BLAST annotation	Protein Domains	Functionality
DhNV_001	+	1	3270	1048	PmNV YP_009051843.1	38.74	98	0.00E+00	DNA Polymerase - DNA polymerase type-B family catalytic domain.	DNA-directed DNA polymerase, family B, multifunctional domain; YprB ribonuclease H-like domain	DNA processing
DhNV_002	-	3286	3528	80	PmNV YP_009051845.1	50	67	3.00E-11	hypothetical protein HgNV_003; similar to KN57_gp007	cytplasmic domain; non-cytoplasmic domain; Tmhelix; Transmembrane region	
DhNV_003	-	3528	4193	221	HgNV QBB28609.1	36.14	89	1.00E-26	Ac92-like protein	ERV/ALR sulfhydryl oxidase domain; cytoplasmic domain; non-cytoplasmic domain; transmembrane	Packaging, assembly, and release
DhNV_004	-	4196	6235	679	PmNV YP_009051847.1	29.31	99	2.00E-78	Vp91; chitin binding peritrophin-A domain	invertebrate chitin-binding families; baculovirus Vp91, capsid protein, N-terminal; chitin bidning domain; non-cytoplasmic	*Per os* infectivity
DhNV_005	+	6492	7799	435	HgNV QBB28611.1	37.7	96	3.00E-103	ODV-E56 (PIF-5); Baculo E56; chondroitin AC/alginate lyase	non-cytoplasmic domain; cytoplasmic domain; disorder prediction; polar; Tmhelix, transmembrane	*Per os* infectivity
DhNV_006	+	8029	8577	182	PmNV YP_009051850.1	22.03	87	3.00E-06	hypothetical protein, similar to KN57_gp012; similar to HgNV_007	coil; disorder prediction; negative polyelectrolyte	
DhNV_007	+	8583	9794	403	PmNV YP_009051852.1	29.51	97	9.00E-45	P47	no predictions	RNA transcription
DhNV_008	-	9841	11061	406	HgNV QBB28614.1	52.33	83	8.00E-122	Pif-2; per os infectivity factor 2	Per os infectivity factor 2	*Per os* infectivity
DhNV_009	-	11067	11762	231	PmNV YP_009051855.1	26.88	62	3.00E-09	putative DUTPase (partial); uridine kinase, similar to HZV115	no predictions	
DhNV_010	-	11944	12615	223	-	-	-	-	-	no predictions	
DhNV_011	-	12844	14154	436	-	-	-	-	-	disorder prediction; polar	
DhNV_012	+	14410	15732	440	PmNV YP_009051858.1	34.23	90	2.00E-59	hypothetical protein PmNV_20; DNA excision repair enzyme; KN57_gp020	PIN-like domain superfamily; 3.40.50.1010; PIN_FEN1	DNA processing
DhNV_013	+	15886	16473	195	HgNV QBB28619.1	27.07	65	8.00E-09	hypothetical protein HgNV_014; similar to PmNV_021	no predictions	
DhNV_014	-	16477	17412	311	PmNV YP_009051860.1	27.49	84	2.00E-26	hypothetical protein KN57_gp022	no predictions	Packaging, assembly, and release
DhNV_015	+	17597	20908	1103	HgNV QBB28621.1	38.64	93	0.00E+00	Late expression factor 8; bifunctional DNA-directed RNA polymerase subunit beta	DNA-directed RNA polymerase, subunit 2, hybrid-binding domain superfamily; RNA polymerase, beta subunit, conserved site; beta and beta-prime subunits of DNA dependent RNA-polymerase	RNA transcription
DhNV_016	+	21049	22233	394	PmNV YP_009051862.1	25.47	75	2.00E-24	p51	no predictions	
DhNV_017	-	22278	22409	43	-	-	-	-	-	no predictions	
DhNV_018	-	22555	24585	676	PmNV YP_009051874.1	39.04	89	9.00E-134	occlusion-derived virus envelope protein E66	Chondroitin AC/alginate lyase; 1.50.10.100 CATH	Packaging, assembly, and release
DhNV_019	-	24890	25366	158	-	-	-	-	-	cytoplasmic domain; disorder prection; transmembrane; Tmhelix; coil; polar; polyampholyte; non-cytoplasmic domain;	
DhNV_020	-	25533	26012	159	-	-	-	-	-	disorder prediction; polyampholyte; coil	
DhNV_021	-	26124	26219	31	-	-	-	-	-	no predictions	
DhNV_022	-	27368	27673	101	-	-	-	-	-	coil; disorder prediction	
DhNV_023	-	27858	29510	550	PmNV YP_009051874.1	37.15	97	1.00E-120	occlusion-derived virus envelope protein E66	Chondroitin AC/alginate lyase; 1.50.10.100 CATH	Packaging, assembly, and release
DhNV_024	-	29883	33107	1074	HgNV QBB28623.1	24.26	27	2.00E-19	hypothetical protein HgNV_018; similar to KN57_gp057	disorder prediction, positive polyelectrolyte, coil	
DhNV_025	-	33291	33911	206	-	-	-	-	-	non-cytoplasmic domain; signal IP TM; coil; cytoplasmic domain; non-cytoplasmic domain; Tmhelix; transmembrane; signal peptide; signal peptide h-region; signal peptide c-region; signal peptide h-region; signal peptide n-region	

(continued)

DhNV_026	+	34981	35223	80	-	-	-	-	-	coil; disorder prediction; polyampholyte	
DhNV_027	+	37053	37406	117	-	-	-	-	-	no predictions	
DhNV_028	-	37433	37735	100	-	-	-	-	-	non-cytoplasmic domain; tmhelix; transmembrane; cytoplasmic domain	
DhNV_029	+	37970	39130	386	Sucra jujuba NPV YP_009186763.1	35.19	11	8.00E-04	cg30-1 RING-finger (Really Interesting New Gene) domain, a specialized type of Zn-finger of 40 to 60 residues that binds two atoms of zinc	zinc finger, RING-type; ring 2; ZF ring 2; ring/u-box; ring-hc	
DhNV_030	+	39327	40460	377	-	-	-	-	-	no predictions	
DhNV_031	-	40523	42310	595	PmNV YP_009051886.1	27.36	97	5.00E-53	hypothetical protein KN57_gp048, similar to HzNV-2 orf10	DNA ligase/mRNA capping enzyme, catalytic domain	DNA processing
DhNV_032	-	42336	43307	323	PmNV YP_009051885.1	20.72	93	1.00E-16	hypothetical protein KN57_gp047, HgNV_035	no predictions	
DhNV_033	-	43379	44578	399	PmNV YP_009051884.1	37.35	78	8.00E-58	p-loop NTPase	P-loop containing nucleoside triphosphate hydrolase	
DhNV_034	-	44702	45529	275	HgNV QBB28638.1	33.47	80	2.00E-31	PmV-like protein	disorder prediction; polar	
DhNV_035	-	45536	46240	234	-	-	-	-	-	no predictions	
DhNV_036	-	46412	47215	267	-	-	-	-	-	no predictions	
DhNV_037	+	47188	49464	758	HgNV QBB28635.1	27.8	82	9.00E-55	hypothetical protein HgNV_030, similar to KN57_gp042	disorder prediction; proline-rich; coil	
DhNV_038	+	49439	49765	108	-	-	-	-	-	no predictions	
DhNV_039	-	49852	51474	540	PmNV YP_009051877.1	38.5	98	5.00E-131	PIF-1, Per-os infectivity factor 1; PIF; TM	cytoplasmic domain; transmembrane; Tmhelix; non-cytoplasmic domain	*Per os* infectivity
DhNV_040	+	51578	51877	99	-	-	-	-	-	no predictions	
DhNV_041	+	52001	52738	245	-	-	-	-	-	no predictions	
DhNV_042	+	53166	54284	372	-	-	-	-	-	coil; disorder prediction	
DhNV_043	+	54521	55633	370	-	-	-	-	-	no predictions	
DhNV_044	+	55966	56925	319	-	-	-	-	-	Tmhelix	
DhNV_045	+	57077	57916	279	-	-	-	-	-	Zinc finger, RING-type; ZF ring 2; ZF ring 1 ring/u-box	
DhNV_046	+	58153	58782	209	-	-	-	-	-	no predictions	
DhNV_047	-	58798	59061	87	-	-	-	-	-	P-loop containing nucleoside triphosphate hydrolase	
DhNV_048	+	59144	60145	333	HgNV QBB28632.1	43.84	60	2.00E-49	guanosine monophosphate kinase	no predictions	
DhNV_049	-	60222	60737	171	-	-	-	-	-	no predictions	
DhNV_050	+	60727	60936	69	HgNV QBB28643.1	48.08	75	1.00E-09	hypothetical protein HgNV_038, similar to KN57_gp051	Cytoplasmic domain; Tmhelix; transmembrane, non-cytoplasmic domain	Packaging, assembly, and release
DhNV_051	+	61069	61893	274	PmNV YP_009051890.1	30.39	37	8.00E-06	late expression factor 5; Baculo_LEF-5_C	disorder prediction; negative polyelectrolyte; coil; polyampholyte	RNA transcription
DhNV_052	-	61952	62119	55	-	-	-	-	-	no predictions	
DhNV_053	+	62358	63065	235	-	-	-	-	-	no predictions	
DhNV_054	+	63071	64027	318	PmNV YP_009051893.1	42.09	92	1.00E-79	integrase	DNA breaking-rejoining enzyme, catalytic core; Integrase-like, catalytic domain superfamily; integrase, catalyric domain	DNA processing
DhNV_055	+	64170	65024	284	PmNV YP_009051894.1	35	96	1.00E-36	very late expression factor 1	Integrase-like, catalytic domain superfamily; DNA breaking-rejoining enzyme, catalytic core; integrase, catalytic domain	Packaging, assembly, and release
DhNV_056	-	65044	66786	580	PmNV YP_009051896.1	43.74	85	6.00E-147	LEF-9, DNA-directed RNA polymerase	beta and beta-prime subunits of DNA dependent RNA-polymerase	RNA transcription
DhNV_057	+	67123	68409	428	-	-	-	-	-	non-cytoplasmic domain; cytoplasmic domain; transmembrane	
DhNV_058	+	68549	68701	50	-	-	-	-	-	no predictions	
DhNV_059	+	68708	69325	205	-	-	-	-	-	inhibitor of apoptosis repeat	

(continued)

DhNV_060	+	69426	70427	333	PmNV YP_009051897.1	30.74	84	4.00E-32	38K protein	HAD-like superfamily, protein of unknown function DUF694	Packaging, assembly, and release
DhNV_061	-	70410	70763	117	-	-	-	-	-	no predictions	
DhNV_062	+	70765	71562	265	HgNV QBB28654.1	30.8	92	1.00E-36	hypothetical protein HgNV_049, similar to KN57_gp061	no predictions	
DhNV_063	+	71583	71930	115	PmNV YP_009051900.1	39.34	53	2.00E-07	hypothetical protein KN57_gp062	Cytoplasmic domain; Tmhelix; transmembrane, non-cytoplasmic domain	
DhNV_064	-	71956	72429	157	-	-	-	-	-	no predictions	
DhNV_065	-	72644	72991	115	-	-	-	-	-	cytoplasmic domain; disorder prediction; transmembrane; Tmhelix; negative polyelectrolyte; non-cytoplasmic domain	
DhNV_066	-	73071	74423	450	PmNV YP_009051903.1	32.04	76	3.00E-50	hypothetical protein KN57_gp065	coil	
DhNV_067	+	74531	75928	465	PmNV YP_009051904.1	28.4	84	4.00E-36	hypothetical protein KN57_gp066; HgNV_054	coil; disorder prediction	
DhNV_068	-	75972	77471	499	-	-	-	-	-	no predictions	
DhNV_069	+	77585	78340	251	PmNV YP_009051928.1	36.89	96	4.00E-36	very late expression factor 1	no predictions	Packaging, assembly, and release
DhNV_070	-	78362	79234	290	Pyricularia oryzae 70-15 XP_003712544.1	41.18	44	3.00E-04	hypothetical protein MGG_16847	proline rich extensin signature; non-cytoplasmic domain; disorder prediction; proline-rich; Tmhelix; signal peptide; signal peptide h-region; polar; signal peptide c-region; signal peptide N-region	
DhNV_071	+	79418	81550	710	HgNV QBB28668.1	46.9	99	0.00E+00	P74	Baculoviridae p74; Baculoviridae p 74 N-terminal; non-cytoplasmic domain; cytoplasmic domain; Tmhelix; Transmembrane	*Per os* infectivity
DhNV_072	+	81646	83016	456	-	-	-	-	-	disorder prediction; polyampholyte; coil	
DhNV_073	+	83086	83835	249	PmNV YP_009051908.1	25.68	87	7.00E-11	hypothetical protein KN57_gp070	no predictions	
DhNV_074	+	83823	84806	327	PmNV YP_009051907.1	29.39	79	7.00E-26	hypothetical protein KN57_gp069	no predictions	
DhNV_075	+	84839	85549	236	PmNV YP_009051906.1	27.23	77	3.00E-12	hypothetical protein KN57_gp068	disorder prediction; polyampholyte; polar	
DhNV_076	+	85655	87085	476	HgNV QBB28676.1	32.35	85	7.00E-63	LOC108666550-like protein	family not named [PTHR38566]	
DhNV_077	+	87333	88022	229	-	-	-	-	-	disorder prediction; polar	
DhNV_078	+	88348	89223	291	-	-	-	-	-	disorder prediction; polar	
DhNV_079	-	89242	89673	143	HgNV QBB28683.1	40.17	81	4.00E-26	Ac68-like protein	non-cytoplasmic domain; tmhelix; transmembrane; cytoplasmic domain	*Per os* infectivity
DhNV_080	-	89702	91621	639	PmNV YP_009051925.1	24.91	87	1.00E-35	hypothetical protein KN57_gp087; HgNV_077	no predictions	
DhNV_081	-	91632	92108	158	PmNV YP_009051924.1	48.32	94	4.00E-45	Ac81-like protein	Protein AC81, baculovirus; cytoplasmic domain; Tmhelix; transmembrane; non-cytoplasmic domain	Packaging, assembly, and release
DhNV_082	+	92187	92621	144	HgNV QBB28680.1	26.9	100	2.00E-07	hypothetical protein HgNV_075	no predictions	
DhNV_083	+	92627	93763	378	-	-	-	-	-	no predictions	
DhNV_084	-	93788	94495	235	-	-	-	-	-	SignalP-TM	
DhNV_085	-	94601	95626	341	-	-	-	-	-	no predictions	
DhNV_086	-	95803	98343	846	-	-	-	-	-	coil	
DhNV_087	-	98355	100577	740	PmNV YP_009051917.1	34.25	68	6.00E-85	helicase 2	P-loop containing nucleoside triphosphate hydrolase	DNA processing
DhNV_088	+	100650	101324	224	HgNV QBB28673.1	34.63	100	6.00E-24	hypothetical protein HgNV_068	Ribonuclease H-like superfamily; Ribonuclease H superfamily; coil	
DhNV_089	-	101369	102316	315	-	-	-	-	-	no predictions	
DhNV_090	-	102399	104207	602	PmNV YP_009051914.1	39.4	97	2.00E-127	helicase 2	P-loop containing nucleoside triphosphate hydrolase	DNA processing
DhNV_091	-	104337	104852	171	-	-	-	-	-	no predictions	
DhNV_092	-	104990	106453	487	HgNV QBB28686.1	33.94	89	4.00E-75	Late expression factor 4 (LEF-4)	no predictions	RNA transcription
DhNV_093	-	106608	106967	119	-	-	-	-	-	no predictions	
DhNV_094	-	107160	107741	193	HgNV QBB28688.1	45.08	97	2.00E-55	Per os infectivity factor 3	Per os infectivity factor 3; non cytoplasmic domain; Tmhelix; transmembrane; cytoplasmic domain	*Per os* infectivity

(continued)

DhNV_095	-	107731	111603	1290	PmNV YP_009051932.1	31.31	93	4.00E-179	helicase	cytoplasmic domain; non-cytoplasmic domain; transmembrane; coil	DNA processing
DhNV_096	+	111616	112278	220	PmNV YP_009051934.1	47.98	78	3.00E-50	ODV-E28 (PIF-4)	Per os infectivity factor 4; cytoplasmic domain; Tmhelix; transmembrane; non-cytoplasmic domain	*Per os* infectivity
DhNV_097	-	112294	113100	268	-	-	-	-	-	no predictions	
DhNV_098	-	113105	113848	247	PmNV YP_009051935.1	29.38	78	3.00E-18	hypothetical protein PmNV_097	no predictions	
DhNV_099	+	113847	114779	310	PmNV QBB28692.1	44.61	86	6.00E-60	esterase, acetyl esterase/lipase	alpha/beta hydrolase fold; cytoplasmic domain; non-cytoplasmic domain; Tmhelix; transmembrane	
DhNV_100	-	114785	116131	448	PmNV YP_009051937.1	29.75	31	2.00E-17	hypothetical protein PmNV_099; GrBNV_gp67-like; EF_HAND_1; coiled-coil domain	disorder prediction	
DhNV_101	-	116291	116596	101	PmNV YP_009051938.1	46.81	93	2.00E-29	11K virion structural protein; TONB_DEPENDENT_REC_1; TM	cytoplasmic domain; non-cytoplasmic domain; Tmhelix; transmembrane	Structural
DhNV_102	-	116706	116969	87	-	-	-	-	-	no predictions	
DhNV_103	+	117073	117909	278	PmNV YP_009051940.1	31.11	93	1.00E-29	hypothetical protein KN57_gp102; HgNV_091	no predictions	
DhNV_104	-	118038	118757	239	PmNV YP_009051945.1	34.26	88	2.00E-36	hypothetical protein KN57_gp107; HgNV_096	coil	
DhNV_105	+	118759	118854	31	-	-	-	-	-	no predictions	
DhNV_106	+	119066	119647	193	HgNV QBB28702.1	31.93	81	7.00E-14	hypothetical protein HgNV_097; KN57_gp108	no predictions	

Table 1. Table displaying identified ORFs of DhNV. Positive/negative strand, left end, right end, and protein length are listed. Hit/Accession come from the top BLASTp result with an e-value < 0.001. Similarity, coverage, and exact e-values are displayed from said results. BLAST repository notes from each similar protein are shown along with identified protein domains and commonly identified nudivirus functional groups. Core genes (n = 24) are presented with a darker orange background and genes with similarity and function only detected in the crustacean viruses (n = 11) are noted in yellow

represents a homolog of the *Iap* nudivirus gene in DhNV, where an inhibitor of apoptosis repeat domain was detected but is relatively different from existing Baculovirus homologues [1]. In addition to family-level gene conservation, we identified 11 "crustacean-infecting nudivirus" genes that are conserved among those that infect crustaceans. Using a gene-block approach, we identified that PmNV and HgNV share gene synteny, where the DhNV genome exhibits three reorganization events, termed 'X', 'Y' and 'Z' (Fig. 3). These rearrangements are visible only in this virus, alongside a low average protein similarity of ~ 50%, and may indicate that a fifth nudivirus genus could be erected to hold peracarid-infecting nudiviruses. We suggest *Epsilonnudivirus*. In further work, greater genomic availability of viruses from peracarid hosts could help to better define these demarcation criteria.

Further genomic characterisation of peracarid-infecting nudiviruses may also help to identify the evolutionary history of DhNV, especially with regards to genes that show relatedness outside the *Nudiviridae*. Examples include DhNV_029, which shares 35.19% similarity to the *cg30-1* gene (YP_009186763) from *Sucra jujuba nucleopolyhedrovirus* (Table 1), a butterfly-infecting baculovirus. This is the first of two ORFs with zinc finger, RING-type domains in the DhNV genome; with DhNV_045 being the second (Table 1). These do show some relation to HgNV and PmNV, where both HgNV and PmNV contain three proteins with Zinc finger, RING-type domains: HgNV_019, 064, and 067 and KN57gp_003, 033, and 049 respectively [1,2]. DhNV_070 also shows high similarity to a non-nudivirus organism. A hypothetical protein from *Pyricularia oryzae* (Table 1), the fungal pathogen that causes rice blast disease, shows 41.18% similarity to DhNV_070. Protein domain analysis using InterProScan revealed mainly cytoplasmic and disorder protein domains from the *P. oryzae* sequence (XP_003712544) while DhNV_070 yielded a detailed signature match to proline rich extensin, commonly found in plant cell walls. This extensin domain does not appear in any *Gammanudivirus* protein. Finally, DhNV_076 shows some similarity to a homologue of HgNV (LOC108666550-like protein); however, both also show high levels of similarity to ORFs of invertebrate taxa, which lacks an identified domain or function. Such a conserved gene encoded by these viruses may reflect an ancient horizontal gene acquisition from a host during their evolutionary history.

New perspectives surrounding the *Nudiviridae*. Nudivirus infections often delay development of their arthropod hosts, eventually causing death [4]. However, high prevalence of nudiviruses in hosts apparently displaying few clinical signs of infection may also suggest some host benefit of retaining such sub-clinical infections [2,5,10,11]. While the exact relationship between DhNV and host survival still requires testing, a significant association with increased activity may subsequently increase invasive capabilities of the host [11]. Examining the genome of DhNV revealed several conserved and convergent traits of crustacean nudiviruses, highlighting potential genes for diagnostic development and further research into functional roles during host infection and survival within the environment. Further sequencing and characterisation of many hypothetical proteins will provide more insight into the evolutionary history and host relationship of DhNV relative to other *Nudiviridae*. Through genomic analysis, phylogeny, and virion morphology it is evident that the *Nudiviridae* in Crustacea are highly derived from their insect relatives and a great diversity of currently undescribed taxa likely reside in other arthropod hosts on land and in water.

Materials and methods

Collection of infected *Dikerogammarus haemobaphes* and next generation sequencing. *Dikerogammarus haemobaphes* were collected, dissected and underwent DNA extraction as explained by Bojko et al. [11], who also explore virion morphology and pathology associated with the discovery of a novel nudivirus. Stored DNA from a single individual displaying the characteristic pathology of bacilliform virus infection was selected for next generation sequencing using Illumina MiSeq and Illumina HiSeq. The DNA extract underwent library preparation for Illumina MiSeq using the NEXTERA XT library preparation kit, according to manufacturer's protocol (Illumina, UK). The library underwent quality screening using a bioanalyzer (Agilent), was quantified using a QuantiFluor fluorimeter (Promega), was denatured using sodium hydroxide and diluted to 10 pM in Illumina HT1 hybridisation buffer for sequencing via an Illumina V3-600 cartridge. The same DNA extract was used to produce a library for Illumina HiSeq using the Illumina TruSeq DNA PCR-Free library preparation kit, according to manufacturer's protocols. The library underwent quality screening using a bioanalyzer (Agilent), was quantified using a QuantiFluor fluorimeter (Promega), was denatured using sodium hydroxide and diluted to 10 pM in Illumina HT1 hybridisation buffer for sequencing on an Illumina HiSeq 3,000 with a 2×150 cartridge.

MiSeq and HiSeq outputs were trimmed in silico using Illuminaclip v0.32 (Trimmomatic: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) [14] and pooled into paired and unpaired sequence files. The paired sequence data from each technique were paired-end-combined using PEAR v0.9.8 (settings: overlap similarity minimum, 20 bp) [15] to increase the read length of paired reads by combining the reads into single sequence reads. These reads were assembled using SPAdes v3.13.0 [16] with default parameters and k-mer lengths 21, 33, 55, 77, 99 and 127, to produce 228,433 scaffolds with a maximum read length of 119,824 bp and minimum read length of 128 bp.

Identification and annotation of the viral genome. Scaffolds above 100,000 bp were extracted from the dataset and annotated using PROKKA v1.11 [17]. The subsequent output was assessed for similarity to existing sequence data using NCBI, Blastp nr database. This identified a raw contiguous sequence of 119,824 bp as the genome of DhNV, which was subsequently circularized and checked for average coverage using CLC Genomics workbench v11 to result in a genome of 119,754 bp (coverage: 157.93X). PROKKA v1.11 [17] and GeneMarkS [18] was used to annotate the viral genome (parameters: virus). A combination of these two tools resulted in 95 identical open reading frames (ORFs), 8 frames with high similarity but different gene size and three ORFs identified just by PROKKA. Combined, this provided 106 ORFs for annotation. The protein product of the 106 ORFs were compared to existing information using BLASTp via the NCBI repository (GenBank) with a cut-off e-value of < 0.001. The protein sequences were also assessed using the InterProScan tool (ebi.ac.uk/interpro/) to identify domains and predicted function. Twenty-one conserved core baculovirus/nudivirus genes were identified; however, P6.9 was not found within the genome of DhNV after analysis using BLASTp, ExPASy [19], GeneMarkS and InterProScan.

The gene synteny data for DhNV was compared to two crustacean-infecting viruses, *Homarus gammarus nudivirus* (HgNV) and *Penaeus monodon nudivirus* (PmNV) (*Gammanudiviridae*), and an insect-infecting virus, *Tipula oleracea nudivirus* (ToNV) (*Deltanudiviridae*), whose data were obtained from NCBI accessions: KJ184318, MK439999, NC_026242, respectively. The data were plotted using Circa (omgenomics.com/circa).

The annotated viral genome is available through NCBI accession: MT488302.

Phylogenetic analysis of DhNV among the *Nudiviridae*. A concatenated phylogenetic tree was developed using 18 of the 21 identified nudivirus core proteins from DhNV and seven other nudiviruses: *Drosophila innubila nudivirus* (DiNV), *Oryctes rhinocerous nudivirus* (OrNV), *Gryllus bimaculatus nudivirus* (GbNV), *Helicoverpa* (syn. *Heliothis*) *zea nudivirus*-2 (HzNV2), HgNV, PmNV, and ToNV. A baculovirus outgroup, *Lonomia obliqua multiple nucleopolyhedrovirus* (LoobMNPV) was used to root the tree. The p47 (missing from GbNV), Helicase 2, vlf-1, and p6.9 were not included as they are not present in the genomes of all the tested nudiviruses or are duplicated in DhNV. The remaining conserved proteins, *38 k*, *ac81, DNA polymerase, helicase, lef-4, lef-5, lef-8, lef-9, ac92 (p33), p74 (pif-0), pif-1, pif-2, pif-3, odv-e28* (*pif-4*), *odv-e56 (pif-5), pif-6, vp91 (pif-8)* and *31 K (vp39),* were aligned using Geneious v10 using MAFFT with default parameters. In HzNV2, Lef-9 was trimmed using Geneious due to its fusion with p47 [1]. IQtree was used to produce the maximum likelihood phylogenetic tree, which included 13,795 positions using the VT + F + I + G4 model (according to BIC) with 1,000 bootstrap replicates. Subsequently, the tree was imported into Figtree v1.4.3 for annotation. Transmission electron micrographs from each nudivirus were used to create illustrations of the virions with approximations of nucleocapsid size [2,11,20–25]. A second concatenated tree was produced using putative *iap* and *pif-2* genes from DhNV and seven other nudiviruses: OrNV, GbNV, HzNV2, HgNV, PmNV, ToNV, in addition to recently obtained *Macrobrachium rosenbergii nudivirus* CN-SL2011 (MrNV) (NCBI:txid1217568) sequences, which includes just two protein coding genes (*iap* and *pif-2*). DiNV was excluded from this tree as it lacks an identifiable *iap* ORF. The baculovirus, *Lonomia obliqua multiple nucleopolyhedrovirus* (LoobMNPV) was used to root the tree. Genes were trimmed using Geneious v10 and aligned using MAFFT with default parameters. IQtree produced a phylogenetic tree using the Blosum62 + G4 model (according to BIC) with 1,000 bootstrap replicates. The tree was imported into Figtree v1.4.4 for final annotation.

Data availability

Sequence data from this study are available through NCBI as stated herein. Biological materials from the host are available from the Cefas Aquatic Registry and Repository.

Received: 6 February 2020; Accepted: 4 August 2020
Published online: 09 September 2020

References

1. Yang, Y. T. *et al.* The genome and occlusion bodies of marine *Penaeus monodon nudivirus* (PmNV, also known as MBV and PemoNPV) suggest that it should be assigned to a new nudivirus genus that is distinct from the terrestrial nudiviruses. *BMC Genomics* **15**, 628 (2014).
2. Holt, C. C. *et al.* The first clawed lobster virus Homarus gammarus nudivirus (HgNV n. sp.) expands the diversity of the Nudiviridae. *Sci. Rep.* **9**, 10086 (2019).
3. Adams, M. J., Lefkowitz, E. J., King, A. M. Q. & Carstens, E. B. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2014). *Arch. Virol.* **159**(10), 2831–2841 (2014).
4. Bézier, A. *et al.* The genome of the nucleopolyhedrosis causing virus from *Tipula oleracea* (ToNV) sheds new light into the *Nudiviridae* family. *J. Virol.* **89**, 3008 (2014).
5. Stentiford, G. D., Bateman, K. & Feist, S. W. Pathology and ultrastructure of an intranuclear bacilliform virus (IBV) infecting brown shrimp *Crangon crangon* (Decapoda: Crangonidae). *Dis. Aquat. Organ.* **58**, 89–97 (2004).
6. Bateman, K. S. & Stentiford, G. D. Cancer pagurus bacilliform virus (CpBV) infecting juvenile European edible crabs C pagurus from UK waters. *Dis. Aquat. Organ.* **79**, 147–151 (2008).
7. Bojko, J. & Ovcharenko, M. Pathogens and other symbionts of the Amphipoda: taxonomic diversity and pathological significance. *Dis. Aquat. Organs.* **136**, 3–36 (2019).
8. Bojko, J. *et al.* Parasites, pathogens and commensals in the "low-impact" non-native amphipod host *Gammarus roeselii*. *Parasit. Vect.* **10**, 193 (2017).
9. Bateman, K. S. & Stentiford, G. D. A taxonomic review of viruses infecting crustaceans with an emphasis on wild hosts. *J. Invertebr. Pathol.* **147**, 86–110 (2017).
10. Bojko, J. *et al.* Green crab *Carcinus maenas* symbiont profiles along a North Atlantic invasion route. *Dis. Aquat. Organ.* **128**(2), 147–168 (2018).
11. Bojko, J. *et al.* Pathogens of *Dikerogammarus haemobaphes* regulate host activity and survival, but also threaten native amphipod populations in the UK. *Dis. Aquat. Organ.* **136**, 63–78 (2019).
12. Bojko, J. *et al.* Baseline histopathological survey of a recently invading island population of 'killer shrimp' Dikerogammarus villosus. *Dis. Aquat. Organ.* **106**, 241–253 (2019).
13. Beperet, I., Simón, O., Williams, T., López-Ferber, M. & Caballero, P. The '11K' gene family members sf68, sf95 and sf138 modulate transmissibility and insecticidal properties of Spodoptera frugiperda multiple nucleopolyhedrovirus. *J. Invertebr. Pathol.* **127**, 101–109 (2015).
14. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15), 2114–2120 (2014).
15. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2013).
16. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
17. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
18. Besemer, J., Lomsadze, A. & Borodovsky, M. GeneMark S (2001) a self-training method for prediction of gene starts in microbial genomes Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **29**, 2607–2618 (2001).
19. Gasteiger, E. *et al.* ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788 (2003).
20. Lightner, D. V., Redman, R. M. & Bell, T. A. Observations on the geographic-distribution, pathogenesis and morphology of the baculovirus from *Penaeus monodon* Fabricius. *Aquaculture* **32**(3–4), 209–233 (1983).
21. Huger, A. M. A new virus disease of crickets (Orthoptera: Gryllidae) causing macronucleosis of fatbody. *J. Invertebr. Pathol.* **45**, 108–111 (1985).
22. Raina, A. K. & Adams, J. R. Gonad-specific virus of corn earworm. *Nature* **374**, 770–770 (1995).
23. Wolff, J. L. C. *et al.* Identification and characterization of a baculovirus from *Lonomia obliqua* (Lepidoptera: Saturniidae). *J. Invertebr. Pathol.* **79**, 137–145 (2002).
24. Huger, A. M. The Oryctes virus: Its detection, identification, and implementation in biological control of the coconut palm rhinoceros beetle, *Oryctes rhinoceros* (Coleoptera: Scarabaeidae). *J. Invertebr. Pathol.* **89**, 78–84 (2005).
25. Unckless, R. L. A DNA Virus of *Drosophila*. *PLoS ONE* **6**, e26564 (2011).

Acknowledgements

The authors acknowledge NERC funding to J.B. (Award #: 1368300); GDS and DB were funded by Cefas contract #FB002, which supported the collection of NGS data along with personal funds from DCB (University of Florida).

Author contributions

J.B., G.D.S. and D.B. designed the study. J.B. collected the animals. J.B. and T.A. conducted the bioinformatic analyses, including phylogenetics and annotation. T.A., G.D.S., D.B., D.C.B. and J.B. contributed to the text of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.B.

Reprints and permissions information is available at www.nature.com/reprints.

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