Virus Research 278 (2020) 197864

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Short communication

Complete genome sequence of *Oryctes rhinoceros nudivirus* isolated from the coconut rhinoceros beetle in Solomon Islands

Kayvan Etebari[a,*,1], Igor Filipović[a,b,1], Gordana Rašić[b], Gregor J. Devine[b], Helen Tsatsia[c], Michael J. Furlong[a]

[a] School of Biological Sciences, The University of Queensland, Brisbane, Australia
[b] QIMR Berghofer Medical Research Institute, Royal Brisbane Hospital, Brisbane, Australia
[c] Ministry of Agriculture and Livestock, Honiara, Solomon Islands

ARTICLE INFO

Keywords:
Oryctes rhinoceros nudivirus (OrNV)
Genome
Oxford Nanopore Technologies (ONT)
sequencing
Oryctes rhinoceros
Scarabaeidae
Coconut rhinoceros beetles
Solomon Islands

ABSTRACT

Oryctes rhinoceros nudivirus (OrNV) has been an effective biocontrol agent against the insect pest *Oryctes rhinoceros* (Coleoptera: Scarabaeidae) for decades, but there is evidence that resistance could be evolving in some host populations. We detected OrNV infection in *O. rhinoceros* from Solomon Islands and used Oxford Nanopore Technologies (ONT) long-read sequencing to determine the full length of the virus genomic sequence isolated from an individual belonging to a mitochondrial lineage (CRB-G) that was previously reported as resistant to OrNV. The complete circular genome of the virus consisted of 125,917 nucleotides, 1.698 bp shorter than the originally-described full genome sequence of Ma07 strain from Malaysia. We found 130 out of 139 previously annotated ORFs (seven contained interrupted/non-coding sequences, two were identified as duplicated versions of the existing genes), as well as a putatively inverted regions containing four genes. These results demonstrate the usefulness of a long-read sequencing technology for resolving potential structural variations when describing new virus isolates. While the Solomon Islands isolate exhibited 99.41 % nucleotide sequence identity with the originally described strain, we found several genes, including a core gene (*vlf-1*), that contained multiple amino acid insertions and/or deletions as putative polymorphisms of large effect. Our complete annotated genome sequence of a newly found isolate in Solomon Islands provides a valuable resource to help elucidate the mechanisms that compromise the efficacy of OrNV as a biocontrol agent against the coconut rhinoceros beetle.

Members of the family *Nudiviridae* form a monophyletic but highly diverse group of unassigned, large circular double-stranded DNA viruses, with enveloped and rod-shaped virions that are pathogenic for a wide range of arthropods (Burand, 1998; Wang et al., 2012). Currently, very few insect nudiviruses have been described, but the coconut rhinoceros beetle virus (*Oryctes rhinoceros nudivirus*, *OrNV*) has been used as a pest biocontrol agent for decades (Bedford, 1986). This virus, previously described as *Oryctes baculovirus*, was isolated from Malaysian populations of *O. rhinoceros* in the early 1960s and was successfully introduced into the coconut rhinoceros beetle populations across the South Pacific islands in order to reduce pest damage to coconut palms (Bedford, 2013; Huger, 2005; Marschal, 1970). OrNV infection can be fatal to *O. rhinoceros* larvae, pupae and adults, but the disease is chronic and external symptoms in adult insects can be obscure (Burand, 1998).

The geographic range of the coconut rhinoceros beetle, *O. rhinoceros*, in the Pacific has recently expanded beyond its previous distribution (Fiji, Papua New Guinea, Samoa and Tonga) and now includes Guam (2007), Hawaii (~2013), Solomon Islands (2015), and more recently Vanuatu (2019) and New Caledonia (2019). It has been suggested that this resurgence is related to an *O. rhinoceros* mitochondrial lineage, termed the CRB-G haplotype, that is considered resistant to OrNV (Marshall et al., 2017; Reil et al., 2018). To obtain a better insight into the geographic diversity in OrNV, we sequenced the genomic DNA material of a virus-infected *O. rhinoceros* adult beetle collected in Solomon Islands.

A female adult beetle was collected from Guadalcanal, Solomon Islands using a pheromone trap (Oryctalure, P046-Lure, ChemTica Internacional, S. A., Heredia Costa Rica) in January 2019 and preserved in 95 % ethanol. Total DNA was extracted from mid-gut tissue using the Qiagen Blood and Tissue DNA extraction kit, and was used for OrNV

* Corresponding author at: School of Biological Sciences, Goddard Bldg. (No. 8), The University of Queensland, St Lucia, QLD, 4072, Australia.
E-mail addresses: k.etebari@uq.edu.au (K. Etebari), i.filipovic@uq.net.au (I. Filipović), Gordana.Rasic@qimrberghofer.edu.au (G. Rašić), Greg.Devine@qimrberghofer.edu.au (G.J. Devine), Helen.Tsatsia@sig.gov.sb (H. Tsatsia), m.furlong@uq.edu.au (M.J. Furlong).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.virusres.2020.197864
Received 7 November 2019; Received in revised form 7 January 2020; Accepted 12 January 2020
Available online 13 January 2020

detection and beetle mitochondrial DNA amplification and sequencing. The *O. rhinoceros* mitochondrial haplotype was established as CRB-G (Marshall et al., 2017) *via* Sanger sequencing of the partial *cox1* gene sequence that was amplified using the universal barcode primers LCO1490 and HCO2198 (Folmer et al., 1994). Presence of the OrNV was confirmed by successful amplification of a 945 bp product using the OrV15 primers that target *OrNV-gp083* gene (Richards et al., 1999). The PCR product was visualized by gel electrophoresis and validated by Sanger sequencing. For ultra-long read DNA sequencing, we extracted high-molecular weight DNA from four legs and partial thorax tissue using a magnetic (SPRI) bead-based protocol. Specifically, smaller pieces of tissue (50 mm^3) were each incubated in a 1.7 ml eppendorf tube with 360 µL ATL buffer, 40 µL of proteinase K (Qiagen Blood and Tissue DNA extraction kit) for 3 h at RT, while rotating at 1 rpm (end-over-end). Then, 400 µL of AL buffer was added for 10 min, followed by 8 µL of RNase A for 5 min. Tissue debris was spun down quickly (1 min at 10,000 rcf) and 600 µL of homogenate was transferred to a fresh tube, where 600 µL of SPRI bead solution was added and incubated for 30 min while rotating at 1 rpm (end-over-end). After two washes with 75 % ethanol, a final volume of 50 µL TE buffer was added. DNA quality was assessed on the 4200 Tapestation system (Agilent) and the concentration was determined using the Qubit broad-range DNA kit. DNA size selection (enrichment of DNA > 10 kb) was done using the Circulomics Short Read Eliminator XS kit. Library preparation was done with 1 µg of size-selected HMW DNA, using the Ligation Sequencing Kit SQK-LSK109 (Oxford Nanopore Technologies, Cambridge UK) following manufacturer's guidelines. In total, four libraries were loaded onto the MinION sequencing devices using the Flow Cell model R9.4.1 (Oxford Nanopore Technologies,) and sequenced by running the ONT MinKNOW Software. High-accuracy base calling on the raw sequence data was done with the Guppy base caller ONT v.3.2.4. High-quality sequences (= > Phred 13) were used for the genome assembly with Flye v.2.5 (metagenome assembly mode) (Kolmogorov et al., 2019).

Open reading frames in the resulting full-length circular consensus sequence were identified using CLC Genomics Workbench ver. 12.0 (QIAGEN) and then visually inspected and compared with the previously generated OrNV reference sequence (NC_011588). The NC_011588 reference genome sequence (named Ma07) was produced by multiple displacement amplification (MDA) of the whole viral genome from the homogenized mid-gut tissue of six infected insects collected from Johor, Peninsular Malaysia in 2007 (Wang et al., 2008). Finally, Nanopore long-read homopolymer stretches longer than 4 nucleotides were checked for artificial (sequencing technology-based) deletions or insertions and corrected based on the reference sequence.

Here, we produce 125,917 bp of the complete circular genome of OrNV isolate using the Oxford Nanopore Technologies (ONT) long-read sequencing platform, with a median depth of 1,196 × . The annotated genomic sequence of this Solomon Islands isolate has been deposited in GenBank under the accession number MN623374. ONT produced tens of single molecule reads ≥50 kbp that each span 40 % or more of the entire virus genome (the longest read 98,6 kbp spanned 78 % of the genome), ensuring unambiguous contiguity of the final assembly.

This new version of OrNV genome is 1,698 bp shorter than the genome previously reported by Wang et al. (2008). We identified 130 genes out 139 annotated coding regions in the genome of the Solomon Island isolate of OrNV. Seven previously annotated hypothetical proteins (*gp032*, *gp068*, *gp050*, *gp081*, *gp082*, *gp066* and *gp091*) did not produce any coding regions due to the presence of multiple stop codons and frame shifting in the new assembled sequence. We also found a region with a rearrangement (inversion) of four genes when compared to the previous assembly (Fig. 1), supported by 446 single molecule reads (10–98.5 kbp long) that each span this entire region. This region also lacked two genes (*OrNV gp129* and *gp130*) coding for the previously reported hypothetical proteins (Fig. 1), which we determined to be duplicated versions of *OrNV gp135* and *gp136,* respectively, in our new assembly. These structural differences suggest heterogeneous virus isolates (with and without inversion), MDA-related sequencing or assembly errors in the previous annotation that was prepared using shorter-read technology, and indicate that long-read sequencing of non-amplified (raw) DNA extracts from single individual may be preferred when genomes of new DNA virus isolates are described in non-model organisms. Without re-sequencing the original Ma07 strain with the long-read technology, however, we cannot rule out the possibility that these changes represent a true inversion and gene loss in our isolate.

In total, 352 single nucleotide polymorphisms (SNPs) have been found in 89 genes, with 138 amino acid modifications in 53 coding regions when compared to the previously reported OrNV sequence (Wang et al., 2008). Among these amino acid modifications, we found 77 amino acid substitutions, as well as 37 deletions and 24 insertions (Supplementary Table 1). 60 % of identified genes in the newly sequenced OrNV isolate have identical amino acid sequences with Ma07, and only eight genes have more than four amino acid modifications - mostly deletions and insertions rather than substitutions (Fig. 1, Supplemental Table 1). The greatest number of amino acid modifications has been found in the *GrBNV_gp28-like protein* and *GrBNV_gp48-like protein* genes, with 18 and 10 changes, respectively (Fig. 1). *GrBNV_gp28-like* gene has been identified as one of 12 genes with a higher rate of amino acid substitution between OrNV and Drosophila innubila nudivurs (DiNV), with a suspected role in adaptation to a novel host system (Hill and Unckless, 2018). Previously analyzed nudivirus genome data suggest that *vlf-1*, *pif-1*, *pif-3* may also be important in adapting to a new host (Hill and Unckless, 2018). While our isolate did not contain any changes in the amino acid sequences for *pif-1* or *pif-3* genes when compared to the Ma07 isolate, *vlf-1* gene was among the most modified (with four additions and one deletion, Fig. 1). *Vlf-1* is a one of core genes required for late and very late viral gene transcription and normal capsid assembly, and serves an essential function during the final stages of the DNA packaging process (Vanarsdall et al., 2006). Five of the eight most differing genes in the Solomon Islands isolate are located within or adjacent to the putatively inverted genomic region (Fig. 1) pointing to a region of interest that could be undergoing a larger structural change (within linked genes).

We also compared the newly-sequenced isolate to other available partial sequences from different OrNV isolates. A pairwise comparison of a genomic region with 40 ORFs (36,920 bp) revealed 99.85 % nucleotide sequence identity between the Solomon Islands isolate and PV505 isolate from the Philippines (accession number AH015832), and 99.77 % identity between PV505 and the Malaysian isolate Ma07 (accession number NC_011588). A partial sequence of the *dna polymerase B* gene was available for additional nine OrNV isolates from Indonesia (accession numbers from MK241540 to MK241548) for the phylogenetic reconstruction. Multiple sequence alignment with CLUSTAL Omega (Sievers et al., 2011) produced the alignment lengths between 3,744 bp and 4,989 bp that were used to generate a phylogenetic tree with a maximum likelihood approach implemented in PhyMLv.3.0 (Guindon et al., 2010) and the Bayesian method in MrBayes (Ronquist et al., 2003), assuming the GTR substitution model and estimation of all parameters from the data. Both methods produced identical tree topologies, with the Solomon Island isolate being most closely related to PV505 from the Philippines, and belonging to a lineage that also includes Ma07 from Malaysia and two isolates from Indonesia (Sulteng and Riau) (Fig. 2, Supplementary Fig. 1).

The presence of genetic variation among field isolates of OrNV has also been previously demonstrated with the restriction fragment length polymorphisms (Moslim et al., 2011). We produced the *in silico* restriction endonuclease cleavage site map of the Solomon Islands OrNV isolate using CLC genomic Workbench version 12 that contained 43 EcoRI, 27 HindIII, 21 BamHI and 7 PstI fragments, respectively (Supplemental Fig. 3). In comparison, the previous version of the OrNV genome (Ma07) contains 44 EcoRI, 26 Hind III, 22 BamHI and 7 PstI cleavages sites (Wang et al., 2008).

Crawford and Zelazny (1990) did not find any changes in the *O.*

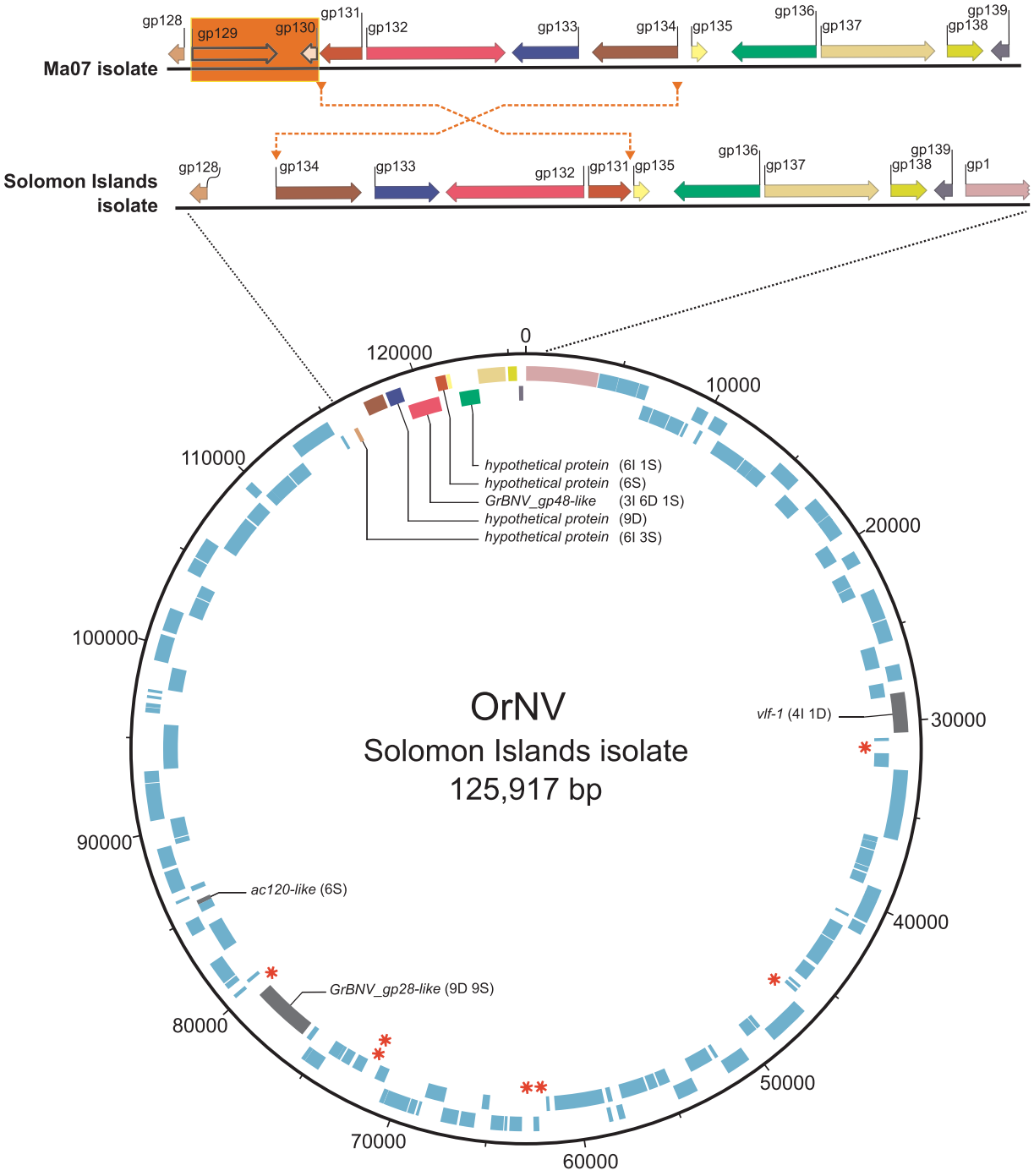

Fig. 1. Circular genome diagram of OrNV Solomon Islands (SI) isolate. Outer ring (cyan) represents genes on the forward strand and inner ring represents genes on the reverse strand. The SI isolate of OrNV is 1,698 bp shorter than Ma07. Highlighted arrangement that differs between the original (Ma07) and the new isolate (Solomon Islands isolate) contains four genes within a putative inversion, and two hypothetical proteins (*gp129 and gp130*) that are not present in the new assembly (they are identified as duplicated versions of *gp135 and gp136*). Gene names (with the number and type of amino acid modifications in brackets) are given for the eight most differentiated genes. Red asterisks indicate the positions of the previously annotated hypothetical proteins that did not produce full coding regions in the new assembled sequence due to the presence of multiple stop codons and frame shifting.

rhinoceros viral genome < 2 years after the virus was introduced into in the beetle population in the Maldives. However, they reported some modifications (a point mutation and recombination) over a longer, 4-year period (Crawford and Zelazny, 1990).

A study conducted on samples collected from Guadalcanal, Solomon Islands only three years prior to our collection did not detect OrNV infection (Marshall et al., 2017), suggesting very recent virus invasion into this host population. Although closely related to the OrNV isolate PV505 from the Philippines, the source of the virus found in our 2019 collection is not known; it could have been introduced deliberately as a biological control agent or through accidental incursion of infected beetles from neighboring islands. Interestingly, it was found in an individual from the CRB-G mitochondrial lineage, whose members have a low mortality rate from the OrNV infection (Marshall et al., 2017). Decreased mortality, however, is not necessarily related to host resistance to the virus, as many other factors could be involved in this phenomenon. During the course of evolution, viruses coevolve with their hosts to overcome host resistance and gain the upper hand in the

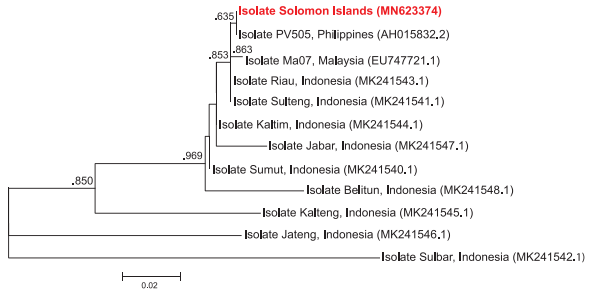

Fig. 2. Maximum likelihood phylogeny based on the partial sequence of the *dna_polB* gene from 12 OrNV isolates. Branch support (%) was obtained from 1000 bootstraps in PhyML v.3.0.

evolutionary arms race (Hill and Unckless, 2018). While the Solomon Islands isolate exhibited high nucleotide and amino acid sequence identity with the originally described isolate Ma07, we did find several genes, including one core gene (*vlf-1*), that contained multiple amino acid insertions and/or deletions as putatively large-effect mutations (Zia et al., 2011). Further sequence analyses of OrNV isolates in combination with the characterization of its effects on the host across multiple populations is needed to understand if such mutations are polymorphisms that compromise the efficacy of OrNV as a biocontrol agent that has kept the coconut rhinoceros beetle at bay across the Pacific for several decades. Our complete sequence and annotation of a newly found isolate in Solomon Islands will provide a valuable resource for such studies.

Authors statement

Kayvan Etebari: Conceptualization, Methodology, Investigation, Data curation, Visualization, Writing- Original draft preparation.

Igor Filipović: Conceptualization, Methodology, Investigation, Data curation, Visualization, Writing- Original draft preparation.

Gordana Rašić: Methodology, Investigation, Resources, Writing-Reviewing and Editing.

Gregor J. Devine: Resources, Supervision, Writing - Reviewing and Editing.

Helen Tsatsia: Resources, Investigation.

Michael J. Furlong: Conceptualization, Writing - Reviewing and Editing, Funding acquisition, Project administration, Supervision.

Acknowledgements

This project was supported by the Australian Centre for International Agricultural Research funding (HORT/2016/185), the University of Queensland (UQECR2057321) and by core funds from the Mosquito Control Laboratory at QIMR Berghofer MRI.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.virusres.2020.197864.

References

Bedford, G.O., 1986. Biological control of the rhinoceros beetle (*Oryctes rhinoceros*) in the South Pacific by baculovirus. Agric. Ecosyst. Environ. 15 (2–3), 141–147.
Bedford, G.O., 2013. Biology and management of palm dynastid beetles: recent advances. In: In: Berenbaum, M.R. (Ed.), Annual Review of Entomology Vol. 58. pp. 353–372 Annual Reviews, Palo Alto.
Burand, J.P., 1998. Nudiviruses. In: Miller, L.K., Ball, L.A. (Eds.), The Insect Viruses. Springer US, Boston, MA, pp. 69–90.
Crawford, A.M., Zelazny, B., 1990. Evolution in *Oryctes* baculovirus: rate and types of genomic change. Virology 174 (1), 294–298.
Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Marine Biol. Biotechnol. 3 (5), 294–299.
Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59 (3), 307–321.
Hill, T., Unckless, R.L., 2018. The dynamic evolution of *Drosophila innubila* Nudivirus. Infect. Genet. Evol. 57, 151–157.
Huger, A.M., 2005. 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 (1), 78–84.
Kolmogorov, M., Yuan, J., Lin, Y., Pevzner, P.A., 2019. Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. 37 (5), 540–546.
Marschal, K., 1970. Introduction of a new virus disease of the coconut Rhinoceros beetle in Western Samoa. Nature 225 (5229), 288–289.
Marshall, S.D.G., Moore, A., Vaqalo, M., Noble, A., Jackson, T.A., 2017. A new haplotype of the coconut rhinoceros beetle, *Oryctes rhinoceros*, has escaped biological control by *Oryctes rhinoceros* nudivirus and is invading Pacific Islands. J. Invertebr. Pathol. 149, 127–134.
Moslim, R., Kamarudin, N., Ghani, I.A., Wahid, M.B., Jackson, T.A., Tey, C.C., Ahdly, M., 2011. Molecular approaches in the assessment of *Oryctes rhinoceros* virus for the control of rhinoceros beetle in oil palm plantations. J. Oil Palm Res. 23, 1096–1109.
Reil, J.B., Doorenweerd, C., San Jose, M., Sim, S.B., Geib, S.M., Rubinoff, D., 2018. Transpacific coalescent pathways of coconut rhinoceros beetle biotypes: resistance to biological control catalyses resurgence of an old pest. Mol. Ecol. 27 (22), 4459–4474.
Richards, N.K., Glare, T.R., Aloali'i, I., Jackson, T.A., 1999. Primers for the detection of *Oryctes* virus from Scarabaeidae (Coleoptera). Mol. Ecol. 8 (9), 1552–1553.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: bayesian phylogenetic inference under mixed models. Bioinformatics 19 (12), 1572–1574.
Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W.Z., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539.
Vanarsdall, A.L., Okano, K., Rohrmann, G.F., 2006. Characterization of the role of very late expression factor 1 in baculovirus capsid structure and DNA processing. J. Virol. 80 (4), 1724–1733.
Wang, Y.J., Bininda-Emonds, O.R.P., Jehle, J.A., 2012. Nudivirus genomics and phylogeny. In: Garcia, M.L., Romanowski, R. (Eds.), Viral Genomes - Molecular Structure, Diversity, Gene Expression Mechanisms and Host-Virus Interactions, pp. 33–52.
Wang, Y.J., Kleespies, R.G., Ramle, M.B., Jehle, J.A., 2008. Sequencing of the large dsDNA genome of *Oryctes rhinoceros* nudivirus using multiple displacement amplification of nanogram amounts of virus DNA. J. Virol. Methods 152 (1–2), 106–108.
Zia, A., Moses, A.M., 2011. Ranking insertion, deletion and nonsense mutations based on their effect on genetic information. BMC Bioinform. 12, 299.