Identification and full genomic sequence of nerine yellow strip virus

Beddoe, N., Adams, I. P., McGreig, S. and Forsyth, A.

Title page for Annotated Sequence Record submission

Author names

Naomi Beddoe, Ian P. Adams, Sam McGreig and Alec Forsyth

Title

Identification and full genomic sequence of nerine yellow stripe virus

Author affiliations and email addresses

Naomi Beddoe

Canterbury Christ Church University, North Holmes Road, Canterbury, Kent, CT1 1QU, UK

naomi.beddoe@canterbury.ac.uk

Alec Forsyth

Canterbury Christ Church University, North Holmes Road, Canterbury, Kent, CT1 1QU, UK

alec.forsyth@canterbury.ac.uk

Ian Adams

Fera Science Limited, The National Agri-Food Innovation Campus, Sand Hutton, York, YO41 1LZ

ian.adams@fera.co.uk

Sam McGreig, The National Agri-Food Innovation Campus, Sand Hutton, York, YO41 1LZ

sam.mcGreig@fera.co.uk

Corresponding author

Alec Forsyth

16-digit ORCID of the author(s)

Naomi Beddoe 0000-0002-7766-891X

Ian P. Adams 0000-0002-5589-6460

Alec Forsyth 000-0002-4283-3583

Funding

Canterbury Christ Church University and the EUPHRESCO project, VirusCurate funded this work.

Conflicts of interest/Competing interests

The authors declare that they have no conflict of interest.

Availability of data and material

GenBank accession MT396083

Code availability (software application or custom code)

Sequence data were aligned and visualised using Sequencher (Gene Codes).

MinION data were recalled using Guppy3.3 (Oxford Nanopore Technologies) quality checked using NanoPlot (https://github.com/wdecoster/NanoPlot), NanoQC (https://github.com/wdecoster/nanoQC) and Nanofilt (https://github.com/wdecoster/nanofilt) and analysed using the Pomoxis pipeline (https://github.com/nanoporetech/pomoxis).

MiSeq data were trimmed using Sickle version 1.33 (https://github.com/najoshi/sickle) and assembled with Trinity version 2.8.4 (https://github.com/trinityrnaseq/trinityrnaseq).

Phylogenetic analyses were performed using MEGAX (https://www.megasoftware.net).

Recombination analyses were performed using RDP4.100 (http://web.cbio.uct.ac.za/~darren/rdp.html).

Abstract

This study reports the first complete genome sequence of nerine yellow stripe virus (NeYSV, GenBank MT396083). The genome consists of 10165 nucleotides, excluding the 3' terminal poly(A) tail. A single open reading frame encodes a large polyprotein of 3294 amino acids with typical potyvirus features. The nuclear inclusion b and coat protein region shares 95% identity with previously reported NeYSV partial sequence (NC_043153.1). Phylogenetic analysis of polyprotein amino acid sequence showed that NeYSV clustered with hippeastrum mosaic virus (YP_006382256.1).

Main text

Nerines are native to Southern Africa, valued for their striking flowers and are popular within horticultural and floristry industries. They are grown commercially in Europe, internationally traded and are becoming increasingly important in the UK. Nerines are bulbous flowering plants, members of the *Amaryllidaceae* and include hardy and tender varieties. Due to the vegetative nature of propagation, accumulated virus infection can persist through generations. Nerine yellow stripe virus (NeYSV), a virus in the *Potyviridae*, was previously identified (Wendy Monger personal communication to Andrew Eames, 2007) in nerine plants growing in the UK but the full sequence was not determined. NeYSV host range appears to be confined to the *Amaryllidaceae*, with infections reported in multiple countries and symptoms including mild to severe mosaic, darker green oval spots, chlorotic/yellow stripes on leaves/stalks and occasional flower breaking [1-5]. NCBI currently holds partial sequences of NeYSV derived from various plant genera: these include two sequences from a nerine plant (Netherlands, NC_043153.1, EF362621.1), seven partial sequences from *Crinum* genus plants (USA, MG012805.1 and India, KJ886934.2, KJ886933.2, KM066971.1, KM066970.1, KM066969.1, KM066968.1), two partial sequences from *Amaryllis* genus plants (USA, JX865782.1 and New Zealand, FJ618537.1) and one partial sequence each from a *Stenomasson* genus plant (Netherlands, EU042758.1), a *Hymenocallis* genus plant (Netherlands, EF362622.1) and a *Cyrtanthus* genus plant (previously *Vallotta* genus) (New Zealand, DQ407932.1). A plant growing in a pot in the UK (*Nerine bowdenii* × *N. sarniensis* 'Hera') displayed virus-like symptoms, including chlorotic leaf patches and was used as source of virus material. Subsequent analysis revealed the plant was also infected with other viruses (data not shown), which could have contributed to the chlorotic leaf patch symptoms. Total RNA was extracted using RNeasy Plant Mini Kit (Qiagen) and cDNA synthesised using RNA to cDNA EcoDry premix double primed (TakaraBio) following manufacturer's instructions. NeYSV presence was confirmed by amplifying cDNA using DreamTaq Green 2x (Invitrogen), with coat protein (CP) specific primers (forward 5-GAATGGTATTGAGCTAGAGCAG-3 and reverse 5-CTCACACCAAGAAGTGAGTG-3) designed to GenBank accessions EF362622.1/ EF362621.1. Purified (QIAquick PCR purification Kit, Qiagen) fragments were ligated to pCR2.1 TA vector (Invitrogen) and Sanger sequenced (DBS Genomics, Durham). Sequence data were compared to NCBI using blastn/blastx and shared 96% identity with previously reported partial NeYSV sequence EF362622.1. A MinION cDNA library was prepared and sequenced (SQK-LSK108 kit/ R9 flowcell, Oxford Nanopore Technologies) as per manufacturer's instructions (Supplementary data 1 'MinION-1DcDNA-sequencing-by-ligation_v2'). To improve accuracy, reads were recalled with Guppy3.3 (Oxford Nanopore Technologies) with high accuracy

base calling selected, quality checked using NanoPlot (https://github.com/wdecoster/NanoPlot), NanoQC (https://github.com/wdecoster/nanoQC) and Nanofilt (https://github.com/wdecoster/nanofilt) and analysed using the Pomoxis pipeline (https://github.com/nanoporetech/pomoxis). A 3675 nucleotide (nt) contig from the original base called data shared 62.2% amino acid (aa) identity with hippeastrum mosaic virus (HiMV) cytoplasmic inclusion protein (CI) protein (YP_006390061.1) and polyprotein (YP_006382256.1, AFJ92920.1), covering the P3, 6K1, CI and 6K2 regions. An 1884 nt contig from the Guppy 3.3 recalled data (Q5 quality filter) shared 94% identity with previously reported partial NeYSV sequence (NC_043153) covering the NIb and CP regions. A 4417 nt contig from Guppy 3.3 recalled data (Q7 quality filter) shared 54% identity with HiMV (YP_006382256.1, AFJ92920.1) covering part of the CI, second 6-kDa protein (6K2), viral protein genome-linked (VPg), nuclear inclusion a protein/protease (NIa), NIb and CP regions. A MiSeq cDNA library was created, sequenced and analysed using the method described by Fribourg *et al*. [6]. Reads were assembled with Trinity (https://github.com/trinityrnaseq/trinityrnaseq) version 2.8.4, producing a single complete contig of nerine yellow stripe virus. The 10148 nt contig shared a 95% nt identity with NeYSV (NC_043153.1) over the NIb and CP regions and gave a full potyvirus-like open reading frame (ORF) in the 2nd frame. The MiSeq contig was validated by RT-PCR with primers designed to the MiSeq data (Supplementary data 2) in overlapping regions and the resulting amplicons were Sanger sequenced. Seven basepair differences were identified between the MiSeq assembly and the cloned sequence data that were derived from RT-PCR using primers designed to the MiSeq assembly, three of which were coding differences (data not shown). These differences may be down to the MiSeq assembly producing a consensus sequence of the viral diversity within the sample extract, as opposed to data derived from individual clones from different leaf extracts.

The 3675, 1884 and 4417 nt MinION contigs shared 35%, 18% and 43% cover and 92.26%, 99.01% and 99.09% identity (all with E-value 0.0) respectively with the MiSeq contig. Coverage of the MiSeq contig in relation to the full genome of nerine yellow stripe virus was 99.9% and average depth was 71.8. The 5' and 3' untranslated regions (UTR) were identified and confirmed by rapid amplification of cDNA ends (TakaraBio) as per manufacturer's instructions. The genome (10165 nt) organisation (Fig.1) is typical of a potyvirus, including a 106 nt 5' UTR, a single ORF encoding a 3294 aa polyprotein and a 174 nt 3' UTR with a poly(A) tail. Potybox 'a' was not identified in the 5' UTR, however potybox 'b' [7] was identified ^{49}CAAGCA^{54}. The initiation codon sequence starting at nt position 107 is ^{103}CAAC<u>ATG</u>TC^{111}. Nine putative polyprotein cleavage sites were predicted [8] for ten mature proteins: first protein/protease (P1), helper component protease (HC-Pro), third protein (P3), first 6-kDa protein (6K1), CI, 6K2, VPg, NIa-Pro, NIb and CP and pretty interesting potyvirus open reading frame (PIPO) was predicted within P3 [8,9,10]. PIPO starts from a conserved region ^{3496}GGAAAAAAA^{3504} and gives a 256 nt/85 aa product. Several conserved motifs common to potyviruses [9] were identified (Supplementary data 3).

The ability of potyviruses to be transmitted by aphid vectors relates to conserved HC-Pro and CP region motifs. Several conserved aphid transmissibility-related motifs in the HC-Pro have been previously identified including KITC, KLSC, CCC and PTK [10]. ^{519}KITC^{522}, ^{751}KLSG^{754}, ^{760}CCCVT^{764} and ^{778}PTK^{780} were identified in the HC-Pro region of the NeYSV sequence (Supplementary data 3). CP motifs of NeYSV were identified and are discussed further; DAG motifs have been identified in

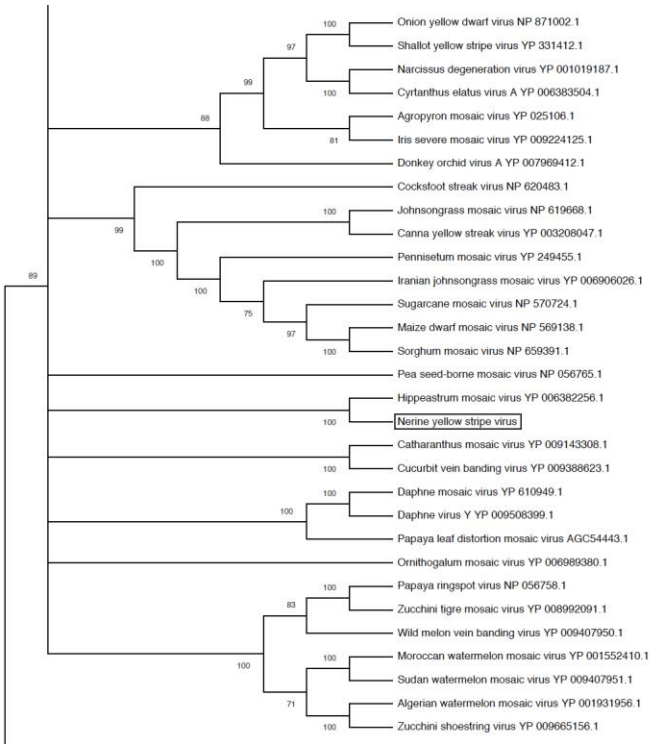
many potyvirus CP regions [11]. Nigam *et al*. highlighted two potential DAG motifs (one proximal and one distal to the CP N terminus) and showed that motif sequence and location variation are possible in this hypervariable region outside the CP core domain [11]. The NeYSV sequence exhibits a ^{6}NAG^{8} motif (proximal to CP N terminus) and a ^{28}NVG^{30} in the distal location (numbered from the beginning of the CP region). Previously published partial NeYSV sequences show similar sequence/location combinations and the closest virus from phylogenetic analyses, HiMV, exhibits a ^{3}DAG^{5}, ^{21}DAG^{23} combination.

NeYSV and 126 complete potyvirus polyprotein aa sequences listed on NCBI were analysed using ClustalW (alignment) and MegaX (tree generation) [12]. NeYSV clustered with HiMV using both Maximum Likelihood (ML) (Fig.2) and Neighbor-joining (NJ) (data not shown) estimates. Similar results were generated using complete potyvirus nt sequences (data not shown) and predicted cleaved NeYSV protein sequences (except NIa, which was too divergent to align). When the NeYSV sequence was compared to the closest related virus on the phylogenetic tree, HiMV (YP_006382256.1), the following aa percentage identities were returned using blastp. Polyprotein 53.84%, P1 31.85%, HC-Pro 55.90%, P3 35.90%, PIPO 34.48%, 6K1 47.06%, CI 67.24%, 6K2 50.94%, VPg 48.40%, NIa 59.67%, NIb 65.45% and CP 56.59%. For the two viruses either side of this branch on the phylogenetic tree, polyprotein and CP aa comparisons were made. Pea seed-borne mosaic virus (NP_056765.1) polyprotein 45.56% and CP 51.14%. Sorghum mosaic virus (NP_659391.1) polyprotein 42.82%, CP 54.51%. On the other side of the NeYSV/HiMV branch, catharanthus mosaic virus (YP_009143308.1) polyprotein shared 44.58% and CP 57.08%. Cucurbit vein banding virus (YP_009388623.1) polyprotein shared 42.80% and CP 57.08%. The NeYSV aa sequence does not pass the percentage identity threshold suggested by ICTV for either polyprotein (<82%) or CP (<80%) when compared to HiMV, pea seed-borne mosaic virus, sorghum mosaic virus, catharanthus mosaic virus or cucurbit vein banding virus. This supports the existing situation that NeYSV is a distinct potyvirus as listed in the ICTV Master Species List 2019.v1 [14]. Full genome nucleotide potyvirus sequences (Supplementary data 4, using linked complete nucleotide GenBank accessions) were tested for the presence of phylogenetic anomalies using the full suite of options in RDP4 with default parameters [15,16]. No nerine yellow stripe virus recombination events were detected by four or more methods. Additionally, previously reported partial nerine yellow stripe virus sequences (Supplementary data 4 b) were tested alongside the CP region of the MiSeq sequence and no recombination events were detected.

Fig.1 Schematic representation of the genome organisation of NeYSV. The 5’ and 3’ untranslated regions (UTRs) are shown as bold lines. The oval represents a single ORF encoding a polyprotein, the numbers below the oval show the first and last nucleotide position of the ORF and the total nucleotide length exclusive of the poly(A) tail. The polyprotein is predicted to encode ten mature proteins, typical of potyvirus organisation; P1 (first protein/protease), HC-Pro (helper component protease), P3 (third protein), 6K1 (first 6-kDa protein), CI (cytoplasmic inclusion protein), 6K2 (second 6-kDa protein), VPg (viral protein genome-linked), NIa (nuclear inclusion a protein/protease), Nib (nuclear inclusion b) and CP (coat protein). The numbers above the box show the amino acid position of the predicted cleavage sites by the viral proteinases. A small ORF is shown above the

box indicating PIPO (pretty interesting potyvirus open reading frame) encoded by frameshift or transcriptional slippage from the P3 region.

Fig.2 Partial phylogenetic tree showing NeYSV polyprotein sequence clustered with hippeastrum mosaic virus (YP_006382256.1). Complete potyvirus polyprotein sequences were aligned using ClustalW (parameters for pairwise alignment were Gap opening penalty 10.00, Gap extension penalty 0.10 and for multiple alignment were Gap opening penalty 10.00, Gap extension penalty 0.20; Use negative matrix was off, delay divergent cutoff % was 30 and keep predefined gap was unchecked) and an unrooted tree was generated in MegaX using ML, tested to 1000 bootstrap replicates. ML aa tree substitution model was Jones-Taylor-Thornton [13] using uniform rates on all sites, heuristic method was Nearest-Neighbor-Interchange, initial tree was set by default to NJ/BioNJ and no branch swap filter was set. Condensed bootstrap tree generated by collapsing branches of support values lower than 70%. The NeSYV sequence was analysed with 126 potyvirus sequences, this figure shows 31 sequences, including NeYSV. A list of potyvirus sequences used in the phylogenetic analysis is available in Supplementary data 4 and the full phylogenetic tree image is available in Supplementary data 5.



Acknowledgements

Canterbury Christ Church University and the EUPHRESCO project, VirusCurate funded this work.

References

[1]. Pearson, MN, Cohen, D, Cowell, SJ, Jones, D, Blouin, A, Lebas, BSM, Shiller, JB and Clover, GRG (2009) A survey of viruses of flower bulbs in New Zealand. Australasian Plant Pathology 38:305–309

[2]. Limbath, DP, Blanchfield, AL and Wilson, CR (2000) Incidence and pathology of common viruses of nerine in Tasmania. Annals of Applied Biology 136:21-25

[3]. Loebenstein, G Lawson, RH and Brunt, AA (1995) Virus and virus-like diseases of bulb crops. John Wiley & Sons, Chichester

[4]. Guaragna, MA, Lamborn, J, Groth-Helms, D, Mollov, D, Lockhart, B, van Schadewijk, L, Hammond, J and Jordan, R (2013) First report of Nerine yellow stripe virus in Amaryllis in the United States. Plant Disease 97(10):1389

[5]. Kumar, S, Raj, R, Kaur, C and Raj, SK (2015) Association of Nerine yellow stripe virus with mosaic disease of Crinum asiaticum ornamental plants in India. Plant Disease 99(11):1655

[6] Fribourg, CE, Gibbs, AJ, Adams, IP, Boonham, N. and Jones, RA, 2019 Biological and Molecular Properties of Wild potato mosaic virus Isolates from Pepino (Solanum muricatum). Plant disease 103(7):1746-1756

[7] Chung, BY.-W, Miller, WA, Atkins, JF and Firth, AE (2008) An overlapping essential gene in the Potyviridae. PNAS Microbiology 105(15):5897-5902. https://doi/10.1073/pnas.0800468105

[8] Adams, MJ, Antoniw, JF and Beaudoin, F (2005) Overview and analysis of the polyprotein cleavage sites in the family Potyviridae Molecular Plant Pathology 6(4):471-487

[9] Worrall, EA, Hayward, AC, Fletcher, SJ and Mitter, N (2019) Molecular characterization and analysis of conserved potyviral motifs in bean common mosaic virus (BCMV) for RNAi-mediated protection. Archives of Virology 164:181-194

[10] Dombrovsky, A., Huet, H., Chejanovsky, N., and Raccah, B. (2005). Aphid transmission of a potyvirus depends on suitability of the helper component and the N terminus of the coat protein. Arch. Virol. 150 (2), 287–298. doi: 10.1007/ s00705-004-0407-7

[11] Nigam D, LaTourrette K, Souza PFN and Garcia-Ruiz H (2019) Genome- Wide Variation in Potyviruses. Front. Plant Sci. 10:1439. https://doi/10.3389/fpls.2019.01439

[12] Kumar, S, Stecher, G, Li, M, Knyaz, C and Tamura, K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35:1547-1549

[13] Jones, DT, Taylor, WR and Thornton, JM (1992) The rapid generation of mutation data matrices from protein sequences. Computer Applications in the biosciences 8:275-282

[14] ICTV (2019) ICTV Master Species List 2019.v1 Available online https://talk.ictvonline.org/files/master-species-lists/m/msl/9601 Accessed 10 July 2020

[15] Martin DP, Murrell B, Golden M, Khoosal A, & Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evolution 1: vev003 doi: 10.1093/ve/vev003

[16] The RDP method: Martin, D. & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. Bioinformatics 16: 562-563.