Supporting Information.

**Characterization of a New *Chlorovirus* Type with Permissive and Non-permissive Features on Phylogenetically Related Algal Strains**

Cristian F. Quispea,c,1#, Ahmed Esmaela,d,1, Olivia Sondermana,c, Michelle McQuinna, Irina Agarkovaa,b, Mohammed Battahd, Garry A. Duncane, David D. Dunigana,b, Timothy P.L. Smithf, Cristina DeCastrog,Immacolata Specialeh, Fangrui Mai, and James L.Van Ettena,b

aNebraska Center for Virology, bDepartment of Plant Pathology and cSchool of Biological Science, University of Nebraska-Lincoln, NE 68583-0900. dBotany Department, Faculty of Science, Benha University, Qalubiya Governorate, 13511, Egypt, eDepartment of Biology, Nebraska Wesleyan University, Lincoln, Nebraska, USA, fAgricultural Research Service, United States Department of Agriculture, U.S. Meat Animal Research Center, Clay Center, NE 68933-0166, gDepartment of Agricultural Sciences, University of Napoli, Via Universita 100, 80055 Portici (NA) Italy, hDepartment of Chemical Sciences, University of Napoli, Napoli, Italy, ICenter for Biotechnology, University of Nebraska, Lincoln, NE 68583-0900

#Current address: Carrera de Ciencia y Producción Agropecuaria. Escuela Agrícola Panamericana Zamorano, Honduras, CA.

**Contents**

**Page**

**Glycopeptide isolation and chemical composition** 2

**Sup. Table 1.** Annotation of the OSy-NE5 genome 4

**Sup. Table 2.** OSy-NE5 encoded tRNAs 12

**Sup. Table 3.** OSy-NE5 color encoded CDSs used in the genome map (Fig. 6) 13

**Sup. Table 4.** Proton and carbon chemical shifts (600 MHz) recorded in

deuterated water at 310 K with acetone as internal standard 17

**Sup. Figure 1**. Expansion of OSy-NE5 glycopeptide HSQC spectrum

measured at 600 MHz, 310 K. 18

**Sup. Figure 2.** Expansion of OSy-NE5 glycopeptide T-ROESY spectrum

measured at 600 MHz, 310 K. 19

**Glycopeptide isolation and chemical composition.**

*Glycopeptide isolation*

The major capsid protein was isolated, protease digested and the glycopeptide purified as described previously (De Castro et al., 2013). Monosaccharide absolute configuration D for all residues, but fucose, was determined by preparing the fully acetylated 2-(-)-octyl derivatives as reported for PBCV-1 (De Castro et al., 2013).In order to determine the absolute configuration of the 3 O Me-xylose, the acetylated octylglycoside mixture was methylated following the procedure described for virus NY-2A (De Castro et al., 2016). Also in this case, the xylose was D configured.

*NMR spectroscopy*

NMR experiments were recorded in D2O at 310 K on a Bruker DRX-600 instrument equipped with a cryo-probe. Homonuclear experiments were recorded using 512 FIDs of 2048 complex with 40 scans per FID and mixing time of 100 ms and 300 ms were used for TOCSY and T-ROESY spectra acquisition, respectively. Heteronuclear experiments (HSQC, HMBC and HSQC-TOCSY) were acquired with 512-256-240 FIDs of 2018 complex point with 60-120-200 scans for FID, respectively. The mixing time of 80 ms was used for HSQC-TOCSY spectrum. Standard Bruker software (Topspin 3.1) was used for process and analysis of all spectra.

*NMR characterization*

The 1H-NMR spectrum (projection in Sup. Fig. 1, chemical shift values in Sup. Table 4) showed six signals in the anomeric region from 5.7 to 4.4 ppm, that were labeled with a letter **A** to **F** in order of their decreasing chemical shift, a carbinolic region (from 4.3 to 3.10 ppm) containing one O-Me signal at 3.63 ppm and two methyl signals at around 1.3 ppm, typical of 6-deoxyresidues. Inspection of the HSQC spectrum (Sup. Fig. 1, Sup. Table 4) disclosed that the **D** was *N*-linked to the protein due to its carbon chemical shift at 80.7 ppm, and its pattern in the TOCSY spectrum was consistent with the *gluco* stereochemistry. Integrating TOCSY information with those from the other spectra (COSY and HSQC) identified this residue as a beta-glucose, substituted at both O3 and O4, similarly to what was reported for the PBCV-1 glycopeptide (De Castro et al., 2013). The same approach was used to identify all other residues.

Regarding **A**, TOCSY experiment displayed a correlation from the anomeric proton (5.63 ppm) with H-2, H-3 and H-4, furthermore COSY spectrum did not display any intense correlation between H-3 and H-4, as typically occurs for *galacto* configured residue. T-ROESY experiment (Sup. Fig. 2) was used to determine the position of H-5 proton that was in turn correlated with a methyl group at 1.30 ppm. Thus, A was an alpha-fucose residue, and the low field values of the carbon C-2, C-3 and C-4 with respect to standard values determined that it was fully substituted as in PBCV-1. The spectroscopic pattern of **B** was very similar to that of **A**, with the difference that H-6 of **B** correlated with a carbon at 62.3 ppm, thus it was a galactose unit, not further substituted and α configured at the anomeric center on the basis of the carbon chemical shift values determined.

With regard to **C**, its anomeric proton displayed only one COSY and TOCSY correlation (H-1/H-2), whereas the connections up to the H-6 (a methyl group) were visible from H-2: this information coupled to HSQC spectrum, identified E as a terminal rhamnose. The spectroscopic pattern of **E** was similar to that of **F** and the efficient propagation of the magnetization in the TOCSY spectrum identified them as two beta-xylose units. Of note, H-3 of **F** resonated at (1H/13C) 3.23/86.8 ppm suggesting that this position was *O*-methylated as confirmed by the corresponding HMBC and T-ROESY correlations. Moreover, the T-ROESY experiment (Sup. Fig. 2) revealed that **E** was linked at *O*-4 of **A**, while **F** at *O*-4 of **D**. These two units take also the name distal and proximal, respectively, owing to the different distance from the N-glycosylation site.

Completion of the OSy-NE5 glycopeptide structure was inferred by analyzing the T-ROESY spectrum (Sup. Fig. 2), which disclosed that **A** was linked at *O*-3 of **D** and in turn had **B** at *O*-2, **C** at *O*-3 and **E** at *O*-4. Combining NMR information with those from absolute configuration permitted the construction of the structure of the glycan as depicted in Fig. 4B.

**References**

De Castro C, Molinaro A, Piacente F, Gurnon JR, Sturiale L, Palmigiano A, Lanzetta R, Parrilli M, Garozzo D, Tonetti M, Van Etten JL *PNAS.* **2013,** *110*, 13956-13960. Doi: 10.1073/pnas.1313005110.

De Castro C, Speciale I, Duncan G, Dunigan DD, Agarkova I, Lanzetta R, Sturiale L, Palmigiano A, Garozzo D, Molinaro A, Tonetti M, Van Etten JL *Angew. Chem. Int. Ed.* **2016**, 55, 654-65

**Sup. Table 1.** Annotation of the OSy-NE5 genome

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CDS** | **start** | **end** | **ORF length** | **Putative activity** | **PBCV-1 gene tag (or other Chlorovirus) of Osy-NE5 homolog** | **NCBI RefSeq/GenBank Accesion Number: NC64A\_PBCV-1 Sequence ID or other Chlorovirus** | **Cluster ID** | **#PBCV-1 homolog is active** |
| OS5\_001L | 478 | 1044 | 188 | hypothetical protein | SAG3.83\_TN603.4.2\_002L | AGE59612.1 | CL0468 |  |
| OS5\_002L | 1898 | 2083 | 61 |  | a004L |  | CL0757 |  |
| OS5\_003R | 2693 | 3874 | 393 | hypothetical protein, many Glu, Asp, and Glys | NSSF |  | n/d |  |
| OS5\_004R | 3950 | 4876 | 308 | Ser/Thr protein kinase | A034R | NP\_048382.1 | CN0002 | yes |
| OS5\_005L | 4881 | 6590 | 569 | DNA binding protein | A035L | NP\_048383.2 | CL0822 |  |
| OS5\_006L | 6610 | 6924 | 104 | hypothetical protein | A037L | NP\_048385.1 | CL0002 |  |
| OS5\_007L | 6964 | 7410 | 148 | SKP1-like protein | A039L | NP\_048387.1 | n/d | yes |
| OS5\_008R | 7537 | 8742 | 401 | hypothetical protein | A041R | NP\_048389.1 | CL0511 |  |
| OS5\_009L | 8739 | 10574 | 611 | AAA ATPase | A044L | NP\_048392.1 | CL0262 |  |
| OS5\_010R | 10642 | 11019 | 125 | hypothetical protein | A048R | NP\_048396.1 | CL0274 |  |
| OS5\_011L | 11016 | 11645 | 209 | Glycerophosphoryl diesterase phosphodiesterase | A049L | NP\_048397.1 | CL0285 |  |
| OS5\_012L | 11692 | 12117 | 141 |  | A050L |  | CN0039 |  |
| OS5\_013R | 12221 | 12349 | 42 | hypothetical protein | A050aL | YP\_004678878.1 | n/d |  |
| OS5\_014L | 12337 | 13146 | 269 | Aquaglyceroporin | Pbi\_OR0704.2.2\_056R | AGE58959.1 | CL0466 | yes |
| OS5\_015R | 13282 | 14370 | 362 | D-lactate dehydrogenase | A053R | NP\_048401.1 | CL0334 |  |
| OS5\_016L | 14558 | 14758 | 66 | ribonuclease E | NC64A\_IL-5-2s1\_196R | AGE54132.1 | CL0535 |  |
| OS5\_017L | 14822 | 15877 | 351 | Glycosyltransferase domain, GTB type | A659L | NP\_049015.2 | CL0446 |  |
| OS5\_018R | 15974 | 17461 | 495 | hypothetical protein, IgGFc-binding protein, Pro, Lys rich | A057aR | YP\_004678879.1 | CL0898 |  |
| OS5\_019R | 17494 | 18237 | 247 | Methyltransferase FkbM family | NC64A\_CvsA1\_020R | AGE52378.1 | CL0071 |  |
| OS5\_020L | 18334 | 18960 | 208 | hypothetical protein | A437aR | YP\_004678954.1 | n/d |  |
| OS5\_021R | 19035 | 19946 | 303 | hypothetical protein | NC64A\_CvsA1\_025R | AGE52380.1 | CL0063 |  |
| OS5\_022R | 19971 | 20762 | 263 | hypothetical protein | NC64A\_CvsA1\_027R | AGE52381.1 | CL0537 |  |
| OS5\_023R | 20878 | 21855 | 325 | hypothetical protein | A067R | NP\_048415.1 | CL0530 |  |
| OS5\_024R | 21888 | 22952 | 354 | hypothetical protein | A071R | NP\_048419.1 | CN0017 |  |
| OS5\_025L | 22962 | 23801 | 279 | Exostosin | A075L | NP\_048423.1 | CL0413 |  |
| OS5\_026L | 24085 | 24354 | 89 | No hits to anything | NSSF |  | n/d |  |
| OS5\_027L | 24418 | 24693 | 91 | hypothetical protein | A076L | NP\_048424.1 | CL0590 |  |
| OS5\_028L | 24686 | 24964 | 92 | hypothetical protein | A077L | NP\_048425.1 | CL0430 |  |
| OS5\_029R | 25083 | 25979 | 298 | N-carbamoylputrescine amidase | A078R | NP\_048426.1 | CL0441 | yes |
| OS5\_030R | 26750 | 27478 | 242 | hypothetical protein | A079R | NP\_048427.2 | n/d |  |
| OS5\_031L | 27482 | 28051 | 189 | hypothetical protein | A081L | NP\_048429.1 | n/d |  |
| OS5\_032L | 28124 | 28654 | 176 | hypothetical protein | A084L | NP\_048432.1 | CL0472 |  |
| OS5\_033R | 28800 | 29480 | 226 | Prolyl 4-hydroxylase | A085R | NP\_048433.1 | CL0483 | yes |
| OS5\_034R | 29529 | 30443 | 304 | hypothetical protein | A088R | NP\_048436.1 | CL0342 |  |
| OS5\_035R | 30553 | 31041 | 162 | hypothetical protein | A088R | NP\_048436.1 | CL0342 |  |
| OS5\_036L | 31123 | 32424 | 433 | PBCV-specific basic adaptor domain-containing protein | A092/093L | NP\_048441.3 | CN0002 |  |
| OS5\_037L | 32462 | 33562 | 366 | Beta 1,3-glucanase | A094L | NP\_048442.1 | CL0520 | yes |
| OS5\_038R | 33716 | 35407 | 563 | Hyaluronan synthase | A098R | NP\_048446.1 | CL0902 | yes |
| OS5\_039R | 35562 | 37349 | 595 | Glucosamine--fructose-6-phosphate aminotransferase | A100R | NP\_048448.2 | CL0534 | yes |
| OS5\_040R | 37518 | 38510 | 330 | mRNA-capping enzyme (mRNA guanylyltransferase) | A103R | NP\_048451.1 | n/d | yes |
| OS5\_041L | 38497 | 39315 | 272 | Ubiquitin carboxyl-terminal hydrolase | A105L | NP\_048453.1 | CL0551 |  |
| OS5\_042L | 39404 | 40276 | 290 | Transcription factor IIB | A107L | NP\_048455.1 | CL0559 |  |
| OS5\_043L | 40378 | 40863 | 161 | GIY-YIG catalytic domain-containing endonuclease | A108bL | NP\_048457.2 | CL0566 |  |
| OS5\_044R | 40988 | 43570 | 860 | Glycosyl transferase | A111/114R | NP\_048459.2 | CL0569 |  |
| OS5\_045R | 43618 | 44658 | 346 | GDP-D-mannose dehydratase | A118R | NP\_048466.1 | CL0577 | yes |
| OS5\_046R | 44678 | 44992 | 104 | Tir Gfp DNA mobile protein | A121R | NP\_048469.2 | CL0584 |  |
| OS5\_047R | 45042 | 48848 | 1268 | hypothetical glycoprotein | A122/123R | NP\_048471.2 | CL0231 |  |
| OS5\_048L | 48850 | 49392 | 180 | Transcription factor IIS | A125L | NP\_048472.1 | CL0603 |  |
| OS5\_049R | 49426 | 50142 | 238 | hypothetical protein | A127R | NP\_048475.1 | CL0611 |  |
| OS5\_050R | 50214 | 51368 | 384 | hypothetical protein | A154L | NP\_048502.1 | n/d |  |
| OS5\_051R | 51388 | 51705 | 105 | hypothetical protein | A130R | NP\_048478.1 | n/d |  |
| OS5\_052L | 51698 | 52105 | 135 | hypothetical protein | A131L | NP\_048479.1 | CL0619 |  |
| OS5\_053L | 52246 | 52764 | 172 | GIY-YIG catalytic domain-containing endonuclease | A134L | NP\_048482.1 | n/d |  |
| OS5\_054R | 52810 | 53250 | 146 | hypothetical protein | A136R | NP\_048484.1 | CL0633 |  |
| OS5\_055R | 53326 | 53532 | 68 | hypothetical protein | A137R | NP\_048485.1 | CL0540 |  |
| OS5\_056R | 53578 | 54384 | 268 | hypothetical protein | A138R | NP\_048486.1 | n/d |  |
| OS5\_057L | 54381 | 54692 | 103 | hypothetical protein | A139L | NP\_048487.2 | CL0641 |  |
| OS5\_058R | 54770 | 58264 | 1164 | hypothetical protein, surface protein? | A140/145R | NP\_048488.3 | CN0001 |  |
| OS5\_059L | 58650 | 59090 | 146 | hypothetical protein | A150L | NP\_048498.1 | CL0659 |  |
| OS5\_060R | 59204 | 60580 | 458 | ATP-dependent RNA helicase | A153R | NP\_048501.1 | CL0676 |  |
| OS5\_061L | 60583 | 61395 | 270 | GIY-YIG catalytic domain-containing endonuclease | A315L | NP\_048671.1 | n/d |  |
| OS5\_062L | 61476 | 61814 | 112 | hypothetical protein | A157L | NP\_048505.1 | CL0686 |  |
| OS5\_063L | 61862 | 62182 | 106 | hypothetical protein | A158L | NP\_048506.1 | CL0688 |  |
| OS5\_064R | 62338 | 62775 | 145 | hypothetical protein | A161R | NP\_048509.1 | CL0786 |  |
| OS5\_065L | 62776 | 63996 | 406 | hypothetical protein, PBCV-1, ?glutamate receptor | A162L | NP\_048510.1 | n/d |  |
| OS5\_066R | 64037 | 65383 | 448 | Ligand-gated ion channel | A163R | NP\_048511.2 | CL0699 |  |
| OS5\_067R | 65408 | 65584 | 58 | hypothetical protein | A164aR | YP\_004678903.1 | CL0708 |  |
| OS5\_068L | 65610 | 65948 | 112 | hypothetical protein | A165L | NP\_048513.2 | CL0712 |  |
| OS5\_069L | 65979 | 66440 | 153 | hypothetical protein | A165aL | YP\_004678904.1 | n/d |  |
| OS5\_070R | 66503 | 67309 | 268 | ygaJ-like virus recombinase | A166R | NP\_048514.2 | CL0724 |  |
| OS5\_071R | 67348 | 67845 | 165 | hypothetical protein | A168R | NP\_048516.1 | CL0733 |  |
| OS5\_072R | 67957 | 68928 | 323 | Aspartate carbamoyltransferase | A169R | NP\_048517.1 | CL0739 | yes |
| OS5\_073R | 68959 | 70134 | 391 | hypothetical protein | A171R | NP\_048519.1 | CL0511 |  |
| OS5\_074L | 70137 | 70964 | 275 | Patatin-like phospholipase | A173L | NP\_048521.1 | CL0752 |  |
| OS5\_075L | 71085 | 71294 | 69 | PBCV-specific basic adaptor domain-containing protein | A176L | NP\_048524.1 | CN0002 |  |
| OS5\_076L | 71315 | 71617 | 100 | PBCV-specific basic adaptor domain-containing protein | A176L | NP\_048524.1 | CN0002 |  |
| OS5\_077R | 72123 | 72860 | 245 | hypothetical protein | A177R | NP\_048525.1 | n/d |  |
| OS5\_078R | 72948 | 73820 | 290 | Zn-finger containing protein? | A478L | NP\_048834.1 | n/d |  |
| OS5\_079L | 73815 | 74099 | 94 | Potassium ion channel protein | A250R | NP\_048599.1 | CN0009 | yes |
| OS5\_080L | 74122 | 74988 | 288 | Calcium/calmodulin dependent protein kinase type 1G | A248R | NP\_048597.1 | n/d | yes |
| OS5\_081R | 75204 | 78062 | 952 | Chlorovirus glycoprotein repeat domain-containing protein | A122/123R | NP\_048471.2 | CL0231 |  |
| OS5\_082R | 78104 | 82468 | 1454 | Chlorovirus glycoprotein repeat domain-containing protein | A122/123R | NP\_048471.2 | CL0231 |  |
| OS5\_083R | 82497 | 86882 | 1461 | Ser/Thr-protein phosphatase, 6 regulatory ankyrin repeats | A025/027/029L | NP\_048355.2 | n/d |  |
| OS5\_084R | 86897 | 91348 | 1483 | Chlorovirus glycoprotein repeat domain-containing protein | A025/027/029L | NP\_048373.2 | n/d |  |
| OS5\_085R | 91445 | 95965 | 1506 | Chlorovirus glycoprotein repeat domain-containing protein | A122/123R | NP\_048471.2 | CL0231 |  |
| OS5\_086R | 96011 | 97246 | 411 | Major capsid protein Vp54 (I) | A430L (weak match) | NP\_048787.1 | n/d |  |
| OS5\_087L | 97259 | 98143 | 294 | Chitin binding domain-containing protein, many PAPK repeats | A246aR | NP\_048594.2 | CN0022 |  |
| OS5\_088L | 98171 | 100348 | 725 | DEAD/DEAH box II RNA helicase | A241R | NP\_048589.1 | CL0122 |  |
| OS5\_089R | 100512 | 100949 | 145 | hypothetical protein | A239L | NP\_048587.2 | CL0113 |  |
| OS5\_090L | 100954 | 102504 | 516 | Homospermidine synthase | A237R | NP\_048585.1 | CL0111 | yes |
| OS5\_091R | 102577 | 102903 | 108 | hypothetical protein | A234L | NP\_048582.1 | CL0105 |  |
| OS5\_092L | 102900 | 103223 | 107 | hypothetical protein | A233R | NP\_048581.1 | CL0103 |  |
| OS5\_093R | 103273 | 104412 | 379 | hypothetical protein | A231L | NP\_048579.1 | CN0012 |  |
| OS5\_094L | 104426 | 105016 | 196 | hypothetical protein | A230R | NP\_048578.1 | CL0085 |  |
| OS5\_095R | 105041 | 105274 | 77 | hypothetical protein | A229L | NP\_048577.1 | n/d |  |
| OS5\_096R | 105296 | 105709 | 137 | hypothetical protein | A227L | NP\_048575.1 | CL0079 |  |
| OS5\_097L | 105718 | 107613 | 631 | Cellulose synthase catalytic subunit (UDP-forming) or N-acetylglucosaminyl transferase | A219/222/226R | NP\_048569.4 | CL0775 |  |
| OS5\_098R | 107723 | 108877 | 384 | Amine oxidase | A217L | NP\_048564.1 | CL0058 |  |
| OS5\_099R | 108897 | 109832 | 311 | Alginate lyase | A215L | NP\_048562.2 | CN0001 | yes |
| OS5\_100L | 110285 | 111085 | 266 | GIY-YIG catalytic domain-containing endonuclease | A495R | NP\_048851.1 | n/d |  |
| OS5\_101R | 111222 | 111641 | 139 | hypothetical protein | A214L | NP\_048561.1 | n/d |  |
| OS5\_102R | 111682 | 112128 | 148 | hypothetical protein | A213L | NP\_048560.1 | CL0030 |  |
| OS5\_103L | 112125 | 112301 | 58 | Hypothetical protein | A212R | NP\_048559.2 | CL0073 |  |
| OS5\_104L | 112325 | 113197 | 290 | hypothetical protein | A208R | NP\_048555.2 | CL0793 |  |
| OS5\_105L | 113325 | 114443 | 372 | Arginine/ornithine decarboxylase | A207R | NP\_048554.1 | CL0013 | yes |
| OS5\_106L | 114506 | 115033 | 175 | PBCV-specific basic adaptor domain-containing protein | A205R | NP\_048552.1 | CL0003 |  |
| OS5\_107L | 115175 | 115819 | 214 | hypothetical protein | A203R | NP\_048550.1 | CL0907 |  |
| OS5\_108R | 115885 | 116226 | 113 | hypothetical protein | A202L | NP\_048549.1 | CL0897 |  |
| OS5\_109R | 116245 | 116529 | 94 | hypothetical protein | A201L | NP\_048548.1 | CL0886 |  |
| OS5\_110L | 116552 | 116908 | 118 | Cytidine and deoxycytidylate deaminase | A200R | NP\_048547.1 | CL0877 |  |
| OS5\_111L | 117010 | 117312 | 100 | hypothetical protein | A199R | NP\_048546.1 | n/d |  |
| OS5\_112R | 117360 | 117818 | 152 | hypothetical protein | A196L | NP\_048543.1 | CL0463 |  |
| OS5\_113R | 117823 | 118629 | 268 | PCNA | A193L | NP\_048540.1 | n/d |  |
| OS5\_114L | 118632 | 122762 | 1376 | SWI/SNF chromatin remodeling complex? | A189/192R | NP\_048536.3 | CN0029 |  |
| OS5\_115L | 122805 | 123501 | 232 | DNA polymerase | A185R | NP\_048532.2 | CN0021 |  |
| OS5\_115L | 123601 | 125644 | 681 | DNA polymerase | A185R | NP\_048532.2 | CN0021 |  |
| OS5\_116R | 125720 | 125947 | 75 | hypothetical protein, no hits to anything | NSSF |  | n/d |  |
| OS5\_117L | 126119 | 127696 | 525 | Chitinase | A181/182R | NP\_048529.2 | CL0887 | yes |
| OS5\_118R | 128104 | 128553 | 149 | Hypothetical protein | A253R | NP\_048604.1 | CL0165 |  |
| OS5\_119R | 128584 | 130101 | 505 | Chitinase | A260R | NP\_048613.3 | n/d | yes |
| OS5\_120R | 130135 | 130683 | 182 | Chloramphenicol acetyltransferase | NC64A\_CviKI\_281R | AGE51474.1 | CL0462 |  |
| OS5\_121L | 130667 | 131422 | 251 | GIY-YIG catalytic domain-containin endonuclease | NC64A\_NY2A\_B389R | YP\_001497585.1 | CN0005 |  |
| OS5\_122L | 131487 | 132137 | 216 | hypothetical protein | A262/263L | NP\_048617.2 | n/d |  |
| OS5\_123L | 132158 | 132907 | 249 | Poxvirus A22-like protein | A265L | NP\_048619.2 | n/d |  |
| OS5\_124L | 132921 | 133157 | 78 | hypothetical protein | A273L | NP\_048627.1 | CN0036 |  |
| OS5\_125L | 133217 | 134389 | 390 | hypothetical protein | unclassified-chlorovirus\_CV-XZ6E | S26850 | n/d |  |
| OS5\_126R | 134530 | 135663 | 377 | DNA adenine methyltransferase | NC64A\_NY-2B\_261L | AGE58281.1 | CN0030 | yes |
| OS5\_127R | 135696 | 137411 | 571 | hypothetical protein, no hits [possible restriction endonuclease?] | NSSF |  | n/d |  |
| OS5\_128R | 137433 | 138548 | 371 | DNA adenine methyltransferase | Pbi\_CVB-1\_012R | AGE50697.1 | CN0030 | yes |
| OS5\_129R | 138579 | 139757 | 392 | DNA adenine methyltransferase | NC64A\_NY2A\_B230L | YP\_001497426.1 | CN0030 | yes |
| OS5\_130R | 139853 | 140869 | 338 | hypothetical protein | A328L | NP\_048684.1 | CL0283 |  |
| OS5\_131R | 140999 | 141577 | 192 | protein of unknown function (DUF1390) | A647R | NP\_049003.1 | n/d |  |
| OS5\_132R | 141778 | 142521 | 247 | hypothetical protein | A275R | NP\_048629.1 | n/d |  |
| OS5\_133L | 142506 | 143345 | 279 | Ser/Thr-protein kinase | A277L | NP\_048631.1 | CN0015 | yes |
| OS5\_134L | 143421 | 145166 | 581 | protein kinase, PBCV-1-specific basic adaptor domain-containing protein | A278L | NP\_048632.2 | CN0002 |  |
| OS5\_135L | 145323 | 146162 | 279 | Amidase | A284L | NP\_048638.1 | CL0240 |  |
| OS5\_136R | 146229 | 147335 | 368 | hypothetical protein | A286R | NP\_048640.1 | CL0250 |  |
| OS5\_137L | 147389 | 148243 | 284 | Ser/Thr protein kinase | A289L | NP\_048643.1 | CL0780 | yes |
| OS5\_138L | 148329 | 149342 | 337 | Chitosanase | A292L | NP\_048646.1 | CL0264 | yes |
| OS5\_139L | 149354 | 150292 | 312 | GDP-L-fucose synthase 2 | A295L | NP\_048649.2 | n/d | yes |
| OS5\_140L | 150323 | 150568 | 81 | No homologs | NSSF |  | n/d |  |
| OS5\_141R | 150567 | 150821 | 84 | hypothetical protein | A296R | NP\_048650.2 | n/d |  |
| OS5\_142L | 150856 | 151806 | 316 | Fibernectin binding protein | NC64A\_CviKI\_028L | AGE51367.1 | CL0531 |  |
| OS5\_143R | 151960 | 152883 | 307 | hypothetical protein | A067R | NP\_048415.1 | CL0530 |  |
| OS5\_144L | 152921 | 153445 | 174 | Phosphoglycerate mutase | A297L | NP\_048651.1 | CL0265 |  |
| OS5\_145R | 153565 | 154026 | 153 | hypothetical protein | NC64A\_NE-JV-4\_441L | AGE57332.1 | CL0235 |  |
| OS5\_146L | 154050 | 154724 | 224 | hypothetical protein | A298L | NP\_048652.1 | CL0266 |  |
| OS5\_147L | 154745 | 155479 | 244 | hypothetical protein | A301L | NP\_048655.1 | n/d |  |
| OS5\_148R | 155534 | 155770 | 78 | hypothetical protein | A304R | NP\_048658.1 | CL0268 |  |
| OS5\_149L | 155811 | 156425 | 204 | Dual specificity protein phosphatase 16 / mitogen | A305L | NP\_048659.1 | CL0269 |  |
| OS5\_150L | 156450 | 156710 | 86 | hypothetical protein | A306L | NP\_048660.1 | CL0488 |  |
| OS5\_151L | 156747 | 157100 | 117 | hypothetical protein | A308L | NP\_048663.2 | CL0270 |  |
| OS5\_152L | 157229 | 157561 | 110 | Fibernectin-binding protein | A180R | NP\_048528.1 | CL0794 |  |
| OS5\_153L | 157624 | 158136 | 170 | hypothetical protein | A310L | NP\_048665.1 | CL0272 |  |
| OS5\_154L | 158204 | 158920 | 238 | hypothetical protein | A312L | NP\_048668.2 | n/d |  |
| OS5\_155L | 159128 | 159343 | 71 | hypothetical protein | A313L | NP\_048669.1 | CL0553 |  |
| OS5\_156R | 159423 | 159665 | 80 | hypothetical protein | A314R | NP\_048670.1 | CL0554 |  |
| OS5\_157R | 159689 | 161221 | 510 | hypothetical protein | A316R | NP\_048672.2 | n/d |  |
| OS5\_158R | 161279 | 161644 | 121 | hypothetical protein | A320R | NP\_048676.1 | CL0277 |  |
| OS5\_159R | 161681 | 162040 | 119 | hypothetical protein | A321R | NP\_048677.1 | CL0278 |  |
| OS5\_160L | 162151 | 162681 | 176 | hypothetical protein | A322L | NP\_048678.1 | CL0280 |  |
| OS5\_161L | 162733 | 164067 | 444 | hypothetical protein | A324L | NP\_048680.1 | CL0281 |  |
| OS5\_162L | 164148 | 164786 | 212 | hypothetical protein | A326L | NP\_048682.1 | CL0282 |  |
| OS5\_163L | 164817 | 165884 | 355 | hypothetical protein | A328L | NP\_048684.1 | n/d |  |
| OS5\_164R | 165974 | 166264 | 96 | hypothetical protein | A329R | NP\_048685.1 | CL0284 |  |
| OS5\_165R | 167717 | 168559 | 280 | hypothetical protein | A267L | NP\_048621.1 | CL0490 |  |
| OS5\_166R | 169037 | 169216 | 59 | hypothetical protein | a329cR | YP\_004678937.1 | n/d |  |
| OS5\_167L | 169229 | 170410 | 393 | hypothetical protein | A333L | NP\_048688.2 | n/d |  |
| OS5\_168L | 170451 | 171251 | 266 | hypothetical protein | A337L | NP\_048693.2 | CL0291 |  |
| OS5\_169L | 171389 | 171940 | 183 | hypothetical protein | A337L | NP\_048693.2 | n/d |  |
| OS5\_170L | 172002 | 172409 | 135 | hypothetical protein | A341L | NP\_048698.2 | CL0292 |  |
| OS5\_171L | 172496 | 173350 | 284 | GIY-YIG catalytic domain-containing endonuclease | A315L | NP\_048671.1 | n/d |  |
| OS5\_172L | 173430 | 175124 | 564 | hypothetical protein | A342L | NP\_048699.1 | CL0293 |  |
| OS5\_173L | 175222 | 175533 | 103 | hypothetical protein | A349L | NP\_048706.2 | CL0294 |  |
| OS5\_174R | 175644 | 176012 | 122 | hypothetical protein | A350R | NP\_048707.1 | CL0296 |  |
| OS5\_175L | 176126 | 176749 | 207 | hypothetical protein | A352L | NP\_048709.1 | CL0297 |  |
| OS5\_176L | 176813 | 177826 | 337 | hypothetical protein | A357L | NP\_048714.2 | n/d |  |
| OS5\_177R | 177896 | 181510 | 1204 | Superfamily II helicase | A363R | NP\_048717.2 | CN0016 |  |
| OS5\_178R | 181634 | 182059 | 141 | hypothetical protein | A373R | NP\_048730.1 | CL0301 |  |
| OS5\_179R | 182179 | 183123 | 314 | hypothetical protein | A007/008L | NP\_048352.1 | CN0006 |  |
| OS5\_180L | 183237 | 183962 | 241 | hypothetical protein | A378L | NP\_048735.2 | CN0001 |  |
| OS5\_181L | 183986 | 184615 | 209 | hypothetical protein | A379L | NP\_048736.1 | CL0304 |  |
| OS5\_182R | 184788 | 186248 | 486 | Major capsid protein (IV) | A383R | NP\_048739.2 | CN0004 |  |
| OS5\_183R | 186272 | 186457 | 61 | hypothetical protein | A384bL | YP\_004678944.1 | CL0307 |  |
| OS5\_184L | 186540 | 188390 | 616 | Chitin binding domain-containing protein, Capsid protein (V) | A384dL | NP\_048747.2 | n/d |  |
| OS5\_185R | 188481 | 189245 | 254 | DNA packaging ATPase | A392R | NP\_048749.2 | n/d |  |
| OS5\_186R | 189281 | 189490 | 69 | Hypothetical protein | NC64A\_NY2A\_B537R | YP\_001497733.1 | n/d |  |
| OS5\_187R | 189502 | 189906 | 134 | hypothetical protein | A394R | NP\_048751.1 | CL0310 |  |
| OS5\_188L | 189992 | 191476 | 494 | hypothetical protein, no hits with anything | NSSF |  | n/d |  |
| OS5\_189R | 191560 | 191811 | 83 | hypothetical protein | A395R | NP\_048752.1 | n/d |  |
| OS5\_190L | 191964 | 192422 | 152 | hypothetical protein | A396L | NP\_048753.2 | CL0312 |  |
| OS5\_191L | 192477 | 192833 | 118 | hypothetical protein | A398L | NP\_048755.1 | CL0314 |  |
| OS5\_192R | 192906 | 193490 | 194 | RNase H | A399R | NP\_048756.1 | CL0315 |  |
| OS5\_193R | 193523 | 193876 | 117 | hypothetical protein | A400R | NP\_048757.1 | n/d |  |
| OS5\_194R | 193914 | 194768 | 284 | hypothetical protein | A401R | NP\_048758.1 | CL0317 |  |
| OS5\_195R | 194788 | 195489 | 233 | Lipoprotein lipase? | A402R | NP\_048759.1 | CL0318 |  |
| OS5\_196R | 195608 | 195889 | 93 | hypothetical protein | A403R | NP\_048760.1 | CL0319 |  |
| OS5\_197R | 195920 | 196507 | 195 | hypothetical protein | A404R | NP\_048761.2 | CL0320 |  |
| OS5\_198L | 196498 | 196656 | 52 | hypothetical protein | A404aL | YP\_004678950.1 | n/d |  |
| OS5\_199R | 196715 | 198421 | 568 | hypothetical protein, Pro rich protein | A405R | NP\_048762.1 | CL0822 |  |
| OS5\_200L | 198424 | 199056 | 210 | hypothetical protein | A407L | NP\_048764.1 | CL0322 |  |
| OS5\_201L | 199094 | 199861 | 255 | hypothetical protein | A408L | NP\_048765.1 | n/d |  |
| OS5\_202L | 199867 | 200196 | 109 | hypothetical protein | A410L | NP\_048767.1 | n/d |  |
| OS5\_203R | 200284 | 200790 | 168 | hypothetical protein | A411R | NP\_048768.1 | n/d |  |
| OS5\_204R | 200799 | 201365 | 188 | hypothetical protein | A412R | NP\_048769.1 | CL0326 |  |
| OS5\_205L | 201366 | 202070 | 234 | hypothetical protein | A413L | NP\_048770.2 | CL0327 |  |
| OS5\_206R | 202149 | 202367 | 72 | hypothetical protein | A414R | NP\_048771.1 | CL0501 |  |
| OS5\_207R | 202443 | 203006 | 187 | Deoxynucleoside kinase | A416R | NP\_048773.1 | CL0427 |  |
| OS5\_208L | 202982 | 204271 | 429 | Replication factor C | A417L | NP\_048774.1 | CL0426 |  |
| OS5\_209L | 204303 | 204515 | 70 | hypothetical protein | A420L | NP\_048777.1 | CL0496 |  |
| OS5\_210R | 204560 | 204859 | 99 | hypothetical protein | A421R | NP\_048778.1 | CL0425 |  |
| OS5\_211R | 204883 | 205077 | 64 | hypothetical protein | A422aR | YP\_004678951.1 | CL0498 |  |
| OS5\_212R | 205088 | 205576 | 162 | hypothetical protein | A423R | NP\_048780.1 | CL0424 |  |
| OS5\_213R | 205604 | 205987 | 127 | hypothetical protein, no hits | NSSF |  | n/d |  |
| OS5\_214R | 206022 | 206366 | 114 | hypothetical protein | A426R | NP\_048783.1 | CL0423 |  |
| OS5\_215L | 206363 | 206704 | 113 | Thioredoxin | A427L | NP\_048784.1 | CN0033 |  |
| OS5\_216L | 206778 | 207182 | 134 | Choline-binding protein A | A428L | NP\_048785.1 | n/d |  |
| OS5\_217L | 207208 | 208569 | 453 | Ankyrin repeat-containing protein | A429L | NP\_048786.2 | CL0421 |  |
| OS5\_218R | 208720 | 209466 | 248 | GIY-YIG catalytic domain-containing endonuclease | A315L | NP\_048671.1 | n/d |  |
| OS5\_219L | 209523 | 210836 | 437 | Major capsid protein VP54 (I) | A430L (strong match) | NP\_048787.1 | CN0004 |  |
| OS5\_220L | 210916 | 211734 | 272 | GIY-YIG catalytic domain-containing endonuclease | A315L | NP\_048671.1 | n/d |  |
| OS5\_221R | 211852 | 212298 | 148 | hypothetical protein | A432R | NP\_048789.2 | CL0418 |  |
| OS5\_222L | 212295 | 212456 | 53 | PBCV-specific basic adaptor domain-containing protein | A436L | NP\_048793.2 | CL0349 |  |
| OS5\_223L | 212524 | 212850 | 108 | Chromosomal protein MC1 | A437L | NP\_048794.2 | CL0417 |  |
| OS5\_224L | 212879 | 213115 | 78 | Glutaredoxin | A438L | NP\_048795.1 | CN0032 | yes |
| OS5\_225R | 213138 | 213476 | 112 | hypothetical protein | A439R | NP\_048796.1 | n/d |  |
| OS5\_226R | 213483 | 213644 | 53 | hypothetical protein | A439aR | YP\_004678955.1 | n/d |  |
| OS5\_227L | 213633 | 214046 | 137 | hypothetical protein | A441L | NP\_048798.1 | CL0414 |  |
| OS5\_228R | 214187 | 215113 | 308 | hypothetical protein | A443R | NP\_048800.1 | CL0412 |  |
| OS5\_229L | 215128 | 215790 | 220 | SAM-dependent methyltransferase | NC64A\_CviKI\_481L | AGE51569.1 | CL0065 |  |
| OS5\_230L | 215845 | 216132 | 95 | hypothetical protein | A444L | NP\_048801.1 | CL0411 |  |
| OS5\_231L | 216225 | 217613 | 462 | 2-poly prenylphenol 6-hydrolase domain, protein kinase, glycosyltransferase [A445L?], ABC transporter | A445L | NP\_048802.2 | CL0410 |  |
| OS5\_232L | 217674 | 217994 | 106 | protein disulfide isomerase | A448L | NP\_048805.1 | CL0409 |  |
| OS5\_233R | 218144 | 218728 | 194 | mRNA capping enzyme beta chain, RNA triphosphatase | A449R | NP\_048806.1 | CL0407 | yes |
| OS5\_234R | 218962 | 219720 | 252 | hypothetical protein | A450R | NP\_048807.1 |  |  |
| OS5\_235L | 219757 | 220581 | 274 | Lysophospholipase, AB abhydrolase | A271L | NP\_048625.2 | CL0192 |  |
| OS5\_236L | 220697 | 221110 | 137 | hypothetical protein (possible intein) | A452L | NP\_048809.2 | CL0406 |  |
| OS5\_237L | 221278 | 222108 | 288 | hypothetical protein | A454L | NP\_048811.1 | CL0405 |  |
| OS5\_238L | 222175 | 224136 | 653 | Phage/plasmid primase, ATPase, superfamily III helicase | A456L | NP\_048813.1 | CN0026 |  |
| OS5\_239R | 224270 | 224455 | 61 | hypothetical protein | A461R | YP\_004678958.1 | CL0067 |  |
| OS5\_240R | 224489 | 225301 | 270 | RNase III | A464R | NP\_048820.1 | CL0402 | yes |
| OS5\_241R | 225339 | 225695 | 118 | Thiol oxidoreductase | A465R | NP\_048821.1 | CN0028 |  |
| OS5\_242L | 225720 | 226658 | 312 | hypothetical protein | A467L | NP\_048823.1 | CL0399 |  |
| OS5\_243R | 226807 | 228132 | 441 | DNA primase | A468R | NP\_048824.1 | CL0398 |  |
| OS5\_244R | 228196 | 228771 | 191 | hypothetical protein | A470R | NP\_048826.1 | CL0397 |  |
| OS5\_245R | 228823 | 229344 | 173 | hypothetical protein | A471R | NP\_048827.1 | CL0396 |  |
| OS5\_246R | 229482 | 230456 | 324 | Ribonucleoside-diphosphate reductase small subunit | A476R | NP\_048832.1 | n/d |  |
| OS5\_247L | 230437 | 231321 | 294 | hypothetical protein | A490L | NP\_048846.1 | n/d |  |
| OS5\_248L | 231357 | 231632 | 91 | hypothetical protein | A480L | NP\_048836.1 | CL0394 |  |
| OS5\_249L | 231660 | 232343 | 227 | Ring finger ubiquitin ligase | A481L | NP\_048837.1 | CL0271 |  |
| OS5\_250R | 232422 | 233066 | 214 | MYM-type zinc finger with FCS sequencing motif, VLTF2 type transcription factor | A482R | NP\_048838.1 | CL0393 |  |
| OS5\_251L | 233061 | 233528 | 155 | hypothetical protein | A484L | NP\_048840.1 | CL0392 |  |
| OS5\_252R | 233611 | 234054 | 147 | hypothetical protein | A485R | NP\_048841.1 | n/d |  |
| OS5\_253R | 234068 | 234334 | 88 | hypothetical protein | Pbi\_Can18-4\_476L | AGE49834.1 | CL0494 |  |
| OS5\_254R | 234378 | 235337 | 319 | hypothetical protein | A488R | NP\_048844.1 | CL0390 |  |
| OS5\_255R | 235387 | 235617 | 76 | hypothetical protein | A491R | NP\_048847.1 | CL0389 |  |
| OS5\_256L | 235614 | 236159 | 181 | hypothetical protein | A492L | NP\_048848.1 | n/d |  |
| OS5\_257R | 236201 | 237295 | 364 | Pox VLTF3 late transcription factor | A494R | NP\_048850.1 | CL0386 |  |
| OS5\_258R | 237349 | 237789 | 146 | hypothetical protein | A497R | NP\_048853.1 | CL0385 |  |
| OS5\_259L | 237839 | 238891 | 350 | hypothetical protein, lot of PAPKs | A500L | NP\_048856.2 | CL0383 |  |
| OS5\_260L | 238925 | 239212 | 95 | hypothetical protein | A502L | NP\_048858.1 | CL0381 |  |
| OS5\_261L | 239258 | 240097 | 279 | hypothetical protein | A503L | NP\_048859.1 | CL0380 |  |
| OS5\_262L | 240176 | 241675 | 499 | hypothetical protein | A505L | NP\_048861.1 | CL0379 |  |
| OS5\_263R | 241859 | 242422 | 187 | hypothetical protein | Pbi\_MT325\_m295L | ABT13849.1 | CL0485 |  |
| OS5\_264L | 242444 | 242692 | 82 | hypothetical protein | A519L | NP\_048875.1 | CL0377 |  |
| OS5\_265L | 242697 | 242996 | 99 | hypothetical protein | A520L | NP\_048876.2 | CL0376 |  |
| OS5\_266L | 243014 | 243562 | 182 | Zn metallopeptidase | A521L | NP\_048877.2 | CL0374 |  |
| OS5\_267L | 243591 | 244208 | 205 | Putative metallopeptidase WLM | A521aL | YP\_004678965.1 | n/d |  |
| OS5\_268R | 244263 | 244796 | 177 | hypothetical protein | A523R | NP\_048879.1 | CL0372 |  |
| OS5\_269R | 244839 | 245279 | 146 | hypothetical protein | A526R | NP\_048882.2 | CL0371 |  |
| OS5\_270R | 245257 | 245571 | 104 | hypothetical protein | A527R | NP\_048883.2 | CL0370 |  |
| OS5\_271R | 245686 | 245958 | 90 | Hypothetical protein | a528R | NP\_048884.1 | n/d |  |
| OS5\_272R | 245922 | 246974 | 350 | DNA cytosine-specific methyltransferase | A530R | NP\_048886.2 | n/d | yes |
| OS5\_273L | 246971 | 247165 | 64 | hypothetical protein | A531L | NP\_048887.1 | CL0777 |  |
| OS5\_274L | 247197 | 247436 | 79 | hypothetical protein | A532L | NP\_048888.1 | CL0369 |  |
| OS5\_275L | 247463 | 247615 | 50 | hypothetical protein | A532aL | YP\_004678967.1 | n/d |  |
| OS5\_276R | 247715 | 249310 | 531 | hypothetical protein | A533R | NP\_048889.1 | CL0368 |  |
| OS5\_277L | 249312 | 249536 | 74 | hypothetical protein | A535L | NP\_048891.1 | CL0367 |  |
| OS5\_278L | 249602 | 249856 | 84 | hypothetical protein | A536L | NP\_048892.1 | CL0366 |  |
| OS5\_279L | 249861 | 250676 | 271 | hypothetical protein | A537L | NP\_048893.1 | CL0364 |  |
| OS5\_280R | 250760 | 251305 | 181 | GIY-YIG catalytic domain-containing endonuclease | A539R | NP\_048895.1 | CL0363 |  |
| OS5\_281L | 251321 | 254722 | 1133 | Chlorovirus glycoprotein repeat domain-containing, virus tail fiber | A540L | NP\_048896.2 | CN0011 |  |
| OS5\_282R | 254843 | 255739 | 298 | ATP-dependent DNA ligase | A544R | NP\_048900.1 | CL0362 |  |
| OS5\_283L | 255721 | 256950 | 409 | glycosyl transferase | A546L | NP\_048902.4 | CL0279 |  |
| OS5\_284L | 256937 | 258307 | 456 | ATP-dependent SNF2 helicase | A548L | NP\_048904.2 | CL0361 |  |
| OS5\_285L | 258399 | 258836 | 145 | dUTP diphosphatase | A551L | NP\_048907.1 | CL0360 | yes |
| OS5\_286R | 258963 | 259916 | 317 | transcription factor TFIID | A552R | NP\_048908.3 | CL0359 |  |
| OS5\_287L | 259931 | 261427 | 498 | tRNA(Ile)-lysidine synthase | A554/556/557L | NP\_048910.2 | CL0357 |  |
| OS5\_288L | 261527 | 262729 | 400 | Capsid protein (III) | A558L | NP\_048914.1 | CN0004 |  |
| OS5\_289L | 262842 | 263513 | 223 | hypothetical protein | A559L | NP\_048915.1 | CL0354 |  |
| OS5\_290L | 263551 | 265575 | 674 | DNA binding protein, lysine activity | A561L | NP\_048917.1 | CN0001 |  |
| OS5\_291R | 265665 | 266477 | 270 | GIY-YIG catalytic containing endonuclease | A287R | NP\_048641.2 | n/d |  |
| OS5\_292R | 266499 | 268562 | 687 | hypothetical protein | A565R | NP\_048921.2 | CN0001 |  |
| OS5\_293L | 268575 | 269000 | 141 | hypothetical protein | A567L | NP\_048923.1 | n/d |  |
| OS5\_294L | 269023 | 269577 | 184 | hypothetical protein | A568L | NP\_048924.1 | CL0350 |  |
| OS5\_295L | 269601 | 269813 | 70 | hypothetical protein | A570L | NP\_048926.1 | n/d |  |
| OS5\_296R | 270059 | 270409 | 116 | PBCV-specific basic adaptor domain-containing protein | A571R | NP\_048927.1 | n/d |  |
| OS5\_297R | 270423 | 270968 | 181 | hypothetical protein | A572R | NP\_048928.1 | CL0348 |  |
| OS5\_298L | 270974 | 271720 | 248 | PCNA | A574L | NP\_048930.1 | CN0019 |  |
| OS5\_299L | 271795 | 272301 | 168 | hypothetical protein | A575L | NP\_048931.1 | CN0020 |  |
| OS5\_300L | 272408 | 272815 | 135 | hypothetical protein | A577L | NP\_048933.2 | n/d |  |
| OS5\_301L | 272831 | 276016 | 1061 | DNA topoisomerase II | A583L | NP\_048939.2 | CL0343 | yes |
| OS5\_302L | 276146 | 277255 | 369 | hypothetical protein | A590L | NP\_048946.2 | CL0917 |  |
| OS5\_303R | 277480 | 278118 | 212 | hypothetical protein | NC64A\_KS1B\_271R | AGE54561.1 | CL0500 |  |
| OS5\_304R | 278391 | 279254 | 287 | hypothetical protein | A593R | NP\_048949.2 | CL0341 |  |
| OS5\_305R | 279287 | 279715 | 142 | dCMP deaminase | A596R | NP\_048952.1 | CL0340 | yes |
| OS5\_306R | 279802 | 280368 | 188 | hypothetical protein | NC64A\_NY2A\_B871R | YP\_001498067.1 | CL0455 |  |
| OS5\_307L | 280363 | 281517 | 384 | Histidine decarboxylase | A598L | NP\_048954.1 | CL0339 |  |
| OS5\_308R | 281594 | 282511 | 305 | ADP-ribosylglycohydrolase | NSSF |  | n/d |  |
| OS5\_309R | 282565 | 282864 | 99 | hypothetical protein | A601R | NP\_048957.1 | CL0338 |  |
| OS5\_310L | 282854 | 283105 | 83 | Ubiquitin, It is probably 3 amino acids too long | NSSF |  | n/d |  |
| OS5\_311L | 283120 | 283542 | 140 | hypothetical protein | A602L | NP\_048958.2 | CL0337 |  |
| OS5\_312R | 283647 | 283961 | 104 | hypothetical protein | a603R | NP\_048959.1 | CL0336 |  |
| OS5\_313R | 284002 | 284145 | 47 | hypothetical protein | a603bR | YP\_004678984.1 | CL0069 |  |
| OS5\_314L | 284149 | 284640 | 163 | Zn metallopeptidase? | A604L | NP\_048960.1 | CL0335 |  |
| OS5\_315L | 284651 | 285127 | 158 | hypothetical protein | A605L | NP\_048961.1 | CL0333 |  |
| OS5\_316R | 285187 | 286548 | 453 | Cell envelope protein – Rhizobium | Pbi\_NE-JV-1\_839L | AGE56446.1 | CL0034 |  |
| OS5\_317R | 286578 | 287750 | 390 | Ankyrin repeat-containing protein | A607R | NP\_048963.2 | n/d |  |
| OS5\_318L | 287759 | 288928 | 389 | UDP-glucose dehydrogenase | A609L | NP\_048965.1 | CL0287 | yes |
| OS5\_319L | 289010 | 289369 | 119 | Histone H3K27 methylase | A612L | NP\_048968.1 | CL0331 | yes |
| OS5\_320L | 289423 | 291135 | 570 | Ser/Thr-protein kinase? | A614L | NP\_048970.2 | CN0002 | yes |
| OS5\_321R | 291206 | 292177 | 323 | Ser/Thr protein kinase | A617R | NP\_048973.2 | n/d | yes |
| OS5\_322L | 292192 | 292602 | 136 | hypothetical protein | A618L | NP\_048974.1 | CL0328 |  |
| OS5\_323L | 292619 | 293305 | 228 | Transposase | A619L | NP\_048975.1 | CL0382 |  |
| OS5\_324L | 293345 | 293596 | 83 | hypothetical protein | A620L | NP\_048976.1 | n/d |  |
| OS5\_325L | 293616 | 293972 | 118 | hypothetical protein | A621L | NP\_048977.2 | CL0432 |  |
| OS5\_326L | 294030 | 295607 | 525 | capsid protein (II) | A622L | NP\_048978.1 | CN0004 |  |
| OS5\_327R | 295894 | 296262 | 122 | Putative membrane protein | A624R | NP\_048980.1 | CL0428 |  |
| OS5\_328R | 296281 | 297597 | 438 | hypothetical protein | A627R | NP\_048983.1 | CN0012 |  |
| OS5\_329L | 297616 | 297939 | 107 | hypothetical protein | A628L | NP\_048984.1 | n/d |  |
| OS5\_330R | 298102 | 300405 | 767 | Ribonucleotide reductase large subunit | A629R | NP\_048985.1 | CL0434 |  |
| OS5\_331R | 300485 | 300808 | 107 | hypothetical protein | A633R | NP\_048989.1 | CL0435 |  |
| OS5\_332L | 300809 | 301228 | 139 | hypothetical protein | A634L | NP\_048990.1 | CL0436 |  |
| OS5\_333R | 301294 | 302916 | 540 | CTP synthetase | NSSF |  | n/d |  |
| OS5\_334R | 303079 | 303342 | 87 | hypothetical protein | A635R | NP\_048991.1 | n/d |  |
| OS5\_335R | 303412 | 303837 | 141 | hypothetical protein | A637R | NP\_048993.1 | CL0439 |  |
| OS5\_336R | 303931 | 305247 | 438 | hypothetical protein | A643R | NP\_048999.2 | CL0442 |  |
| OS5\_337R | 305286 | 305801 | 171 | hypothetical protein | A644R | NP\_049000.2 | CL0443 |  |
| OS5\_338R | 305895 | 306272 | 125 | hypothetical protein | A645R | NP\_049001.1 | n/d |  |
| OS5\_339R | 306560 | 307333 | 257 | protein of unknown function (DUF1390) | A649R | NP\_049005.1 | n/d |  |
| OS5\_340L | 307340 | 307792 | 150 | Polyamine acetytransferase | A654L | NP\_049010.1 | CL0445 | yes |
| OS5\_341L | 308001 | 308648 | 215 | hypothetical protein | A656L | NP\_049012.2 | n/d |  |
| OS5\_342L | 308835 | 309407 | 190 | hypothetical protein | A659L | NP\_049015.2 | n/d |  |
| OS5\_343L | 309431 | 309946 | 171 | Mpv17/PMP22 family protein, mito & peroxisome membrane | A662L | NP\_049018.1 | CL0447 |  |
| OS5\_344L | 310032 | 310493 | 153 | hypothetical protein | A664L | NP\_049020.2 | CL0449 |  |
| OS5\_345L | 310634 | 311140 | 168 | hypothetical protein | A348R | NP\_048705.1 | CL0556 |  |
| OS5\_346L | 311347 | 311814 | 155 | hypothetical protein | A665L | NP\_049021.2 | CL0449 |  |
| OS5\_347L | 311849 | 314461 | 870 | Calcium-transporting ATPase | NC64A\_KS1B\_652L | AGE54749.1 | CL0504 | yes |
| OS5\_348R | 314621 | 315064 | 147 | hypothetical protein | NSSF |  | n/d |  |
| OS5\_349L | 315072 | 315335 | 87 | hypothetical protein | NC64A\_KS1B\_658L | AGE54750.1 | CL0953 |  |
| OS5\_350L | 315381 | 318224 | 947 | Translation elongation factor 3 | A666L | NP\_049022.1 | CL0450 |  |
| OS5\_351R | 318294 | 318917 | 207 | Ankyrin repeat PH and SEC7 domain containing protein | A672R | NP\_049028.1 | n/d |  |
| OS5\_352R | 319153 | 319803 | 216 | Thymidylate synthase X | A674R | NP\_049030.1 | CL0453 | yes |
| OS5\_353R | 319842 | 320954 | 370 | PBCV-specific basic adaptor domain-containing protein | A676R | NP\_049032.2 | CN0002 |  |
| OS5\_354R | 321051 | 323288 | 745 | Ser/Thr protein phosphatase 6 regulatory ankyrin repeat subunit B like | A330R | NP\_048686.1 | n/d |  |
| OS5\_355R | 323337 | 324515 | 392 | Alpha-L-arabinofuranosidase | SAG3.83\_ATCV1\_Z832L | YP\_001427313.1 | CL0539 |  |
| OS5\_356R | 325064 | 325249 | 61 | hypothetical protein | a004L | NP\_048352.1 | CL0757 |  |
| OS5\_357R | 326103 | 326669 | 188 | hypothetical protein | SAG3.83\_TN603.4.2\_002L | AGE59612.1 | CL0468 |  |

NSSF = no significant sequence found in BLASTp search

n/d = none determined

#Recombinant proteins have been produced from Chlorovirus PBCV-1 homologs of these CDSs and shown to have the expected activity.

**Sup. Table. 2. OSy-NE5 encoded tRNAs**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| tRNA | Anticodon | tRNA#a | Begin | End |
| Leu | TAA | 1 | 166543 | 166625 |
| Leu | TAA | 2 | 166734 | 166816 |
| Leu | TAA | 3 | 166925 | 167007 |
| Asn | GTT | 4 | 167117 | 167190 |
| Gly | TCC | 5 | 167194 | 167264 |
| Asn | GTT | 6 | 167288 | 167361 |
| Lys | CTT | 7 | 167365 | 167437 |
| Arg | TCT | 8 | 167465 | 167536 |
| Arg | TCT | 9 | 167564 | 167635 |
| Tyrb | GTA | 10 | 168635 | 168719 |
| Asp | GTC | 11 | 168747 | 168818 |
| Leu | CAA | 12 | 168820 | 168900 |
| Thr | TGT | 13 | 168904 | 168975 |

a Order of the tRNA in the OSy-NE5 genome.

b tRNA contains a predicted intron.

**Sup. Table. 3. Genes used in the color-coded map in Figure 6.**

**CDS no. Size annotation**

**A). DNA replication, recombination and repair**

OS5\_115L 913 DNA polymerase –681+232 aa

OS5\_243R 441 DNA primase

OS5\_282R 298 ATP-dependent DNA ligase

OS5\_208L 429 Replication factor C

OS5\_113R 262 PCNA-sliding clamp

OS5\_298L 248 PCNA-sliding clamp

OS5\_301L 1061 DNA topoisomerase II

OS5\_192R 194 RNase H

OS5\_185R 254 ATPase (DNA packaging)

OS5\_012L 141 Pyrimidine dimer-specific glycosylase

OS5\_009L 611 AAA ATPase

OS5\_238L 653 Phage/plasmid primase

OS5\_019R 247 Methyltransferase FkbM family

OS5\_005L 568 DNA binding protein

OS5\_290L 674 DNA binding protein

**B). Transcription**

OS5\_040R 330 mRNA-capping enzyme guanylyltransferase

OS5\_233R 194 RNA triphosphatase

OS5\_042L 290 transcription factor TFIIB

OS5\_286R 317 transcription factor TFIID

OS5\_048L 180 transcription factor TFIIS

OS5\_250R 214 VLTF2-type transcription factor

OS5\_257R 364 Pox\_VLTF3 superfamily transcription factor

OS5\_284L 456 ATP-dependent SNF 2 helicase

OS5\_060R 464 ATP dependent RNA helicase

OS5\_088L 725 DEAE/DEAH box II RNA helicase

OS5\_177R 1204 Superfamily II helicase

OS5\_114L 1376 SWI/SNF chromatin remodeling complex

OS5\_110L 118 Cytidine and deoxycytidylate deaminase

OS5\_240R 270 RNase III

OS5\_319L 119 Histone H3K27 methylase

OS5\_287L 498 tRNA (ile)-lysine synthase

**C). Protein synthesis, modification and degradation**

OS5\_007L 148 SKP1-like protein

OS5\_033R 226 prolyl 4-hydroxylase

OS5\_041L 272 ubiquitin carboxyl-terminal hydrolase

OS5\_232L 106 Protein disulfide isomerase

OS5\_310L 81 ubiquitin - present in 4 other CVs

OS5\_350L 947 Translation elongation factor 3

OS5\_241R 118 Thiol oxidoreductase

OS5\_249L 227 Ring finger ubiquitin ligase

OS5\_266L 182 Zn metallopeptidase

OS5\_314L 163 Zn metallopeptidase

**D). DNA modification**

OS5\_126R 377 Adenine-specific methyltransferase

OS5\_128R 371 Adenine-specific methyltransferase

OS5\_129R 392 Adenine-specific methyltransferase

OS5\_272R 350 cytosine-specific methyltransferase

**E). Integration and transposition**

OS5\_043L 161 GIY-YIG catalytic domain-containing endonuclease

OS5\_053L 164 GIY-YIG catalytic domain-containing endonuclease

OS5\_061L 270 GIY-YIG catalytic domain-containing endonuclease

OS5\_100L 266 GIY-YIG catalytic domain-containing endonuclease

OS5\_121L 251 GIY-YIG catalytic domain-containing endonuclease

OS5\_171L 284 GIY-YIG catalytic domain-containing endonuclease

OS5\_218R 248 GIY-YIG catalytic domain-containing endonuclease

OS5\_220L 272 GIY-YIG catalytic domain-containing endonuclease

OS5\_280R 202 GIY-YIG catalytic domain-containing endonuclease

OS5\_291R 270 GIY-YIG catalytic domain-containing endonuclease

OS5\_046R 104 Tir 6Fp DNA mobile protein

OS5\_323L 228 transposase

**F). Lipid manipulation**

OS5\_235L 264 Lysophospholipase

OS5\_074L 275 Patatin-like phospholipase

OS5\_011L 209 glycerophosphoryl diestrase phosphodiesterase

OS5\_144L 174 phosphoglycerate mutase

OS5\_195R 227 Lipoprotein lipase

**G). Nucleotide metabolism**

OS5\_072R 323 Aspartate transcarbamylase

OS5\_207R 187 Deoxynucleoside kinase

OS5\_215L 113 thioredoxin

OS5\_224L 78 Glutaredoxin

OS5\_246R 324 ribonucleoside reductase small subunit

OS5\_330R 767 Ribonucleotide reductase large subunit

OS5\_285L 141 dUTP pyrophosphatase

OS5\_305R 142 dCMP deaminase

OS5\_333R 540 CTP synthetase

OS5\_352R 216 Thymidylate synthase X

**H). Sugar metabolism and manipulation**

OS5\_038R 563 hyaluronan synthase

OS5\_039R 595 Glucosamine synthetase

OS5\_318L 389 UDP-glucose dehydrogenase

OS5\_045R 346 GDP-D-mannose dehydratase

OS5\_139L 312 GDP-L-fucose synthase 2

OS5\_017L 333 Glycosyltransferase domain

OS5\_044R 860 Glycosyltransferase

OS5\_097L 631 Cellulose synthase catalytic subunit

OS5\_283L 395 glycosyl transferase A546L homolog

OS5\_025L 279 exostosin

OS5\_015R 362 D-lactate dehydrogenase

**I). Cell wall degradation**

OS5\_037L 366 beta 1,3 glucanase

OS5\_117L 525 Bifunctional chitinase/lysozyme

OS5\_119R 505 Chitinase

OS5\_138L 337 Chitosanase

OS5\_099R 311 Alginate lyase

OS5\_290L 674 Lysin

**J). Signaling**

OS5\_004R 308 Ser/Thr protein kinase

OS5\_014L 269 aquaglyceroporin

OS5\_066R 448 Ligand-gated ion channel

OS5\_079L 94 Potassium ion channel protein

OS5\_149L 204 Dual specific phosphatase

OS5\_080L 288 Calcium/calmodulin dependent protein kinase

OS5\_133L 279 Ser/Thr protein kinase

OS5\_134L 614 Ser/Thr protein kinase

OS5\_137L 284 Ser /Thr protein kinase

OS5\_179R 314 Ser/Thr protein phosphatase 6 Has 5 ank repeats

OS5\_231L 462 Protein kinase domain

OS5\_320L 570 Ser/Thr protein kinase

OS5\_321R 323 Ser/Thr protein kinase

**K). Miscellaneous**

OS5\_029R 298 N-carbamoylputrescine amidohydrolase

OS5\_340L 197 Polyamine acetyltransferase

OS5\_090L 516 Homospermidine synthase

OS5\_105L 372 Arginine/ornithine decarboxylase

OS5\_098R 384 Amino oxidase

OS5\_120R 182 Chloramphenicol acetyltransferase

OS5\_135L 279 amidase

OS5\_307L 365 histidine decarboxylase

OS5\_142L 316 Fibronectin binding protein

**L). Structural proteins**

OS5\_086R 411 Major capsid protein

OS5\_182R 486 Major capsid protein

OS5\_219L 437 Major capsid protein Vp54

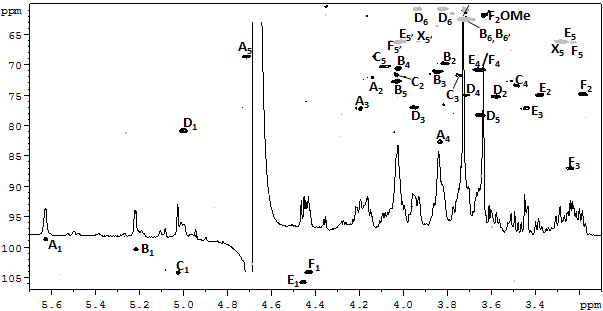
OS5\_288L 400 Major capsid protein

OS5\_326L 525 Major capsid protein

**Sup. Table. 4.** Proton and carbon chemical shifts (600 MHz) recorded in deuterated water at 310 K with acetone as internal standard. **F** is substituted at position *O*-3 with a methyl group (chemical shift 1H/13C: 3.64/61.6 ppm).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **1** | **2** | **3** | **4** | **5 (5eq- 5ax)** | **6-6'** |
| **A** | 1H | 5.63 | 4.15 | 4.20 | 3.83 | 4.712 | 1.30 |
| **2,3,4- α-L-Fuc** | 13C | 98.5 | 71.9 | 77.0 | 82.4 | 68.5 | 16.18 |
| **B** | 1H | 5.22 | 3.81 | 3.84 | 4.02 | 4.02 | 3.73-3.70 |
| **α-D-Gal** | 13C | 100.2 | 69.5 | 70.9 | 70.3 | 72.5 | 62.3 |
| **C** | 1H | 5.02 | 4.03 | 4.09 | 3.74 | 3.49 | 1.28 |
| **α-L-Rha** | 13C | 104.0 | 71.5 | 71.5 | 73.2 | 70.0 | 17.9 |
| **D** | 1H | 4.99 | 3.57 | 3.95 | 3.71 | 3.65 | 3.94-3.82 |
| **3,4-α-D-Glc** | 13C | 80.7 | 75.1 | 76.8 | 74.8 | 78.1 | 60.6 |
| **E** | 1H | 4.46 | 3.38 | 3.44 | 3.66 | 4.03-3.24 | - |
| **β-D-Xyl** | 13C | 105.5 | 74.7 | 76.9 | 70.6 | 66.0 | - |
| **F** | 1H | 4.43 | 3.18 | 3.234 | 3.66 | 4.00-3.23 | - |
| **β-D-Xyl 3OMe** | 13C | 103.9 | 74.6 | 86.8 | 70.6 | 66.0 | - |

**Sup. Fig. 1.** Expansion of OSy-NE5 glycopeptide HSQC spectrum measured at 600 MHz, 310 K. Carbons bearing two hydrogen atoms have densities of opposite sign compared to the others and are colored in grey.

****

**Sup. Fig. 2.** Expansion of OSy-NE5 glycopeptide T-ROESY spectrum measured at 600 MHz, 310 K.

