Qualitative proteomic analysis of *Tipula oleracea* nudivirus occlusion bodies

Annie Bézier,1,* Grégoire Harichaux,2 Karine Musset,1 Valérie Labas2 and Elisabeth A. Herniou1

**Abstract**

Nudiviruses are arthropod-specific large double-stranded circular DNA viruses, related to baculoviruses, which replicate in the nucleus of the cells they infect. To date, six fully sequenced nudiviral genomes are available in databases, and the protein profile from nudivirus particles was mainly characterized by PAGE. However, only a few direct matches have been completed between genomic and proteomic data, with the exception of the major occlusion body protein from *Peneaus monodon* nudivirus and four nucleocapsid proteins from *Helicoverpa* *zea* nudivirus-2. The function of predicted nudiviral proteins is still inferred from what is known from baculoviruses or endogenous nudiviruses (i.e. bracoviruses). *Tipula oleracea* nudivirus (*ToNV*) is the causative agent of crane fly nucleopolyhedrosis. Along with *Peneaus monodon* nudivirus, *ToNV* is the second fully sequenced nudivirus to be described as forming occlusion bodies. The protein profile revealed by Coomassie-stained SDS-PAGE is very similar to those observed for other nudiviruses, with five major protein bands of about 75, 48, 35, 25 and 12 kDa. Proteomic analysis, using on-line nanoflow liquid chromatography in tandem with high-resolution mass spectrometry, revealed that *ToNV* occlusion bodies are composed of 52 viral proteins, the most abundant of which are the functional homologue of baculovirus polyhedrin/granulin and the homologues of three *Helicoverpa zea* nudivirus-2 predicted proteins: the two virion structural proteins 34K (Hz2V052, the baculovirus capsid protein VP39 homologue) and 11K (Hz2V025), and the hypothetical protein Hz2V079, a newly identified nudivirus core gene product.

**INTRODUCTION**

Formerly described as non-occluded baculoviruses [1], nudiviruses form a highly diverse group that shares biological, structural and genomic features with other nuclear arthropod large circular dsDNA viruses (baculoviruses, hytrosaviruses, nimaviruses and filamentous viruses). However, they also share lineage-specific characters [2–5], which justified the creation of the new *Nudiviridae* family by the International Committee on Taxonomy of Viruses [6].

Nudiviruses display a wider host range than baculoviruses, hytrosaviruses or nimaviruses, they infect more insect orders, as well as crustaceans. They have been described in Lepidoptera with *Heliothis zea* nudivirus-1 (*HzNV-1*) [7, 8] and *H. zea* nudivirus-2 (*HzNV-2*) [9–11]; in Diptera with *Drosophila innubia* nudivirus [12], *Tipula paludosa* nudivirus (*TpNV*) [13, 14] and *Tipula oleracea* nudivirus (*ToNV*) [4]; in Orthoptera with *Grillus bimaculatus* nudivirus (*GbNV*) [1, 15, 16]; in Coleoptera with *Oryctes rhinoceros* nudivirus (*OrNV*) [17–20]; and in Decapoda with *Peneaus monodon* nudivirus (*PmNV*) [21, 22], *Peneaus vannamrei* single nucleopolyhedrovirus (*PvNV*) [23] and possibly Crangon crangon intranuclear bacilliform virus (CcrBV) [24]. Other nudiviruses are also suspected in Trichoptera, Siphonaptera, Neuroptera, Homoptera, Thysanura, Acarina and Araneina [1]. To date, only six fully sequenced genomes are available in databases: HzNV-1 (AF451898; [8]) and HzNV-2 (NC_004156; [11]), GbNV (EF203088; [16]), OrNV (EU747721; [20]), ToNV (KM610234; [4]) and PmNV (KJ184318; [22]).

Nudiviruses are transmitted sexually or orally. All developmental stages (larvae, pupae, adults) can become infected. Depending on the host, the infection can be asymptomatic in both larvae and adults (*HzNV-2*), can be lethal in larvae and rather chronic in adults (GbNV and OrNV), and can also induce malformations and sterility (*HzNV-2*) (reviewed by Burand [25]). The tissue tropism is variable and concerns...
reproductive tissues (HzNV-2), fat body (GbNV and OrNV), mid-gut (OrNV), haemocytes (TpNV and PmNV) or hepatopancreas (PmNV). Nudiviruses are particularly prone to endogenization, as they are able to integrate their genome in host cells [7, 26] and to infect reproductive tissues [9]. Recent reports highlight striking cases of nudivirus domestication, especially the symbiotic bracoviruses associated with braconid wasps [27] since 100 million years ago [28], which likely derive from a nudivirus more related to ToNV [4], or the virus-like particles (VLPs) from the ichneumon wasp Venturia canescens, derived from an Alpha-nudivirus [29]. To date, there is no evidence that the endogenous alphavirus of the brown planthopper, Nilaparvata lugens [30], retains any functional activity.

A point of particular interest in the Nudiviridae is their phenotypic diversity, as within this family some viruses are transmitted as enveloped occlusion bodies (OBs), such as ToNV and TpNV [4, 14], or non-enveloped OBs, such as PmNV [22, 31, 32]; as non-occluded virions, such as Penaeus nudiviruses [22, 33], HzNV-1 [7] or GbNV [1, 7], or as facultatively occluded virions, such as OrNV [1, 9]. The genetic bases of OrNV facultative OBs still remain to be determined [4, 34]. The protein profile obtained by PAGE [32, 35, 36] estimates nudivirus particles to contain fewer than 30 proteins. With the exception of the major OB protein (MOBP) from PmNV [31] and four viral structural proteins (11K, 15K, 28K and 34K) from HzNV-2 [37], no direct match has ever been made between complete genomic and proteomic data of nudiviruses.

The genome of ToNV is 145,704 bp and contains 131 predicted ORFs [4]. Comparative genomic analyses revealed that ToNV shares 32 core genes with Nudiviridae and 21 with Baculoviridae [4]. Based on sequence similarity with baculoviruses, these genes are involved in DNA replication and processing, transcription, packaging and assembly, and per os infectivity (reviewed by Rohrmann [38]). Scanning and transmission electron microscopy further revealed that ToNV-enveloped OBs contain a large number of rod-shaped virions embedded within a dense protein matrix [4]. In this study, we describe the first comprehensive qualitative proteomic analysis of an occluded nudivirus. By using online nanoflow liquid chromatography in tandem with high-resolution mass spectrometry (nanoLC-HR-MS/MS), we identified the MOBPs that comprise ToNV-occluded particles and further characterized all other components.

**RESULTS AND DISCUSSION**

**Coomassie-stained SDS-PAGE profile revealed five major protein bands**

Separation of approximately $4 \times 10^6$ OBs on a 12.5% SDS-PAGE gel revealed five major protein bands (Fig. 1). The apparent molecular mass of the major band was estimated at 25 kDa. Four other major bands were also identified with apparent molecular mass of 75, 48, 35 and 12 kDa. A low-complexity profile was also observed for the PvNV [23], with a unique major band corresponding to the major polypeptide (polypehedrin), and for the non-occluded baculovirus-like agent (the Chinese baculovirus isolate named CBV), most probably a nudivirus also pathogenic to penaeid shrimp, with four major bands that were approximately 75, 27.5, 23.5 and 19 kDa [39]. Replication studies on other nudiviruses in insect cells revealed the presence of about nine major bands corresponding to structural proteins from about 14–153 kDa in HzNV-1 [35], 13.5–190 kDa in HzNV-2 [37] and 11.5–69 kDa in OrNV [36]. Protein analyses of occluded circular dsDNA viruses, like baculoviruses, revealed more or less complex profiles, with proteins ranging from about 6 to over 200 kDa according to genera, virion morphotypes, techniques used to separate proteins and staining methods [40–45]. On the other hand, gradient SDS-PAGE analysis of purified Glossina pallidipes salivary gland hypertrophy virus (GpSGHV) with Coomassie blue staining revealed a complex profile with at least 35 proteins ranging from 10 to over 130 kDa, but only around 10 distinctive bands [46]. Considering the lower sensitivity of the Coomassie blue staining approach compared to radiolabelling, silver or colloidal blue staining, the ToNV protein profile appeared similar to those observed for nudiviruses, baculoviruses or hytrosaviruses.

![Fig. 1. Coomassie-stained SDS-PAGE of ToNV OBs analysed by LC-MS/MS. Five microlitres of purified OB (lane 2) were denatured and then run on a 12.5% SDS-PAGE along with Kaleidoscope protein standard (Bio-Rad, lane 1) at 70 V for 3 h. Major bands and their apparent molecular mass (in kDa) are indicated by arrows.](Image)
**ToNV OBs contain at least 48 viral proteins**

To allow the identification of all proteins that composed the ToNV OBs, major bands were excised individually from polyacrylamide gel and the remaining gel lane was divided into 20 other gel sections (Fig. 1). After in-gel trypsin digestion, peptides were extracted from each slice and subjected to bottom-up proteomics [47] using nanoLC-HR-MS/MS, which is a highly sensitive proteomic method well suited for complex samples, available in limited quantities and for non-sequenced organisms [48, 49]. Protein identification with a minimal 95.0% probability against the ToNV-predicted protein database revealed that ToNV OBs were composed of at least 48 viral proteins (Table 1). Among these, 41 were identified with two or more exclusive peptides and 7 were unambiguously identified from a unique peptide with a minimal 95.0% probability (Table 1). Four additional proteins could be identified with a unique peptide at the same confidence level (>95%) but a lower protein threshold probability (50–79%). These corresponded to GbNVORF19, ORF080, P47 and ORF122. Genes encoding these proteins were distributed throughout the genome, contrary to what was observed in hytrosaviruses, in which genes involved in virions production appear confined to a specific genomic region [46, 50].

ToNV OBs thus harboured between 48 and 52 viral proteins, which is the largest number of components detected in nudiviruses. In HzNV-1, HzNV-2 and OrNV, 28, 16 and 27 virus structural proteins were identified, respectively, after infection of insect cell lines using pulse-labelling or silver-staining approaches [35, 36, 51]. This high number of viral proteins identified in ToNV OBs might reflect that ToNV is an occluded nudivirus, or the high sensitivity of nanoLC-HR-MS/MS enables the detection of proteins present in very small amounts. On the other hand, this number was in the range of what has been observed in baculoviruses with equivalent techniques [40, 42–45, 52–55]. For baculoviruses, proteomic analyses have led to a general understanding of virus particle structure for both virion morphotypes, the occlusion-derived viruses (ODVs) that initiate primary infection and the budded viruses that spread the infection from cell to cell leading to a systemic infection [38, 56, 57]. ODVs are composed of 16–73 protein products [52, 55] and budding viruses are composed of 23–51 protein products [43, 45]. However, a higher number of proteins, up to 86, were detected by LC-MS/MS in hytrosaviruses [46, 58], which are enveloped but not occluded nuclear arthropod large circular dsDNA viruses.

**Comparison with other arthropod-specific large double-stranded circular DNA viruses**

Of the 48 identified proteins with a 95% protein threshold, 21 were already known viral particle components in baculoviruses (NPV and/or GV) (reviewed by Rohrmann [38]), nudiviruses [31, 37], hytrosaviruses [46, 50] or endogenous nudiviruses such as bracoviruses [27, 59, 60] or VLPs of the wasp Venturia canescens (VcVLP) [29], and 27 were hitherto unique viral components of ToNV OBs. The previously identified polyhedrin/granulin homologue (MOBP, ORF059) [4] appeared as the major protein component of OBs in ToNV (Table 1), as it does for baculoviruses. This nudiviral polyhedrin/granulin homologue might thus form the matrix of OBs in which ToNV virions are embedded [4] and participate in long-term viroin conservation in the environment, although it might confer less effective protection than baculovirus polyhedra [57, 61].

VP39, 38K and VLFL-1 homologues found associated with both baculovirus viroin morphotypes [40, 43, 45], where they play a crucial role in nucleocapsid formation and assembly, or with particles of endogenous nudiviruses [27, 29, 59, 60] were identified as ToNV OB components (Table 1). P33, which was identified as baculovirus ODV [40, 42, 44, 45], bracovirus particles or VcVLP component [29, 60], was also present. Additionally, two nudiviral core proteins, the integrase INT and the FLAP endonuclease FEN-1, most probably involved in DNA metabolism [62–64], were identified. However, INT is the only one that has previously been identified as particle-associated protein in bracoviruses [60] and hytrosaviruses (GpSGHVORF30; [46]). In combination with VLF-1, which is also an integrase/recombinase, the nudiviral INT and FEN-1 might play a critical role in viral DNA and/or viroin maturation, but also more generally in viral replication [60, 65–67]; they might also be involved in nudiviral DNA integration and its excision from the host genome.

All PIF proteins (P74, PIF-1 to PIF-6 and VP91), which are essential for the oral infectivity of baculoviruses [68–70], were also identified. PIFs are conserved in exogenous nudiviruses [4] and, to a lesser extent, in hytrosaviruses [46, 50] and filamentous viruses [5]. Recent studies revealed that these proteins are also conserved particle components of endogenous nudiviruses, which are not transmitted orally [27, 29, 60]. As for baculoviruses, PIF proteins must therefore play a crucial role in infectivity of nudivirus viroins.

Surprisingly, in ToNV OBs we detected neither of the homologues of Ac81 or P6.9, which were detected in both baculoviruses [40, 43, 45, 71] and GpSGHV for Ac81 [46, 58]. The absence of P6.9 might be due to technical issues. ToNV P6.9 (ORF051) is composed of 119 aa, and a theoretical trypsin digestion using the PeptideCutter Expasy tool (http://web.expasy.org/peptide_cutter) revealed 33 cleavage sites generating 22 unique peptides, ranging in size from 1 aa to a maximum of 11 aa (not shown). Only two peptides of 10 and 11 aa were theoretically just within the detection limits of the mass spectrometry method. This could explain why this protein was not detected in ToNV OBs.

LEF-4 and P47 homologues were present in ToNV OBs. Although the presence of LEF-4 has not been previously described as a viral particle component, the presence of one or more of the other RNA polymerase subunits is not unusual, as it has already been observed in baculoviruses [42, 55] and hytrosaviruses [46, 58].
Table 1. ToNV major and associated OB proteins identified by nanoLC-HR-MS/MS ranked by decreasing predicted molecular mass

The four major components with known homologues in other nudiviruses are highlighted in grey, with colour intensity based on normalized total spectrum count. Homologous proteins were those present (+) or identified (·) as particle components for at least one representative of the referenced viral families. NPV, nucleopolyhedrovirus; GV, granulovirus; NV, nudivirus; SGHV, salivary gland hypertrophy virus; BV, bracovirus; VcVLP, Ventura canescens virus-like particles; MM, predicted molecular mass; pI, isoelectric point.

<table>
<thead>
<tr>
<th>ORF no.</th>
<th>Protein designation</th>
<th>Length (aa)</th>
<th>MM (kDa)</th>
<th>pI</th>
<th>No. of unique peptides</th>
<th>Protein coverage (%)</th>
<th>Total spectrum count</th>
<th>Present (+) or identified (·) as particle component in</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>ORF125</td>
<td>808</td>
<td>94.99</td>
<td>5.02</td>
<td>46</td>
<td>49</td>
<td>857</td>
<td>NPV-GV, NV, SGHV, BV, VcVLP</td>
</tr>
<tr>
<td>130</td>
<td>ORF130</td>
<td>765</td>
<td>88.28</td>
<td>4.92</td>
<td>36</td>
<td>50</td>
<td>415</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>P74</td>
<td>763</td>
<td>87.20</td>
<td>9.91</td>
<td>8</td>
<td>13</td>
<td>40</td>
<td>+, +, +, +</td>
</tr>
<tr>
<td>16</td>
<td>VP91/95</td>
<td>742</td>
<td>86.95</td>
<td>5.33</td>
<td>15</td>
<td>25</td>
<td>111</td>
<td>+, +, +</td>
</tr>
<tr>
<td>57</td>
<td>CcS0C22.5</td>
<td>690</td>
<td>81.88</td>
<td>5.13</td>
<td>15</td>
<td>26</td>
<td>107</td>
<td>+, +</td>
</tr>
<tr>
<td>108</td>
<td>ORF108</td>
<td>704</td>
<td>81.16</td>
<td>5.89</td>
<td>47</td>
<td>45</td>
<td>904</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>PmNVORF62</td>
<td>597</td>
<td>68.85</td>
<td>9.13</td>
<td>42</td>
<td>66</td>
<td>2442</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>HzNVORF128</td>
<td>561</td>
<td>66.18</td>
<td>8.55</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>+, +</td>
</tr>
<tr>
<td>69</td>
<td>PIF-1</td>
<td>527</td>
<td>60.04</td>
<td>8.01</td>
<td>10</td>
<td>21</td>
<td>101</td>
<td>+, +</td>
</tr>
<tr>
<td>25</td>
<td>LEF-4*</td>
<td>494</td>
<td>58.96</td>
<td>4.86</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>+, +, +, +</td>
</tr>
<tr>
<td>120</td>
<td>ORF120</td>
<td>497</td>
<td>57.75</td>
<td>6.00</td>
<td>3</td>
<td>7</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>FEN-1</td>
<td>479</td>
<td>56.85</td>
<td>8.43</td>
<td>17</td>
<td>38</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>PIF-5c</td>
<td>450</td>
<td>51.94</td>
<td>4.22</td>
<td>17</td>
<td>34</td>
<td>262</td>
<td>+, +, +, +</td>
</tr>
<tr>
<td>112</td>
<td>HzNVORF64</td>
<td>408</td>
<td>47.81</td>
<td>4.64</td>
<td>5</td>
<td>15</td>
<td>12</td>
<td>+</td>
</tr>
<tr>
<td>84</td>
<td>ORF084*</td>
<td>421</td>
<td>47.74</td>
<td>5.60</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>ORF080†</td>
<td>388</td>
<td>45.23</td>
<td>4.10</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>VL1-1</td>
<td>385</td>
<td>44.81</td>
<td>5.93</td>
<td>6</td>
<td>22</td>
<td>15</td>
<td>+, +</td>
</tr>
<tr>
<td>15</td>
<td>PmNVORF47</td>
<td>364</td>
<td>43.40</td>
<td>4.89</td>
<td>10</td>
<td>26</td>
<td>285</td>
<td>+</td>
</tr>
<tr>
<td>96</td>
<td>PIF-5b</td>
<td>398</td>
<td>43.27</td>
<td>8.11</td>
<td>15</td>
<td>30</td>
<td>278</td>
<td>+, +, +, +</td>
</tr>
<tr>
<td>7</td>
<td>PIF-2</td>
<td>362</td>
<td>40.96</td>
<td>6.34</td>
<td>14</td>
<td>47</td>
<td>148</td>
<td>+, +, +, +</td>
</tr>
<tr>
<td>107</td>
<td>ORF107</td>
<td>336</td>
<td>39.58</td>
<td>7.76</td>
<td>16</td>
<td>45</td>
<td>606</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>INT*</td>
<td>331</td>
<td>39.15</td>
<td>8.31</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>+, +</td>
</tr>
<tr>
<td>115</td>
<td>P47†</td>
<td>333</td>
<td>39.09</td>
<td>7.75</td>
<td>1</td>
<td>2</td>
<td>20</td>
<td>+, +, +</td>
</tr>
<tr>
<td>119</td>
<td>PIF-4</td>
<td>327</td>
<td>38.31</td>
<td>5.76</td>
<td>14</td>
<td>48</td>
<td>88</td>
<td>+, +, +</td>
</tr>
<tr>
<td>87</td>
<td>VP39</td>
<td>324</td>
<td>37.77</td>
<td>6.19</td>
<td>47</td>
<td>78</td>
<td>4999</td>
<td>+, +, +</td>
</tr>
<tr>
<td>49</td>
<td>HzNVORF106</td>
<td>307</td>
<td>36.30</td>
<td>4.88</td>
<td>13</td>
<td>32</td>
<td>74</td>
<td>+</td>
</tr>
<tr>
<td>63</td>
<td>38K</td>
<td>291</td>
<td>35.02</td>
<td>9.03</td>
<td>7</td>
<td>19</td>
<td>25</td>
<td>+, +</td>
</tr>
<tr>
<td>99</td>
<td>P33</td>
<td>289</td>
<td>34.70</td>
<td>8.85</td>
<td>3</td>
<td>11</td>
<td>30</td>
<td>+, +, +</td>
</tr>
<tr>
<td>27</td>
<td>GbNVORF19†</td>
<td>291</td>
<td>34.12</td>
<td>8.35</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>+, +</td>
</tr>
<tr>
<td>31</td>
<td>PmV</td>
<td>272</td>
<td>31.78</td>
<td>5.75</td>
<td>13</td>
<td>44</td>
<td>200</td>
<td>+</td>
</tr>
<tr>
<td>54</td>
<td>CcS0C22.6</td>
<td>254</td>
<td>30.12</td>
<td>8.51</td>
<td>8</td>
<td>26</td>
<td>169</td>
<td>+</td>
</tr>
<tr>
<td>59</td>
<td>MOBP‡</td>
<td>241</td>
<td>27.36</td>
<td>6.31</td>
<td>44</td>
<td>87</td>
<td>32188</td>
<td>+</td>
</tr>
<tr>
<td>117</td>
<td>HzNVORF9</td>
<td>237</td>
<td>27.31</td>
<td>4.82</td>
<td>8</td>
<td>34</td>
<td>40</td>
<td>+, +</td>
</tr>
<tr>
<td>26</td>
<td>ORF026*</td>
<td>226</td>
<td>27.00</td>
<td>4.97</td>
<td>1</td>
<td>7</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>ORF062*</td>
<td>224</td>
<td>26.31</td>
<td>6.43</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>ORF023</td>
<td>225</td>
<td>26.20</td>
<td>6.19</td>
<td>19</td>
<td>75</td>
<td>894</td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>ORF106</td>
<td>218</td>
<td>25.29</td>
<td>9.82</td>
<td>13</td>
<td>59</td>
<td>651</td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>ORF089</td>
<td>203</td>
<td>23.37</td>
<td>4.88</td>
<td>10</td>
<td>52</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>ORF032</td>
<td>205</td>
<td>22.47</td>
<td>6.14</td>
<td>10</td>
<td>38</td>
<td>566</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>PIF-3</td>
<td>193</td>
<td>21.86</td>
<td>4.83</td>
<td>6</td>
<td>25</td>
<td>43</td>
<td>+, +, +, +</td>
</tr>
<tr>
<td>124</td>
<td>ORF124</td>
<td>180</td>
<td>21.27</td>
<td>8.24</td>
<td>7</td>
<td>34</td>
<td>138</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>ORF034</td>
<td>188</td>
<td>20.64</td>
<td>7.16</td>
<td>5</td>
<td>24</td>
<td>1139</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>PIF-6</td>
<td>143</td>
<td>16.81</td>
<td>5.60</td>
<td>2</td>
<td>15</td>
<td>13</td>
<td>+, +, +</td>
</tr>
<tr>
<td>30</td>
<td>ORF030*</td>
<td>144</td>
<td>16.77</td>
<td>7.25</td>
<td>1</td>
<td>9</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>114</td>
<td>PmNVORF64*</td>
<td>139</td>
<td>15.76</td>
<td>4.36</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>+</td>
</tr>
<tr>
<td>122</td>
<td>ORF122†</td>
<td>133</td>
<td>15.68</td>
<td>9.92</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Twenty-one core nudiviral proteins were detected in ToNV OBs (Fig. 2, in black). Seven proteins (H2N0RF128, PmN0RF47, H2N0RF413, Cc50C22.6, Cc50C22.5, PmN0RF7 and PmN0RF64) were shown for the first time to be present in OBs. Five nudiviral proteins (PmN0RF62, PmV, H2N0RF106, H2N0RF64 and H2N0RF9) have previously exclusively been characterized as particle components of endogenous nudiviruses [27, 29, 60]. The 11K protein had been detected in endogenous nudiviruses and HzNV-2 [37].

Lastly, the four additional proteins (GbN0RF19, ORF080, P47 and ORF122) appeared to be relevant components of ToNV OBs, although below the filter thresholds. In particular, P47 needs to be confirmed as its detection relies on a 100 % peptide sequence identity based on a sequence also found in bacteria and eukaryota (data not shown). However, P47 and GbN0RF19 have previously been identified as viral particle components of Clostera anachoreta granulovirus [55] and VcVLP [29], respectively, which suggests that they are genuine particle components. ORF080 and ORF122 both appear specific to ToNV and are therefore described for the first time as OB-associated proteins.

Overall, from these 52 identified proteins, 21 were encoded by nudivirus core genes, 11 had homologues in at least one exogenous or endogenous nudivirus and 20 are currently specific to ToNV [4].

### Protein abundance

On the basis of the total spectrum count obtained by Scaffold software (Proteome Software), the relative quantity of each OB protein can be estimated (Table 1 and Fig. 2) [72, 73]. Among ToNV OB proteins, the most abundant is encoded by orf087, the functional homologue of the baculovirus polyhedrin [4]. MOBP represented about 65 % of all identified spectrum counts, and MOBP peptides were found within all gel slices subjected to nanoLC-HR-MS/MS (not shown), including all major bands. The identification of single proteins in several bands is not unusual, as this was previously observed for the NPV and GV major structural proteins [41–44]. Moreover, protein abundance estimated by calculating the emPAI using Scaffold Q+ software 4.6.1 [74] further revealed that MOBP represented ~95 % of the viral proteins present in the 25 kDa major band. The second component forming this band was ORF023 (~4 %).

The other main OB components were encoded by orf087, orf028 and orf019, which are respectively homologues to three HzNV-2 predicted proteins corresponding to the virion structural proteins 34K (H2V052, the baculovirus capsid protein VP39 homologue) and 11K (H2V025) [37] and to the hypothetical protein H2V079, which was previously determined as a nudivirus core gene [4]. Of these proteins, VP39 was identified as the second major protein in ToNV OBs (~10 % of all identified spectrum counts) and also the second main viral component of the major band estimated at 35 kDa (~38 %; MOBP accounts for ~61 % of that band). VP39 was also identified as a highly abundant structural protein in baculoviruses [75, 76] or exogenous [37] and endogenous [27, 59, 60] nudiviruses. 11K is a nudiviral specific core protein. This protein was found in the 12 kDa band in the same proportion (~0.5 %) as VP39 and two other proteins of close estimated molecular weight (ORF041, 8.7 kDa and ORF053, 10.7 kDa). The HzNV-2 11K homologue was suspected as a possible viral envelope component [37]; in contrast, bracovirus homologues (H2N0RF124-like) were not found as particle components [59, 60]. The level of PmN0RF62 (ORF019) was found in low proportion in both remaining major bands, mainly in combination with MOBP and ORF023 in the 48 kDa band (~0.2 %) and with MOBP and VP39 in the 75 kDa band (~2 %).

All the per os infectivity factors were present in ToNV OBs, but only two (b and c) out of the three members of the PIF-5 ToNV multigenic family were found. PIFs were not highly represented, with only 13–278 total spectra, accounting for only ~2 % of all identified viral peptides. However, such low abundance was also observed in baculoviruses [68, 77–79], hytrosaviruses [58] and endogenous nudiviruses [59].

Eight proteins, encoded by orf019 (PmN0RF62), orf027 (GbN0RF19), orf028 (11K), orf031 (PmV), orf049 (H2N0RF106), orf059 (MOBP), orf012 (H2N0RF64) and orf017 (H2N0RF9), had homologues previously identified as particle components with no other known function.

### Table 1. cont.

<table>
<thead>
<tr>
<th>ORF no.</th>
<th>Protein designation</th>
<th>Length (aa)</th>
<th>MM (kDa)</th>
<th>pI</th>
<th>No. of unique peptides</th>
<th>Protein coverage (%)</th>
<th>Total spectrum count</th>
<th>Present (+) or identified (+) as particle component in</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>ORF042*</td>
<td>105</td>
<td>12.65</td>
<td>9.70</td>
<td>1</td>
<td>21</td>
<td>14</td>
<td>NPV-GV, NV, SGVH, BV, VcVLP</td>
</tr>
<tr>
<td>28</td>
<td>11K</td>
<td>115</td>
<td>12.39</td>
<td>10.13</td>
<td>9</td>
<td>44</td>
<td>1325</td>
<td>+</td>
</tr>
<tr>
<td>100</td>
<td>PmN0RF7</td>
<td>92</td>
<td>10.73</td>
<td>6.41</td>
<td>3</td>
<td>23</td>
<td>35</td>
<td>+</td>
</tr>
<tr>
<td>53</td>
<td>ORF053</td>
<td>93</td>
<td>10.70</td>
<td>9.38</td>
<td>2</td>
<td>41</td>
<td>11</td>
<td>+</td>
</tr>
<tr>
<td>41</td>
<td>ORF041</td>
<td>74</td>
<td>8.70</td>
<td>9.37</td>
<td>5</td>
<td>49</td>
<td>107</td>
<td>+</td>
</tr>
<tr>
<td>52</td>
<td>HzN0RF143</td>
<td>53</td>
<td>6.53</td>
<td>7.76</td>
<td>3</td>
<td>51</td>
<td>11</td>
<td>+</td>
</tr>
</tbody>
</table>

*Proteins unambiguously identified with a unique peptide, a minimal protein threshold of 95 %.
†Proteins unambiguously identified with a unique peptide, a protein threshold between 50 and 94 % but a peptide threshold above or equal to 95 %.
in nudiviruses [31, 37], bracoviruses [27, 59, 60] and VcVL [29] (Table 1 and Fig. 2). These represented 73.23% of all identified spectrum counts. Finally, 20 proteins had no known viral homologue and currently, therefore, have no known function, except putative associated or viral structural proteins in ToNV OBs. Among these, ORF034 belongs
to the ToNV-specific ORF032 multigenic family [4], from which only two (ORF032 and ORF034) out of the three members were detected as OB components. These 20 proteins specific to ToNV represented 12.90% of all identified spectrum counts and about one sixth of all identified spectrum counts corresponding to accessory gene products (data not shown). In fine, from the 52 identified proteins, nudivirus core genes and accessory genes represented 20 and 80%, respectively, of all identified spectrum counts (data not shown).

**Associated arthropod proteins**

Nine host cell-derived proteins were identified with confidences as being associated with ToNV OBs (over 99%). These mainly concern cytoskeleton proteins, proteins involved in DNA or mRNA processing, and potential oxidative stress response proteins (Table 2). Such host proteins might have been co-purified with ToNV OBs and consequently represent contaminants, but they might also be intrinsically associated with the OBs or viral particles as was shown for other phylogenetically related [43, 45, 54] or unrelated [80, 81] enveloped viruses. They might therefore play a crucial role in the viral life cycle. Similarly, Helicoverpa armigera NPV (HearNPV) budded virus and ODV were shown to be associated with 101 and 21 host proteins, respectively [45]. Furthermore, the presence of such cellular proteins is not unusual, even among nudiviruses in general, as we found that ultra-deep purified particles (sucrose and caesium chloride gradients) of the endogenous nudivirus, Cotesia congregata bracovirus, contain heat shock proteins, ATP synthases, ATPases, tubulin and actin (unpublished data).

In particular, actin, beta tubulin, histone and heat shock protein 70 were identified in baculovirus ODV [43, 45, 54, 55, 82], herpes viruses [81], human immunodeficiency virus [89], SODs were described as virion components in nucleocyttoplasmic large DNA viruses such as vaccinia virus [90, 91] or in the Megavirus chilensis [92]. Sod homologues were found in the genome of entomopoxviruses [93] and almost all lepidopteran baculoviruses [94]. They were described as being associated with ODV in Chrysoideixis chalcites NPV [53] and HearNPV [45]. In this viral context, SODs are suspected to counteract the host stress response, thus increasing virus survival, or to inhibit or regulate cellular SOD function [90, 95]. Concerning ToNV, the insect SOD protein might have been captured by the virus by chance, while infected cells contained many cellular SODs, or by orientation in such a way as to play a similar protective role in virus infection.

Hrp59 is a nuclear protein that belongs to the heterogeneous nuclear ribonucleoprotein (hnRNP) M type. Such proteins are evolutionarily conserved from yeasts to humans. They are involved in packaging nascent transcripts into ribonucleoprotein complexes, presenting gene target specificity depending on the considered type [96, 97]. Human M type homologue is implicated in pre-mRNA splicing regulation under stress conditions [98]. HnRNPs seem to be recurring targets for many RNA viruses [99–101], but to our knowledge, nothing is known about the hnRNP association with DNA virus particles or hnRNP exploitation by DNA viruses.

Other host proteins might also be associated with ToNV OBs that could not be identified because the crane fly genome is

<table>
<thead>
<tr>
<th>Identified protein</th>
<th>Gene ontology– biological process (GO)</th>
<th>Total spectra</th>
<th>Protein coverage (%)</th>
<th>GI no. (integer sequence identifiers)</th>
<th>Species</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu,Zn superoxide dismutase</td>
<td>Response to oxidative stress</td>
<td>14</td>
<td>8</td>
<td>1019906</td>
<td>Drosophila saltans</td>
<td>Diptera</td>
</tr>
<tr>
<td>Histone H4</td>
<td>Nucleosome assembly</td>
<td>7</td>
<td>13</td>
<td>194772468</td>
<td>Drosophila ananassae</td>
<td>Diptera</td>
</tr>
<tr>
<td>Muscle myosin heavy chain Actin</td>
<td>Myosin filament organization</td>
<td>5</td>
<td>1</td>
<td>183979376</td>
<td>Papilio xuthus</td>
<td>Lepidoptera</td>
</tr>
<tr>
<td>ADP, ATP carrier protein</td>
<td>Response to oxidative stress</td>
<td>4</td>
<td>11</td>
<td>195438665</td>
<td>Drosophila willistoni</td>
<td>Diptera</td>
</tr>
<tr>
<td>ADP, ATP carrier protein 2</td>
<td>Response to oxidative stress</td>
<td>3</td>
<td>11</td>
<td>751238593</td>
<td>Solenopsis invicta</td>
<td>Hymenoptera</td>
</tr>
<tr>
<td>Hrp59 protein</td>
<td>mRNA splicing</td>
<td>4</td>
<td>2</td>
<td>50880296</td>
<td>Chromatomyces tentans</td>
<td>Diptera</td>
</tr>
<tr>
<td>Tubulin beta chain</td>
<td>Microtubule-based process</td>
<td>3</td>
<td>2</td>
<td>157108656</td>
<td>Aeles aegypti</td>
<td>Diptera</td>
</tr>
<tr>
<td>Hsp 70 BDI</td>
<td>Stress response</td>
<td>3</td>
<td>4</td>
<td>311797659</td>
<td>Bactroceras dorsalis</td>
<td>Diptera</td>
</tr>
</tbody>
</table>
not sequenced. Finally, to be confirmed as ToNV OB components, immuno-gold labelling analyses of these proteins should be done. Although cytoskeleton and structural host proteins could be involved in virus entry and other proteins in preventing stress response, the real implication of these host factors in the nudivirus life cycle remains largely unknown.

Conclusions

Comprehensive qualitative proteomic analysis using online nanoLC-HR-MS/MS revealed, for the first time, the exhaustive composition of nudivirus OBs. The occluded nudivirus infecting *Tipula oleracea* contained 52 viral proteins and 9 host proteins. About half of the viral proteins were previously identified as particle components in other exogenous or endogenous nudiviruses, as well as in baculoviruses and hytrosaviruses. These mainly concern baculovirus homologues involved in oral infectivity and packaging, assembly and release. Almost all host proteins were previously identified as particle components in other enveloped viruses. Thus, conservation of the main protein components through various viral entities, in particular the nudiviruses, the both demonstrated cases of nudivirus endogenization (bracoviruses and VcVLP), the baculoviruses and the hytrosaviruses, strengthens the idea of a similar basic cell entry mechanism within the insect. However, each virus retains specificity, as over half of the proteins identified to date are specific for ToNV. This study does not provide any information regarding whether these proteins are associated with the nucleocapsids, envelopes or protein matrix of the ToNV. To discriminate further between these different compartments, it would be necessary to treat the OBs with, for example, Nonidet P-40, to compare SDS-PAGE profiles between treated and non-treated occluded particles and to proceed to new protein sequencing. The immuno-gold strategy, using candidate proteins and scanning transmission electron microscopy, could also reveal protein localization.

METHODS

Viral sample

The ToNV archival sample (*Nudivirus*, unclassified) was originally obtained from the historical insect virus collection held at the Natural Environment Research Council, Centre for Ecology and Hydrology (Wallingford, UK). The occluded particle solution used in this study was the same as that used for previous genomic analyses [4].

SDS-PAGE and gel section excision

Five microlitres of purified virus (approximately 3 µg total proteins quantified using Qubit Protein AssayKit, Life Technologies) were denatured for 30 min at 95 °C in 1× PBS and SDS-PAGE sample [102] buffers and then placed on ice for a few minutes. The sample was then electrophoresed for 3 h at 70 V on a 12.5 % polyacrylamide running gel combined with a 5 % stacking gel, along with a Kaleidoscope molecular weight marker (Bio-Rad). After staining with Coomassie blue, major bands were first individually excised from the polyacrylamide gel and the remaining lane was then divided into 20 slices to include all present components in the analysis.

In-gel digestion

Each band was washed in water/acetonitrile (1 : 1) for 5 min followed by a second wash in acetonitrile for 10 min. Cysteine reduction and alkylation were performed by successive incubation in solutions of 10 mM DTT in 50 mM sodium bicarbonate for 30 min at 56 °C and 55 mM iodoacetamide in 50 mM sodium bicarbonate for 20 min at room temperature in the dark, respectively. Gel slices were washed by incubation in 50 mM sodium bicarbonate/acetoniitrile (1 : 1) for 10 min followed by incubation in acetonitrile for 15 min. Proteins were digested overnight in 25 mM sodium bicarbonate with 12.5 ng µl⁻¹ trypsin (sequencing grade, Roche). The resulting peptides were extracted from the gel using incubation in 0.1 % formic acid/acetoniitrile (1 : 1) for 10 min followed by incubation for 5 min in acetonitrile. The collected extractions were pooled, dried with a Speed-Vac (Thermo Fisher), reconstituted with 12 µl of 0.1 % formic acid and 2 % acetonitrile, and sonicated for 10 min.

NanoLC-HR-MS/MS approach

Peptide mixtures were analysed by nanoLC-HR-MS/MS. All experiments were performed on a LTQ Orbitrap Velos Mass Spectrometer (Thermo Fisher Scientific) coupled to an Ultimate 3000 RSLC Ultra High Pressure Liquid Chromatographer ( Dionex) controlled by Chromeleon software (version 6.8 SR11; Dionex). Five microlitres of each sample were injected and loaded on an LCPackings trap column (Acclaim PepMap 100 C₁₈, 100 µm inner diameter × 2 cm long, 3 µm particles, 100 Å pores). Mobile phases consisted of (A) 0.1 % formic acid, 97.9 % water, 2 % acetonitrile (by vol.) and (B) 0.1 % formic acid, 15.9 % water, 84 % acetonitrile (by vol.). Peptides were desalted and pre-concentrated for 10 min at 5 µl min⁻¹ with 4 % solvent B. The peptide separation was conducted using a LCPackings nano-column (Acclaim PepMap C₁₈, 75 µm inner diameter × 50 cm long, 3 µm particles, 100 Å pores). The gradient consisted of 0–55 % B for 90 min, 55–99 % B for 1 min, constant 99 % B for 20 min and 4 % B for 1 min. The column was re-equilibrated for 15 min at 4 % B between runs. The nanoflow rate was set to 300 nl min⁻¹. Standard mass spectrometric conditions for all experiments were spray voltage 1.2 kV, no sheath and auxiliary gas flow; heated capillary temperature, 200 °C; predictive automatic gain control enabled, and an S-lens RF level of 50 %.

Data were acquired using Xcalibur software (version 2.1; Thermo Fisher Scientific). The LTQ Orbitrap Velos instrument was operated in positive ion mode in data-dependent mode to switch between high-resolution full-scan MS spectra collected in profile mode and low-resolution MS/MS spectra in centroid mode. Resolution in the Orbitrap was set to R=60 000. In the scan range of m/z 300–1800, the 20 most intense peptide ions with charge states ≥2 were sequentially isolated (isolation width, 2 m/z; 1 microscan) and fragmented...
in the high-pressure linear ion trap by CID (collision-induced dissociation) with normalized collision energy of 35% and wideband activation enabled. Ion selection threshold was 500 counts for MS/MS, and the maximum permitted ion accumulation times were 200 ms for full scans and 50 ms for CID-MS/MS measurements. An activation \( q = 0.25 \) and activation time of 10 ms were used. Dynamic exclusion was activated for 30 s with a repeat count of 1. The lock mass was enabled for accurate mass measurements. Polydimethylcyclosiloxane \( [m/z, 445.1200025, (Si(CH_3)_2O)_6] \) ions were used for internal recalibration of the mass spectra.

**Protein identification and data validation**

Raw data files were converted to MGF (Mascot generic format) with Proteome Discoverer software (version 1.3; Thermo Fischer Scientific). Precursor mass range of 350–5000 Da and signal to noise ratio of 1.5 were the criteria used for generation of peak lists. In order to identify the proteins, the peptide and fragment masses obtained were matched automatically against a local database comprising all the ToNV predicted proteins; such a comparison with a specific database was all the more necessary because overall nudi viral homologous sequences are relatively distant from one virus to another. MS/MS ion searches were performed using the MASCOT Daemon and search engine (version 2.3; Matrix Science). The parameters used for database searches include trypsin as a protease with two missed cleavages allowed, carbamidomethylation (+57 Da), oxidation of methionine (+16) and N-terminal protein acetylation (+42) as variable modifications. The tolerance of the ions was set to 5 ppm for parent and 0.8 Da for fragment ion matches. Mascot results obtained from the target and decoy database searches were incorporated to Scaffold software (version 3.6.4, Proteome Software), a bioinformatic tool that integrates algorithms to validate identified peptides and proteins by search engine and thus reduce the number of false positives [72]. Peptide identifications were accepted if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm [103]. Protein identifications were accepted if they could be established at greater than 95.0% probability as specified by the Protein Prophet algorithm [104]. Supplementary peptides with 95.0% peptide probability and 50–79% protein probability were also examined and validated manually against the non-redundant protein sequences using blastp (National Center for Biotechnology Information).

**Funding information**

This study was funded by the European Research Council starting grant GENOVIR (205206). The high-resolution mass spectrometer was financed (SMHART project) by the European Regional Development Fund (ERDF), the Conseil Régional du Centre, the French National Institute for Agricultural Research (INRA) and the French National Institute of Health and Medical Research (Inserm).

**Acknowledgements**

Lucie Combes-Soia (PAIB2, INRA) is thanked for helping with data analysis.

**Conflicts of interest**

The authors declare that there are no conflicts of interest.

**References**


22. Yang YT, Lee DY, Wang Y, Hu JM, Li WH et al. The genome and occlusion bodies of marine *Peneaus 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. *BMCMicrobiol* 2014;15:629.


