Phylogenetic relationships of endornaviruses in common bean from the western highlands of Kenya and global sequences

James M Wainaina 1, Elijah Ateka 2, Timothy Makori 2, Monica A Kehoe 3, Laura M Boykin Corresp. 1

1 School of Molecular Sciences/ARC CoE Plant Energy Biology, The University of Western Australia
2 P.O. Box 62000-00200, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya
3 Department of Primary Industries and Regional Development, South Perth, Australia

Corresponding Author: Laura M Boykin
Email address: laura.boykin@uwa.edu.au

**Background:** Endornaviruses are non-pathogenic viruses infecting multiple agricultural important crops including legumes, with global distribution. However, there is an absence on the complete genome of endornaviruses from legumes in particular with the sub-Saharan region. In this study, we report the first complete genomes of PvEV1 and PvEV2, and the evolutionary relationship of these genomes. **Methods:** Viral symptomatic common beans (*Phaseolus vulgaris*) showing Bean common mosaic necrosis virus (BCMNV) symptoms from Vihiga county, in the western highlands of Kenya were collected during field survey’s in the region. High throughput sequencing (RNA-Seq) was carried out on total RNA isolated from symptomatic leaf samples. Subsequently, *de novo* assembly and reference mapping was carried out to obtain the complete genomes of PvEV-1 and PvEV-2. **Results:** We identified the complete genome of *Phaseolus vulgaris endornavirus* 1 and 2 (PvEV-1 and PvEV-2) from sub-Saharan Africa (SSA). The average genome size of PvEV-1 was ~13,890 nucleotides (nt) while PvEV-2 was ~14,698 nt, encoding a single open reading frame (ORF). Single ORFs ranged from 4,632 to 4,633 aa in PvEV-1 and from 4,899 – to 4,954 aa in PvEV-2. Both ORFs encoded for the RNA-dependent RNA polymerase (*RdRP*) gene. The percentage sequence similarity between PvEV-1, PvEV-2 from this study GenBanks sequences was 29 % to 99 %. Bayesian phylogenetic analysis resolved in two well-supported monophyletic clades, with isolates from this study clustering with those from Brazil sequences. **Discussion:** This study provides the first insights into the evolutionary relationships of PvEV from SSA diverse and contributes towards filling the current knowledge gaps on *endornaviruses*
Phylogenetic relationships of endornaviruses in common bean from the western highlands of Kenya and global sequences

James M. Wainaina¹, Elijah Ateka², Timothy Makori², Monica A. Kehoe³ and Laura M. Boykin¹*

¹School of Molecular Sciences and Australian Research Council Centre of Excellence in Plant Energy Biology, University of Western Australia, Crawley, Perth, WA 6009, Australia
²Department of Horticulture, Jomo Kenyatta University of Agriculture and Technology, Nairobi P.O. Box 62000-00200, Kenya
³Department of Primary Industries and Regional Development, Diagnostic and Laboratory Service, South Perth WA 6151, Australia

*Correspondence: Dr. Laura M. Boykin

Email: laura.boykin@uwa.edu.au
Abstract

Background: Endornaviruses are non-pathogenic viruses infecting multiple agricultural important crops including legumes, with global distribution. However, there is an absence on the complete genome of endornaviruses from legumes in particular with the sub-Saharan region. In this study, we report the first complete genomes of PvEV1 and PvEV2, and the evolutionary relationship of these genomes.

Methods: Viral symptomatic common beans (Phaseolus vulgaris) showing Bean common mosaic necrosis virus (BCMNV) symptoms from Vihiga county, in the western highlands of Kenya were collected during field survey’s in the region. High throughput sequencing (RNA-Seq) was carried out on total RNA isolated from symptomatic leaf samples. Subsequently, de novo assembly and reference mapping was carried out to obtain the complete genomes of PvEV-1 and PvEV-2.

Results: We identified the complete genome of Phaseolus vulgaris endornavirus 1 and 2 (PvEV-1 and PvEV-2) from sub-Saharan Africa (SSA). The average genome size of PvEV-1 was ~13,890 nucleotides (nt) while PvEV-2 was ~14,698 nt, encoding a single open reading frame (ORF). Single ORFs ranged from 4,632 to 4,633 aa in PvEV-1 and from 4,899 – to 4,954 aa in PvEV-2. Both ORFs encoded for the RNA-dependent RNA polymerase (RdRP) gene. The percentage sequence similarity between PvEV-1, PvEV-2 from this study GenBanks sequences was 29 % to 99 %. Bayesian phylogenetic analysis resolved in two well-supported monophyletic clades, with isolates from this study clustering with those from Brazil sequences.

Discussion: This study provides the first insights into the evolutionary relationships of PvEV from SSA diverse and contributes towards filling the current knowledge gaps on endornaviruses

Keywords: Common bean, dsRNA, Seed-borne viruses, sub-Saharan Africa, smallholder farms

Introduction
The family *endornaviridae* are non-viron producing viruses that infect plants, fungi and oomycetes (Hacker, Brasier and Buck, 2005; Roossinck et al., 2011). They are double-stranded RNA (*dsRNA*) viruses and are non-encapsulated (Khalifa & Pearson, 2014). Their genome size is approximately 13.5–17.6 kilobase pairs (kbp) with a site specific nick on the 5’ region (Okada *et al.*, 2013; Khalifa and Pearson, 2014). The entire genome is composed of a single open reading frame (ORF) that codes the RNA dependent RNA polymerase (*RdRP*). The family *Endornaviridae* is composed of one genus *Endornavirus*. Species within the genus *Endornavirus* have been isolated from various important agricultural crops such as; common bean (*Phaseolus vulgaris L.*) (Okada *et al.*, 2013; Nordenstedt *et al.*, 2017), barley (*Hordeum vulgare*) (Zabalgogeazcoa & Gildow, 1992) and bell pepper (*Capsicum annuum*) (Chen & Bernards, 2015). They are transmitted across plants through vertical transmission, mainly through infected seeds (Okada *et al.*, 2013). However, they seem to have minimal effect on the host plant’s phenotype. To date their role within the plant ecosystem remains unknown (Hacker, Brasier and Buck, 2005; Roossinck *et al.*, 2011).

One species within the genus *Endornavirus* that has been isolated previously in common bean is *Phaseolus vulgaris endornavirus* (PvEV) (Okada *et al.*, 2013). PvEv are of two forms; *PvEV-1* and *PvEV-2* (Okada *et al.*, 2013). These two viruses are found to co-infect beans, although they do not cause any known viral symptoms (Mackenzie, Pring, & Bassett, 1988). Currently there are limited genomic resources on *PvEV* that would facilitate our understanding of their evolutionary relationships. Especially within sub-Saharan Africa (SSA), where there are limited genomic resources for *PvEV*, despite the large legume production in Tanzania and Kenya (Nordenstedt *et al.*, 2017). Currently the only available genomic resources of PvEV from SSA has been from small RNAs (Nordenstedt *et al.*, 2017).
In this study, we obtained complete genomes of two PvEV-1 and PvEV-2 isolates from common bean. In addition, we assessed their evolutionary relationships along with other previously published endornaviruses.

**Materials & Methods**

Ethical approval to conduct this study was obtained from the University of Western Australia (RA/4/1/7475). In addition, permission to access all privately owned farms was obtained through signed consent forms by the head of each household. We then sampled symptomatic common bean (*Phaseolus vulgaris*) leaves from the western highlands of Kenya as part of a larger surveillance study on the heterogeneous cropping system in the western highlands of Kenya. RNA extraction, cDNA library preparation and subsequent RNA sequencing on the Illumina Hiseq 2500 were carried out as previously described (Wainaina et al., 2018).

Raw reads were trimmed and assembled using CLC Genomics workbench (CLCGW ver 7.0.5) (Qiagen). Trimmed reads were assembled using the following parameters: quality score limit set to 0.01, maximum number of ambiguities was set to two and read lengths less than 100 nt were discarded. Contigs were assembled using the *de novo* assembly function on CLCGW with default automatic word size, and automatic bubble size parameters. Minimum contig length was set to 500, mismatch cost two, insertion cost three, deletion cost three, length fraction 0.5 and similarity fraction 0.9. All the contigs were subjected to Blastn and Blastx (NCBI) on the Magnus Supercomputer at Pawsey. Contigs that matched *endornavirus* were identified and exported to Geneious 8.1.8 (Biomatters). Reference-based mapping was then carried out using complete genomes retrieved from GenBank reference (PvEV-1 KT456287 and PvEV-2 KT456288). Mapping parameters were set as follows: minimum overlap 10%, minimum overlap identity 80%,
allow gaps 10% and fine tuning iteration up to 10 times. The consensus contig from the mapping was aligned using MAFFT (Katoh and Standley 2013) to the *de novo* contig of interest. The resulting alignments were manually inspected for ambiguities, which were corrected with reference to the original assembly or mapping. The open reading frame and annotation of the final sequences was done in Geneious 8.1.8 (Biomatters). Sequences were referred to as nearly complete if the entire coding region was present, and complete if the entire genome including untranslated regions were present.

Bayesian phylogenetic analysis was carried out using GTR + I + G as the optimal evolutionary model based on the jModel test (Darriba *et al.*, 2012). MrBayes was run for 50 million generations on four chains, with trees sampled every 1000 generations using GTR+I+G as the evolutionary model. In each of the runs, the first 25% (2,500) of the sampled trees were discarded as burn-in. Convergence and mixing of the chains was evaluated using Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/) with the consensus tree visualised in Fig tree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/)

### Results and Discussion

We obtained two complete genomes of PvEV-1 and PvEV-2. The PvEV-1 genome size ranged from 13,897 to 13,890 nucleotides (nt) with GC composition of 37.3 %. The ORF of the PvEV-1 isolates were between 4,632 - 4,633 amino acids (aa). On the other hand, PvEV-2 genome sizes ranged between 14,698 -14,780 nts, with GC composition of 44.8 %. The ORF of PvEV-2 isolates was between 4,899 – 4,954 aa. Both PvEV-1 and PvEV-2 contained a single ORF encoding a *RNA dependent RNA polymerase (RdRP)* protein. This was confirmed using NCBI conserved domain search of the protein sequences of the ORF. Percentage nucleotide sequence similarity in
PvEV-1 and PvEV-2 from this study was between 96% and 99% when compared to the Brazilian sequences, which are the only other complete genome sequences available (Table 1). Although the common bean samples showed BCMNV-like symptoms, we did not identify BCMNV viral reads within these samples. It is plausible that the symptoms observed could have been due to nutritional deficiencies within the host plants or DNA viruses that were not detected by RNA-Seq.

The optimal genome tree resolved two well-supported clades with posterior support of greater than 0.70 (Fig. 1). PvEv-1 sequences from this study grouped in clade II clustering with Brazilian isolates; while similarly PvEv-2 sequences clustered in clade I also with Brazilian isolates (Fig. 1). Percentage sequence similarity between the PvEv-1 isolates from Kenya was 98 and 99% compared to the Brazilian sequences (Table 1). While, PvEv-2 nucleotide sequence similarity of the Kenyan and Brazilian isolates (Table 1) was between 96% and 98%.

Conclusions

In this study, we have identified the first time *Phaseolus vulgaris* endornavirus whole genome from SSA. Subsequent evolutionary relationships and percentage sequence similarity indicated a close similarity between the Kenyan and Brazilian sequences. This lends support to the primary theory that infected seed stocks could be a possible driver of dispersal for these viruses across large geographical regions. More sequences will help to shed further light on the origins and distribution of these viruses.

Figure and Table legends

**Figure 1:** Consensus tree sampled in a Bayesian analysis of the whole genome tree of *Endornavirus* with carrot yellow leaf curl virus as the outgroup in MrBayes 3.2.2. **Key:** BaEV:
Basella alba endornavirus (host Malabar spinach), OsEV: (Oryza sativa endornavirus (host rice), OrEV: Oryza rufipogon endornavirus (host rice) PaEV: Persea americana endornavirus PhEv1: Phytophthora endornavirus, (host: Douglas fir) TaEV: Tuber aestivum endornavirus (host: summer truffle), BPEV: bell pepper endornavirus (host: Bell pepper), PvEv1 PvEv-2: Phaseolus vulgaris endornavirus (host common bean), GEEV: grapevine endophyte endornavirus (host:Grapevine) CYLV: Carrot Yellow leaf curl virus

Table 1: Percentage sequence similarity between 15 endornavirus genomes with the carrot yellow leaf curl virus as the out-group in Geneious 8.1.5

GenBank Accession Numbers

MF281669-MF281672

Acknowledgment: JMW is supported by the Australian Award Scholarships from the Department of Foreign Affairs and Trade (DFAT) and this work is part of his PhD research. Fieldwork activities were coordinated through Cassava Diagnostics Project Kenyan node. Pawsey Supercomputing Centre provided supercomputer resources for data analysis with funding from the Australian Government and the Government of Western Australia.

References

Phytophthora species is related to the plant endornaviruses and contains a putative UDP

http://doi.org/10.1099/vir.0.80808-0

Katoh, K., & Standley, D. M. 2013. MAFFT Multiple Sequence Alignment Software Version 7:


http://doi.org/10.1016/j.virusres.2014.06.010

Phaseolus vulgaris L. are not associated with cytoplasmic male sterility. *Theoretical and

Nordenstedt, N., Marcenaro, D., Chilagane, D., Mwaipopo, B., Nchimbi-msolla, S., Njau, P. J.
R., & Rajama, M. 2017. Pathogenic seedborne viruses are rare but Phaseolus vulgaris
dornaviruses are common in bean varieties grown in Nicaragua and Tanzania, PLOS
ONE 12 (8): 1–18.

2013. Molecular characterization of two evolutionarily distinct endornaviruses co-infecting

http://doi.org/10.1099/vir.0.044487-0


Figure 1 (on next page)

Consensus tree sampled in a Bayesian analysis of the whole genome tree of *Endornavirus* with carrot yellow leaf curl virus as the outgroup in MrBayes 3.2.2.

Consensus tree sampled in a Bayesian analysis of the whole genome tree of *Endornavirus* with carrot yellow leaf curl virus as the outgroup in MrBayes 3.2.2. **Abbreviation Key**

BaEV: *Basella alba* endornavirus (host Malabar spinach), OsEV: (*Oryza sativa* endornavirus (host rice), OrEV: *Oryza rufipogon* endornavirus (host rice) PaEV: *Persea americana* endornavirus PhEv1: *Phytophthora* endornavirus, (host: *Douglas fir*) TaEV: *Tuber aestivum* endornavirus (host: summer truffle), BPEV: *bell pepper* endornavirus (host: Bell pepper), PvEv1 PVEv-2: *Phaseolus vulgaris* endornavirus (host common bean), GEEV: *grapevine endophyte* endornavirus (host:Grapevine) CYLV: *Carrot Yellow leaf curl virus*
Table 1. Percentage nucleotide sequence similarity between 15 *Endornavirus* genomes and the carrot yellow leaf curl virus as an outgroup. Highlighted in are *Endornavirus* sequences from this study compared to GenBank isolates.
Table 1: Percentage nucleotide sequence similarity between 15 *Endornavirus* genomes and the carrot yellow leaf curl virus as an outgroup. Highlighted in are *Endornavirus* sequences from this study compared to GenBank isolates.

<table>
<thead>
<tr>
<th></th>
<th>BaEV_AB844264.1</th>
<th>OeEV_NC_007647.1</th>
<th>OrEV_AB014344.1</th>
<th>PvEv1_NC023_678.1_Brazil</th>
<th>PvEv1_SR_F2_Kenya</th>
<th>PvEv1_SR_F36_Kenya</th>
<th>PvEv1_KHR456_287.1_Kenya</th>
<th>PaEV_JN880414.1</th>
<th>BPEV_NC_015781.2</th>
<th>PvEv2_AB719_SR_F2_Kenya</th>
<th>PvEv2_SR_F36_Kenya</th>
<th>PvEv2_KHR456_288.1_Brazil</th>
<th>PhEv1_NC_007069.1</th>
<th>GEEV_NC_019493.1</th>
<th>CYLV_NC_013007.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>BaEV_AB844264.1</td>
<td>45.5</td>
<td>44.1</td>
<td>45.3</td>
<td>44.4</td>
<td>44.2</td>
<td>44.2</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
</tr>
<tr>
<td>OeEV_NC_007647.1</td>
<td>45.5</td>
<td>44.1</td>
<td>45.3</td>
<td>44.4</td>
<td>44.2</td>
<td>44.2</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
</tr>
<tr>
<td>OrEV_AB014344.1</td>
<td>45.3</td>
<td>44.4</td>
<td>45.3</td>
<td>44.2</td>
<td>44.2</td>
<td>44.2</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
</tr>
<tr>
<td>PvEv1_SR_F2_Kenya</td>
<td>45.3</td>
<td>44.4</td>
<td>44.2</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
</tr>
<tr>
<td>PvEv1_SR_F36_Kenya</td>
<td>45.2</td>
<td>44.3</td>
<td>44.2</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
<td>49.3</td>
</tr>
<tr>
<td>PhEv1_NC_007069.1</td>
<td>31.9</td>
<td>32.4</td>
<td>32.1</td>
<td>31.6</td>
<td>31.6</td>
<td>31.6</td>
<td>31.6</td>
<td>31.6</td>
<td>31.6</td>
<td>31.6</td>
<td>31.6</td>
<td>31.6</td>
<td>31.6</td>
<td>31.6</td>
<td>31.6</td>
</tr>
<tr>
<td>CYLV_NC_013007.1</td>
<td>26.3</td>
<td>27</td>
<td>26.9</td>
<td>27.3</td>
<td>27.2</td>
<td>27.2</td>
<td>27.2</td>
<td>27.2</td>
<td>27.2</td>
<td>27.2</td>
<td>27.2</td>
<td>27.2</td>
<td>27.2</td>
<td>27.2</td>
<td>27.2</td>
</tr>
</tbody>
</table>