



Genomic characterisation and evolutionary relationships of groundnut rosette virus from the western highlands of Kenya

James M. Wainaina¹ · Jagger Harvey² · Elijah Ateka³ · Timothy Makori³ · David Karanja⁴ · Monica A. Kehoe⁵ · Laura M. Boykin¹

Received: 8 February 2018 / Accepted: 4 July 2018 / Published online: 9 August 2018
© Sociedade Brasileira de Fitopatologia 2018

Abstract

Viral symptomatic groundnut (*Arachis hypogaea* L.) leaf samples were collected in the heterogeneous agro-ecosystem of the western highlands of Kenya. High throughput RNA sequencing (RNA-Seq) was carried out on total plant RNA using the Illumina HiSeq platform. Subsequently, *de novo* assembly and sequence similarity searches identified the complete genome of *Groundnut rosette virus* (GRV). The genome consisted of 4298 nucleotides, encoding three open reading frames (ORF1-ORF3) including an RNA-dependent RNA polymerase gene. A sequence similarity search against GenBank reference sequences matched with 84% identity to a Malawian isolate (Z69910). Bayesian evolutionary relationships using the ORF3 and ORF4 genomic regions clustered the Kenyan isolates within a distinct clade. This study provides the first complete genome of GRV from Kenya and provides a genomic resource for the development of molecular diagnostic tools.

Keywords *Groundnut rosette virus* · *Tombusviridae* · Next-generation sequencing · Sub-Saharan Africa

Groundnuts are predominantly grown in semiarid, tropic and subtropic ecoregions. In sub-Saharan Africa (SSA) groundnuts are grown for their nutritional content, providing proteins and vitamins to humans (Naidu et al. 1999) as well as a raw material in animal feeds (Kana et al. 2013). The cultivation of groundnuts is mainly for subsistence purposes within smallholder farms with no irrigation practiced. One of the limitations of production are viral diseases, in particular groundnut

rosette. *Groundnut rosette virus* (GRV) is a positive, single stranded RNA virus classified within the genus *Umbravirus* (King et al. 2012). GRV is closely associated with a satellite RNA (sat-RNA) also referred to as helper RNA. The sat-RNA assists in the encapsulation of GRV into a coat protein that is lacking from GRV. This allows for the persistent transmission of GRV by the aphid *Aphis craccivora* (Naidu et al. 1999). Interestingly, GRV is confined to the SSA region (Waliyar et al. 2007; Okello et al. 2017). The estimated losses attributed to groundnut rosette disease caused by the combination of GRV, sat-RNA and *Groundnut rosette assistor virus* (GRAV, family *Luteoviridae*) range from \$5 to \$250 million USD per year in Nigeria and Zambia respectively (Waliyar et al. 2007). However, to date there have been limited efforts or opportunities to obtain the genomes of GRV. An increase in the availability of the GRV genomes in public databases would facilitate the development of molecular diagnostic tools for use by diagnosticians and plant breeders. In this study, we carried out two field surveys in the western highlands of Kenya across two seasons (2015/2016).

Groundnut leaf samples with chlorotic rosette-like symptoms were collected and stored in silica gel, and subsequently transported to the laboratory for analysis. Total RNA extraction was carried out using the Zymo RNA Miniprep kit (Zymo Research) according to the manufacturer's instructions.

Section Editor: Michael Goodin

✉ James M. Wainaina
james.wainaina@research.uwa.edu.au; jwmbora@yahoo.com

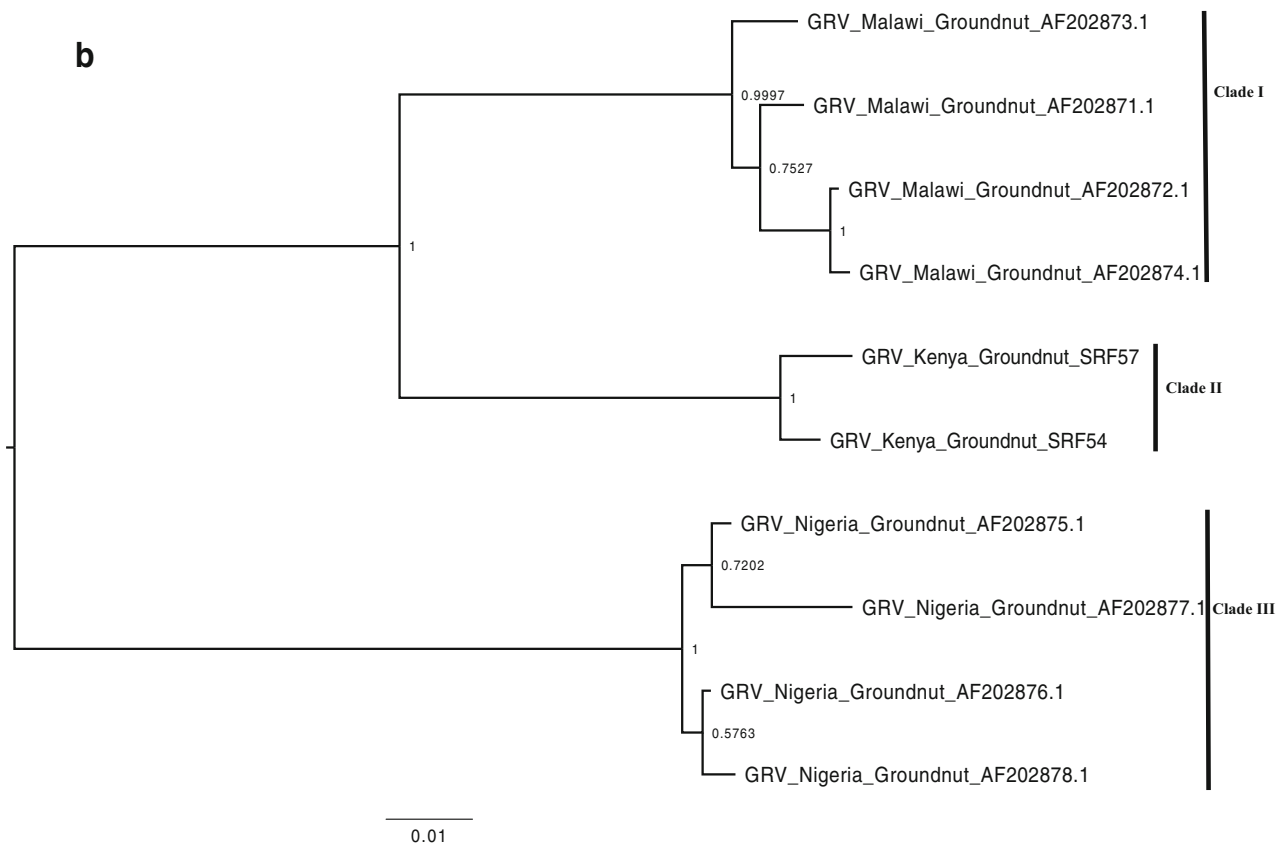
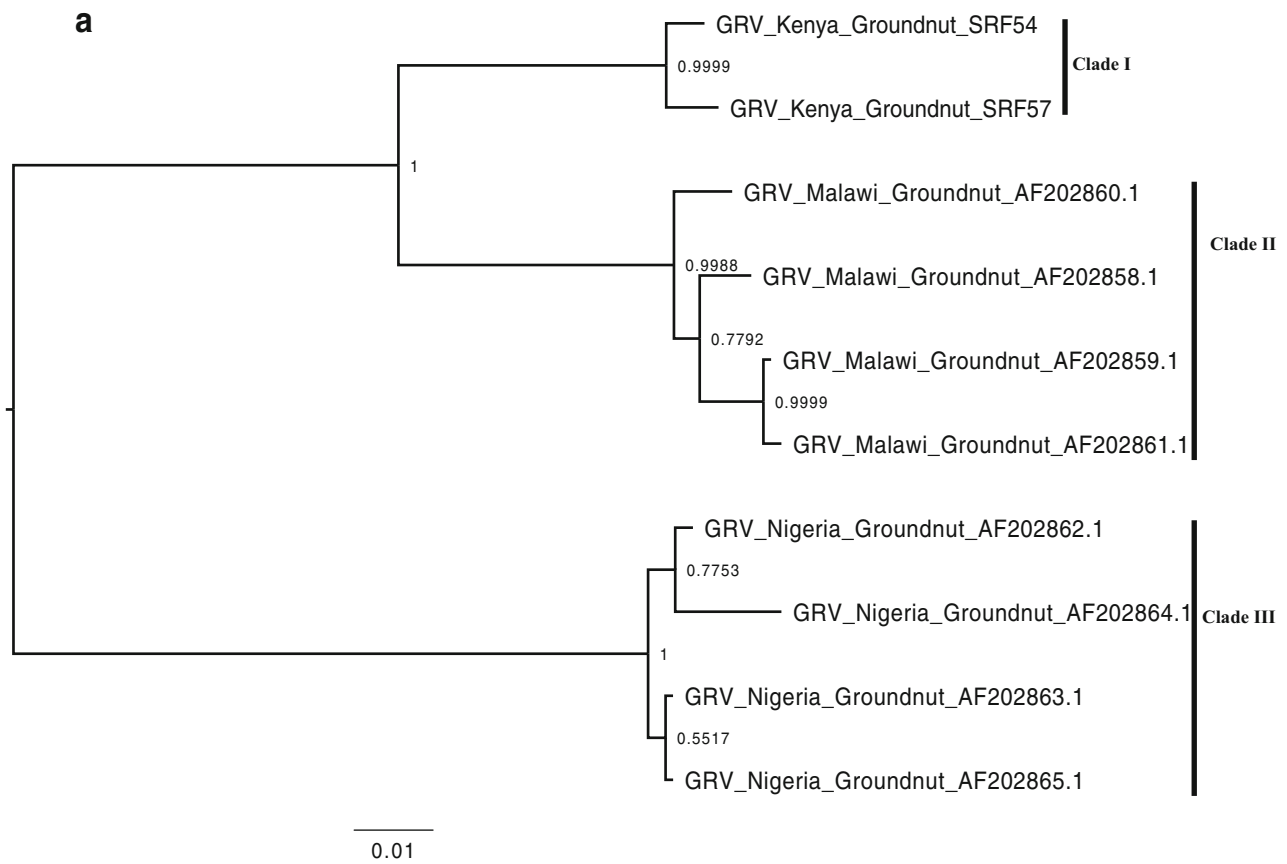
¹ School of Molecular Sciences and Australian Research Council Centre of Excellence in Plant Energy Biology, University of Western Australia, Crawley, Perth, WA 6009, Australia

² Feed the Future Innovation Lab for the Reduction of Post-Harvest Loss, Kansas State University, Manhattan, Kansas, USA

³ Department of Horticulture, Jomo Kenyatta University of Agriculture and Technology, P.O. Box 62000-00200, Nairobi, Kenya

⁴ Kenya Agricultural and Livestock Research Organization (KARLO), Box 340-90100, Machakos, Kenya

⁵ Department of Primary Industries and Regional Development, Diagnostic and Laboratory Service, South Perth, WA 6151, Australia



◀ **Fig. 1** Bayesian phylogenetic relationships of groundnut rosette virus (GRV) using nucleotide sequences of the open reading frames 3 (a) and 4 (b). Clade I, groundnut rosette virus (GRV) Malawian isolates; clade II, GRV Kenyan isolates; clade III, GRV Nigerian isolates. Accession numbers for GRV Kenyan isolates: SFR 54, MG646923; SRF 57, MG646922

Individual samples were used for cDNA library preparation using the Illumina Truseq Stranded Total RNA kit (Illumina) with plant ribozero depletion. Each library was dual indexed with unique adaptors and subsequently enriched with a PCR stage. Libraries were subsequently quantified and the correct insert size determined using the Tape Station Agilent 2200 (Agilent) using a D1000 screen tape and ladder (Agilent). Libraries with correct insert size and concentration were normalised, pooled and sequenced on rapid mode run using a single flow cell on the Illumina HiSeq 2500 at Macrogen Korea. We obtained a final genome of 4298 nt in length with an average coverage of 4487 times. The complete genome was deposited in GenBank (accession number MG646922). Assembly of this genome was from a total of 13,439,450 reads obtained after trimming to remove low quality bases. *De novo* assembly of the trimmed reads was carried out in CLC genomic workbench 8.5.1 (CLCGW) using set parameters as previously described (Kehoe et al. 2014). All contigs were subjected to NCBI BLASTn searches. GRV contigs were subsequently mapped to a reference genome (Z69910) using Geneious v. 8.1.8 (Biomatters). In addition, trimmed reads were imported into Geneious and mapped against reference GenBank sequences. Mapping parameters were set as follows: minimum overlap 10%, minimum overlap identity 80%, allow gaps 10% with fine tuning, iteration of up to 25 times. To confirm that the genome sequence was complete we compared the genomes obtained against those in GenBank. A partial genome was also generated in this study (SRF 54; MG646923) with 10,525,186 reads after trimming and only the ORF 3 and ORF 4 were used for phylogenetic analysis.

Sequence similarity search using BLASTn showed nucleotide sequence similarity of 84% to GRV isolates from Malawi (Z69910). Species demarcation within the genus *Umbravirus* is based on natural host range and a percent nucleotide identity of less than 70% (King et al. 2012). Based on these criteria, the GRV isolate from Kenya was considered to be a member of the *Groundnut rosette virus* species in genus *Umbravirus*. The complete genome length of 4298 nucleotides was larger than current reference genomes of 4019 nucleotides from GenBank. The extra 279 bases were found on the 3' end outside the main ORF. Evolutionary relationships of GRV isolates from this study and reference sequences from GenBank were based on nucleotide sequences of ORF3 and

ORF4. These two gene regions are the main markers used in the molecular diagnosis of GRV. Bayesian phylogenetic relationship analysis was carried out using the optimal evolutionary model GTR + I + G on MrBayes 3.2.2 (Ronquist et al. 2012) for 50 million generations, with sampling every 1000 generations with 25% of the sub optimal trees discarded. The consensus tree was viewed in FigTree v1.4.2. The sequences formed distinct clades based on geographical sampling locations (Fig. 1). These genome resources will have a direct impact in the development of molecular diagnostic tools in particular within SSA, where the virus remains pandemic.

Acknowledgments An Australian Award Scholarship by the Department of Foreign Affairs and Trade (DFAT) supports JW, and this work forms part of his PhD research. Fieldwork activities were coordinated through the Cassava Diagnostic project Kenyan node. A Rising Star grant from the Faculty of Science alumni at the University of Western Australia to LMB provided laboratory reagents and consumables. Pawsey Supercomputing Centre provided supercomputer resources for data analysis with funding from the Australian Government and the Government of Western Australia.

References

- Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61:539–542. <https://doi.org/10.1093/sysbio/sys029>
- Kana JR, Gnonlonfin BGJ, Harvey J, Wainaina J, Wanjuki I, Skilton RA, Tegui A (2013) Assessment of aflatoxin contamination of maize, peanut meal and poultry feed mixtures from different agroecological zones in Cameroon. *Toxins* 5:884–894
- Kehoe MA, Coutts BA, Buirchell BJ, Jones RAC (2014) Plant virology and next generation sequencing: experiences with a potyvirus. *PLoS One* 9
- King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (2012) *Virus taxonomy*. Ninth report of the international committee on taxonomy of viruses. Elsevier/Academic Press, London
- Naidu RA, Kimmins FM, Deom CM, Subrahmanyam P, Chiyembekeza AJ, Van der Merwe PJ (1999) Groundnut rosette: a virus disease affecting groundnut production in sub-Saharan Africa. *Plant Disease* 83:700–709
- Okello DK, Ugen MA, Tukamuhabwa P, Ochwo M, Odong TL, Adriko J, Kiconco F, Male A, Deom CM (2017) Molecular diagnostics of groundnut rosette disease agents in Uganda: implications on epidemiology and management of groundnut rosette disease. *Journal of Plant Breeding and Crop Sciences* 9:63–70
- Waliyar F, Kumar PL, Ntare BR, Monyo E, Nigam SN, Reddy AS, Osiru M, Diallo, AT (2007) A century of research on groundnut rosette disease and its management. *Information Bulletin no. 75*. International Crops Research Institute for the Semi-Arid Tropics, Patancheru