Diversity of viruses and viroids in the rhizosphere of common bean cultivars differing in resistance to the fungal root pathogen *Fusarium oxysporum*

Lucas P P Braga, Andrew J Tanentzap, Benjamin Lee, Siu Mui Tsai, Jos M Raaijmakers, Rodrigo Mendes and Lucas W Mendes

Supplementary Information

Material and Methods

Greenhouse experiment

To obtain rhizosphere soil, we conducted a greenhouse experiment at the Center for Nuclear Energy in Agriculture, University of Sao Paulo (CENA/USP). We used three common bean cultivars with contrasting levels of resistance to *Fusarium oxysporum*: resistant IAC Milenio (Carbonell et al., 2014), moderately resistant BRS Estilo (Melo et al., 2010) and susceptible IAC Alvorada (Carbonell et al., 2008). According to the breeding program that generated the cultivars, resistance is inferred to be associated with restriction of fungal colonization by chemical and structural barriers favored by root exudate metabolites (Carbonell et al., 2014, 2008; Melo et al., 2010). The plants were grown in mesocosms made of ceramic pots (30 cm high x 20 cm diameter) with a stone layer (5 cm) on the bottom and were filled with 8 kg of soil. We used Amazon Dark Earth soils, which are anthropogenic horizons built-up by the Indigenous Pre-Colombian populations (c.a. 500-8,700 years ago). These soils have high fertility and microbial diversity (Brossi et al., 2014), and are composed of a mean (±SE) 573.3±10 $g.kg^{-1}$ sand, 91±5.5 $g.kg^{-1}$ silt, 335.7±13.3 $g.kg^{-1}$ clay, pH 4.8±0.3, organic matter 54.7±0.3 $g.kg^{-1}$ and total nitrogen 2758±144.6 $mg.kg^{-1}$. To simulate realistic field conditions for rhizosphere selection of plant-microbe interactions, we did not inoculate *Fusarium* in the experiment. Each cultivar was grown in independent pots (n=3) with three common bean seeds sowed in each pot. Seeds germinated at 28/19ºC (day/night) with a 12 h photoperiod. Moisture and temperature were constantly monitored to ensure optimal growth conditions. Plants were harvested at the early flowering stage (i.e. R1). The roots were carefully removed from the pots and transported on ice to be processed in the same day at the CENA/USP. The roots were firstly shaken to remove all loosely adhering soil. Then, the firmly attached rhizosphere soil was collected with sterile brushes and immediately frozen at -80ºC.

RNA extraction and sequencing

RNA extraction of 2 g of rhizosphere soils per replicate was performed using the PowerSoil Total RNA Isolation Kit (MoBio, Carlsbad, USA) according to the manufacturer's instructions. Quality and concentration of extracted RNA were determined on an Agilent 2100 Bioanalyzer (Agilent Technologies, Germany). Next, we used a Ribo-Zero rRNA Removal Kit (Illumina, USA) for ribossomal RNA depletion. The libraries were prepared with TruSeq Stranded mRNA Kit (Illumina) and sequenced on a HiSeq 2500 (Illumina), producing nine metatrascriptomes (3 cultivars x 3 replicates). Metatranscriptome sequencing produced on average (±SE) 19,163,158±428,331 reads after quality control considering an average quality of Q35±3.

Viral community analysis

Metatranscriptomes were quality filtered according to the JGI RQCFilter pipeline (Arkin et al., 2018) and coassembled for the three replicates of each of the three cultivars with IDBA-UD in KBase (Arkin et al., 2018; Peng et al., 2012). The contigs (>3,000bp) were then screened with VirSorter2 (Guo et al., 2021), Seeker (Auslander et al., 2020) and DeepVirFinder (Ren et al., 2017) to detect viral sequences. The classified contigs were clustered at >95% average nucleotide identity across 80% of the length of the shorter contig (Roux and Emerson, 2022) using GenomeCluster at CyVerse (https://cyverse.org). Only sequences classified by at least two of these tools were considered further. The classified sequences were then subjected to a further quality check following the standard operational procedure (dx.doi.org/10.17504/protocols.io.btv8nn9w) using CheckV (Nayfach et al., 2021) and manually screening the annotations of functional domains that they encoded. A functional domains annotation analysis was performed in KBase (Arkin et al., 2018). We additionally ran Palmscan to classify viral sequences representing RNA viruses based on the presence of viral RNA-dependent RNA polymerase (RdRp) (Edgar et al., 2022). We used RdRp amino acid sequences to search for similarity against the NCBI Virus database (Hatcher et al., 2017) and infer taxonomic classification and putative host prediction based on the best hits that were retrieved (see Supplementary Table S1). Viral sequences not encoding RdRp could represent either transcripts of DNA viruses or fragmented genomes of RNA viruses, so we used the whole genome in the similarity search for these sequences. Viroid sequences were identified and classified according to their ribozyme annotations using Vdsearch (Lee et al., 2022). The read mapping step using the annotated viral and viroid sequences as references was performed using bbmap (https://sourceforge.net/projects/bbmap/), considering matches with >95% of sequence identity and 70% of sequence coverage, and counts were normalized by sequence length and library depth for relative abundance.

Statistical analysis

We tested for pairwise differences in viral and viroid communities among the cultivars. For this purpose, we used the linear discriminant analysis effect size (LEfSE) algorithm (Segata et al., 2011), which is specifically designed to detect changes in relative abundance among groups of microbiome data (Segata et al., 2011). For each feature (i.e., virus or viroid sequence), the LEfSE algorithm first assesses if the observed abundances are differentially distributed between pairwise groups using a Kruskal-Wallis sum-rank test ($\alpha = 0.05$). Next, the features that were detected to be differentially abundant are subjected to a linear discriminant analysis (LDA). The LDA estimates the effect size of each differentially abundant feature. The threshold LDA score to identify differential features and the normalization of relative abundance were set to the default parameters in the LEfSE software (Segata et al., 2011). To gain further insight into virus-viroid interactions in the rhizosphere, we used a network co-occurrence approach. We inter-correlated the relative abundances of all viruses and viroids and adjusted for multiple comparisons with Bonferroni correction ($p = 0.0004$), which meant that only absolute Spearman rank correlations >0.99 were considered statistically significant. Network graphs were analyzed using the R package 'igraph'. We used betweenness centrality and degree to assess network stability. Betweenness centrality measures the number of shortest paths between pairs of vertices that pass through a given vertex (Freeman et al., 1991) A node with high betweenness centrality has a high control over the flow of information in the network (Kirkley et al., 2018; Mahyar et al., 2018; Röttjers and Faust, 2018). Degree measures the number of edges connected to a node. A node with high degree is considered a hub in the network (Röttjers and Faust, 2018). Nodes with high betweenness centrality and degree are the ones most impacting network structure and, therefore, can be considered as keystone (Barrat et al., 2004; Röttjers and Faust, 2018). We selected the top-10 largest values for betweenness centrality and degree as cut-off for consistent comparison across cultivar-specific networks. Statistically significant correlations were visualized with Cytoscape 3.8 using the CoSe layout (Shannon et al., 2003).

References

Arkin, A.P., Cottingham, R.W., Henry, C.S., Harris, N.L., Stevens, R.L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., Sneddon, M.W., Henderson, M.L., Riehl, W.J., Murphy-Olson, D., Chan, S.Y., Kamimura, R.T., Kumari, S., Drake, M.M., Brettin, T.S., Glass, E.M., Chivian, D., Gunter, D., Weston, D.J., Allen, B.H., Baumohl, J., Best, A.A., Bowen, B., Brenner, S.E., Bun, C.C., Chandonia, J.-M., Chia, J.-M., Colasanti, R., Conrad, N., Davis, J.J., Davison, B.H., DeJongh, M., Devoid, S., Dietrich, E., Dubchak, I., Edirisinghe, J.N., Fang, G., Faria, J.P.,

Frybarger, P.M., Gerlach, W., Gerstein, M., Greiner, A., Gurtowski, J., Haun, H.L., He, F., Jain, R., Joachimiak, M.P., Keegan, K.P., Kondo, S., Kumar, V., Land, M.L., Meyer, F., Mills, M., Novichkov, P.S., Oh, T., Olsen, G.J., Olson, R., Parrello, B., Pasternak, S., Pearson, E., Poon, S.S., Price, G.A., Ramakrishnan, S., Ranjan, P., Ronald, P.C., Schatz, M.C., Seaver, S.M.D., Shukla, M., Sutormin, R.A., Syed, M.H., Thomason, J., Tintle, N.L., Wang, D., Xia, F., Yoo, H., Yoo, S., Yu, D., 2018. {KBase}: The United States Department of Energy Systems Biology Knowledgebase. Nat. Biotechnol. 36, 566–569.

Auslander, N., Gussow, A.B., Benler, S., Wolf, Y.I., Koonin, E. V, 2020. Seeker: alignment-free identification of bacteriophage genomes by deep learning. Nucleic Acids Res. 48, e121–e121. https://doi.org/10.1093/nar/gkaa856

Barrat, A., Barthélemy, M., Pastor-Satorras, R., Vespignani, A., 2004. The architecture of complex weighted networks. Proc. Natl. Acad. Sci. 101, 3747–3752. https://doi.org/10.1073/pnas.0400087101

Brossi, M.J. de L., Mendes, L.W., Germano, M.G., Lima, A.B., Tsai, S.M., 2014. Assessment of Bacterial bph Gene in Amazonian Dark Earth and Their Adjacent Soils. PLoS One 9, e99597. https://doi.org/10.1371/journal.pone.0099597

Carbonell, S.A.M., Chiorato, A.F., Bolonhezi, D., Barros, V.L.N.P. de, Borges, W.L.B., Ticelli, M., Gallo, P.B., Finoto, E.L., Santos, N.C.B. dos, 2014. “IAC Milênio” - Common bean cultivar with high grain quality. Crop Breed. Appl. Biotechnol. 14, 273–276. https://doi.org/10.1590/1984-70332014v14n4c44

Carbonell, S.A.M., Chiorato, A.F., Ito, M.F., Perina, E.F., Gonçalves, J.G.R., Souza, P.S., Gallo, P.B., Ticelli, M., Colombo, C.A., Azevedo Filho, J.A., 2008. IAC-Alvorada and IAC-Diplomata: new common bean cultivars. Crop. Breed. Appl. Biotechnol. 8, 163–166. https://doi.org/10.12702/1984-7033.v08n02a10

Edgar, R.C., Taylor, J., Lin, V., Altman, T., Barbera, P., Meleshko, D., Lohr, D., Novakovsky, G., Buchfink, B., Al-Shayeb, B., Banfield, J.F., de la Peña, M., Korobeynikov, A., Chikhi, R., Babaian, A., 2022. Petabase-scale sequence alignment catalyses viral discovery. Nature 602, 142–147. https://doi.org/10.1038/s41586-021-04332-2

Freeman, L.C., Borgatti, S.P., White, D.R., 1991. Centrality in valued graphs: A measure of betweenness based on network flow. Soc. Networks 13, 141–154. https://doi.org/10.1016/0378-8733(91)90017-N

Guo, J., Bolduc, B., Zayed, A.A., Varsani, A., Dominguez-Huerta, G., Delmont, T.O., Pratama, A.A., Gazitúa, M.C., Vik, D., Sullivan, M.B., Roux, S., 2021. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome 9, 37.

https://doi.org/10.1186/s40168-020-00990-y
Hatcher, E.L., Zhdanov, S.A., Bao, Y., Blinkova, O., Nawrocki, E.P., Ostapchuck, Y., Schäffer, A.A., Brister, J.R., 2017. Virus Variation Resource – improved response to emergent viral outbreaks. Nucleic Acids Res. 45, D482–D490. https://doi.org/10.1093/nar/gkw1065
Kirkley, A., Barbosa, H., Barthelemy, M., Ghoshal, G., 2018. From the betweenness centrality in street networks to structural invariants in random planar graphs. Nat. Commun. 9, 2501. https://doi.org/10.1038/s41467-018-04978-z
Lee, B.D., Neri, U., Roux, S., Wolf, Y.I., Camargo, A.P., Krupovic, M., Simmonds, P., Kyrpides, N., Gophna, U., Dolja, V. V, Koonin, E. V, 2022. A vast world of viroid-like circular RNAs revealed by mining metatranscriptomes. bioRxiv 2022.07.19.500677. https://doi.org/10.1101/2022.07.19.500677
Mahyar, H., Hasheminezhad, R., Ghalebi K., E., Nazemian, A., Grosu, R., Movaghar, A., Rabiee, H.R., 2018. Compressive sensing of high betweenness centrality nodes in networks. Phys. A Stat. Mech. its Appl. 497, 166–184. https://doi.org/10.1016/j.physa.2017.12.145
Melo, L.C., Peloso, M.J. Del, Pereira, H.S., Faria, L.C. de, Costa, J.G.C. da, Díaz, J.L.C., Rava, C.A., Wendland, A., Abreu, Â. de F.B., 2010. BRS Estilo: common bean cultivar with Carioca grain, upright growth and high yield potential. Crop Breed. Appl. Biotechnol. 10, 377–379. https://doi.org/10.1590/S1984-70332010000400015
Nayfach, S., Camargo, A.P., Schulz, F., Eloe-Fadrosh, E., Roux, S., Kyrpides, N.C., 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat. Biotechnol. 39, 578–585. https://doi.org/10.1038/s41587-020-00774-7
Peng, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28, 1420–1428. https://doi.org/10.1093/bioinformatics/bts174
Ren, J., Ahlgren, N.A., Lu, Y.Y., Fuhrman, J.A., Sun, F., 2017. {VirFinder}: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. Microbiome 5, 69.
Röttjers, L., Faust, K., 2018. From hairballs to hypotheses–biological insights from microbial networks. FEMS Microbiol. Rev. 42, 761–780. https://doi.org/10.1093/femsre/fuy030
Roux, S., Emerson, J.B., 2022. Diversity in the soil virosphere: to infinity and beyond? Trends Microbiol. 30, 1025–1035. https://doi.org/https://doi.org/10.1016/j.tim.2022.05.003
Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., Huttenhower, C., 2011. Metagenomic biomarker discovery and explanation. Genome Biol. 12, R60. https://doi.org/10.1186/gb-2011-12-6-r60
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski,

B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504.