Arch Virol (2018) 163:263–267
https://doi.org/10.1007/s00705-017-3571-2

ANNOTATED SEQUENCE RECORD

Discovery of the first maize-infecting mastrevirus in the Americas using a vector-enabled metagenomics approach

Rafaela S. Fontenele[1,2] · Dione M. T. Alves-Freitas[1] · Pedro I. T. Silva[1] · Josemar Foresti[3] · Paulo. R. Silva[3] · Márcio T. Godinho[1] · Arvind Varsani[2,4] · Simone G. Ribeiro[1]

Received: 4 June 2017 / Accepted: 4 September 2017 / Published online: 27 September 2017

Abstract The genus *Mastrevirus* (family *Geminiviridae*) is composed of single-stranded DNA viruses that infect mono- and dicotyledonous plants and are transmitted by leafhoppers. In South America, there have been only two previous reports of mastreviruses, both identified in sweet potatoes (from Peru and Uruguay). As part of a general viral surveillance program, we used a vector-enabled metagenomics (VEM) approach and sampled leafhoppers (*Dalbulus maidis*) in Itumbiara (State of Goiás), Brazil. High-throughput sequencing of viral DNA purified from the leafhopper sample revealed mastrevirus-like contigs. Using a set of abutting primers, a 2746-nt circular genome was recovered. The circular genome has a typical mastrevirus genome organization and shares <63% pairwise identity with other mastrevirus isolates from around the world. Therefore, the new mastrevirus was tentatively named "maize striate mosaic virus". Seventeen maize leaf samples were collected in the same field as the leafhoppers, and ten samples were found to be positive for this mastrevirus. Furthermore, the ten genomes recovered from the maize samples share >99% pairwise identity with the one from the leafhopper. This is the first report of a maize-infecting mastrevirus in the Americas, the first identified in a non-vegetatively propagated mastrevirus host in South America, and the first mastrevirus to be identified in Brazil.

Handling Editor: F. Murilo Zerbini.

Electronic supplementary material The online version of this article (doi:10.1007/s00705-017-3571-2) contains supplementary material, which is available to authorized users.

✉ Arvind Varsani
arvind.varsani@asu.edu

✉ Simone G. Ribeiro
simone.ribeiro@embrapa.br

[1] Embrapa Recursos Genéticos e Biotecnologia, Brasília, Brazil

[2] The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine School of Life Sciences, Arizona State University, Tempe, AZ, USA 85287

[3] Faculdade de Agronomia e Medicina Veterinária, Campus Darcy Ribeiro, Universidade de Brasília, Brasília, DF, Brasil

[4] Structural Biology Research Unit, Department of Clinical Laboratory Sciences, University of Cape Town, Observatory, Cape Town, South Africa

The use of viral metagenomics approaches and high-throughput sequencing (HTS) has led to a rapid increase in the discovery of novel plant viruses [37]. The vector-enabled metagenomics (VEM) approach is based on the isolation and sequencing of viral particles directly from the insect vectors. VEM coupled with HTS has become a powerful tool to characterize the viral diversity in a variety of ecosystems. The VEM approach was first used to identify viruses in whiteflies [34] and subsequently has been used on other insect vectors to identify both novel and known plant viruses [20, 35, 39–42]. HTS of total nucleic acids from plants has also lead to the discovery of diverse geminiviruses [2, 5, 26, 27]. Thus, these techniques are very attractive for viral surveillance and to identify viral communities circulating in ecosystems.

As part of a program of surveillance of plant-infecting DNA viruses in Brazil, we used a vector-enable metagenomics approach and sampled leafhoppers (*Dalbulus maidis,* Hemiptera: Cicadellidae) feeding on maize in Itumbiara (State of Goiás). The leafhoppers (~20 individuals) were homogenized in 500 µl of SM buffer (50 mM Tris-HCl, 10 mM $MgSO_4$, 0.1 M NaCl, pH 7.5). The homogenate was

centrifuged at 10000 × *g*, and the supernatant was filtered through a 0.22-µm syringe filter. The filtrate (200 µl) was used for extraction of viral nucleic acid using a ZR Viral DNA/RNA Kit™ (Zymo Research, USA). The viral DNA was amplified by rolling-circle amplification (RCA) using an Illustra TempliPhi™ Amplification Kit (GE Healthcare, USA). The RCA products were sequenced on an Illumina Nextseq500 sequencing platform (Illumina, USA) at the Genomics Core at the Biodesign Institute, Arizona State University (USA).

The Illumina sequencing paired-end reads were assembled *de novo* using ABySS 1.9 [45] with k-mer = 64. BLASTx [3] was used to analyze the *de novo*-assembled contigs. One hundred thirty-seven contigs were identified with similarities to mastreviruses, with the largest being 388 nt. Mastreviruses are plant-infecting circular single-stranded DNA viruses that are encapsidated in geminate particles [19]. The genus *Mastrevirus* is part of the family *Geminiviridae,* and its members are transmitted by leafhoppers (order Hemiptera, family Cicadellidae). Mastreviruses can infect either monocot or dicotyledonous plants, and their genomes encode at least four open reading frames (ORFs). Two of these, the movement protein (MP) and the capsid protein (CP), are encoded on the virion-sense strand, and the replication-associated protein (Rep) and the RepA are encoded on the complementary-sense strand.

A set of abutting primers were designed based on the largest mastrevirus-like (similarities to the CP gene sequences) contig (J455-F, 5' – ACC CTT CTT AAC TTC CAC CAC GGC AGA A – 3' and J455-R, 5' – GGT AAT TGT CTG ATG GTT ACC TCC TAC A – 3') to recover the full-length genome from the sample. The viral RCA product (0.5 µl) was used as template for amplification with the abutting primers using Kapa Hifi Hotstart DNA polymerase (KAPA Biosystems, USA) using the following thermal cycling conditions: 98 °C for 3 min, 25 cycles of 98 °C for 15 s, 60 °C for 15 s, 72 °C for 3 min and a final extension of 72 °C for 3 min. An amplicon of ~2.7 kb was resolved in a 0.7% agarose gel, excised and gel purified. The purified amplicon was cloned into pJET 1.2 plasmid (ThermoFisher Scientific, USA), Sanger sequenced by primer walking at Macrogen Inc. (Korea), and assembled using Geneious 10 [21].

The 2746-nt genome recovered from the leafhoppers contains the nonanucleotide motif (TAATATT↓AC) and has a typical mastrevirus genome organization [49]. The Rep amino acid sequence contains motifs I, II and II in its N-terminus, which are essential for rolling-circle replication [23, 38], and the GRS (geminivirus Rep sequence) domain [33]. In addition, closer to the C-terminal region, we identified the RBR interaction domain, which interacts with retinoblastoma-related protein in the host [4], the RXL motif, and the Walker A, Walker B and motif C regions, which bind to dNTP and have helicase activity [9, 15]. Comparison of the motifs found in the Brazilian mastrevirus with those of other mastreviruses revealed a variation in the GRS domain consisting of eight extra amino acids in the central portion of the motif. The GRS motif is known to be less conserved in mastreviruses than in other geminiviruses [33].

Since the leafhoppers were collected in the same field as the maize samples, the abutting primers J455-F/-R were used to screen and recover similar viruses from these plants by PCR using the same thermal cycling conditions as were used for recovery from the leafhoppers described above. Total DNA from the seventeen maize samples was individually extracted using the CTAB protocol [10]. The circular DNA was enriched using RCA, and screened using abutting primers, and the 2.7-kb PCR-generated amplicons were cloned using the same protocol as that used for the mastrevirus from the leafhoppers. Of the seventeen maize samples collected, ten were found to be infected with the new mastrevirus identified in the leafhopper.

Analysis of genome-wide pairwise identities using SDT v1.2 [31] revealed that the genomes of the mastrevirus from the ten maize plants and the one from the leafhopper share >99% identity. An analysis of representative genome sequences from members of different mastrevirus species showed that the new mastrevirus isolates from Brazil share <63% pairwise identity (Supplementary Data 1). Based on the currently accepted threshold of species demarcation for mastreviruses [30], the new Brazilian mastrevirus belong to a new species, and we have tentatively named this virus "maize striate mosaic virus" (MSMV) based on the mild striation and mosaic symptoms observed on the maize leaves of the 10 positive samples.

Three datasets were created that comprised 1) genome sequences, 2) Rep amino acid sequences, and 3) CP amino acid sequences of the 11 MSVMs from this study together with 38 representative sequences of mastreviruses and two sequences as outgroups (Eragrostis curvula streak virus, ECSV, FJ665630; beet curly top Iran virus, BCTIV, EU273816). The datasets were aligned individually using MUSCLE [11]. The genome sequence alignment was used to infer a neighbor-joining (NJ) phylogenetic tree with the Jukes-Cantor nucleotide substitution model and 1000 bootstrap iterations in MEGA5 [47]. Branches with less than 60% branch support were collapsed in TreeGraph2 [46]. The Rep and CP amino acid sequence alignments were used to infer maximum-likelihood (ML) phylogenetic trees using approximate likelihood ratio test (aLRT) with the model LG+I+G and LG+I+G+F for Rep and CP, respectively. All branches with <80% aLRT branch support were collapsed using TreeGraph2 [46]. The NJ phylogenetic tree of the genome sequences and the ML phylogenetic trees of the

Rep and CP sequences show that the MSMV sequences group with the monocot-infecting mastreviruses (Fig. 1). The Rep sequences of MSMVs share <54% and the CP sequences share <49% identity with those of other mastreviruses (Supplementary Data 1).

Here, we report the first monocot-infecting mastrevirus to be identified South America (and the first mastrevirus in Brazil), which was found infecting maize using a vector-enabled metagenomics approach. This finding is very important, since mastreviruses have been known to be involved in plant disease outbreaks throughout the world. Maize streak disease (MSD) associated with the mastrevirus maize streak virus A causes high yield losses in maize crops in Africa [29, 36, 44]. In addition, mastreviruses cause damage to other economically important cultivated plants such as wheat, chickpeas, lentils and oats [25, 32, 43, 48].

Monocot-infecting mastreviruses in the Americas have been identified in switchgrass in the USA [1] and in sugarcane germplasm in the USA (Florida) and the Caribbean (Barbados and Guadalupe) [6, 7]. Nonetheless, mastreviruses sequences in the New World (partial sequences) were first identified in South America using HTS of small RNAs extracted from Peruvian sweet potatoes [24]. More recently, a mastrevirus, named sweet potato symptomless virus (SpSV), has been identified in sweet potatoes from various countries, including Uruguay [8].

Maize is an economically important crop in Brazil and has already been associated with viral diseases caused by strains of potyvirus sugar cane mosaic virus [13], viruses from the family *Luteoviridae*, the luteovirus barley yellow dwarf virus-MAV [28], a new putative polerovirus, maize yellow mosaic virus [14], and a rhabdovirus, maize mosaic virus [22]. Moreover, a leafhopper-transmitted virus with

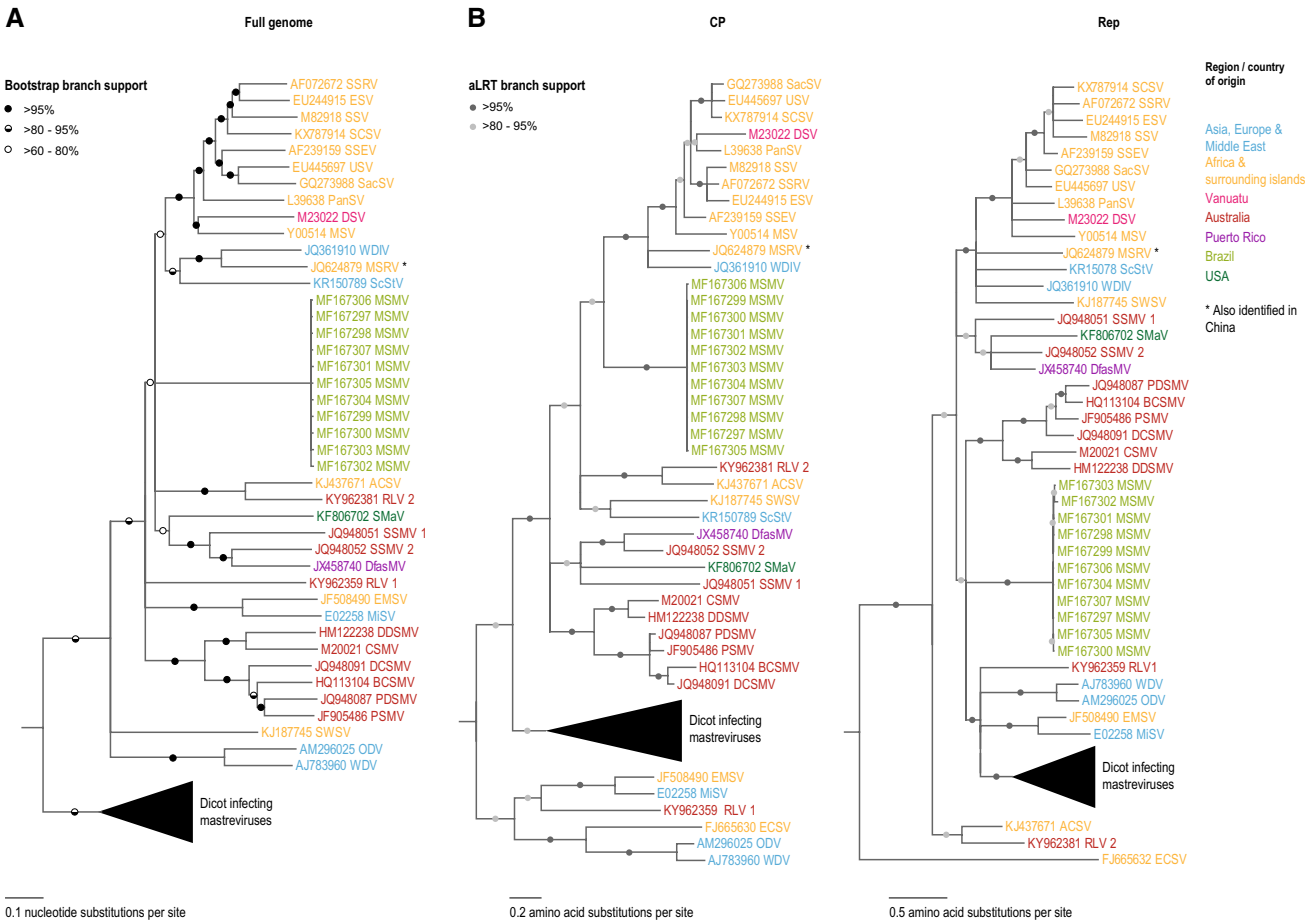

Fig. 1 (A) Neighbor-joining phylogenetic tree of the complete genome sequences of the MSMVs and representative mastreviruses rooted with ECSV (FJ665630) and BCTIV (EU273816) sequences. Branches with <60% bootstrap support have been collapsed. (B) Maximum-likelihood (ML) phylogenetic tree of the CP and Rep amino acid sequences of the MSMVs and representative mastreviruses using the models LG+I+G+F and LG+I+G for the CP and Rep, respectively, with the approximate-likelihood test (aLRT). The ML phylogenetic trees were rooted with the BCTIV (EU273816) sequence, and branches with <80% aLRT support have been collapsed

a positive ssRNA genome, marafivirus mayze rayado fino virus (family *Tymoviridae*) is a component of the corn stunt disease complex, which causes high yield losses to maize crops in Brazil [16–18].

The VEM approach has proven to be useful in viral discovery; e.g., the first mastrevirus identified in the Caribbean was identified from VEM of dragonflies, which are known insect predators [39]. Moreover, HTS was used to identify other geminivirus that had never been previously reported in Brazil [12]. The discovery of the first mastrevirus in Brazil through a VEM approach attests to the importance of this technique, not only for the discovery of novel viruses but also as a monitoring system to prevent as well as to discover new viruses that could potentially lead to plant disease outbreaks.

Acknowledgements This research was partly supported by internal Grants from Embrapa (Brazil) and a Grant (314732/2014-4) from CNPq (Brazil). RSF, DMTAF, MTG were supported by fellowships from CNPq.

GenBank accession numbers MF167297–MF167307.

Compliance with ethical standards

Conflicts of interest The authors declare there are no conflicts of interest.

Research involving human participants and/or animals The research did not involve human participants or animals.

Informed consent The research did not involve human participants or animals.

Data The data used for the analyses described in this manuscript are publicly available in the GenBank database.

References

1. Agindotan BO, Domier LL, Bradley CA (2015) Detection and characterization of the first North American mastrevirus in switchgrass. Arch Virol 160:1313–1317
2. Al Rwahnih M, Alabi OJ, Westrick NM, Golino D, Rowhani A (2017) Description of a novel monopartite geminivirus and its defective subviral genome in grapevine. Phytopathology 107:240–251
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
4. Arguello-Astorga G, Lopez-Ochoa L, Kong LJ, Orozco BM, Settlage SB, Hanley-Bowdoin L (2004) A novel motif in geminivirus replication proteins interacts with the plant retinoblastoma-related protein. J Virol 78:4817–4826
5. Bernardo P, Muhire B, Francois S, Deshoux M, Hartnady P, Farkas K, Kraberger S, Filloux D, Fernandez E, Galzi S, Ferdinand R, Granier M, Marais A, Monge Blasco P, Candresse T, Escriu F, Varsani A, Harkins GW, Martin DP, Roumagnac P (2016) Molecular characterization and prevalence of two capulaviruses: Alfalfa leaf curl virus from France and *Euphorbia caput*-medusae latent virus from South Africa. Virology 493:142–153
6. Boukari W, Alcalá-Briseño RI, Kraberger S, Fernandez E, Filloux D, Daugrois J-H, Comstock JC, Lett J-M, Martin DP, Varsani A (2017) Occurrence of a novel mastrevirus in sugarcane germplasm collections in Florida, Guadeloupe and Réunion. Virol J 14:146
7. Candresse T, Filloux D, Muhire B, Julian C, Galzi S, Fort G, Bernardo P, Daugrois J-H, Fernandez E, Martin DP (2014) Appearances can be deceptive: revealing a hidden viral infection with deep sequencing in a plant quarantine context. PloS ONE 9:e102945
8. Cao M, Lan P, Li F, Abad J, Zhou C, Li R (2017) Genome characterization of sweet potato symptomless virus 1: a mastrevirus with an unusual nonanucleotide sequence. Arch Virol 162:1–4
9. Clerot D, Bernardi F (2006) DNA helicase activity is associated with the replication initiator protein rep of tomato yellow leaf curl geminivirus. J Virol 80:11322–11330
10. Doyle JJ, Doyle JL (1987) A rapid isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15
11. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797
12. Fontenele RS, Lamas NS, Lacorte C, Lacerda ALM, Varsani A, Ribeiro SG (2017) A novel geminivirus identified in tomato and cleome plants sampled in Brazil. Virus Res. 240:175–179
13. Goncalves MC, Galdeano DM, Maia ID, Chagas CM (2011) Genetic variability of sugarcane mosaic virus causing maize mosaic in Brazil. Pesqui Agropecu Bras 46:362–369
14. Gonçalves MC, Godinho MT, Alves-Freitas DMT, Varsani A, Ribeiro SG (2017) First report of maize yellow mosaic virus infecting maize in Brazil. Plant Disease. doi:10.1094/PDIS-04-17-0569-PDN
15. Gorbalenya AE, Koonin EV, Wolf YI (1990) A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. FEBS Lett 262:145–148
16. Hammond RW, Bedendo IP (2001) Role of Maize rayado fino virus in the etiology of "Red Stunt" disease in Brazil. Plant Dis 85:99
17. Hammond RW, Ramirez P (2001) Molecular characterization of the genome of *Maize rayado* fino virus, the type member of the genus Marafivirus. Virology 282:338–347
18. Hammond RW, Bedendo IP (2005) Molecular confirmation of maize rayado fino virus as the Brazilian corn streak virus. Sci Agric 62:601–603
19. Harrison BD (1985) Advances in geminivirus research. Annu Rev Phytopathol 23:55–82
20. Kamali M, Heydarnejad J, Pouramini N, Masumi H, Farkas K, Kraberger S, Varsani A (2017) Genome sequences of *Beet curly top* Iran virus, Oat dwarf virus, Turnip curly top virus, and wheat dwarf virus identified in leafhoppers. Genome Announc 5:e01674-16
21. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649
22. Kitajima E, Costa A (1982) The ultrastructure of the corn chlorotic vein banding (Brazilian maize mosaic) virus-infected corn leaf tissues and viruliferous vector. Fitopatol Bras 7:247–259
23. Koonin Eugene V, Ilyina Tatyana V (1992) Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. J General Virol 73:2763–2766

24. Kreuze JF, Perez A, Untiveros M, Quispe D, Fuentes S, Barker I, Simon R (2009) Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. Virology 388:1–7
25. Kumar J, Kumar J, Singh SP, Tuli R (2014) Association of satellites with a mastrevirus in natural infection: complexity of wheat dwarf india virus disease. J Virol 88:7093–7104
26. Liang P, Navarro B, Zhang Z, Wang H, Lu M, Xiao H, Wu Q, Zhou X, Di Serio F, Li S (2015) Identification and characterization of a novel geminivirus with a monopartite genome infecting apple trees. J Gen Virol 96:2411–2420
27. Ma Y, Navarro B, Zhang Z, Lu M, Zhou X, Chi S, Di Serio F, Li S (2015) Identification and molecular characterization of a novel monopartite geminivirus associated with mulberry mosaic dwarf disease. J Gen Virol 96:2421–2434
28. Mar TB, Lau D, Schons J, Yamazaki-Lau E, Nhani A (2013) Molecular identification based on coat protein sequences of the Barley yellow dwarf virus from Brazil. Sci Agric 70:428–434
29. Martin DP, Willment JA, Billharz R, Velders R, Odhiambo B, Njuguna J, James D, Rybicki EP (2001) Sequence diversity and virulence in Zea mays of Maize streak virus isolates. Virology 288:247–255
30. Muhire B, Martin DP, Brown JK, Navas-Castillo J, Moriones E, Zerbini FM, Rivera-Bustamante R, Malathi VG, Briddon RW, Varsani A (2013) A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus Mastrevirus (family Geminiviridae). Arch Virol 158:1411–1424
31. Muhire BM, Varsani A, Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PloS ONE 9:e108277
32. Nahid N, Amin I, Mansoor S, Rybicki EP, van der Walt E, Briddon RW (2008) Two dicot-infecting mastreviruses (family Geminiviridae) occur in Pakistan. Arch Virol 153:1441–1451
33. Nash TE, Dallas MB, Reyes MI, Buhrman GK, Ascencio-Ibanez JT, Hanley-Bowdoin L (2011) Functional analysis of a novel motif conserved across geminivirus Rep proteins. J Virol 85:1182–1192
34. Ng TF, Duffy S, Polston JE, Bixby E, Vallad GE, Breitbart M (2011) Exploring the diversity of plant DNA viruses and their satellites using vector-enabled metagenomics on whiteflies. PloS one 6:e19050
35. Ng TF, Willner DL, Lim YW, Schmieder R, Chau B, Nilsson C, Anthony S, Ruan Y, Rohwer F, Breitbart M (2011) Broad surveys of DNA viral diversity obtained through viral metagenomics of mosquitoes. PloS ONE 6:e20579
36. Pande D, Madzokere E, Hartnady P, Kraberger S, Hadfield J, Rosario K, Jaschke A, Monjane AL, Owor BE, Dida MM, Shepherd DN, Martin DP, Varsani A, Harkins GW (2017) The role of Kenya in the trans-African spread of maize streak virus strain A. Virus Res 232:69–76
37. Roossinck MJ, Martin DP, Roumagnac P (2015) Plant virus metagenomics: advances in virus discovery. Phytopathology 105:716–727
38. Rosario K, Duffy S, Breitbart M (2012) A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. Arch Virol 157:1851–1871
39. Rosario K, Padilla-Rodriguez M, Kraberger S, Stainton D, Martin DP, Breitbart M, Varsani A (2013) Discovery of a novel mastrevirus and alphasatellite-like circular DNA in dragonflies (Epiprocta) from Puerto Rico. Virus Res 171:231–237
40. Rosario K, Capobianco H, Ng TF, Breitbart M, Polston JE (2014) RNA viral metagenome of whiteflies leads to the discovery and characterization of a whitefly-transmitted carlavirus in North America. PloS ONE 9:e86748
41. Rosario K, Seah YM, Marr C, Varsani A, Kraberger S, Stainton D, Moriones E, Polston JE, Duffy S, Breitbart M (2015) Vector-enabled metagenomic (VEM) surveys using whiteflies (Aleyrodidae) reveal novel begomovirus species in the new and old worlds. Viruses 7:5553–5570
42. Rosario K, Marr C, Varsani A, Kraberger S, Stainton D, Moriones E, Polston JE, Breitbart M (2016) Begomovirus-associated satellite dna diversity captured through vector-enabled metagenomic (VEM) surveys using whiteflies (Aleyrodidae). Viruses 8:36
43. Schwinghamer MW, Thomas JE, Schilg MA, Parry JN, Dann EK, Moore KJ, Kumari SG (2010) Mastreviruses in chickpea (Cicer arietinum) and other dicotyledonous crops and weeds in Queensland and northern New South Wales, Australia. Australas Plant Pathol 39:551–561
44. Shepherd DN, Martin DP, Van Der Walt E, Dent K, Varsani A, Rybicki EP (2010) Maize streak virus: an old and complex 'emerging' pathogen. Mol Plant Pathol 11:1–12
45. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19:1117–1123
46. Stover BC, Muller KF (2010) TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics 11:7
47. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739
48. Thomas JE, Parry JN, Schwinghamer MW, Dann EK (2010) Two novel mastreviruses from chickpea (Cicer arietinum) in Australia. Arch Virol 155:1777–1788
49. Zerbini FM, Briddon RW, Idris A, Martin DP, Moriones E, Navas-Castillo J, Rivera-Bustamante R, Roumagnac P, Varsani A, Ictv Report C (2017) ICTV virus taxonomy profile: geminiviridae. J Gen Virol 98:131–133