CAPÍTULO 5

CLADISTIC ANALISYS IN GEMINIVIRIDAE: AN EVIDENCE OF MULTISPECIFICITY FOR CULTIVARS HOSTS

Rafael Trindade Maia
Universidade Federal de Campina Grande, Centro de Desenvolvimento do Semiárido
Sumé, Paraíba. E-mail: rafael.rafatrin@gmail.com
Corresponding author

Aparecida Yasmim Silva de Azevedo
Universidade Federal de Campina Grande, Centro de Desenvolvimento do Semiárido.
Sumé, Paraíba.

Maria Bartira Chaves de Souza Silva
Universidade Federal de Campina Grande, Centro de Desenvolvimento do Semiárido.
Sumé, Paraíba.

Ana Verônica Silva do Nascimento
Universidade Federal de Campina Grande, Centro de Desenvolvimento do Semiárido.
Sumé, Paraíba.

ABSTRACT: In recent years the *Geminiviridae* family virus has been intensively studied due the severity of the diseases caused in several cultures of economic importance; like bean, cotton, corn, tomato and cassava. The objective of this work was to do a cladistic inference of *Begomovirus* populations through computational tools. Viral genome sequences were obtained from NCBI (https://www.ncbi.nlm.nih.gov/), totaling 297 sequences from different countries. The sequences were aligned with the Bioedit program, using the ClustalW algorithm. After the alignment, the cladogram was obtained through the Maximum Likelihood method in the MEGA program, with bootstrap analysis of 1000 replicates. ModelTest was used to select the most suitable evolutionary model (Reversible G + I) for the data set. The results showed that the bootstrap values in the clades ranged from 9 to 100. Pepper, potato and watermelon virus sequences were used with outgroup, but they were grouped into clades shared by the viruses obtained in the Tomato strain, revealing an evidence that the mosaic virus in these cultivars belong to the same species and same genus.
KEYWORDS: Maximum Likelihood; Mosaic virus; Tomato.

1 I INTRODUCTION

The *Geminiviridae* family is structurally characterized by the geminated morphology of the viral particle, with 18-30 nm and genetically by having a single-stranded circular single-stranded DNA molecule (ssDNA). Each of the molecules has 2500-3000 nucleotides (nt) encapsidated by a single structural protein that is arranged in the form of 22 capsomers forming two incomplete icosahedrons which confers the geminated appearance of the viral particles, typical characteristic of this family (BROWN et al., 2011).

The *Geminiviridae* family is composed by seven genus: *Begomovirus, Mastrevirus, Curtovirus, Becurtovirus, Turncuntovirus, Eragrovirus and Topocovirus* (2016). In addition, some representatives of the *Begomovirus* genus are associated with DNA satellites called alpha and beta-satellite. *Begomoviruses* (*Geminiviridae* family) and their associated satellite DNAs, known as alphasatellites and betasatellites, form complexes that cause devastating diseases in agricultural systems (LEKE et al., 2015). Except the *Begomovirus* genus, which includes viruses with a mono or bisegmented genome, all other genus of the family include only viruses with a monosegmented genome (PASSOS, 2016).

The DNA-A has five or six genes, one or two in the viral sense (CP and, in Old World viruses, AV2) and four in the complementary sense (Rep, TrAP, REn and AC4). The role of Rep (Replication-associated protein) in the coupling to the origin of replication for the initiation of DNA synthesis. The TrAP (Trans-Activating Protein) protein is responsible for activating transcription of the CP and NSP genes, as well as acting in the suppression of plant defense responses. REn (Replication-Enhancer protein) enhances replication, increasing the concentration of viral DNA. AC4 protein is related to suppression of gene silencing and, in monosegmented *Begomoviruses*, in viral movement. In some species an AC5 ORF is present, a recent study has shown that the encoded product plays an important role in suppressing plant defense responses. The CP protein is responsible for the viral encapsidation and specificity with the insect vector and, in the monosegmented *Begomoviruses*, it acts on the viral movement. The AV2 gene is present only in the Begomovirus of the Old World and its function is related to the viral movement (BRIDDON et al., 2010, PASSOS, 2016, VANITHARANI et al., 2004).

According to Passos (2016), of the three new genera, established by the ICTV, the *Becurtovirus* presents two species: Beet curly top Iran virus (BCTIV) and Spinach curly top Arizona virus (SCTAV). The transmission of BCTIV is by spittlebugs, while the SCTAV has not yet had the vector identified. *Eragrostis curvula* streak virus (ECSV) and Turnip curly top virus (TCTV), respectively, with the first one with no known vector and the second one transmitted by *Eragrovirus* and *Turncurtovirus*, spittlebugs Varsani et al. (2014) reports that a difference observed between the genus *Becurtovirus* and *Eragrovirus* and the other viruses in the family refers to the conserved sequence associated with the viral replication start site. Generally *Geminiviruses* have the TAATATTAC nine-nucleotide, however, viruses of both genus have the TAAGATTCC sequence.

Although the family *Geminiviridae* presents great diversity of genus, one of these has been highlighting in the recent research, for triggering several diseases that affect the tomato, belonging to the genus *Begomovirus*. It is currently the most relevant for the high incidence of viruses, with reports of losses up to 100% in tomato crops (BRINDDON et al., 2015).

Begomovirus Genus

Viruses of the *Begomovirus* genus has mostly of the genome divided into two single-stranded circular DNA components, called DNA A and DNA B, each with approximately 2.6 kbp. DNA-A encodes all the proteins necessary for the transcription, replication and encapsidation of both DNAs, whereas DNA-B contributes with the necessary functions for the movement of the virus in the plant and the development of symptoms (ROJAS et al., 2005). They have an intergenic region of approximately 200 bp, highly conserved between the two components of the same virus, which is essential for the recognition of the replication and transcription processes of the two genomes, thus maintaining the fidelity of the bipartite genome (BROWN et al., 2011 FAUQUET et al., 2005, JESKE, 2009).

The DNA-A has five or six genes, one or two in the viral sense (CP and, in Old World viruses, AV2) and four in the complementary sense (Rep, TrAP, REn and AC4). The role of Rep (Replication-associated protein) in the coupling to the origin of replication for the initiation of DNA synthesis. The TrAP (Trans-Activating Protein) protein is responsible for activating transcription of the CP and NSP genes, as well as acting in the suppression of plant defense responses. REn (Replication-Enhancer protein) enhances replication, increasing the concentration of viral DNA. AC4 protein is related to suppression of gene silencing and, in monosegmented *Begomoviruses*, in viral movement. In some species an AC5 ORF is present, a recent study has shown that the encoded product plays an important role in suppressing plant defense responses. The CP protein is responsible for the viral encapsidation and specificity with the insect vector and, in the monosegmented *Begomoviruses*, it acts on the viral movement. The AV2 gene is present only in the *Begomovirus* from Old World and its function is related to the viral movement (BRIDDON ET AL., 2010, PASSOS, 2016 AND VANITHARANI et al., 2004).

DNA-B has the NSP genes in the viral sense and MP in the complementary sense. Nuclear Shuttle Protein (NSP) is responsible for transporting viral DNA from the nucleus to the cytoplasm, and MP (Movement Protein) protein transports the viral cell-to-cell DNA through the plasmodesmas. In monopartic Begomoviruses, intracellular transport is performed by CP. Movement protein (MP) is encoded by the BC1 ORF in bipartite *Begomoviruses* or by V2 in single-pores, although the two ORFs do not share sequence identity. This protein is responsible for cell-to-cell viral movement, increasing the limit of exclusion of the plasmodesms (NOUEIRY et al., 1994; RÊGO, 2016; UNSELD et al., 2001).

Begomovirus replication

Begomoviruses are transmitted by *Bemisia tabaci* in a persistent circulatory way. Viral particles are acquired via stylet during the feeding process in infected plants, enter

the esophagus and the filter chamber, and are subsequently transported through the wall of the intestine to the hemolymph where they circulate until they enter the salivary gland, from which are transmitted to new plants during insect feeding (CZOSNEK et al., 2002; GHANIM et al., 2007; GILBERTSON et al., 2015).

DNA replication occurs in three phases: initiation, elongation and termination. The *Geminivirus* Rep catalyzes the initiation and termination of circle-cycle replication by cleavage and binding of the viral DNA at a conserved site within the viral genome. Similar to a number of small DNA viruses, geminiviruses do not encode their own DNA polymerases and therefore depend on host polymerases and associated factors (collectively, the so-called host replication) for viral DNA synthesis during the elongation step. In healthy plants, the availability of host repliastomy is heavily regulated by cell cycle and developmental controls, which must be reprogrammed before geminiviruses can replicate their genomes. The virions penetrate the plant cells during the feeding process of the virulent insect-vector. Inside the cell, the genetic material is routed to the nucleus. It is unclear whether the virus moves into the encapsulated or uncapsidated nucleus, but it is believed that CP is involved in this process, interacting with the host's transport chain (Gafni and Epel, 2002; HANLEY-BOWDOIN et al., 2013).

The replicative cycle of *Geminiviruses* can be subdivided into two functionally distinct stages, characterized by specific events. The first stage of the infectious cycle involves the consertion of the circular genomic ssDNA into a covalently closed circular supercoiled dsDNA intermediate called replicative form I (RFI), which is completed with the exclusive participation of the proteins of the infected cells since the dsDNA is the transcriptionally active template. The second stage consists of the use of dsDNA as a template for the ssDNA amplification by the rolling circle mechanism, in which the Rep viral protein is required. It is responsible for the initiation reaction involving a cut within the conserved nonanucleotide (5'-TAATATT - AC-3 ') in all *Geminiviruses*, located in the loop of the staple structure present in the intergenic region, this protein has several biochemical activities, including single and double stranded DNA binding property and DNA binding in specific sequences, ATP hydrolysis and initiation of the circle-cycle replication mechanism. After its inception the factors necessary to complete the rolling circle phase are of cellular origin. Later, the production and encapisidation of mature circular genomic ssDNA into viral particles occurs (FERNANDES, 2010, HANLEY-BOWDOIN et al., 2013).

These mechanisms present high complexity and according to the genus it can present different modes of transmission and replication. Therefore, the study of phylogeny presents one of the effective tools in the identification and characterization of these virus.

PHYLOGENY

The world is made up of a huge variety of organisms. In order to study and understand this diversity, it became necessary to group beings by their common characteristics, that is, to classify them. There are several criteria that can provide a basis for a classification system for any set of entities. Therefore, the fundamental principle of any attempt to classify objects or organisms was to speed communication between people who use information about these organisms (CABRAL, 2011).

The first attempt of classification according to MIYAKI et al. (2001) occurred from the studies of the species of Charles Darwin and from this the concept of phylogeny arose after the concept of ancestrality between species, through its first diagram published, representing the similarity between species. Thus, the phylogenies are nothing more than the indication of the supposed ancestral relations for a set of species. The term molecular phylogeny is evident, which is understood as the study of the evolutionary relations between organisms by the use of molecular data, such as sequences of nucleic acids and proteins, or other molecular markers. The logic of phylogenetic inference for molecular and morphological characters is identical, but the two have different properties, methods, and concepts. The main problems faced in the analysis of morphological data are also faced with molecular data (LIMA, 2003; RIDLEY, 2004).

The evolutionary inferences made from molecular data allows an inference about the evolutionary relations between species. Several results obtained from the molecular phylogeny have aided in the structuring and the current conception of the taxonomic classification of the species (SAKAMOTO, 2016).

The phylogenetic trees are constructed through data and parameters, being these defined like graphs that they have a hierarchical structure. In these trees, nodes are called taxonomic units, which may represent, depending on the data analyzed, species, populations, genes or proteins. The nodes are classified in terminals (leaves), when these are in the end of the tree, or in internal, when of them depart one or more descending branches. The terminal nodes represent the same samples used for inference of the tree and are therefore also called operational taxonomic units (OTUs), which in turn correspond to the basic unit (species, population, gene or protein) to be studied and compared. Internal nodes represent evolutionary events that depict the divergence of the taxonomic unit under analysis, such as speciation events, if the taxonomic unit is population, or gene duplication events, if the taxonomic unit is a gene or a protein. As the determination of internal nodes are products of a phylogenetic inference, we also call them hypothetical taxonomic units (HTUs). The branches are elements that connect the nodes and their size represents the estimated time of the evolutionary relation between the taxonomic units. The smaller the distance of the branches between the taxonomic units in comparison, the closer they are to evolutionarily (SOKAl, 1966).

Phylogenetic trees may also assume different conformations depending on the arrangement of the branches along the tree. These different conformations that a tree can assume are called topologies. The concept of topology is of great importance in phylogenetic studies because it represents the basis of all interpretation of evolutionary histories among the samples under analysis. (SAKAMOTO, 2016).

METHODOLOGY

Blast algorithm was used to search *Geminivirus* genome sequences. A total of 297 entries were selected from NCBI genbank database (https://www.ncbi.nlm.nih.gov/genbank/) from differents countries around the world. Multiple sequences aligments were prepared with Bioedit (HALL et al. 1999) program (ClustalW algorithm). After aligment, Modeltest was runned for the sequence set and the evolutive model was choosen (Reversible G + I). The Maximum Likelihood to build the cladogram in the MEGA 7.0 software (http://www.megasoftware.net). The Maximum Likelihood method is a ponctual estimation method, which consists in estimating the unknown parameters so that the probability of observing the Y data is greater (or maximum). Thus, a (point) estimator of θ at any function (Y) is one that makes use of the information contained in a population sample f (Y; θ) to obtain a set of numbers that can be considered to represent approximately the unknown value of the parameters in θ. Thus, the estimation takes place from the realization of the function θ (Y) for a given sample (GUJARATI AND PORTER, 2011).

Due these features, this method has the best statistical properties, under large samples (consistency, non-bias, efficiency and normal distribution). However, in practice its use requires the use of numerical methods, starting with an initial solution arbitrated by the user numerical calculation and a possibly large number of interactions (MILAGRES, 2015).

RESULTS AND DISCUSSION

The bootstrap values ranged from 9 to 100. Although we get some uncertain branches in our analisys based on bootstrapp support (<70), it is clear that here in this results we can conclude that *Geminivirus (Begomovirus)* displays a multispecificity for interesting ecocomic crops. Here, in the partial tree (Fig 1) is clear to conclude that the *Begomovirus* does not displays a great specificity for cultivars. Some studies determinates that *Geminiviridae* are classified by genome organization, hosts distribuition and vetor specificity.

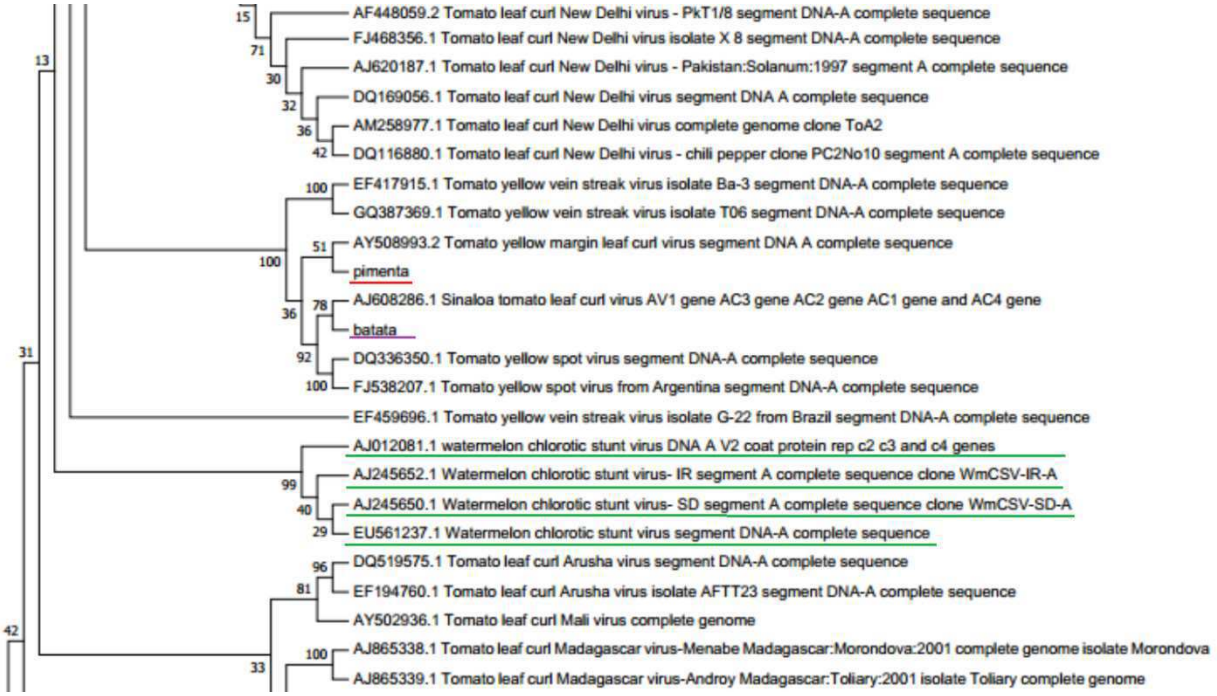

Figure 1. Partial view of the *Geminiviridae* tree. Labeled in red is the pepper sequence, while potato are labeled in purple and watermelon in green.

This is an interesting result, and corroborates with the multispecificity for hosts. We can also see, by accessing the genbank code (AY508993.2), that Geminiviridae from pepper is more related to a strain obtained from Merida region (Venezuela country), while potato seems to be more related to a strain from Sinaloa, Mexico. The watermelon sequences has a more closed relationship with a *Tomato* strain from Brazil.

In another segment of the *Geminiviridae* tree (Figure 2), the watermelon grouped in an adjacent clade shared by pepper and potato. However, they grouped with different strains from those observed in Figure 1. The most closed related clade to Watermelon was a clade constituted by Ramie (*Boehmeria nivea*) strains from China. The pepper and potato strains shared a clade with *Tomato* sequence obtained from Sinaloa strains. Although we found some relevant insights about the *Geminivirus* multispecificity, many aspects of its distributioin and philogeny are still uncertainly. Future studies of genetic population and distribution of these strains should be applied to improve a better understand of thes virus family.

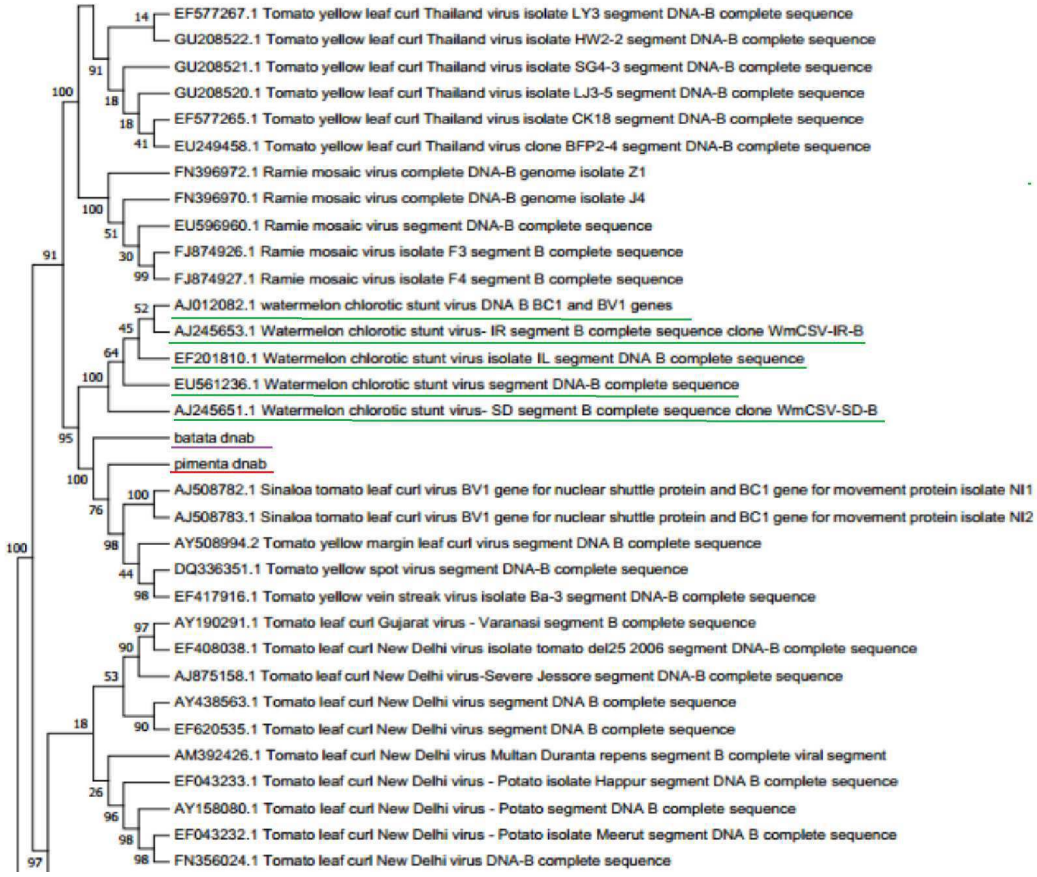

Figure 2. Partial tree view. Watermelon is labeled in green; potato in purple and pepper in red.

CONCLUSIONS

In the context of our results, and considering the *Geminiviridae* distribuition, we can conclude that *Begomovirus* does not present a straigh specificity to its hosts. It means, as discussed before, that the *Geminiviridae* viruses specificity is more assotiated with the vector than the plant hosts, wich represents an evidence of multispecificity for host infection. These results will be useful for future crop manegement and mosaic disease control.

REFERÊNCIAS

BRIDDON, R. W.; PATIL, B. L.; BAGEWADI, B.; NAWAZ–UL-REHMAN, M. S.; FAUQUET, C. M. Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. **BMC Evolutionary Biology**, v. 10, n. 1, p. 97-108, 2010.

BRIDDON, R. W. Geminiviridae. In: Encyclopedia Life Sciences (eLS), **John Wiley & Sons**, 1–12, Ltd: Chichester, 2015.

BROWN, J. K.; FAUQUET, C. M.; BRIDDON, R. W.; ZERBINI, M.; MORIONES, E.; NAVASCASTILLO, J. Geminiviridae. In: KING A.M.Q., ADAMS M.J., CARSTENS E.B.; LEFKOWITZ E.J. (eds). Virus Taxonomy - Ninth Report of the International Committee on Taxonomy of Viruses. Associated Press, Elsevier Inc.; London, UK. pp. 351– 373, 2012.

CABRAL, T. S. Filogenia molecular de fungos gasteroides das ordens Phallales e Geastrales (Phallomycetidae). Dissertação apresentada ao Programa de Pós-graduação em Ciências Biológicas da Universidade Federal do Rio Grande do Norte, 2011.

CZOSNEK, H.; GHANIM, M. The circulative pathway of begomoviruses in the whitefly vector Bemisia tabaci: Insights from studies with Tomato yellow leaf curl virus. **Annals of Applied Biology**, v. 140, p. 215-232, 2002.

FERNANDES, N. A. N. **Variabilidade genômica e geográfica de espécies de Begomovirus em tomateiro e em dois gêneros de plantas daninhas do Brasil.** Dissertação apresentada ao Programa de Pós-Graduação em Fitopatologia da Universidade de Brasília, 2010.

GAFNI, Y.; EPEL, B. L. The role of host and viral proteins in intra- and inter-cellular trafficking of geminiviruses. **Physiol. Mol. Plant Pathol**. 2002, 60**:** 231-241.

GILBERTSON, R.L.; BATUMAN, O.; WEBSTER, C.G.; ADKINS, S. Role of the insect supervectors Bemisia tabaci and Frankliniella occidentalis in the emergence and global spread of plant viruses. **Annual Review of Virology**, v. 2, p. 67-93, 2015.

GHANIM, M.; SOBOL, I.; GHANIM, M.; CZOSNEK, H. Horizontal transmission of begomoviruses between Bemisia tabaci biotypes. **Arthropod-Plant Interactions**, v. 1, p. 195-204, 2007.

GARCIA, M. **Uma filogenia mitocondrial de metazoários.** 212f. Petrópolis, RJ. Dissertação (Mestrado em Modelagem Computacional com Ênfase em Bioinformática) - Programa de Pós-graduação em Modelagem Computacional, Universidade Federal do Rio de Janeiro, 2007.

GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**. 5ª ed. Editora: AMGH, Rio de Janeiro, 2011.

HALL, T. A. BioEdiT: a user-friendly biological sequence aligment editor and analisys programfor windows 95/98/NT. **Nucleic Acids Symposium Series**, 41, 95-98. 1999.

HANLEY-BOWDOIN, L.; BEJARANO, E. R.; ROBERTSON, D.; MANSOOR, S. Geminiviruses: masters at redirecting and reprogramming plant processes. **Nat Rev Microbiol.** 2013 Nov;11(11):777-88.

LEKE, W.; MIGNOUNA, D.; BROWN, J.; KVARNHEDEN, A. Begomovirus disease complex: Emerging threat to vegetable production systems of West and Central Africa. **Agric. Food Secur**. 2015.

LIMA, J. P. M. **Estudos taxonômicos moleculares no táxon Phaseoleae DC. (Leguminosae, Papilionoideae) utilizando sequências de DNA ribossômico (rDNA).** 151 f. (Dissertação de Mestrado), Universidade Federal do Ceará, Fortaleza, 2003.

MILAGRES, T.G. P. S. Otimização da função de verossimilhança de modelos GAS com aplicações em séries de vento. Relatório apresentado ao Departamento de Engenharia Elétrica da PUC – Rio, 2015.

MIYAKI, C. Y., RUSSO, C. A. de M. & PEREIRA, S. L. Reconstrução Filogenética. Introdução e o Método da Máxima Parcimônia. In: MATIOLI, S. R. (ed.) **Biologia Molecular e Evolução**. p. 97-107. Holos editora. Ribeirão Preto, SP. 2001.

NOUEIRY, A. O.; LUCAS, W. J.; GILBERTSON, R. L. Two Proteins of a Plant DNA Virus Coordinate Nuclear and Plasmodesmal Transport. **Cell**, v.76, n. 5, p. 925-932, 1994.

PASSOS, L. S. **Diversidade de espécies e filogenia de begomovírus que infectam plantas não-cultivadas nos estados do Ceará e Piauí**. Dissertação apresentada ao Programa de Pós-Graduação em Genética e Melhoramento da Universidade Federal do Piauí, 2016.

RÊGO, C. M. **Diversidade genômica de Begomovírus em tomateiros resistente (brs sena) e susceptível (h-9553)**. Dissertação apresentada ao Programa de Pós- Graduação em Fitopatologia da Universidade de Brasília , 2016.

SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Mol Biol Evol**. 1987 Jul;4(4):406-25.

SAKAMOTO, T. **Ferramentas para análise filogenética e de distribuição taxonômica de genes ortólogos**. Tese apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, 2016.

SANDES, E. F. O.; MELO, A. C. M. A. Algoritmos paralelos exatos e otimizações para alinhamento de sequências biológicas longas em plataformas de alto desempenho. **XXXVI Congresso da Sociedade Brasileira de Computação**, 2016.

SILVA, J. C. F. **Data warehouse enriquecido com métodos de aprendizado de máquina para a família Geminiviridae**. Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Viçosa, 2016.

SOKAL, R. R. Numerical Taxonomy. **Scientific American**, v. 215, p. 106–116, 1966.

SOUZA, L. H. T. **Análises da entropia do fluxo de informação em redes de interação proteica associadas as ferramentas de enriquecimento funcional revelam genes de interesse prognóstico em glioblastomas**. Dissertação apresentada ao Programa de Pós-Graduação em Ciências Biológicas da Universidade Federal do Rio Grande do Sul, 2015.

UNSELD, S.; HOHNLE, M.; RINGEL, M.; FRISCHMUTH, T. Subcellular targeting of the coat protein of African cassava mosaic geminivirus. **Virology** 286: 373-383, 2001.

VARSANI, A.; NAVAS-XASTILLO, J.; MORIONES, E.; HERNÁNDEZ-ZEPEDA, C.; IDRIS, A.; BROWN, J. K.; ZERBINI, F. M.; MARTIN, D. P. Establishment of three new genera in the family Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. **Archives of Virology**, v. 159, n. 8, p. 2193-2203, 2014.

VANITHARANI, R.; CHELLAPPAN, P.; PITA, J. S.; FAUQUET, C. M. Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of posttranscriptional gene silencing. **Journal of Virology** 78:9487-9498, 2004.