ORIGIN, EVOLUTION AND MOLECULAR EPIDEMIOLOGY OF PAPAYA RINGSPOT VIRUS

Olarte Castillo, X.A.[1], Sierra, R.[1], Bernal, A.J.[1], Fermín, G.[2] & Restrepo, S.[1*]

[1] Laboratorio de Micología y Fitopatología, Departamento de Ciencias Biológicas, Universidad de los Andes, Bogotá D.C., Colombia.

[2] Facultad de Ciencias, Universidad de los Andes, Mérida, Venezuela.

*Corresponding author

ABSTRACT

Papaya ringspot virus (PRSV) is the most important virus affecting papaya and cucurbit plants in tropical and subtropical areas. PRSV is divided into two groups: the W type, which infects papaya *(Carica papaya)* and plants belonging to the Cucurbitaceae family, and the P type, which only infects papaya. The study of molecular variation of this virus is of great relevance, especially to develop transgenic resistance. In order to have a better understanding about the variation, origin, evolution and epidemiology of this virus it is necessary to construct a robust phylogenetic inference. Moreover, it is essential to include a comprehensive collection of sequences of poorly studied regions, especially South America. For this purpose, the sequences of the Capsid Protein (CP) have been used. For this gene, a phylogenetic approach using a molecular clock is presented, including sequences reported previously from different parts of the world and new ones obtained in Colombia. In order to calibrate the tree, the known dates of collection from different isolates were used. The advantage of this method is that divergence times can be accurately calculated. These dates can be used to further define the origin and epidemiology of PRSV. Our analysis shows that apparently, movement of the virus between regions contributes to the observed population variation. The hypothesis of the origin of PRSV-P biotype from PRSV-W is revised, because evidence for the opposite process is found in several instances.

Keywords: *Papaya ringspot virus*, Relaxed Molecular Clock, Bayesian Phylogeny

INTRODUCTION

Papaya ringspot virus (PRSV; Family *Potyviridae*, genus *Potyvirus*) is an important pathogen of papaya and cucurbits (Purcifull et al., 1984). PRSV particles are flexuous filaments of 780 x 12nm with a single stranded RNA genome of positive polarity with ca. 10.000 nucleotides that encode a polyprotein within a single ORF (De La Rosa and Lastra, 1983). Plants infected with PRSV develop symptoms ranging from mosaic, chlorosis and distortion of leaves, to ring spot and streaking on fruits and water-soaked oily streaks on stem and petioles (Purcifull et al., 1984). The virus is transmitted non-persistently by several species of aphids (Purcifull et al., 1984), a process requiring the coat protein (CP) and the helper component (HC-Pro) (Maia, Haenni, and Bernardi, 1996; Pirone, 1991; Pirone and Blanc, 1996; Wang et al., 1998).

PRSV isolates are divided into two groups: the P and W biotypes (Purcifull et al., 1984). PRSV-W and PRSV-P can infect plants in the *Cucurbitaceae* family, however only PRSV-P naturally infects papaya *(Carica papaya)* (Gonsalves, 1998; Purcifull et al., 1984). Both isolates are serologically indistinguishable (Purcifull et al., 1984) and experimentally can infect *Chenopodium quinoa* and *C. amaranticolor*, producing local lesions (Gonsalves, 1998).

PRSV-P has been isolated in papaya-producing areas including Africa, the Caribbean, Australia, Tropical Asia, Latin America and the South Pacific (Tennant et al., 1994). Due to its large distribution, this virus is known as the major limiting factor for papaya production throughout the tropics and subtropics (Gonsalves, 1998). This includes Colombia, where the disease produced by PRSV-P has been recorded in several Departments, since 1952 (Sánchez de Luque and Martínez López, 1998) and it is considered as the most important disease of papaya in this country (Páez, 2003). Control of the disease in Colombia includes cultural practice methods such as isolation of infected cultivars and removal of young infected trees (Páez, 2003). The use of the

attenuated isolate HA 5-1 developed in Hawaii (Yeh and Gonsalves, 1984) has also been used, showing a low efficiency of protection (Páez, 2003). Furthermore, various studies in different countries where the attenuated isolate or transgenic protection were not useful, have shown that resistance depends on the level of RNA similarity between the transgene or mild strain and the incoming virus (Gonsalves, 1998; Tennant et al., 1994; Tripathi et al., 2008). Studies evaluating the viability of virus mild strains for cross-protection have also been achieved for PRSV-W, showing similar results (Rezende and Pacheco, 1998; Rezende et al., 1994). Therefore to obtain an adequate protection, virus populations in each country must be investigated to either find a local attenuated virus or to further understand the local virus variability (Gonsalves, 1998; Silva-Rosales et al., 2000; Tripathi et al., 2008).

The molecular variability and phylogenetic relationships of PRSV were investigated by Batenson et al. (2002), including sequences from Indochina, Japan, Australia, United States, Brasil, Mexico and Puerto Rico. The general conclusion was that PRSV populations show high variability, and that variability was related to the geographical origin of the strains. More importantly, evidence of the origin of PRSV-P from PRSV-W, previously proposed Batenson et al., 1994 either by mutation or recombination, could not be clearly resolved. According to this, in order to increase the knowledge on the origin, evolution and epidemiology of PRSV, it is necessary to obtain more sequences from different countries, especially from regions that were missing in previous studies such as Central and South America (Bateson et al., 2002).

To help resolve these issues, we examined the evolutionary dynamics of PRSV including recently published sequences from different countries and new sequences from PRSV-P obtained in Colombia. For this purpose, numerous sequences of the Capsid Protein (CP) were employed and a phylogenetic approach using a relaxed molecular clock is presented. In order to calibrate the tree, the known dates of

collection from different isolates were used. The advantage of this method is that divergence times can be accurately calculated and hence the viral diversity across time and space can be studied. This type of analysis has been performed with diverse kinds of viruses, and has been proved to be helpful to solve epidemiological issues (Rambaut et al., 2008).

MATERIALS AND METHODS

Virus isolates

Virus isolates used in this study were collected from papaya fields in Colombia. The isolates were collected between 2008 and 2009 in the following Departments: Casanare, Santander, Valle del Cauca, Risaralda, Caldas and Cundinamarca (Figure 1). Whenever possible, every location was georeferenced with the Magellan SporTrak™ or 12XL Garmin™ GPS. The names and origin of the isolates are listed in Supplementary Table 1.

Young papaya leaves showing or not symptoms of PRSV infection were collected. These samples were placed in plastic bags with an absorbent slightly moist paper towel slightly moist. To preserve the samples until the day of its processing plastic bags were kept open on ice. Once in the laboratory, samples were macerated in liquid nitrogen with sterile mortars and pistils. Finally, all samples were stored at -80ºC until RNA extraction.

RNA extraction, cDNA synthesis and PCR

In order to avoid RNases, all surfaces were cleaned with RNase away (Molecular BioProducts, San Diego, California, USA). Total RNA was extracted from macerated tissue using the Concert™ Plant RNA Reagent (Invitrogen, California, USA) according to manufacturer's instructions. The resulting pellet was resuspended in 20µl of sterile water treated with Diethilpirocarbonate (DEPC; 0,1%v/v). Four µl of each extracted RNA was denatured by heating at 65ºC for 10 minutes. Denatured RNA was subjected

to electrophoresis in a denaturing 1.2% agarose- formaldehyde gel, using a constant voltage of 70V for 25 minutes.

cDNA was synthesized from total RNA using the complementary primer 04-04 (Chin et al., 2007). For this purpose either the SuperScript™ III RT/Taq mix Platinum® (Invitrogen) or the iScript™ Select cDNA Synthesis Kit (Biorad) were used. To avoid RNA degradation, 2µl of RNasin® Plus RNase Inhibitor (Promega) were added to all of the reaction mixtures mentioned above.

Subsequently, double-stranded DNA fragments of the CP and HC-Pro genes were generated. For the CP gene, the following primers were used: 04-04 and 04-02 (Chin *et al,* 2007) or MB11M (modified from Bateson *et al,* 1994; removing 15 nucleotides not relevant for this study). PCR conditions included an initial denaturation cycle of 2 min at 94 °C followed by 39 cycles of denaturation for 30 s at 94 °C, annealing for 30 s at 46.5ºC and extension for 30 s at 72 °C, with a final extension of 6 min at 72 °C. For the amplification of the HC-Pro gene, the primers FHCP-566 and RHCP-566 were used. The PCR amplification scheme was the same as described formerly for the CP gene except for the annealing temperature (56°C for the HC-Pro gene). All primer information is shown in Figure 2.

Both genes were amplified using either GoTaq® Green Master Mix (Promega) or a mix of *Pfu* DNA polymerase (Fermentas) and Fermentas *Taq* DNA polymerase. For the CP gene 3µl of cDNA and 0.5µl of each primer (10µM) were used in a final volume of 25µl. For HC-Pro gene, 3µl of cDNA were also used in a 20 µl reaction mixture containing 1X Buffer, 2.5mM $MgSO_4$, 0.2mM dNTPs, 0.2µM Primers, 0,25U of *Taq*,, 1U of *Pfu* and 5% DMSO. Amplified products were analyzed in 1% agarose gel electrophoresis in TAE buffer.

Cloning and Sequencing

PCR fragments were purified from the agarose gel with the Wizard® SV Gel and PCR Clean-Up System (Promega). The purified PCR products were polyadenilated in a 10µl final volume reaction containing Fermentas *Taq* DNA polymerase (5U), *Taq* DNA Buffer (10X), $MgCl_2$ (25mM) and dATP (0,2mM). The polyadenilated products were cloned into pGEM®-T (Promega) according to manufacturer's instructions.

Competent cells of *Escherichia coli* DH5α were prepared by a series of washes as previously described (Maniatis, Fritsch, and Sambrook, 1982). Forty µl aliquots of these cells were were maintained at -80ºC, and 1.5µl of the vector ligation was electroporated in these cells at 1.25V in a MicroPulser™ (Biorad) electroporator. Transformed cells were cultured in plates with LB medium supplemented with ampicillin (100µg/ml) and X-gal (20µg/ml). White colonies were selected to perform PCR with the Universal M-13 primers, using GoTaq Green Master Mix (Promega). After visualization of the amplified products in a 1% agarose gel, the PCR fragments were gel-purified with the Clean-Up System described above. The purified products were sequenced at Macrogen (Korea). A total of three clones per virus isolate per gene were sequenced in both directions (3' and 5').

Sequence analysis

Nucleotide sequences for the CP and HC-Pro genes for other PRSV isolates were retrieved from Genbank (http://www.ncbi.nlm.nih.gov/Genbank/). The information regarding these PRSV isolates is listed in Supplementary Table 2.

The quality of the obtained sequences was assessed in BioEdit (Hall, 1999), with which the DNA sequencing chromatographs were analyzed. A VecScreen filter was performed in order to clean sequences of vector contamination (available at http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html). Subsequently, sequences were assembled in Sequencher™ (Gene Codes Corporation) and aligned using the

MUSCLE algorithm (Edgar, 2004) in the EMBL-EBI web page (http://www.ebi.ac.uk/Tools/muscle/). A recombination analysis between and within groups of sequences (*i.e.,* Countries) was performed in RecombiTEST (Piganeau, Gardner, and Eyre-Walker, 2004), using the Maximum Chi-square test (Smith, 1992). On the other hand, a total of 13 genomes of PRSV-P and PRSV-W had been completely sequenced (AY231130, S46722, DQ374153, DQ374152, DQ340771, DQ340769, DQ340770, AY027810, X97251, AY162218, AY010722, EF017707, X67673). These 13 sequences were also aligned with MUSCLE as described previously with the CP and HC-Pro sequences obtained from Colombia. The nucleotide distance was calculated using MEGA (Tamura et al., 2007), and the nucleotide diversity (π), the Segregant Sites (S), the Watterson's θ (θ_W), Tajima's D (Tajima, 1989) and the proportion of non-synonymous to synonymous substitutions (ω = dN/dS), were implemented using DNAsp (Rozas and Rozas, 1995).

Phylogenetic analyses

The Maximum Likelihood (ML) analysis was implemented using PAUP* 4.0b10 (Swofford, 2002). The nucleotide substitution model was obtained with ModelTest3.7 (Posada and Crandall, 1998), following the Akaike Information Criteria (AIC) (Akaike, 1974; Akaike, Petrov, and Csaki, 1973). The branch support was calculated with a 1000 replicates bootstrap. This analysis was performed for the sequences of CP and HC-Pro, and the obtained topologies were compared with a Shimodaira-Hasegawa statistical test (Shimodaira and Hasegawa, 1999). Potato virus Y (PVY, Accession number AY884985) was used as outgroup.

A phylogeny under a relaxed-clock model was constructed for the CP gene using BEAST v 1.4.8 (Drummond et al., 2006). In this Bayesian phylogenetic inference, an Uncorrelated Lognormal molecular clock model was chosen and the tree prior was the model of Constant Size. The known dates of collection were used as calibration points.

An arbitrary date was set in cases where the collection dates were unknown. The substitution model was GTR with gamma and invariant sites as a model for site heterogeneity, with 4 gamma categories. The length of the MCMC was 20.000.000 with a screen every 1.000th generation. The convergence of this analysis was studied with Tracer v 1.4 (Avaible at http://tree.bio.ed.ac.uk/software/tracer/), in order to examine the robustness of the obtained model.

RESULTS

Sample collection

Typical symptoms of PRSV were observed in Colombian papaya fields, including reduction in foliar area, mosaic, distortion, and chlorosis, as well as fruit ringspot (Figure 3).

For the amplification of the CP gene, the primer pair 04-02 and 04-04 did not produce an amplicon in many cases. The best result was obtained with primers 04-02 and MB11M. A total of nine new sequences of the CP gene were obtained. Nevertheless, virus mechanical inoculation in indicator plants (i.e. in *Chenopodium*) needs to be performed to discard the presence of the virus in the other collection samples.

Sequence analysis

Sequences from Colombian samples, and 118 additional sequences obtained in GenBank of PRSV-P and PRSV-W from different parts of the world, were aligned using MUSCLE. The recombination analysis showed two possible events of recombination ($p<0,05$), both involving as recombinant and donors Indian isolates (data not shown). The recombinant sequence, India_UPLK, has as a major parent the isolate India_UPV. The other case of recombination is the India_WB, having as a major parent the isolate India_HP.

The nucleotide diversity (π) ranged from $\pi=0,01102$ (Jamaican isolates) to $\pi=0,07977$ (Indian isolates) followed by $\pi=0,07177$ (Mexican isolates). For the 15 sequences of

each CP and HC-Pro, HC-Pro was found to be more diverse (π=0,15277) than CP (π=0,10718). For both genes, the observed value of Tajima's D within groups and in the overall alignment was significantly negative, indicating that they are not under neutral evolution. In order to identify which selection force is acting on each gene, the proportion of non synonymous/synonymous substitutions (ω) was calculated, obtaining for both genes a ω greater than 1, indicating positive selection (Supplementary Table 3).

Phylogenetic analysis

A phylogeny of the total CP sequences was constructed using an uncorrelated LogNormal Molecular Clock and the resulting topology situated PRSV virus origin approximately between 75 and 150 years ago (Figure 4). In concordance with previous reports (Bateson et al., 2002), the origin of the virus is placed in South Asia, more precisely in India. However, this basal clade is formed only by isolates of PRSV-P and no isolates of PRSV-W are found there. Moreover, Asian isolates are grouped into two clusters. One group is monophyletic and includes eastern isolates (China, Thailand, Vietnam, Japan, Taiwan and Korea). The other is the paraphyletic and basal group of South Asia, including isolates from India and Bangladesh. This group is the closest ancestor of the Americas-Australia monophyletic group. In the Americas-Australia group, Jamaica and Australia are monophyletic in contrast with the Mexican isolates that appear to have several origins. Indeed, one of the Mexican isolates is basal to the Australian clade (Figure 4, black arrow).

Regarding the Colombian isolates of PRSV, the ones from Valle del Cauca, Cundinamarca and Santander (Figure 1) grouped together in a monophyletic clade (Figure 4, pink arrow). However, isolates from Casanare were found within the Venezuela clade (Figure 4, pink and orange), showing two different origins for the Colombian isolates of PRSV-P.

For the comparison of the CP and HC-Pro genes, ML phylogenetic trees were constructed. A heuristic search of the ML analysis for CP-gene generated an optimal tree under the GTR+I+G model (Figure 5.A), and for HC-Pro-gene the generated tree was under the TIM+I+G model (Figure 5.B). Both trees showed Brazil and Hawaii as monophyletic groups. However, only the HC-Pro tree could resolve the relationships of the Taiwan isolates, whereas the CP tree resolved those for Thailand. Although for both cases, these branches were not highly supported. Neither tree could resolve the relationships of Colombian and Mexican isolates; however these were grouped with Brazilian and Hawaiian isolates. According to the SH test, for topology comparison, these topologies differed significantly from each other ($p < 0.05$).

DISCUSSION

The phylogeny constructed under a relaxed molecular clock showed a supported topology, with the exception of a few poorly supported clades (Figure 4, branches in red). However, the range of dates appears to be calculated accurately. For example, PRSV-W was first described in Australia 31 years ago (Greber, 1978), date that is included in the range of 30 to 45 years ago calculated in this study to be the origin of that clade. For the cases of Colombia and Hawaii, PRSV-P was first described around 50 years ago (Jensen, 1949; Sánchez de Luque and Martínez López, 1998) and in Venezuela it was reported 60 years ago (Muller, 1941). In the constructed phylogeny, the dates for these groups were placed between 5 to 45 years ago. The deviation of these obtained dates and those reported, may be due to the use of relatively recent sequences (the sequences included from Colombia are from the last two years), which can lead to underestimation of divergence times. Moreover the collection dates of Hawaii where unknown, which can also influence the estimation of divergence dates. For the Mexican isolates, collected during a wider time range (from 1994 to 2000), the estimated date of origin is between 65 to 45 years ago, a range that covers the first

report in this country in 1975 (Téliz-Ortiz et al., 1991). However, as there appears to be multiple origins for the Mexican, Colombian and Venezuela isolates (non-monophyletic groups), the calculus for the origin of this isolates can be deviated, as probably no every isolate has been recorded. Nevertheless it is noted that the first report of a disease might also be an underestimation of the date of origin of a virus in a given place.

Previous reports sustained that PRSV appears to be originated in Asia (Batenson et al., 2002), and the present study also supports this hypothesis. This is also in accordance with the high nucleotide diversity found in Asia, more specifically in India, the basal clade. However, neither Indian nor Asian clades are monophyletic. More interesting, it has been proposed that PRSV-P evolved from PRSV-W, but in this study there is evidence that the opposite process is also possible and that it actually has occurred many times. The first evidence that PRSV-P evolved from PRSV-W came from Australian populations of the virus (Bateson et al., 1994). As for this case, PRSV-W was first described in Australia in 1978 (Greber, 1978) at least 20 years before PRSV-P appeared (Thomas and Dodman, 1993). When that study was carried out, only few sequences of PRSV (all of them from Asia and USA), were available. The final result revealed that PRSV-P from Australia arose from an isolate of an Australian PRSV-W. However, the inclusion of more sequences from Central and South America showed that an isolate of PRSV-P from Mexico is the closest ancestor to the PRSV-W (Figure 4, black arrow). Moreover, when more sequences of PRSV-W from Australia were included, some of them appeared to have arisen from PRSV-P isolates (Figure 4). Later studies that confirmed the hypothesis of the origin of PRSV-P from PRSV-W (Bateson et al., 2002; Inoue-Nagata et al., 2007), did not include the sequences of PRSV-P from India used in this study (Jain et al., 2004). When included in this and in Jain et al.

(2004) studies, these sequences are situated in the most basal part of the tree, showing PRSV-P from India as the basal ancestor either for PRSV-P than for PRSV-W. For the Colombian case, the isolates obtained from Casanare did not group together with the isolates of the rest of Colombia but with isolates from Venezuela (Figure 4, pink arrow). Of all the departments where the collections were made, Casanare is the closest to Venezuela (Figure 1), implying a probable migration of infected plants between these two countries. This shows that the virus movement is important in shaping the molecular population evolution of PRSV, as this is also observed in the overall phylogeny with many non-monophyletic groups. Unfortunately, migration analyses cannot be performed with the CP gene because it is under positive selection, which can bias this type of analysis. Because of this, a highly supported phylogeny as the one constructed here, shows to be a good approach to infer the migration patterns for PRSV. Nonetheless, a better approach that can be done with a calibrated phylogeny is the fact that with this one one can take in account evolutionary history not only in space but also in time. For this, the study of population dynamics of genetic diversity in time (Drummond et al., 2005), has been proposed for all genes of the virus genome. However only 14 complete genome sequences have been reported and only 4 on them are from the W biotype. Another good approach could be phylogeographic

The tree obtained with the HC-Pro gene was better at resolving the relationships between the isolates than the one using CP. Even if the overall tree topology was very similar, they were significantly different. In a previous study of PRSV-P isolates, neighbor joining trees of the CI, 6K and HC-Pro proteins showed a different topology from the other five proteins of PRSV (Noa-Carrazana, González-de-León, and Silva-Rosales, 2007). Interestingly, these genes are expected to have a similar evolutionary history because they are involved in aphid transmission (Urcuqui-Inchima, Haenni, and Bernardi, 2001). However, HC-Pro is a multifunctional protein involved not only on

aphid propagation but also in long distance and cell to cell movement, silencing suppression, virus synergism in co-infection and development of symptoms (Ballut et al., 2005; Urcuqui-Inchima, Haenni, and Bernardi, 2001). To study the coevolution between these two genes, the conserved parts involved in aphid transmission should be used. In this case the regions around the DAG motif in CP (close to the N-terminus), which are implicated in aphid transmission (Shukla, Ward, and Brunt, 1994), have shown considerable conservation (Bateson et al., 1994; Bateson et al., 2002). For HC-Pro the regions around the KITC (N-terminus) and PTK (C-terminus) seemed to be involved in aphid transmission (Revers et al., 1999) and so they also should be used for the study of coevolution between these two genes. As the HC-Pro gene gave a more resolved phylogeny, this gene could be used to resolve the relationships among PRSV isolates.

Finally, the PRSV evolutionary history is very complex, including several changes in biotype, which include the change of host range. Molecular studies should be done to understand why PRSV-P appears to have lost its natural ability to infect cucurbits despite all this host changes. For this reason, this virus could be used as a biological model to study this process. Moreover it could be the model to study different selective pressures on virus, including how a virus population is influenced by the presence of resistant cultivars.

ACKNOWLEDGEMENTS

To Adriana Castañeda and her ICA partners and to Aníbal Tapiero and his CORPOICA partners, for accompanying me and giving me the necessary resources for the papaya collection in Valle del Cauca and Casanare, respectively. To my uncle Jaime Olarte for helping me in the adventurous collection in Santander.

Figure 1. Collection places of papaya samples used in this study. **A.** Colombian map showing the departments where the isolates were collected. Green: Casanare, red: Cundinamarca, pink: Santander, violet: Caldas, yellow: Risaralda and orange: Valle del Cauca. The number inside or near each department indicates the number of collected samples. **B.** Close up of the map showing in green the specific collection sites at each department.

Figure 2. Annealing sites of primers used in this study, over the genome sequence of the PRSV isolate from Hawaii (PRSVP-HA). The primer sequences and melting temperature is also shown. Primers that amplify the HC-Pro gene are shown in green. Primers that amplify the CP gene are shown in purple (Image modified from Tripathi, *et al.,* 2008).

Figure 3. Symptoms observed in papaya fields in Colombia. **A.** Papaya tree showing foliar area reduction. **B.** Fruit ringspot. **C.** Leaf Mosaic. **D.** Distortion of young leaves. **E.** Chlorosis.

Figure 4. Phylogenetic tree of the CP gene using an Uncorrelated LogNormal relaxed molecular clock model. Black dots represent PRSV-W isolates, gray dots represent unknown isolates suspected to belong to the PRSV-W group. Black arrow indicate the Mexican isolate that gave origin to the Australian clade. In pink, Colombian isolates. In orange, the isolates from Venezuela. Blue horizontal bars represent the 95% HPD interval for the divergence time estimates. Branches colored in red are not well supported (posterior probabilities lower than 0,65).

Figure 5. A. ML tree under the GTR+I+G for the CP gene. **B.** ML tree under the TIM+I+G for the HC-Pro gene. Above the branches, the bootstrap values for a 1000 replicates is shown. External group: Potato Virus Y (PVY).

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control* 19(6), 716-723.

Akaike, H., Petrov, B. N., and Csaki, F. (1973). Second international symposium on information theory. *Information Theory and an Extension of the Maximum Likelihood Principle. Akademiai Kiado: Budapest. p* 267.

Ballut, L., Drucker, M., Pugniere, M., Cambon, F., Blanc, S., Roquet, F., Candresse, T., Schmid, H. P., Nicolas, P., and Gall, O. L. (2005). HcPro, a multifunctional protein encoded by a plant RNA virus, targets the 20S proteasome and affects its enzymic activities. *Journal of General Virology* 86(9), 2595.

Bateson, M. F., Henderson, J., Chaleeprom, W., Gibbs, A. J., and Dale, J. L. (1994). Papaya ringspot potyvirus: isolate variability and the origin of PRSV type P (Australia). *Journal of General Virology* 75(12), 3547-3553.

Bateson, M. F., Lines, R. E., Revill, P., Chaleeprom, W., Ha, C. V., Gibbs, A. J., and Dale, J. L. (2002). On the evolution and molecular epidemiology of the potyvirus Papaya ringspot virus. *Journal of General Virology* 83(10), 2575.

Chin, M., Rojas, Y., Moret, J., Fermin, G., Tennant, P., and Gonsalves, D. (2007). Varying genetic diversity of Papaya ringspot virus isolates from two time-separated outbreaks in Jamaica and Venezuela. *Archives of virology* 152(11), 2101-2106.

De La Rosa, M., and Lastra, R. (1983). Purification and partial characterization of papaya ringspot virus. *Phytopathol. Z* 106, 329-336.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4(5), 699.

Drummond, A. J., Rambaut, A., Shapiro, B., and Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* 22(5), 1185.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics* 5(1), 113.

Gonsalves, D. (1998). Control of papaya ringspot virus in papaya: a case study. *Annual Review of Phytopathology* 36(1), 415-437.

Greber, R. S. (1978). Watermelon mosaic virus 1 and 2 in Queensland cucurbit crops. *Australian Journal of Agricultural Research* 29, 1235–1245.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp.* Ser. 41:95-98.

Inoue-Nagata, A. K., Franco, C. M., Martin, D. P., Rezende, J. A. M., Ferreira, G. B., Dutra, L. S., and Nagata, T. (2007). Genome analysis of a severe and a mild isolate of Papaya ringspot virus-type W found in Brazil. *Virus Genes* 35(1), 119-127.

Jain, R. K., Sharma, J., Sivakumar, A. S., Sharma, P. K., Byadgi, A. S., Verma, A. K., and Varma, A. (2004). Variability in the coat protein gene of Papaya ringspot virus isolates from multiple locations in India. *Archives of virology* 149(12), 2435-2442.

Jensen, D. D. (1949). Papaya virus diseases with special reference to papaya ringspot. *Phytopathology* 39, 191-211.

Maia, I. G., Haenni, A., and Bernardi, F. (1996). Potyviral HC-Pro: a multifunctional protein. *Journal of General Virology 77*(7), 1335-1341.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). "Molecular cloning." Cold Spring Harbor Laboratory Cold Spring Harbor, NY.

Muller, A. S. (1941). El reconocimiento de las enfermedades de las plantas cultivadas en Venezuela, 1937-1941. *Boletín Sociedad Venezolana de Ciencias Naturales* 8(48), 99-113.

Noa-Carrazana, J. C., González-de-León, D., and Silva-Rosales, L. (2007). Molecular characterization of a severe isolate of papaya ringspot virus in Mexico and its relationship with other isolates. *Virus Genes* 35(1), 109-117.

Páez, A. R. (2003). Manejo del Virus de la Mancha Anular de la Papaya en la Región Caribe Colombiana. *Corpoica, Boletín Técnico.* 8.

Piganeau, G., Gardner, M., and Eyre-Walker, A. (2004). A broad survey of recombination in animal mitochondria. *Molecular Biology and Evolution* 21(12), 2319-2325.

Pirone, T. P. (1991). Viral genes and gene products that determine insect transmissibility. *Seminars in Virology* 2, 81-87.

Pirone, T. P., and Blanc, S. (1996). Helper-dependent vector transmission of plant viruses. *Annual Review of Phytopathology* 34(1), 227-247.

Posada, D., and Crandall, K. A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* 14(9), 817.

Purcifull, D. E., Edwardson, J., Hiebert, E., and Gonsalves, D. (1984). Papaya ringspot virus. *CMI/AAB Descriptions of plant viruses* 292(8).

Rambaut, A., Pybus, O. G., Nelson, M. I., Viboud, C., Taubenberger, J. K., and Holmes, E. C. (2008). The genomic and epidemiological dynamics of human influenza A virus. *Nature* 453(7195), 615-619.

Revers, F., Le Gall, O., Candresse, T., and Maule, A. J. (1999). New advances in understanding the molecular biology of plant/potyvirus interactions. *Molecular Plant-Microbe Interactions* 12(5), 367-376.

Rezende, J. A. M., and Pacheco, D. A. (1998). Control of papaya ringspot virus-type W in zucchini squash by cross-protection in Brazil. *Plant Disease* 82(2), 171-175.

Rezende, J. A. M., Yuki, V. A., Vega, J., Scagliusi, S. M. M., Borba, L. F., and Costa, A. S. (1994). Isolados fracos do potyvirus causador do mosaico da abobrinha presentes em bolhas atuam na premunização. *Fitopatologia Brasileira* 18(1), 55-61.

Rozas, J., and Rozas, R. (1995). DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *Bioinformatics* 11(6), 621-625.

Sánchez de Luque, C., and Martínez López, G. (1998). El virus de la mancha anular de la papaya (Papaya ringspot potyvirus, PRSV) en Colombia.

Shimodaira, H., and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16, 1114-1116.

Shukla, D. D., Ward, C. W., and Brunt, A. A. (1994). "The potyviridae." CAB International Wallingford, UK.

Silva-Rosales, L., Becerra-Leor, N., Ruiz-Castro, S., Teliz-Ortiz, D., and Noa-Carrazana, J. C. (2000). Coat protein sequence comparisons of three Mexican isolates of papaya ringspot virus with other geographical isolates reveal a close relationship to American and Australian isolates. *Archives of virology* 145(4), 835-843.

Smith, J. M. (1992). Analyzing the mosaic structure of genes. *Journal of Molecular Evolution* 34(2), 126-129.

Swofford, D. L. (2002). PAUP*: phylogenetic analysis using parsimony (* and other methods), version 4.0 b10. *Sinauer, Sunderland, Massachusetts, USA*.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3), 585-595.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24(8), 1596.

Téliz-Ortiz, D., Mora-Aguilera, A., Nieto-Angel, D., Gonsalves, D., García-Pérez, E., Matheis-Toledano, L., and Ávila-Resendiz, C. (1991). Mancha anular del papayo en México. *Revista Mexicana de Fitopatología, México* 9(1), 64-68.

Tennant, P. F., Gonsalves, C., Ling, K. S., Fitch, M., Manshardt, R., Slightom, J. L., and Gonsalves, D. (1994). Differential protection against papaya ringspot virus isolates in coat protein gene transgenic papaya and classically cross-protected papaya. *Phytopathology (USA)*.

Thomas, J. E., and Dodman, R. L. (1993). The first record of papaya ringspot virus-type P from Australia. *Australasian Plant Pathology* 22, 2-2.

Tripathi, S., Suzuki, J. Y., Ferreira, S. A., Gonsalves, D., and Usda, A. R. S. (2008). Papaya ringspot virus-P: characteristics, pathogenicity, sequence variability and control.

Urcuqui-Inchima, S., Haenni, A. L., and Bernardi, F. (2001). Potyvirus proteins: a wealth of functions. *Virus Research* 74(1-2), 157-175.

Wang, R. Y., Powell, G., Hardie, J., and Pirone, T. P. (1998). Role of the helper component in vector-specific transmission of potyviruses. *Journal of General Virology* 79(6), 1519-1524.

1 Yeh, S. D., and Gonsalves, D. (1984). Evaluation of induced mutants of papaya ringspot virus
2 for control by cross protection. *Phytopathology* 74(9), 1086-1091.

Figure 1.

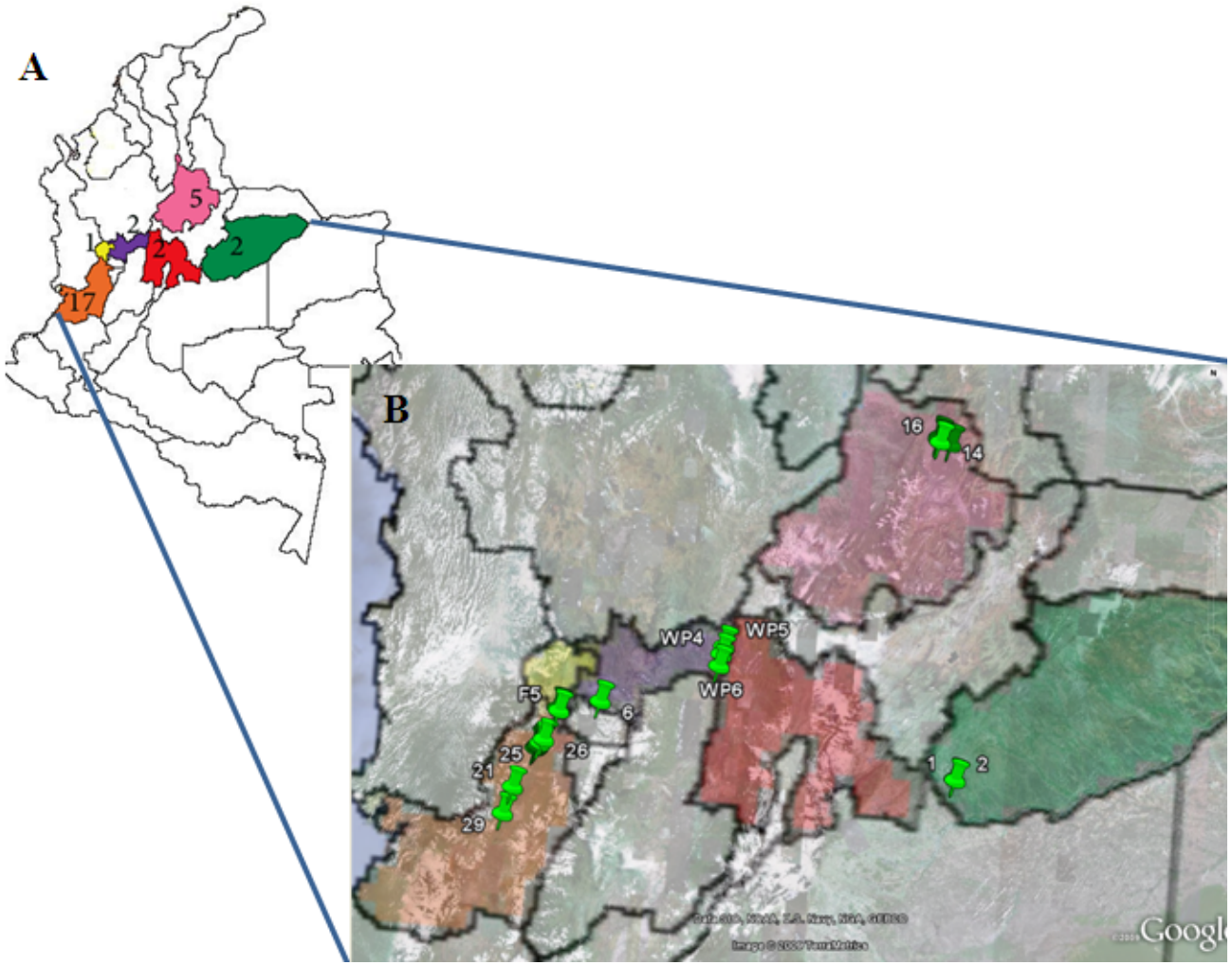


Figure 2.

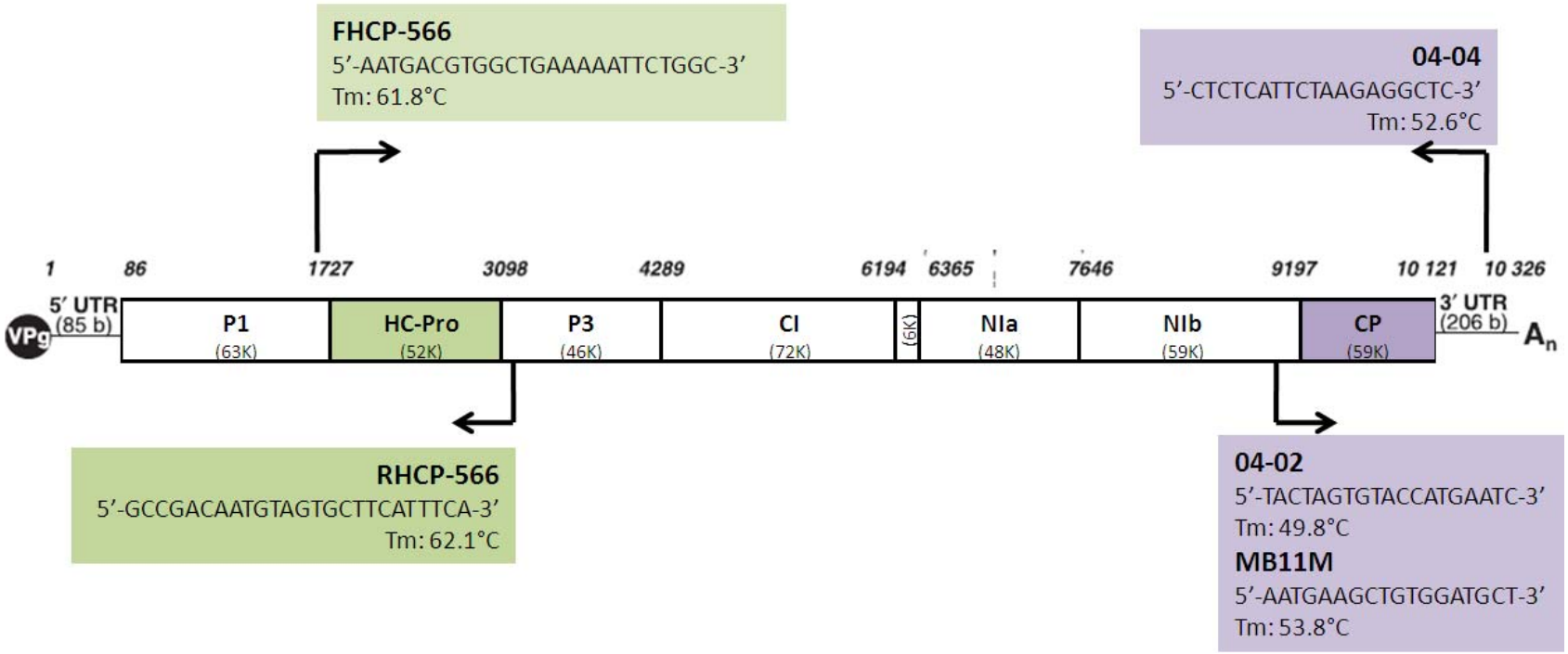

FHCP-566
5'-AATGACGTGGCTGAAAAATTCTGGC-3'
Tm: 61.8°C
04-04
5'-CTCTCATTCTAAGAGGCTC-3'
Tm: 52.6°C
1
86
1727
3098
4289
6194
6365
7646
9197
10 121
10 326
5' UTR
(85 b)
VPg
P1
(63K)
HC-Pro
(52K)
P3
(46K)
CI
(72K)
(6K)
NIa
(48K)
NIb
(59K)
CP
(59K)
3' UTR
(206 b)
An
RHCP-566
5'-GCCGACAATGTAGTGCTTCATTTCA-3'
Tm: 62.1°C
04-02
5'-TACTAGTGTACCATGAATC-3'
Tm: 49.8°C
MB11M
5'-AATGAAGCTGTGGATGCT-3'
Tm: 53.8°C

Figure 3.

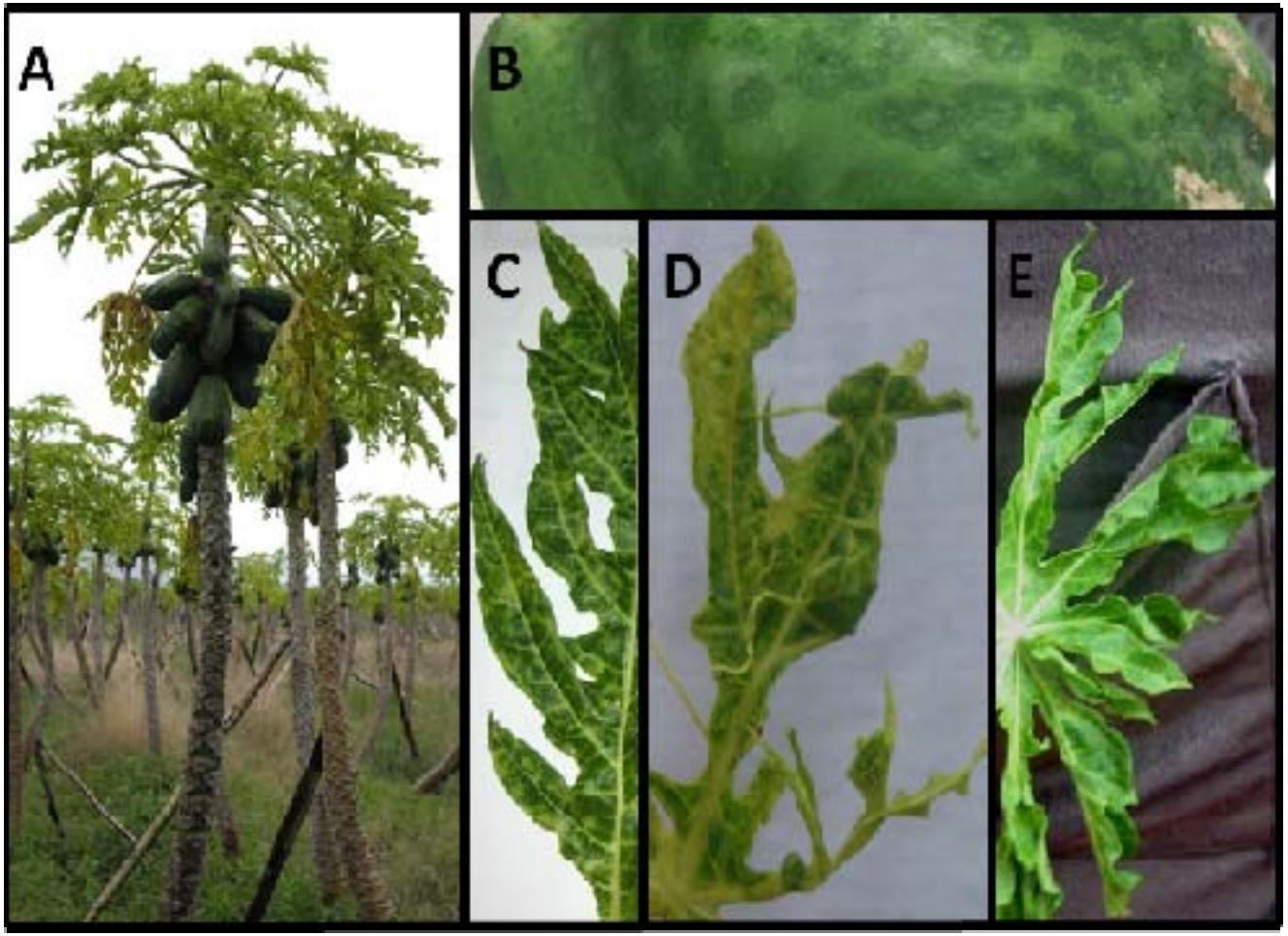


Figure 4.

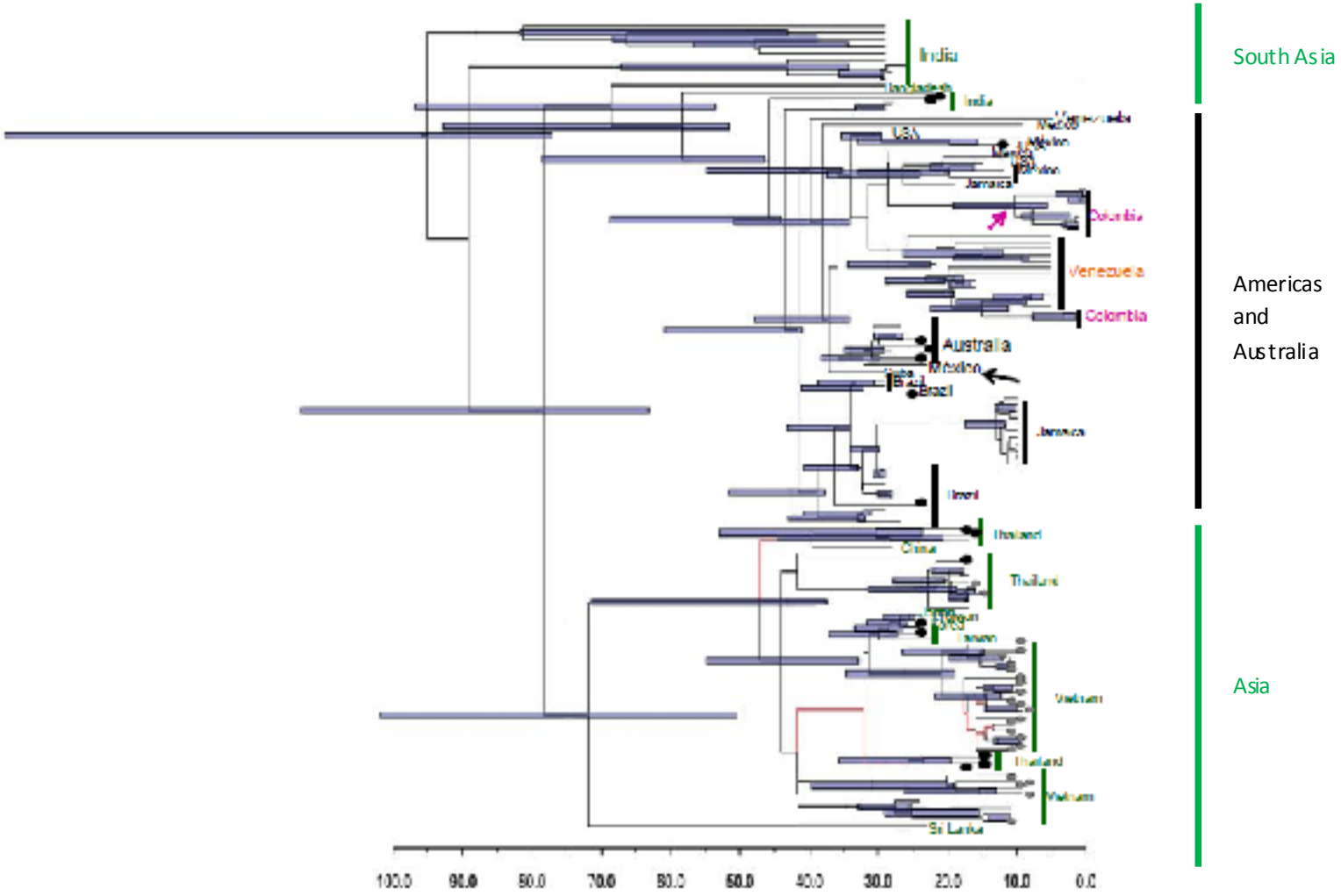
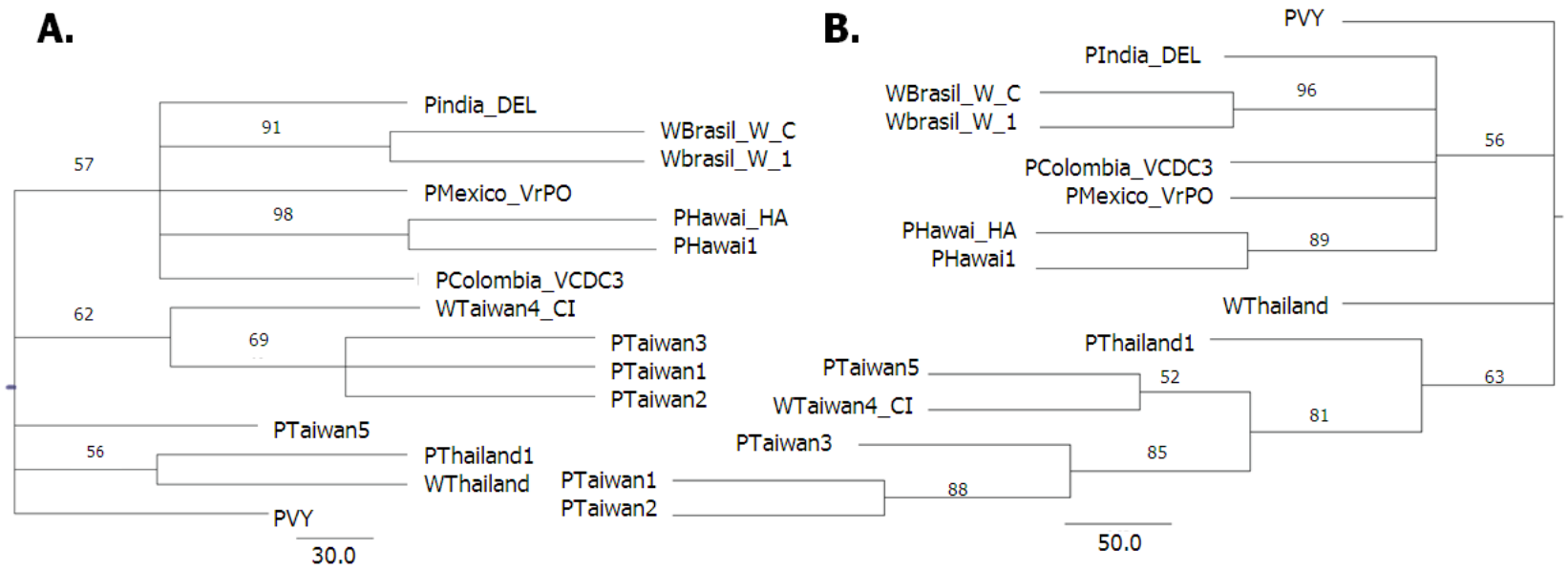

India
USA
Jamaica
Colombia
Venezuela
Colombia
Australia
Brazil
Jamaica
China
Thailand
Vietnam
Thailand
Vietnam
Sri Lanka
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
South Asia
Americas and Australia
Asia

Figure 5.


A.
Pindia_DEL
WBrasil_W_C
Wbrasil_W_1
PMexico_VrPO
PHawai_HA
PHawai1
PColombia_VCDC3
WTaiwan4_CI
PTaiwan3
PTaiwan1
PTaiwan2
PTaiwan5
PThailand1
WThailand
PVY
91
57
98
62
69
56
30.0
B.
PVY
PIndia_DEL
WBrasil_W_C
Wbrasil_W_1
PColombia_VCDC3
PMexico_VrPO
PHawai_HA
PHawai1
WThailand
PThailand1
PTaiwan5
WTaiwan4_CI
PTaiwan3
PTaiwan1
PTaiwan2
96
56
89
63
52
81
85
88
50.0

SUPPLEMENTARY MATERIAL

Table1. Description and characteristics of collections places.

Location		Georeference	Collection date
Casanare Between Barranca de Upia (Meta) and Villa Nueva, (Casanare)	Finca 1 Finca El Estepo	4°28'29''N, 72°52'35''W, 262 MASL	February 2008
	Finca 2 Finca Villa Arizona	4°28'09''N, 72°52'27''W, 262 MASL	
Norte de Santander	Finca 3	----------	
Risaralda	Finca 6 Finca La Carmelita. After Santa Rosa de Cabal.	4°51'31"N, 75°39'06''W, 1466 MASL	April 2008
Caldas	WP004 Over the road between Honda-La Dorada	05° 17'33.5'' N74°44'46.4'' W, 227 MASL	July 2008
	WP005	05° 21'15.6''N 74° 43'46.3'' W, 214 MASL	
Cundinamarca	WP006 Outside Honda, going to Mariquita, near the Gualí river.	05°12'09.0'' N 74° 45'21.3'' W, 270 MASL	
	17 Near the Tolemaida military base	----------	March 2009
Santander	13 Finca Girón On the road between Lebrija and Girón	07°05'15.15'' N73° 11'13.62'' W, 940 MASL	April 2009
	14 Outside Girón	07°04'19.75'' N 73° 10'38.53'' W, 751 MASL	
	15 Finca de Cecilia Outside Floridablanca	07°03'24.14'' N 73° 05'22.45'' W, 751 MASL	
	16Entering Girón	07°04'06.13'' N 73° 10'21.05'' W, 897 MASL	
	18 On the road to Bucaramanga's Airport (Palonegro)	07°06'22.85'' N 73° 11'03.92'' W, 1105 MASL	
	Finca 4 Finca Santa Rosa	4°46'42''N, 75°58'45''W, 938 MASL	April 2008
	Finca 5	4°,45'44''N, 75°58'3"W, 939 MASL	
Valle del Cauca (Municipio de Roldanillo, Corregimiento de Santa Rita)	21 Before Roldanillo	4°,08'51''N, 76°16'47''W, 935 MASL	April 2009
	22 Before Roldanillo	----------	
	23 After Roldanillo	4°,27'40''N, 76°7'34"W, 914 MASL	
	24 After Roldanillo	----------	
	25 After Roldanillo	4°,30'14''N, 76°5'29"W, 925 MASL	
	27 After Roldanillo, arriving to La Unión	4°,29'08''N, 76°6'19"W, 909 MASL	
	28 After Roldanillo	4°,28'26''N, 76°6'48"W, 912 MASL	
	26 After La Unión	4°,31'56''N, 76°4'3''W, 907 MASL	
	29 Between Roldanillo – Buga on the road to Buenaventura	3°56'17"N, 76°20'29''W, 960 MASL	
	Muestra de Adriana Bernal	----------	----------
	ICV	----------	February 2009
	Cachipay distributor	----------	

Table 2. PRSV sequences used in this study. When known, the year of collection is shown in parenthesis under each acronym or geographic origin.

Acronym	Geographic origin	Accesion number	Reference
USA_Haw	USA	X67673	Wang *et al*, (1994)
USA_Florida (1997)		AF196839	Davis, M. J. & Ying, Z. (1999)
WUS_FL**	Australia	D00594	Quemada *et al*. (1990)
Aust1		U14736	Bateson *et al*. (1994)
Aust2		U14737	
Aust3		U14738	
WAust_NT**		U14744	
WAust_DB1**		S89893	
WAust_GAT**		U14739	
Thail1	Thailand (1992)	AB044340	Unpublished
Thail2		U14743	Bateson *et al*. (1994)
Thail3		AF506898	Bateson *et al*. (2002)
Thail4		AF506899	
Thail5		AF506900	
Thail6		AF506901	
Thail7		AF506902	
WThail3**		AF506895	
WThail4**		AF506894	
WTahil6**		AF506892	
WTahil5**		AF506893	
WTahil7**		AF506891	
WTahil8**		AF506890	
UTahil10*		AF506896	
UTahil9*		AF506897	
Taiwan1	Taiwan	AB044341	Unpublished
Taiwan2		X78557	Wang *et al*, (1994)
WTaiwCI**		AY027810	
Jap1	Japan	AB044339	Unpublished
China	China	AF243496	
Cuba	Cuba	S46722	
Bangla	Bangladesh	AY423557	Jain *et al.* (2004)
WSKorea	South Korea	AB369277	
Sri1	Sri Lanka	U14741	Bateson *et al*. (1994)
Vietnam1	Vietnam	U14742	Bateson *et al*. (1994)
Vietnam2	Vietnam (1998)	AF506862	Bateson *et al*. (2002)
Vietnam3		AF506889	
UVietnam50*		AF506841	
UVietnam48*		AF506843	
UVietnam46*		AF506903	
UVietnam45*		AF506846	
UVietnam43*	Vietnam (1998)	AF506848	Bateson *et al*. (2002)
UVietnam41*		AF506850	
UVietnam40*		AF506851	
UVietnam39*		AF506852	
UVietnam3*		AF506853	
UVietnam37*		AF506854	
UVietnam36*		AF506855	
UVietnam35*		AF506856	
UVietnam32*		AF506859	
UVietnam31*		AF506860	
UVietnam30*		AF506861	
UVietnam42*		AF506849	
UVietnam47*		AF506844	
UVietnam44*		AF506847	
UVietnam33*		AF506858	

Table 2. ***(Continued)***

Acronym	Geographic origin	Accesion number	Reference
India	India	AF063220	Jain *et al.* (1998)
WIndia**		AF063221	
India_AP		AY238880	Jain *et al.* (2004)
India_CG		AY491011	
India_DL		AY238883	
India_HP		AY458617	
India_JK		AY458619	
India_KA1		AY238884	
India_KA3		AY458618	
India_UPLK		AY458620	
India_UPV		AY238882	
India_WB		AY238885	
WIndia1**		AF506845	Bateson *et al.* (2002)
WIndia2**		EU475877	
Mexico1 (1997)	Mexico	AJ012649	Silva-Rosales *et al.* (2000)
Mexico2 (1994)		AJ012099	
Mexico3 (1995)		AJ012650	
Mexico4 (1998)		AF309968	Unpublished
Mexico5 (2000)		AF319493	
Mexico6		AF319499	
Mexico7 (2000)		AF319502	
Jama1 (1990)	Jamaica	DQ104823	Tennant, (1996)
Jama2 (1999)		DQ104822	Chin *et al.* (2007)
Jama3 (1999)		DQ104821	
Jama4 (1999)		DQ104820	
Jama5 (1999)		DQ104819	
Jama6 (1999)		DQ104818	
Jama7 (1999)		DQ104817	
JamWoodburne (1999)		DQ104816	
JamSpring_Field (1999)		DQ104815	
JamPondside (1999)		DQ104814	
JamPamphret (1999)		DQ104813	
JamBelvedere (1999)		DQ104812	
Brasil1	Brasil	AF344640	Lima *et al.* (2002)
Brasil2		AF344647	
Brasil_BP		AF344645	
Brasil_PE		AF344646	
Brasil_BA_CA		AF344641	
Brasil_BA_IT1		AF344639	
Brasil_DF		AF344650	
Brasil_SP		AF344642	
Brasil_ES	Brasil	AF344644	Lima *et al.* (2002)
Brasil_PR		AF344643	
WBras1**		DQ374153	
WBrasC**		DQ374152	
Ve_MerVigia (1993)	Venezuela	DQ339576	Chin *et al.* (2007)
Ve_MerLagunillas (1993)		DQ339577	
Ve_Lara (2004)		DQ339580	
Ve_Merida4 (2004)		EF189730	
Ve_Mer5 (2004)		EF189731	
Ve_Mer6 (2004)		EF189736	
Ve_Mer8 (2004)		EF189733	
Ve_Tachira (2004)		DQ339579	
Ve_Trujillo (2004)		EF189734	
Ve_Tru2 (2004)		DQ339578	
Ve_Trujillo5 (2004)		EF189735	
Ve_Zulia (2004)		DQ339581	
Ve_Zul7 (2004)		EF189732	

*= Virus biotype not confirmed, but PRSV-W is suspected.

**= PRSV-W

Table 3. Nucleotide diversity (π), Segregant Sites (S), Watterson's θ (θ_W) and Tajima's D (Tajima, 1989). NC = No Calculated, due to the need of a minimum of 4 sequences to do that analysis. The number of haplotypes is included in parenthesis.

Population		Number of sequences*	S	π	θ_W	D
India	All	13 (12)	101	0,07977	0,06823	-0,06992
	P	11 (10)	92	0,07681	0,06585	-0,11355
	W	2 (2)	46	0,07199	0,07199	NC
Brazil	All	12 (12)	76	0,03052	0,03920	-0,09384
	P	10 (10)	68	0,03202	0,03744	-0,05404
	W	2 (2)	16	0,02492	0,02492	NC
Venezuela		13 (13)	107	0,04565	0,05371	-0,07820
Jamaica		12 (5)	40	0,01102	0,02063	-0,10956
Mexico		7 (7)	129	0,07177	0,07587	-0,12496
Thailand	All	15 (15)	137	0,05331	0,06594	-0,06165
	P	7 (7)	64	0,03301	0,04088	-0,05563
	W	8 (8)	108	0,06455	0,06518	-0,05412
Colombia		9 (9)	61	0,03915	0,03507	-0,09172
USA		3 (3)	31	0,03219	0,03219	NC
Australia	All	6 (6)	29	0,011508	0,01981	-0,05823
	P	3 (3)	12	0,01248	0,01248	NC
	W	3 (3)	17	0,01765	0,01765	NC
Vietnam	All	25 (23)	138	0,03974	0,05693	-0,10078
	P	3 (3)	27	0,02804	0,02804	NC
	W	22 (20)	132	0,03670	0,05640	-0,12481
Taiwan		3 (3)	21	0,02191	0,2191	NC
Asia	All	61 (57)	186	0,06983	0,08332	-0,14325
	P	27 (26)	174	0,08401	0,07560	-0.02547
	W	35 (33)	204	0,05319	0,07752	-0,10005
Americas and Australia		63 (56)	234	0,05322	0,07771	-0,09784
All sequences		124 (112)	243	0,07533	0,09134	-0,11985