RESEARCH ARTICLE

Molecular adaptation within the coat protein-encoding gene of Tunisian almond isolates of *Prunus* necrotic ringspot virus

MONCEF BOULILA[1]*, SAWSSEN BEN TIBA[2†] and SAOUSSEN JILANI[3†]

[1]*Institut de l'Olivier, B.P. 14, 4061 Sousse Ibn-Khaldoun, Tunisia*
[2]*Institut Supérieur de Biotechnologie de Monastir, Avenue Tahar Hadded. B.P. 74, Monastir 5019, Tunisia*
[3]*Centre Régional de Transfusion Sanguine de Sousse, Centre Hospitalo-Universitaire Farhat Hached, Avenue Ibn El Jazzar-Sousse 4000, Tunisia*

Abstract

The sequence alignments of five Tunisian isolates of *Prunus* necrotic ringspot virus (PNRSV) were searched for evidence of recombination and diversifying selection. Since failing to account for recombination can elevate the false positive error rate in positive selection inference, a genetic algorithm (GARD) was used first and led to the detection of potential recombination events in the coat protein-encoding gene of that virus. The Recco algorithm confirmed these results by identifying, additionally, the potential recombinants. For neutrality testing and evaluation of nucleotide polymorphism in PNRSV CP gene, Tajima's *D*, and Fu and Li's *D* and *F* statistical tests were used. About selection inference, eight algorithms (SLAC, FEL, IFEL, REL, FUBAR, MEME, PARRIS, and GA branch) incorporated in HyPhy package were utilized to assess the selection pressure exerted on the expression of PNRSV capsid. Inferred phylogenies pointed out, in addition to the three classical groups (PE-5, PV-32, and PV-96), the delineation of a fourth cluster having the new proposed designation SW6, and a fifth clade comprising four Tunisian PNRSV isolates which underwent recombination and selective pressure and to which the name Tunisian outgroup was allocated.

[Boulila M., Ben Tiba S. and Jilani S. 2013 Molecular adaptation within the coat protein-encoding gene of Tunisian almond isolates of *Prunus* necrotic ringspot virus. *J. Genet.* **92**, 11–24]

Introduction

Prunus necrotic ringspot virus (PNRSV) is one of the most economically important and widely distributed *Prunus* virus. It exists as a number of strains that cause a diversity of symptoms on many hosts (Minck 1980; Minck and Aichele 1984). The severity of damage also depends on plant species, cultivar, and environmental conditions. Damage caused by infection from PNRSV may result in death of buds and roots, reduced growth in young trees (of special concern to nursery growers), and reduction in the number of fruit (Minck 1991; Nemeth 1986). PNRSV is a positive-sense RNA plant virus with a tripartite genome that belongs to genus *Ilarvirus* (Rybicki 1995), which has the same genomic organization, encoding functionally similar translation products, as members of the genera Bromovirus, Cucumovirus, Anulavirus, Oleavirus and Alfamovirus, belong to the family Bromoviridae (King *et al.* 2011). The RNA 1 and RNA 2 are monocistronic and encode nonstructural proteins (P1 and P2) involved in RNA replication (Van Vloten-Doting *et al.* 1981). RNA 3 is bicistronic and encodes two polypeptides, a polypeptide with homology to proteins required for cell-to-cell movement of plant viruses (ORF 3a) and the coat protein (ORF 3b) (Bachman *et al.* 1994; Guo *et al.* 1995; Hammond and Crosslin 1995; Scott and Ge 1995). The movement protein (MP or P3) is translated from RNA 3, while the coat protein (CP) synthesis occurs via a subgenomic monocistronic mRNA (RNA 4), which is collinear with the 3′ end of RNA 3 and is encapsidated. Di Terlizzi *et al.* (2001) reported that they found a small RNA (RNA 5) which was purified and determined as a biologically inactive copy of the 3′-UTR of the genomic RNA 3.

In order to survive, viruses must evolve and adapt. RNA viruses infecting cultivated plants have high evolutionary potential. This potential results from a large mutation and recombination rates and, in case of multipartite genomes,

*For correspondence. E-mail: boulila.moncef@yahoo.fr.
†These authors contributed equally to this work.

Keywords. bioinformatics; molecular genetics; evolution; sequence; phylogeny; *Prunus* necrotic ringspot virus.

such as PNRSV, the reassortment of segments during mixed infections. Evolution of RNA viruses has revealed that RNA recombination is a widespread phenomenon that has shaped numerous viruses by rearranging viral genomes or disseminating functional modules. Genetic recombination, the formation of chimeric molecules from segments previously separated on the same molecules or present on different parental molecules, is one of the most important mechanisms for generating novel genomes that may have selective advantages over parental genomes. Intermolecular RNA recombination can proceed via two fundamentally different mechanisms: replicative and nonreplicative. Nagy and Simon (1997) classified RNA recombinants into three major classes. (i) Class one is the base-pairing-dependent (similarity required) recombinants, which are generated by precise alignment between the primer RNA (the incomplete nascent strand derived from the donor template) and the acceptor RNA prior to reinitiation of primer-dependent-RNA synthesis by the viral replicase. Thus, sequence identity of 15–30 nt or longer between the donor and acceptor RNAs facilitates the generation of class one recombinants. (ii) In class two, recombinants are generated by a base-pairing-independent process (similarity-nonessential). During these recombination events *cis*-acting replication elements may facilitate rebinding the viral replicase and primer RNA to the acceptor RNA prior to reinitiation of primer-dependent RNA synthesis by the viral replicase. Generation of class two recombinants does not require sequence similarity between the donor and acceptor RNAs. (iii) In the third class of recombinants, the base-pairing-assisted recombinants are formed by a mechanism with features combining the first two classes.

Analysing the instantaneous rate of nonsynonymous (amino acid-changing) and synonymous (silent) nucleotide substitutions in protein-coding molecular sequences can give important clues to understand how they evolved. In particular, the ratio of the nonsynonymous and synonymous fixation rates have been used to measure the level of selective pressure on proteins (McDonald and Kreitman 1991). Synonymous mutations do not change the encoded protein and so are often assumed to be selectively neutral. If a nonsynonymous mutation does not affect the fitness of a protein, it would become fixed within the population at the same rate as a synonymous mutation, giving a nonsynonymous/synonymous rate ratio (ω) of 1. If a nonsynonymous change makes the protein more or less fit on average, then ω will be greater or less than 1, respectively (Yang and Bielawski 2000). If a site has $\omega > 1$, it is unusually variable and is said to have evolved under positive selection. Similarly, a site with $\omega < 1$ is unusually conserved and is said to have been subject to purifying selection.

The objective of this study was to estimate selection pressures and recombination signatures in the capsid-coding gene of five Tunisian isolates of PNRSV infecting almond as well as the evolutionary relationships among them and to other isolates occurring elsewhere.

Materials and methods

Virus source

Surveys were recently carried out in a collection variety, and samples were collected from five almond varieties (Louzi, Montrone, Texas, Mollese and Ghernghezel) grown in Sfax district (southern Tunisia), exhibiting typical almond mosaic characterized by a variety of symptoms ranging from bright chrome-yellow (calico) to chlorotic discolourations, localized necrosis of the leaf blade, leaf curling, bud failure, fasciations, rosetting, stunting, and bushy growth. Each sample consisted 10 shoots bearing leaves with virus symptoms and taken from the four compass points as well as from the internal part of the tree. In addition, the sampled 20 years-old trees were distant from one another at a space of 7 × 7 m.

Sample preparation

The plant samples (leaves) were ground (1:5, w/v) in a PBS-T buffer (1.44 g $Na_2HPO_4.2H_2O$, 8 g NaCl, 0.2 g KH_2PO_4, 0.2 g KCl, 0.2 g NaN_3, 0.5 mL Tween 20) routinely used for ELISA assays, supplemented with 2% polyvinylpyrrolidone K25 (PVP K25) and 0.2% sodium diethyldithiocarbamate (DIECA) in individual plastic bags containing gauze (Bioreba) using a rolling grinder (Homex 6) (Bioreba, Reinach BL1, Switzerland). After grinding, the extracts were rapidly clarified by centrifugation (825 × g for 5 min at 4°C) using Sigma 1–15 K (DJB Labcare, Buckinghamshire, UK).

Primer design

The PNRSV primers were designed to flank the CP gene. The selected primers spanned part of the intercistronic region, the entire gene, and part of the 3′-UTR in the RNA 3. The primers were designed in a conserved region identified by using the similarity search tools WU-BLAST 2, and selected by the use of Primer3 software (Rozen and Skaletsky 2000).

The primers of PNRSV were located between nucleotides 989 and 1895 in the RNA 3 (Guo *et al.* 1995) with the sequences: 5′-TGTGACCGTTAAGGTCGGTA-3′, for the sense primer, and 5′-CGCAGGTAAGATTTCCAAGC-3′ for the antisense primer, to amplify specifically a 907-bp product.

Immunocapture-RT-PCR assay

One-hundred microlitres of clarified plant extract were submitted to an immunocapture (IC) performed directly in tubes used for reverse transcription PCR reaction (Wetzel *et al.* 1992). The coating of the IC tubes was carried out using 2 μg of immunoglobulins per millilitre. Incubation was carried out overnight at +4°C in both steps. Following three washes using PBS-Tween buffer, the tubes were dried and further processed for amplification. Fifty microlitres of RT-PCR mix (6 μL 1% Triton X-100, 5 μL 5× Green *GoTaq*

DNA polymerase buffer (Promega, Madison, USA), 1.5 μL 25 mM $MgCl_2$, 2.5 μL formamide 5%, 1 μL 10 mM dNTPs (2.5 mM dGTP, 2.5 mM dATP, 2.5 mM dCTP, 2.5 mM dTTP), 15 pmol of each of the two primers, 2.5 U AMV reverse transcriptase (Promega), 1 U *GoTaq* DNA polymerase (Promega), 10 U RNasin (Promega), 31 μL sterile water were added to each tube. The cycling scheme using the cycler PTC 100 (MJ Research, Watertown, USA) was the following: 50 min at 46°C (RT-reaction), 5 min at 94°C to denature the templates and the reverse transcriptase followed by 35 cycles of amplification: 60 s at 94°C (denaturation), 45 s at 46°C (annealing) and 75 s at 72°C (DNA synthesis). A final 15 min elongation step at 72°C was performed at the end of 35 cycles. Amplification products were analysed by electrophoresis of 10 μL aliquots on 1.5% agarose gel, in 1× Tris-Borate-EDTA buffer (Sambrook *et al.* 1989). Bands were visualized by ethidium bromide staining (5 $\mu g.mL^{-1}$) and photographed using a UV transilluminator (ETX 20.M) at a wavelength of 312 nm and a Vilber Lourmat photo-print system (model DP-001.FDC) (Vilber Lourmat, France).

Sequencing of PCR products

After cleanup of PCR products with wizard cleanup system (Promega), amplicons were subject to nucleotide sequencing by the dideoxy chain termination (Sanger *et al.* 1977) method using the Big Dye Terminator Ready Reaction mix provided by Applied Biosystems (Foster City, USA) in an automated capillary sequencer 3130. Sequencing was carried out in both directions using the primers mentioned above. The resulting sequences were analysed using Chromas Lite (Technelysium, Helenvale, Australia).

Computer-assisted-sequence analysis

Sequence analysis and phylogeny: Databank search for homologies to PNRSV were performed using the FASTA (Pearson and Lipman 1988) and WU-BLAST 2, based on the Basic Local Alignment Search Tool algorithm (Altschul *et al.* 1990, 1994) programs.

The nucleotide sequences of Tunisian isolates added to those retrieved from GenBank were aligned using ClustalW 2.1, ClustalX 2.0.9 (Larkin *et al.* 2007) and MultAlign (Corpet 1988) softwares with default configuration. Their phylogenetic relationships were determined using neighbour joining (NJ) and maximum likelihood (ML) algorithms incorporated in MEGA program v5 (Tamura *et al.* 2011). Based on the evaluation of best fit substitution model executed in MEGA5, ML tree was reconstructed under assumption of the substitution model K2 coupled to a discrete gamma distribution (+G) with five rate categories (Kimura 1980). The substitution model parameters estimated were (i) base frequencies: $f(A) = f(T) = f(C) = f(G) = 0.25$, (ii) substitution rates: $r(AT) = 0.045$, $r(AC) = 0.045$, $r(AG) = 0.161$, $r(CG) = 0.045$, $r(CT) = 0.161$, $r(GT) = 0.045$, and (iii) transition/tranversion ratios: $R = 1.81$. The Bayesian information criterion value (BIC = 10211.055) with K2+G model was the lowest among 24 models tested. Bootstrap analysis for 500 replicates was performed to assess the robustness of the branches.

Recombination analysis: Potential recombination events between diverged nucleotide sequences were explored using two programs: genetic algorithm for recombination detection (GARD) (Kosakovsky Pond *et al.* 2006a, b), and Recco (Maydt and Lengauer 2006). GARD makes use of statistical approach to search recombination breakpoints from multiple-sequence alignments of homologous sequences. Potential breakpoints are identified by improvement of the small-sample corrected Akaike information criterion (AICc) (Akaike 1974) for phylogenetic trees constructed of individual recombinant fragments. Based on the outcome of the analysis, a level of support is assigned and expressed as a breakpoint placement score (Kosakovsky Pond *et al.* 2006a, b). Identified breakpoints by GARD were then assessed for significance by using KH test (Kishino and Hasegawa 1989) in HyPhy (hypothesis testing using phylogenies) package (Kosakovsky Pond *et al.* 2005). Recco, an algorithm developed and described by Maydt and Lengauer (2006) as a fast, simple and sensitive method for detecting recombination in a set of sequences and locating putative recombination breakpoints is based on cost minimization. This method has only two tunable parameters, recombination and mutation cost. In practice, the only parameter considered is α, representing the cost of mutation relative to recombination. When α changes from 0 to 1, the cost of mutation weighted by α increases, and the cost for recombination weighted by $1 - \alpha$ decreases. In other words, the parameter α controls the ambiguity between mutation and recombination.

RNA polymorphism and evolution: DnaSP v5.10.01 (Rozas *et al.* 2003) was used to estimate Tajima's *D* (Tajima 1989), and Fu and Li's *D* and *F* (Fu and Li 1993) statistical tests to examine the hypothesis of neutrality operating on the PNRSV CP sequences. We also estimated several population genetic parameters including nucleotide polymorphism (Π estimated by the average number of nucleotide differences between two random sequences in a population), haplotype diversity (*Hd*, the frequency and number of haplotypes in a population), the statistic θ from the number of segregation sites (*S*) (Watterson 1975), the average rate of synonymous and nonsynonymous substitutions, , Δ *Hd* (the variance of haplotype diversity), *K* (average of number of pairwise nucleotide differences). The distribution of *d*S and *d*N along the coding regions was analysed using the SNAP program (http://www.hiv.lanl.gov; Korber 2000). Based on the results obtained by the statistical tests mentioned above, examination for selection was performed using codon-based ML methods i.e., the single-likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), internal fixed

effects likelihood (IFEL), random effects likelihood (REL), mixed effects model of episodic selection (MEME), fast unbiased Bayesian approximation (FUBAR), genetic algorithm (GA) branch models (Kosakovsky Pond *et al.* 2005) and the partitioning approach for robust inference of selection (PARRIS) (Scheffler *et al.* 2006) implemented at http://www.datamonkey.org, the web server of HyPhy package (Delport *et al.* 2010).

Results

Immunocapture-reverse transcription-polymerase chain reaction

IC-RT-PCR successfully amplified the targeted genome portion of the five isolates collected. The expected size was obtained (907 bp) (figure 1) and the CP gene size was 686 bp as revealed in the sequences (figure 2).

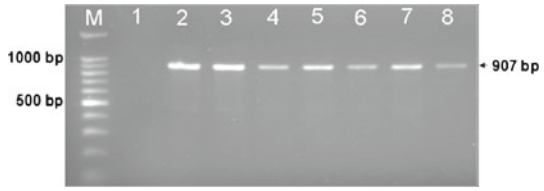

Figure 1. Agarose gel of *Prunus* necrotic ringspot virus PCR products. M, marker (100-bp ladder); lane 1, negative control; lane 2, positive control; lane 3, Montrone; lane 4, Louzi; lane 5, Ghernghezel; lane 6, Mollese; lane 7, Texas; lane 8 Mollese replicate. Bands correspond to amplicons having a size of 907 bp.

Detection of recombination events in PNRSV CP gene

GARD was used to search for putative breakpoint recombination delimiting sequence regions having distinct phylogenies. GARD placed breakpoints at base pairs 48, 122, 144 and 335 in aligned sequences of CP gene based on AICc goodness of fit. Based on Shimodaira–Hazegawa (SH) test using three P values (threshold of significance at $P = 0.1$, $P = 0.05$, $P = 0.01$) flanking topologies were not significantly discordant suggesting a priori that other processes (e.g. substitution rate heterogeneity in the CP gene) may be contributing to phylogenetic variation before and after the breakpoint (table 1). Besides detection of potential recombination sites, we also searched for the possible recombinants; to this end, we used the Recco program. The analysis yielded one (Montrone, Ghernghezel, Texas), and five (Mollese, Louzi) recombination signals. It is noteworthy that the breakpoint interval varied from a single-point recombination (point mutation) such as for Ghernghezel (145 nt), Mollese (691 and 694 nts), and Louzi (475, 476 and 480 nts); double point for Mollese (593–594), and Louzi (570–571), to much wider intervals, i.e. varying from three (Louzi (471–473), Mollese (603–605)), six (Mollese (114–119)), eight (Montrone (537–544)), to 50 (Texas (305–354)).

Sequence divergence comparison and ML estimate of substitution matrix

Sequence comparisons of the CP gene among Tunisian isolates revealed divergence of 79–89% at nucleotide level.

```
                1                                                                                                   130
      Texas ATGGTTTGCC GAATTTGCAA TCATACCCAC --GCTGGTGG ATGCCGTTCT TGCAAGAAGT GCCATCCGAA TGGAGCTCTG GTCCCACTCA GGGCTCAACA GAGGGCTGCG AATAACCCGA ATAGAAA---
 GhernGheze ATGGTTTGCC GAATCTGCAA TCATACCCAC CCTCTGAAAA GTGCCGTTCT TGCAAGAAGT GCCATCCGAA TGATGCTCTG ATCCCACTCA GGGCTCAACA AAGGGCTGCG AATAACCCGA ATAGGAATAG
      Louzi ATGGTTTGCC GAATTTGCAA TCATACCCAC --GCTGGTGG GTGCCGTTCT TGCAAGAAGT GCCATCCGAA TGGAGCTCTG GTCCCACTCA GGGCTCAACA GAGGGCTGCA AATAACCCGA ATAGAAA---
    Mollese ATGGTTTGCC GAATTTGCAA TCATACTCAC --GCTGGTGG ATGCCGTTCT TGCAAGAAGT GCCATCCGAA TGGAGCTCTG GTCCCACTCA GGGCTCAACA GAGGGCTGCA AATAACCCGA ATAGAAA---
   Montrone ATGGTTTGCC GAATTTGCAA TCATACCCAC --GCTGGTGG ATGCCGTTCT TGCAAGAAGT GCCATCCGAA TGGAGCTCTG GTCCCACTCA GGGCTCAACA GAGGGCTGCA AATAACCCGA ATAGAAA---
  Consensus ATGGTTTGCC GAATtTGCAA TCATAccCAC ..gCTGgtgg aTGCCGTTCT TGCAAGAAGT GCCATCCGAA TGgaGCTCTG gTCCCACTCA GGGCTCAACA gAGGGCTGCa AATAACCCGA ATAGaAA...

                131                                                                                                 260
      Texas ---CCCGAAT AGGGCTTCGA GTGG-TACCG GACCAGTGGT CCGACCACAA CCGGTCGTGA AGACCACCTG GACCGTGAGA GGTCCGAATG TGCCTCCCCG AATTCCTAAG GGG-TTTGTA GCACACAATC
 GhernGheze GAACCCGAAT AGGGTTTCGA GCGG-TATAG GACCTGCGGT CCGACCGCAA TCGGTCGTTA AGACCACTTG GACCGTGAGG GGTCCGAATG TGCCTCCCCG AATCCCCAAG GGT-TATGTA GCACATAATC
      Louzi ---CCCGAAT AGGGCTTCGA GTGGGTACCG GACCAGTGGT CCGACCACAA CCGGTCGTGA AGACCACCTG GACCGTGAGA GGTCCGAATG TGCCTCCCCG AATTCCTATG GGGGTTTGTA CCACACAATC
    Mollese ---CCCGAAT AGGGCTTCGA GTGGGTACCG GACCAGTGGT CCGACCACAA CCGGTCGTGA AGACCACCTG GACCGTGAGA GGTCCGAATG TGCCTCCCCG AATTCCTAAG GGG-TTTGTA GCACACAGTC
   Montrone ---CCCGAAT AGGGCTTCGA GTGG-TACCG GACCAGTGGT CCGACCACAA CCGGTCGTGA AGACCACCTG GACCGTGAGA GGTCCGAATG TGCCTCCCCG AATTCCTAAG GGG-TTTGTA GCACATAATC
  Consensus ...CCCGAAT AGGGcTTCGA GtGG.TAccG GACCaGtGGT CCGACCaCAA cCGGTCGTgA AGACCACcTG GACCGTGAGa GGTCCGAATG TGCCTCCCCG AATtCCtAaG GGg.TtTGTA gCACAcAaTC

                261                                                                                                 390
      Texas ACCGAGAGGT GACGACGACA GAGGCAGTGA AGTACTTGAG CATAGACTTC ACGACCACTC TCCC-TCAGT TGATGGGTCA GAATTT-GAC CCTATT-GAC GGT-ATACTC CGAATGGA-T CTATGAGTTC
 GhernGheze ACCGAGAGGT GACGACGACA GAGGCAGTGA AGTACTTGAG CATTGACTTC ACGACCACTC TCCC-TCAGT TGATGGGTCA GAATTT-GAC CTTATT-AAC TGTCATAGTC CGAATGAACT CTATGAGTTC
      Louzi ACCGAGAGGT GCCGACGACA GAGGCAGTGG AGTACTTCAG CATTGACTTC ACGACCACTC TCCCCTCAGT TGATGGGTCA GAATTT-GAC CCTATT-GAC TGTTATAGTC CGAATGAACT CTATGAGTTC
    Mollese ACCGAGAGGT GACGACGACA GAGGCAGTGA AGTACTTGAG CATAGACTTC ACGACCACTC TCCC-TCAGT TGATGGGTCA GAATTT-GAC CCTATT-GAC TGTTATAGTC CGAATGAACT CTATGAGTTC
   Montrone ACCGAGAGGT GACGACGACA GAGGCAGTGA AGTACTTGAG CATAGACTTC ACGACCACTC TCCC-TCAGT TGATGGGTCA GAATTTTGAC CCTTTTTGGC TGTTATACTC CGAATGTACT CTAAGAATTC
  Consensus ACCGAGAGGT GaCGACGACA GAGGCAGTGa AGTACTTgAG CATaGACTTC ACGACCACTC TCCC.TCAGT TGATGGGTCA GAATTT.GAC CcTaTT.gaC tGTtATAgTC CGAATGaAcT CTAtGAgTTC

                391                                                                                                 520
      Texas GAA-TGGT-- GGATGGGATG G-TGGA-GGA CTATAAGGTG GAACAACCTG ATG-TCC-AA TG--CCTGTC TAGGA---GG GTT-TTGCAG GA--CAAC-- GAGAGGT--G GCGTTCAACT CCTCCAATTA
 GhernGheze GAA-TGGTTG GATTGGGATG G-TGGAGGAC TTATACGGGG GATCA-CCTG ATGGTCCGAA TG-CCCTGTC TAGGAAG-GG GTTCTTGAAG GACCCAACCA AGGAGGTTGG GCAGTTCGAA CCTCCCTTCC
      Louzi GAA-GGGTGG GATGGGGATG G-TGGA-GGA CTATAAGGGG AACAACCCTG ATGGTCCGAA TG-CCCTGTC CAGGAAA-GG GTTCTTGAAG GAACCAACCA AAGAGGTGGG GCAATTCGAA CCTCCTTCCC
    Mollese GAA-TGGTTG GATTGGGATG G-TGGA-GGA CTATAAGGTG GAACAACCTG ATGGTCCGAA TG-CCCTGTC TAGGAAA-GG GTTCTTGGAG GAACCAACCG AGGAGGTTGG GCAGTTCGAA CCTCCCTTCC
   Montrone GAAAGGGTTG GATTGGGATG GGTGGAAGGA CTATAGGGGG GAACAACCTG CTGGTCCGAA TGGCCCTGTT TAGGAAAAGG GTTCTTGAAG GA-CCAACCG AA-AGGTTGG GAGTTCCAAA CCTCCTT-CC
  Consensus GAA.tGGTtg GattGGGATG G.TGGA.Gga cTATAaGGgG gAacAaCCTG aTGgTCCgAA TG.cCCTGTc tAGGAaa.GG GTTcTTGaAG GA.cCAACc. aagAGGTtgG Gca.Ttcgaa CCTCC.ttcc

                521                                                                                                 650
      Texas GACTCGACCT TTGCCGTCGC TCGGCCTTAT GAATTCAGAC AAATGCCCTG GGCCAAGTCT GGTGGGATTT GTCTATGTAA GCCCAGTGTT CCACTGAGGC TGCTGCATTA TACTGCTGGA ATTAAAATGA
 GhernGheze GATTTAGATT TCGACACTTT T-GCGCGTAC GCATCGGG-- ------CGTC ATCGAAT--C AGAACGAAGT GCCCGCTGGG GGCCAGGTCT TGGTTATGG- -----ATTTG TAACGTACTG GTTAAGTGAC
      Louzi GATTTAGATT TCGAACCTTT T-GCGCGAAC GCATCGGG-- ----GCGGTT TTCGAATTCA AGAACCAAAT GCCCCCGGGG --CCAAGTTT TGGTTA-GG- ----GATTTG TATCTTGGGG GTTACACGGA
    Mollese GATTTAGATT TCGACACTTT T-GCGCGTAC GCATCGTG-- ----GTCGTT ATCGAATTCA AGAACGGAGT GCCCGCTGGG GCCCAAGGTC TTGTTA-GG- ----GATTTG TACCGTAGTG GTTAAGCGAA
   Montrone AATTTACATT TCGACCCTTT TTGCCCGACC GCATCG-G-- ----GCCTTT TTCAAATTCA AGACCAAAG- GCCCCCTGGG G-CCAAGGTT TGGTTATGG- ----AATTTG TACAGTAGGG GTAATGCGAA
  Consensus gAtTtagatT TcGac.cttt T.GcgCgtac GcATcg.G.. ....gccgTt .tCgAAttca aGaacgaagt GcCc.ctggg g.CCAaggtt tggtTa.GG. ....gatTTg TAccgtaggg gTtAagcgaa

                651                                                                                                 720
      Texas TAGCACATCG ATTTCTCGAT GAGACTGCGC CACCAATTGT ATTCCAATTA TCCCAAGGCT TCAGCATTAG
 GhernGheze TTACCCC--- GATGCAAATT CCGAACTGAA GTCCT-TGGC TGTCAATATG AACCT--GCT GGAATCTAG
      Louzi TTTTCCCCGA GGGCAAATTC CCAACTGGAA GTCCT-GGGG GGTCATAAAA GACCT--GGT TGAATTTAG
    Mollese TTACCACAGA AGTGCAATTT CCAACTGGAT GTCCT-TGGT GGTCGATGAG AACCT--GCT GGGATCTAA
   Montrone TTACCAC-GA GGTGCAATTT CCAACCGGTT CTGCT-TGGT GGTC-CAGAA AACTC--GTT TAAACTTAG
  Consensus Ttacccc.ga ggtgcaattt ccaActgga. gtcCt.tgGt ggTC.aa.aa aaCct..GcT tgaat.Tag.
```

Figure 2. Nucleotide sequence alignment of CP-encoding gene of five almond isolates of *Prunus* necrotic ringspot virus collected in Tunisia performed by using MultiAlign program. Dashes indicate alignment optimization.

Table 1. Evidence of recombination across PNRSV CP gene determined by GARD a genetic algorithm. Khishino–Hasegawa (KH) tests verified the significance of breakpoints estimated by GARD analysis.

AICc	ΔAICc	Breakpoint location	*P* value (LHS)	*P* value (RHS)	Signficance
7756.96	3.21357	48	1.00000	1.00000	NS
		122	1.00000	1.00000	NS
		144	1.00000	1.00000	NS
		335	1.00000	1.00000	NS

KH test was used in both directions to compare phylogenies constructed from the alignment segment to the left hand side (LHS) and right hand side (RHS) of each estimated breakpoint. All *P* values have been adjusted by Bonferroni correction. AICc, Akaike information criterion corrected, a measure of model accuray; ΔAICc, difference between two AIC corrected values for two nested models. NS, not significant.

The highest divergence was observed between Texas and Ghernghezel isolates (identity rate = 79%), whereas the lowest divergence was between Mollese and Louzi isolates (identity rate = 89%). Likewise, compared to 82 foreign isolates (table 2), Texas revealed to be the most divergent (on average: 75.42% of identity), while Mollese the least divergent (on average: 90.84% of identity).

According to fit best model assessment, substitution pattern and rates were estimated under the Kimura2-parameter model coupled to discrete +G (five categories, parameter value for rate variation among sites = 0.7725). Rates of different transitional substitutions were 9.36 and rates of different transversional substitutions were 7.82. The nucleotide frequencies were A=T/U=C=G=0.25. The maximum log

Table 2. Description of the accessions used (82 foreign isolates along with five Tunisian isolates) in this study according to the clustering arisen from the phylogeny inference.

Group	Accession	Isolate	Host	Country
PV-32	AY684271	Yunnan	Rose	China
	AY948441	RM-5	Rose	India
	DQ003584	I-23	Rose cv. 'Queen Elizabeth	Poland
	AY948440	RM-2	Rose	Chettalli, Karnataka, India
	DQ983498	143	Rose cv. 'Montezuma	Poland
	U57046	30/4	*Prunus persica* L. Batsch	USA
	/	Ghernghezel	Almond	Tunisia
	HQ833194	Pch-a	Peach	Hubei, China
	HQ833197	Nec-g	Nectarine	Hubei, China
	HQ833196	Nec-f	Nectarine	Hubei, China
	HQ833198	Fchr-r	Flower cherry	Shandong, China
	Y07568	PV32	Malus sp.	Spain
	AF170156	l/l3	Cherry var. Seneca	Holovousy, Czech Republic
	AF170158	PS 12/16	Cherry var. Stark Hardy Giant	Holovousy, Czech Republic
	AF170159	PS 14/22	Flowering cherry var. Shirofuga Kanzan	Holovousy, Czech Republic
	FJ610343	/	Cherry	Xi'an, China
	FJ610344	/	Rose	Haining, China
	JQ005049	Pch3	Peach	Canada
	JQ005048	Pch2	Peach	Canada
	EF565259	PchCl.sum1	Peach cv. Summer Lady	Chile
	JN416776	Pch12	Peach	Canada
	EF565261	PlmCl.bla1	Plum cv. Black Ambar	Chile
	EF565264	PchBr.unk1	Peach	Brazil
	EF565265	PchBr.unk2	Peach	Brazil
PE-5	FJ546090	PchMx.Azt	*Prunus persica* cv. Aztecade Oro	Mexico
	FJ546091	PchMx.Azt2	*Prunus persica* cv. Aztecade Oro	Mexico
	AF034994	CH30	Cherry	USA
	HQ833199	Pch-b	Peach	Hubei, China
	L38823	PE-5	*Prunus persica*	USA
	HQ833200	Fpch-j	Ornamental peach	Zhejiang, China
	EF565253	NctCl.aug1	Nectarine cv. August Fire	Chile
	AJ133207	PchIt.mry1	Peach, variety Marylin	Italy
	AJ133210	ChrIt.bla1	Cherry, variety Black giant	Italy

Table 2 (*contd*.)

Group	Accession	Isolate	Host	Country
SW6	EF565256	PchCl.loa2	Peach cv. Loadell	Chile
	AF013287	SW6	Sweet cherry	USA
	AM408909	/	*Prunus persica*	India
	AM494934	/	Plum	H.P., India
	AM408910	/	*Prunus persica* var. *nucipersica*	Kullu, India
	AM920668	/	Cherry	Palampur, India
PV-96	JQ005056	Pch10	Peach	Canada
	JQ005055	Pch9	Peach	Canada
	JQ005054	Pch8	Peach	Canada
	JQ005057	Pch11	Peach	Canada
	JQ005050	Pch4	Peach	Canada
	JQ005053	Pch7	Peach	Canada
	JQ005058	Pch13	Peach	Canada
	AF170171	Valticka	Peach	Czech Republic
	AF332616	NRSiz7/Kirka II/7	Plum	Poland
	AF170170	Na hrbu	*Prunus domestica* ssp. *insititia*	CeskeBudejovice, Czech Republic
	AF170168	Sss	Prune	CeskeBudejovice, Czech Republic
	AF170167	UH1	Prune	Holovousy, Czech Republic
	AF332615	NRSiz6/F12/1 237	Cherry	Poland
	AY037791	KU	Sour cherry	Slovakia
	AF170162	PS 7/12	Sour cherry var. MuskatnajaPrazskaja	Holovousy, Czech Republic
	AF170161	PS 7/11	Sour cherry var. MuskatnajaPrazskaja	Holovousy, Czech Republic
	AF170160	06-54	Prune var. Anna Spath	Holovousy, Czech Republic
	EF565262	PlmCl.fri1	Plum cv. Friar	Chile
	S78312	PV-96	*Prunus mahaleb*	Germany
	JQ005045	Chr17	Cherry	Canada
	JQ005046	Chr18	Cherry	Canada
	JQ005042	Chr14	Cherry	Canada
	JQ005038	Chr10	Cherry	Canada
	JQ005043	Chr15	Cherry	Canada
	HQ833191	Chr-p	Cherry	Shandong, China
	HQ833192	Chr-q	Cherry	Shandong, China
	AF332612	NRSiz1/dzik	Sweet cherry	Poland
	AF170164	7/20	Cherry var. Valesca	Czech Republic
	AJ133205	PchIt.may1	Peach, variety May crest	Italy
	AF170165	4/8	Cherry var. Precoce Nero Di Vignale	Holovousy, Czech Republic
	FJ231736	ChrT133	*Prunus avium* L.	USA
	FJ231734	ChrT54	*Prunus cerasus* L	USA
	HQ833193	Chr-m	Cherry	Henan, China
	JQ005047	Pch1	Peach	Canada
	EU368738	PNRSV-Mk	Sour cherry	Hungary
	EU368736	PNRSV-AL17	*Prunus dulcis* cv. Tardy nonPareil	Italy
Tunisian outgroup	/	Louzi	Almond	Tunisia
	/	Montrone	Almond	Tunisia
	/	Texas	Almond	Tunisia
	/	Mollese	Almond	Tunisia

likelihood for the computation was −4255.895. There were a total of 686 positions in the final dataset. In all these analyses, codon positions included were first + second + third + noncoding. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). Analyses were conducted using MEGA5.

Inference about selection

Seeking for determining the selective forces that might be operating during the evolutionary diversification of PNRSV, we used the following approaches: (i) Tajima's D, and Fu and Li's D and F statistical tests to examine whether the number of segregating sites in the sequences departs from the neutral expectation. The DnaSP software v5.10.01 allowed the calculation of Tajima's D as well as Fu and Li's D and F statistical tests to assess the neutrality and influence of demographic forces on the population which are as follows: Tajima's $D = -1.50551$ (significance at $P < 0.001$), Fu and Li's $D = -0.51802$ and Fu and Li's $F = -0.78076$ (no significance at $P > 0.1$) (table 3). Calculation was based on the total number of mutations. The significantly

Table 3. Population genetic parameters and neutrality tests calculated for the *CP* gene.

Population statistics							Test of neutrality+			Synonymous and nonsynonymous statistics after SNAP algorithm*		
M	*S*	ϑ	Π	*Hd*	Δ*Hd*	*K*	Tajima's statistics	Fu and Li's *F* statistics	Fu and Li's *D* statistics	*d*N	*d*S	*d*N/*d*S
5	626	0.73959	0.59402	1.00	0.016	407.500	−1.50551	−0.780716	−0.51802	1.5732	1.2434	1.2

M, number of sequences; *S*, number of segregating sites; θ, the statistic θ from the number of segregation sites (*S*) (Watterson θ estimator, 1975) and the average of synonymous and nonsynonymous substitutions, Π, nucleotide diversity (estimated by the average number of nucleotide differences between two random sequences in a population), *Hd*, haplotype diversity; Δ*Hd*, the variance of haplotype diversity; *K*, average number of pairwise nucleotide differences; +Tajima's *D* and Fu and Li's *D* and *F* tests measure the departure from neutrality for all mutation in CP cistron. *Average of all pairwise comparisons (http://www.hiv.lanl.gov/cgi-bin/SNAP/WEBSNAP/SNAP.cgi).

negative values of Tajima's *D*, and Fu and Li's *D* and *F* statistical tests for CP sequences discounted the neutral hypothesis suggesting the occurrence of purifying selection and demographic expansion of PNRSV population; (ii) the second approach consisted of evaluating of synonymous and nonsynonymous substitution rates per site by using SNAP software which, on the contrary, indicated an adaptive evolution (table 3). Further, the selection profiles on the CP gene were determined by submitting the sequence alignment to the HyPhy package which implements various models of evolution. We investigated the signature of selective pressure based on *d*N/*d*S ratio by applying the SLAC, FEL, IFEL, REL and FUBAR methods which incorporate not only nonsynonymous but also synonymous rate variation among codon sites explicitly. The SLAC method identified one site under negative selection at 0.1 significance level (codon 82 corresponding to amino acid residue His) (table 4a). The FEL approach detected two and seven positively and negatively selected sites, respectively at 0.1 significance level. The IFEL model found 13 and five adaptively and negatively selected sites, respectively at the same significance level. More sites were identified as positively selected with IFEL (13) than with FEL (two), suggesting that selective pressure was occurring primarily at the population level (internal branches) (table 4a). The REL method detected 21 and 55 codons under positive and purifying selection, respectively. The statistical significance of each value was evaluated by its Bayesian posterior probability ($P < 0.05$; table 4d). Recently, Murrell *et al*. (2012, 2013). developed two supplementary models (FUBAR and MEME). They stated that FUBAR is a much faster and statistically more robust than REL. In our case, this newly developed method identified one site under negative selection at posterior probability ≥ 0.9 (codon 82) and one positively selected site at posterior probability ≥ 0.9 (codon 207) (table 4b). Interestingly, SLAC, FEL, IFEL, and FUBAR models found a consensus on codon 82 which seemingly was under purifying selection. Similarly, both FEL and REL models found that codon 103 was under purifying selection too. On the other hand, MEME is a recommended method by the authors mentioned above to detect positive selection. In fact, they stated that it can find signatures of episodic selection even when the majority of lineages are subject to purifying selection. In our study, MEME found 13 sites with evidence of episodic diversifying selection at 0.1 significance level (table 4e). The PARRIS method demonstrated evidence of positive selection in aligned sequences as given by inferred distribution rates for the null (M1) and alternative models (M2) mentioned in table 4c. The GA branch was used to test the hypothesis that different selective environments were acting on the branches of the phylogeny. In estimating *d*N/*d*S, three rate classes were found to be supported by 46 models in the 95% confidence set. It is worth noting that the lineage Texas had the highest probability conferring the highest *d*N/*d*S ratio (*d*N/*d*S = 1.451 at $P = 100\%$), followed by Louzi (*d*N/*d*S = 1.151 at $P = 92.5\%$), and

Table 4. Positively and negatively selected sites in CP gene estimated by a) SLAC, FEL, and IFEL models; b) FUBAR model; c) PARRIS model; d) REL model; e) MEME model.

a

Selection	Model	Codon	Amino acid residue	Normalized *d*N-*d*S	*P* value
Negative	SLAC	82	His	−2.0067	0.074
Positive	FEL	190	Ala	0.29065	0.06099
		222	Leu	1.14227	0.07653
Negative		49	Gly	−0.32121	0.09042
		58	Pro	−0.43387	0.07096
		69	Pro	−2179.54	0.09617
		82	His	−2179.60	0.0087
		103	Thr	−0.3295	0.0972
		114	Leu	−0.8826	0.0840
		220	Gly	−1.7273	0.0668
Positive	IFEL	3	Ser	0.511845	0.0633
		4	Gln	1.37049	0.0623
		43	Pro	0.6689	0.0660
		47	Ser	0.58441	0.0363
		48	Ser	0.8685	0.0774
		129	Arg	3121.75	0.0386
		132	Gly	1.18164	0.0790
		157	Gly	3121.75	0.0531
		169	Pro	238.078	0.0201
		170	Arg	22.5912	0.022
		172	Ile	2.8246	0.009
		181	Ser	4.1558	0.063
		213	Asn	3121.75	0.066
Negative		69	Pro	−2179.79	0.055
		76	Pro	−0.64066	0.093
		82	His	−2179.79	0.010
		95	Tyr	−2179.79	0.079
		107	Leu	−1.5565	0.062

b

Selection	Model	Codon	Amino acid residue	α (synonymous)	β (nonsynonymous)	$\beta - \alpha$	Posterior probability
Negative	FUBAR	82	His	3.96568	0.138587	−3.82709	0.941112
Positive	FUBAR	207	Gly	0.60473	3.5516	2.94687	0.920789

c

Inferred rate distributions	Synonymous rate				ω (*d*N/*d*S) ratio			
	Rate class	1	2	3	Rate class	1	2	3
Null model (M1)	*d*S	0.11	1.78	3.53	ω	0.58	1.00	–
	Probability	0.672	0.133	0.195	Probability	0.802	0.198	–
Alternative model (M2)	*d*S	0.77	3.22	8.26	ω	0.61	1.00	8.95
	Probability	0.969	0.0	0.031	Probability	0.652	0.0	0.348

d

Selection	Model	Codon	Amino acid residue	Normalized *d*N-*d*S	Posterior probability	Bayes factors
Positive	REL	120	Ser	8.40186	0.9924	227.13
		122	Ser	8.23754	0.9737	64.228
		124	Leu	8.3526	0.9869	129.761
		137	Thr	8.23905	0.9741	64.661
		140	Gly	8.18357	0.9678	51.7508
		143	Leu	8.41226	0.9936	269.498
		144	Pro	8.2071	0.9705	56.5449
		145	Gly	8.33781	0.98527	114.872
		147	Met	8.39446	0.99166	204.209

Table 4 (*contd*.)

d

Selection	Model	Codon	Amino acid residue	Normalized *d*N-*d*S	Posterior probability	Bayes factors
		154	Val	8.25057	0.9754	68.1715
		155	Leu	8.21638	0.97157	58.6916
		166	Asn	8.41148	0.9935	265.812
		173	Ser	8.42801	0.99544	375.347
		180	Pro	8.37972	0.9899	169.983
		207	Gly	8.33204	0.9846	109.934
		210	Thr	8.3992	0.9922	218.561
		212	Asp	8.3209	0.9833	101.501
		214	Thr	8.20566	0.97036	56.226
		217	Val	8.42212	0.9947	327.354
		221	Arg	8.19681	0.96936	54.337
Negative		2	Val	−0.36719	0.99657	178.211
		3	Ser	−0.36602	0.99661	171.223
		5	Ile	−0.37022	0.99708	199.213
		7	Asn	−0.37729	0.99788	274.432
		8	His	−0.35952	0.9958	140.685
		10	His	−0.376546	0.9977	263.879
		12	Gly	−0.37050	0.99711	201.35
		14	Cys	−0.39878	0.9993	833.808
		15	Arg	−0.29591	0.98870	50.9587
		17	Cys	−0.38987	0.9993	833.808
		18	Lys	−0.39256	0.9996	1474.45
		19	Lys	−0.3819	0.9984	364.188
		20	Cys	−0.38987	0.9993	833.808
		21	His	−0.37991	0.9981	318.945
		22	Pro	−0.36058	0.9959	144.903
		23	Asn	−0.3820	0.9984	367.017
		24	Glu	−0.3820	0.9984	368.636
		25	Ala	−0.32559	0.9920	72.668
		26	Leu	−0.3752	0.9996	247.715
		27	Val	−0.34786	0.9945	106.519
		28	Pro	−0.3895	0.9992	808.107
		29	Leu	−0.34130	0.9938	93.6741
		30	Arg	−0.3758	0.9977	254.587
		31	Ala	−0.3255	0.9920	72.668
		32	Gln	−0.39319	0.9996	1796.57
		33	Gln	−0.38954	0.9992	791.072
		34	Arg	−0.3758	0.9977	254.587
		35	Ala	−0.325590	0.9920	72.668
Negative	REL	37	Asn	−0.38882	0.9991	711.952
		38	Asn	−0.3896	0.9992	800.157
		39	Pro	−0.36058	0.9959	144.903
		40	Asn	−0.38882	0.9991	711.952
		41	Arg	−0.36694	0.9967	176.705
		42	Asn	−0.3896	0.9992	800.157
		43	Pro	−0.36742	0.9967	179.644
		44	Asn	−0.38882	0.9991	711.952
		47	Ser	−0.38262	0.9984	383.55
		48	Ser	−0.3199	0.9914	67.221
		49	Gly	−0.3730	0.9974	223.631
		51	Gly	−0.3732	0.9974	225.876
		58	Pro	−0.3790	0.9980	303.411
		65	Thr	−0.35516	0.9953	125.51
		78	Gly	−0.34434	0.9941	99.220
		84	His	−0.38559	0.9995	544.213
		86	Glu	−0.39241	0.9995	1413.11
		90	Thr	−0.354790	0.9953	124.485
		91	Glu	−0.32201	0.9916	69.124
		93	Val	−0.31154	0.9904	60.4894

Table 4 (*contd*.)

d

Selection	Model	Codon	Amino acid residue	Normalized *d*N-*d*S	Posterior probability	Bayes factors
		103	Thr	−0.36579	0.9965	169.934
		104	Leu	−0.3589	0.9958	138.521
		105	Pro	−0.33484	0.9930	83.7351
		106	Gln	−0.3803	0.9982	328.106
		108	Met	−0.3490	0.9946	109.134
		196	Leu	−0.306447	0.98989	57.0152

e

Codon	α	β^-	Pr ($\beta = \beta^-$)	β^+	Pr ($\beta = \beta^+$)	*P* value	*Q* value
60	5.10^{-9}	0	0.60299	5.67409	0.39701	0.06633	1
62	5.10^{-9}	0	0.505579	2.08516	0.494421	0.097150	1
65	5.10^{-9}	0	0.747168	1.4724	0.252832	0.052088	1
71	5.10^{-9}	$3.75.10^{-16}$	0.587201	4.44734	0.412799	0.0620545	1
84	5.10^{-9}	0	0.771351	1.46554	0.228549	0.0614695	1
87	5.10^{-9}	$3.75.10^{-16}$	0.38098	0.895283	0.61902	0.086845	1
99	0	0	0.54346	1.94847	0.45654	0.0607456	1
129	5.10^{-9}	$3.75.10^{-16}$	0.369641	8.877306	0.530359	0.0669854	1
131	0.203762	0	0.442355	9.696	0.557645	0.0927061	1
139	0	0	0.350192	3.48554	0.549808	0.0694875	1
174	5.10^{-9}	0	0.625979	8.42081	0.374021	0.0137528	1
190	5.10^{-9}	5.10^{-9}	0.0834492	0.959865	0.916551	0.0728026	1
197	5.10^{-9}	6.10^{-15}	0.446369	26.07000	0.553631	0.0864072	1

The *P* value is derived using a mixture of χ^2 distribution and *q* values using Simes's procedure which controls the false discovery rate under the strict neutral null hypothesis (likely to be conservative).

Montrone, Ghernghezel, and Mollese isolates (dN/dS = 0.013 at $P = 0$).

Phylogenetic inferences

Phylogenetic analysis of the full-length CP gene sequence of PNRSV isolates comprising those collected from Tunisia (this study) and those extracted from GenBank showed that they were split into five distinct clusters. The resulting topologies of the trees reconstructed by both NJ and ML algorithms were similar. In addition to the three classical groups (PE-5, PV-32, and PV-96), two supplementary clades were delineated. The first one encompassed accessions reported from India (AM494934, AM408909, AM408910, and AM920668) and Chile (EF565255 and EF565256) as well as the isolate SW6 (AF013287) described long time before. The second which we called Tunisian outgroup, was formed by four Tunisian isolates (Texas, Montrone, Mollese, and Louzi) (figure 3). According to our results, these isolates were highly recombinant and also underwent selective pressure making them genetically different compared to all members of the other groups. Interestingly, Ghernghezel isolate behaved differently and was genetically close to representatives of PV-32 group. The tree topology had the support of the identity percentage determined at the nucleotide level. Indeed, Ghernghezel diverged from Texas (the lowest, 79%) to Mollese (the highest, 89%); meanwhile, with several members of PV-32 group, the identity rate reached 90%.

Discussion

Recombination is one of the major influences on genetic diversity in many biological entities. Understanding well the part it has to play is crucial to many population genetic analyses. Although all genetic variation is most importantly created through mutation, recombination can create new variants by combining types already present in the population. Despite the availability of a few reports dealt with recombination in the genome of PNRSV (Boulila 2009a, b), to date, the study of these important evolutionary forces is still limited and calls for a more extended and an in-depth description. Tempting to enhance the knowledge of these evolution traits potentially occurring in the CP gene of PNRSV, we used the genetic algorithm GARD to point out the existence of recombination breaking points. This method predicted that recombination could occur in aligned sequences. Since, GARD cannot provide supplementary information regarding the identity of the recombinants, further analyses using the Recco algorithm allowed

Figure 3. Dendrogram depicting phylogenetic relationships among Tunisian isolates compared to other isolates available in databases. Five clusters were delineated: three classical groups (PE-5, PV-32, and PV-96), a proposed new group baptized SW6 and a fifth group composed of Tunisian isolates underwent recombination and selection pressure baptized Tunisian outgroup. The tree was produced by the ML algorithm under assumption of the Model K2+G an option of MEGA5 software. Bootstrap analysis of 500 replicates was performed. The numbers above the branches indicate the bootstrap confidence value. Numbers above critical branches are significant bootstrap values (> 65%). The scale bar shows the number of substitution per nucleotide.

the identification of the potential recombinants which were all five Tunisian isolates studied here. Additionally, recombination means that a substitution pattern took place. In this respect, we estimated the residue substitution matrix under the Kimura2-parameter model coupled to discrete +G in the CP gene.

When the sequences under consideration have recombined, it is no longer the case that there is a single phylogeny describing their evolution. Instead, different phylogenies are required to describe the evolutionary relationships of the segments defined by recombination breakpoints. This feature was taken into account when dealt with selection inference. Nucleotide polymorphism and neutrality in PNRSV CP was evaluated by Tajima's *D* (Tajima 1989) and Fu and Li's *D* and *F* (Fu and Li 1993) statistical tests. They indicated that the sequences did not fit the neutral theory model at equilibrium between mutation and genetic drift. The negative values obtained from these tests signified that demographic expansion as well as a purifying selection occurred. Conversely, the SNAP software used in this study pointed out that PNRSV CP gene was under positive selection. Clearly, according to these sole results, a discrepancy arose. Therefore and undeniably, further in-depth analyses were rendered necessary. HyPhy package provides a broad array of tools for the analysis of genetic sequences using ML for estimating selection on site-by-site basis. In estimating site-by-site variation in rates, a counting method (SLAC), fixed-effects likelihood (FEL) and random-effects likelihood (REL) were used for site-by-site inference of selection and they provided a consistent divergence. In fact, although SLAC showed a sole codon (82) under negative selection, FEL and REL detected, on the one hand, two and 20 residues under positive and negative selection, respectively; and, on the other hand, seven and 55 positively and negatively selected sites, respectively. The IFEL method indicated the presence of 13 codons and five codons under positive and negative selection, respectively, suggesting a variation within the lineages. It is worth mentioning that FEL, SLAC, and IFEL methods found a consensus on codon 82 which was under negative selection in all these situations. Similarly, both FEL and REL models found that codon 103 was under purifying selection too. It should be noted that REL may suffer from false-positive rate for a small dataset as mentioned by the authors themselves (Murrell *et al*. 2013). As a consequence and alternatively, we decided to use FUBAR model which was recently made available in the datamonkey server and follow the recommendation of the authors who stated that is a more reliable tool than REL. Interestingly, FUBAR detected the codon 82 as negatively selected by analogy to the three methods mentioned earlier besides the detection of the codon 207 which was adaptively selected. Moreover, Murrell *et al*. (2012) proposed a new model called MEME which can detect signatures of episodic selections even when the majority of lineages are subject to negative selection. In this respect, our analysis showed that 13 codons had signatures of episodic diversifying selection. The PARRIS method found evidence for positive selection in aligned sequences by inferring the rate distribution for the null model M1 (nearly neutral) and the model M2 (positive selection). To test the hypothesis that different selective environments were operating on the branches of the phylogeny, we used GA branch algorithm which delineated three class rates. Moreover, using this algorithm, the topology of the reconstructed tree was in accordance with the general topology shown in figure 3. In evaluating the sign and intensity of selection that might be occurring in the CP of South American isolates of PNRSV, Fiore *et al.* (2008) found only 13 predicted sites under negative selection by using only REL model. Compared to these results, our study demonstrated that positive selection may also occur in the CP gene of PNRSV even by using only the REL method; thus, suggesting that environmental conditions including climate may interfere on the molecular adaptation of that virus.

On phylogeny, earlier studies (Boulila 2010; Fiore *et al.* 2008; Oliver *et al.* 2009) reported that many PNRSV isolates have been characterized and phylogenetically grouped into three subgroups named PV-32, PV-96 and PE5. Fiore *et al.* (2008) and Oliver *et al.* (2009) reported that phylogenetic analysis did not indicate a clear relationship between genetic variability of PNRSV isolates and their geographical origin. Similarly, no relationship was found between genetic variability and host genotype. Nevertheless, the majority of isolates from phylogroups PV-96 and PV-32 tend to exhibit latent/mild or chlorotic/necrotic symptoms, respectively. In the current study, we could segregate the analysed isolates into five different clusters. It was clearly pointed out that genetic events such as recombination and selection pressure may influence the genetic affiliation and relativity of any isolate predisposed to such events. Four Tunisian isolates (Montrone, Louzi, Mollese and Texas) constituted a distinct group (outgroup) from all other groups described so far precisely because they underwent recombination and selective pressure. Such events were favoured undeniably by environmental conditions namely virus–host interactions, climate and especially due to a probable co-infection. This is strongly plausible because the Tunisian isolates dealt with here were associated with almond mosaic syndrome which was reported to be as a disease complex (association with *Prune* dwarf virus) (Boulila 2009c). In contrast, Ghernghezel isolate was genetically similar to members of PV-32 group and supported by the fact that it was the lesser frequently recombining isolate as determined by RECCO. It worth mentioning that all members of the other groups were tested for recombination; it was found that no recombination events occurred.

Further, an additional group was delineated: the proposed new group which was phylogenetically separated from all other groups mentioned earlier. The clustering was made up of PNRSV isolates collected from India and Chile. They were included with the SW6 isolate (AF013287) described by Scott *et al.* (1998). Being the oldest of the group and similarly to PV-32, PV-96 and PE-5 clusters, we propose that SW6 as the head of this new group. It should be noted

that the Indian and Chilean PNRSV isolates including the American SW6 cherry strain were assigned by Oliver *et al.* (2009) to group CH 30. CH 30 is also a cherry strain whose CP gene was partially sequenced by Hammond and Crosslin (1998) and deposited in GenBank under the accession number, AF034994. Glasa *et al.* (2002) grouped the partially sequenced accessions CH 71, CH 30 and NRSiz5 under the appellation CH 30 and our phylogenetic investigations supported that statement and showed that CH 30 was clearly closely related to members of PE-5 rather than to the newly proposed SW6 group (figure 3). Likewise, Fiore *et al.* (2008) mentioned that SW6 belongs to PV-96 phylogroup. In contrast, our results demonstrated that is quite distant from all members forming PV-96 cluster (bootstrap support = 68%) (figure 3). Therefore, the assignation of the name CH 30 by Oliver *et al.* (2009) was inappropriate. Also, we do not find that SW6 cherry strain belongs to PV-96 phylogroup, contrary to the result of Fiore *et al.* (2008)

Besides recombination and adaptive evolution, earlier phylogenetic analyses showed that genetic exchange by reassortment of genomic segments may have an important role in virus evolution (Roossinck 1997). Reassortment has also been invoked as an important process to avoid the accumulation of deleterious mutation which could compensate in RNA viruses for the biological cost of segmented genomes (Chao 1988). Being a segmented virus, PNRSV can undergo reassortment to shape it genetic variation. To date, there is lack of knowledge about this mechanism of evolution and constitutes definitely a fertile area of research to be explored in PNRSV.

References

Akaike H. 1974 A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723.

Altschul S. F., Gish W., Miller W., Meyers E. W. and Lipman D. J. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Altschul S. F., Boguski M. S., Gish W. and Wooton J. C. 1994 Issues in searching molecular sequence databases. *Nat. Genet.* **6**, 119–129.

Bachman E. J., Scott S. W., Xin G. E. and Vance, V. B. 1994 The complete nucleotide sequence of *Prune dwarf virus* RNA3: implication for coat protein activation of genome replication in ilarviruses. *Virology* **201**, 127–131.

Boulila M. 2009a Molecular evidence for recombination in *Prunus necrotic ringspot virus*. *Plant Mol. Biol. Rep.* **27**, 189–198.

Boulila M. 2009b Recombination structure and genetic relatedness among members of the family *Bromoviridae* based on their RNAs 1 and 2 sequence analyses. *Virus Genes* **38**, 435–444.

Boulila M. 2009c Molecular characterization of an almond isolate of *Prune dwarf virus* in Tunisia: putative recombination breakpoints in the partial sequences of the coat protein-encoding gene in isolates from different geographic origin. *Phytopathol. Mediterr.* **48**, 411–421.

Boulila M. 2010 Putative recombination events and evolutionary history of five economically important viruses of fruit trees based on the coat protein-encoding gene sequence analysis. *Biochem. Genet.* **48**, 357–375.

Chao L. 1988 Evolution of sex in RNA viruses. *J. Theor. Biol.* **133**, 99–112.

Corpet F. 1988 Multiple sequence alignment with hierarchical clustering. *Nucleic. Acids Res.* **16**, 10881–10890.

Delport W., Poon A. F. Y., Frost S. D. W. and Kosakovsky Pond S. L. 2010 Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **29**, 2455–2457.

Di Terlizzi B., Skrzeczkowski L. J., Mink G. I., Scott S. W. and Zimmerman M. T. 2001 The RNA 5 of *Prunus necrotic ringspot virus* is a biologically inactive copy of the 3′-UTR of the genomic RNA 3. *Arch. Virol.* **146**, 825–833.

Fiore N., Fajardo T. V. M., Prodan S., Herranz M. C., Aparicio F., Montealegre J. *et al*. 2008 Genetic diversity of the movement and coat protein genes of South American isolates of *Prunus necrotic ringspot virus*. *Arch. Virol.* **153**, 909–919.

Fu Y. X. and Li W. H. 1993 Statistical tests of neutrality of mutations. *Genetics* **133**, 639–709.

Glasa M., Betinova E., Kudela O. and Subr Z. 2002 Biological of Prunus necrotic ringspot virus isolates and possible approaches to their phylogenetic and molecular characterisation typing. *Ann. Appl. Biol.* **140**, 279–283.

Guo D., Maiss E., Adam G. and Casper R. 1995 *Prunus necrotic ringspot* ilarvirus: nucleotide sequence of RNA 3 and the relationship to other ilarviruses based on coat protein comparison. *J. Gen. Virol.* **76**, 1073–1079.

Hammond R. W. and Crosslin J. M. 1995 The complete nucleotide sequence of RNA 3 of a peach isolate of *Prunus necrotic ringspot virus*. *Virology* **208**, 349–353.

Hammond R. W. and Crosslin J. M. 1998 Virulence and molecular polymorphism of Prunus necrotic ringspot virus isolates. *J. Gen. Virol*. **78**, 1815–1823.

Kimura M. 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120.

King A. M. Q., Adams M. J., Carstens E. B. and Lefkowitz E. J. 2011 Classification and nomenclature of viruses. In *Virus taxonomy*, pp. 1327. Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press.

Kishino H. and Hasegawa M. 1989 Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29**, 170–179.

Korber B. 2000 HIV signatures and similarities. In *Computational and evolutionary analysis of HIV molecular sequences* (ed. A. G. Rodrigo and G. H. Jr Learn), pp. 55–72. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Kosakovsky Pond S. L., Frost S. D. W. and Muse S. V. 2005 HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679.

Kosakovsky Pond S. L., Posada D., Gravenor M. B., Woelk C. H. and Frost S. D. 2006a GARD: a genetic algorithm for recombination detection. *Bioinformatics* **22**, 3096–3098.

Kosakovsky Pond S. L., Posada D., Gravenor M. B., Woelk C. H. and Frost S. D. 2006b Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* **23**, 1891–1901.

Larkin M. A., Blackshileds G., Brown N. P., Chenna R., McGettigan P. A., McWilliam H. *et al.* 2007 Clustal W and clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.

Maydt J. and Lengauer T. 2006 Recco: recombination analysis using cost optimization. *Bioinformatics* **22**, 1064–1071.

McDonald J. H. and Kreitman M. 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654.

Minck G. I. 1980 Identification of rugose mosaic-diseased cherry trees by enzyme-linked immunosorbent assay. *Plant Dis.* **64**, 691–694.

Minck G. I. 1991 *Prunus necrotic ringspot virus* plant diseases of international importance. In *Diseases of fruit crops*. (ed. J. Kumar, H. S. Chaube, U. S. Singh and A. N. Mukhopadhyay), vol. 3, pp. 335–356. Prentic Hall, Englwood Cliffs, USA.

Minck G. I. and Aichele M. D. 1984 Use of enzyme-linked immunosorbent assay results in efforts to control orchard spread of cherry rugose mosaic disease in Washington. *Plant Dis.* **68**, 207–210.

Murrell B., Wertheim J. O., Moola S., Weighill T., Scheffler K. and Kosakovsky Pond S. L. 2012 Detecting individual sites subject to episodic diversifying selection. *PLOs Genet.* **8** (doi: 10.1371/journal.pgen.1002764).

Murrell B., Moola S., Mabona A., Weighill T., Sheward D., Kosakovsky Pond S. L. and Scheffler K. 2013 FUBAR: A fast, unconstrained Bayesian approximation for inferring selection. *Mol. Biol. Evol*. (doi: 10.1093/molbev/mst030).

Nagy P. D. and Simon A. E. 1997 New insights into the mechanisms of RNA recombination. *Virology* **235**, 1–9.

Nemeth M. 1986 *Virus, mycoplasma and rickettsia diseases of fruit trees*, pp. 840. Martinus Nijhoff, The Hague, The Netherlands.

Oliver J. E., Freer J., Andersen R. L., Cox K. D., Robinson T. L. and Fuchs M. 2009 Genetic diversity of *Prunus necrotic ringspot virus* isolates within a cherry orchard in New York. *Plant Dis.* **93**, 599–606.

Pearson W. R. and Lipman D. J. 1988 Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.

Roossinck M. J. 1997 Mechanisms of plant virus evolution. *Ann. Rev. Phytopathol.* **35**, 191–209.

Rozas J., Sanchez-DelBarrio J. C., Messeguer X. and Rozas R. 2003 Dna*SP*, DNA polymorphism analyses by the coaslescent and other methods. *Bioinformatics* **19**, 2496–2497.

Rozen S. and Skaletsky H. J. 2000 Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics methods and protocols: methods in molecular biology*, pp. 365–386. (ed. S. Krawetz and S. Misener) Humana Press, Totowa, USA.

Rybicki E. P. 1995 *The Bromoviridae*. In *virus taxonomy, sixth report of the international committee on taxonomy of viruses*. (ed. F. A. Murphy, C. M. Fauquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli *et al.*), pp. 405–457. Springer-Verlag, Vienna, Austria.

Sambrook J., Fritsch E. F. and Maniatis T. 1989 Molecular cloning: A laboratory, 2nd edition. Cold Spring Harbor Laboratory Press. Cold Spring Harbor. New York, USA.

Sanger F., Nicklen S. and Coulson A. R. 1977 DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.

Scheffler K., Martin D. P. and Seoighe C. 2006 Robust inference of positive selection from recombining coding sequences. *Bioinformatics* **22**, 2493–2499.

Scott S. W. and Ge X. 1995 The complete nucleotide sequence of RNA 3 of *Citrus leaf rugose* and *Citrus variegation* ilarviruses. *J. Gen. Virol.* **76**, 957–963.

Scott S. W., Zimmerman M. T., Ge X. and Mackenzie D. J. 1998 The coat proteins and putative movement proteins of isolates of *Prunus necrotic ringspot virus* from different host species and geographic origins are extensively conserved. *Eur. J. Plant Pathol.* **104**, 155–161.

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

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. and Kumar S. 2011 MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.

Van Vloten-Doting L., Francki R. I. B., Fulton R. W., Kaper J. M. and Lane L. C. 1981 *Tricornaviridae*- a proposed family of plant viruses with tripartite single-stranded RNA genomes. *Intervirology* **15**, 198–203.

Watterson G. A. 1975 On the number of segregating sites in general models without recombination. *Theor. Popul. Biol.* **7**, 256–276.

Wetzel T., Candresse T., Macquaire G., Ravelonandro M. and Dunez J. 1992 A highly sensitive immunocapture polymerase chain reaction method for plum pox potyvirus detection. *J. Virol. Methods* **39**, 27–37.

Yang Z. and Bielawski J. P. 2000 Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–503.

Received 13 August 2012, in revised form 8 October 2012; accepted 10 October 2012
Published on the Web: 13 March 2013